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

Zhongjun Qu† Boston University

Denis Tkachenko National University of Singapore

December 19, 2012; July 10, 2013

Abstract

This paper presents a unified framework for analyzing local and global identification in log linearized DSGE models that encompasses both determinacy and indeterminacy. The analysis is conducted from a frequency domain perspective. First, for local identification, it presents necessary and sufficient conditions for the identification of the structural parameters along with the sunspot parameters, the identification of the former irrespective of the latter and the identification of the former conditional on the latter. The conditions apply to both singular and nonsingular systems and permit analysis using a subset of frequencies. Second, for global identification, the paper utilizes the Kullback-Leibler distance between two stationary vector linear processes and shows that global identification fails if and only if the minimized distance equals zero. Consequently, it delivers a set of parameter values that yield observational equivalence under identification failure. The global identification condition requires a nonsingular system but can be applied across models with different structures. Third, to understand the strength of identification, the paper proposes a measure of the empirical distance between two DSGE models. It gauges the possibility of distinguishing one model from another using likelihood ratio tests based on a finite number of observations generated by the models. Finally, although the theory is developed under a DSGE setup, it is applicable to other dynamic linear models with well defined spectra such as structural (factor augmented) vector autoregressive moving average models.

Key words: Dynamic stochastic general equilibrium models, global identification, multiple equilibria, spectral density.

*We thank Frank Kleibergen, Arthur Lewbel, Eric Renault and seminar participants at BC, Brown for helpful comments and suggestions.

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

1 Introduction

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

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

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

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

The results on local identification cover three situations. (1) The identification of both the structural and the sunspot parameters. This is motivated by the possibility of locally determining both the parameters describing technology and preferences and those governing the equilibrium beliefs of agents. (2) The identification of the structural parameters without making statements about the identification of the sunspot parameters. This shows whether it is possible to locally determine the parameters describing technology and preferences, even though the equilibrium beliefs may be unidentifiable. (3) The identification of the structural parameters conditional on the value of the sunspot parameters. This reveals whether it is possible to locally determine the parameters describing technology and preferences once we select a mechanism for equilibrium belief formation.

In each of the three situations, the paper provides necessary and sufficient conditions for identification. In practice, the conditions can be applied sequentially. If the first condition is violated, then the other two can be used to better understand the sources of identification failure. All the results allow arbitrary relationships between the number of observables and shocks, therefore are applicable to both singular and nonsingular DSGE systems. They permit latent state variables and measurement errors and are relatively straightforward to compute. In particular, no simulation is needed.

Next, the paper considers global parameter identification. Here the main challenges stem from: (1) the solutions are typically not representable as simultaneous equation systems with a finite number of predetermined regressors, and (2) the parameters interact in a highly nonlinear fashion. These two features make the results in Rothenberg (1971, Sections 4-6) inapplicable. Exploiting a different route, this paper derives identification conditions by further exploiting the general equilibrium feature and vector linear structure of the solutions. First, it obtains the Kullback–Leibler distance between two DSGE models. This distance depends only on the spectral densities and the dimension of the observables, thus can be computed without simulation or any reference to any data. Second, it shows that global identification fails if and only if the distance reaches zero when

minimized over the relevant parameter region. When that occurs, the condition delivers a set of parameter values generating observational equivalence. When identification holds, the minimized value is still informative about the strength of identification. This feature is further exploited in the paper to deliver a measure for the empirical closeness of models.

The local and global identification conditions can be compared along three dimensions. First and broadly, identification analysis can involve two types of questions with increasing levels of complexity. They are: (1) Parameter identification conditional on the model structure. In our context, this asks whether there exists a different parameter value within the same DSGE structure that can lead to the same dynamic properties. (2) Identification permitting different structures. In our context, this asks whether an alternatively specified DSGE structure (for example, a model with a different monetary policy rule or with different shock processes) can generate the same dynamic properties as some benchmark structure. (As a simplified example, for any non-invertible Gaussian MA(1) process there is an invertible model that is observationally equivalent.) The local identification condition, here and elsewhere in the literature, can only be used to address the first type of identification questions. In contrast, the global condition proposed here can also tackle the second question. This is the first and only condition in the DSGE literature with such a property. Second, the computational cost associated with the global identification condition can be substantially higher, especially when the dimension of the parameter vector is high. Third, the global identification condition requires the model to be nonsingular. Nevertheless, for singular systems, the condition can be applied to nonsingular subsystems as illustrated in the paper.

Even if global identification holds, there may exist parameter values delivering similar dynamic properties through the lens of finite sample sizes. This is known to be an important issue for DSGE models. For example, Del Negro and Schorfheide (2008) considered a New Keynesian DSGE model and observed quantitative similarity between a model with moderate price and trivial wage rigidity and one in which both rigidities are high. More generally, DSGE models that are different in specifications may bear similarity in quantitative implications. Such issues can often be analyzed fruitfully from a Bayesian perspective. Taking a frequentist approach, this paper suggests a measure for the empirical closeness of models. It is based on the power of the likelihood ratio test for the default against an alternative model specification using hypothetical samples of finite sizes generated by the two models. It has three features. (1) It is straightforward to compute for general DSGE models. The main computation cost is in solving the two models once and computing the spectral densities. (2) Its value always falls between 0 and 1 with a higher value indicating a

greater difference. (3) It is flat in the sample size if the Kullback–Leibler distance is zero and monotonically approaches one if the Kullback–Leibler distance is nonzero. The work here is related to Hansen (2007) who, working from a Bayesian decision making perspective, proposed to use Chernoff's (1952) information measure to quantify statistical challenges for distinguishing between two models.

As an illustration, the paper applies the results to a model considered in An and Schorfheide (2007). Previously, Qu and Tkachenko (2012) showed that the Taylor rule parameters are locally unidentified at a parameter value that yields determinacy. Here, we consider a range of different values under indeterminacy and find consistently that the Taylor rule parameters are locally identified. This finding confirms that indeterminacy offers richer dynamics than determinacy (see Lubik and Schorfheide, 2004). Such a distinction in parameter identification has previously only been documented for models with analytical solutions. The methods developed here allows such analysis in more sophisticated and empirically relevant models.

The paper also considers the model analyzed in Lubik and Schorfheide (2004) and Clarida, Galí and Gertler (1999). It finds the parameters are globally identified at the posterior means given in Lubik and Schorfheide (2004), both under determinacy and indeterminacy. When measuring the empirical distance between models with different monetary policy rules, we find there exists an expected inflation rule that leads to identical behavior dynamics as the current inflation rule. Meanwhile, there exists a output growth rule that leads to similar, though not identical, behavior dynamics as the output gap rule. Both findings hold under both determinacy and indeterminacy. The application to the medium scale model of Smets and Wouters (2007) is in progress.

The work substantially expands the early literature on identification in rational expectations models. There, notable contributions include Wallis (1980), Pesaran (1981) and Blanchard (1982). Wallis (1980, p.63) provided a rank condition for the local identification of the structural parameters in a model that features observable exogenous processes with lagged expectations (i.e., the structural equations at time t include only conditional expectations formed at $t-1$). These two features make the model effectively static. Pesaran (1981, Theorem 1) considered the same model and provided a necessary and sufficient condition for global identification of the structural parameters in a particular structural equation. He also considered models under indeterminacy and emphasized that parameter identification crucially depends on the model's implication for the conditional expectations. However, no formal identification conditions were provided for this case. Blanchard (1982) extended Wallis' (1980) analysis to first-order dynamic linear models. The above

analyses share a common feature. That is, the solutions are finite order static regressions or vector autoregressions. This makes identification conditions for standard simultaneous equations system applicable. Therefore, although their results shed some light on the problem, their methods are not applicable to general DSGE models.

The remainder of the paper is structured as follows. Section 2 discusses the model solution and its spectral density under indeterminacy. Sections 3 and 4 present results on local and global identification, respectively. Section 5 proposes a measure of empirical distance between DSGE models. Section 6 discusses the application of the results to (factor augmented) VARMA models. Section 7 provides some illustrative applications while Section 8 concludes. The paper has two appendices. The first contains computational details on solving the model under indeterminacy using Lubik and Schorfheide (2003). The second contains the proofs.

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

2 The spectrum of a DSGE model under indeterminacy

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

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

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

Under indeterminacy, the structural parameters alone do not uniquely determine the dynamics (i.e., the spectral density or the autocovariances) of the observables. This constitutes a conceptual challenge for analyzing identification. Next, we take three key steps to overcome this obstacle. First,

the result in Lubik and Schorfheide (2003) is applied to obtain a representation for the full set of solutions under indeterminacy. We also amend the procedure to allow the degree of indeterminacy to exceed one. Second, the parameter space is augmented to include the sunspot parameters in addition to the structural parameters. Third, utilizing the augmented parameter space, we compute the unique spectral density which fully characterizes the second order properties of the observables. This spectral density is easily computable, serving as the main instrument for the identification analysis in this paper.

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

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

or equivalently,

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

In Appendix A, we provide some computational details and also give an alternative derivation to this representation using an elementary result in matrix algebra (see (A.4)). The matrices Θ_1, Θ_ε and Θ_ϵ depend only on Γ_0, Γ_1, Ψ and Π, therefore are functions of the structural parameters alone. The dimension of Θ_ϵ, called the degree of indeterminacy, is also determined by the structural parameters. As in the literature, the vector ϵ_t is labeled as sunspot shocks.

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

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

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

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

Step 2. Parameter augmentation. Let θ^D denote the p-by-1 vector of structural parameters. Let

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

denote the q-by-1 vector of sunspot parameters. Define the augmented parameter vector

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

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

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

with

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

In the above notation, the observables are explicitly indexed by θ to signify that their dynamic properties will depend on both the structural and the sunspot parameters. As commented in Qu and Tkachenko (2012), the matrix $A(L)$ permits analyzing models with latent endogenous variables such as those in Smets and Wouters (2003, 2007). In these two models, S_t includes variables from a frictionless economy unobservable to the econometrician. Such variables can be excluded simply by assigning zero entries in $A(L)$.

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

The spectral density of $Y_t(\theta)$ is, using (2)-(4),

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

with

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

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

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

where $H(\cdot)$ is again defined by (4).

Remark 2 *The dimension of $f_\theta(\omega)$ is always n_Y-by-n_Y with n_Y being the dimension of $Y_t(\theta)$. In particular, its dimension does not depend on the number of shocks or whether there is indeterminacy.*

Komunjer and Ng (2011) and Qu and Tkachenko (2012) discussed challenges for studying identifications in DSGE models under determinacy. Similar challenges persist under indeterminacy. First, as stated in Remark 1, $Y_t(\theta)$ in (3) in general does not admit a finite order vector autoregressive representation. This implies that the conventional approach of first locating the identified finite dimensional reduced form parameters and then using them to back out the structural ones breaks down. Second, because this is a general equilibrium model, it precludes any observed exogenous processes. Therefore, the approach taken in the static context, in particular that of Wallis (1980, p.63), also breaks down. Finally, depending on the number of shocks in the model, the system can be singular. Singularity implies that the conventional information matrix is not well defined, therefore the rank condition in Rothenberg (1971) is not applicable.

Our identification analysis builds on Qu and Tkachenko (2012). The key idea is that, under indeterminacy, the solutions continue to be representable as a stationary vector linear process whose second order properties continue to be fully characterized by the spectral density matrix $f_\theta(\omega)$ over $\omega \in [-\pi, \pi]$. The elements of $f_\theta(\omega)$ can then be treated as mappings from the finite dimensional augmented parameter space to a space of complex valued functions defined over $[-\pi, \pi]$. Local or global identification holds if and only if the overall mapping is locally or globally injective. The results on identification of subsets of θ, such as θ^D, then follow as corollaries.

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

Assumption 2 *Assume $\theta^D \in \Theta^D \subset \mathbb{R}^q$ 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 is an interior point of Θ under indeterminacy. $\|H(\exp(-i\omega); \theta)\| \leq C < \infty$ for all $\theta \in \Theta$ and $\omega \in [-\pi, \pi]$.*

Note that for local identification, the compactness is not required. Also $\|H(\exp(-i\omega); \theta)\| \leq C$ only needs to hold in an open neighborhood of θ_0 over $\omega \in [-\pi, \pi]$.

3 Local identification

This section extends the results in Qu and Tkachenko (2012, Section 3) to allow for indeterminacy. We start with the identification of the augmented parameter vector θ. Intuitively, if θ_0 is locally identifiable, then it is potentially possible to locally determine both parameters describing technology and preferences (θ^D) and those governing the equilibrium beliefs of agents (θ^U).

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

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

Influenced by Koopmans (1949, p. 125), the identification analysis in econometrics is often embedded in a hierarchical or two step formulation. First, the joint distribution of the observations is written in terms of reduced form equations in which the parameters are always identified. 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, in DSGE models, the reduced form does not possess a natural definition and the parameters involved can be unidentifiable. This motivates us to adopt Haavelmo's (1944, p. 99) formulation to write the joint distribution directly in terms of the structural parameters.

More specifically, we treat the elements of $f_\theta(\omega)$ as mappings from the finite dimensional parameter space to complex valued functions defined over $[-\pi, \pi]$. The mappings associated with the elements of $f_\theta(\omega)$ are infinite dimensional and difficult to analyze directly. However, as in Qu and

Tkachenko (2012), for local identification it suffices to analyze a finite dimensional matrix based on $f_\theta(\omega)$. To state this precisely, we start with the following assumption.

Assumption 3 *Let*

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

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

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.*

Remark 3 *The dimension of $G(\theta)$ is always $(p+q)$-by-$(p+q)$. In particular, it is invariant to the number of equations, observables and shocks in the model. This feature is advantageous for analyzing identification in models with a high number of equations.*

If the regular point assumption in the theorem is dropped, the condition is still sufficient but not necessary for local identification. This assumption can be examined using the nonidentification curves proposed in Qu and Tkachenko (2012). First, consider the simple case where $G(\theta_0)$ has one zero eigenvalue. Then, compute a curve that solves the following differential equation in a neighborhood of θ_0

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

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

maintained assumption in the local identification literature, our paper appears to be the first that suggests a method to check its relevance in practice.

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

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

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

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

have the same rank.

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

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

Corollary 2 *Under Assumptions 1-3, θ^D is conditionally locally identifiable from the spectrum of $\{Y_t\}$ at a point θ_0 if and only if the following matrix is nonsingular:*

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

Finally, we highlight two extensions considered in Qu and Tkachenko (2012, Corollaries 1 and 5) that also hold under indeterminacy. We focus on the identification of θ_0, although partial

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

and conditional identification can be formulated analogously. The proofs are omitted. The first extension is about identification based on a subset of frequencies, e.g., those corresponding to business cycle fluctuations. Let $W(\omega)$ be an indicator function symmetric about zero to select such frequencies.

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

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

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

The second extension is about identification in nonsingular systems.

Assumption 4 *The eigenvalues of $f_\theta(\omega)$ are uniformly bounded away from zero for all ω in an open neighborhood of θ.*

Under the above assumption and normality, the information matrix at θ_0 is

$$I(\theta_0) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta} \right)^* \left(f_{\theta_0}^{-1}(\omega)' \otimes f_{\theta_0}^{-1}(\omega) \right) \frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'} d\omega. \tag{8}$$

Corollary 4 *Under Assumptions 1-4, θ is locally identified at θ_0 from the second order properties of $\{Y_t\}$ if and only if $I(\theta_0)$ is nonsingular.*

Therefore, Rothenberg's (1971) condition applies to DSGE models under indeterminacy when the system is nonsingular. The matrix $I(\theta_0)$ is invariant to the units of measurement of Y_t (θ). This feature is not shared by $G(\theta_0)$. Thus it can be preferable to $G(\theta_0)$ when applicable.

4 Global identification

We now consider global identification of θ at θ_0.

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

The criterion we propose is based on the Kullback-Leibler distance computed in the frequency domain. To motivate the idea, suppose that there is a realization from the DSGE model, denoted by $Y_t(\theta)$ $(t = 1, ..., T)$, see (3). Then, as $T \to \infty$, the Fourier transform at ω satisfies

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

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

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

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

$$KL(\theta_0, \theta_1) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\text{tr}\left(f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)\right) - \log\det\left(f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)\right) - n_Y\right\} d\omega. \tag{9}$$

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

The quantity $KL(\theta_0, \theta_1)$ plays an important role in the information theory literature. Pinsker (1964, p.198, Theorem 10.5.1) first obtained $KL(\theta_0, \theta_1)$ as the entropy rate of one vector stationary Gaussian process with respect to another (in our notation, of $Y_t(\theta_0)$ with respect to $Y_t(\theta_1)$). Later, using his result, Parzen (1983, p. 232, Theorem 2) showed that the autoregressive spectral density estimator satisfies the maximum entropy principle. Our paper appears to be the first that uses $KL(\theta_0, \theta_1)$ to study parameter identification in dynamic models.

Assumption 5 $\inf_{\theta\in\Theta} \inf_{\omega\in[-\pi,\pi]} \|f_\theta(\omega)\| > K$ *for some* $K > 0$.

Theorem 2 (Global identification) *Under Assumptions 1, 2 and 5,* θ *is globally identified from the second properties of* $\{Y_t\}$ *at* θ_0 *if and only if*

$$\inf_{\theta_1\in\Theta, \theta_1\neq\theta_0} KL(\theta_0, \theta_1) > 0.$$

[2]The above derivation is kept informal to ease understanding. A more formal derivation for $KL(\theta_0, \theta_1)$ proceeds as follows. Let $L(\theta)$ denote the frequency domain approximate likelihood based on $Y_t^d(\theta)$, see (11). Then, it can be shown that

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

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

$KL(\theta_0, \theta_1)$ is a deterministic function of θ. Its calculation is simple because $f_\theta(\omega)$ follows from (5) and the integral can be accurately approximated by an average. The main function of the theorem is to reduce the problem of global identification to minimizing a deterministic function. For the latter there is an array of methods available. The incurred computational cost greatly depends on the dimension of θ. If it is high, then the recently developed global optimization algorithms (such as parallel simulated annealing) can be quite useful.

In practice, the global condition can be applied jointly with the local conditions to deliver informative results. If the local conditions show that the parameters are locally identified, then $KL(\theta_0, \theta_1)$ can be computed over Θ excluding a small neighborhood of θ_0. More targeted searches can also be implemented depending on the issue of interest. For example, suppose that the focus is on global identification of the monetary policy rule parameters while fixing others. Then, one can search over the values of such parameters while keeping the rest fixed. In that case the computation may be feasible even with a grid search.

In the above analysis, the DSGE models differ only in terms of parameter values θ. In fact, the framework can permit them to have different structures. For example, they can have different monetary policy rules or shock processes. We call them different structures if they differ more than in just parameter values. More precisely, suppose $Y_t(\theta)$ and $Z_t(\phi)$ are two vector linear processes generated by two DSGE structures (Structures 1 and 2) with spectral densities $f_\theta(\omega)$ and $h_\phi(\omega)$, where $\theta \in \Theta, \phi \in \Phi$, and Θ and Φ are finite dimensional and compact. Suppose we fully embrace Structure 1 with $\theta = \theta_0$ and are interested in whether Structure 2 can generate the same dynamic properties.

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

Let

$$KL_{fh}(\theta, \phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \operatorname{tr}\left(h_\phi^{-1}(\omega) f_\theta(\omega) \right) - \log\det\left(h_\phi^{-1}(\omega) f_\theta(\omega) \right) - n_Y \right\} d\omega. \tag{10}$$

Corollary 5 *Let Assumptions 1, 2 and 5 hold for $Y_t(\theta)$, $Z_t(\phi)$ and their respective spectral densities. Then, Structure 2 is distinct from Structure 1 at $\theta = \theta_0$ if and only if*

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

In practice, different model structures may involve overlapping but different sets of observables. This is permitted by specifying matrices of lag polynomials $C_1(L)$ and $C_2(L)$ such that

$C_1(L)Y_t(\theta)$ and $C_2(L)Z_t(\phi)$ return the common observables. Then, the above corollary can be applied to them with $f_\theta(\omega)$ and $h_\phi(\omega)$ replaced by $C_1(\exp(-i\omega))f_\theta(\omega)C_1(\exp(-i\omega))^*$ and $C_2(\exp(-i\omega))h_\phi(\omega)C_2(\exp(-i\omega))^*$. Similarly, for singular DSGE systems, this technique can be used to apply Corollary 5 to nonsingualr subsystems.

Corollary 5 can be extended to consider a chosen subset of frequencies, say, the business cycle frequencies. Define

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

Then, under Assumptions 1, 2 and 5, Structure 2 is distinct from Structure 1 at $\theta=\theta_0$ within the frequencies selected by $W(\omega)$ if and only if $KL_{fh}^{W}(\theta_0,\phi)>0$ for all $\phi\in\Phi$.

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

5 Empirical distance between two models

Suppose we aim to measure the feasibility of distinguishing a model structure with spectral density $h_{\phi_0}(\omega)$ from a structure with $f_{\theta_0}(\omega)$ constrained by a hypothetical sample size T.

The frequency domain approximate Gaussian log likelihoods (up to a constant addition) are

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

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

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

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

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

The first summation on the right hand side converges to $-KL_{fh}(\theta_0, \phi_0)$. The second term, after multiplication by $T^{1/2}$, satisfies a central limit theorem under the null hypothesis. A similar decomposition can be used to analyze the log likelihood ratio under the alternative hypothesis.

Assumption 6 *The elements of $f_\theta(\omega)$ and $h_\phi(\omega)$ belong to the Lipschitz class of degree $\beta > 1/2$ with respect to ω.*[3] *Let $\widetilde{\varepsilon}_{ta}$ denote the a-th element of $\widetilde{\varepsilon}_t = (\varepsilon_t, \epsilon_t)$. Assume cum $(\widetilde{\varepsilon}_{ta}, \widetilde{\varepsilon}_{sb}, \widetilde{\varepsilon}_{uc}, \widetilde{\varepsilon}_{vd}) = 0$ for all $1 \le a, b, c, d \le q$ and $-\infty < t, s, u, v < \infty$.*

Theorem 3 *Let Assumptions 1, 2, 5 and 6 hold for both $f_\theta(\omega)$ and $h_\phi(\omega)$. Then, under the null hypothesis that $f_{\theta_0}(\omega)$ is the true spectral density,*

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

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

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

where $KL_{fh}(\theta_0, \phi_0)$ is given in (10), $KL_{hf}(\phi_0, \theta_0)$ is as in (10) but with h_ϕ and f_θ reversed,

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

The main contribution of the above result is to allow straightforward calculation of the approximate power of the likelihood ratio test $L_h(\phi_0) - L_f(\theta_0)$ of $f_{\theta_0}(\omega)$ against $h_{\phi_0}(\omega)$. To proceed, select a significance level α. The first result (the null limiting distribution) in the theorem implies that the critical value for the test is equal, up to $o(1)$, to

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

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

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

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

where $Z \sim N(0,1)$. We label $p_{fh}(\theta_0, \phi_0, \alpha, T)$ as the measure of the empirical distance of the model with spectral density $h_{\phi_0}(\omega)$ from the one with $f_{\theta_0}(\omega)$.

To obtain $p_{fh}(\theta_0, \phi_0, \alpha, T)$, 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 spectral densities. No simulation is needed. Other quantities in $p_{fh}(\theta_0, \phi_0, \alpha, T)$ refer only to the $N(0,1)$ distribution and are trivial to obtain.

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

A range of issues can be tackled using the measure. First, the values of θ and ϕ can be varied to measure the empirical distance between two families of models. For example, it is simple and informative to calculate, for some pre-specified sets S_1 and S_2,

$$\inf_{\theta \in S_1, \phi \in S_2} p_{fh}(\theta, \phi, \alpha, T)$$

and

$$\sup_{\theta \in S_1, \phi \in S_2} p_{fh}(\theta, \phi, \alpha, T).$$

Second, the measure can be treated as a function of T to illustrate to what extent distinguishing models become feasible as T increases. Third, the measure can be calculated over different subsets of frequencies to decompose to what extent the quantitative differences between models are driven by, for example, the business cycle frequencies as opposed to others.

6 Application to (factor augmented) VARMA models

Consider

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

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

Assume all the roots of $|A(z)| = 0$ lie outside of the unit circle. Then, Y_t has the following vector moving average representation

$$Y_t = H_1(L;\theta)\varepsilon_t + H_2(L;\theta)\zeta_t, \tag{12}$$

where

$$\begin{aligned} H_1(L;\theta) &= A(L)^{-1}B(L)\Sigma^{1/2}, \\ H_2(L;\theta) &= A(L)^{-1}\lambda(L)[I - \Gamma(L)L]^{-1}. \end{aligned}$$

The spectral density is

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

Some of the identification restrictions, such as the diagonality of Σ or factor loading restrictions on elements of $\lambda(L)$ can be directly incorporated into $H_1(L;\theta)$, other forms of restrictions, such as the long run restrictions, can be written as constraints on θ. Let

$$\psi(\theta) = 0$$

be the collection of such constraints one wishes to impose. Theorem 2 can then be used for checking global identification. There, global identification at θ_0 holds if and only if

$$\inf_{\theta_1 \in \Theta, \theta_1 \neq \theta_0, \psi(\theta_1) = 0} KL(\theta_0, \theta_1) > 0.$$

The procedures in Section 5 are also applicable. For example, we can contrast models with and without factor augmentation, models with different identification restrictions, or comparing a (factor augmented) VARMA with a DSGE model.

7 Illustrative examples

7.1 An and Schorfheide (2007)

Previously, Qu and Tkachenko (2012) documented lack of identification under determinacy in a model considered in An and Schorfheide (2007). Specifically, they showed that the parameters in

the Taylor rule are locally unidentified and obtained parameter values that yielded observational equivalence. This section conducts a further analysis under indeterminacy. It first examines local and global identification under indeterminacy. Next, it compares behavioral implications of alternative monetary policy rules. Then, it illustrates how the proposed measure of empirical distance can be used to gauge the feasibility of distinguishing between different model specifications.

The log linearized model is:

$$\begin{aligned} y_t &= E_t y_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau}(r_t - E_t \pi_{t+1} - E_t z_{t+1}), \\ \pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - g_t), \\ r_t &= \rho_r r_{t-1} + (1-\rho_r)\psi_1 \pi_t + (1-\rho_r)\psi_2 (y_t - g_t) + \varepsilon_{rt}, \\ g_t &= \rho_g g_{t-1} + \varepsilon_{gt}, \\ z_t &= \rho_z z_{t-1} + \varepsilon_{zt}, \end{aligned} \tag{13}$$

where y_t denotes output, π_t is inflation, r_t is the nominal interest rate, g_t is government spending, z_t is a disturbance in the growth rate of technology and $\varepsilon_{rt} \sim N(0, \sigma_r^2)$, $\varepsilon_{gt} \sim N(0, \sigma_g^2)$ and $\varepsilon_{zt} \sim N(0, \sigma_z^2)$ are serially and mutually uncorrelated structural shocks. The vector of structural parameters is

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

The vector of state variables is

$$S_t = (\pi_t, y_t, r_t, g_t, z_t, E_t \pi_{t+1}, E_t y_{t+1})'$$

and the vector of observables is

$$Y_t = (r_t, y_t, \pi_t)'.$$

Under indeterminacy, the sunspot shock (ϵ_t) is allowed to be correlated with the fundamental shocks:

$$\epsilon_t = M_{r\epsilon}\varepsilon_{rt} + M_{g\epsilon}\varepsilon_{gt} + M_{z\epsilon}\varepsilon_{zt} + \widetilde{\epsilon}_t,$$

where $\widetilde{\epsilon}_t$ is uncorrelated with the fundamental shocks with standard deviation σ_ϵ. Consequently, the sunspot parameter vector is

$$\theta^U = (M_{r\epsilon}, M_{g\epsilon}, M_{z\epsilon}, \sigma_\epsilon)'. \tag{14}$$

7.1.1 Local identification

Qu and Tkachenko (2012) considered local identification at the following parameter values taken from Table 3 in An and Schorfheide (2007) that are in the region of determinacy: $\theta^D = (2,$ 0.9975, 0.33, 1.5, 0.125, 0.75, 0.95, 0.9, 0.4, 3.6, $0.9)'$. They showed that $(\psi_1, \psi_2, \rho_r, \sigma_r)$ are locally unidentified. They obtained a nonidentification curve to depict the parameter combinations yielding identical behavior dynamics. Here we first expand their analysis to indeterminacy. Then, we revisit the determinacy region to further illustrate that the difference in the identification properties under determinacy and indeterminacy is not limited to a few points but is rather a generic feature.

In order to obtain empirically relevant values of θ under indeterminacy, the model is first estimated with the dataset from Smets and Wouters (2007) for the period 1960:I-1979:II (in the literature this period is often regarded as a regime with indeterminacy arising from passive monetary policy) using the frequency domain quasi-Bayesian procedure based on the second order properties suggested in Qu and Tkachenko (2012). The prior distributions are as in Table 1 in Lubik and Schorfheide (2004). When minimizing the log-posterior, β is reparameterized as $r^* = 400(1/\beta - 1)$ with r^* representing the annualized steady-state real interest rate. In a similar fashion, for the analysis under determinacy conducted later, the model is estimated using the post-1982 subsample of the same dataset. The results, together with the prior distributions, are reported in Table 1.

First, consider local identification at the posterior mean obtained using the pre-Volcker subsample (i.e., indeterminacy):

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

The smallest eigenvalue obtained after applying Theorem (1) equals 4.5E-05, while the Matlab default tolerance level is 6.8E-12, suggesting that θ_0 is locally identified. For further verification, we compute the curve (7) that corresponds to the smallest eigenvalue and compare the spectral densities with $f_{\theta_0}(\omega)$. The following three measures, suggested in Qu and Tkachenko (2012), are used to gauge the differences (Let $f_{\theta hl}(\omega)$ denote the (h, l)-th element of the spectral density matrix. Note that when computing the second measure, the denominator is evaluated at the same frequency

that maximizes the numerator.)

$$\begin{aligned} \text{Maximum absolute deviation:} \quad & \max_{\omega\in[0,\pi]} \mid f_{\theta hl}(\omega) - f_{\theta_0 hl}(\omega)\mid, \\ \text{Maximum absolute deviation in relative form} \quad : \quad & \frac{\max_{\omega\in[0,\pi]} \mid f_{\theta hl}(\omega) - f_{\theta_0 hl}(\omega)\mid}{\mid f_{\theta_0 hl}(\omega)\mid}, \\ \text{Maximum relative deviation:} \quad & \max_{\omega\in[0,\pi]} \frac{\mid f_{\theta hl}(\omega) - f_{\theta_0 hl}(\omega)\mid}{\mid f_{\theta_0 hl}(\omega)\mid}. \end{aligned} \tag{15}$$

The three measures take values 5.5E-04, 2.4E-04 and 0.01 respectively when $\|\theta - \theta_0\|$ reaches 0.045. This result confirms that θ_0 is indeed locally identified.

Next, to further the analysis, we draw parameter values from the posterior distribution in Table 1 and check the identification at each point. The following bounds on parameter values are imposed:

$$\begin{aligned} \tau &\in [0.01, 10], \beta \in [0.9, 0.9999], \kappa \in [0.01, 5], \psi_1 \in [0.01, 0.9], \psi_2 \in [0.01, 5], \rho_r \in [0.1, 0.99], \\ \rho_g &\in [0.1, 0.99], \rho_z \in [0.1, 0.99], \sigma_r \in [0.001, 3], \sigma_g \in [0.001, 3], \sigma_z \in [0.001, 3], M_{r\epsilon} \in [-3, 3], \\ M_{g\epsilon} &\in [-3, 3], M_{z\epsilon} \in [-3, 3], \sigma_\epsilon \in [0.001, 3]. \end{aligned}$$

Out of the 4000 draws, the smallest eigenvalue are above the default tolerance levels for all but two cases, all with ρ_r close to the boundary value 0.99. For these two points, the curves (7) corresponding to the smallest eigenvalue reveal that they are in fact also locally identified. Specifically, the maximum relative deviations along the curves exceeds E-02 when $\|\theta - \theta_0\|$ reaches 0.010 and 0.046. The results show that the above local identification property is not confined to a particular point but rather a generic feature of this model under indeterminacy.

The above local identification property is in sharp contrast with that under determinacy. Specifically, the posterior mean obtained using the post-1982 subsample equals

$$\theta_0^D = (2.04, 0.995, 0.93, 2.12, 0.25, 0.65, 0.93, 0.98, 0.22, 0.69, 0.13)'. \tag{16}$$

At this value, the smallest eigenvalue of $G(\theta_0^D)$ equals 5.1E-13 with the corresponding default tolerance level being 1.6E-10. This suggests that θ_0^D is locally unidentified. Because there is one zero eigenvalue, there is one subset of parameters causing the identification failure. They correspond to the parameters in the Taylor rule:

$$(\psi_1, \psi_2, \rho_r, \sigma_r).$$

This finding is in line with the result in Qu and Tkachenko (2012), who considered identification properties at a set of parameter values from Table 3 in An and Schorfheide (2007). To gain further

insight into whether this identification feature pertaining to the monetary policy rule is a common feature of the determinacy region, we draw parameter values from the posterior distribution in Table 1 and apply Theorem 1 to each point. The same bounds for parameter values are used except $\psi_1 \in [1.1, 5]$ and the elements in θ^U are no longer present. Out of the 4000 draws, there are all but 2 cases with the smallest eigenvalue below the default tolerance level. At those 2 points, the eigenvalues are very small, both being of order E-11, and barely exceed the tolerance levels. The values of the three measures along the curves (7) remain negligible in both cases, the largest still being of order E-05 after $\left\|\theta^D - \theta_0^D\right\|$ reaches 1. This shows that the two parameter values are also locally unidentified. In addition, in all cases considered, the lack of identification is caused by the four parameters in the Taylor rule.

Such a feature that some parameters are identified under indeterminacy but not under determinacy has been documented in the literature (see, e.g., Beyer and Farmer (2004)). However, such a property has only been shown to exist in simple models that can be solved analytically. Our paper is the first to show that this phenomenon occurs in models that are of empirical relevance.

7.1.2 Global identification under indeterminacy

We now consider global identification properties of θ at θ_0. Specifically, we search for models closest to θ_0 in terms of the KL criterion over the parameter space excluding a small neighborhood of θ_0, that is, over

$$\{\theta : |\theta - \theta_0|_\infty \geq c\}, \tag{17}$$

where $|.|_\infty$ returns the maximum absolute difference between the elements of θ and θ_0 and c can be set to different values to exclude neighborhoods of different sizes. We consider $c = (0.1, 0.5, 1)$. Although the results are for (17), other definitions of neighborhoods and c can also be considered, for example (Euclidean distance) $\{\theta : \|\theta - \theta_0\| \geq c\}$, or (percentage difference) $\{\theta : |(\theta - \theta_0)./\theta_0|_\infty \geq c\}$. The experimentation with the latter two norms (omitted here) yielded qualitatively similar results.

Case 1: c=0.1. The parameter vector minimizing the KL criterion is given by

$$\theta_{0.1} = (\underbrace{\mathbf{2.61}, 0.995, 0.49, \mathbf{0.64}, \mathbf{0.20}, 0.87, 0.66, 0.60, 0.27, \mathbf{0.57}, \mathbf{0.63}}_{\theta^D}, \underbrace{\mathbf{0.54}, -0.06, 0.26, 0.19}_{\theta^U})',$$

where, here and below, bolded elements are the ones that differ from θ_0 within the second decimal place. The resulting KL criterion equals 5.36E-07. As this appears to be a small value,

we compare the spectral densities of the two models across frequencies from 0 to π and compute $p_{ff}(\theta_0, \theta_{0.1}, 0.05, T)$ for $T = [80, 150, 200, 1000]$ to gain more insight into the nature of identification. Tables 2 and 3 report the deviations of the spectral density at $\theta_{0.1}$ from that at θ_0 and the empirical distance measures. The differences between spectra are small, with the largest relative deviations across all frequencies lying between 0.14% and 2.15%. Note that although these values appear small, they are much larger than the deviations typically observed along the nonidentification curves when lack of local identification is truly present. For example, in the case of points on the non identification curve obtained under determinacy reported above, the deviations are of order E-05 after $\left\|\theta^D - \theta_0^D\right\|$ reaches 1. The empirical distance measures in Table 3 further confirm that the two models are not observationally equivalent and can in principle be differentiated between based on their spectra. But it also shows that it will be very hard to do so with sample sizes typically available for macroeconomic data, as the distance is only 0.051 for $T = 80$ and increases just to 0.053 when $T = 1000$. Overall, the findings imply: (1) θ is globally identified at θ_0, (2) the model with $\theta = \theta_{0.1}$ is difficult to distinguish from $\theta = \theta_0$ in practice, and (3) the model minimizing the KL criterion is obtained by mainly shifting τ, and small adjustments in ψ_1, ψ_2, σ_g and σ_z, while the only sunspot parameter that noticeably changes is $M_{r\epsilon}$.

Case 2: c=0.5. The following point is closest in terms of the KL criterion to the model at θ_0

$$\theta_{0.5} = (\underbrace{\mathbf{3.01}, \mathbf{0.996}, 0.49, \mathbf{0.68}, \mathbf{0.06}, \mathbf{0.86}, 0.66, \mathbf{0.58}, 0.27, \mathbf{0.57}, \mathbf{0.65}}_{\theta^D}, \underbrace{\mathbf{0.57}, \mathbf{-0.08}, \mathbf{0.27}, 0.19}_{\theta^U})'.$$

The KL criterion equals 1.57E-05. Tables 2 and 3 report the spectral deviations and empirical distance measures. The maximum relative deviations are between 0.64% and 11.55%. The empirical distance measures increase only slightly compared to the case with $c = 0.1$, ranging between 0.055 at $T = 80$ and 0.071 at $T = 1000$, suggesting that it is still hard to distinguish the models based on reasonable sample sizes. Again, the most notable change is observed for τ, which increases by 0.5 compared to its value at θ_0. The parameters ψ_1, σ_z, $M_{r\epsilon}$ continue increasing, while ψ_2 further decreases. Additionally, more sunspot parameters change, as $M_{g\epsilon}$ goes down and $M_{z\epsilon}$ increases slightly.

Case 3: c=1. The minimum KL criterion is achieved at

$$\theta_{1.0} = (\underbrace{\mathbf{3.51}, \mathbf{0.994}, \mathbf{0.52}, \mathbf{0.70}, \mathbf{0.01}, \mathbf{0.85}, 0.66, \mathbf{0.57}, 0.27, \mathbf{0.56}, \mathbf{0.69}}_{\theta^D}, \underbrace{\mathbf{0.60}, \mathbf{-0.10}, \mathbf{0.28}, \mathbf{0.18}}_{\theta^U})'.$$

The distance equals 7.57E-05. As expected, its value gradually increases as a progressively larger neighborhood is excluded from the search. Differences between model spectra in Table 2 are higher than in previous cases for most elements, the largest relative deviation being 22.39%. The increase in empirical distances for the sample sizes considered is more moderate. For $T = 80$ the measure equals 0.062, and gradually climbs to 0.104 at $T = 1000$. Therefore, it should still be hard to differentiate between the two models in empirical work. As before, the largest change in the parameter vector is the unit increase in τ from its initial θ_0 value. Most of the remaining parameter also change, albeit in much smaller magnitudes, to maintain the closest KL criterion to the model at θ_0.

In summary, no observationally equivalent model to that at θ_0 is found when some neighborhoods around it are excluded from the search. Still, the models generated by $\theta_{0.1}$, $\theta_{0.5}$ and $\theta_{1.0}$ produce dynamics quite similar to those at θ_0 and will be difficult to differentiate from the latter from second order properties in typical sample sizes.

The analysis shows that an important channel for near observational equivalence is the weak identification of τ. Then does the identification improve substantially when τ is kept fixed? The corresponding results will be included in a later version.

7.1.3 Identification of policy rules

We now consider the KL criterion between the model at θ_0 and a model with the same structure, except that the central bank responds to expected inflation rather than current inflation:

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_1 E_t \pi_{t+1} + (1-\rho_r)\psi_2 (y_t - g_t) + \varepsilon_{rt}. \tag{18}$$

The parameter vector in this model is denoted by ϕ.

Starting with the case of indeterminacy, $\inf_{\phi\in\Phi} KL_{fh}(\theta_0, \phi)$ is reached at

$$\phi = (\underbrace{2.51, 0.995, 0.49, 0.63, \mathbf{0.54}, 0.87, 0.66, 0.60, 0.27, 0.58, 0.62}_{\phi^D}, \underbrace{0.53, -0.06, 0.26, 0.19}_{\phi^U})', \tag{19}$$

where, here and below, the bold fonts denote the parameter values that differ from θ_0 within the second decimal place. The minimized value KL criterion equals

3.35E-14.

To verify whether the two models are indeed observationally equivalent, we look at the three measures of discrepancies between $f(\theta_0)$ and $h(\phi)$, using (15) with $h_{\phi hl}(\omega)$ replacing $f_{\theta hl}(\omega)$. The

three measures equal 4.32E-06, 4.21E-06 and 4.21E-06 respectively. The empirical distance measure stays at 5% when the sample size is increased to 1000. These results, summarized in Tables 4 and 5, show that the expected inflation rule coupled with parameter values in (19) yields identical behavior dynamics as the current inflation rule under θ_0**.** It is impossible to distinguish between them even with an infinite sample size.

We now consider whether such equivalence also arises under determinacy at $\theta^D = \theta_0^D$. The search for an equivalent model yields

$$\phi^D = (2.04, 0.995, 0.93, \mathbf{2.11}, \mathbf{2.23}, 0.65, 0.93, 0.98, 0.22, 0.69, 0.13)'.$$

The minimized KL criterion equals

6.11E-14.

The three measures for spectra deviations from Table 4 are mostly of order E-06 and smaller, with only three elements having deviations of order E-05. Again, the empirical distance measure stays at 5% when the sample size is increased to 1000**.** Thus, policy equivalence also arises under determinacy.

Next, we examine an alternative model whose monetary rule reacts to output growth instead of output gap. Such a rule is also considered in An and Schorfheide (2007):

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

As before, we conduct the search under indeterminacy with the benchmark model given by θ_0 and under determinacy with the base model at θ_0^D.

Under indeterminacy, $\inf_{\phi\in\Phi} KL_{fh}(\theta_0, \phi)$ is achieved at

$$\phi = (\underbrace{\mathbf{2.80}, \mathbf{0.997}, \mathbf{0.48}, 0.\mathbf{69}, \mathbf{0.01}, \mathbf{0.86}, 0.66, \mathbf{0.59}, 0.27, \mathbf{0.57}, \mathbf{0.63}}_{\phi^D}, \underbrace{\mathbf{0.55}, -\mathbf{0.07}, \mathbf{0.27}, 0.19}_{\phi^U})'.$$

The KL criterion between the two models equals

1.94E-05.

Table 6 shows that the largest absolute deviation between spectra equals 0.066, and also happens to be the largest in relative terms at 0.115. It should be noted that for the rest of the elements relative deviations are of orders E-03 and E-04. Table 7 reports the empirical distance measures for different sample sizes. The measure equals 5.6% when the sample size is 80, and slowly increases to 7.4%

when the sample size is 1000. These results suggest that the two models under these two monetary policy rules are not observationally equivalent. However, it is still very hard to distinguish between them in typical sample sizes. Under determinacy, the $\inf_{\phi\in\Phi} KL_{fh}(\theta_0^D, \phi^D)$ is reached at

$$\phi^D = (\mathbf{1.97}, \mathbf{0.999}, \mathbf{0.90}, \mathbf{2.12}, \mathbf{0.01}, \mathbf{0.63}, 0.93, 0.98, 0.22, 0.69, 0.13)'.$$

The resulting distance equals

4.60E-05.

The largest values of the three measures of deviations equal 0.061, 0.022 and 0.117 respectively. The empirical distance measure equals 6.0% and 9.0% for sample sizes of 80 and 1000. Overall, the results are similar to the indeterminacy case.

7.1.4 Empirical distances

This subsection includes some exercises to further illustrate the informativeness of the empirical distance measure by considering a range of models different from (13)-(16). Throughout the experimentation, (13)-(16) is considered as the default specification and the significance level is set to 5%. The results are reported in Table 8.

First, consider a situation where it is known that the alternative model is difficult to distinguish from the default even with a large sample size. Such an alternative model can be constructed by assigning $(\psi_1, \psi_2, \rho_r, \sigma_r)$ values close to the nonidentification curve obtained under determinacy, and keeping the rest fixed at θ_0^D. To this end, we take a parameter vector from the nonidentification curve computed previously for θ_0^D within the analysis in the Subsection 7.1.1, such that $\left\|\theta^D - \theta_0^D\right\| = 1$, but truncate the values of the nonidentified parameters to leave 2 decimal places: $(2.07, 1.24, 0.69, 0.24)$. As expected, the KL criterion between the two models is very small (9.54E-05), and the value of the empirical distance measure is 0.061 with 80 observations. With 150 and 200 observations, the distance measure values are still low at 0.067 and 0.070 respectively, and even with 1000 observations it increases only moderately to 0.109. This exercise is interesting as it provides an idea for the magnitude of the values of $p_{fh}(\theta_0, \phi_0, 0.05, T)$ that one could expect to find in practice when the models are close to being observationally equivalent.

Second, consider the case where all of the parameter values in θ^D are the same as in θ_0^D except for the discount factor β, which is now lowered to 0.9852, implying a change in the discount rate from 2% to 6% on an annual basis. The values of the distance measure are 0.054 for samples of 80, 0.055 for samples of 150, 0.056 for samples of 200, and 0.067 for samples of 1000, which are quite

similar to the previous case. This confirms the empirical fact that it is hard to estimate β to any precision using the dynamic properties of aggregate data on consumption and interest rates.

Third, consider changing the Taylor rule weights ψ_1 and ψ_2 in θ_0^D to 1.5 and 0.5 respectively, while keeping the other parameters fixed at their original values. The results suggest that it is feasible to differentiate between the two models with commonly used sample sizes. With only 80 observations, the empirical distance measure is already markedly higher than in the previous cases with a value of 0.950, and it further improves to 0.998 when the sample size reaches 200.

Finally, it is interesting to study the empirical distance between θ_0 and θ_0^D, and, more generally, between θ_0 and all the permissible values in the determinacy region. Such results will be reported in a later version.

7.2 Lubik and Schorfheide (2004)

This model has a similar structure to that of An and Schorfheide (2007). However, as seen below, the identification properties can be different. The analysis below parallels that of An and Schorfheide (2007) for a better illustration of the similarity and differences between the two models**.** The log linearized model is

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

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

The vector of structural parameters is

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

The vector of state variables is

$$S_t = (\pi_t, y_t, r_t, g_t, z_t, E_t\pi_{t+1}, E_t y_{t+1})'$$

and the vector of observables is

$$Y_t = (r_t, y_t, \pi_t)'.$$

Lubik and Schorfheide (2004) uses the following representation for solutions under indeterminacy:

$$S_t = \widetilde{\Theta}_1 S_{t-1} + \widetilde{\Theta}_\varepsilon \varepsilon_t + \widetilde{\Theta}_\epsilon \epsilon_t,$$

where $\widetilde{\Theta}_1$ and $\widetilde{\Theta}_\epsilon$ have the same definition as Θ_1 and Θ_ϵ in this paper's computational appendix and

$$\widetilde{\Theta}_\varepsilon = \Theta_\varepsilon + \Theta_\epsilon \left(\Theta_\epsilon' \Theta_\epsilon\right)^{-1} \Theta_\epsilon' \left(\Theta_\varepsilon^b - \Theta_\varepsilon\right), \tag{20}$$

where again Θ_ϵ and Θ_ε have the same definition as in the computational appendix and Θ_ε^b is the counterpart of Θ_ε when the ψ_1 element in θ^D replaced by a value on the boundary of the determinacy and indeterminacy

$$\widetilde{\psi}_1 = 1 - \frac{\beta \psi_2}{\kappa}\left(\frac{1}{\beta} - 1\right).$$

Finally, the sunspot shock ϵ_t and the sunspot parameter θ^U are specified in the same way as (14).

7.3 Local identification

First, consider local identification at the posterior mean obtained in Lubik and Schorfheide (2004) using pre-1980 observations:

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

Applying Theorem 1, the smallest eigenvalue equals 6.2E-04. The Matlab default tolerance level is 1.9E-09, suggesting that all parameters are locally identified at θ_0. To confirm identification, we compute the spectra deviations along the curve (7) corresponding to the smallest eigenvalue. The absolute and relative deviations exceed E-03 after $\|\theta - \theta_0\|$ reaches 0.06 and 0.08 respectively, indicating that the posterior mean is indeed locally identified.

Next, we draw parameter values from the posterior distribution of Lubik and Schorfheide (2004). The following parameter bounds are imposed:

$$\begin{aligned}
\tau &\in [0.1, 1], \beta \in [0.9, 0.9999], \kappa \in [0.01, 5], \psi_1 \in [0.01, 0.9], \psi_2 \in [0.01, 5], \rho_r \in [0.1, 0.99], \\
\rho_g &\in [0.1, 0.99], \rho_z \in [0.1, 0.99], \rho_{gz} \in [-0.9, 0.9], \sigma_r \in [0.001, 3], \sigma_g \in [0.001, 3], \sigma_z \in [0.001, 3], \\
M_{r\epsilon} &\in [-3, 3], M_{g\epsilon} \in [-3, 3], M_{z\epsilon} \in [-3, 3], \sigma_\epsilon \in [0.001, 3].
\end{aligned}$$

Out of 4000 draws, the smallest eigenvalues are higher than the default tolerance level for all but 4 cases. For the 4 points, the deviation measures increase quickly along the curve (7), with absolute deviations exceeding E-03 after $\|\theta - \theta_0\|$ reaches 0.08, and relative deviations reaching E-03 when $\|\theta - \theta_0\|$ reaches 0.45. These results indicate that the points are locally identified[4]. Therefore, the local identification property is not confined to the posterior mean θ_0, but rather is a more generic feature.

Then, consider local identification properties under determinacy. Start with the posterior mean in Lubik and Schorfheide (2004) obtained using post-1982 observations:

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

Applying Theorem 1, the smallest eigenvalue equal 2.5E-06. The Matlab default tolerance level is 1.1E-11, suggesting that all parameters are locally identified at θ_0^D. Consistently, the relative deviations increase quickly reaching E-03 when $\|\theta - \theta_0\|$ reaches 0.055, confirming local identification**.** We also draw parameter values from the posterior distribution of Lubik and Schorfheide (2004). Theorem 1 is then applied to all the resulting values. Out of 4000 draws, 3 points have their smallest eigenvalues below the default tolerance level. The corresponding eigenvectors consistently point to β. After either fixing it or including the first order properties into the analysis, the eigenvalues become clearly above the tolerance level. Therefore, like in the case of the indeterminacy region discussed above, the local identification property of θ^D is not confined to just a few points in the parameter space.

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

7.4 Global identification

This subsection considers global identification properties under both determinacy and indeterminacy. Start with indeterminacy.

Case 1: c=0.1. Searching over the parameter subspace given by (17) with $c = 0.1$ yields

$$\theta_{0.1} = (\underbrace{0.69, \mathbf{0.998}, 0.77, 0.77, 0.17, 0.60, 0.68, 0.82, 0.23, 0.27, 1.13, \mathbf{0.15}}_{\theta^D}, \underbrace{\mathbf{-0.69}, \mathbf{1.84}, \mathbf{-0.70}, \mathbf{0.13}}_{\theta^U})'.$$

[4]The complete results will be made availabe in the online supplement.

The KL criterion between the two models is 4.06E-07. Table 9 reports the three measures of deviations between the two spectral densities. The values are small, with the largest relative differences between 0.04% and 0.11% across all elements. The empirical distance measures reported in Table 10 are consistent with these findings for all sample sizes considered, starting from 0.0507 for $T = 80$ and increasing only to 0.0529 for $T = 1000$. As in the global identification analysis for the An and Schorfheide's (2007) model, although the deviations are small, they are much larger than the deviations typically observed along the nonidentification curves when lack of local identification is present. It is notable that here $\theta_{0.1}$ differs from θ_0 mainly in the sunspot parameters, with the largest change in $M_{g\epsilon}$ that makes the neighborhood constraint binding, accompanied by small increases in the values of β and ρ_{gz}.

Case 2: c=0.5 The minimization of the KL criterion yields

$$\theta_{0.5} = (\underbrace{0.69, \mathbf{0.999}, 0.77, 0.77, \mathbf{0.16}, 0.60, 0.68, 0.82, 0.23, \mathbf{0.29}, 1.13, \mathbf{0.11}}_{\theta^D}, \underbrace{\mathbf{-0.64}, \mathbf{1.24}, \mathbf{-0.65}, \mathbf{0.39}}_{\theta^U})'.$$

The corresponding KL criterion is 1.22E-05. Tables 9 and 10 report the deviations between the spectra of the two models and the empirical distance measures. As a reflection of the larger KL criterion, the discrepancies between spectral densities are greater compared to the previous case, with relative deviations lying between 0.22% and 0.87%. However, the empirical distance measures reported in Table 10 show that the two models are still difficult to distinguish in typical sample sizes. For $T = 80$, the measure equals 0.055 and grows only to 0.069 when $T = 1000$. The parameters that are different from θ_0 are the same as in the previous case, adding an increase in σ_g and a decrease in ψ_2, with the largest change again observed for $M_{g\epsilon}$.

Case 3: c=1. The parameter vector minimizing the KL criterion is

$$\theta_{1.0} = (\underbrace{\mathbf{0.70}, \mathbf{0.999}, 0.77, \mathbf{0.78}, \mathbf{0.13}, 0.60, 0.68, 0.82, 0.23, \mathbf{0.31}, \mathbf{1.14}, \mathbf{0.07}}_{\theta^D}, \underbrace{\mathbf{-0.59}, \mathbf{0.74}, \mathbf{-0.61}, \mathbf{0.50}}_{\theta^U})'.$$

The KL criterion increases to 5.41E-05. Tables 9 and 10 show the measures of discrepancies between spectral densities and empirical distances for the considered sample sizes. The relative deviations increase to lie between 0.44% to 2.01%. Still, the increase in empirical distances for intermediate sample sizes is small. For samples of 80 and 200 observations the distances are 0.060 and 0.067 respectively, while for samples of 1000 observations the distance is 0.094. We also see that as the neighborhood around θ_0 that is excluded from search grows, more parameters in θ^D have to change

in addition to those in θ^U to minimize the KL criterion to the model at θ_0. Specifically, there are additional changes in τ, ψ_1, and σ_z, with $M_{g\epsilon}$ still having the largest change that makes the neighborhood constraint binding.

The global identification analysis in this section highlights two points. On one hand, it reinforces the local identification results in Section 7.3 as no model observationally equivalent to that at θ_0 is found outside a relatively small neighborhood around it. On the other hand, when the excluded neighborhood is substantially enlarged, we still find models that are very difficult to distinguish in empirical work.

The analysis shows that an important channel for near observational equivalence is the weak identification of $M_{g\epsilon}$. Then does the identification improve substantially when $M_{g\epsilon}$ is kept fixed? The corresponding results will be included in a later version.

7.5 Identification of policy rules

As for the An and Schorfheide (2007) model, consider a policy rule targeting expected inflation:

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

Denote the parameter vector in the alternative model as ϕ.

Starting with the case of indeterminacy, $\inf_{\phi\in\Phi} KL_{fh}(\theta_0,\phi)$ is obtained at

$$\phi = (\underbrace{0.69, 0.997, 0.77, 0.77, \mathbf{0.76}, 0.60, 0.68, 0.82, 0.23, 0.27, 1.13, 0.14}_{\phi^D}, \underbrace{\mathbf{-0.58}, \mathbf{1.54}, \mathbf{-0.61}, 0.20}_{\phi^U})'.$$

The minimized value equals

2.00E-13.

The three measures of deviations between $f(\theta_0)$ and $h(\phi)$ reported in Table 11 show that the absolute deviations are of order E-05 or smaller, while the relative deviations are of order E-06 or smaller. The empirical distance measure, as shown in Table 12, stays at 5% when the sample size is increased to 1000. These findings further confirm observational equivalence between the two models. This is again a strong result, showing that there exists an expectation rule that generates the identical dynamics as the current expectation rule under indeterminacy.

Under determinacy, $\inf_{\phi\in\Phi} KL_{fh}(\theta_0^D,\phi^D)$ is obtained at

$$\phi^D = (0.54, 0.992, 0.58, \mathbf{2.17}, \mathbf{1.57}, 0.84, 0.83, 0.85, 0.18, 0.18, 0.64, 0.36)'.$$

The minimized value equals

1.64E-13.

Table 11 shows the absolute deviations are of order E-06 or smaller, while the relative deviations are of order E-05 or smaller. Table 12 shows the empirical distance measure stays at 5% when the sample size is increased to 1000. Thus, the two models are indeed observationally equivalent. This implies the expectation rule at ϕ^D can lead to identical behavior dynamics as the current inflation rule at θ_0^D under determinacy. Therefore, the policy equivalence persists under both determinacy and indeterminacy.

Next, as in the An and Schorfheide (2007) example, we examine an alternative model with the monetary rule reacting to output growth instead of output gap:

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

The baseline models are given by θ_0 under indeterminacy and by θ_0^D under determinacy. The results on spectra deviations and empirical distance measures are summarized in Tables 13 and 14 respectively.

Searching over the region of indeterminacy, $\inf_{\phi\in\Phi} KL_{fh}(\theta_0, \phi)$ is obtained at

$$\phi = (\underbrace{\mathbf{0.68}, \mathbf{0.999}, \mathbf{0.76}, \mathbf{0.83}, \mathbf{0.01}, \mathbf{0.58}, \mathbf{0.69}, 0.82, 0.23, \mathbf{0.35}, \mathbf{1.15}, \mathbf{-0.02}}_{\phi^D}, \underbrace{\mathbf{-1.83}, \mathbf{2.41}, \mathbf{-1.04}, \mathbf{0.56}}_{\phi^U})'.$$

The minimized KL distance equals

6.03E-04.

Table 13 shows that the largest values of the three measures of deviations between $f(\theta_0)$ and $h(\phi)$ equal 4.912, 0.045 and 0.045 respectively. The empirical distance measure equals 8.75% for the sample size of 80, and increases to 28.51% when sample size is increased to 1000. Therefore, there is no policy equivalence between the output growth and output gap formulation of the monetary rule. In the case of determinacy, $\inf_{\phi\in\Phi} KL_{fh}(\theta_0, \phi)$ is obtained at

$$\phi^D = (0.54, \mathbf{0.900}, \mathbf{0.62}, \mathbf{2.49}, \mathbf{0.01}, 0.84, 0.83, 0.85, 0.18, 0.18, 0.64, \mathbf{0.35})'.$$

The resulting KL distance equals

1.65E-05.

The largest of the three measures of deviations equal 0.015, 0.076 and 0.088 respectively, demonstrating that the models are not observationally equivalent. The empirical distance measure equals

5.6% for a sample of 80 observations, and rises gradually to 7.2% when the sample size considered is 1000. This result is qualitatively similar to the one obtained with the An and Schorfheide (2007) model in that the output growth policy rule, although in principle identifiable, is difficult to distinguish from the output gap rule in practice.

8 Conclusion

This paper has presented a unified framework for analyzing identification in DSGE models that encompasses both determinacy and indeterminacy. In addition to providing necessary and sufficient conditions for local and global identification, it also proposes a measure of the empirical distance between two DSGE models. Empirical illustrations suggest that the methods are informative while being simple to implement. Although the analysis has focused on the second order properties for the ease of exposition, the information from the steady states can be easily incorporated. Studies involving medium scale DSGE models are currently in progress.

References

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

Benati, L., and P. Surico (2010): "VAR Analysis and the Great Moderation," *American Economic Review*, 99, 1636–1652.

Benhabib, J., and R. Farmer (1999): "Indeterminacies and Sector Specific Externalities," *Journal of Monetary Economics*, 37, 421–444.

Beyer, A., and R. 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 Economic Studies*, 88, 445–462.

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

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.

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. (2011): "Determinacy and identification with Taylor rules," *Journal of Political Economy*, 119, 565–615.

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

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

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

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

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

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

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

Negro, M. D., 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.

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

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

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

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

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

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

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

Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–1175.

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

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

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 when the degree of indeterminacy exceeds one.

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

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

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

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

where an ordering has been imposed such that the diagonal elements of Λ_{11} (Λ_{22}) are greater (smaller) than those of Ω_{11} (Ω_{22}) in absolute values. Assume the zero diagonal elements of Λ lie on different rows from those of Ω such that Ω_{22} is invertible. Because the generalized eigenvalues corresponding to Λ_{22} and Ω_{22} are unstable, the block of equations corresponding to $w_{2,t}$ has a stable solution if and only if

$$w_{2,0} = 0$$

and

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

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

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

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

where $\begin{bmatrix} U_{\cdot 1} & U_{\cdot 2} \end{bmatrix}$ and $\begin{bmatrix} V_{\cdot 1} & V_{\cdot 2} \end{bmatrix}$ are unitary matrices and D_{11} is nonsingular. $U_{\cdot 1}$ and $V_{\cdot 1}^*$ are unique up to multiplication by a unit-phase factor $\exp(i\varphi)$ (for the real case up to sign). The spaces spanned by $U_{\cdot 2}$ and $V_{\cdot 2}$ are also unique, although the matrices themselves are not if their column dimensions exceed one. We use the reduced column echelon form for $V_{\cdot 2}$ when implementing the relevant procedures when the dimensions exceed one. [5]

[5] Note that matrices with the same column space have the same unique reduced column echelon form. To see this, suppose $\widetilde{V}_{\cdot 2}$ and $\bar{V}_{\cdot 2}$ are two different choices of $V_{\cdot 2}$ in the singular value decomposition of $Q_{2\cdot}\Pi$. Because they span the same column space, they must satisfy $\widetilde{V}_{\cdot 2} = \bar{V}_{\cdot 2} N$ with N being a nonsingular matrix. (In fact, because $\bar{V}_{\cdot 2}^* \bar{V}_{\cdot 2} = \widetilde{V}_{\cdot 2}^* \widetilde{V}_{\cdot 2} = I$ and $\bar{V}_{\cdot 2} \bar{V}_{\cdot 2}^* = \widetilde{V}_{\cdot 2} \widetilde{V}_{\cdot 2}^* = I - V_{\cdot 1} V_{\cdot 1}^*$, we can deduce a stronger result that N is unitary.) Let $\widetilde{V}_{\cdot 2}^R$ be the unique reduced column echelon form of $\widetilde{V}_{\cdot 2}$. Then, there exists a sequence of elementary matrices $E_1, ..., E_K$ such that $\widetilde{V}_{\cdot 2}^R = \widetilde{V}_{\cdot 2} E_1 ... E_K . \widetilde{V}_{\cdot 2}^R = \widetilde{V}_{\cdot 2} E_1 ... E_K$. Thus, $\widetilde{V}_{\cdot 2}^R = \bar{V}_{\cdot 2} N E_1 ... E_K$. The matrix N itself equals the product of elementary matrices because it is nonsingular. Therefore, $\widetilde{V}_{\cdot 2}^R$ is also the unique reduced column echelon form of $\widetilde{V}_{\cdot 2}^R$.

Apply this decomposition, (A.2) can be equivalently represented as

$$(U_{\cdot 1} D_{11} V_{\cdot 1}^*)\, \eta_t = (-Q_{2\cdot}\Psi)\, \varepsilon_t \quad \text{for all } t > 0. \tag{A.3}$$

For any t, (A.3) can be considered as a system of linear equations in the form $Ax = b$ with $A = U_{\cdot 1} D_{11} V_{\cdot 1}^*$, $x = \eta_t$ and $b = (-Q_{2\cdot}\Psi)\, \varepsilon_t$. The full set of solutions for such a system is given by (let p be a specific solution to $Ax = b$)

$$\{p + v : v \text{ is any solution to } Ax = 0\}. \tag{A.4}$$

Apply (A.4) to (A.3) with $p = -V_{\cdot 1} D_{11}^{-1} U_{\cdot 1}^* Q$, $v = V_{\cdot 2}\epsilon_t$ and ϵ_t being an arbitrary vector conformable with $V_{\cdot 2}$. The full set of solutions to (A.3) can then be represented as

$$\left\{\eta_t : \eta_t = -V_{\cdot 1} D_{11}^{-1} U_{\cdot 1}^* Q_{2\cdot}\Psi\varepsilon_t + V_{\cdot 2}\epsilon_t \text{ with } E_{t-1}\epsilon_t = 0\right\}. \tag{A.5}$$

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

We now provide some computational details on how to obtain the solutions of S_t using (A.5) as in Sims (2002). Define a matrix Φ as the projection coefficients of the rows of $Q_{1\cdot}\Pi$ onto those of $Q_{2\cdot}\Pi$:

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

Note that

$$Q_{1\cdot}\Pi - \Phi Q_{2\cdot}\Pi = Q_{1\cdot}\Pi - Q_{1\cdot}\Pi V_{\cdot 1} V_{\cdot 1}^* = Q_{1\cdot}\Pi\left(I - V_{\cdot 1} V_{\cdot 1}^*\right),$$

which equals zero under determinacy. Multiplying (A.1) by

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

leads to

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

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

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

Further, using the expression (A.5), the last term satisfies

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

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

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

Multiply both sides of the equation with Z times the inverse of the most right hand side matrix, call it G_0 and, using $w_t = Z^* S_t$, we have

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

where

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

Because

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

the above matrices can also be rewritten by exploiting the zero blocks

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

Appendix B. Proofs

The proofs for Theorem 1 and Corollary 2 in Section 3 are essentially the same as those for Theorem 1 and Corollary 3 in Qu and Tkachenko (2012, supplementary material). This is because after parameter augmentation, θ determines the second order properties of the process. Here we still include the full detail to make the paper self-contained.

Let

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

then

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

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

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

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

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

which implies

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

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

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

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

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

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

we have

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

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

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

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

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

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

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

Then,

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

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

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

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

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

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

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

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

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

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

which implies

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

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

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

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

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

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

but

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

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

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

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

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

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

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

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

Therefore,

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

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

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

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

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

$$f_\theta^{-1}(\omega) f_{\theta_0}(\omega) = V^*(\omega)\Lambda(\omega)V(\omega). \tag{B.4}$$

Let $\lambda_j(\omega)$ be the j-th largest eigenvalue, then

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

For arbitrary $\lambda_j(\omega) > 0$, we always have $\lambda_j(\omega) - \log\lambda_j(\omega) - 1 \geq 0$ with the equality holding if and only if $\lambda_j(\omega) = 1$. Integrating over $[-\pi, \pi]$, we obtain $KL(\theta_1, \theta_0) \geq 0$ with the equality holding if and only if $\Lambda(\omega) = I$ for all $\omega \in [-\pi, \pi]$. However, this implies, using (B.4),

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

Thus,

$$f_\theta(\omega) = f_{\theta_0}(\omega)$$

for all $\omega \in [-\pi, \pi]$.

Proof of Theorem 3. Start with the distribution under the null hypothesis. First, from Assumption 5,

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

Second,

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

This satisfies a multivariate central limit theorem. The limiting variance is given by

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

Under the alternative hypothesis, the same argument can be applied.

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

	Prior Distribution				Pre-Volcker Posterior			Post-1982 Posterior		
	Distribution	Mean	SD		Mode	Mean	90% interval	Mode	Mean	90% interval
τ	Gamma	2	0.5	τ	2.342	2.514	[1.782, 3.376]	2.009	2.037	[1.265, 2.963]
r^*	Gamma	2	1	β	0.996	0.995	[0.991, 0.998]	0.996	0.995	[0.991, 0.998]
κ	Gamma	0.5	0.2	κ	0.444	0.490	[0.267, 0.765]	0.829	0.933	[0.551, 1.396]
ψ_1	Gamma	1.1	0.5	ψ_1	0.575	0.633	[0.311, 0.944]	2.037	2.124	[1.641, 2.668]
ψ_2	Gamma	0.25	0.15	ψ_2	0.149	0.231	[0.059, 0.492]	0.159	0.247	[0.063, 0.528]
ρ_r	Beta	0.5	0.2	ρ_r	0.887	0.871	[0.761, 0.959]	0.662	0.649	[0.545, 0.736]
ρ_g	Beta	0.7	0.1	ρ_g	0.669	0.659	[0.496, 0.810]	0.932	0.928	[0.889, 0.962]
ρ_z	Beta	0.7	0.1	ρ_z	0.623	0.603	[0.467, 0.728]	0.990	0.978	[0.938, 0.999]
σ_r	Inv. Gamma	0.31	0.16	σ_r	0.266	0.270	[0.236, 0.309]	0.208	0.224	[0.182, 0.278]
σ_g	Inv. Gamma	0.38	0.2	σ_g	0.640	0.576	[0.326,0.801]	0.663	0.695	[0.588, 0.823]
σ_z	Inv. Gamma	1	0.52	σ_z	0.562	0.624	[0.453, 0.820]	0.117	0.129	[0.092, 0.170]
$M_{r\epsilon}$	Normal	0	1	$M_{r\epsilon}$	0.554	0.528	[0.251, 0.792]			
$M_{g\epsilon}$	Normal	0	1	$M_{g\epsilon}$	0.059	-0.061	[-0.472, 0.185]			
$M_{z\epsilon}$	Normal	0	1	$M_{z\epsilon}$	0.327	0.259	[0.055, 0.492]			
σ_ϵ	Inv. Gamma	0.25	0.13	σ_ϵ	0.157	0.185	[0.121, 0.268]			

Note. The posterior means were obtained by generating 500000 draws from the log-posterior using the Random Walk Metropolis algorithm with the scaled inverse negative Hessian evaluated at the posterior mode as the proposal covariance matrix. The first 100000 draws were discarded as burn-in. Complete details on this computation can be found in An and Schorfheide (2007) and Tkachenko and Qu (2012). $\beta = (r^*/400+1)^{-1}$. The Inverse Gamma priors are as in Lubik and Schorfheide (2004): $p(\sigma|v,s) \propto \sigma^{-v-1}e^{-vs^2/2\sigma^2}$, where $v = 4$ and s equals 0.25, 0.3, 0.6 and 0.2 respectively.

Table 2. Deviations of model spectra at θ_c from θ_0, AS(2007) model.

	Six highest deviations across all elements and frequencies					
	1	2	3	4	5	6
	Maximum absolute deviations					
$c = 0.1$	0.0153	0.0122	0.0095	0.0037	0.0036	0.0005
$c = 0.5$	0.0730	0.0576	0.0441	0.0200	0.0194	0.0025
$c = 1.0$	0.1064	0.0791	0.0554	0.0387	0.0375	0.0039
	Maximum absolute deviations in relative form					
$c = 0.1$	0.0215	0.0136	0.0022	0.0019	0.0017	0.0014
$c = 0.5$	0.1155	0.0100	0.0092	0.0079	0.0725	0.0064
$c = 1.0$	0.2239	0.1400	0.0154	0.0135	0.0108	0.0081
	Maximum relative deviations					
$c = 0.1$	0.0215	0.0136	0.0022	0.0019	0.0017	0.0014
$c = 0.5$	0.1155	0.0100	0.0092	0.0079	0.0725	0.0064
$c = 1.0$	0.2239	0.1400	0.0154	0.0135	0.0108	0.0081

Note. θ_0 is the posterior mean from the pre-Volcker subsample; θ_c is the minimizer of the KL criterion over $\{\theta : |\theta - \theta_0|_\infty \geq c\}$.

Table 3. Empirical distances of the model at θ_c from θ_0, AS(2007) model.

	$KL_{fh}(\theta_0, \theta_c)$	$T = 80$	$T = 150$	$T = 200$	$T = 1000$
$c = 0.1$	5.36E-07	0.0509	0.0513	0.0515	0.0534
$c = 0.5$	1.57E-05	0.0553	0.0574	0.0586	0.0710
$c = 1.0$	7.57E-05	0.0621	0.0672	0.0703	0.1041

Note. The empirical distance measures are computed for $\alpha = 0.05$.

Table 4. Changes in spectra switching to the expected inflation rule, AS(2007) model.

	Six highest deviations across all elements and frequencies					
	1	2	3	4	5	6
	Maximum absolute deviations					
ϕ	4.32E-06	3.64E-06	3.05E-06	7.28E-07	6.97E-07	1.20E-07
ϕ^D	1.79E-05	8.50E-06	8.38E-06	4.85E-06	3.93E-06	1.81E-06
	Maximum absolute deviations in relative form					
ϕ	4.21E-06	2.60E-06	5.48E-07	4.99E-07	4.79E-07	4.45E-07
ϕ^D	4.75E-05	4.57E-05	1.08E-06	5.21E-07	5.15E-07	5.06E-07
	Maximum relative deviations					
ϕ	4.21E-06	2.60E-06	5.48E-07	4.99E-07	4.79E-07	4.45E-07
ϕ^D	4.75E-05	4.57E-05	1.08E-06	5.21E-07	5.06E-07	1.08E-06

Note. ϕ and ϕ^D are the parameters under the expected inflation rule that minimize the KL criterion to θ_0 and θ_0^D under indeterminacy and determinacy, respectively.

Table 5. Empirical distances of models at ϕ, ϕ^D from θ_0, θ_0^D, AS(2007) model.

	$KL_{fh}(\theta_0, \theta_c)$	$T = 80$	$T = 150$	$T = 200$	$T = 1000$
ϕ	3.35E-14	0.0500	0.0500	0.0500	0.0500
ϕ^D	6.11E-14	0.0500	0.0500	0.0500	0.0500

Note. The empirical distance measures are computed for $\alpha = 0.05$.

Table 6. Changes in spectra switching to the output growth rule, AS(2007) model.

	Six highest deviations across all elements and frequencies					
	1	2	3	4	5	6
	Maximum absolute deviations					
ϕ	0.0657	0.0535	0.0428	0.0198	0.0192	0.0029
ϕ^D	0.0607	0.0286	0.0135	0.0084	0.0070	0.0037
	Maximum absolute deviations in relative form					
ϕ	0.1147	0.0714	0.0117	0.0083	0.0073	0.0062
ϕ^D	0.0223	0.0199	0.0038	0.0037	0.0037	0.0005
	Maximum relative deviations					
ϕ	0.1147	0.0714	0.0117	0.0083	0.0073	0.0062
ϕ^D	0.1175	0.0867	0.0127	0.0038	0.0037	0.0005

Note. ϕ and ϕ^D are the parameters under the output growth rule that minimize the KL criteria to θ_0 and θ_0^D under indeterminacy and determinacy, respectively

Table 7. Empirical distances of models at ϕ, ϕ^D from θ_0, θ_0^D, AS(2007) model.

	$KL_{fh}(\theta_0, \theta_c)$	$T = 80$	$T = 150$	$T = 200$	$T = 1000$
ϕ	1.94E-05	0.0560	0.0583	0.0597	0.0738
ϕ^D	4.60E-05	0.0600	0.0638	0.0661	0.0905

Note. The empirical distance measures are computed for $\alpha = 0.05$.

Table 8. $p_{fh}(\theta_0^D, \phi_0, 0.05, T)$ in the AS(2007) example

Alternative model	$KL_{fh}(\theta_0, \phi_0)$	$T = 80$	$T = 150$	$T = 200$	$T = 1000$
$(\psi_1, \psi_2, \rho_r, \sigma_r) = (2.07, 1.24, 0.69, 0.24)$	9.54E-05	0.061	0.067	0.070	0.109
$\beta = 0.9852$	1.20E-05	0.054	0.055	0.056	0.067
$\psi_1 = 1.5$, $\psi_2 = 0.5$	0.362	0.950	0.992	0.998	1.000

Note. The default model is as in (13). The alternative model column shows specification changes with respect to the default model.

Table 9. Deviations of model spectra at θ_c from θ_0, LS(2004) model.

	Six highest deviations across all elements and frequencies					
	1	2	3	4	5	6
	Maximum absolute deviations					
$c = 0.1$	0.1576	0.1136	0.0100	0.0815	0.0076	0.0008
$c = 0.5$	1.2670	0.9696	0.7417	0.0404	0.0357	0.0029
$c = 1.0$	2.9222	2.2894	1.7910	0.0870	0.0765	0.0067
	Maximum absolute deviations in relative form					
$c = 0.1$	0.0010	0.0011	0.0010	0.0008	0.0008	0.0002
$c = 0.5$	0.0087	0.0082	0.0077	0.0034	0.0033	0.0016
$c = 1.0$	0.0201	0.0194	0.0186	0.0074	0.0070	0.0037
	Maximum relative deviations					
$c = 0.1$	0.0010	0.0011	0.0010	0.0009	0.0008	0.0004
$c = 0.5$	0.0087	0.0082	0.0077	0.0051	0.0034	0.0022
$c = 1.0$	0.0201	0.0194	0.0186	0.0095	0.0074	0.0044

Note. θ_0 is the posterior mean from the pre-Volcker subsample; θ_c is the minimizer of the KL criterion over $\{\theta : |\theta - \theta_0|_\infty \geq c\}$.

Table 10. Empirical distances of the model at θ_c from θ_0, LS(2004) model.

	$KL_{fh}(\theta_0, \theta_c)$	$T = 80$	$T = 150$	$T = 200$	$T = 1000$
$c = 0.1$	4.06E-07	0.0507	0.0510	0.0512	0.0529
$c = 0.5$	1.22E-05	0.0550	0.0569	0.0580	0.0686
$c = 1.0$	5.41E-05	0.0604	0.0646	0.0671	0.0941

Note. The empirical distance measures are computed for $\alpha = 0.05$.

Table 11. Changes in spectra switching to the expected inflation rule, LS(2004) model.

	Six highest deviations across all elements and frequencies					
	1	2	3	4	5	6
	Maximum absolute deviations					
ϕ	1.00E-04	9.99E-05	7.54E-05	2.96E-05	2.42E-05	1.76E-06
ϕ^D	2.25E-06	2.04E-06	1.86E-06	1.75E-06	1.35E-06	3.88E-07
	Maximum absolute deviations in relative form					
ϕ	2.28E-06	2.22E-06	9.14E-07	8.47E-07	7.83E-07	5.22E-07
ϕ^D	1.15E-05	9.95E-07	8.98E-07	6.67E-07	5.50E-07	2.83E-07
	Maximum relative deviations					
ϕ	2.28E-06	2.22E-06	9.14E-07	8.47E-07	7.83E-07	5.22E-07
ϕ^D	1.18E-05	1.76E-06	9.95E-07	8.98E-07	6.67E-07	2.83E-07

Note. ϕ and ϕ^D are the parameters under the expected inflation rule that minimize the KL criterion to θ_0 and θ_0^D under indeterminacy and determinacy, respectively.

Table 12. Empirical distances of models at ϕ, ϕ^D from θ_0, θ_0^D, LS(2004) model.

	$KL_{fh}(\theta_0, \theta_c)$	$T = 80$	$T = 150$	$T = 200$	$T = 1000$
ϕ	2.00E-13	0.0500	0.0500	0.0500	0.0500
ϕ^D	1.64E-13	0.0500	0.0500	0.0500	0.0500

Note. The empirical distance measures are computed for $\alpha = 0.05$.

Table 13. Changes in spectra switching to the output growth rule, LS(2004) model.

	Six highest deviations across all elements and frequencies					
	1	2	3	4	5	6
	Maximum absolute deviations					
ϕ	4.9122	4.6734	4.3012	0.0580	0.0425	0.0363
ϕ^D	0.0149	0.0114	0.0067	0.0048	0.0032	0.0015
	Maximum absolute deviations in relative form					
ϕ	0.0447	0.0396	0.0337	0.0108	0.0063	0.0043
ϕ^D	0.0759	0.0117	0.0041	0.0017	0.0011	0.0008
	Maximum relative deviations					
ϕ	0.0447	0.0396	0.0337	0.0268	0.0183	0.0126
ϕ^D	0.0882	0.0814	0.0106	0.0041	0.0023	0.0011

Note. ϕ and ϕ^D are the parameters under the output growth rule that minimize the KL criteria to θ_0 and θ_0^D under indeterminacy and determinacy, respectively.

Table 14. Empirical distances of models at ϕ, ϕ^D from θ_0, θ_0^D, LS(2004) model.

	$KL_{fh}(\theta_0, \theta_c)$	$T = 80$	$T = 150$	$T = 200$	$T = 1000$
ϕ	6.03E-04	0.0875	0.107	0.120	0.285
ϕ^D	1.65E-05	0.0560	0.0582	0.0595	0.0723

Note. The empirical distance measures are computed for $\alpha = 0.05$.