Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models

ZHONGJUN QU
Department of Economics, Boston University

DENIS TKACHENKO
Department of Economics, Boston University

This paper considers issues related to identification, inference, and computation in linearized dynamic stochastic general equilibrium (DSGE) models. We first provide a necessary and sufficient condition for the local identification of the structural parameters based on the (first and) second order properties of the process. The condition allows for arbitrary relations between the number of observed endogenous variables and structural shocks, and is simple to verify. The extensions, including identification through a subset of frequencies, partial identification, conditional identification, and identification under general nonlinear constraints, are also studied. When lack of identification is detected, the method can be further used to trace out nonidentification curves. For estimation, restricting our attention to nonsingular systems, we consider a frequency domain quasi-maximum likelihood estimator and present its asymptotic properties. The limiting distribution of the estimator can be different from results in the related literature due to the structure of the DSGE model. Finally, we discuss a quasi-Bayesian procedure for estimation and inference. The procedure can be used to incorporate relevant prior distributions and is computationally attractive.

KEYWORDS. Infinite dimensional mapping, local identification, MCMC, nonidentification curve, rank condition, spectral domain.

JEL CLASSIFICATION. C10, C13, C30, E1.

1. INTRODUCTION

The formal quantitative analysis of dynamic stochastic general equilibrium (DSGE) models has become an important subject of modern macroeconomics. It is typically

conducted in the time domain using a state space representation with the aid of Kalman or particle filtering; see An and Schorfheide (2007) and Fernández-Villaverde (2010) for reviews of related literature. This paper considers issues related to identification, inference, and computation from a spectral domain perspective. The goal is to present a unified framework for identifying and estimating linearized DSGE models based on the mean and the spectrum of the underlying process.

The identification of DSGE models is important for both calibration and formal statistical analysis, although the relevant literature is relatively sparse. Substantial progress has been made recently, notably by Iskrev (2010) and Komunjer and Ng (2011), and by Canova and Sala (2009), Consolo, Favero, and Paccagnini (2009), and Fukac, Waggoner, and Zha (2007). Komunjer and Ng (2011) documented that an inherent difficulty in the identification analysis is that the reduced form parameters (i.e., the ones appearing directly in the solution of the model) are in general not identifiable, thus the traditional approach of identifying structural parameters from the reduced form breaks down. Also, the solution system of a DSGE model can be singular (i.e., when the number of observed endogenous variables is greater than the number of exogenous shocks), which constitutes an additional layer of conceptual difficulty. They provided necessary and sufficient conditions for the local identification of the dynamic parameters by exploiting the dynamic structure of the model. Our identification analysis is distinctly different from theirs and other related work in the literature. Specifically, we work in the frequency domain, treating the spectral density as an infinite dimensional mapping, and delivering simple identification conditions applicable to both singular and nonsingular DSGE systems without relying on a particular (say, the minimal state) representation.

We first focus on the identification of the dynamic parameters from the spectrum. We treat the elements of the spectral density matrix as mappings from the structural parameter space to complex valued functions defined over $[-\pi, \pi]$ in a Banach space. Then the parameters are locally identified if and only if the overall mapping is locally injective (that is, if any local change in parameter values leads to a different image). This leads to a necessary and sufficient rank condition for local identification, which depends on the first order derivative of the spectral density matrix with respect to the structural parameters of interest. Depending on the model at hand, the resulting condition can be easily evaluated analytically or numerically. The result is general because the assumptions mainly involve the uniqueness of the DSGE solution (i.e., determinacy) and the continuity and smoothness of the spectral density matrix. Note that although the identification condition is formulated in the spectral domain, it has a time domain interpretation as well. Specifically, under some regularity condition that ensures a one-to-one mapping between the spectral density matrix and the autocovariance functions, the condition is also necessary and sufficient for local identification through the complete set of autocovariances. Next, we incorporate the steady state parameters into the analysis and study identification through both the first and second order properties of the process. The result we obtain is analogous to the previous case with the addition of an extra term depending on the steady state parameters. When interpreted in the time domain, this condition is necessary and sufficient for local identification through the mean and the complete set of autocovariances.

We discuss various extensions of these two identification results. (i) We study identification through a subset of frequencies. This is relevant for situations where it is desirable to construct estimators based on a subset of frequencies to minimize the effect of unmodeled seasonality or measurement errors. (ii) We consider partial identification, that is, identifying a subset of parameters without making identification statements about the rest. (iii) We give a necessary and sufficient condition for conditional identification, that is, the identification of a subset of parameters while holding the values of the other parameters fixed at some known value. (iv) We also study identification under general nonlinear parameter constraints. For example, this allows us to constrain some monetary shocks to have no long run effect on real variables, which can be easily formulated as a set of restrictions on the spectral density matrix at frequency zero. The second and third extensions are motivated by Komunjer and Ng (2011), although the assumptions they used are different. The first extension is new. It provides the identification foundation for inference based on a subset of frequencies studied later in the paper.

Furthermore, when lack of identification is detected, our method can be used to trace out parameter values that yield processes with identical (first and) second order properties. We summarize the path of these values via nonidentification curves and provide a simple algorithm to obtain them. It appears that our paper is the first to deliver such curves. They can serve three purposes. First, because they showcase which parameters are unidentified and their equivalent parameter values, they are useful for building a DSGE model. Second, because they characterize the size of the nonidentified local neighborhood, they are useful for inference. In particular, if the neighborhood is very small, then the lack of local identification arguably may not be a great threat to inference that assumes identification nonetheless; otherwise, serious thoughts should be given. Third, the curves can be embedded into a procedure to ensure the robustness of the identification analysis. This point is elaborated using an example in Section 3.2.

We illustrate the proposed method using a model considered by An and Schorfheide (2007) and document a serious concern about the identification of the parameters in the Taylor rule equation. The result shows that when varying parameters in this equation along a certain path, the (mean and) spectrum of the observables stay the same; thus it is impossible to uniquely pin down the parameter values even with an infinite sample. The values on the curve suggest that in this model it is impossible to distinguish between a hawkish rule (a long run policy coefficient of 1.57 on inflation and 0.00 on output, resulting in respective Taylor rule weights of 0.41 and 0.00) and a more dovish rule (0.99 on inflation and 1.00 on output, with Taylor rule weights of 0.20 on each). To our knowledge, the current paper is the first to document such an identification feature about the Taylor rule parameters.

As will become clear, our results, as well as their proofs, are closely connected to Rothenberg (1971), who considered identification of parametric econometric models from the density functions and provided rank conditions based on the information matrix. However, there exists an important difference. Namely, in our analysis, the spectral density is a complex valued matrix that may be singular. Under singularity, the conventional information matrix does not exist. This generates some conceptual and technical difficulties that do not arise in Rothenberg (1971). Consequently, our condition is based

on a criterion function different from the information matrix. We further show that when restricting to the nonsingular special case, our condition is equivalent to evaluating the rank of the information matrix. Therefore, the condition of Rothenberg (1971) still applies, albeit only to nonsingular models.

An identification result is useful only if it corresponds to an estimator. This motivates the consideration of the frequency domain quasi-maximum likelihood (FQML) estimation in this paper. The FQML approach was first proposed by Whittle (1951). Its statistical properties have been studied by, among others, Dunsmuir and Hannan (1976), Dunsmuir (1979), and Hosoya and Taniguchi (1982) in the statistics literature. In the economics literature, Hansen and Sargent (1993) derived the FQML as an approximation to the time domain Gaussian quasi-maximum likelihood (QML) and used it to understand the effect of seasonal adjustment in estimating rational expectations models. Diebold, Ohanian, and Berkowitz (1998) laid out a general framework for estimation and model diagnostics based on a full second order comparison of the model and data dynamics. Their criterion function includes FQML as a special case.

The contribution of the current paper in the area of FQML estimation is threefold. First, we formally establish the link between the identification result and the property of the estimator by showing that the rank condition derived is necessary and sufficient for the estimator to be asymptotically locally unique. Therefore, the identification result is empirically relevant. Second, we derive the limiting distribution of the estimator under mild conditions. Finally, we discuss a computationally attractive method to obtain the estimates, following the approach of Chernozhukov and Hong (2003). In addition to the computational advantage, it allows us to impose priors on the parameters, thus having a (quasi-) Bayesian interpretation. Note that the above results allow for estimation using only a subset of frequencies.

In addition to the above mentioned papers, there exists a small but growing literature that exploits the merits of estimation and diagnosis of econometric models in the spectral domain. Engle (1974) considered band spectrum regressions and demonstrated their value in dealing with errors in variables and seasonality. Altug (1989) applied FQML to estimate models with additive measurement errors. Watson (1993) suggested plotting the model and data spectra as one of the most informative diagnostics. Berkowitz (2001) considered the estimation of rational expectation models based on the spectral properties of the Euler residuals. Also see Christiano, Eichenbaum, and Marshall (1991) and Christiano and Vigfusson (2003) for applications of FQML to various problems. We believe that the identification, estimation, and computational results obtained in this paper can be useful to further develop the literature in this field and to facilitate estimation and comparison of more sophisticated models.

The paper is organized as follows. The structure of the DSGE solution is discussed in Section 2. Section 3 considers the local identification of the structural parameters together with an algorithm to trace out nonidentification curves and an illustrative example. The FQML estimator and its asymptotic properties are studied in Section 4. The discussion on interpretation of the estimates in misspecified models is also included. Section 5 presents a quasi-Bayesian approach for computation and inference. Section 6

concludes. All proofs are contained in the Appendix available in a supplementary file on the journal website, http://qeconomics.org/supp/126/supplement.pdf.

The following notation is used: $|z|$ is the modulus of z; the imaginary unit is denoted by i; X^* stands for the conjugate transpose of a complex valued matrix X. For a random vector x_t, x_{ta} denotes its ath element. For a matrix A, A_{ab} stands for its (a, b)th entry. If $f_\theta \in \mathbb{R}^k$ is a differentiable function of $\theta \in \mathbb{R}^p$, then $\partial f_{\theta_0}/\partial \theta'$ is a $k \times p$ matrix of partial derivatives evaluated at θ_0. We use the notation "$\to^p$" and "$\to^d$" to denote convergence in probability and in distribution, and $O_p(\cdot)$ and $o_p(\cdot)$ are the usual symbols for stochastic orders of magnitude.

2. The model

Suppose a discrete time DSGE model has been solved and log-linearized around the steady state. Assume the solution is unique. Let $Y_t^d(\theta)$ be the log deviations of endogenous variables from their steady states, with θ being a finite dimensional structural parameter vector containing the dynamic parameters. The log deviations $Y_t^d(\theta)$ can be represented in various ways, and our method does not rely on a particular representation. To maintain generality, we only assume that they are representable as

$$Y_t^d(\theta) = \sum_{j=0}^{\infty} h_j(\theta)\epsilon_{t-j}, \tag{1}$$

where $h_j(\theta)$ $(j = 0, \ldots, \infty)$ are real valued matrices of constants and $\{\epsilon_t\}$ is a white noise process of unobserved structural shocks. The dimensions of the relevant variables and parameters are

$$Y_t^d(\theta) : n_Y \times 1, \qquad \epsilon_t : n_\epsilon \times 1, \qquad h_j(\theta) : n_Y \times n_\epsilon, \qquad \theta : q \times 1.$$

Let $H(L; \theta)$ denote the matrix of lagged polynomials, that is,

$$H(L; \theta) = \sum_{j=0}^{\infty} h_j(\theta)L^j. \tag{2}$$

Then $Y_t^d(\theta)$ can be written concisely as

$$Y_t^d(\theta) = H(L; \theta)\epsilon_t. \tag{3}$$

REMARK 1. We work directly with the vector moving average representation (3) without assuming invertibility, that is, $\epsilon_t = \sum_{j=0}^{\infty} g_j(\theta)Y_{t-j}^d(\theta)$ for some $g_j(\theta)$. Invertibility is restrictive because it requires $n_Y \geq n_\epsilon$. Consequently, we allow for both $n_Y \geq n_\epsilon$ and $n_Y < n_\epsilon$. Note that the system is singular if $n_Y > n_\epsilon$.

ASSUMPTION 1. *The process $\{\epsilon_t\}$ satisfies $E(\epsilon_t) = 0$, $E(\epsilon_t \epsilon_t') = \Sigma(\theta)$ with $\Sigma(\theta)$ being a finite $n_\epsilon \times n_\epsilon$ matrix for all θ, and $E(\epsilon_t \epsilon_s') = 0$ for all $t \neq s$: $\sum_{j=0}^{\infty} \text{tr}(h_j(\theta)\Sigma(\theta)h_j(\theta)') < \infty$.*

Assumption 1, along with (1), implies that $Y_t^d(\theta)$ is covariance stationary and has a spectral density matrix $f_\theta(\omega)$ that can be written as

$$f_\theta(\omega) = \frac{1}{2\pi} H(\exp(-i\omega); \theta) \Sigma(\theta) H(\exp(-i\omega); \theta)^*, \tag{4}$$

where X^* denotes the conjugate transpose of a generic complex matrix X. To illustrate the flexibility of the above framework, we consider the following two examples.

EXAMPLE 1. Consider a linear rational expectations system as in Sims (2002) (in this example and the next, we omit the dependence of the parameters on θ to simplify notation),

$$\Gamma_0 S_t = \Gamma_1 S_{t-1} + \Psi Z_t + \Pi \eta_t, \tag{5}$$

where S_t is a vector of model variables that includes the endogenous variables and the conditional expectation terms, Z_t is an exogenously evolving, possibly serially correlated, random disturbance, and η_t is an expectational error. Models with more lags or with lagged expectations can be accommodated by expanding the S_t vector accordingly. Then, under some conditions (Sims (2002, p. 12)), the system can be represented as

$$S_t = \Theta_1 S_{t-1} + \Theta_0 Z_t + \Theta_S \sum_{j=1}^\infty \Theta_f^{j-1} \Theta_Z E_t Z_{t+j}, \tag{6}$$

where Θ_0, Θ_1, Θ_S, Θ_f, and Θ_Z are functions of Γ_0, Γ_1, Ψ, and Π. Assuming Z_t follows a vector linear process (for example, $Z_{t+1} = \Phi Z_t + \epsilon_{t+1}$), we then have $S_t = \Theta_1 S_{t-1} + B(L)\epsilon_t$ for some lag polynomial matrix $B(L)$, implying $S_t = (I - \Theta_1 L)^{-1} B(L)\epsilon_t$.

Let $A(L)$ be a matrix of finite order lag polynomials that specifies the observables such that

$$Y_t^d = A(L)S_t.$$

Then we have

$$Y_t^d = A(L)(I - \Theta_1 L)^{-1} B(L)\epsilon_t.$$

Therefore, the spectral density of Y_t^d is given by (4) with $H(L; \theta) = A(L)(I - \Theta_1 L)^{-1} \times B(L)$.

REMARK 2. In the above example, the matrix $A(L)$ offers substantial flexibility since it allows us to study identification and estimation based on a subset of variables (equations) or a linear transformation of them. To see this, suppose S_t includes two endogenous variables x_t and w_t. Then $A(L)$ can be chosen such that Y_t^d includes only x_t but not w_t, or includes $x_t - x_{t-1}$ but not x_t. Consequently, it is straightforward to analyze DSGE models with latent endogenous variables simply by assigning 0s and 1s to the entries of $A(L)$. We illustrate the specification of $A(L)$ in Section 3.2 through a concrete example. Note that such analysis is permitted because we do not impose restrictions on the relation between n_Y and n_ϵ.

EXAMPLE 2. Another representation used in the literature by, among others, Uhlig (1999), is

$$k_{t+1} = Pk_t + Qz_t,$$
$$w_t = Rk_t + Sz_t,$$
$$z_{t+1} = \Psi z_t + \epsilon_{t+1},$$

where k_t is a vector of observed endogenous (state) variables whose values are known at time t, w_t is a vector of observed endogenous (jump) variables, z_t has the same definition as in the previous example, and P, Q, R, S, and Ψ are matrices of constants depending on the structural parameter θ. Let

$$Y_t^d = \begin{pmatrix} k_t \\ w_t \end{pmatrix}. \tag{7}$$

Then the spectral density of Y_t^d is given by (4) with

$$H(L; \theta) = \begin{pmatrix} L^{-1}[I - PL] & 0 \\ -R & I \end{pmatrix}^{-1} \begin{pmatrix} Q \\ S \end{pmatrix} [I - \Psi L]^{-1}.$$

Again, one can study identification and estimation based on a subset of equations or a linear combination of them by picking an appropriate $A(L)$ and considering $Y_t^d = A(L)(k_t', w_t')'$ instead of (7), which corresponds to

$$H(L; \theta) = A(L) \begin{pmatrix} L^{-1}[I - PL] & 0 \\ -R & I \end{pmatrix}^{-1} \begin{pmatrix} Q \\ S \end{pmatrix} [I - \Psi L]^{-1}. \tag{8}$$

As becomes clear later, if estimating the dynamic parameters is the main objective, then it is not necessary to specify the steady states of the DSGE solution. However, in some cases one may be interested in estimating the dynamic and steady state parameters jointly, for example, for conducting welfare analyses. Our framework permits this. First, recall that θ denotes the dynamic parameter vector. Importantly, parameters that affect both the steady states and the log deviations are treated as dynamic, and thus are included in θ. Next, let α denote the parameters that affect only the steady states, which is possibly a null set in some DSGE models. Finally, define the augmented parameter vector

$$\bar{\theta} = (\theta', \alpha')'$$

and assume that the observables (Y_t) are related to the log deviations $(Y_t^d(\theta))$ and the steady states $(\mu(\bar{\theta}))$ via

$$Y_t = \mu(\bar{\theta}) + Y_t^d(\theta).$$

The above expression acknowledges that in DSGE models, the constant term μ typically depends on both θ and α. In the remainder of the paper, we examine the identification and estimation of θ based on the properties of $f_\theta(\omega)$ alone, and of $\bar{\theta}$ based jointly on $\mu(\bar{\theta})$ and $f_\theta(\omega)$.

3. Local identification of structural parameters

We first consider the identification of θ at some θ_0 and subsequently of $\bar{\theta}$ at some $\bar{\theta}_0$. The next assumption imposes some restrictions on the parameter space.

ASSUMPTION 2. *We have $\theta \in \Theta \subset \mathbb{R}^q$ and $\bar{\theta} \in \bar{\Theta} \subset \mathbb{R}^{p+q}$ with Θ and $\bar{\Theta}$ being compact and convex. Assume θ_0 and $\bar{\theta}_0$ are interior points of Θ and $\bar{\Theta}$, respectively.*

Note that for identification analysis alone, we do not require the compactness and convexity assumptions on Θ and $\bar{\Theta}$. However, they are needed to study the asymptotic properties of the parameter estimates.

The concept for location identification is defined in the same way as in Rothenberg (1971, see his Definition 3).

DEFINITION 1. The dynamic parameter vector θ is said to be locally identifiable from the second order properties of $\{Y_t\}$ at a point 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 $\theta_0 = \theta_1$.

The above concept is formulated in the frequency domain. However, there is an equivalent formulation in the time domain in terms of autocovariance functions. Specifically, suppose $\{Y_t\}$ satisfy Assumption 1 with autocovariance function $\Gamma(k)$ ($k = 0, \pm 1, \ldots$) satisfying $\Gamma(k) = \Gamma(-k)$ and that $f_\theta(\omega)$ is continuous in ω. Then Theorem $1''$ in Hannan (1970, p. 46) implies that there is a one-to-one mapping between $\Gamma(k)$ ($k = 0, \pm 1, \ldots$) and $f_\theta(\omega)$ ($\omega \in [-\pi, \pi]$) given by

$$\Gamma(k) = \int_{-\pi}^{\pi} \exp(ik\omega) f_\theta(\omega)\, d\omega.$$

Therefore, θ is locally identifiable from $f_\theta(\omega)$ if and only if it is locally identifiable from the complete set of autocovariances $\{\Gamma(k)\}_{k=-\infty}^{\infty}$ of Y_t.

The spectral density matrix has n_Y^2 elements. Each element can be viewed as a map from Θ to complex valued functions defined over $[-\pi, \pi]$ in a Banach space. Therefore, the parameters are locally identified at θ_0 if and only if the overall mapping is locally injective (i.e., any local change in parameter values will lead to a different image for some element). The mappings are infinite dimensional and difficult to analyze directly. However, it turns out the identification can be characterized by a finite dimensional matrix. To state this precisely, we start with the following assumption.

ASSUMPTION 3. *The elements of $f_\theta(\omega)$ are continuous in ω, and continuous and differentiable in θ. The elements of the derivatives $\partial \operatorname{vec}(f_\theta(\omega))/\partial\theta'$ are continuous in θ and ω. Let*

$$G(\theta) = \int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec}(f_\theta(\omega)')}{\partial\theta'}\right)' \left(\frac{\partial \operatorname{vec}(f_\theta(\omega))}{\partial\theta'}\right) d\omega. \tag{9}$$

Assume there exists an open neighborhood of θ_0 in which $G(\theta)$ has a constant rank.

This first part of the assumption requires the spectral density to be smooth with continuous first order derivatives. The second part requires θ_0 to be a regular point of the matrix $G(\theta)$. These assumptions are quite mild. Note that in the definition of $G(\theta)$, the primes (′) denote simple transposes rather than conjugate transposes. Alternatively, we can also write $G(\theta)$ as

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

where the asterisk ($*$) now denotes the conjugate transpose.

REMARK 3. The dimension of $G(\theta)$ is always $q \times q$ and independent of n_Y or n_ϵ. Its (j, k)th element is given by

$$G_{jk}(\theta) = \int_{-\pi}^{\pi} \operatorname{tr}\left\{ \frac{\partial f_\theta(\omega)}{\partial \theta_j} \frac{\partial f_\theta(\omega)}{\partial \theta_k} \right\} d\omega.$$

We use this representation to compute $G(\theta)$ in the application in Section 3.2. Lemma A.1 in the Appendix provides another representation, showing explicitly that the integrand of $G(\theta)$, therefore $G(\theta)$ itself, is real, symmetric, and positive semidefinite. This feature is useful for proving the subsequent theoretical results.

THEOREM 1. *Let Assumptions 1–3 hold. Then θ is locally identifiable from the second order properties of $\{Y_t\}$ at a point θ_0 if and only if $G(\theta_0)$ is nonsingular.*

The main computational work in obtaining $G(\theta_0)$ is to evaluate the first order derivatives and to compute the integral. This is typically straightforward using numerical methods. First, divide the interval $[-\pi, \pi]$ into N subintervals to obtain $(N + 1)$ frequency indices. Let ω_s denote the sth frequency in the partition. Then $\partial f_{\theta_0}(\omega_s)/\partial \theta_j$ can be computed numerically using a simple two-point method,

$$\frac{f_{\theta_0 + \mathbf{e}_j h_j}(\omega_s) - f_{\theta_0}(\omega_s)}{h_j}, \quad j = 1, \dots, N+1,$$

where $\mathbf{e}_j$ is a $q \times 1$ unit vector with the jth element equal to 1 and h_j is a step size that can be parameter dependent. In practice, to obtain the right hand side quantity, we only need to solve the DSGE model twice, once using $\theta = \theta_0$ and once with $\theta = \theta_0 + \mathbf{e}_j h_j$. After this is repeated for all parameters in θ, we can compute $G_{jk}(\theta_0)$ using

$$\frac{2\pi}{N+1} \sum_{s=1}^{N+1} \operatorname{tr}\left\{ \frac{\partial f_\theta(\omega_s)}{\partial \theta_j} \frac{\partial f_\theta(\omega_s)}{\partial \theta_k} \right\}.$$

Note that no simulation is needed in this process. For the model considered in Section 3.2 (An and Schorfheide (2007)), the computation takes less than a minute to finish with $N = 9999$.

Because $G(\theta)$ is real, symmetric, and positive semidefinite, its eigendecomposition always exists. Therefore, the rank of $G(\theta_0)$ can be evaluated using an algorithm for eigenvalue decomposition and counting the number of nonzero eigenvalues.

Theorem 1 is closely related to Theorem 1 in Rothenberg (1971), who considered identification in parametric models. In his case, $f_\theta(\omega)$ is replaced by the parametric density function and $G(\theta)$ is simply the information matrix. Since the information matrix describes the local curvature of the log-likelihood as a function of θ, its rank naturally provides a measure for identification, for lack of identification is simply the lack of sufficient information to distinguish between alternative structures. In our case, the result is equally intuitive, since the parameters are locally identified if and only if any deviation of the parameters from θ_0 leads to different mappings for $f_\theta(\omega)$. We now state a result that formally establishes the link with Rothenberg's (1971) condition. Note that under Gaussianity the information matrix is given by[1]

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

which is defined only if the system is nonsingular. We restrict our attention to such a situation.

COROLLARY 1. *Let Assumptions 1–3 hold. In addition, assume $f_{\theta_0}(\omega)$ has full rank for all $\omega \in [-\pi, \pi]$. Then $G(\theta_0)$ and $I(\theta_0)$ have the same rank. Also, for any $c \in \mathbb{R}^q$, $G(\theta_0)c = 0$ if and only if $I(\theta_0)c = 0$.*

Therefore, Rothenberg's (1971) condition applies to DSGE models, albeit only to nonsingular systems. Because $G(\theta_0)$ and $I(\theta_0)$ share the same null space, they deliver the same information about nonidentification. The issue of nonidentification is further addressedin Section 3.1.

Given the insight conveyed by Theorem 1, it becomes straightforward to study the identification of $\bar{\theta}$ based on both first and second order properties of the process.

DEFINITION 2. *The parameter vector $\bar{\theta}$ is said to be locally identifiable from the first and the second order properties of $\{Y_t\}$ at a point $\bar{\theta}_0$ if there exists an open neighborhood of $\bar{\theta}_0$ in which $\mu(\bar{\theta}_1) = \mu(\bar{\theta}_0)$ and $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$ implies $\bar{\theta}_0 = \bar{\theta}_1$.*

ASSUMPTION 4. *The elements of $\mu(\bar{\theta})$ are continuously differentiable with respect to $\bar{\theta}$. Let*

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

Assume there exists an open neighborhood of $\bar{\theta}_0$ in which $\bar{G}(\bar{\theta})$ has a constant rank.

---

[1]Under Gaussianity, $I(\theta_0)^{-1}$ is the asymptotic covariance matrix of the FQML estimator based on the full spectrum; see Section 4, in particular Theorem 3 and the expression (18) that follows.

REMARK 4. The dimension of $\bar{G}(\bar{\theta})$ is $(p+q) \times (p+q)$. The first term is a bordered matrix, consisting of $G(\theta)$ with p rows and columns of 0s appended to it. Both terms are positive semidefinite, hence taking the sum cannot decrease the rank. Also note that the (j,k)th element of $\bar{G}(\bar{\theta})$ is given by

$$\bar{G}_{jk}(\bar{\theta}) = \int_{-\pi}^{\pi} \mathrm{tr}\left\{ \frac{\partial f_\theta(\omega)}{\partial \bar{\theta}_j} \frac{\partial f_\theta(\omega)}{\partial \bar{\theta}_k} \right\} d\omega + \frac{\partial \mu(\bar{\theta})'}{\partial \bar{\theta}_j} \frac{\partial \mu(\bar{\theta})}{\partial \bar{\theta}_k}.$$

THEOREM 2. *Let Assumptions 1–4 hold. Then $\bar{\theta}$ is locally identifiable from the first and second order properties of $\{Y_t\}$ at a point $\bar{\theta}_0$ if and only if $\bar{G}(\bar{\theta}_0)$ is nonsingular.*

Theorems 1 and 2 can be further extended in various directions. In what follows, we discuss four such extensions.

DSGE models are often designed to explain business cycle movements, not very long run or very short run fluctuations. At the latter frequencies, such models can be severely misspecified. It is therefore important to consider estimation and inference based on business cycle frequencies only. Such consideration may also arise due to concerns about unmodeled seasonality or measurement errors; see Hansen and Sargent (1993), Diebold, Ohanian, and Berkowitz (1998), and Berkowitz (2001). We now present a result that lays the identification foundation for such an analysis. Let $W(\omega)$ denote an indicator function defined on $[-\pi, \pi]$ that is symmetric around 0 and equal to 1 over a finite number of closed intervals. Extend the definition of $W(\omega)$ to $\omega \in [\pi, 2\pi]$ by using $W(\omega) = W(2\pi - \omega)$.[2] Define the matrices

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

$$\bar{G}^W(\bar{\theta}) = \left\{ \int_{-\pi}^{\pi} W(\omega) \left(\frac{\partial \, \mathrm{vec}(f_\theta(\omega)')}{\partial \bar{\theta}'} \right)' \left(\frac{\partial \, \mathrm{vec}(f_\theta(\omega))}{\partial \bar{\theta}'} \right) d\omega \right\} + \frac{\partial \mu(\bar{\theta})'}{\partial \bar{\theta}} \frac{\partial \mu(\bar{\theta})}{\partial \bar{\theta}'}.$$

COROLLARY 2 (Identification From a Subset of Frequencies).

(i) *Let Assumptions 1–3 hold, but with $G(\theta)$ replaced by $G^W(\theta)$. Then θ is locally identifiable from the second order properties of $\{Y_t\}$ through the frequencies specified by $W(\omega)$ at a point θ_0 if and only if $G^W(\theta_0)$ is nonsingular.*

(ii) *Let Assumptions 1–4 hold, but with $\bar{G}(\bar{\theta})$ replaced by $\bar{G}^W(\bar{\theta})$. Then $\bar{\theta}$ is locally identifiable from the first and second order properties of $\{Y_t\}$ through the frequencies specified by $W(\omega)$ at a point $\bar{\theta}_0$ if and only if $\bar{G}^W(\bar{\theta}_0)$ is nonsingular.*

The proof is the same as for Theorems 1 and 2, because $W(\omega)$ is a nonnegative real valued function; therefore, it is omitted. Note that because the quantities

$$\left(\frac{\partial \, \mathrm{vec}(f_\theta(\omega)')}{\partial \theta'} \right)' \left(\frac{\partial \, \mathrm{vec}(f_\theta(\omega))}{\partial \theta'} \right)$$

---

[2]This extension is needed for FQML estimation since the objective function involves summation over $\omega_j = 2\pi/T, \ldots, 2\pi(T-1)/T$; see (15).

are positive semidefinite for any $\omega \in [-\pi, \pi]$, the difference $G(\theta_0) - G^W(\theta_0)$ is always positive semidefinite. This ensures that if θ_0 is identified using a subset of frequencies, it is also identified if considering the full spectrum. The converse does not necessarily hold. The same statement can be made about the relation between $\bar{G}(\bar{\theta}_0)$ and $\bar{G}^W(\bar{\theta}_0)$.

The second extension concerns the identification of a subset of parameters without making identification statements about the rest (partial identification). Specifically, let θ^s be a subset of parameters from θ. We say it is locally identified from the second order properties of $\{Y_t\}$ 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 $\theta_0^s = \theta_1^s$. 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 θ^s arbitrarily close in the sense of $\|\theta_0^s - \theta_1^s\|/\|\theta_0 - \theta_1\|$ being arbitrarily small. Analogously, we can define the identification of a subset of $\bar{\theta}$, say $\bar{\theta}^s$, based on the first and second order properties. The following result is a consequence of Theorem 8 in Rothenberg (1971), which can be traced back to Wald (1950) and Fisher (1966).

COROLLARY 3 (Partial Identification).

(i) *Let Assumptions 1–3 hold. Then θ^s is locally identifiable from the second order properties of $\{Y_t\}$ at a point θ_0^s if and only if $G(\theta_0)$ and*

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

have the same rank.

(ii) *Let Assumptions 1–4 hold. Then $\bar{\theta}^s$ is locally identifiable from the first and second order properties of $\{Y_t\}$ at a point $\bar{\theta}_0^s$ if and only if $\bar{G}(\bar{\theta}_0)$ and*

$$\bar{G}^a(\bar{\theta}_0) = \begin{bmatrix} \bar{G}(\bar{\theta}_0) \\ \partial\bar{\theta}_0^s/\partial\bar{\theta}' \end{bmatrix}$$

have the same rank.

The proof is given in the Appendix. Furthermore, one may be interested in studying the identification of a subset of parameters while keeping the values of the others fixed at θ_0 (conditional identification). The result for this extension is formally stated below.

COROLLARY 4 (Conditional Identification).

(i) *Let Assumptions 1–3 hold. Then a subvector of θ, θ^s, is conditionally locally identifiable from the second order properties of $\{Y_t\}$ at a point θ_0 if and only if*

$$G(\theta_0)^s = \int_{-\pi}^{\pi} \left(\frac{\partial \mathrm{vec}(f_{\theta_0}(\omega)')}{\partial \theta^{s\prime}} \right)' \left(\frac{\partial \mathrm{vec}(f_{\theta_0}(\omega))}{\partial \theta^{s\prime}} \right) d\omega$$

is nonsingular.

(ii) *Let Assumptions 1–4 hold. Then a subvector of $\bar{\theta}$, $\bar{\theta}^s$, is conditionally locally identifiable from the first and second order properties of $\{Y_t\}$ at a point $\bar{\theta}_0$ if and only if*

$$\bar{G}(\bar{\theta}_0)^s = \int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec}(f_{\theta_0}(\omega)')}{\partial \bar{\theta}^{s\prime}} \right)' \left(\frac{\partial \operatorname{vec}(f_{\theta_0}(\omega))}{\partial \bar{\theta}^{s\prime}} \right) d\omega + \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}^s} \frac{\partial \mu(\bar{\theta}_0)}{\partial \bar{\theta}^{s\prime}}$$

is nonsingular.

The proof is the same as for Theorems 1 and 2 because $G(\theta_0)^s$ and $\bar{G}(\bar{\theta}_0)^s$ have the same structure as $G(\theta_0)$ and $\bar{G}(\bar{\theta}_0)$, but with derivatives taken with respect to a subset of parameters. Therefore the detail is omitted. Comparison between Corollaries 3 and 4 suggests that the latter is often practically more relevant and its result is also simpler to interpret; we therefore expect it to be more frequently applied in practice.

Next, we consider identification under general constraints on the parameters. One potential example is that shocks to monetary variables have no long term effect on real variables, which can be formulated as a set of restrictions on the spectral density at frequency zero.

COROLLARY 5 (Identification Under General Constraints).

(i) *Let Assumptions 1–3 hold. Suppose θ_0 satisfies $\psi(\theta_0) = 0$ with $\psi(\theta)$ a $k \times 1$ constraint vector continuously differentiable in θ. Define the Jacobian matrix $\Psi(\theta)$ with the (j, l)th element given by*

$$\Psi_{jl}(\theta) = \partial \psi_j(\theta)/\partial \theta_l.$$

Suppose θ_0 is a regular point of both $G(\theta)$ and $\Psi(\theta)$. Then θ satisfying $\psi(\theta) = 0$ is locally identified from the second order properties of $\{Y_t\}$ at a point θ_0 if and only if

$$\begin{bmatrix} G(\theta_0) \\ \Psi(\theta_0) \end{bmatrix}$$

has full column rank equal to q.

(ii) *Let Assumptions 1–4 hold and let the other conditions stated in part* (i) *of this corollary hold with θ replaced by $\bar{\theta}$. Then $\bar{\theta}$ satisfying $\psi(\bar{\theta}) = 0$ is locally identified from the first and second order properties of $\{Y_t\}$ at a point $\bar{\theta}_0$ if and only if*

$$\begin{bmatrix} \bar{G}(\bar{\theta}_0) \\ \bar{\Psi}(\bar{\theta}_0) \end{bmatrix}$$

has rank $(q + p)$.

Note that Corollary 5 can also be used to study conditional identification, because the latter is a special case of simple linear restrictions. However, Corollary 4 is simpler to apply, especially if the dimension of θ^s is much smaller compared to that of θ. Clearly, Corollaries 3–5 can be applied in conjunction with Corollary 2 to study identification through a subset of frequencies.

We now compare the above analysis with those of Iskrev (2010) and Komunjer and Ng (2011). Iskrev (2010) suggested to identify the parameters from the mean and the first T autocovariances of the observables. Because his result (Theorem 2) assumes T is finite, the resulting conditions are sufficient but not necessary. Meanwhile, the key differences between our work and Komunjer and Ng (2011) can be summarized along five aspects. First, the perspective is different. Komunjer and Ng (2011) regarded the solution of a DSGE model as a minimal system with miniphase. Their condition effectively exploits the implication of the latter two features for identification. Instead, we regard the spectrum of a DSGE model as an infinite dimensional mapping. The analysis studies its property under local perturbation of the structural parameter vector. Second, the assumption is different. We do not require the solution system to have minimal phase. Therefore, we permit the rank of the spectral density matrix to vary across frequencies. This is practically relevant. For example, in Smets and Wouters (2007), the rank of the spectral density is lower at frequency zero because the first differences of stationary variables are considered. Third, the system representation requirement is different. Komunjer and Ng (2011) required a minimal state representation, while we do not. Whatever is the state representation under which the model is solved (S_t in the GENSYS algorithm, for example), the spectral density can be computed and that is all that is needed. Fourth, the treatment of stochastic singularity is different. Komunjer and Ng (2011) gave separate results for singular and nonsingular systems, while our single condition applies to both. Intuitively, this follows because the dimension of our criterion function is independent of those of the observation vector and the vector of innovations, but only depends on that of the structural parameter vector. Finally, the computation is different. Although both methods require numerical differentiation, it is applied to different objects. In Komunjer and Ng (2011), it is applied to the coefficient matrices in the state space representation, while in our case, we compute the derivative of the spectral density with respect to the structural parameter vector.

3.1 *Tracing out nonidentification curves*

In this section, the discussion focuses on θ because for $\bar{\theta}$ the procedure works in the same way. Suppose Theorem 1 or Corollary 2 shows that θ is locally unidentifiable.

First, consider the simple case where $G(\theta_0)$ has only one zero eigenvalue. Let $c(\theta_0)$ be a corresponding real eigenvector satisfying $\|c(\theta_0)\| = 1$. Then $c(\theta_0)$ is unique up to multiplication by -1 and thus can be made unique by restricting its first nonzero element to be positive. This restriction is imposed in the subsequent analysis. Let $\delta(\theta_0)$ be an open neighborhood of θ_0. Under Assumptions 1–3, $G(\theta)$ is continuous and has only one zero eigenvalue in $\delta(\theta_0)$, while $c(\theta)$ is continuous in $\delta(\theta_0)$. As in Rothenberg (1971), define a curve χ using the function $\theta(v)$, which solves the differential equation

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

where v is a scalar that varies in a neighborhood of 0 such that $\theta(v) \in \delta(\theta_0)$. Then along χ, θ is not identified at θ_0 because

$$\frac{\partial \operatorname{vec}(f_{\theta(v)}(\omega))}{\partial v} = \frac{\partial \operatorname{vec}(f_{\theta(v)}(\omega))}{\partial \theta(v)'} c(\theta) = 0 \qquad (10)$$

for all $\omega \in [-\pi, \pi]$, where the last equality uses Assumption 3 and the fact that $c(\theta)$ is the eigenvector corresponding to the zero eigenvalue (cf. (A.3) in the Appendix). We call χ the nonidentification curve.

Clearly, this curve is continuous in v. It is also locally unique, in the sense that there does not exist another continuous curve containing θ_0 and satisfying $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$. We state this result as a corollary:

COROLLARY 6. *Let Assumptions 1–3 hold and let* $\operatorname{rank}(G(\theta_0)) = q - 1$. *Then, in a small neighborhood of* θ_0, *there exists precisely one curve passing through* θ_0 *that satisfies* $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ *for all* $\omega \in [-\pi, \pi]$.

Corollary 6 is not a trivial result because it involves infinite dimensional maps. The key idea in the proof is to reduce the problem to a finite dimensional one by considering projections of $f_\theta(\cdot)$ associated with finite partitions of $[-\pi, \pi]$. Then a standard constant rank theorem can be applied. The details of the proof are in the Appendix.

The nonidentification curve can be evaluated numerically in various ways. The simplest example is the Euler method. First, obtain $c(\theta_0)$ as described above. Then compute recursively

$$\begin{aligned}
\theta(v_{j+1}) &\approx \theta(v_j) + c(\theta(v_j))(v_{j+1} - v_j), \quad v_{j+1} \geq v_j \geq 0, j = 0, 1, \ldots, \\
\theta(v_{j-1}) &\approx \theta(v_j) + c(\theta(v_j))(v_{j-1} - v_j), \quad v_{j-1} \leq v_j \leq 0, j = 0, -1, \ldots,
\end{aligned} \qquad (11)$$

where $|v_{j+1} - v_j|$ is the step size, which can be set to some small constant, say h. The associated approximation error in each step is of $O(h^2)$ if $\theta(v)$ has bounded first and second derivatives. Therefore, the cumulative error over a finite interval is $O(h)$. It is important to note that because $\delta(\theta_0)$ is usually unknown, so is the domain of the curve. However, this is not a problem in practice, because we can first obtain a curve over a wide support, then resolve the model and compute the spectral density using points on this curve. The curve can then be truncated to exclude the points that violate determinacy, the natural bounds of the parameters (e.g., the discount rate, stationary autoregressive coefficients), and those yielding $f_\theta(\omega)$ different from $f_{\theta_0}(\omega)$.

Next, consider the case where $G(\theta_0)$ has multiple zero eigenvalues. Then, in general, there exists an infinite number of curves satisfying (10), because any linear combination of the eigenvectors points to a direction of nonidentification. It is not useful to try reporting all such curves. To see this, suppose $\theta_0 = (\theta_0^1, \theta_0^2)'$ and that changing θ^1 along a certain curve χ_1 while keeping θ^2 fixed at θ_0^2 yields identical spectral densities. Also suppose the same property holds when we vary θ^2 and fix θ^1 at θ_0^1, yielding a curve χ_2. Suppose the rank of $G(\theta)$ stays constant in a local neighborhood of θ_0. Then changing θ^1 and θ^2 simultaneously can also generate new curves and there are infinitely many of

them. In this example, χ_1 and χ_2 contain essentially all the information, as the rest of the curves are derived from them, and thus it suffices to report only two of them. Motivated by the above observation, we propose a simple four-step procedure that delivers a finite number of nonidentification curves. The key idea underlying this procedure is to distinguish between separate sources of nonidentification by using Corollary 4. More specifically, we apply the rank condition recursively to subsets of parameters to find the ones that are not identified and depict their observationally equivalent values using curves.

Step 1. Apply Theorem 1 to verify whether all the parameters in the model are locally identified. Proceed to Step 2 if lack of identification is detected.

Step 2. Apply Corollary 4 to each individual parameter. If a zero eigenvalue of $G(\theta)^s$ evaluated at θ_0 is found, then it implies that the corresponding parameter is not locally conditionally identified. Apply the procedure outlined above to obtain a nonidentification curve (changing only this element and fixing the value of the others at θ_0). Repeating this for all individual parameters, we obtain a finite number of curves, with each curve being a scalar valued function of v.

Step 3. Increase the number of parameters in the considered subsets of θ_0 by one at a time. Single out the subsets with the following two properties: (i) it does not include the subset detected in previous steps as a proper subset and (ii) when applying Corollary 4, it reports only one zero eigenvalue. Repeat the procedure outlined above for all such subsets to obtain nonidentification curves. Note that if the subset has k elements, then the associated curve is a $k \times 1$ vector valued function of v.

Step 4. Continue Step 3 until all subsets are considered. Solve the model using parameter values from the curves to determine the appropriate domain for v. Truncate the curves obtained in Steps 1–4 accordingly.

Step 2 returns nonidentification curves resulting from changing only one element in the parameter vector. In Step 3, the number of elements is increased sequentially. For each iteration, the algorithm first singles out parameter subvectors whose elements are not separately identified. Then only subvectors satisfying the two properties outlined in Step 3 are further considered. The first property is to rule out redundancy because if a k-element subset constitutes a nonidentification curve, including any additional element (fixing its value or varying it if it itself is not conditionally identifiable) by definition constitutes another such curve, but it conveys no additional information. The second property serves the same purpose, because if some subvector yields a $G(\theta)^s$ with multiple zero eigenvalues, then it must be a union of subvectors identified in previous steps and containing fewer elements. To see that this is necessarily the case, suppose that for a given subvector, two zero eigenvalues are reported. Then there exists a linear combination of the two corresponding eigenvectors that makes the first element of the resulting vector zero. Similarly, there is a combination that makes the second element zero. The two resulting vectors are valid eigenvectors; however, they correspond to lower dimensional subvectors of θ. Now apply Corollary 4 to these two subvectors. If single zero eigenvalues are reported, then it implies that they have already been considered in the previous steps. Otherwise, the dimension of the subvectors can be further

reduced by using the same argument, eventually leading to the conclusion that they have been previously considered. The general case with more than two zero eigenvalues can be analyzed similarly.

In Steps 3 and 4, we do not remove any parameter from θ after nonidentification curves are found. Otherwise, we may fail to detect some curves. To see this, suppose $\theta \in \mathbb{R}^4$, and that the subvectors (θ_1, θ_2) and $(\theta_1, \theta_3, \theta_4)$ form two nonidentification curves. If we removed θ_1 and θ_2 from θ after considering two-parameter subsets, then we would miss $(\theta_1, \theta_3, \theta_4)$. Finally, in Step 4, the truncation narrows down the domain of the nonidentification curve, which can be used, for example, to exclude parameter values that are incompatible with the economic theory. This is computationally simple to implement in practice because the domain of any curve is always one dimensional. For illustration, consider the curve $(\theta_1(v), \theta_2(v))$ and suppose that the economic theory requires the value of θ_1 to be nonnegative. Then we simply chop off those v with $\theta_1(v) \leq 0$. If the theory also imposes restriction on θ_2, then we simply drop those v over which at least one restriction is violated.

This procedure delivers a finite number of curves with the following two features. First, the curves are minimal in the sense that, for each curve, all elements in the corresponding subvector have to change to generate nonidentification. Fixing the value of any element shrinks the corresponding curve to a single point. Second, the curves are sufficient in the sense that, for any subvector that can generate a nonidentification curve passing through θ_0, it or one of its subsets are already included. Finally, the procedure is simple to implement because it mainly involves repeated applications of Corollary 4. This simplicity is achieved because we start with the lowest dimension, thus there is no need to directly handle the situation with multiple zero eigenvalues. It should also be noted that, apart from evaluating the nonidentification curves, the procedure is not computationally demanding. Once $G(\theta)$ is computed in Step 1, the $G(\theta)^s$ for any subvector considered can be obtained by simply picking out relevant elements of $G(\theta)$ (cf. Remark 3). Specifically, suppose we are interested in a particular k-element subvector of θ. If we number parameters inside θ, and let Φ be a set of parameter numbers of interest (i.e., if we want to vary only parameters 1, 2, and 5, then $\Phi = \{1, 2, 5\}$), then the (i, j)th element of $G(\theta)^s$ is given by

$$G(\theta)_{i,j}^s = G(\theta)_{\Phi_i, \Phi_j}, \quad i = 1, 2, \ldots, k; j = 1, 2, \ldots, k. \tag{12}$$

Also note that in the case of Theorem 2, the same logic applies to the term $[\partial\mu(\bar{\theta}_0)'/\partial\bar{\theta}^s] \times [\partial\mu(\bar{\theta}_0)/\partial\bar{\theta}^{s'}]$, that is, having computed it once, one can repeatedly apply Corollary 4 by selecting relevant elements from it and $\overline{G}(\bar{\theta})^s$ in the same fashion as in (12).

3.2 *An illustrative example*

To provide a frame of reference, we consider a DSGE model from An and Schorfheide (2007) whose identification also was studied by Komunjer and Ng (2011). We consider identification based on the (first and) second order properties and also obtain nonidentification curves.

The log-linearized solutions are given by

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

$$\pi_t = \beta E_t \pi_{t+1} + \frac{\tau(1-\nu)}{\nu \overline{\pi}^2 \phi}(y_t - g_t),$$

$$c_t = y_t - g_t,$$

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_1 \pi_t + (1-\rho_r)\psi_2(y_t - g_t) + e_{rt},$$

$$g_t = \rho_g g_{t-1} + \epsilon_{gt},$$

$$z_t = \rho_z z_{t-1} + \epsilon_{zt},$$

where $e_{rt} = \epsilon_{rt}$, $\epsilon_{rt} \sim WN(0, \sigma_r^2)$, $\epsilon_{gt} \sim WN(0, \sigma_g^2)$, and $\epsilon_{zt} \sim WN(0, \sigma_z^2)$ are mutually uncorrelated shocks, and $\overline{\pi}$ is the steady state inflation rate. The vector of parameters to be identified is

$$\theta = (\tau, \beta, \nu, \phi, \overline{\pi}^2, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r^2, \sigma_g^2, \sigma_z^2).$$

We use parameter values[3]

$$\theta_0 = (2, 0.9975, 0.1, 53.6797, 1.008^2, 1.5, 0.125, 0.75, 0.95, 0.9, 0.4, 3.6, 0.9),$$

as given in Table 3 of An and Schorfheide (2007).

We first describe how to compute the spectrum for a given parameter vector. We can write the model as in (5) with

$$S_t = (z_t, g_t, r_t, y_t, \pi_t, c_t, E_t(\pi_{t+1}), E_t(y_{t+1}))'. \tag{13}$$

The exact formulations of the matrices Γ_0, Γ_1, Ψ, and Π are omitted here.[4] We use the GENSYS algorithm provided by Sims (2002) to obtain the model solution numerically in the form of (6), specifically

$$S_t = \Theta_1 S_{t-1} + \Theta_0 \epsilon_t,$$

where Θ_1 and Θ_0 are functions of θ. The spectral density, as noted before, can then be computed using (4) with

$$H(L; \theta) = A(L)(I - \Theta_1 L)^{-1}\Theta_0.$$

---

[3]Note that we scale the values for the variances $(\sigma_r^2, \sigma_g^2, \sigma_z^2)$ from An and Schorfheide (2007) by 10^5. This scaling is merely to ensure numerical stability and does not affect any of our conclusions.

[4]Please refer to the MATLAB code available from the authors' web pages for details.

Given the S_t in (13) and $Y_t^d = (r_{t-1}, y_t, \pi_t, c_t)'$, the matrix $A(L)$ is given by[5]

$$\begin{pmatrix} 0 & 0 & L & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Note that the results in this example do not rely on using the solution algorithm of Sims (2002). Other algorithms considered in the literature (e.g., that in Uhlig (1999)), can be used to obtain the same conclusions. The algorithm will produce the P, Q, R, S representation as in (7), with $k_{t+1} = r_t$, $w_t = (y_t, \pi_t, c_t)'$, and $z_t = (e_{rt}, g_t, z_t)'$. The spectrum can then be computed as in (8).

3.2.1 *Analysis based on the second order properties.* To compute $G(\theta_0)$, the integral in $G(\theta_0)$ is approximated numerically by averaging over 10,000 Fourier frequencies from $-4999\pi/5000$ to $4999\pi/5000$ and multiplying by 2π. The results reported are robust to varying the number of frequencies between 5000 and 10,000. The step size for the numerical differentiation[6] is set to $10^{-7} \times \theta_0$. The rank of $G(\theta_0)$ is computed as the number of nonzero eigenvalues, using the MATLAB default tolerance set at $\text{tol} = \text{size}(G)\,\text{eps}(\|G\|)$, where eps is the floating point precision of G. We obtain $\text{rank}(G(\theta_0)) = 10$. Because $q = 13$, this means that the entire parameter vector cannot be identified from the spectrum. In addition, this suggests that three parameters have to be fixed to achieve identification.

Since the model is not identified, we can follow the procedure outlined in Section 3.1 to pinpoint the sources of nonidentification. In Step 2, we apply Corollary 4 to all one-element subsets of θ which, as noted above in (12), simply amounts to checking whether any diagonal elements of $G(\theta_0)$ are zero. None is found, hence we continue to Step 3 and consider all two-element subvectors of θ. We find three subvectors that yield $G^s(\theta_0)$ with one zero eigenvalue: (ν, ϕ), $(\nu, \overline{\pi}^2)$, and $(\phi, \overline{\pi}^2)$. This finding is very intuitive, since all of these parameters enter the slope of the Phillips curve equation and thus are not separately identifiable, as noted by An and Schorfheide (2007). We do not report the non-identification curves for these cases, as they are trivial and can be eliminated by reparameterizing the model with $\kappa \equiv \tau(1 - \nu)/(\nu\overline{\pi}^2\phi)$ as a new parameter instead. However, highlighting them does play a useful part in illustrating our procedure at work.

Before we continue, we exclude all three-parameter subvectors that contain either of the three nonidentification sets identified above as proper subsets. Considering all remaining three-element subvectors of θ yields no new nonidentification sets. However, there is one four-element subvector which has one zero eigenvalue:

$$(\psi_1, \psi_2, \rho_r, \sigma_r^2).$$

Interestingly, all of these parameters enter the Taylor rule equation in the model.

---

[5]Considering r_t instead of r_{t-1} in Y_t^d yields the same result. We only need to replace the lag operator in the first row of $A(L)$ by 1. Such a feature is true in general.

[6]A simple two-point method is used. In our experience, using higher order methods did not change the conclusions.

(a) ψ_1

(b) ψ_2

(c) ρ_r

(d) σ_r^2
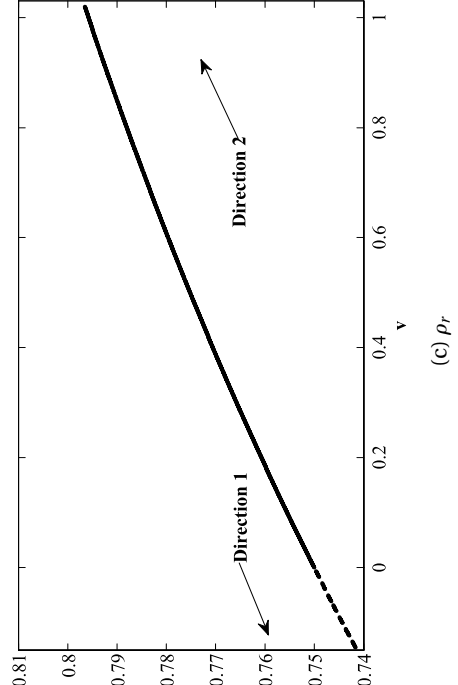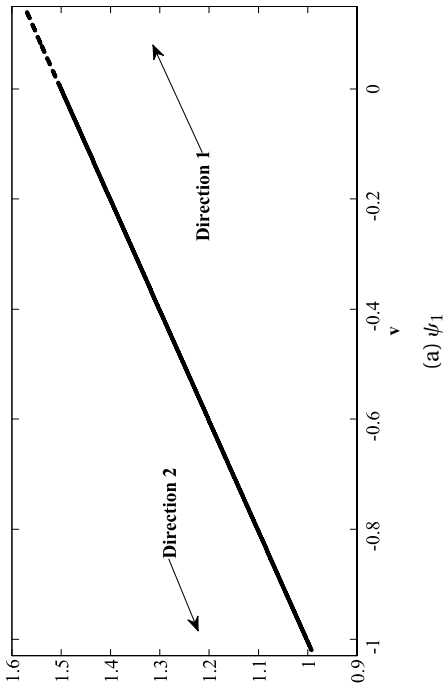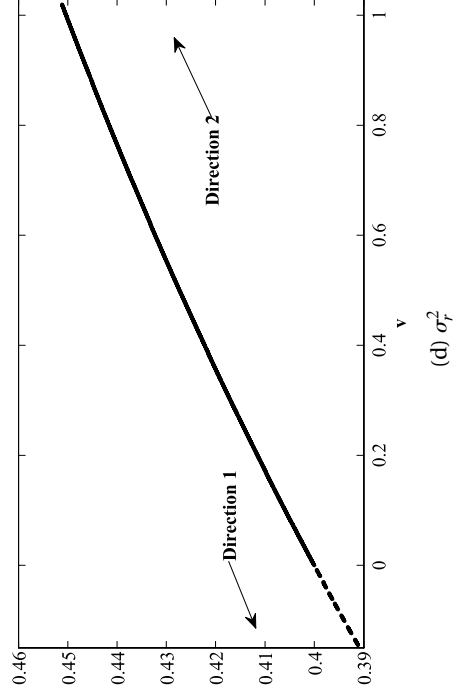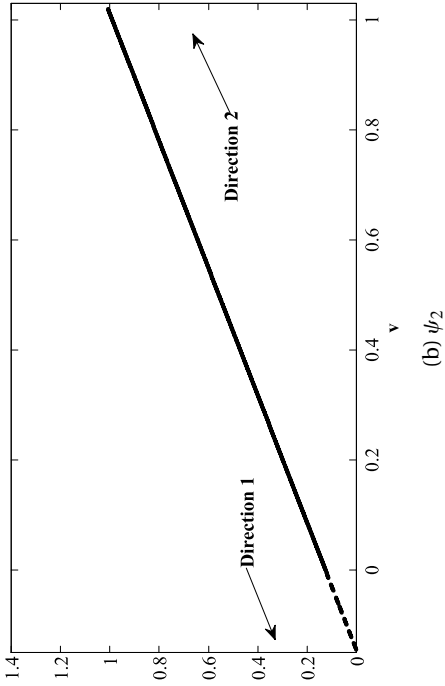
Having excluded all subvectors containing the nonidentification parameter sets above and repeating Step 4 with more parameters, we do not find any more sources of nonidentification in this model. The result implies that to achieve identification, it is necessary and sufficient to fix two parameters out of ν, ϕ and $\overline{\pi}^2$, and one parameter out of ψ_1, ψ_2, ρ_r, and σ_r^2.

The above finding is further confirmed when we repeat the exercise by considering a reparameterization of the model with κ as defined above: θ is still not identified and $G(\theta_0)$ has only one zero eigenvalue. Note that the reparameterization amounts to fixing two parameters out of ν, ϕ, and $\overline{\pi}^2$. This leaves only one direction of nonidentification, which turns out to be, not surprisingly, along the $(\psi_1, \psi_2, \rho_r, \sigma_r^2)$ subvector.

We then proceed to evaluate the nonidentification curve, consisting of combinations of ψ_1, ψ_2, ρ_r, and σ_r^2, using the Euler method with step size $h = 10^{-5}$ in a small neighborhood around θ_0. The result is presented in Figure 1. The figure shows the nonidentification curve pertaining to each parameter. The initial value is at θ_0 and the curve is extended in each direction using (11). The directions are marked on the graph by bold and dotted lines. Note that ψ_2, which governs the output weight in the Taylor rule and must be nonnegative, is decreasing along direction 1. Therefore, we truncate the curve at the closest point to zero where ψ_2 is still positive. Along direction 2, we reach an indeterminacy region before any natural bounds on parameter values are violated, and hence truncate the curve at the last point that yields a determinate solution. Therefore, this case also provides an illustration of how to narrow down the domain of the nonidentification curve in practice.

To give a quantitative idea of the parameter values on the curve, we also present a sample of values from various points on the curve in Table 1. Specifically, 10 points were taken at regularly spaced intervals from θ_0 in the positive and negative direction.

Of course it is necessary to verify that the points on the curve result in identical spectral densities. We do this by computing the $f_\theta(\omega)$ at half of the Fourier frequencies used in the computation of $G(\theta_0)$ (i.e., 5000 frequencies between 0 and π)[7] for each point on the curve and then compare it to the ones computed at θ_0. Due to numerical error involved in solving the model, the computation of the G matrix, and the approximation method for the differential equation, small discrepancies between the spectra computed

---

FIGURE 1. The nonidentification curve $(\psi_1, \psi_2, \rho_r, \sigma_r^2)$. The nonidentification curve is given by $\partial \theta(v)/\partial v = c(\theta)$, $\theta(0) = \theta_0$, where $c(\theta)$ is the eigenvector corresponding to the only zero eigenvalue of $G(\theta)$. The curve is computed recursively using the Euler method, so that $\theta(v_{j+1}) = \theta(v_j) + c(\theta(v_j))h$, where h is the step size, fixed at 1e–05. The variables $(\psi_1, \psi_2, \rho_r, \sigma_r^2)$ change simultaneously along the curve in the indicated directions. Directions 1 and 2 are obtained by restricting the first element of $c(\theta)$ to be positive or negative, respectively. Since a negative Taylor rule weight contradicts economic theory, direction 1 is truncated at the last point where ψ_2 is nonnegative. Direction 2 is truncated at the boundary of the determinacy region. Consequently, the curve is extended from θ_0 for 14,475 steps in direction 1 and for 101,972 steps in direction 2.

[7]There is no need to consider $\omega \in [-\pi, 0]$ because $f_\theta(\omega)$ is equal to the conjugate of $f_\theta(-\omega)$.

Table 1. Parameter values and the corresponding two smallest eigenvalues along the nonidentification curve.[a]

	ψ_1	ψ_2	ρ_r	σ_r^2	λ_1	λ_2
θ_0	1.500000	0.125000	0.750000	0.400000	7.09E–10	3.251348
Panel (a): Direction 1						
θ_1	1.507156	0.112571	0.749192	0.399139	1.47E–10	3.266554
θ_2	1.514316	0.100134	0.748378	0.398272	4.73E–10	3.281960
θ_3	1.521476	0.087698	0.747559	0.397401	9.56E–10	3.297558
θ_4	1.528636	0.075262	0.746735	0.396525	1.15E–09	3.313348
θ_5	1.535796	0.062827	0.745905	0.395644	5.33E–10	3.329337
θ_6	1.542955	0.050392	0.745070	0.394758	1.79E–09	3.345526
θ_7	1.550114	0.037958	0.744229	0.393868	1.90E–09	3.361918
θ_8	1.557272	0.025524	0.743383	0.392973	1.82E–10	3.378520
θ_9	1.564431	0.013091	0.742531	0.392073	1.80E–09	3.395333
θ_{10}	1.571589	0.000659	0.741674	0.391168	1.79E–10	3.412362
Panel (b): Direction 2						
θ_{-1}	1.449285	0.213085	0.755581	0.405975	2.19E–10	3.148993
θ_{-2}	1.398558	0.301193	0.760920	0.411732	1.30E–11	3.054759
θ_{-3}	1.347819	0.389321	0.766031	0.417282	5.23E–13	2.967750
θ_{-4}	1.297070	0.477467	0.770930	0.422636	1.12E–12	2.887193
θ_{-5}	1.246311	0.565629	0.775628	0.427803	3.63E–12	2.812419
θ_{-6}	1.195543	0.653807	0.780138	0.432793	6.18E–12	2.742843
θ_{-7}	1.144767	0.741998	0.784471	0.437615	3.12E–12	2.677957
θ_{-8}	1.093985	0.830202	0.788638	0.442275	3.33E–12	2.617315
θ_{-9}	1.043195	0.918417	0.792647	0.446783	4.15E–12	2.560521
θ_{-10}	0.992400	1.006643	0.796507	0.451145	3.76E–12	2.507230

[a]The parameters θ_j represent equally spaced points taken from the nonidentification curve extended from θ_0 for 14,475 steps in direction 1 and for 101,972 steps in direction 2. The quantities λ_1 and λ_2 represent the smallest and the second smallest eigenvalues of $G(\theta_i)^s$, respectively. The step size of the approximation is 10^{-5}. Along direction 1, the curve is truncated at the closest point to zero where ψ_2 is still positive, as it determines the output weight in the Taylor rule and must be nonnegative. Along direction 2, the curve is truncated at the last point yielding a determinate solution. Results are rounded to the nearest sixth digit to the right of decimal.

at θ_0 and the points on the curve should be expected. We therefore consider three different measures of the discrepancies (let $f_{\theta hl}(\omega)$ denote the (h, l)th element of the spectral density matrix with parameter θ and let Ω be the set that includes the 5000 frequencies between 0 and π):

$$\text{maximum absolute deviation}: \quad \max_{\omega_j \in \Omega} |f_{\theta hl}(\omega_j) - f_{\theta_0 hl}(\omega_j)|,$$

$$\text{maximum absolute deviation in relative form}: \quad \frac{\max_{\omega_j \in \Omega} |f_{\theta hl}(\omega_j) - f_{\theta_0 hl}(\omega_j)|}{|f_{\theta_0 hl}(\omega_j)|},$$

$$\text{maximum relative deviation}: \quad \max_{\omega_j \in \Omega} \frac{|f_{\theta hl}(\omega_j) - f_{\theta_0 hl}(\omega_j)|}{|f_{\theta_0 hl}(\omega_j)|}.$$

Note that when computing the second measure, the denominator is evaluated at the same frequency that maximizes the numerator. To save space, we only report results for

TABLE 2. Deviations of spectra across frequencies (direction 1).[a]

| | | | | Spectral Density Matrix Element Number | | | | | |
	$(1,1)$	$(2,1)$	$(3,1)$	$(4,1)$	$(2,2)$	$(3,2)$	$(4,2)$	$(3,3)$	$(4,3)$	$(4,4)$
					Measure 1: Maximum absolute deviations across frequencies					
θ_1	1.49E–07	1.68E–08	9.85E–08	1.68E–08	1.99E–08	1.26E–08	1.99E–08	5.80E–08	1.26E–08	1.99E–08
θ_2	2.96E–07	3.40E–08	1.97E–07	3.40E–08	3.98E–08	2.52E–08	3.98E–08	1.16E–07	2.52E–08	3.98E–08
θ_3	4.43E–07	5.11E–08	2.94E–07	5.11E–08	5.83E–08	3.68E–08	5.83E–08	1.75E–07	3.68E–08	5.83E–08
θ_4	5.93E–07	7.13E–08	3.97E–07	7.13E–08	7.76E–08	4.87E–08	7.76E–08	2.36E–07	4.87E–08	7.76E–08
θ_5	7.35E–07	8.51E–08	4.88E–07	8.51E–08	9.78E–08	6.18E–08	9.78E–08	2.89E–07	6.18E–08	9.78E–08
θ_6	8.82E–07	1.02E–07	5.86E–07	1.02E–07	1.18E–07	7.43E–08	1.18E–07	3.47E–07	7.43E–08	1.18E–07
θ_7	1.04E–06	1.24E–07	6.92E–07	1.24E–07	1.37E–07	8.64E–08	1.37E–07	4.11E–07	8.64E–08	1.37E–07
θ_8	1.19E–06	1.37E–07	7.88E–07	1.37E–07	1.59E–07	1.01E–07	1.59E–07	4.64E–07	1.01E–07	1.59E–07
θ_9	1.34E–06	1.57E–07	8.91E–07	1.57E–07	1.79E–07	1.13E–07	1.79E–07	5.27E–07	1.13E–07	1.79E–07
θ_{10}	1.49E–06	1.76E–07	9.94E–07	1.76E–07	1.99E–07	1.25E–07	1.99E–07	5.89E–07	1.25E–07	1.99E–07
					Measure 2: Maximum absolute deviations across frequencies in relative form					
θ_1	6.66E–09	2.11E–09	7.03E–09	2.11E–09	8.19E–10	7.02E–09	9.83E–09	6.34E–09	7.02E–09	9.83E–09
θ_2	1.32E–08	4.28E–09	1.40E–08	4.28E–09	1.64E–09	1.40E–08	1.97E–08	1.26E–08	1.40E–08	1.97E–08
θ_3	1.98E–08	6.43E–09	2.10E–08	6.43E–09	2.44E–09	2.06E–08	2.89E–08	1.91E–08	2.06E–08	2.89E–08
θ_4	2.65E–08	8.97E–09	2.83E–08	8.97E–09	3.32E–09	2.75E–08	3.87E–08	2.58E–08	2.75E–08	3.87E–08
θ_5	3.28E–08	1.07E–08	3.48E–08	1.07E–08	4.08E–09	3.45E–08	4.85E–08	3.15E–08	3.45E–08	4.85E–08
θ_6	3.94E–08	1.29E–08	4.18E–08	1.29E–08	4.91E–09	4.15E–08	5.83E–08	3.78E–08	4.15E–08	5.83E–08
θ_7	4.62E–08	1.56E–08	4.93E–08	1.56E–08	5.80E–09	4.85E–08	6.83E–08	4.49E–08	4.85E–08	6.83E–08
θ_8	5.29E–08	1.73E–08	5.62E–08	1.73E–08	6.60E–09	5.62E–08	7.89E–08	5.07E–08	5.62E–08	7.89E–08
θ_9	5.98E–08	1.97E–08	6.35E–08	1.97E–08	7.46E–09	6.31E–08	8.87E–08	5.75E–08	6.31E–08	8.87E–08
θ_{10}	6.66E–08	2.22E–08	7.09E–08	2.22E–08	8.34E–09	7.01E–08	9.86E–08	6.43E–08	7.01E–08	9.86E–08
					Measure 3: Maximum relative deviations across frequencies					
θ_1	7.57E–09	3.01E–08	2.01E–08	3.01E–08	4.64E–09	9.15E–09	1.20E–08	6.34E–09	9.15E–09	1.20E–08
θ_2	1.48E–08	6.36E–08	4.14E–08	6.36E–08	9.33E–09	1.83E–08	2.41E–08	1.26E–08	1.83E–08	2.41E–08
θ_3	2.25E–08	8.82E–08	5.91E–08	8.82E–08	1.36E–08	2.68E–08	3.53E–08	1.91E–08	2.68E–08	3.53E–08
θ_4	2.96E–08	1.27E–07	8.27E–08	1.27E–07	1.82E–08	3.56E–08	4.72E–08	2.58E–08	3.56E–08	4.72E–08
θ_5	3.69E–08	1.54E–07	1.01E–07	1.54E–07	2.29E–08	4.50E–08	5.93E–08	3.15E–08	4.50E–08	5.93E–08
θ_6	4.42E–08	1.89E–07	1.23E–07	1.89E–07	2.76E–08	5.41E–08	7.13E–08	3.78E–08	5.41E–08	7.13E–08
θ_7	5.13E–08	2.31E–07	1.48E–07	2.31E–07	3.23E–08	6.31E–08	8.34E–08	4.49E–08	6.31E–08	8.34E–08
θ_8	5.91E–08	2.60E–07	1.68E–07	2.60E–07	3.74E–08	7.33E–08	9.66E–08	5.07E–08	7.33E–08	9.66E–08
θ_9	6.67E–08	2.92E–07	1.89E–07	2.92E–07	4.20E–08	8.22E–08	1.08E–07	5.75E–08	8.22E–08	1.08E–07
θ_{10}	7.42E–08	3.28E–07	2.12E–07	3.28E–07	4.67E–08	9.13E–08	1.21E–07	6.43E–08	9.13E–08	1.21E–07

[a]The parameters θ_1 to θ_{10} are as defined in Table 1. The upper triangular elements are omitted due to symmetry.

the points in Table 1, as the rest are very similar. Both Tables 2 and 3 show that even the largest observed deviations are quite modest (recall that the Euler method involves a cumulative approximation error that is of the same order as the step size, in this case 10^{-5}). This confirms that the spectral density is constant along the curve.

Note that all four parameters in $(\psi_1, \psi_2, \rho_r, \sigma_r^2)$ have to change simultaneously to generate nonidentification. This can be further verified as follows. Suppose fixing σ_r^2 still leaves (ψ_1, ψ_2, ρ_r) unidentified. Then this subvector should generate a nonidentification curve. However, using the procedure outlined above yields a curve, the points on which produce much larger deviations from $f_{\theta_0}(\omega)$ than those reported in Tables 2

TABLE 3. Deviations of spectra across frequencies (direction 2).[a]

	(1, 1)	(2, 1)	(3, 1)	(4, 1)	(2, 2)	(3, 2)	(4, 2)	(3, 3)	(4, 3)	(4, 4)

Spectral Density Matrix Element Number

Measure 1: Maximum absolute deviations across frequencies

	(1, 1)	(2, 1)	(3, 1)	(4, 1)	(2, 2)	(3, 2)	(4, 2)	(3, 3)	(4, 3)	(4, 4)
θ_{-1}	8.49E–07	8.20E–08	5.00E–07	8.20E–08	1.45E–07	9.87E–08	1.45E–07	2.52E–07	9.87E–08	1.45E–07
θ_{-2}	1.69E–06	1.59E–07	1.01E–06	1.59E–07	2.75E–07	1.86E–07	2.75E–07	5.28E–07	1.86E–07	2.75E–07
θ_{-3}	2.52E–06	2.34E–07	1.53E–06	2.34E–07	3.95E–07	2.64E–07	3.95E–07	8.18E–07	2.64E–07	3.95E–07
θ_{-4}	3.35E–06	3.07E–07	2.06E–06	3.07E–07	5.04E–07	3.34E–07	5.04E–07	1.13E–06	3.34E–07	5.04E–07
θ_{-5}	4.17E–06	3.83E–07	2.60E–06	3.83E–07	6.02E–07	3.96E–07	6.02E–07	1.46E–06	3.96E–07	6.02E–07
θ_{-6}	4.99E–06	4.64E–07	3.16E–06	4.64E–07	6.91E–07	4.50E–07	6.91E–07	1.80E–06	4.50E–07	6.91E–07
θ_{-7}	5.80E–06	5.58E–07	3.72E–06	5.58E–07	7.72E–07	4.98E–07	7.72E–07	2.17E–06	4.98E–07	7.72E–07
θ_{-8}	6.62E–06	6.76E–07	4.30E–06	6.76E–07	8.44E–07	5.39E–07	8.44E–07	2.55E–06	5.39E–07	8.44E–07
θ_{-9}	7.43E–06	8.17E–07	4.89E–06	8.17E–07	9.10E–07	5.75E–07	9.10E–07	2.95E–06	5.75E–07	9.10E–07
θ_{-10}	8.26E–06	9.74E–07	5.50E–06	9.74E–07	9.67E–07	6.04E–07	9.67E–07	3.38E–06	6.04E–07	9.67E–07

Measure 2: Maximum absolute deviations across frequencies in relative form

	(1, 1)	(2, 1)	(3, 1)	(4, 1)	(2, 2)	(3, 2)	(4, 2)	(3, 3)	(4, 3)	(4, 4)
θ_{-1}	3.79E–08	1.62E–08	3.56E–08	1.62E–08	3.65E–09	4.78E–08	6.30E–08	2.75E–08	4.78E–08	6.30E–08
θ_{-2}	7.56E–08	3.07E–08	7.22E–08	3.07E–08	7.67E–09	9.22E–08	1.23E–07	5.76E–08	9.22E–08	1.23E–07
θ_{-3}	1.13E–07	4.37E–08	1.09E–07	4.37E–08	1.18E–08	1.34E–07	1.79E–07	8.93E–08	1.34E–07	1.79E–07
θ_{-4}	1.50E–07	5.55E–08	1.47E–07	5.55E–08	1.62E–08	1.73E–07	2.33E–07	1.23E–07	1.73E–07	2.33E–07
θ_{-5}	1.86E–07	6.55E–08	1.86E–07	6.55E–08	2.07E–08	2.09E–07	2.84E–07	1.59E–07	2.09E–07	2.84E–07
θ_{-6}	2.23E–07	7.42E–08	2.25E–07	7.42E–08	2.54E–08	2.42E–07	3.32E–07	1.97E–07	2.42E–07	3.32E–07
θ_{-7}	2.59E–07	8.06E–08	2.65E–07	8.06E–08	3.01E–08	2.72E–07	3.76E–07	2.37E–07	2.72E–07	3.76E–07
θ_{-8}	2.96E–07	8.50E–08	3.07E–07	8.50E–08	3.47E–08	3.00E–07	4.17E–07	2.79E–07	3.00E–07	4.17E–07
θ_{-9}	3.32E–07	1.03E–07	3.49E–07	1.03E–07	3.92E–08	3.25E–07	4.55E–07	3.22E–07	3.25E–07	4.55E–07
θ_{-10}	3.69E–07	1.22E–07	3.92E–07	1.22E–07	4.39E–08	3.48E–07	4.90E–07	3.69E–07	3.48E–07	4.90E–07

Measure 3: Maximum relative deviations across frequencies

	(1, 1)	(2, 1)	(3, 1)	(4, 1)	(2, 2)	(3, 2)	(4, 2)	(3, 3)	(4, 3)	(4, 4)
θ_{-1}	4.78E–08	1.32E–07	9.81E–08	1.32E–07	3.22E–08	6.66E–08	8.37E–08	5.00E–08	6.66E–08	8.37E–08
θ_{-2}	9.58E–08	2.46E–07	1.89E–07	2.46E–07	6.14E–08	1.27E–07	1.60E–07	9.41E–08	1.27E–07	1.60E–07
θ_{-3}	1.43E–07	3.59E–07	2.78E–07	3.59E–07	8.84E–08	1.82E–07	2.31E–07	1.34E–07	1.82E–07	2.31E–07
θ_{-4}	1.89E–07	4.65E–07	3.64E–07	4.65E–07	1.13E–07	2.32E–07	2.96E–07	1.69E–07	2.32E–07	2.96E–07
θ_{-5}	2.34E–07	5.67E–07	4.48E–07	5.67E–07	1.36E–07	2.78E–07	3.57E–07	2.00E–07	2.78E–07	3.57E–07
θ_{-6}	2.80E–07	6.66E–07	5.31E–07	6.66E–07	1.56E–07	3.19E–07	4.12E–07	2.27E–07	3.19E–07	4.12E–07
θ_{-7}	3.24E–07	7.62E–07	6.12E–07	7.62E–07	1.75E–07	3.56E–07	4.63E–07	2.50E–07	3.56E–07	4.63E–07
θ_{-8}	3.69E–07	8.55E–07	6.92E–07	8.55E–07	1.92E–07	3.89E–07	5.09E–07	2.79E–07	3.89E–07	5.09E–07
θ_{-9}	4.13E–07	9.47E–07	7.71E–07	9.47E–07	2.07E–07	4.19E–07	5.51E–07	3.22E–07	4.19E–07	5.51E–07
θ_{-10}	4.57E–07	1.04E–06	8.51E–07	1.04E–06	2.21E–07	4.44E–07	5.90E–07	3.69E–07	4.44E–07	5.90E–07

[a]The parameters θ_{-1} to θ_{-10} are as defined in Table 1. The upper triangular elements are omitted due to symmetry.

and 3. Specifically, maximum relative and absolute deviations in both directions are of order 10^{-4} at the very first point away from θ_0, which is already higher than the implied approximation error, then reach order 10^{-2} for most elements of the spectrum in under 4000 steps away from θ_0, and keep growing fast as the curve is extended further. We also experimented with other three-parameter subsets of $(\psi_1, \psi_2, \rho_r, \sigma_r^2)$ and reached similar findings. These findings provide further support for our result.

3.2.2 *Analysis based on the first and second order properties.* We now extend the analysis to incorporate the steady state parameters. Consider the measurement equations

from An and Schorfheide (2007) that relate the output growth, the inflation, and the interest rate observed quarterly to the steady states and the elements of S_t:

$$YGR_t = \gamma^{(Q)} + 100(y_t - y_{t-1} + z_t),$$

$$INFL_t = \pi^{(A)} + 400\pi_t,$$

$$INT_t = \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400r_t,$$

where

$$\gamma^{(Q)} = 100(\gamma - 1), \qquad \pi^{(A)} = 400(\overline{\pi} - 1), \qquad r^{(A)} = 400\left(\frac{1}{\beta} - 1\right),$$

and γ is a constant in the technological shock equation. The parameter vector becomes

$$\overline{\theta} = (\tau, \beta, \nu, \phi, \overline{\pi}, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r^2, \sigma_g^2, \sigma_z^2, \gamma^{(Q)}),$$

where $\gamma^{(Q)}$ is the only nondynamic parameter. Thus, we have

$$\mu(\overline{\theta}) = \begin{pmatrix} \gamma^{(Q)} \\ 400(\overline{\pi} - 1) \\ 400(\overline{\pi} - 1) + 400(\frac{1}{\beta} - 1) + 4\gamma^{(Q)} \end{pmatrix}$$

and the $A(L)$ matrix in this case is

$$\begin{pmatrix} 100 & 0 & 0 & 100 - 100L & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 400 & 0 & 0 & 0 \\ 0 & 0 & 400 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Setting $\gamma^{(Q)} = 0.55$ as in An and Schorfheide (2007), we consider identification at

$$\overline{\theta}_0 = (2, 0.9975, 0.1, 53.6797, 1.008, 1.5, 0.125, 0.75, 0.95, 0.9, 0.4, 3.6, 0.9, 0.55).$$

Note that $\mu(\overline{\theta})$ can be easily differentiated analytically in this case.

Applying Theorem 2, we find $\text{rank}(\overline{G}(\overline{\theta}_0)) = 12$. Hence, $\overline{\theta}_0$ is not identifiable from the first and second order properties of the observables either. After applying the procedure from Section 3.1, we find two subvectors, (ν, ϕ) and $(\psi_1, \psi_2, \rho_r, \sigma_r^2)$, which account for nonidentification. Intuitively, we no longer detect $(\nu, \overline{\pi})$ and $(\phi, \overline{\pi})$, as $\overline{\pi}$ enters $\mu(\overline{\theta})$ and hence is identifiable from the mean. Since the two nonidentification curves are exactly the same as in the dynamic parameter case, they are not reported here.

REMARK 5. This example shows that in this model the Taylor rule parameters are not separately identifiable from the (first and) second order properties of observables at $\overline{\theta}_0$. Such a finding, first documented in the current paper, was also more recently documented in Komunjer and Ng (2011). This constitutes a serious concern for estimation in this and similar DSGE models.

REMARK 6. The results also have direct implications for Bayesian inference. Suppose we impose a tight prior on one of the four parameters, say ψ_1, while using flat priors on the rest. Then the posterior distributions of ψ_2, ρ_r, and σ_r^2 most often become concentrated due to their relation with ψ_1. Therefore, simply comparing the marginal priors and the posteriors may give the false impression that the parameters are separately (or even strongly) identified and may overstate the informativeness of the data about the parameters.

3.2.3 *A procedure to ensure robustness.*   In the above discussion, we used a particular step size for numerical differentiation and the default tolerance level for deciding the ranks of $G(\theta_0)$ and $\overline{G}(\overline{\theta}_0)$. We now examine the sensitivity of the results to a range of numerical differentiation steps (from 10^{-2} to 10^{-9}) and tolerance levels (from 10^{-2} to 10^{-10}). The results are reported in Table 4. We can see that the results are robust over a wide range of step sizes and tolerance levels. Discrepancies start to occur when the step size is very small or very large, and when the tolerance level is very stringent. This is quite intuitive, as when the step size is too large, the numerical differentiation induces a substantial error, since the estimation error for the two-point method is of the same order as the step size. When the step size is too small, the numerical error from solving the model using GENSYS is large relative to the step size; therefore, the rank will also be estimated imprecisely. Our choice of the step size of $10^{-7} \times \theta_0$ can therefore be seen as balancing the trade-off between derivative precision and robustness of the rank computations to tolerance levels as low as 10^{-10}.

Furthermore, the nonidentification curve can be embedded into a procedure to reduce the reliance on the step size and tolerance level. Specifically, we can consider the following:

Step 1. Compute the ranks of $G(\theta_0)$ and $\overline{G}(\overline{\theta}_0)$ using a wide range of step sizes and tolerance levels (such as those in Table 4). Locate the outcomes with the smallest rank.

Step 2. Derive the nonidentification curves conditioning on the smallest rank reported. Compute the discrepancies in spectral densities using values on the curve.

The purpose of Step 1 is to avoid falsely reporting identification when the parameters are unidentified or, more generally, to overstating identification. However, it may incorrectly label identified parameters as unidentified, which is further addressed in Step 2. The idea is, if this indeed occurred, then some curves reported in Step 2 will, in fact, correspond to parameter subsets that are identifiable. Therefore, the discrepancy surfaces as we move along such curves away from θ_0 and $\overline{\theta}_0$. Note that applying this procedure, with step sizes and tolerance levels stated in Table 4, leads to the same results discussed in Sections 3.2.1 and 3.2.2.

REMARK 7. Based on the evidence reported here and our experimentation with other models, we suggest using $10^{-7} \times \theta_0$ (or similar magnitudes) and $\text{size}(G)\,\text{eps}(\|G\|)$ as the default step size and tolerance level when implementing the methods, followed by the two-step procedure outlined above to ensure robustness.

TABLE 4. Rank sensitivity analysis.[a]

	Differentiation Step Size $\times \theta_0$							
TOL	1E–02	1E–03	1E–04	1E–05	1E–06	1E–07	1E–08	1E–09
	Rank of $G(\theta_0)$ in the 13-parameter model							
1E–02	10	10	10	10	10	10	10	10
1E–03	10	10	10	10	10	10	10	10
1E–04	11	10	10	10	10	10	10	10
1E–05	11	10	10	10	10	10	10	10
1E–06	11	11	10	10	10	10	10	11
1E–07	12	11	10	10	10	10	10	11
1E–08	12	12	11	10	10	10	11	12
1E–09	12	12	11	10	10	10	11	12
1E–10	12	12	12	11	10	10	12	12
Default	12	12	11	10	10	10	11	12
	Rank of $G(\theta_0)$ in the 11-parameter model							
1E–02	10	10	10	10	10	10	10	10
1E–03	10	10	10	10	10	10	10	10
1E–04	11	10	10	10	10	10	10	10
1E–05	11	10	10	10	10	10	10	10
1E–06	11	11	10	10	10	10	10	11
1E–07	11	11	10	10	10	10	10	11
1E–08	11	11	11	10	10	10	11	11
1E–09	11	11	11	10	10	10	11	11
1E–10	11	11	11	11	10	10	11	11
Default	11	11	11	10	10	10	10	11

[a]TOL refers to the tolerance level used to determine the rank. Default refers to the MATLAB default tolerance level.

4. FQML ESTIMATION

We first present a brief derivation of the FQML estimators and then study their asymptotic properties in both well specified and misspecified models. The subsequent analysis assumes that the system is nonsingular, that is, $n_Y \leq n_\epsilon$.

4.1 *The estimators*

For the sole purpose of deriving the quasi-likelihood function, assume that the process $\{Y_t\}$ is Gaussian. Let ω_j denote the Fourier frequencies, that is, $\omega_j = 2\pi j/T$ ($j = 1, 2, \ldots, T-1$). The discrete Fourier transforms are given by

$$w_T(\omega_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} Y_t \exp(-i\omega_j t), \quad j = 1, 2, \ldots, T-1.$$

Note that replacing Y_t by $Y_t - \mu(\bar{\theta})$ does not affect the value of $w_T(\omega_j)$ at these frequencies. The transforms $w_T(\omega_j)$ have a complex valued multivariate normal distribution

and for large T are approximately independent, each with the probability density function (see Hannan (1970, pp. 223–225))

$$\frac{1}{\pi^{n_Y} \det(f_\theta(\omega_j))} \exp\bigl[-\operatorname{tr}\{f_\theta^{-1}(\omega_j)w_T(\omega_j)w_T^*(\omega_j)\}\bigr], \quad j = 1, 2, \dots, T-1.$$

Therefore, an approximate log-likelihood function of θ based on observations $Y_1, \dots, Y_T$ is given, up to constant multiplication, by

$$-\sum_{j=1}^{T-1}\bigl[\log\det(f_\theta(\omega_j)) + \operatorname{tr}\{f_\theta^{-1}(\omega_j)I_T(\omega_j)\}\bigr], \tag{14}$$

where $I_T(\omega_j) = w_T(\omega_j)w_T^*(\omega_j)$ denotes the periodogram. Letting $W(\omega_j)$ be an indicator function as defined in the previous section, we consider the generalized version of (14):

$$L_T(\theta) = -\sum_{j=1}^{T-1} W(\omega_j)\bigl[\log\det(f_\theta(\omega_j)) + \operatorname{tr}\{f_\theta^{-1}(\omega_j)I_T(\omega_j)\}\bigr]. \tag{15}$$

Then the FQML estimator for θ is given by

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} L_T(\theta). \tag{16}$$

Thus, the above procedure allows us to estimate the dynamic parameters based on the second order properties of $\{Y_t\}$ without any reference to the steady state parameters. Compared with the time domain QML, the estimate here can be obtained without demeaning the data.

It is also simple to estimate both dynamic and steady state parameters jointly. Let

$$w_{\bar{\theta},T}(0) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} Y_t - \mu(\bar{\theta}) \quad \text{and} \quad I_{\bar{\theta},T}(0) = w_{\bar{\theta},T}(0)w_{\bar{\theta},T}(0)'.$$

Noticing that $w_{\bar{\theta},T}(0)$ has a multivariate normal distribution with asymptotic variance $f_\theta(0)$ and is asymptotically independent of $w_T(\omega_j)$ for $j = 1, 2, \dots, T-1$, we arrive at the approximate log-likelihood function of $\bar{\theta}$:

$$\bar{L}_T(\bar{\theta}) = L_T(\theta) - \bigl[\log\det(f_\theta(0)) + \operatorname{tr}\{f_\theta^{-1}(0)I_{\bar{\theta},T}(0)\}\bigr].$$

Then the FQML estimator for $\bar{\theta}$ is given by

$$\widehat{\bar{\theta}}_T = \arg\max_{\bar{\theta} \in \bar{\Theta}} \bar{L}_T(\bar{\theta}). \tag{17}$$

4.2 Asymptotic properties of the FQML estimators

The asymptotic properties of the estimator (16), with $W(\omega_j) = 1$ for all ω_j, have been studied under various data generating processes in the statistics literature; see, for

example, Dunsmuir (1979) and Hosoya and Taniguchi (1982). The estimator (17) received less attention. One exception is Hansen and Sargent (1993), who formally established that $T^{-1}\bar{L}_T(\bar{\theta})$ converges to the same limit as the time domain Gaussian quasi-maximum likelihood function for $\bar{\theta}$ uniformly in $\bar{\theta} \in \bar{\Theta}$. Their result allows for non-Gaussianity and model misspecification. This section can be viewed as a further development of their work in the following sense. First, we formally establish the relationship between the identification condition and the asymptotic properties of the estimator. Second, we explicitly derive the limiting distribution of the estimator, which is important for inference and model comparison.

We gradually tighten the assumptions to obtain increasingly stronger results. To analyze the first issue, the following assumptions are imposed on the second and fourth order properties of the observed process $\{Y_t\}$.

ASSUMPTION 5. (i) *The process $\{Y_t\}$ is generated by*

$$Y_t = \mu(\bar{\theta}_0) + Y_t^d(\theta_0)$$

with $Y_t^d(\theta)$ satisfying (1). (ii) *The spectral density matrix $f_\theta(\omega)$ is positive definite with eigenvalues bounded away from 0 and ∞ uniformly in ω for all $\theta \in \Theta$. The elements of $\partial \mathrm{vec}(f_\theta(\omega))/\partial\theta'$ are bounded away from ∞ uniformly in ω for all $\theta \in \Theta$. The elements of $f_\theta(\omega)$ belong to* $\mathrm{Lip}(\beta)$ *with respect to ω, the Lipschitz class of degree β, $\beta > 1/2$.*

ASSUMPTION 6. *The process ϵ_t is fourth order stationary. Let $Q_{h,l,g,k}(j_1, j_2, j_3)$ be the joint cumulant of ϵ_{th}, $\epsilon_{(t+j_1)l}$, $\epsilon_{(t+j_2)g}$, and $\epsilon_{(t+j_3)k}$. Assume $\sum_{j_1,j_2,j_3=-\infty}^{\infty} |Q_{h,l,g,k}(j_1, j_2, j_3)| < \infty$ for any $1 \le h, l, g, k \le n_\epsilon$.*

The first part of Assumption 5 states that the model is correctly specified. This is relaxed in Section 4.3. The second part strengthens the first condition in Assumption 3. It is satisfied by stationary finite order vector autoregressive moving average processes with finite error covariance matrices, which are the forms that the solutions to linearized DSGE models typically take. In Assumption 6, the summability of the fourth cumulant is weaker than the independence assumption: a sufficient condition is provided in Andrews (1991, Lemma 1).

We now define the concept of a locally unique maximizer.

DEFINITION 3. Let $L(\varphi)$ be some generic criterion function. We say φ_0 is a locally unique maximizer of $L(\varphi)$ if there exists an open neighborhood of φ_0 such that $L(\varphi) < L(\varphi_0)$ for all φ different from φ_0 in this neighborhood.

Define the following quantities as the limits of $T^{-1}L_T(\theta)$ and $T^{-1}\bar{L}_T(\bar{\theta})$:

$$L_\infty(\theta) = -\frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)\big[\log\det(f_\theta(\omega)) + \mathrm{tr}\{f_\theta^{-1}(\omega)f_{\theta_0}(\omega)\}\big]\,d\omega,$$

$$\bar{L}_\infty(\bar{\theta}) = L_\infty(\theta) - \frac{1}{2\pi}(\mu(\bar{\theta}_0) - \mu(\bar{\theta}))'f_\theta^{-1}(0)(\mu(\bar{\theta}_0) - \mu(\bar{\theta})).$$

LEMMA 1. *Let Assumptions* 1–3, 5, *and* 6 *hold.*

  (i) *Then* $T^{-1}L_T(\theta) \to^P L_\infty(\theta)$ *uniformly over* $\theta \in \Theta$.

  (ii) *The parameter vector* θ_0 *is a locally unique maximizer of* $L_\infty(\theta)$ *if and only if it is locally identified. Furthermore, if* θ_0 *is globally identified,*[8] *then it is the unique maximizer of* $L_\infty(\theta)$.

  (iii) *The convergence* $\hat{\theta}_T \to^P \theta_0$ *holds if one of the following two conditions is satisfied:* (a) θ_0 *is globally identified or* (b) θ_0 *is locally identified and the maximization is carried over the corresponding small neighborhood of identification, say* $\delta(\theta_0)$, *instead of* Θ.

  (iv) *Let Assumptions* 1–6 *hold. Then properties* (i)–(iii) *hold when* θ, θ_0, $\hat{\theta}_T$, $L_T(\theta)$, *and* $L_\infty(\theta)$ *are replaced by* $\bar{\theta}$, $\bar{\theta}_0$, $\hat{\bar{\theta}}_T$, $\bar{L}_T(\bar{\theta})$, *and* $\bar{L}_\infty(\bar{\theta})$, *respectively.*

The first result is essentially due to Lemma A.3.3(1) in Hosoya and Taniguchi (1982). Their result is pointwise in θ and is established with $W(\omega) = 1$. Our result strengthens theirs to uniform convergence, which is important for showing property (iii). The second result formally establishes the close link between the identification conditions and the asymptotic properties of the FQML estimator. The result is quite intuitive ex post, however, it is worth documenting given that the identification property is derived without explicitly referring to the likelihood function. The first two results lead directly to property (iii) by a uniform weak law of large numbers. Property (iv) holds based on the same arguments.

To derive the limiting distribution of the estimators, the assumptions on $\{\epsilon_t\}$ need to be further strengthened.

ASSUMPTION 7. (i) *The process* $\{\epsilon_t\}$ *is a vector of martingale difference sequences with respect to the* σ-*field generated by* $\epsilon_s : s \le t$. *We have* $E(\epsilon_{ta}\epsilon_{tb}|\mathcal{F}_{t-\tau}) = \Sigma_{ab}$, $E(\epsilon_{ta}\epsilon_{tb}\epsilon_{tc}|\mathcal{F}_{t-\tau}) = \xi_{abc}$, *and* $E(\epsilon_{ta}\epsilon_{tb}\epsilon_{tc}\epsilon_{td}|\mathcal{F}_{t-\tau}) = \zeta_{abcd}$ *almost surely with* $\Sigma_{aa} > 0$ *and* $\zeta_{aadd} > 0$ *for all* $1 \le a, b, c, d \le n_\epsilon$. (ii) *Let* $c(t, r) = \epsilon_t \epsilon'_{t+r} - E(\epsilon_t \epsilon'_{t+r})$. *Assume* $\lim_{T\to\infty} T^{-1} \sum_{r=0}^L \sum_{t=1}^T E[c_{ab}(t, r)^2 \mathbf{1}\{c_{ab}(t, r)^2 > \varepsilon T\}] < \varepsilon$ *holds for any* $\varepsilon > 0$, $L < \infty$, *and all* $1 \le a, b \le n_\epsilon$.

Part (i) of Assumption 7 imposes restrictions on the conditional moments up to the fourth order, and $\Sigma_{aa} > 0$ and $\zeta_{aadd} > 0$ are the usual positive variance conditions. It is essentially the same as Assumption C2.3 in Dunsmuir (1979). This part can be further relaxed to allow some conditional heteroskedasticity at the cost of some technical and notational complications; see Theorem 3.1 in Hosoya and Taniguchi (1982). Part (ii) is a Lindeberg-type condition. It ensures that the sample autocovariances $T^{-1/2} \sum_{t=1}^{T-r} c(t, r)$ $(r = 0, 1, \ldots, L)$ satisfy a central limit theorem for any finite fixed L. It can be replaced by other sufficient conditions that serve the same purpose. The next result states the limiting distributions of $\hat{\theta}_T$ and $\hat{\bar{\theta}}_T$.

---

[8]The parameter vector θ is said to be globally identifiable from the second order properties of $\{Y_t\}$ at a point θ_0 if for any $\theta_1 \in \Theta_\theta$, $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$ implies $\theta_0 = \theta_1$.

THEOREM 3. *Suppose θ_0 and $\bar{\theta}_0$ are globally identified or the maximizations (16) and (17) are over convex compact sets in which they are locally identified and are interior points.*

(i) *Let Assumptions 1–3 and 5–7 hold. Then*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \to^d N(0, M^{-1}VM^{-1}),$$

where M and V are $q \times q$ matrices, with the (j, l)th element given by

$$M_{jl} = \int_{-\pi}^{\pi} W(\omega) \operatorname{tr}\left\{ f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_j} f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_l} \right\} d\omega,$$

$$V_{jl} = 4\pi M_{jl} + \sum_{a,b,c,d=1}^{n_\epsilon} \kappa_{abcd} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_j} H(\omega) \, dw \right]_{ab}$$

$$\times \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_l} H(\omega) \, dw \right]_{cd},$$

where $[\cdot]_{ab}$ denotes the (a, b)th element of the matrix, κ_{abcd} is the fourth cross-cumulant of ϵ_{ta}, ϵ_{tb}, ϵ_{tc}, and ϵ_{td}, $H(\omega) = H(\exp(-i\omega); \theta_0) = \sum_{j=0}^{\infty} h_j(\theta_0) \exp(-i\omega j)$ (cf. (3)), and $H^*(\omega)$ is its conjugate transpose.

(ii) *Let Assumptions 1–7 hold. Then $\sqrt{T}(\hat{\bar{\theta}}_T - \bar{\theta}_0) \to^d N(0, \bar{M}^{-1}\bar{V}\bar{M}^{-1})$, where $\bar{M}$ and $\bar{V}$ are $(q + p) \times (q + p)$ matrices, with the (j, l)th element given by*

$$\bar{M}_{jl} = \int_{-\pi}^{\pi} W(\omega) \operatorname{tr}\left\{ f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_j} f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_l} \right\} d\omega$$

$$+ 2 \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}_j} f_{\theta_0}^{-1}(0) \frac{\partial \mu(\bar{\theta}_0)}{\partial \bar{\theta}_l},$$

$$\bar{V}_{jl} = 4\pi \bar{M}_{jl} + \sum_{a,b,c,d=1}^{n_\epsilon} \kappa_{abcd} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_j} H(\omega) \, dw \right]_{ab}$$

$$\times \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_l} H(\omega) \, dw \right]_{cd} + A_{jl} + A_{lj}$$

with $A_{jl} = 2 \sum_{a,b,c=1}^{n_\epsilon} \xi_{abc} \{ \int_{-\pi}^{\pi} W(\omega) [H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_j} H(\omega)]_{ab} \, d\omega \} \times [\frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}_l} f_{\theta_0}^{-1}(0) H(0)]_c$ and $\xi_{abc} = E(\epsilon_{ta} \epsilon_{tb} \epsilon_{tc})$.

When $W(\omega) = 1$, the first result reduces to Corollary 2.2 in Dunsmuir (1979, p. 497) and Proposition 3.1 in Hosoya and Taniguchi (1982), which were obtained in the context of parameter estimation in stationary vector time series models. The generalization to a more general $W(\omega)$ is new. The limiting distribution depends on the fourth order properties of the process. For DSGE models, this is because the same set of parameters affects both the conditional mean and the conditional covariance of the process Y_t^d in (1). Technically, the term $h_0(\theta)$ is in general not an identity matrix, but rather depends

on unknown parameters. This causes the second term in V_{jl} to be in general nonzero. However, in the important special case where ϵ_t are Gaussian with diagonal covariance matrix, $\kappa_{abcd} = 0$ and the limiting distribution depends only on the second order property of the process. This holds for different specifications of $W(\omega)$. Specifically, we have $M^{-1}VM^{-1} = 4\pi M^{-1}$ with

$$[M]_{jl} = \int_{-\pi}^{\pi} W(\omega) \operatorname{tr}\left[f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_j} f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_l} \right] d\omega$$

or, in matrix notation,

$$M^{-1}VM^{-1} = \left[\frac{1}{4\pi} \int_{-\pi}^{\pi} W(\omega) \frac{\partial \operatorname{vec}(f_{\theta_0}(\omega)')'}{\partial \theta} \right.$$
$$\left. \times (f_{\theta_0}^{-1}(\omega)' \otimes f_{\theta_0}^{-1}(\omega)) \frac{\partial \operatorname{vec}(f_{\theta_0}(\omega))}{\partial \theta'} d\omega \right]^{-1}. \tag{18}$$

The second result in the theorem is new in the literature even for the case with $W(\omega) = 1$. The inclusion of the steady state parameter makes the limiting distribution dependent on the third order properties of Y_t, namely ξ_{abc}. Again, in the important special case with Gaussianity and a diagonal covariance matrix, $\xi_{abc} = 0$ and only the second order property matters.

To construct the confidence interval, $f_{\theta_0}(\omega)$, $H(\omega)$, and $H^*(\omega)$ ($\omega \in [-\pi, \pi]$) can be consistently estimated by replacing θ_0 and $\bar{\theta}_0$ with $\hat{\theta}_T$ and $\widehat{\bar{\theta}}_T$, and applying (2) and (4). The derivatives and the integrals can be evaluated numerically. The cumulants ξ_{abc} and κ_{abcd} can be replaced by their sample counterparts.

4.3 *Misspecified models*

We consider the interpretation of the parameter estimates when the DSGE models are viewed as approximations. The next assumption allows the true data generating process to be different from that implied by the DSGE solution.

Assumption MI. *The observations $\{Y_t\}_{t=1}^T$ follow a covariance stationary process given by $Y_t - \mu_0 = \sum_{j=0}^{\infty} h_{0j}\varepsilon_{t-j}$, whose mean μ_0 and spectral density $f_0(\omega)$ are possibly different from $\mu(\bar{\theta}_0)$ and $f_{\theta_0}(\omega)$. Also, Y_t satisfies Assumptions 5(ii) with $f_\theta(\omega)$ replaced by $f_0(\omega)$ and Assumptions 6 and 7 with ϵ_t replaced by ε_t.*

Suppose the estimates $\hat{\theta}_T$ and $\widehat{\bar{\theta}}_T$ are constructed in the same way as before and define the pseudo-true values

$$\theta_0^m = \arg\max_{\theta \in \Theta} L_\infty^m(\theta) \quad \text{and} \quad \bar{\theta}_0^m = \arg\max_{\bar{\theta} \in \bar{\Theta}} \bar{L}_\infty^m(\bar{\theta}),$$

where

$$L_\infty^m(\theta) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega)\left[\log\det(f_\theta(\omega)) + \operatorname{tr}\{f_\theta^{-1}(\omega)f_0(\omega)\}\right] d\omega,$$

$$\bar{L}_\infty^m(\bar{\theta}) = L_\infty^m(\theta) - \frac{1}{2\pi}(\mu_0 - \mu(\bar{\theta}))' f_\theta^{-1}(0)(\mu_0 - \mu(\bar{\theta})).$$

Suppose θ_0^m and $\bar{\theta}_0^m$ lie in the interior of Θ and $\bar{\Theta}$.

COROLLARY 7. *Suppose θ_0^m and $\bar{\theta}_0^m$ are globally identified or the maximizations* (16) *and* (17) *are over convex compact sets in which they are locally identified and are interior points. Let Assumption* MI *hold.*

(i) *Assume the DSGE solution $Y_t^d(\theta)$ satisfies Assumptions* 1–3 *and* 5(ii)*. Then*

$$\sqrt{T}(\hat{\theta}_T - \theta_0^m) \to^d N(0, \Omega^{-1}\Pi\Omega^{-1})$$

with

$$\Omega = \int_{-\pi}^{\pi} W(\omega)\left[\frac{\partial^2}{\partial\theta\,\partial\theta'}\log\det\big(f_{\theta_0^m}(\omega)\big) + \frac{\partial^2}{\partial\theta\,\partial\theta'}\operatorname{tr}\{f_{\theta_0^m}^{-1}(\omega)f_0(\omega)\}\right]d\omega,$$

$$\Pi_{jl} = 4\pi\int_{-\pi}^{\pi} W(\omega)\operatorname{tr}\left\{f_0(\omega)\frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial\theta_j}f_0(\omega)\frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial\theta_l}\right\}d\omega$$

$$+ \sum_{a,b,c,d=1}^{n_\varepsilon}\kappa_{abcd}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)H_0^*(\omega)\frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial\theta_j}H_0(\omega)\,dw\right]_{ab}$$

$$\times\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)H_0^*(\omega)\frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial\theta_l}H_0(\omega)\,dw\right]_{cd},$$

where κ_{abcd} is the fourth cross-cumulant of ε_{ta}, ε_{tb}, ε_{tc}, and ε_{td}, and $H_0(\omega) = \sum_{j=0}^{\infty} h_{0j} \times \exp(-i\omega j)$.

(ii) *Assume the DSGE solution is given by $\mu(\bar{\theta}) + Y_t^d(\theta)$ and satisfies Assumptions* 1–4 *and* 5(ii)*. Then $\sqrt{T}(\hat{\theta}_T - \bar{\theta}_0^m) \to^d N(0, \bar{\Omega}^{-1}\bar{\Pi}\bar{\Omega}^{-1})$ with*

$$\bar{\Omega} = \int_{-\pi}^{\pi} W(\omega)\left[\frac{\partial^2}{\partial\bar{\theta}\,\partial\bar{\theta}'}\log\det\big(f_{\theta_0^m}(\omega)\big) + \frac{\partial^2}{\partial\bar{\theta}\,\partial\bar{\theta}'}\operatorname{tr}\{f_{\theta_0^m}^{-1}(\omega)f_0(\omega)\}\right]d\omega$$

$$+ 2\frac{\partial\mu(\bar{\theta}_0^m)'}{\partial\bar{\theta}}f_{\theta_0^m}^{-1}(0)\frac{\partial\mu(\bar{\theta}_0^m)}{\partial\bar{\theta}'},$$

$$\bar{\Pi}_{jl} = 4\pi\left\{\int_{-\pi}^{\pi} W(\omega)\operatorname{tr}\left\{f_0(\omega)\frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial\bar{\theta}_j}f_0(\omega)\frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial\bar{\theta}_l}\right\}d\omega\right.$$

$$\left.+ 2\frac{\partial\mu(\bar{\theta}_0^m)'}{\partial\bar{\theta}_j}f_{\theta_0^m}^{-1}(0)\frac{\partial\mu(\bar{\theta}_0^m)}{\partial\bar{\theta}_l}\right\}$$

$$+ \sum_{a,b,c,d=1}^{n_\varepsilon}\kappa_{abcd}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)H_0^*(\omega)\frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial\bar{\theta}_j}H_0(\omega)\,dw\right]_{ab}$$

$$\times\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)H_0^*(\omega)\frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial\bar{\theta}_l}H_0(\omega)\,dw\right]_{cd} + A_{jl} + A_{lj}$$

with

$$A_{jl} = 2 \sum_{a,b,c,d=1}^{n_\varepsilon} \xi_{abc} \left\{ \left[\int_{-\pi}^{\pi} W(\omega) \left[H_0^*(\omega) \frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial \bar{\theta}_j} H_0(\omega) \right]_{ab} d\omega \right] \right.$$

$$\times \left. \left[\frac{\partial \mu(\bar{\theta}_0^m)'}{\partial \bar{\theta}_l} f_{\theta_0^m}^{-1}(0) H_0(0) \right]_c \right.$$

and $\xi_{abc} = E(\varepsilon_{ta} \varepsilon_{tb} \varepsilon_{tc})$.

Misspecification in general affects both the mean and the variance of the estimate. Note that when only estimating the dynamic parameters, misspecifying $\mu(\bar{\theta})$ has no effect on the estimate $\hat{\theta}_T$.

5. Quasi-Bayesian inference

This section extends the above framework to incorporate prior distributions on the DSGE parameters. It also discusses a computationally attractive procedure to obtain parameter estimates. The analysis is motivated by Chernozhukov and Hong (2003). We focus on θ_0 because the procedure is identical for $\bar{\theta}_0$.

Consider the function

$$p_T(\theta) = \frac{\pi(\theta) \exp(L_T(\theta))}{\displaystyle\int_\Theta \pi(\theta) \exp(L_T(\theta)) \, d\theta}, \tag{19}$$

where $L_T(\theta)$ is the same as in (15) and $\pi(\theta)$ can be a proper prior probability density or, more generally, a weight function that is strictly positive and continuous over Θ. Because $\exp(L_T(\theta))$ is a more general criterion function than the likelihood, $p_T(\theta)$ is in general not a true posterior in the Bayesian sense. However, it is a proper distribution density over the parameters of interest, and is termed quasi-posterior in Chernozhukov and Hong (2003).

The estimate for θ_0 can be taken to be the quasi-posterior mean

$$\hat{\theta}_T = \int_\Theta \theta p_T(\theta) \, d\theta.$$

To compute the estimator, we can use Markov chain Monte Carlo (MCMC) methods, such as the Metropolis–Hastings algorithm, to draw a Markov chain

$$S = \left(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(B)} \right)$$

whose marginal density is approximately given by $p_T(\theta)$, and $\hat{\theta}_T$ can be computed as

$$\hat{\theta}_T = \frac{1}{B} \sum_{j=1}^B \theta^{(j)}.$$

Meanwhile, for a given continuously differentiable function $g : \Theta \to \mathbb{R}$, for example, an impulse response at a given horizon, its estimate can be obtained via

$$g(\hat{\theta}_T) = \frac{1}{B} \sum_{j=1}^{B} g(\theta^{(j)}).$$

Here we omit the details on the construction of the Markov chains, since they follow standard procedures. One may refer to Chernozhukov and Hong (2003, Section 5) or An and Schorfheide (2007) for more details.

The next result provides an asymptotic justification for the estimator under correct model specification.

THEOREM 4. *Suppose θ_0 ($\bar{\theta}_0$) is globally identified or $\pi(\theta)$ ($\pi(\bar{\theta})$) is strictly positive only over a compact convex neighborhood of θ_0 ($\bar{\theta}_0$) in which they are locally identified and are interior points. Then $\hat{\theta}_T$ ($\widetilde{\bar{\theta}}_T$) has the same limiting distribution as in Theorem 3 under the corresponding assumptions stated there.*

Consider the construction of confidence intervals for the elements of θ_0 or, more generally, of $g(\theta_0)$. In the important special case of Gaussianity with $\Sigma(\theta)$ being diagonal, the confidence intervals can be obtained directly from the quantiles of the MCMC sequence $(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(B)})$. Such intervals are asymptotically valid because $\kappa_{abcd} = 0$ and therefore $M = V$. The same result holds for $\bar{\theta}_0$ because $\xi_{abc} = 0$, thus $\bar{M} = \bar{V}$. In the general case, because $\exp(L_T(\theta))$ is a more general criterion function, implying $M \neq V$, such an interval is not necessarily asymptotically valid, as clearly demonstrated in Chernozhukov and Hong (2003). However, valid large sample inference can still be easily carried out using the Delta method, as suggested in Chernozhukov and Hong (2003, Theorem 4). Specifically, let $\hat{M}^{-1}$ be T times the variance–covariance matrix of the MCMC sequence $(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(B)})$. Let $\hat{V}$ be an estimator for V, which can be obtained using the formula in Theorem 3 by replacing $H(\omega)$, κ_{abcd}, and $\partial f_{\theta_0}^{-1}(\omega)/\partial \theta_j$ $(j = 1, 2, \dots, q)$ with their consistent estimates. Then a valid $(1 - \alpha)$ percent confidence interval for $g(\theta_0)$ is given by

$$[c_{g,T}(\alpha/2), c_{g,T}(1 - \alpha/2)],$$

where

$$c_{g,T}(\alpha) = g(\hat{\theta}_T) + q_\alpha T^{-1/2} \sqrt{\frac{\partial g(\hat{\theta}_T)}{\partial \theta'} \hat{M}^{-1} \hat{V} \hat{M}^{-1} \frac{\partial g(\hat{\theta}_T)}{\partial \theta}}$$

with q_α being the α-quantile of the standard normal distribution. Analogous argument can be applied to construct confidence intervals for $g(\bar{\theta}_0)$. The asymptotic validity of such intervals can be verified using the same argument as in Chernozhukov and Hong (2003, Theorem 4). Therefore, the details are omitted here.

Under misspecification, a result analogous to Theorem 4 can be obtained, with the true value replaced by the pseudo-true values and the covariance matrix modified accordingly.

The key computational difference between the above method and the time domain quasi-Bayesian inference is in computing the Kalman filter versus the spectral density at the different parameter values. Therefore, the computation costs are similar. The spectral domain approach has some additional advantages. First, one can exclude some frequencies by specifying an appropriate $W(\omega)$, which is not easy to achieve in the time domain. Second, if the sole interest is in estimating the dynamic parameters, it is not necessary to specify $\mu(\bar{\theta})$ or to demean the data. Third, although not pursued in the current paper, the spectral domain approach can be extended to handle models without requiring log linearizations. The idea is that as long as the spectral density can be computed, analytically or by simulation, a criterion function similar to (14) can be constructed to obtain parameter estimates. Such an idea has been mentioned elsewhere, for example, in Diebold, Ohanian, and Berkowitz (1998), but has not been formally studied. Finally, it provides a platform for conducting hypothesis testing and model diagnosis from the spectral domain, as emphasized by Watson (1993). For example, one can readily obtain estimates and confidence interval for components of the spectral density matrix and contrast them with the observed data. Also, it is simple to construct tests for restrictions imposed on a given frequency component, such as the zero frequency. We plan to explore such developments in future work.

6. Conclusion

We have provided a unified treatment of issues related to identification, inference, and computation in linearized DSGE models in the frequency domain. In addition to presenting a necessary and sufficient condition for local identification of the structural parameters, we also proposed a method to trace out nonidentification curves when lack of identification is detected. The application of our condition is straightforward because it mainly involves computing the first order derivatives of the spectral density. The MATLAB code and the results for a more complex medium size DSGE model are available on our webpage. For estimation, we considered a frequency domain quasi-maximum likelihood (FQML) estimator and showed that it permits incorporation of relevant prior distributions and is computationally attractive.

The current work can be further developed in several directions. First, we have assumed determinacy, but we conjecture that our identification condition can be applied to any selected equilibrium path under indeterminacy, provided that the state vector and the parameter space are augmented accordingly. Second, although we have worked with log-linearized systems, we conjecture the condition can be applied to DSGE models solved with higher order approximations, provided the resulting spectral density and its derivatives can be computed precisely. Although the paper does not consider weak identification, it can be shown that the frequency domain perspective affords a simple and transparent inferential procedure robust to weak identification (see Qu (2011)). We are currently pursuing such research directions and hope to report results in the near future.

References

Altug, S. (1989), "Time-to-build and aggregate fluctuations: some new evidence." *International Economic Review*, 30, 889–920. [98]

An, S. and F. Schorfheide (2007), "Bayesian analysis of DSGE models." *Econometric Reviews*, 26, 113–172. [96, 97, 103, 111, 112, 113, 119, 129]

Andrews, D. W. K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica*, 59, 817–858. [123]

Berkowitz, J. (2001), "Generalized spectral estimation of the consumption-based asset pricing model." *Journal of Econometrics*, 104, 269–288. [98, 105]

Canova, F. and L. Sala (2009), "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics*, 56, 431–449. [96]

Chernozhukov, V. and H. Hong (2003), "An MCMC approach to classical estimation." *Journal of Econometrics*, 115, 293–346. [98, 128, 129]

Christiano, L. J., M. Eichenbaum, and D. Marshall (1991), "The permanent income hypothesis revisited." *Econometrica*, 59, 397–423. [98]

Christiano, L. J. and R. J. Vigfusson (2003), "Maximum likelihood in the frequency domain: The importance of time-to-plan." *Journal of Monetary Economics*, 50, 789–815. [98]

Consolo, A., C. A. Favero, and A. Paccagnini (2009), "On the statistical identification of DSGE models." *Journal of Econometrics*, 150, 99–115. [96]

Diebold, F. X., L. E. Ohanian, and J. Berkowitz (1998), "Dynamic equilibrium economies: A framework for comparing models and data." *The Review of Economic Studies*, 65, 433–451. [98, 105, 130]

Dunsmuir, W. (1979), "A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise." *The Annals of Statistics*, 7, 490–506. [98, 123, 124, 125]

Dunsmuir, W. and E. J. Hannan (1976), "Vector linear time series models." *Advances in Applied Probability*, 8, 339–364. [98]

Engle, R. F. (1974), "Band spectrum regression." *International Economic Review*, 15, 1–11. [98]

Fernández-Villaverde, J. (2010), "The econometrics of DSGE models." *Journal of the Spanish Economic Association*, 1, 3–49. [96]

Fisher, F. M. (1966), *The Identification Problem in Economics*. McGraw-Hill, New York. [106]

Fukac, M., D. F. Waggoner, and T. Zha (2007), "Local and global identification of DSGE models: A simultaneous-equation approach." Working paper, Federal Reserve Bank of Atlanta. [96]

Hannan, E. J. (1970), *Multiple Time Series*. Wiley, New York. [102, 122]

Hansen, L. P. and T. J. Sargent (1993), "Seasonality and approximation errors in rational expectations models." *Journal of Econometrics*, 55, 21–55. [98, 105, 123]

Hosoya, Y. and M. Taniguchi (1982), "A central limit theorem for stationary processes and the parameter estimation of linear processes." *The Annals of Statistics*, 10, 132–153. [98, 123, 124, 125]

Iskrev, N. (2010), "Local identification in DSGE models." *Journal of Monetary Economics*, 57, 189–202. [96, 108]

Komunjer, I. and S. Ng (2011), "Dynamic identification of DSGE models." *Econometrica*, 79, 1995–2032. [96, 97, 108, 111, 119]

Qu, Z. (2011), "Inference and specification testing in DSGE models with possible weak identification." Working paper, Department of Economics, Boston University. [130]

Rothenberg, T. J. (1971), "Identification in parametric models." *Econometrica*, 39, 577–591. [97, 98, 102, 104, 106, 108]

Sims, C. A. (2002), "Solving linear rational expectations models." *Computational Economics*, 20, 1–20. [100, 112, 113]

Smets, F. and R. Wouters (2007), "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American Economic Review*, 97, 586–606. [108]

Uhlig, H. (1999), "A toolkit for analyzing nonlinear dynamic stochastic models easily." In *Computational Methods for the Study of Dynamic Economies* (R. Marimon and A. Scott, eds.), Oxford University Press, New York. [101, 113]

Wald, A. (1950), "Note on the identification of economic relations." In *Statistical Inference in Dynamic Economic Models*. Cowles Commission Monograph 10, Wiley, New York. [106]

Watson, M. W. (1993), "Measures of fit for calibrated models." *Journal of Political Economy*, 101, 1011–1041. [98, 130]

Whittle, P. (1951), *Hypothesis Testing in Time Series Analysis.* Thesis, Uppsala University. Almqvist and Wiksell, Uppsala; Hafner, New York. [98]