BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

FREQUENCY DOMAIN ANALYSIS OF DSGE AND STOCHASTIC VOLATILITY MODELS

by

DENIS TKACHENKO

B.A., Trinity College Dublin, 2005
M.Sc., Trinity College Dublin, 2006

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2012

UMI Number: 3529026

UMI 3529026

Approved by

First Reader

Zhongjun Qu, Ph.D.
Assistant Professor of Economics

Second Reader

Pierre Perron, Ph.D.
Professor of Economics

Third Reader

Hiroaki Kaido, Ph.D.
Assistant Professor of Economics

Acknowledgments

I am sincerely grateful to my main advisor, Zhongjun Qu, whose constant guidance, encouragement, and technical acumen have been an invaluable help during the entire process of completing this dissertation. I deeply appreciate your mentorship, work ethic, and the fact that your door was always open for me during my time at Boston University. I have also been forunate to have Pierre Perron as one of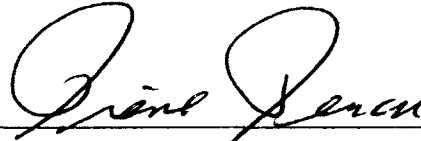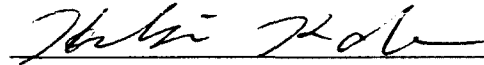 my advisors and my teachers. I would like to thank him for his enthusiasm and conveying unique insights while teaching econometrics that ultimately convinced me to enter this field of economic research. His numerous helpful comments during the seminars and after reading many drafts of the papers comprising this dissertation are also much appreciated. Also, I would like to thank Hiroaki Kaido for his comments and support, especially during the job market period. Thank you all again for your trust and support throughout these years, and for all the invaluable lessons you have taught me, both academic and otherwise.

My thanks also go to my classmates. It has been very fulfilling and stimulating to be part of such a great community that has formed within our Ph.D. program. From educating each other and discussing research ideas to having fun together and winning the intramural soccer tournament for several seasons - it has made tough times a lot less stressful and helped form friendships for life.

Finally, I would like to thank my parents for always being there for me and respecting my educational choices throughout my life, however illogical they may have seemed at the time.

FREQUENCY DOMAIN ANALYSIS OF DSGE AND STOCHASTIC VOLATILITY MODELS

(Order No.   )

DENIS TKACHENKO

Boston University, Graduate School of Arts and Sciences, 2012

Major Professor: Zhongjun Qu, Assistant Professor of Economics

ABSTRACT

In this dissertation, we use frequency domain methods to address issues related to identification and estimation in linearized dynamic stochastic general equilibrium (DSGE) and stochastic volatility models.

The first chapter provides a necessary and sufficient condition for the local identification of the structural parameters based on the (first and) second order properties of the linearized DSGE model. The condition is flexible and simple to verify. It is extended to study identification through a subset of frequencies, partial identification, conditional identification, and constrained identification. When lack of identification is detected. the method can be used to trace out nonidentification curves. For estimation in nonsingular systems, we consider a frequency domain quasi-maximum likelihood (FDQML) estimator and present its asymptotic properties, which can be different from existing results due to the structure of the DSGE model. Finally, we discuss a quasi-Bayesian procedure for estimation and inference that can incorporate relevant prior distributions and is computationally attractive.

The second chapter analyzes a popular medium scale DSGE model of Smets and Wouters (2007) using the framework developed in the previous chapter. For identification,

in addition to checking parameter identifiability, we derive the corresponding nonidentification curve. For estimation and inference, we contrast estimates obtained using the full spectrum with those using only the business cycle frequencies to find notably different parameter values and impulse response functions. A further comparison between the non-parametrically estimated and model implied spectra suggests that the business cycle based method delivers better estimates of the features that the model is intended to capture.

The final chapter proposes an FDQML estimator of the integrated volatility of financial assets in the noisy high frequency data setting. The approach allows for the microstructure noise to be a stationary linear process, and is analytically tractable. In practice, we approximate the noise process by a finite order autoregression, where the order is chosen using the Akaike information criterion (AIC). The simulation study shows that the finite sample performance of the estimator is very similar to its time domain analogue in the case of i.i.d. noise, and is substantially better when more sophisticated noise specifications are considered.

Contents

List of Tables

List of Figures

List of Abbreviations

AIC   Akaike Information Criterion

AR   Autoregressive

ARCH   Autoregressive Conditional Heteroskedasticity

ARMA   Autoregressive Moving Average

BC   Business Cycle

CIR   Cox-Ingersoll-Ross

CLT   Central Limit Theorem

DGP   Data Generating Process

DSGE   Dynamic Stochastic General Equilibrium

FDQML Frequency Domain Quasi-Maximum Likelihood

FOC   First Order Condition

GARCH Generalized Autoregressive Conditional Heteroskedasticity

i.i.d   Independent Identically Distributed

IV   Integrated Volatility

MA   Moving Average

MCMC   Monte Carlo Markov Chain

MLE   Maximum Likelihood Estimator

MN   Mixed Normal

MSE   Mean Squared Error

MSRV	Multi-Scale Realized Volatility
QML	Quasi-Maximum Likelihood
RMSE	Root Mean Squared Error
RV	Realized Variance
RWM	Random Walk Metropolis
SW	Smets and Wouters
TDQML	Time Domain Quasi-Maximum Likelihood
TSRV	Two-Scale Realized Volatility
VAR	Vector Autoregressive
VARMA	Vector Autoregressive Moving Average
US	United States of America

Chapter 1

Identification and Frequency Domain QML Estimation of Linearized DSGE Models (with Zhongjun Qu)

1.1 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 chapter 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 Fukaç, 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, i.e., identifying a subset of parameters without making identification statements about the rest. (iii) We give a necessary and sufficient condition for conditional identification, i.e., 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 chapter.

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 work 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 1.3.2.

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 work 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 in-

formation 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 (FDQML) estimation in this chapter. The FDQML 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 FDQML 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 FDQML as a special case.

The contribution of this chapter in the area of FDQML 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 FDQML 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 FDQML to various problems. We believe that the identification, estimation, and computational results obtained in this chapter can be useful to further develop the literature in this field and to facilitate estimation and comparison of more sophisticated models.

The chapter is organized as follows. The structure of the DSGE solution is discussed in Section 1.2. Section 1.3 considers the local identification of the structural parameters together with an algorithm to trace out nonidentification curves and an illustrative example. The FDQML estimator and its asymptotic properties are studied in Section 1.4. The discussion on interpretation of the estimates in misspecified models is also included. Section 1.5 presents a quasi-Bayesian approach for computation and inference. Section 1.6 concludes. All proofs are contained in the mathematical appendix 1. Section 1.8 contains relevant tables and figures.

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 a-th element. For a matrix A, A_{ab} stands for its (a, b)-th entry. If $f_\theta \in R^k$ is a differentiable function of $\theta \in R^p$, then $\partial f_{\theta_0}/\partial \theta'$ is a k × p matrix of partial derivatives evaluated at θ_0. "$\to^p$" and "$\to^d$" signify convergence in probability and in distribution. And $O_p(\cdot)$ and $o_p(\cdot)$ are the usual symbols for stochastic orders of magnitude.

1.2 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. $Y_t^d(\theta)$ 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

$$Y_t^d(\theta) = \sum_{j=0}^{\infty} h_j(\theta)\epsilon_{t-j}, \tag{1.1}$$

where $h_j(\theta)$ $(j = 0, ..., \infty)$ 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

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

Let $H(L; \theta)$ denote the matrix of lagged polynomials, i.e.,

$$H(L; \theta) = \sum_{j=0}^{\infty} h_j(\theta)L^j. \tag{1.2}$$

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

$$Y_t^d(\theta) = H(L; \theta)\epsilon_t. \tag{1.3}$$

Remark 1.1. *We work directly with the vector moving average representation (1.3) without assuming invertibility, i.e., $\epsilon_t = \sum_{j=0}^{\infty} g_j(\theta) Y_{t-j}^d(\theta)$ 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 1.1. $\{\epsilon_t\}$ *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* $t \neq s$. $\sum_{j=0}^{\infty} tr(h_j(\theta)\Sigma(\theta)h_j(\theta)') < \infty$.

Assumption 1.1, along with (1.1), implies that $Y_t^d(\theta)$ is covariance stationary and has a spectral density matrix $f_\theta(\omega)$ that can be written as

$$f_\theta(\omega) = \frac{1}{2\pi} H(\exp(-i\omega); \theta)\Sigma(\theta)H(\exp(-i\omega); \theta)^*, \tag{1.4}$$

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 1.1. *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),*

$$\Gamma_0 S_t = \Gamma_1 S_{t-1} + \Psi Z_t + \Pi \eta_t, \tag{1.5}$$

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

$$S_t = \Theta_1 S_{t-1} + \Theta_0 Z_t + \Theta_S \sum_{j=1}^{\infty} \Theta_f^{j-1} \Theta_Z E_t Z_{t+j}, \tag{1.6}$$

where $\Theta_0, \Theta_1, \Theta_S, \Theta_f$, 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 (1.4) with $H(L; \theta) = A(L)(I - \Theta_1 L)^{-1} B(L)$.

Remark 1.2. *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 zeros and ones to the entries of $A(L)$. We illustrate the specification of $A(L)$ in Section 1.3.2 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 1.2. *Another representation used in the literature by, among others, Uhlig (1999), is*

$$
\begin{aligned}
k_{t+1} &= P k_t + Q z_t, \\
w_t &= R k_t + S z_t, \\
z_{t+1} &= \Psi z_t + \epsilon_{t+1},
\end{aligned}
$$

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

$$Y_t^d = \begin{pmatrix} k_t \\ w_t \end{pmatrix}. \tag{1.7}$$

Then the spectral density of Y_t^d is given by (1.4) 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 (1.7), which corresponds to

$$H(L;\theta) = A(L) \begin{pmatrix} L^{-1}[I - PL] & 0 \\ -R & I \end{pmatrix}^{-1} \begin{pmatrix} Q \\ S \end{pmatrix} [I - \Psi L]^{-1}. \tag{1.8}$$

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 chapter, 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 $f_\theta(\omega)$.

1.3 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 1.2. $\theta \in \Theta \subset \mathbb{R}^q$ 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 1.1. *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 $\Gamma(k)$ $(k = 0, \pm 1, ...)$ 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 $\Gamma(k)$ $(k = 0, \pm 1, ...)$ 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 1.3. *The elements of $f_\theta(\omega)$ are continuous in ω, and continuous and differentiable in θ. The elements of the derivatives $\partial vec(f_\theta(\omega))/\partial \theta'$ are continuous in θ and ω. Let*

$$G(\theta) = \int\limits_{-\pi}^{\pi} \left(\frac{\partial vec(f_\theta(\omega)')}{\partial \theta'} \right)' \left(\frac{\partial vec(f_\theta(\omega))}{\partial \theta'} \right) d\omega. \tag{1.9}$$

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

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

where the asterisk now denotes the conjugate transpose.

Remark 1.3. *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*

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

We use this representation to compute $G(\theta)$ in the application in Section 1.3.2. Lemma 1.2 in Section 1.7 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 1.1. *Let Assumptions 1.1-1.3 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 s-th frequency in the partition. Then $\partial f_{\theta_0}(\omega_s)/\partial\theta_j$ can be computed numerically using a simple two-point method,

$$\frac{f_{\theta_0 + e_j h_j}(\omega_s) - f_{\theta_0}(\omega_s)}{h_j} \quad j = 1, ..., N + 1,$$

where e_j is a $q \times 1$ unit vector with the j-th 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 $\theta = \theta_0 + e_j h_j$. After this is repeated for all parameters in θ, we can compute $G_{jk}(\theta_0)$ using

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

Note that no simulation is needed in this process. For the model considered in Section 1.3.2 (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 1.1 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 pa-

rameters 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]

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

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

Corollary 1.1. *Let Assumptions 1.1-1.3 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 $c \in R^q$, $G(\theta_0)c = 0$ if and only if $I(\theta_0)c = 0$.*

Therefore, Rothenberg's (1971) condition applies to DSGE models, albeit only to non-singular 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 addressed in Section 1.3.1.

Given the insight conveyed by Theorem 1.1, it becomes straightforward to study the identification of $\bar\theta$ based on both first and second order properties of the process.

Definition 1.2. *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 1.4. *The elements of $\mu(\bar\theta)$ are continuously differentiable with respect to $\bar\theta$. Let*

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

Assume there exists an open neighborhood of $\bar\theta_0$ in which $\bar G(\bar\theta)$ has a constant rank.

Remark 1.4. *$\bar G(\bar\theta)$ is a $(p + q) \times (p + q)$ matrix. The first term is a bordered matrix, consisting of $G(\theta)$ with p rows and columns of zeros appended to it. Both terms are positive*

---

[1]Under Gaussianity, $I(\theta_0)^{-1}$ is the asymptotic covariance matrix of the FDQML estimator based on the full spectrum, see Section 1.4, in particular Theorem 1.3 and the expression (1.18) that follows.

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

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

Theorem 1.2. *Let Assumptions 1.1-1.4 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 1.1 and 1.2 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 mis-specified. 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 zero and equal to one 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

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

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

Corollary 1.2. (Identification from a subset of frequencies)

---

[2]This extension is needed for FDQML estimation since the objective function involves summation over $\omega_j = 2\pi/T, ..., 2\pi(T-1)/T$; see (1.15).

1. *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.*

2. *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 1.1 and 1.2, because $W(\omega)$ is a nonnegative real valued function; therefore, it is omitted. Note that because the quantities

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

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 1.3. (Partial identification)

1. *Let Assumptions 1.1-1.3 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.

2. *Let Assumptions 1.1-1.4 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 provided in Section 1.7. 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 1.4. (Conditional Identification).

1. *Let Assumptions 1.1-1.3 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*

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

is nonsingular.

2. *Let Assumptions 1.1-1.4 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*

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

is nonsingular.

The proof is the same as for Theorems 1.1 and 1.2 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 1.3 and 1.4 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 1.5. (Identification under general constraints)

1. *Let Assumptions 1.1-1.3 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.

2. *Let Assumptions 1.1-1.4 hold and let the other conditions stated in part 1 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 1.5 can also be used to study conditional identification, because the latter is a special case of simple linear restrictions. However, Corollary 1.4 is simpler to apply, especially if the dimension of θ^s is much smaller compared to that of θ. Clearly,

Corollaries 1.3-1.5 can be applied in conjunction with Corollary 1.2 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.

1.3.1 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 1.1 or Corollary 1.2 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 1.1 to 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

$$\frac{\partial \text{vec}\left(f_{\theta(v)}(\omega)\right)}{\partial v} = \frac{\partial \text{vec}\left(f_{\theta(v)}(\omega)\right)}{\partial \theta(v)'}c(\theta) = 0 \tag{1.10}$$

for all $\omega \in [-\pi, \pi]$, where the last equality uses Assumption 1.3 and the fact that $c(\theta)$ is the eigenvector corresponding to the zero eigenvalue (see (1.22) in the mathematical 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 1.6. *Let Assumptions 1.1-1.3 hold and let $rank(G(\theta_0)) = q - 1$. 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 1.6 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 $f_\theta(.)$ associated with finite partitions of $[-\pi, \pi]$. Then a standard constant rank theorem can be applied. The details of the proof are in Section 1.7.

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

$$\begin{aligned}
\theta(v_{j+1}) &\approx \theta(v_j) + c(\theta(v_j))(v_{j+1} - v_j), \ v_{j+1} \geq v_j \geq 0, \ j = 0, 1, ..., & (1.11) \\
\theta(v_{j-1}) &\approx \theta(v_j) + c(\theta(v_j))(v_{j-1} - v_j), \ v_{j-1} \leq v_j \leq 0, \ j = 0, -1, ...,
\end{aligned}$$

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 order $O(h^2)$ 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 (1.10), 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 1.4. 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 1.1 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 1.4 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 1.4, 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 to 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 1.4 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 $\theta \in R^4$, 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 imple-

ment 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 $G(\theta)$ (see Remark 1.3). 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

$$G(\theta)^s_{i,j} = G(\theta)_{\Phi_i, \Phi_j}, \quad i = 1, 2, ..., k; \ j = 1, 2, ..., k. \tag{1.12}$$

Also note that in the case of Theorem 1.2, the same logic applies to the term $[\partial\mu(\bar{\theta}_0)'/\partial\bar{\theta}^s]$ $[\partial\mu(\bar{\theta}_0)/\partial\bar{\theta}^{s\prime}]$, i.e., having computed it once, one can repeatedly apply Corollary 1.4 by selecting relevant elements from it and $\overline{G}(\bar{\theta})^s$ in the same fashion as in (1.12).

1.3.2 An illustrative example

To provide a frame of reference, we consider a DSGE model from An and Schorfheide (2007) whose identification is also 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

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

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

$$c_t = y_t - g_t,$$

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

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

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

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 $\bar{\pi}$ is the steady state inflation rate. The vector of parameters to be identified is

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

We use parameter values

$$\theta_0 = (2, 0.9975, 0.1, 53.6797, 1.008^2, 1.5, 0.125, 0.75, 0.95, 0.9, 0.4, 3.6, 0.9),^3$$

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 (1.5) with

$$S_t = (z_t, g_t, r_t, y_t, \pi_t, c_t, E_t(\pi_{t+1}), E_t(y_{t+1}))'. \tag{1.13}$$

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 (1.6), 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

---

[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.

computed using (1.4) with

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

Given the S_t in (1.13) 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 (1.7), 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 (1.8).

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 $-4,999\pi/5,000$ to $4,999\pi/5,000$ and multiplying by 2π. The results reported are robust to varying the number of frequencies between 5,000 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 $tol = size(G)eps(\|G\|)$, where eps is the floating point precision of G. We obtain $rank(G(\theta_0)) = 10$. Because $q = 13$, this means that the entire parameter vector cannot be

---

[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.

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 1.3.1 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 (1.12), 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) , $(\nu, \bar{\pi}^2)$, and $(\phi, \bar{\pi}^2)$. 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 nonidentification curves for these cases, as they are trivial and can be eliminated by reparameterizing the model with $\kappa \equiv \tau(1-\nu)/(\nu\bar{\pi}^2\phi)$ 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.

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 $\bar{\pi}^2$, 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 $\bar{\pi}^2$. 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 1.1. 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 (1.11). 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 1.1. Specifically, ten 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., 5,000 frequencies between 0 and π)[7] for each point on the

---

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

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 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 5,000 frequencies between 0 and π):

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

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

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

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 the points in Table 1.1, as the rest are very similar. Both Tables 1.2 and 1.3 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 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 4,000 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.

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:

$$
\begin{aligned}
YGR_t &= \gamma^{(Q)} + 100(y_t - y_{t-1} + z_t), \\
INFL_t &= \pi^{(A)} + 400\pi_t, \\
INT_t &= \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400r_t,
\end{aligned}
$$

where

$$
\gamma^{(Q)} = 100(\gamma - 1), \quad \pi^{(A)} = 400(\bar{\pi} - 1), \quad r^{(A)} = 400(\frac{1}{\beta} - 1),
$$

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

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

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

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

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

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

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

Applying Theorem 1.2, we find $rank(\overline{G}(\bar{\theta}_0)) = 12$. Hence, $\bar{\theta}_0$ is not identifiable from the first and second order properties of the observables either. After applying the procedure from Section 1.3.1, 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 $(\nu, \bar{\pi})$ and $(\phi, \bar{\pi})$, as $\bar{\pi}$ enters $\mu(\bar{\theta})$ 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 1.5. *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 $\bar{\theta}_0$. Such a finding, first documented in this chapter, was also more recently documented in Komunjer and Ng (2011). This constitutes a serious concern for estimation in this and similar DSGE models.*

Remark 1.6. *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.*

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 1.4. 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 1.4). 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 $\bar{\theta}_0$. Note that applying this procedure, with step sizes and tolerance levels stated in Table 1.4, leads to the same results discussed in the two previous subsections.

Remark 1.7. *Based on the evidence reported here and our experimentation with other models, we suggest using $10^{-7} \times \theta_0$ and $size(G)eps(\|G\|)$ as the default step size and tolerance level when implementing the methods, followed by the two-step procedure outlined above to ensure robustness.*

1.4 FDQML estimation

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

1.4.1 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, i.e., $\omega_j = 2\pi j/T$ $(j = 1, 2, ..., T - 1)$. The discrete Fourier transforms are given by

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

Note that replacing Y_t by $Y_t - \mu(\bar{\theta})$ does not affect the value of $w_T(\omega_j)$ at these frequencies. $w_T(\omega_j)$ 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))

$$\frac{1}{\pi^{nr} \det(f_\theta(\omega_j))} \exp\left[-\text{tr}\left\{f_\theta^{-1}(\omega_j) w_T(\omega_j) w_T^*(\omega_j)\right\}\right], \quad j = 1, 2, ..., T-1.$$

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

$$-\sum_{j=1}^{T-1} \left[\log \det(f_\theta(\omega_j)) + \text{tr}\left\{f_\theta^{-1}(\omega_j) I_T(\omega_j)\right\}\right], \tag{1.14}$$

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 (1.14)

$$L_T(\theta) = -\sum_{j=1}^{T-1} W(\omega_j) \left[\log \det(f_\theta(\omega_j)) + \text{tr}\left\{f_\theta^{-1}(\omega_j) I_T(\omega_j)\right\}\right], \tag{1.15}$$

Then the FDQML estimator for θ is given by

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} L_T(\theta). \tag{1.16}$$

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

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

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 $j = 1, 2, ..., T-1$, we arrive at the

approximate log likelihood function of $\bar{\theta}$:

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

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

$$\widehat{\bar{\theta}}_T = \arg\max_{\bar{\theta}\in\bar{\Theta}}\bar{L}_T\left(\bar{\theta}\right). \tag{1.17}$$

1.4.2 Asymptotic properties of the FDQML estimators

The asymptotic properties of the estimator (1.16), 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 (1.17) received less attention. One exception is Hansen and Sargent (1993), who formally established that $T^{-1}\bar{L}_T\left(\bar{\theta}\right)$ 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 1.5. *(i)* $\{Y_t\}$ *is generated by*

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

with $Y_t^d(\theta)$ *satisfying (1.1).* *(ii)* $f_\theta(\omega)$ *is positive definite with eigenvalues bounded away*

from 0 and ∞ uniformly in ω for all $\theta \in \Theta$. The elements of $\partial vec(f_\theta(\omega))/\partial\theta'$ are bounded away from ∞ uniformly in ω for all $\theta \in \Theta$. The elements of $f_\theta(\omega)$ belong to $Lip(\beta)$ with respect to ω, the Lipschitz class of degree $\beta, \beta > 1/2$.

Assumption 1.6. ϵ_t *is fourth-order stationary. Let $Q_{h,l,g,k}\,(j_1, j_2, j_3)$ be the joint cumulant of $\epsilon_{th}, \epsilon_{(t+j_1)l}, \epsilon_{(t+j_2)g}$ 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 1.5 states that the model is correctly specified. This is be relaxed in Section 1.4.3. The second part strengthens the first condition in Assumption 1.3. It is satisfied by stationary finite order vector autoregressive moving average (VARMA) processes with finite error covariance matrices, which are the forms that the solutions to linearized DSGE models typically take. In Assumption 1.6, 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.

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 $T^{-1}\bar{L}_T(\bar{\theta})$:

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

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

Lemma 1.1. *Let Assumptions 1.1-1.3, 1.5 and 1.6 hold. Then*

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

2. 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 $L_\infty(\theta)$.

---

[8] *The parameter vector θ is said to be globally identifiable from the second order properties of $\{Y_t\}$ at a*

3. $\hat{\theta}_T \to^p \theta_0$ *if one of the following two conditions is satisfied: i)* θ_0 *is globally identified, or ii)* θ_0 *is locally identified and the maximization is carried over the corresponding small neighborhood of identification, say* $\delta(\theta_0)$*, instead of* Θ*.*

4. Let Assumptions 1.1-1.6 hold. Then Properties 1-3 hold when θ*,* $\theta_0, \hat{\theta}_T, L_T(\theta)$*, 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 3. The second result formally establishes the close link between the identification conditions and the asymptotic properties of the FDQML 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 3 by a uniform weak law of large numbers. Property 4 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 1.7. *(i)* $\{\epsilon_t\}$ *is a vector of martingale difference sequences with respect to the* $\sigma-$*field generated by* $\epsilon_s : s \leq t$. $E(\epsilon_{ta}\epsilon_{tb}|\mathcal{F}_{t-\tau}) = \Sigma_{ab}$, $E(\epsilon_{ta}\epsilon_{tb}\epsilon_{tc}|\mathcal{F}_{t-\tau}) = \xi_{abc}$, $E(\epsilon_{ta}\epsilon_{tb}\epsilon_{tc}\epsilon_{td}|\mathcal{F}_{t-\tau}) = \zeta_{abcd}$ *a.s. with* $\Sigma_{aa} > 0$ *and* $\zeta_{aadd} > 0$ *for all* $1 \leq a, b, c, d \leq n_\epsilon$. *(ii) Let* $c(t, r) = \epsilon_t \epsilon'_{t+r} - E(\epsilon_t \epsilon'_{t+r})$. *Assume* $\lim_{T \to \infty} T^{-1} \sum_{r=0}^{L} \sum_{t=1}^{T} E\left[c_{ab}(t, r)^2 \mathbf{1}\left\{c_{ab}(t, r)^2 > \epsilon T\right\}\right] < \epsilon$ *holds for any* $\epsilon > 0$, $L < \infty$, *and all* $1 \leq a, b \leq n_\epsilon$.

Part (i) of Assumption 1.7 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

---

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* $\theta_0 = \theta_1$.

Lindeberg-type condition. It ensures that the sample autocovariances $T^{-1/2}\sum_{t=1}^{T-r}c(t,r)$

$(r=0,1,...,L)$ 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 $\widehat{\bar{\theta}}_T$.

Theorem 1.3. *Suppose θ_0 and $\bar{\theta}_0$ are globally identified or the maximizations (1.16) and (1.17) are over convex compact sets in which they are locally identified and are interior points.*

1. *Let Assumptions 1.1-1.3 and 1.5-1.7 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

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

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

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

where $[.]_{ab}$ denotes the (a,b)-th element of the matrix, κ_{abcd} is the fourth cross-cumulant of $\epsilon_{ta},\epsilon_{tb},\epsilon_{tc}$, and ϵ_{td}, $H(\omega) = H(\exp(-i\omega);\theta_0) = \sum_{j=0}^{\infty}h_j(\theta_0)\exp(-i\omega j)$ (see (1.3)), and $H^*(\omega)$ is its conjugate transpose.

2. *Let Assumptions 1.1-1.7 hold. Then $\sqrt{T}(\widehat{\bar{\theta}}_T - \bar{\theta}_0) \to^d N(0, \bar{M}^{-1}\bar{V}\bar{M}^{-1})$, where $\bar{M}$ and $\bar{V}$ are $(q+p) \times (q+p)$ matrices, with the (j,l)-th element given by*

$$\bar{M}_{jl} = \int_{-\pi}^{\pi} W(\omega)tr\left\{f_{\theta_0}(\omega)\frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial\bar{\theta}_j}f_{\theta_0}(\omega)\frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial\bar{\theta}_l}\right\}d\omega + 2\frac{\partial\mu(\bar{\theta}_0)'}{\partial\bar{\theta}_j}f_{\theta_0}^{-1}(0)\frac{\partial\mu(\bar{\theta}_0)}{\partial\bar{\theta}_l}.$$

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

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

$$\text{with} \quad A_{jl} \quad = \quad 2\sum_{a,b,c=1}^{n_\epsilon} \xi_{abc} \left\{ \int_{-\pi}^{\pi} W(\omega) \left[H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_j} H(\omega) \right]_{ab} d\omega \right\} \times$$

$$\left[\frac{\partial \mu(\bar\theta_0)'}{\partial \theta_l} f_{\theta_0}^{-1}(0) H(0) \right]_c \quad and \quad \xi_{abc} = E(\epsilon_{ta}\epsilon_{tb}\epsilon_{tc}).$$

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 (1.1). 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

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

or, in matrix notation,

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

$$(1.18)$$

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 $\hat{\bar{\theta}}_T$ and applying (1.2) and (1.4). The derivatives and the integrals can be evaluated numerically. The cumulants ξ_{abc} and κ_{abcd} can be replaced by their sample counterparts.

1.4.3 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 1.5(ii) with $f_\theta(\omega)$ replaced by $f_0(\omega)$ and Assumptions 1.6 and 1.7 with ϵ_t replaced by ε_t.*

Suppose the estimates $\hat{\theta}_T$ and $\hat{\bar{\theta}}_T$ 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

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

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

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

Corollary 1.7. *Suppose θ_0^m and $\bar{\theta}_0^m$ are globally identified or the maximizations (1.16) and (1.17) are over convex compact sets in which they are locally identified and are interior points. Let Assumption MI hold.*

1. *Assume the DSGE solution $Y_t^d(\theta)$ satisfies Assumptions 1.1-1.3 and 1.5(ii). Then*

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

with

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

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

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

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

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

2. *Assume the DSGE solution is given by $\mu(\bar{\theta}) + Y_t^d(\theta)$ and satisfies Assumptions 1.1-1.4 and 1.5(ii). Then, $\sqrt{T}(\hat{\bar{\theta}}_T - \bar{\theta}_0^m) \to^d N(0, \bar{\Omega}^{-1}\bar{\Pi}\bar{\Omega}^{-1})$ with*

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

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

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

$$+ 8\pi\frac{\partial\mu(\bar{\theta}_0^m)'}{\partial\bar{\theta}_j}f_{\theta_0^m}^{-1}(0)\frac{\partial\mu(\bar{\theta}_0^m)}{\partial\bar{\theta}_l}$$

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

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

$$\text{with} \quad A_{jl} \quad = \quad 2\sum_{a,b,c,d=1}^{n_\epsilon} \xi_{abc} \left\{ \int_{-\pi}^{\pi} W(\omega) \left[H_0^*(\omega) \frac{\partial f_{\theta_0^m}^{-1}(\omega)}{\partial \theta_j} H_0(\omega) \right]_{ab} d\omega \right\} \times$$

$$\left[\frac{\partial \mu(\bar\theta_0^m)'}{\partial \theta_l} f_{\theta_0^m}^{-1}(0) H_0(0) \right]_c \quad \text{and} \quad \xi_{abc} = E(\varepsilon_{ta}\varepsilon_{tb}\varepsilon_{tc}).$$

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 $\hat\theta_T$.

1.5 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

$$p_T(\theta) = \frac{\pi(\theta)\exp\left(L_T(\theta)\right)}{\int_\Theta \pi(\theta)\exp\left(L_T(\theta)\right)d\theta}, \tag{1.19}$$

where $L_T(\theta)$ is the same as in (1.15) 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 $\exp\left(L_T(\theta)\right)$ 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

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

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 1.4. *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 $\hat{\theta}_T$ ($\widehat{\bar{\theta}}_T$) 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 the quantiles of the MCMC sequence $\left(\theta^{(1)}, \theta^{(2)}, ..., \theta^{(B)} \right)$. 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 $\left(\theta^{(1)}, \theta^{(2)}, ..., \theta^{(B)}\right)$. Let $\hat{V}$ be an estimator for V, which can be obtained using the formula in Theorem 1.3 by replacing $H(\omega)$, κ_{abcd}, and $\partial f_{\theta_0}^{-1}(\omega)/\partial \theta_j$ $(j = 1, 2, ..., q)$ 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 1.4 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 this

chapter, 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 (1.14) 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.

1.6 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 (FDQML) 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 chapter 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.

1.7 Mathematical appendix 1

The spectral density matrix $f_\theta(\omega)$ is a Hermitian matrix satisfying $f_\theta(\omega)^* = f_\theta(\omega)$. It is in general not symmetric. The following correspondence is useful for understanding and proving the identification results:

$$f_\theta(\omega) \longleftrightarrow f_\theta(\omega)^R \quad \text{with } f_\theta(\omega)^R = \begin{bmatrix} Re(f_\theta(\omega)) & Im(f_\theta(\omega)) \\ -Im(f_\theta(\omega)) & Re(f_\theta(\omega)) \end{bmatrix}, \qquad (1.20)$$

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

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

The following lemma is crucial for proving the subsequent results.

Lemma 1.2. *We have the identity*

$$\left(\frac{\partial vec(f_\theta(\omega)')}{\partial \theta'}\right)' \left(\frac{\partial vec(f_\theta(\omega))}{\partial \theta'}\right) = \frac{1}{2}\left(\frac{\partial R(\omega;\theta)}{\partial \theta'}\right)' \left(\frac{\partial R(\omega;\theta)}{\partial \theta'}\right). \qquad (1.21)$$

Proof of Lemma 1.2. The (j,k)-th element of the term on the left hand side is equal to

$$\left(\frac{\partial vec(f_\theta(\omega)')}{\partial \theta_j}\right)' \left(\frac{\partial vec(f_\theta(\omega))}{\partial \theta_k}\right)$$

$$= \text{tr}\left\{\frac{\partial f_\theta(\omega)}{\partial \theta_j}\frac{\partial f_\theta(\omega)}{\partial \theta_k}\right\} = \text{tr}\left\{Re\left(\frac{\partial f_\theta(\omega)}{\partial \theta_j}\frac{\partial f_\theta(\omega)}{\partial \theta_k}\right)\right\}$$

$$= \frac{1}{2}\text{tr}\left\{\left(\frac{\partial f_\theta(\omega)}{\partial \theta_j}\frac{\partial f_\theta(\omega)}{\partial \theta_k}\right)^R\right\} = \frac{1}{2}\text{tr}\left\{\frac{\partial(f_\theta(\omega)^R)}{\partial \theta_j}\frac{\partial(f_\theta(\omega)^R)}{\partial \theta_k}\right\}$$

$$= \frac{1}{2}\left(\frac{\partial vec\left(f_\theta(\omega)^R\right)}{\partial \theta_j}\right)' \left\{\frac{\partial vec\left(f_\theta(\omega)^R\right)}{\partial \theta_k}\right\},$$

where the first equality is because of the identity $vec(A')'vec(B) = \text{tr}(AB)$ for generic

matrices A and B, the second is because $f_\theta(\omega)$ is Hermitian, thus this term is real valued,

the third equality is because of the definition (1.20), the fourth is because, for generic

complex matrices, if $Z = XY$, then $Z^R = X^R Y^R$ (see Lemma 3.7.1(ii) in Brillinger (2001)),

and the fifth is because $f_\theta(\omega)^R$ is real and symmetric. The last term in the display is simply

the (j,k)-th element of the right hand side term in (1.21). This completes the proof. ■

Proof of Theorem 1.1. Lemma 1.2 implies that $G(\theta)$ defined by (1.9) is real, symmetric,

positive semidefinite, and equal to

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

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

result.

Suppose θ_0 is *not* locally identified. Then there exists an infinite sequence of vectors

$\{\theta_k\}_{k=1}^{\infty}$ approaching θ_0 such that, for each k,

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

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

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

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

$$d_k = \frac{\theta_k - \theta_0}{\|\theta_k - \theta_0\|},$$

then

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

The sequence $\{d_k\}$ is an infinite sequence on the unit sphere and therefore there exists a limit point d (note that d does not depend on j or ω). As $\theta_k \to \theta_0$, d_k approaches d and we have

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

where the convergence result holds because $f_\theta(\omega)$ is continuously differentiable in θ (Assumption 1.3). Because this holds for an arbitrary j, it holds for the full vector $R(\omega; \theta_0)$. Therefore,

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

which implies

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

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

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

Applying Lemma 1.2, because $d \neq 0$, $G(\theta_0)$ is singular.

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

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

we have

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

Since the integrand is continuous in ω and always nonnegative, we must have

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

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

$$\frac{\partial R(\omega; \theta)}{\partial \theta'} c(\theta) = 0 \tag{1.22}$$

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

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

$$\theta(0) = \theta_0.$$

Then

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

for all $\omega \in [-\pi, \pi]$ and $0 \le v \le \bar{v}$, where the last equality uses (1.22). Thus, $R(\omega; \theta)$ is constant on the curve χ. This implies that $f_\theta(\omega)$ is constant on the same curve and that θ_0 is unidentifiable. This completes the proof. ∎

Proof of Corollary 1.1. The statement in the subsequent proof applies to all $\omega \in [-\pi, \pi]$. Using the same argument as in the proof of Lemma 1.2, $I(\theta_0)$ can be rewritten as

$$I(\theta_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\partial R(\omega; \theta_0)}{\partial \theta'} \right)' \left([f_{\theta_0}(\omega)^R]^{-1} \otimes [f_{\theta_0}(\omega)^R]^{-1} \right) \frac{\partial R(\omega; \theta_0)}{\partial \theta'} d\omega. \qquad (1.23)$$

Because spectral density matrices are Hermitian and positive semidefinite, $f_{\theta_0}(\omega)^R$ is real, symmetric, and positive semidefinite (see Lemma 3.7.1 (vii) in Brillinger (2001)). Furthermore, because here $f_{\theta_0}(\omega)$ has full rank, $f_{\theta_0}(\omega)^R$ is in fact positive definite. Thus, $([f_{\theta_0}(\omega)^R]^{-1} \otimes [f_{\theta_0}(\omega)^R]^{-1})$ is positive definite (see Theorem 1 in Magnus and Neudecker (1999, p. 28)).

We now prove $G(\theta_0)$ and $I(\theta_0)$ have the same null space. Since they are both $q \times q$ matrices, the result then implies they have the same rank. First, suppose $c \in R^q$ and $I(\theta_0)c = 0$. Then $c'I(\theta_0)c = 0$ or, explicitly,

$$\int_{-\pi}^{\pi} \left(\frac{\partial R(\omega; \theta_0)}{\partial \theta'} c \right)' \left([f_{\theta_0}(\omega)^R]^{-1} \otimes [f_{\theta_0}(\omega)^R]^{-1} \right) \left(\frac{\partial R(\omega; \theta_0)}{\partial \theta'} c \right) d\omega = 0.$$

Because the integrand is continuous in ω and always nonnegative, we must have

$$\left(\frac{\partial R(\omega; \theta_0)}{\partial \theta'} c \right)' \left([f_{\theta_0}(\omega)^R]^{-1} \otimes [f_{\theta_0}(\omega)^R]^{-1} \right) \left(\frac{\partial R(\omega; \theta_0)}{\partial \theta'} c \right) = 0.$$

Because $\left(\left[f_{\theta_0}(\omega)^R \right]^{-1} \otimes \left[f_{\theta_0}(\omega)^R \right]^{-1} \right)$ is positive definite, this implies

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

Therefore,

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

and, consequently, $G(\theta_0)c = 0$. Next suppose $c \in R^q$ and $G(\theta_0)c = 0$. Applying the same argument that leads to (1.22), we have

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

Then, trivially,

$$\left(\frac{\partial R(\omega; \theta_0)}{\partial \theta'} \right)' \left(\left[f_{\theta_0}(\omega)^R \right]^{-1} \otimes \left[f_{\theta_0}(\omega)^R \right]^{-1} \right) \left(\frac{\partial R(\omega; \theta_0)}{\partial \theta'} c \right) = 0.$$

Upon integration, we have $I(\theta_0)c = 0$. ∎

Proof of Theorem 1.2. Using Lemma 1.2 again, $\bar{G}(\bar{\theta})$ can be equivalently represented as

$$\bar{G}(\bar{\theta}) = \frac{1}{2} \int\limits_{-\pi}^{\pi} \left(\frac{\partial R(\omega; \theta)}{\partial \bar{\theta}'} \right)' \left(\frac{\partial R(\omega; \theta)}{\partial \bar{\theta}'} \right) d\omega + \left(\frac{\partial \mu(\bar{\theta})}{\partial \bar{\theta}'} \right)' \frac{\partial \mu(\bar{\theta})}{\partial \bar{\theta}'}$$

with both terms on the right hand side being real, symmetric, and positive semidefinite.

Let

$$\bar{R}(\omega; \bar{\theta}) = \begin{bmatrix} R(\omega; \theta) \\ \frac{1}{\sqrt{\pi}} \mu(\bar{\theta}) \end{bmatrix},$$

then

$$\bar{G}(\bar{\theta}) = \frac{1}{2} \int\limits_{-\pi}^{\pi} \left(\frac{\partial \bar{R}(\omega; \bar{\theta})}{\partial \bar{\theta}'} \right)' \left(\frac{\partial \bar{R}(\omega; \bar{\theta})}{\partial \bar{\theta}'} \right) d\omega.$$

Using this representation, the proof proceeds in the same way as in Theorem 1.1, with θ

replaced by $\bar{\theta}$ and $R(\omega; \theta)$ replaced by $\bar{R}(\omega; \bar{\theta})$. The detail is omitted. ∎

Proof of Corollary 1.3. We only prove the first result, as the second can be proven analogously using the formulation in the proof of Theorem 1.2.

Suppose the subvector θ_0^s is *not* locally identified. Write $\theta = (\theta^{s\prime}, \theta^{r\prime})'$. There exists an infinite sequence of vectors $\{\theta_k\}_{k=1}^{\infty}$ approaching θ_0 such that

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

By the definition of partial identification, $\{\theta_k^s\}$ can be chosen so that $\|\theta_k^s - \theta_0^s\| / \|\theta_k - \theta_0\| > \varepsilon$, with ε being some arbitrarily small positive number. The values of θ_k^r can either change or stay fixed in this sequence; no restriction is imposed on them besides those in the preceding display. As in the proof of Theorem 1.1, in the limit, we have

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

with $d^s \neq 0$ (where d^s comprises the elements in d that correspond to θ^s). Therefore, on one hand,

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

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

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

which implies

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

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

this result implies $G^a(\theta_0)$ has a higher column rank than $G(\theta_0)$.

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

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

which implies for all $\omega \in [-\pi, \pi]$ and all $\theta \in \delta(\theta_0)$ (see arguments leading to (1.22)),

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

but

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

where $c^s(\theta)$ denotes the elements in $c(\theta)$ that correspond to θ^s. Because $G(\theta)$ is continuous and has constant rank in $\delta(\theta_0)$, the vector $c(\theta)$ is continuous in $\delta(\theta_0)$. As in Theorem 1.1, consider the curve χ defined by the function $\theta(v)$ which solves, for $0 \leq v \leq \bar{v}$, the differential equation

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

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

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

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

Proof of Corollary 1.3. The proof is essentially the same as in Rothenberg (1971,

Theorem 2) and is included for the sake of completeness. Suppose $\Psi(\theta)$ has rank s for all

θ in a neighborhood of θ_0. Then, by the implicit function theorem, there exists a partition

of θ into $\theta^1 \in R^s$ and $\theta^2 \in R^{q-s}$ such that

$$\theta^1 = q(\theta^2)$$

for all solutions of $\psi(\theta) = 0$ in a neighborhood of θ_0 with θ_0^2 being an interior point of that

neighborhood. Consequently, the spectral density can be rewritten as

$$f_\theta(\omega) = f_{q(\theta^2),\theta^2}(\omega),$$

which involves only $q - s$ parameters. Let

$$Q(\theta^2) = \frac{\partial q(\theta^2)}{\partial \theta^{2\prime}} \quad \text{and} \quad \tilde{G}(\theta) = \begin{bmatrix} Q(\theta^2)' & I \end{bmatrix} G(\theta) \begin{bmatrix} Q(\theta^2) \\ I \end{bmatrix}.$$

Then, by Theorem 1.1, θ_0 is identified if and only if $\tilde{G}(\theta_0)$ has full rank.

Suppose there exists a vector $d \in R^{q-s}$ such that

$$\tilde{G}(\theta_0) d = 0. \tag{1.24}$$

Then the structure of $G(\theta)$ (see Lemma 1.2) implies that (1.24) holds if and only if

$$G(\theta_0) \begin{bmatrix} Q(\theta_0^2) \\ I \end{bmatrix} d = 0.$$

Let

$$c = \begin{bmatrix} Q(\theta_0^2) \\ I \end{bmatrix} d.$$

Then we have: (1) $c \neq 0$ if and only if $d \neq 0$, and (2)

$$\left[\begin{array}{c} G(\theta_0) \\ \Psi(\theta_0) \end{array}\right] c = 0$$

if and only if (1.24) holds, where $\Psi(\theta_0)c = 0$ always holds because θ_0 satisfies the constraint $\psi(\theta) = 0$. Thus, the preceding matrix has full rank if and only if θ_0 is identified under the constraints. This completes the proof. ■

Proof of Corollary 1.6. Without loss of generality, assume $n_Y = 1$. Otherwise, the proof can be carried out by analyzing $R(\omega; \theta)$. The map $\theta \longmapsto f_\theta$ is infinite dimensional. The proof therefore involves two steps. The first is to reduce it to a finite dimensional problem. The second is to apply a constant rank theorem (a generalization of the implicit function theorem).

Consider a positive integer N and a partition of the interval $[-\pi, \pi]$ by $\omega_j = (2\pi j/2^N) - \pi$, with $j = 0, 1, ..., 2^N$. Then, the map

$$\theta \longmapsto (f_\theta(\omega_0), ..., f_\theta(\omega_{2^N})).$$ \hfill (1.25)

is finite dimensional. To simplify notation, let $f_{\theta,N} = (f_\theta(\omega_0), ..., f_\theta(\omega_{2^N}))'$. Conventionally, the rank of the above map is defined as the rank of the Jacobian matrix $\partial f_{\theta,N}/\partial \theta'$, which is of dimension $(2^N + 1) \times q$ with rank no greater than $q - 1$ at θ_0, because if the rank equals q, then θ_0 becomes locally identified, contradicting the assumption in the corollary. Note that, for a given N, its rank can be strictly less than $q - 1$.

We now show that there exists a finite N such that $\partial f_{\theta,N}/\partial \theta'$ has rank $q - 1$ at θ_0. Suppose such an N does not exist. Then the rank of $\partial f_{\theta,N}/\partial \theta'$ is at most $q - 2$ for arbitrarily

large N. This implies that the rank of

$$G_N(\theta_0) = \frac{2\pi}{2^N + 1} \sum_{j=0}^{2^N} \left(\frac{\partial f_{\theta_0}(\omega_j)}{\partial \theta'} \right)' \left(\frac{\partial f_{\theta_0}(\omega_j)}{\partial \theta'} \right)$$

is at most $q - 2$ for arbitrarily large N, because vectors orthogonal to $\partial f_{\theta,N}/\partial \theta'$ are also orthogonal to $G_N(\theta)$ by construction. Let $\lambda_{N,j}$ $(j = 1, ..., q)$ be the eigenvalues of $G_N(\theta_0)$ sorted in an increasing order. Then, for any finite N,

$$\lambda_{N,1} = \lambda_{N,2} = 0.$$

On the other hand, because $G_N(\theta_0) \to G(\theta_0)$, so do its eigenvalues. Thus, for any $\varepsilon > 0$, there exists a finite N such that $|\lambda_2 - \lambda_{N,2}| < \varepsilon$, where λ_2 is the second smallest eigenvalue of $G(\theta_0)$. Choosing $\varepsilon = \lambda_2/2$ leads to

$$|\lambda_{N,2}| > \lambda_2/2.$$

Since $\text{rank}(G(\theta_0)) = q - 1$ by Assumption, λ_2 is strictly positive. Thus, we reach a contradiction. Because the convergence of $G_N(\theta) \to G(\theta)$ is uniform in an open neighborhood of θ_0, say $\delta(\theta_0)$, the above analysis also implies there exists an N such that $\partial f_{\theta,N}/\partial \theta'$ has constant rank $q - 1$ in that neighborhood.

Use such an N and consider again the map $\theta \longmapsto f_{\theta,N}$, which is finite dimensional, is continuously differentiable, and has constant rank $q - 1$ in $\delta(\theta_0)$. Define the level set

$$\{\theta \in \delta(\theta_0) : f_{\theta,N} = f_{\theta_0,N}\}.$$

Then the rank theorem (Krantz and Parks (2002, Theorem 3.5.1 and the discussion on p. 56)) implies that the level set constitutes a smooth, parameterized one dimensional manifold. Thus, there exists a unique level curve passing through θ_0 satisfying $f_{\theta,N} = f_{\theta_0,N}$.

Therefore, we have established the result for a particular finite N. Further increasing N leads to finer partitions of $[-\pi, \pi]$. This cannot decrease the rank of the map (1.25), and therefore cannot increase the number of level curves passing through θ_0. Thus, in the limit, there is at most one level curve passing through θ_0. The existence of such a curve for the infinite dimensional case has already been shown in the main text, given by (1.10). This completes the proof. ∎

Proof of Lemma 1.1. Applying Lemma A.3.3 (1) in Hosoya and Taniguchi (1982), for a given $\theta \in \Theta$, we have

$$plim_{T \to \infty} \frac{1}{T} \sum_{j=1}^{T-1} \text{tr} \left\{ W(\omega_j) f_\theta^{-1}(\omega_j) I_T(\omega_j) \right\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ W(\omega) f_\theta^{-1}(\omega) f_{\theta_0}(\omega) \right\} d\omega.$$

To prove stochastic equicontinuity, consider for any $\theta_1, \theta_2 \in \Theta$,

$$\frac{1}{T} \sum_{j=1}^{T-1} \text{tr} \left\{ W(\omega_j) \left(f_{\theta_1}^{-1}(\omega_j) - f_{\theta_2}^{-1}(\omega_j) \right) I_T(\omega_j) \right\}.$$

Apply a first order Taylor expansion

$$\frac{1}{T} \sum_{j=1}^{T-1} \text{tr} \left\{ W(\omega_j) \left(f_{\theta_1}^{-1}(\omega_j) - f_{\theta_2}^{-1}(\omega_j) \right) I_T(\omega_j) \right\}$$

$$= \frac{1}{T} \sum_{j=1}^{T-1} \frac{\partial \text{tr} \left\{ W(\omega_j) f_{\tilde{\theta}}^{-1}(\omega_j) I_T(\omega_j) \right\}}{\partial \theta'} (\theta_1 - \theta_2)$$

$$= -\frac{1}{T} \sum_{j=1}^{T-1} W(\omega_j) vec \left(I_T(\omega_j)' \right)' \{ f_{\tilde{\theta}}^{-1}(\omega_j)' \otimes f_{\tilde{\theta}}^{-1}(\omega_j) \} \frac{\partial vec \left(f_{\tilde{\theta}}(\omega_j) \right)}{\partial \theta'} (\theta_1 - \theta_2) \quad (1.26)$$

where $\tilde{\theta}$ lies between θ_1 and θ_2. The norm of (1.26) is bounded by

$$\frac{1}{T} \sum_{j=1}^{T-1} \left\| vec \left(I_T(\omega_j)' \right) \right\| \left\| \{ f_{\tilde{\theta}}^{-1}(\omega_j)' \otimes f_{\tilde{\theta}}^{-1}(\omega_j) \} \frac{\partial vec \left(f_{\tilde{\theta}}(\omega_j) \right)}{\partial \theta'} \right\| \| \theta_1 - \theta_2 \|.$$

The quantity

$$\left\| (f_{\tilde{\theta}}^{-1}(\omega_j)' \otimes f_{\tilde{\theta}}^{-1}(\omega_j)) \frac{\partial vec\left(f_{\tilde{\theta}}(\omega_j)\right)}{\partial \theta'} \right\|$$

is uniformly bounded by Assumption 1.5(ii). The term $T^{-1} \sum_{j=1}^{T-1} \left\| vec(I_T(\omega_j)') \right\|$ only

depends on θ_0 and is $O_p(1)$ because the diagonal elements of $T^{-1} \sum_{j=1}^{T-1} I_T(\omega_j)$ are positive

and satisfy a law of large numbers (Hosoya and Taniguchi (1982, Lemma A.3.3 (1))), and

the norm of the off-diagonal elements can be bounded by the diagonal elements using the

Cauchy–Schwarz inequality. Therefore, the term (1.26) can be made uniformly small by

choosing a small $\|\theta_1 - \theta_2\|$. Meanwhile,

$$\frac{1}{T} \sum_{j=1}^{T-1} W(\omega_j) \log \det f_\theta(\omega_j) \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) \log \det f_\theta(\omega) dw$$

uniformly in $\theta \in \Theta$. Thus, the first result holds.

For the second result, we first show that θ_0 maximizes $L_\infty(\theta)$. Apply the same argu-

ment as in Hosoya and Taniguchi (1982, p. 149). For every $\omega \in [-\pi, \pi]$,

$$W(\omega) \left[\log \det f_\theta(\omega) + tr \left\{ f_\theta^{-1}(\omega) f_{\theta_0}(\omega) \right\} \right]$$

$$= W(\omega) \log \det f_{\theta_0}(\omega) + W(\omega) \left[tr \left\{ f_\theta^{-1}(\omega) f_{\theta_0}(\omega) \right\} - \log \det \left\{ f_\theta^{-1}(\omega) f_{\theta_0}(\omega) \right\} \right]$$

$$= W(\omega) \log \det f_{\theta_0}(\omega) + W(\omega) \left[\sum_{j=1}^{n_Y} \lambda_j(\omega) - \log \lambda_j(\omega) - 1 \right] + W(\omega) n_Y,$$

where $\lambda_j(\omega)$ is the j-th eigenvalue of $f_\theta^{-1}(\omega) f_{\theta_0}(\omega)$. Because $\lambda_j(\omega) - \log \lambda_j(\omega) - 1 \geq 0$ and

the equality holds if and only if $\lambda_j(\omega) = 1$, $j = 1, ..., n_Y$, this implies

$$L_\infty(\theta) \leq -\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) \left(\log \det f_{\theta_0}(\omega) + n_Y \right) dw,$$

which holds with equality if and only if $\lambda_j(\omega) = 1$ for all $\omega \in [-\pi, \pi]$ $(j = 1, ..., n_Y)$.

However, $\lambda_j(\omega) = 1$ $(j = 1, ..., n_Y)$ implies $f_{\theta_0}(\omega) = f_\theta(\omega)$ because the latter are positive

definite Hermitian matrices. Hence, θ_0 is a global maximizer.

The above result implies that any other parameter vector, say θ_1, is a maximizer if and only if $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$. Now suppose the parameters are locally identified. Then there are no parameter values close to θ_0 satisfying this equality. Thus, θ_0 is the locally unique maximizer. To see the converse, suppose θ_0 is the locally unique maximizer. Then there cannot be any parameter close to θ_0 satisfying $f_{\theta_0}(\omega) = f_\theta(\omega)$ for all ω. Thus, by definition, we have local identification. The argument to establish the result for the global identification proceeds in the same way.

The third result follows directly from the uniform weak law of large numbers. ∎

Proof of Theorem 1.3. We only prove the second result, which includes the first as a special case. The first order condition (FOC) gives

$$
2\pi T^{-1/2} \sum_{j=0}^{T-1} W(\omega_j) \frac{\partial \mathrm{vec}\left(f_{\widehat{\theta}_T}(\omega_j)'\right)'}{\partial \bar{\theta}} \left\{ f_{\widehat{\theta}_T}^{-1}(\omega_j)' \otimes f_{\widehat{\theta}_T}^{-1}(\omega_j) \right\} \mathrm{vec}(I_T(\omega_j) - f_{\widehat{\theta}_T}(\omega_j))
$$
$$
+ 2T^{-1/2} \sum_{t=1}^{T} \frac{\partial \mu(\widehat{\theta}_T)'}{\partial \bar{\theta}} f_{\widehat{\theta}_T}^{-1}(0) \left(Y_t - \mu(\widehat{\theta}_T) \right) = 0.
$$

Note that the first summation starts at $j = 0$ and $I_T(0) = I_{\widehat{\theta}_T, T}(0)$. The above FOC implies

$$
2\pi T^{-1/2} \sum_{j=0}^{T-1} W(\omega_j) \frac{\partial \mathrm{vec}(f_{\theta_0}(\omega_j)')'}{\partial \bar{\theta}} \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j) \right) \mathrm{vec}\left(I_T(\omega_j) - f_{\widehat{\theta}_T}(\omega_j) \right)
$$
$$
+ 2T^{-1/2} \sum_{t=1}^{T} \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}} f_{\theta_0}^{-1}(0)(Y_t - \mu(\widehat{\theta}_T)) = o_p(1),
$$

which holds because $\widehat{\theta}_T \to^p \bar{\theta}_0, f_{\theta_0}(\omega_j)$ and $\mu(\bar{\theta}_0)$ are continuously differentiable, and $f_{\theta_0}^{-1}(\omega_j)$ have bounded eigenvalues. Apply a first order Taylor expansion around $\bar{\theta}_0$. Then

the left hand side of the preceding display is equal to

$$2\pi T^{-1/2} \sum_{j=0}^{T-1} W(\omega_j) \frac{\partial \text{vec}(f_{\theta_0}(\omega_j)')'}{\partial \bar{\theta}} \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j) \right) \tag{1.27}$$

$$\times \text{vec}\left(I_T(\omega_j) - f_{\theta_0}(\omega_j) \right) \quad \text{(I)} \tag{1.28}$$

$$+2T^{-1/2} \sum_{t=1}^{T} \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}} f_{\theta_0}^{-1}(0) \left(Y_t - \mu(\bar{\theta}_0) \right) \quad \text{(II)}$$

$$-2\pi T^{-1} \sum_{j=0}^{T-1} W(\omega_j) \frac{\partial \text{vec}(f_{\theta_0}(\omega_j)')'}{\partial \bar{\theta}} \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j) \right) \frac{\partial \text{vec}\left(f_{\theta_0}(\omega_j) \right)}{\partial \bar{\theta}'}$$

$$\times T^{1/2}(\widehat{\bar{\theta}} - \bar{\theta}_0) \quad \text{(III)} \tag{1.29}$$

$$-2 \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}} f_{\theta_0}^{-1}(0) \frac{\partial \mu(\bar{\theta}_0)}{\partial \bar{\theta}'} T^{1/2}(\widehat{\bar{\theta}} - \bar{\theta}_0) \quad \text{(IV)}$$

$$+o_p(1).$$

First consider term (III). The quantity in front of $T^{1/2}(\widehat{\bar{\theta}} - \bar{\theta}_0)$ converges to

$$\int_{-\pi}^{\pi} W(\omega) \frac{\partial \text{vec}(f_{\theta_0}(\omega)')'}{\partial \bar{\theta}} \left(f_{\theta_0}^{-1}(\omega)' \otimes f_{\theta_0}^{-1}(\omega) \right) \frac{\partial \text{vec}\left(f_{\theta_0}(\omega) \right)}{\partial \bar{\theta}'} d\omega,$$

whose (h, k)-th element is given by

$$\int_{-\pi}^{\pi} \text{tr} \left\{ W(\omega) f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_h} f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_k} \right\} d\omega.$$

Therefore, the above expansion implies (see Theorem 1.3 for the definition of $\bar{M}$)

$$T^{1/2}(\widehat{\bar{\theta}} - \bar{\theta}_0) = \bar{M}^{-1} * \text{(I)} + \bar{M}^{-1} * \text{(II)} + o_p(1).$$

Term (I) satisfies a central limit theorem (CLT), whose covariance matrix has the (h, k)-th element given by (see Theorem 3.1 and Proposition 3.1 in Hosoya and Taniguchi (1982); in particular, their formula for U_{jl})

$$4\pi \int_{-\pi}^{\pi} W(\omega) \text{tr} \left\{ f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_h} f_{\theta_0}(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_k} \right\} d\omega$$

$$+ \sum_{a,b,c,d=1}^{n_t} \kappa_{abcd} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_h} H(\omega) d\omega \right]_{ab} \times$$

$$\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \bar{\theta}_k} H(\omega) d\omega \right]_{cd}.$$

Term (II) also satisfies a CLT, with covariance matrix given by

$$8\pi \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}} f_{\theta_0}^{-1}(0) \frac{\partial \mu(\bar{\theta}_0)}{\partial \bar{\theta}'}.$$

To complete the proof, we only need to verify the covariance matrix between (I) and (II).

Let

$$A = \mathrm{Cov}((I), (II))$$

and consider its (h, k)-th element

$$A_{hk} = 4\pi \mathrm{Cov} \left\{ \begin{array}{c} \mathrm{tr} \left(\frac{1}{\sqrt{T}} \sum_{j=0}^{T-1} W(\omega_j) \frac{\partial f_{\theta_0}^{-1}(\omega_j)}{\partial \bar{\theta}_h} (I_T(\omega_j) - f_{\theta_0}(\omega)) \right), \\ \left(\frac{1}{\sqrt{T}} \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}_k} f_{\theta_0}^{-1}(0) \sum_{t=1}^{T} (Y_t - \mu(\bar{\theta}_0)) \right) \end{array} \right\}.$$

Define

$$\phi^h(\omega_j) = \frac{\partial f_{\theta_0}^{-1}(\omega_j)}{\partial \bar{\theta}_h} \quad \text{and} \quad \psi^k(0) = \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}_k} f_{\theta_0}^{-1}(0).$$

Then

$$A_{hk} =$$

$$= 4\pi \mathrm{Cov} \left\{ \begin{array}{c} \mathrm{tr} \left(\frac{1}{\sqrt{T}} \sum_{j=0}^{T-1} W(\omega_j) \phi^h(\omega_j) (I_T(\omega_j) - f_{\theta_0}(\omega)) \right), \\ \left(\frac{1}{\sqrt{T}} \psi^k(0) \sum_{t=1}^{T} (Y_t - \mu(\bar{\theta}_0)) \right) \end{array} \right\}$$

$$= 4\pi \mathrm{Cov} \left\{ \begin{array}{c} \frac{1}{\sqrt{T}} \sum_{j=0}^{T-1} W(\omega_j) \sum_{a,b=1}^{n_Y} \phi_{ab}^h(\omega_j) (I_{Tba}(\omega_j) - f_{\theta_0 ba}(\omega)), \\ \frac{1}{\sqrt{T}} \sum_{c=1}^{n_Y} \psi_c^k(0) \sum_{t=1}^{T} (Y_{tc} - \mu_c(\bar{\theta}_0)) \end{array} \right\}$$

$$= 4\pi \sum_{a,b,c=1}^{n_Y} \text{Cov} \left\{ \begin{array}{c} \frac{1}{\sqrt{T}} \sum_{j=0}^{T-1} W(\omega_j)\phi_{ab}^h(\omega_j)(I_{Tba}(\omega_j) - f_{\theta_0 ba}(\omega)), \\ \frac{1}{\sqrt{T}} \psi_c^k(0) \sum_{t=1}^{T} (Y_{tc} - \mu_c(\bar{\theta}_0)) \end{array} \right\},$$

where $I_{Tba}(\omega_j)$ is the (b,a)-th element of $I_T(\omega_j)$ and other quantities are defined analogously. Consider the two terms inside the curly brackets separately. Applying the same argument as in Theorem 10.8.5 in Brockwell and Davis (1991), we have

$$\frac{1}{\sqrt{T}} \sum_{j=0}^{T-1} W(\omega_j)\phi_{a,b}^h(\omega_j)(I_{Tba}(\omega_j) - f_{\theta_0 ba}(\omega))$$

$$= \frac{1}{\sqrt{T}} \sum_{j=0}^{T-1} \sum_{f,g=1}^{n_\epsilon} W(\omega_j)\phi_{ab}^h(\omega_j)H_{bf}(\omega_j)\left(I_{Tfg}^\epsilon(\omega_j) - EI_{Tfg}^\epsilon(\omega_j)\right)H_{ga}^*(\omega_j) + o_p(1),$$

where and $I_{Tfg}^\epsilon(\omega_j)$ denote the (f,g)-th element of the periodogram of ϵ_t. Applying Theorem 10.3.1 in Brockwell and Davis (1991), we have

$$\frac{1}{\sqrt{T}} \psi_c^k(0) \sum_{t=1}^{T} (Y_{tc} - \mu_c(\bar{\theta}_0)) = \frac{1}{\sqrt{T}} \sum_{l=1}^{n_\epsilon} \sum_{t=1}^{T} \psi_c^k(0)H_{cl}(0)\epsilon_{tl} + o_p(1),$$

where $H(0) = \sum_{j=0}^{\infty} h_j(\theta_0)$ (see (1.3)). Therefore, their covariance is equal to

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{j=0}^{T-1} \sum_{f,g,l=1}^{n_\epsilon} W(\omega_j)\phi_{ab}^h(\omega_j)H_{bf}(\omega_j)H_{ga}^*(\omega_j)\psi_c^k(0)H_{cl}(0)$$

$$\times E\left\{ \left(I_{Tfg}^\epsilon(\omega_j) - EI_{Tfg}^\epsilon(\omega_j)\right)\epsilon_{tl} \right\} + o_p(1)$$

$$= \frac{1}{T} \sum_{t=1}^{T} \sum_{f,g,l=1}^{n_\epsilon} W(\omega_j)\phi_{ab}^h(\omega_j)H_{bf}(\omega_j)H_{ga}^*(\omega_j)\psi_c^k(0)H_{cl}(0)\xi_{fgl} + o_p(1)$$

$$= \frac{1}{2\pi} \sum_{f,g,l=1}^{n_\epsilon} \left\{ \int_{-\pi}^{\pi} W(\omega)H^*(\omega)_{ga}\phi_{ab}^h(\omega)H_{bf}(\omega_j)d\omega \right\} \times \xi_{fgl} \times \left\{ \psi_c^k(0)H_{cl}(0) \right\} + o_p(1).$$

Some algebra shows that

$$A_{hk} = 2 \sum_{f,g,l=1}^{n_\epsilon} \left[\int_{-\pi}^{\pi} W(\omega)H(\omega)^* \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_h} H(\omega)d\omega \right]_{gf} \times \xi_{gfl} \times \left[\frac{\partial \mu(\bar{\theta}_0)'}{\partial \theta_k} f_{\theta_0}^{-1}(0)H(0) \right]_l \blacksquare$$

Proof of Corollary 1.7. We prove the second result. Because the argument is very similar to Theorem 3 and Taniguchi (1979, Theorem 2), we only provide an outline. The estimate $\widehat{\bar{\theta}}$ solves

$$\frac{\partial \bar{L}_T(\widehat{\bar{\theta}})}{\partial \bar{\theta}} = 0, \tag{1.30}$$

and the pseudo-true value $\bar{\theta}_0^m$ satisfies

$$\frac{\partial \bar{L}_\infty^m(\bar{\theta}_0^m)}{\partial \bar{\theta}} = 0. \tag{1.31}$$

Consider a Taylor expansion of (1.30) around $\bar{\theta}_0^m$:

$$\frac{\partial \bar{L}_T(\bar{\theta}_0^m)}{\partial \bar{\theta}} + \frac{\partial^2 \bar{L}_T(\widetilde{\bar{\theta}})}{\partial \bar{\theta} \partial \bar{\theta}'}(\widehat{\bar{\theta}} - \bar{\theta}_0^m) = 0,$$

where $\widetilde{\bar{\theta}}$ lies between $\widehat{\bar{\theta}}$ and $\bar{\theta}_0^m$. Rearrange terms and apply (1.31):

$$T^{1/2}\left(\widehat{\bar{\theta}} - \bar{\theta}_0^m\right) = \left[-\frac{2\pi}{T}\frac{\partial^2 \bar{L}_T(\widetilde{\bar{\theta}})}{\partial \bar{\theta} \partial \bar{\theta}'}\right]^{-1}\left(2\pi T^{-1/2}\frac{\partial \bar{L}_T(\bar{\theta}_0^m)}{\partial \bar{\theta}} - 2\pi T^{1/2}\frac{\partial \bar{L}_\infty^m(\bar{\theta}_0^m)}{\partial \bar{\theta}}\right).$$

Furthermore,

$$-\frac{2\pi}{T}\frac{\partial^2 \bar{L}_T(\widetilde{\bar{\theta}})}{\partial \bar{\theta} \partial \bar{\theta}'}$$

$$\rightarrow \int_{-\pi}^{\pi} W(\omega)\left[\frac{\partial^2}{\partial \bar{\theta} \partial \bar{\theta}'}\log\det(f_{\theta_0^m}(\omega)) + \frac{\partial^2}{\partial \bar{\theta} \partial \bar{\theta}'}\mathrm{tr}\left\{f_{\theta_0^m}^{-1}(\omega)f_0(\omega)\right\}\right]$$

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

because $\widetilde{\bar{\theta}} \rightarrow^p \bar{\theta}_0^m$ and because of the continuity of the integrand. Also,

$$2\pi T^{-1/2}\frac{\partial \bar{L}_T(\bar{\theta}_0^m)}{\partial \bar{\theta}} - 2\pi T^{1/2}\frac{\partial \bar{L}_\infty^m(\bar{\theta}_0^m)}{\partial \bar{\theta}}$$

$$= -2\pi T^{-1/2}\sum_{j=1}^{T-1}W(\omega_j)\frac{\partial}{\partial \bar{\theta}}\mathrm{tr}\left\{f_{\bar{\theta}_0^m}^{-1}(\omega_j)\left(I_T(\omega_j) - f_0(\omega)\right)\right\}$$

$$+2T^{-1/2} \sum_{t=1}^{T} \frac{\partial \mu(\bar{\theta}_0^m)'}{\partial \bar{\theta}} f_{\theta_0^m}^{-1}(0) \, (Y_t - \mu_0) + o_p(1)$$

$$= \quad (\text{M1}) + (\text{M2}) + o_p(1).$$

Terms (M1) and (M2) satisfy a central limit theorem and can be analyzed in the same way as terms (I) and (II) in (1.27). The limiting covariance matrix can be verified accordingly. The detail is omitted. ■

Proof of Theorem 4: It suffices to verify that Assumptions 1 to 4 in Chernozhukov and Hong (2003) hold under our set of conditions. Relabel these assumptions as CH1 to CH4. CH1 and CH2 are trivial. CH3 is implied by Lemma 1.1(1), 1.1(2) and 1.1(4). To verify CH4, applying a second order Taylor expansion of $L_T(\theta)$ around θ_0 (see CH4(i)):

$$L_T(\theta) - L_T(\theta_0) = (\theta - \theta_0)' \frac{\partial L_T(\theta_0)}{\partial \theta} + \frac{1}{2}(\theta - \theta_0)' \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'}(\theta - \theta_0) + R_T(\theta)$$

with

$$R_T(\theta) = (\theta - \theta_0)' \left\{ \frac{\partial^2 L_T(\tilde{\theta}_T)}{\partial \theta \partial \theta'} - \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'} \right\} (\theta - \theta_0),$$

where $\tilde{\theta}_T$ lies between θ and θ_0. Now

$$T^{-1/2} \frac{\partial L_T(\theta_0)}{\partial \theta} \to^d N(0, V).$$

Therefore, CH4(ii) is satisfied (V corresponds to Ω_n in CH4). For CH4(iii), note that V is nonrandom and positive definite, and that

$$-T^{-1} \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'}$$

$$= \quad T^{-1} \sum_{j=1}^{T-1} W(\omega_j) \left(\frac{\partial vec(f_{\theta_0}(\omega_j)')}{\partial \theta'} \right)' \left\{ f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j) \right\} \left(\frac{\partial vec(f_{\theta_0}(\omega_j))}{\partial \theta'} \right)$$

$$= \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega_j) \left(\frac{\partial vec(f_{\theta_0}(\omega)')}{\partial \theta'} \right)' \left\{ f_{\theta_0}^{-1}(\omega)' \otimes f_{\theta_0}^{-1}(\omega) \right\} \left(\frac{\partial vec(f_{\theta_0}(\omega))}{\partial \theta'} \right) d\omega \quad (1.32)$$

$$+o(1), \tag{1.33}$$

where the leading term on the right hand side is nonrandom and positive definite because $f_{\theta_0}^{-1}(\omega)$, and

$$\int\limits_{-\pi}^{\pi} W(\omega_j) \left(\frac{\partial vec(f_{\theta_0}(\omega)')}{\partial \theta'} \right)' \left(\frac{\partial vec(f_{\theta_0}(\omega))}{\partial \theta'} \right) d\omega$$

are positive definite by Assumption 1.5 and local identification. It is $O(1)$ because the integrand is bounded; see Assumption 1.5. Therefore, CH4(iii) is satisfied. CH4(iv.a) holds because

$$|R_T(\theta)| \le \left\| T^{1/2}(\theta - \theta_0) \right\|^2 \left\| T^{-1} \frac{\partial^2 L_T(\tilde{\theta}_T)}{\partial \theta \partial \theta'} - T^{-1} \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'} \right\|,$$

where the second term can be made arbitrarily small by choosing $\|\theta - \theta_0\|$ small because of (1.32) and the boundedness and continuity of $\partial vec(f_{\theta}(\omega))/\partial \theta'$ and $f_{\theta}^{-1}(\omega)$ in θ (Assumptions 1.3 and 1.5(ii)). CH4(iv.b) holds because of the preceding argument and the fact that $\left\| T^{1/2}(\theta - \theta_0) \right\|^2 = O(1)$.

The proof for $\widehat{\tilde{\theta}}_T$ involves the same argument and is therefore omitted. ∎

1.8 Supplementary materials appendix 1

Table 1.1: Parameter values and the corresponding two smallest eigenvalues along the nonidentification curve

	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

Note. θ_j represent equally spaced points taken from the nonidentification curve extended from θ_0 for 14475 steps in direction 1, and for 101972 steps in direction 2. 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.

Table 1.2: Deviations of spectra across frequencies (direction 1)

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

Note. θ_1 to θ_{10} are as defined in Table 1.1. The omitted elements display identical deviations to the ones reported here.

Table 1.3: Deviations of spectra across frequencies (direction 2)

	Spectral density matrix element number
	(1,1)	(2,1)	(3,1)	(2,2)	(3,2)	(4,2)	(3,3)
Measure 1: Maximum absolute deviations across frequencies
θ_{-1}	8.49E-07	8.20E-08	5.00E-07	1.45E-07	9.87E-08	1.45E-07	2.52E-07
θ_{-2}	1.69E-06	1.59E-07	1.01E-06	2.75E-07	1.86E-07	2.75E-07	5.28E-07
θ_{-3}	2.52E-06	2.34E-07	1.53E-06	3.95E-07	2.64E-07	3.95E-07	8.18E-07
θ_{-4}	3.35E-06	3.07E-07	2.06E-06	5.04E-07	3.34E-07	5.04E-07	1.13E-06
θ_{-5}	4.17E-06	3.83E-07	2.60E-06	6.02E-07	3.96E-07	6.02E-07	1.46E-06
θ_{-6}	4.99E-06	4.64E-07	3.16E-06	6.91E-07	4.50E-07	6.91E-07	1.80E-06
θ_{-7}	5.80E-06	5.58E-07	3.72E-06	7.72E-07	4.98E-07	7.72E-07	2.17E-06
θ_{-8}	6.62E-06	6.76E-07	4.30E-06	8.44E-07	5.39E-07	8.44E-07	2.55E-06
θ_{-9}	7.43E-06	8.17E-07	4.89E-06	9.10E-07	5.75E-07	9.10E-07	2.95E-06
θ_{-10}	8.26E-06	9.74E-07	5.50E-06	9.67E-07	6.04E-07	9.67E-07	3.38E-06
Measure 2: Maximum absolute deviations across frequencies in relative form
θ_{-1}	3.79E-08	1.62E-08	3.56E-08	3.65E-09	4.78E-08	6.30E-08	2.75E-08
θ_{-2}	7.56E-08	3.07E-08	7.22E-08	7.67E-09	9.22E-08	1.23E-07	5.76E-08
θ_{-3}	1.13E-07	4.37E-08	1.09E-07	1.18E-08	1.34E-07	1.79E-07	8.93E-08
θ_{-4}	1.50E-07	5.55E-08	1.47E-07	1.62E-08	1.73E-07	2.33E-07	1.23E-07
θ_{-5}	1.86E-07	6.55E-08	1.86E-07	2.07E-08	2.09E-07	2.84E-07	1.59E-07
θ_{-6}	2.23E-07	7.42E-08	2.25E-07	2.54E-08	2.42E-07	3.32E-07	1.97E-07
θ_{-7}	2.59E-07	8.06E-08	2.65E-07	3.01E-08	2.72E-07	3.76E-07	2.37E-07
θ_{-8}	2.96E-07	8.50E-08	3.07E-07	3.47E-08	3.00E-07	4.17E-07	2.79E-07
θ_{-9}	3.32E-07	1.03E-07	3.49E-07	3.92E-08	3.25E-07	4.55E-07	3.22E-07
θ_{-10}	3.69E-07	1.22E-07	3.92E-07	4.39E-08	3.48E-07	4.90E-07	3.69E-07
Measure 3: Maximum relative deviations across frequencies
θ_{-1}	4.78E-08	1.32E-07	9.81E-08	3.22E-08	6.66E-08	8.37E-08	5.00E-08
θ_{-2}	9.58E-08	2.46E-07	1.89E-07	6.14E-08	1.27E-07	1.60E-07	9.41E-08
θ_{-3}	1.43E-07	3.59E-07	2.78E-07	8.84E-08	1.82E-07	2.31E-07	1.34E-07
θ_{-4}	1.89E-07	4.65E-07	3.64E-07	1.13E-07	2.32E-07	2.96E-07	1.69E-07
θ_{-5}	2.34E-07	5.67E-07	4.48E-07	1.36E-07	2.78E-07	3.57E-07	2.00E-07
θ_{-6}	2.80E-07	6.66E-07	5.31E-07	1.56E-07	3.19E-07	4.12E-07	2.27E-07
θ_{-7}	3.24E-07	7.62E-07	6.12E-07	1.75E-07	3.56E-07	4.63E-07	2.50E-07
θ_{-8}	3.69E-07	8.55E-07	6.92E-07	1.92E-07	3.89E-07	5.09E-07	2.79E-07
θ_{-9}	4.13E-07	9.47E-07	7.71E-07	2.07E-07	4.19E-07	5.51E-07	3.22E-07
θ_{-10}	4.57E-07	1.04E-06	8.51E-07	2.21E-07	4.44E-07	5.90E-07	3.69E-07

Note. θ_{-1} to θ_{-10} are as defined in Table 1.1. The omitted elements display identical deviations to the ones reported here.

Table 1.4: Rank sensitivity analysis

| | Differentiation step size $\times$ θ_0 | | | | | | | |
	1E-02	1E-03	1E-04	1E-05	1E-06	1E-07	1E-08	1E-09
TOL								
	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

Note. TOL refers to the tolerance level used to determine the rank. Default refers to the MATLAB default tolerance level.

a) ψ₁

b) ψ₂

c) ρ_r

d) σ_r²
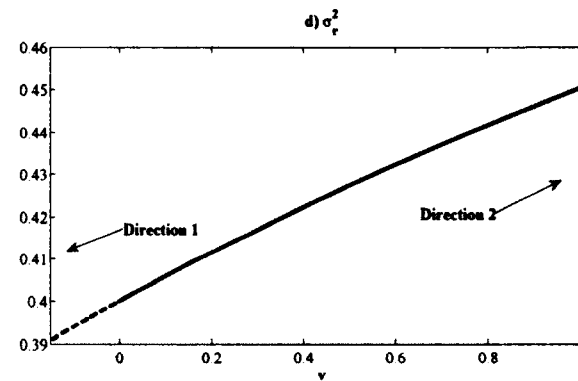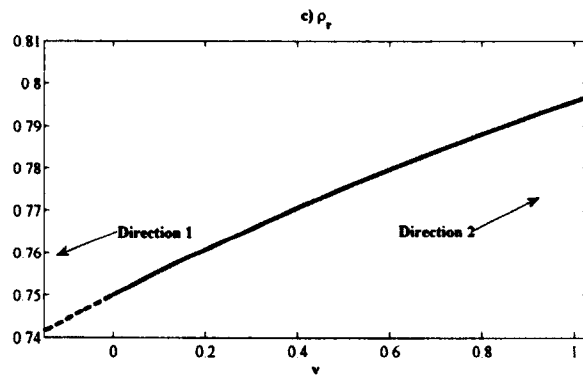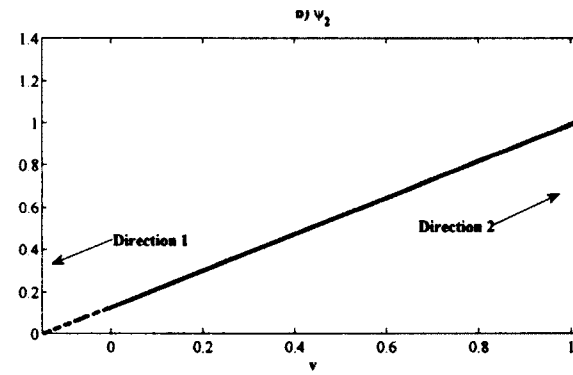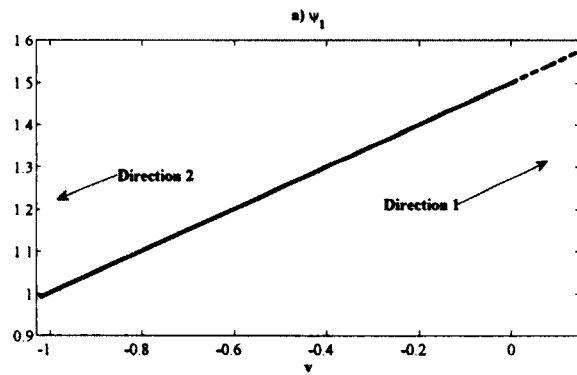
Note. The nonidentification curve is given by $\partial\theta(v)/\partial v = c(\theta)$, $\theta(0) = \theta_0$, where $c(\theta)$ is the eigenvector corresponding to the only zero eigenvalue of $G(\theta)$. The curve is computed recursively using the Euler method, so that $\theta(v_{j+1}) = \theta(v_j) + c(\theta(v_j))h$, where h is the step size, fixed at 1e-05. $(\psi_1, \psi_2, \rho_r, \sigma_r^2)$ 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.

Figure 1·1: The nonidentification curve $(\psi_1, \psi_2, \rho_r, \sigma_r^2)$

Chapter 2

Frequency Domain Analysis of Medium Scale DSGE Models with Application to Smets and Wouters (2007) (with Zhongjun Qu)

2.1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have become a widely applied instrument for analyzing business cycles, understanding monetary policy and for forecasting. Some medium scale DSGE models, such as that of Smets and Wouters (2007) (henceforth SW (2007)), are considered both within academia and by central banks. These models typically feature various frictions, often involving a relatively large number of equations and parameters with complex cross-equation restrictions. Although such sophistication holds promise for delivering rich and empirically relevant results, it also poses substantial challenges for identification, estimation and model diagnostics. This chapter shows how these issues can be tackled from a frequency domain perspective, using the framework developed in the previous chapter (published as Qu and Tkachenko (2012)). We use SW (2007) as the working example throughout the chapter, motivated by the fact that it has become a workhorse model in the DSGE literature. The analysis of other medium scale DSGE models can be conducted in a similar manner.

The identification of DSGE models is important for both model calibration and formal statistical analysis, although the relevant literature has lagged behind relative to estimation

until quite recently. Canova and Sala (2009) marks an important turning point by convincingly documenting the types of identification issues that can surface when analyzing a DSGE model. Iskrev (2010) gives sufficient conditions for the local identification of structural parameters based on the mean and a set of autocovariances. Qu and Tkachenko (2012) and Komunjer and Ng (2011) are the first to provide necessary and sufficient conditions for local identification. In the previous chapter, we have shown that taking a frequency domain perspective can deliver simple identification conditions applicable to both singular and nonsingular DSGE systems without relying on a particular (say, the minimum state) representation.

In this chapter, we show that the methods developed in Chapter 1 of this thesis can be applied in a straightforward manner to SW (2007) to deliver informative results. We structure our identification analysis into the following steps: (1) Identification based on the second order properties. This shows whether the parameters can be identified based solely on the dynamic properties of the system. (2) Identification based on the first (i.e., the mean) and the second order properties. This reveals whether the information from the steady state restrictions can help identification. (3) Identification based on a subset of frequencies. This is motivated by the fact that DSGE models are often designed to model business cycle movements, not very long or very short term fluctuations. Upon completing the above three steps, we find that the parameters in SW (2007) are unidentified without further restrictions. (4) To obtain further insights, we derive the nonidentification curves to depict parameter values that yield observational equivalence. The curves immediately reveal which and how many parameters need to be fixed to yield local identification. Note that the results from Steps (1) and (2) are in accordance with Iskrev (2010) and Komunjer and Ng (2011, the web appendix). Although these findings are not new, the analysis is,

and it also illustrates the simplicity of taking a frequency domain approach in this setting. Issues in Steps (3) and (4) have not been previously considered for medium scale DSGE models.

Next, we consider estimating SW (2007) from a frequency domain perspective using the methodology developed in the previous chapter. The method has two features. First, it allows for estimation and inference using a subset of frequencies, something that is outside the scope of conventional time domain methods. This is important because DSGE models are designed for medium term economic fluctuations, not very short or long term fluctuations. Second, it is straightforward to conduct Bayesian inference and the computation involved is similar to the time domain approach. Although we have analyzed the statistical properties of this method in the previous chapter, we did not provide an application. This chapter provides the first application of the method to a medium scale DSGE model.

Specifically, we follow SW (2007) in specifying the priors and An and Schorfheide (2007) in obtaining the posterior mode and Hessian for the proposal distribution. A Random Walk Metropolis (RWM) algorithm is used to obtain the posterior draws. We start with inference using the mean and the spectrum, then the full spectrum only and finally consider inference using only business cycle frequencies. The same priors are used throughout. For the first two cases, we obtain estimates that are very similar to those of SW (2007). This reflects the close linkage between the time and frequency domain likelihood. However, for the third case, we obtain noticeably different estimates of the parameters governing the exogenous disturbances. At the same time, the parameters governing contemporaneous interactions of the observables remain similar with only a few exceptions. The impulse response functions are noticeably different. To our knowledge, this is the first time such a finding is documented in the DSGE literature.

Then, we contrast the model implied spectrum and absolute coherency with that observed in the data. The analysis is motivated by Watson's (1993) suggestion of plotting the model and data spectra as one of the most informative diagnostics. It is also related to King and Watson (1996), who compared the spectra of the three quantitative rational expectations models with that of the data. Both the business cycle and the full spectrum based estimates do a reasonable job in matching these two key features. The business cycle based estimates achieve a better fit at the intended frequencies. However, they both underestimate the absolute coherency of the interest rate and other four variables (consumption growth, investment growth, output growth, and labor hours). The latter finding suggests a dimension along which the model can be further improved. To our knowledge, this is the first time such analysis is applied to medium scale DSGE models.

Our results suggest that the frequency domain perspective affords substantial depth and flexibility in identification analysis and in estimating the parameters of the model, while remaining simple in application and comparable in terms of computational burden relative to the conventional time domain methods. In practice, we suggest to carry out both the business cycle and the full spectrum based analysis jointly. This allows us to assess to what extent the results are driven by the very low frequency misspecifications, which is a hard task to tackle using a time domain framework.

The remainder of the chapter is structured as follows. Section 2.2 includes a brief description of the SW (2007) model to make the chapter self-contained. Section 2.3 carries out identification analysis and reports nonidentification curves. Section 2.4 presents estimation results. Section 2.5 conducts model diagnostics from a frequency domain perspective. Section 2.6 concludes. A more comprehensive summary of model equations is provided in Section 2.7. All figures and tables are located in Section 2.8.

2.2 The DSGE model of SW (2007)

SW (2007) has become a workhorse model in the DSGE literature and many medium scale

DSGE models consist of modifications or extensions of this model. It is an extended version

of the standard New Keynesian real business cycle model, featuring a number of frictions

and real rigidities. To make this chapter self-contained, we subsequently briefly describe

the structure of the model economy. Note that the discussion is meant to highlight the key

elements in the model, and a more detailed description of the model equations, variables,

and parameters is included in the mathematical appendix.

The model has seven observable endogenous variables with seven exogenous shocks.

In equilibrium, the model has a balanced growth path driven by deterministic labor-

augmenting technological progress. We focus on the log linearized system as in the original

article.

2.2.1 The aggregate resource constraint

The aggregate resource constraint is given by

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

Output (y_t) is composed of consumption (c_t), investment (i_t), capital utilization costs as

a function of the capital utilization rate (z_t), and exogenous spending (ε_t^g). The latter is

assumed to follow a first order autoregressive (AR) model with an i.i.d. Normal error term

(η_t^g) and is also affected by the productivity shock (η_t^a) as follows:

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

The coefficients c_y, i_y and z_y are functions of the structural parameters, as shown in the mathematical appendix.

2.2.2 Households

Households maximize a nonseparable utility function with two arguments (consumption and labor effort) over an infinite life horizon. Consumption appears in the utility function relative to a time-varying external habit variable. The dynamics of consumption follow from the consumption Euler equation

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

where l_t is hours worked, r_t is the nominal interest rate, and π_t is inflation. The disturbance term ε_t^b can be interpreted as a risk premium that households require to hold the one period bond. It follows the stochastic process

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

Households also choose investment given the capital adjustment cost they face. The dynamics of investment are given by

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

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

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

The corresponding arbitrage equation for the value of capital is given by

$$q_t = q_1 E_t q_{t+1} + (1 - q_1) E_t r_{t+1}^k - (r_t - \pi_{t+1}) - \frac{1}{c_3}\varepsilon_t^b \quad \text{with} \quad \varepsilon_t^b = \rho_b \varepsilon_{t-1}^b + \eta_t^b.$$

2.2.3 Final and intermediate goods market

The model has a perfectly competitive final goods market and a monopolistic competitive intermediate goods market. It features a symmetric equilibrium where all firms make identical decisions. At such an equilibrium, the aggregate production function is

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

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

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

The current capital service use (k_t^s) is a function of capital installed in the previous period (k_{t-1}) and the degree of capital utilization (z_t): $k_t^s = k_{t-1} + z_t$. Furthermore, the capital utilization is a positive fraction of the rental rate of capital (r_t^k): $z_t = z_1 r_t^k$. The accumulation of installed capital (k_t) is given by

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

where ε_t^i is the investment specific technology process as defined before.

The price mark-up, defined as the difference between the average price and the nominal marginal cost, satisfies

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

where w_t is the real wage. The firms set prices according to the Calvo model, leading to the following New Keynesian Phillips curve

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

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

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

The MA(1) term is intended to pick up some of the high frequency fluctuations in prices.

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

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

2.2.4 Labor market

Labor is differentiated by an intermediate labor union. The wage mark-up is

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

Real wage w_t adjusts slowly due to the rigidity

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

The wage mark-up disturbance is assumed to follow an ARMA(1,1) process:

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

2.2.5 Government policies

The empirical monetary policy reaction function is

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

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

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

The variable y_t^* stands for a time-varying optimal output level that is the result of a flexible price-wage economy. More generally, we use superscript star to denote variables in this economy. Such an economy needs to be solved along with the sticky price-wage economy for the purposes of identification and estimation.

2.2.6 The model solution

Our analysis requires computing the spectral density matrix of the observed endogenous variables. This is straightforward to obtain using the GENSYS algorithm of Sims (2002), although other methods (e.g., Uhlig (1999)) can also be used.

The GENSYS algorithm requires representing the state variables in the following form:

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

where S_t is a vector of model variables that includes the endogenous variables and the conditional expectation terms, Z_t are exogenously evolving and possibly serially correlated random disturbances, and ζ_t are expectation errors. For SW (2007) (note that the ordering of variables and parameters corresponds to our MATLAB code),

$$\begin{aligned}
S_t = \; & [\eta_t^w, \eta_t^p, z_t^*, r_t^{k*}, k_t^{s*}, q_t^*, c_t^*, i_t^*, y_t^*, l_t^*, w_t^*, r_t^*, k_t^*, \mu_t^w, z_t, r_t^k, k_t^s, q_t, c_t, i_t, y_t, l_t, \pi_t, w_t, r_t, \\
& \varepsilon_t^a, \varepsilon_t^b, \varepsilon_t^g, \varepsilon_t^i, \varepsilon_t^r, \varepsilon_t^p, \varepsilon_t^w, k_t, E(i_{t+1}^*), E(c_{t+1}^*), E(r_{t+1}^{k*}), E(q_{t+1}^*), E(l_{t+1}^*), E(i_{t+1}), \\
& E(c_{t+1}), E(r_{t+1}^k), E(q_{t+1}), E(l_{t+1}), E(\pi_{t+1}), E(w_{t+1})]',
\end{aligned}$$

where the elements 18 to 24 of S_t correspond to the observables used for identification analysis and estimation, which are (we use lower cases to stand for log deviations from the respective steady state) output (y_t), consumption (c_t), investment (i_t), wage (w_t), labor hours (l_t), inflation (π_t) and the interest rate (r_t). The other elements correspond to

model variables in both sticky and flexible price-wage economies, seven shock processes, and twelve expectation terms. See the mathematical appendix for more information on the elements of S_t above. The vector of structural shocks is given by

$$Z_t = (\eta_t^a, \eta_t^b, \eta_t^g, \eta_t^i, \eta_t^r, \eta_t^p, \eta_t^w)',$$

where, as discussed above, η_t^a is a technology shock, η_t^b is a risk premium shock, η_t^g is an exogenous spending shock, η_t^i is an investment shock, η_t^r is a monetary policy shock, η_t^p and η_t^w price and wage mark-up shocks respectively. The elements of ζ_t are all zero except the last twelve entries that correspond to the one period ahead expectation errors of the last twelve terms of S_t. This implies that Π is of dimension 45×12, is an identity matrix for the last twelve rows, and zero otherwise. The coefficients matrices Γ_0, Γ_1, and Ψ are functions of the structural dynamic parameters θ, consisting of

$$\theta = (\rho_{ga}, \mu_w, \mu_p, \alpha, \psi, \varphi, \sigma_c, \lambda, \phi_p, \iota_w, \xi_w, \iota_p, \xi_p, \sigma_l, r_\pi, r_{\Delta y}, r_y, \rho, \rho_a, \rho_b, \rho_g, \rho_i, \rho_r, \rho_p, \rho_w,$$
$$\sigma_a, \sigma_b, \sigma_g, \sigma_i, \sigma_r, \sigma_p, \sigma_w, \gamma, \beta, \delta, g_y, \phi_w, \epsilon_p, \epsilon_w).$$

Under conditions that ensure the existence and uniqueness of the solution (see p.12 in Sims (2002)), the system can be represented as

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

where Θ_1 and Θ_0 are functions of θ [1], which further implies

$$S_t = (I - \Theta_1 L)^{-1} \Theta_0 Z_t. \tag{2.1}$$

From the above vector moving average representation, we can easily obtain the repre-

---

[1]Therefore, a complete notation should be $\Theta_0(\theta)$ and $\Theta_1(\theta)$. We omit such a dependence for simplicity.

sentation for the observable endogenous variables. To see this, suppose that the observable Y_t, up to an unknown mean vector, is given by

$$(c_t - c_{t-1}, i_t - i_{t-1}, y_t - y_{t-1}, l_t, \pi_t, w_t - w_{t-1}, r_t). \tag{2.2}$$

To map this to the solution (2.1), we simply let $A(L)$ be a matrix of finite order lag polynomials that specifies the observables, then we compute

$$A(L)S_t = A(L)(I - \Theta_1 L)^{-1}\Theta_0 Z_t \tag{2.3}$$

with

$$
\underset{7 \times 45}{A(L)} =
\begin{bmatrix}
\overset{(1,1)}{0} & \cdots & \overset{(1,18)}{1-L} & \overset{(1,19)}{0} & \overset{(1,20)}{0} & \overset{(1,21)}{0} & \overset{(1,22)}{0} & \overset{(1,23)}{0} & \overset{(1,24)}{0} & \cdots \\
\overset{(1,45)}{0} & & & & & & & & & \\
\vdots & \cdots & 0 & 1-L & 0 & 0 & 0 & 0 & 0 & \cdots \\
\vdots & & & & & & & & & \\
\vdots & \cdots & 0 & 0 & 1-L & 0 & 0 & 0 & 0 & \cdots \\
\vdots & & & & & & & & & \\
\vdots & \cdots & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots \\
\vdots & & & & & & & & & \\
\vdots & \cdots & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \cdots \\
\vdots & & & & & & & & & \\
\vdots & \cdots & 0 & 0 & 0 & 0 & 0 & 1-L & 0 & \cdots \\
\vdots & & & & & & & & & \\
0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots \\
0 & & & & & & & & &
\end{bmatrix}.
$$

The vector moving average representation (2.3) plays a central role in our analysis. First, it enables straightforward computation of the spectrum of Y_t:

$$f_\theta(\omega) = \frac{1}{2\pi} H(\exp(-i\omega); \theta) \Sigma(\theta) H(\exp(-i\omega); \theta)^*, \qquad (2.4)$$

where the asterisk denotes the conjugate transpose,

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

and $\Sigma(\theta)$ is the variance covariance matrix of Z_t[2]. Second, we can easily compute the impulse response functions and the variance decomposition. Third, the choice of $A(L)$ offers substantial flexibility as we can vary it to study estimation and inference based on different combinations of variables.

For identification and inference based on the spectrum, there is no need to specify the steady state. However, it is also straightforward to incorporate the mean into the analysis. To see this, define an augmented parameter vector $\bar{\theta}$ that includes θ and parameters affecting only the steady state. Then, notice that for log linearized DSGE models the observables Y_t can typically be 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 specification in SW (2007) corresponds to $Y_t^d(\theta)$ given by (2.1) and $\mu(\bar{\theta}) = (\bar{\gamma}, \bar{\gamma}, \bar{\gamma}, \bar{l}, \bar{\pi}, \bar{\gamma}, \bar{r})'$. The parameters $\bar{\gamma}, \bar{\pi}$ and $\bar{r}$ are functions of structural parameters and $\bar{l}$ is a new steady state parameter. The detailed discussion is presented in subsection 2.3.2 below.

---

[2]Note that in our code $\Sigma(\theta)$ is a 7×7 identity matrix, as we incorporate the shock standard deviations into Ψ when we set up the dynamic system.

2.3 Identification analysis

In this section we perform identification analysis based on the (first and) second order properties of the model, consider identification from a subset of frequencies (business cycle frequencies) and implement a robustness check for the results. The corresponding theoretical results have been derived in the previous chapter: see Theorems 1.1-1.2 and Corollaries 1.2-1.6. We conduct our identification analysis by setting θ_0 to the posterior mean from the Table 1A in SW (2007):

$$\theta_0 = (0.52, 0.88, 0.74, 0.19, 0.54, 5.48, 1.39, 0.71, 1.61, 0.59, 0.73, 0.22, 0.65, 1.92, 2.03, 0.22,$$
$$0.08, 0.81, 0.95, 0.18, 0.97, 0.71, 0.12, 0.90, 0.97, 0.45, 0.24, 0.52, 0.45, 0.24, 0.14, 0.24,$$
$$1.0043, 0.9984, 0.025, 0.18, 1.5, 10, 10).$$

We choose the above parameter values for illustration purposes and because, given the analysis, they are empirically reasonable values. In practice, the same analysis can be carried out for other parameter values using the same methodology.

2.3.1 Analysis of SW (2007) 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 $-4,999\pi/5,000$ to $4,999\pi/5,000$ and multiplying by 2π. We keep the step size for the numerical differentiation at $10^{-7} \times \theta_0$, and use the MATLAB default tolerance set at $tol = \max(size(G)eps(\|G\|))$ to decide whether an eigenvalue is zero, where eps is the floating point precision of G. We obtain $\text{rank}(G(\theta_0)) = 36$. Since the dimension of θ_0 is 39, this implies that θ is unidentified at θ_0. Additionally, this result suggests that a minimum of three parameters needs to be fixed to achieve identification.

Since the model is not identified, we can proceed to search for the nonidentified subsets

of parameters. We find no such one-element subsets of θ in Step 2. In the next step, we find two subvectors that yield $G(\theta_0)^s$ with one zero eigenvalue: (ξ_w, ϵ_w) and (ξ_p, ϵ_p). This finding is not surprising, as the parameters in each subset play very similar roles in the model after linearization (they determine the speed of adjustment of prices and wages through the Calvo probability, or the curvature of demand, respectively) and thus are not separately identifiable. SW (2007) recognized that and fixed ϵ_p and ϵ_w in estimation. Iskrev (2010) obtains the same result by applying his condition. We do not report the nonidentification curves for these subsets, as they are trivial and are highlighted here for illustration purposes.

We then exclude all three-parameter subvectors that contain either of the two nonidentification sets identified above as proper subsets and continue the analysis. We find no three- or four-element nonidentification subsets. However, we pinpoint one five-element subvector which has one zero eigenvalue:

$$(\varphi, \lambda, \gamma, \beta, \delta)$$

where φ is the adjustment cost parameter, λ (denoted as h in SW (2007)) is the habit parameter, γ governs the steady state growth rate, β is the discount rate, and δ is the depreciation rate. This result is also in accordance with Iskrev (2010). After excluding all subvectors containing the nonidentification sets highlighted above, we find no further sources of nonidentification in this model. Therefore, our findings imply that fixing one parameter out of each of $(\varphi, \lambda, \gamma, \beta, \delta)$, (ξ_w, ϵ_w) and (ξ_p, ϵ_p) is necessary and sufficient for identification from the second order properties.

We then evaluate the nonidentification curve using the Euler method with step size $h = 10^{-4}$ in a small neighborhood around θ_0. The result is presented in Figure 2.1, which demonstrates how, for each of $\varphi, \lambda, \gamma, \beta$ and δ, the parameters have to change simultaneously

in order to generate nonidentification. The curve is extended using (1.11) in the two directions starting from θ_0, which are marked on the graph by the bold and dotted lines respectively. It should be noted that β is increasing along direction 2. Since it represents the discount rate, it cannot be larger than 1. Therefore, we truncate the direction 2 of the curve at a point where β is closest to 1. This leaves us with only 472 steps in direction 2, which, compared to 670,000 steps computed for direction 1, is very small and hence in the Figure 2.1 values corresponding to direction 2 look like a bold dot rather than a line. Given the number of the steps computed, we did not reach the point where natural bounds on parameters are violated along direction 1, but it is clear that we would truncate it at a point where β reaches zero, λ reaches zero, or δ reaches 1, whichever happens first.

To give an illustration of parameter changes involved, we report ten points taken from the curve at equally spaced intervals in each direction in Table 2.1. In addition, we report the smallest and the second smallest eigenvalues of $G(\theta_0)^3$ to show that its rank stays constant along the curve.

To verify that the points on the curve indeed result in identical spectral densities, we compute three different measures of the discrepancies between $f_\theta(\omega)$ and $f_{\theta_0}(\omega)$ considered in the previous chapter:

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

Maximum absolute deviation in relative form $\quad : \quad \dfrac{\max_{\omega_j \in \Omega} | f_{\theta hl}(\omega_j) - f_{\theta_0 hl}(\omega_j) |}{| f_{\theta_0 hl}(\omega_j) |}$

Maximum relative deviation: $\quad \max_{\omega_j \in \Omega} \dfrac{| f_{\theta hl}(\omega_j) - f_{\theta_0 hl}(\omega_j) |}{| f_{\theta_0 hl}(\omega_j) |},$

where $f_{\theta hl}(\omega)$ denotes the (h, l)-th element of the spectral density matrix with parameter θ and Ω is the set that includes the 5,000 frequencies between 0 and π^3. In order to conserve

---

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

space, we report 8 largest deviations that occur across all 49 elements of $G(\theta)^s$ computed over 5,000 frequencies in descending order for points in Table 2.1. The results can be found in Tables 2.2 and 2.3. They show that even the largest deviations are very small. Given that there are numerical errors involved in the application of the solution algorithm and the computation of the $G(\theta)^s$ matrix, and that the Euler method involves a cumulative approximation error of order 10^{-4} in our case, we can conclude that the spectrum stays the same along the curve.

2.3.2 Analysis of SW (2007) based on the first and second order properties

This subsection extends the analysis to incorporate the steady state of the model. The measurement equations from SW (2007), relating the observables to the means and the log deviations, are as follows:

$$
\begin{aligned}
dlCONS_t &= \bar{\gamma} + c_t - c_{t-1} \\[4pt]
dlINV_t &= \bar{\gamma} + i_t - i_{t-1} \\[4pt]
dlGDP_t &= \bar{\gamma} + y_t - y_{t-1} \\[4pt]
lHOURS_t &= \bar{l} + l_t \\[4pt]
dlP_t &= \bar{\pi} + \pi_t \\[4pt]
dlWAG_t &= \bar{\gamma} + w_t - w_{t-1} \\[4pt]
FEDFUNDS_t &= \bar{r} + r_t,
\end{aligned}
$$

where l and dl stand for 100 times log and log difference, respectively; $\bar{\gamma} = 100(\gamma - 1)$, $\bar{\pi} = 100(\Pi_* - 1)$, and $\bar{r} = 100(\beta^{-1}\gamma^{\sigma_c}\Pi_* - 1) = \beta^{-1}\gamma^{\sigma_c}\bar{\pi} + 100(\beta^{-1}\gamma^{\sigma_c} - 1)$. Among the means, $\bar{\gamma}$ is a function of the dynamic parameter γ, $\bar{\pi}$ and $\bar{r}$ depend on the common steady parameter inflation rate Π_* and $\bar{l}$ is a new parameter. Hence, we can augment θ by two

parameters and the full parameter vector becomes

$$\bar{\theta} = (\theta, \bar{\pi}, \bar{l}).$$

We have 41 parameters in total and $\mu(\bar{\theta})$ is given by

$$\mu(\bar{\theta}) = (\bar{\gamma}, \bar{\gamma}, \bar{\gamma}, \bar{l}, \bar{\pi}, \bar{\gamma}, \bar{r})'.$$

We set $\bar{\pi}_0 = 0.78$ and $\bar{l}_0 = 0.53$ as in Table 1A in SW (2007). $\mu(\bar{\theta})$ can be differentiated analytically in this case, e.g., using MATLAB's symbolic math toolbox.

Applying Theorem 1.2 yields $\text{rank}(G(\bar{\theta}_0)) = 39$. Since now $q = 41$, the result tells us that we cannot identify the parameter vector at $\bar{\theta}_0$ from the first and the second order properties of the observables, and, furthermore, that two parameters need to be fixed to achieve identification. The sources of nonidentification in this case are the two subsets we have detected in the previous subsection, namely (ξ_w, ϵ_w) and (ξ_p, ϵ_p). This result is, again, not surprising and should be expected given the similar role the parameters play in the model, as discussed in the previous subsection. We no longer detect the $(\varphi, \lambda, \gamma, \beta, \delta)$ subset as γ determines the steady state growth rate $\bar{\gamma}$ and hence can be identified from the mean. Once γ is identified, the rest of the four parameters are uniquely determined. Iskrev (2010) reaches the same conclusion. Thus, fixing one parameter from each of (ξ_w, ϵ_w) and (ξ_p, ϵ_p) is necessary and sufficient for identification based on the mean and the spectrum.

2.3.3 Analysis of SW (2007) using a subset of frequencies.

In this subsection we examine identification from a subset of frequencies. Specifically, we focus on business cycle frequencies. We use the conventional definition, i.e., treat all frequencies corresponding to periods between 6 and 32 quarters as business cycle frequencies

(see, e.g., King and Watson (1996)). We compute both $G^W(\theta_0)$ and $\overline{G}^W(\overline{\theta}_0)$ to examine identification from the second, and first and second order properties of the observables. We obtain $\text{rank}(G^W(\theta_0)) = 36$ and $\text{rank}(\overline{G}^W(\overline{\theta}_0)) = 39$, which coincides with the results obtained using all frequencies. All results and conclusions are the same as in the previous two subsections. This shows that for this model business cycle frequencies have the same local identification power at θ_0 and $\overline{\theta}_0$ as the full spectrum.

2.3.4 Robustness checks using nonidentification curves

The results above have been obtained using a particular step size for numerical differentiation and the MATLAB default tolerance level for computing the ranks of the $G(\theta)$ and $G(\overline{\theta})$ matrices. Here, we check the sensitivity of $G(\theta_0)$ to a range of numerical differentiation steps (from 10^{-2} to 10^{-9}) and tolerance levels (from 10^{-3} to 10^{-10}). The results can be found in Table 2.4. Although we report the rank sensitivity analysis results only for $G(\theta)$, similar checks have been performed for all of the matrices computed above to ensure robustness of the reported rank.

It can be seen from the chapter that varying the differentiation step can affect the rank decision. Specifically, the estimated rank changes if the step size is too large or too small, and when the tolerance level is more stringent. This is quite intuitive, as when the step size is too large, the numerical differentiation will induce 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 will be large relative to the step size, therefore the rank will also be estimated imprecisely. In this example, the step size 1e-07$\times\theta_0$ and the MATLAB default tolerance level seem to produce good balance between precision and robustness.

The dependence of the results on the step size and the tolerance level is certainly undesirable. To address this issue, we suggested previously that the nonidentification curve analysis be embedded into the following two-step procedure to reduce the reliance on step size and tolerance level:

- Step 1. Compute the ranks of $G(\theta_0)$ using a wide range of step sizes and tolerance levels. 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 to confirm observational equivalence. If the discrepancies are large, proceed to the outcome with the next smallest rank and repeat the analysis. Continue until spectral densities on the curve are identical or full local identification is established.

In applications, it often suffices to compute as few as 10 points on the nonidentification curve to establish whether spectral densities are identical or not, as in the latter case the deviations become quite large only a few steps away from θ_0, so the computational burden involved is not large. Applying this procedure using the step sizes and tolerance levels in Table 2.4 leads to the same conclusion as stated above. This is because 36 is the smallest rank in the Table 2.4 (Step 1) and the discrepancies between $f_\theta(\omega)$ and $f_{\theta_0}(\omega)$ along the curves are negligible (Step 2). In summary, this example demonstrates another reason why nonidentification curves can be a useful tool for identification analysis.

2.4 Estimation and inference

We also consider estimating the model of SW (2007) from a frequency domain perspective. We start with briefly summarizing the quasi-Bayesian estimation procedure proposed in Section 1.5.

2.4.1 The basic framework

First, we consider the quasi-likelihood functions implied by the linearized DSGE model. Under the assumption that the DSGE system is nonsingular (i.e., $n_Y \leq n_\epsilon$), which is satisfied by the SW (2007) model, the approximate generalized Whittle log likelihood function of θ based on the sample $Y_1, ..., Y_T$ is given by

$$L_T(\theta) = -\sum_{j=1}^{T-1} W(\omega_j) \left[\log \det(f_\theta(\omega_j)) + \text{tr}\left\{ f_\theta^{-1}(\omega_j) I_T(\omega_j) \right\} \right],$$

where $\omega_j = 2\pi j/T$ $(j = 1, 2, ...T - 1)$ denote the Fourier frequencies, $W(\omega_j)$ is the indicator function as defined in the identification section, and $I_T(\omega_j)$ is the sample periodogram. Define the discrete Fourier transform of the data by

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

then the periodogram can be computed as $I_T(\omega_j) = w_T(\omega_j) w_T^*(\omega_j)$. Note that maximizing $L_T(\theta)$ allows us to estimate dynamic parameters based on the spectrum of $\{Y_t\}$ without any reference to the parameters that only enter the steady state. Also, unlike for time domain QML, the estimates can be obtained without demeaning the data, since the values of $w_T(\omega_j)$ at the Fourier frequencies are not affected by replacing Y_t by $Y_t - \mu(\bar{\theta})$ in the definition of $w_T(\omega_j)$ above.

The extension to estimation of both dynamic and steady state parameters jointly is straightforward. Let

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

Since $w_{\bar{\theta},T}(0)$ has a multivariate normal distribution with asymptotic variance $f_\theta(0)$ and is asymptotically independent of $w_T(\omega_j)$ for $j = 1, 2, ..., T - 1$, it can be shown that the

approximate log likelihood function of $\bar{\theta}$ takes the form:

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

Since the direct application of maximum likelihood methods to estimation of DSGE models is plagued by the problem where the obtained estimates are often at odds with economic theory, possibly due to the models' stylized nature and potential misspecification, it has become common practice to use Bayesian methods that introduce information not contained in the observed sample via reweighting the likelihood by the relevant prior density (see An and Schorfheide (2007) for discussion). This motivates us to incorporate prior distributions on the DSGE parameters into our estimation framework following Chernozhukov and Hong (2003).

Specifically, for the dynamic parameter only case, we consider

$$p_T(\theta) = \frac{\pi(\theta)\exp\left(L_T\left(\theta\right)\right)}{\int_\Theta \pi(\theta)\exp\left(L_T\left(\theta\right)\right)d\theta},$$

where $\pi(\theta)$ can be a proper prior density or, more generally, any weight function that is strictly positive and continuous over the parameter space. The function $p_T(\theta)$ is termed quasi-posterior in Chernozhukov and Hong (2003), because, while being a proper distribution density over the parameters, it is in general not a true posterior in the Bayesian sense, as $\exp\left(L_T\left(\theta\right)\right)$ is a more general criterion function than the likelihood. The quasi-posterior mean, given by

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

can be taken as the estimate for θ_0. Computation of an estimate involves drawing a Markov chain $S = \left(\theta^{(1)}, \theta^{(2)}, \ldots \theta^{(B)}\right)$ from the quasi-posterior density using a Markov chain Monte

Carlo (MCMC) algorithm and computing the mean of the draws

$$\hat{\theta}_T = \frac{1}{B} \sum_{j=1}^{B} \theta^{(j)}. \qquad (2.5)$$

Estimates of a given continuously differentiable function g: $\Theta \to \mathbb{R}$, e.g., an impulse response at a given horizon, can be obtained by computing

$$g(\hat{\theta}_T) = \frac{1}{B} \sum_{j=1}^{B} g(\theta^{(j)}). \qquad (2.6)$$

In this chapter, we use the popular Random Walk Metropolis algorithm to generate draws from $p_T(\theta)$. It belongs to the more general class of Metropolis-Hastings algorithms, the first version of which was proposed by Metropolis et al. (1953) and later generalized by Hastings (1970). Schorfheide (2000) and Otrok (2001) were the seminal contributions in using this algorithm for Bayesian estimation of DSGE models. We use the version of the algorithm implemented in Schorfheide (2000). The steps involved and some discussion on their practical implementation are presented below.

- Step 1. Use numerical optimization to maximize $L_T(\theta) + \log(\pi(\theta))$. The maximizer is the posterior mode, denoted $\tilde{\theta}$.

- Step 2. Obtain the inverse of the Hessian computed at the posterior mode, denote it $\tilde{\Sigma}$.

- Step 3. Draw a starting value $\theta^{(0)}$ from $N(\tilde{\theta}, c^2\tilde{\Sigma})$, where c is a scaling parameter, or specify a starting value directly.

- Step 4. For $s = 1, 2..., B$, draw ϑ from the proposal distribution $N(\theta^{(s-1)}, c^2\tilde{\Sigma})$. Accept the draw ($\theta^{(s)} = \vartheta$) with probability $\min\{1, \alpha(\theta^{(s-1)}, \vartheta \mid Y)\}$ and reject it

$(\theta^{(s)} = \theta^{(s-1)})$ otherwise. The acceptance probability is

$$\alpha(\theta^{(s-1)}, \vartheta \mid Y) = \frac{\exp(L_T(\vartheta))\pi(\vartheta)}{\exp\left(L_T(\theta^{(s-1)})\right)\pi(\theta^{(s-1)})}.$$

- Step 5. Compute the posterior mean estimates as in (2.5) and (2.6).

In Step 1, one of the practical problems that may be encountered by an optimization algorithm is the possible lack of existence or uniqueness of the solution for the DSGE model. To circumvent these issues, we use the *csminwel* optimization routine written by Chris Sims (see Leeper and Sims (1994)), which randomly perturbs the search direction if it reaches a cliff caused by indeterminacy or nonexistence. Regarding the prior, we use the same $\pi(\theta)$ as in the Table 1A in SW (2007).

In Step 2, the Hessian matrix, computed assuming Normality, has its (j, l)-th element given by

$$[\tilde{\Sigma}^{-1}]_{jl} = \frac{1}{4\pi} \int\limits_{-\pi}^{\pi} W(\omega) tr \left[f_{\tilde{\theta}}(\omega) \frac{\partial f_{\tilde{\theta}}^{-1}(\omega)}{\partial \theta_j} f_{\tilde{\theta}}(\omega) \frac{\partial f_{\tilde{\theta}}^{-1}(\omega)}{\partial \theta_l} \right] d\omega,$$

which can be estimated by replacing the integral with an average over the Fourier frequencies.

In Step 4, the choice of the scaling parameter c is determined by calibrating the acceptance rate of the Markov chain. Roberts et al. (1997) suggested a heuristic rule to use proposal distributions with an acceptance rate close to 25% for models of dimension higher than two under the assumption that both the target and the proposal distribution are Normal. Since this assumption is not satisfied in our case, we follow the literature by drawing several Markov chains with acceptance rates between 25% and 40%. Therefore, while keeping the seed of the random number generator fixed, we try a range of values for c until we find one that yields the desired acceptance rate. In our experience, for a

given c and a random number generator seed, computing the acceptance rate of a chain of 1,000-5,000 draws gives a good idea about what to expect from a much longer chain. Also, we may draw ϑ that yield indeterminacy or nonexistence of the DSGE solution, or contain parameter values that violate the specified bounds (our bounds are as in the Dynare code of SW (2007)). In such cases, we set $L_T(\theta) + \log(\pi(\theta))$ to a very large negative number $(-1e10)$ so that such draws are always rejected.

We first perform estimation of $\bar{\theta}$ based jointly on the mean and the full spectrum of observables, as this closely mirrors the analysis of SW (2007) conducted from a time domain perspective. In order to enhance comparability of results, five parameters are kept fixed in estimation, as in SW (2007), at the following values

$$\epsilon_p = \epsilon_w = 10, \delta = 0.025, g_y = 0.18, \lambda_w = 1.5.$$

2.4.2 Estimation based on the mean and the full spectrum

The data we use is that used in SW (2007) and we consider the same sample period as in their Dynare code, namely (Q1 1965 - Q4 2004). The prior distribution is kept the same as in SW (2007) and is presented in Table 2.5. For each Markov chain, a sample of 250,000 draws from the posterior distribution is generated, and the first 50,000 draws are discarded as burn-in. We report results for $c = \sqrt{0.15}$ as the scaling constant, which resulted in the acceptance rate of 24%[4]. It should also be noted that the theoretical spectral density at frequency zero is singular, because the observables contain first differences of stationary variables. Computationally, we deal with this problem by using the generalized inverse to calculate $f_\theta^{-1}(0)$ and the product of nonzero eigenvalues of $f_\theta(0)$ to obtain $\det(f_\theta(0))$. For ease of comparison, we report the results for the former case alongside those obtained in

---

[4]Here and below we used several scaling factors yielding the acceptance rates between 25% and 40%, and found that the results are not sensitive to these changes.

SW (2007) in Table 2.6.

Overall, the parameter estimates in Table 2.6 are very similar to their counterparts in SW (2007). The posterior means and modes are close. The 90% probability intervals overlap for 38 out of the 41 parameters. The two exceptions are that our estimate of the technology shock persistence (ρ_a) is higher (0.98 compared to 0.95 in SW (2007)), while the estimated persistence parameter of the exogenous spending shock (ρ_g) is lower (0.92 versus 0.97). For these two parameters the corresponding 90% probability intervals are disjoint. We can also single out a somewhat higher estimate of the elasticity of consumption σ_c (1.81 compared to 1.38), although there is still slight overlap in the 90% intervals, and a lower estimate of the trend growth rate ($\bar{\gamma}$) of 0.27 versus 0.43 in SW (2007).

2.4.3 Estimation based on the full spectrum

We now perform estimation of θ based on the full spectrum of observables. We consider the same data set, prior, and MCMC algorithm, except we choose $c = 0.4$, which produced an acceptance rate of 23%. The results are reported in Table 2.7.

Overall, the parameter estimates in Table 2.7 are very similar to those based on the mean and the full spectrum. The estimated trend growth rate is back in line with the results of SW (2007), but the estimated mean discount rate goes up from 0.76% to 1.04% on an annual basis. The rest of the estimates obtained using the full spectrum are virtually the same as those in Table 2.6. We can also see that overall the estimation results using the full spectrum are, as would be expected, very close to those obtained by SW (2007) using time domain methods.

2.4.4 Estimation using business cycle frequencies

DSGE models are constructed to explain business cycle movements. Schorfheide (2011) emphasized that "many time series exhibit low frequency behavior that is difficult, if not impossible, to reconcile with the model being estimated. This low frequency misspecification contaminates the estimation of shocks and thereby inference about the sources of business cycle". Therefore, it is instructive to consider in what way if any the estimates change when the estimation is carried out using business cycles frequencies only. Our procedure allows for such an investigation. We use the same methodology as in the previous subsection to perform estimation of dynamic parameters, selecting only the frequencies corresponding to cycles of 6 to 32 quarters and changing the variance tuning parameter to $c = \sqrt{0.13}$, which results in an acceptance rate of 23%. The results are reported in the right panel of Table 2.7.

We find that a number of parameter estimates differ substantially from those obtained using the full spectrum. The most notable differences pertain to the parameters governing the exogenous shocks. Specifically, the AR coefficient of the total factor productivity process, ρ_a, drops from 0.98 to 0.84 while the standard deviation of its shock remains unchanged. The parameter governing the impact of productivity shocks on exogenous spending, ρ_{ga}, is almost halved from 0.47 to 0.24. Additionally, the AR coefficient of the wage mark-up process ρ_w comes down from 0.96 to 0.56 and its MA coefficient μ_w drops from 0.92 to 0.32. The standard deviation of its shock decreases but the two posterior intervals overlap. On the other hand, the AR coefficients for risk premium (ρ_b) and monetary policy (ρ_r) shocks rise from 0.21 to 0.75, and from 0.09 to 0.34 respectively. The standard deviations of the respective shocks decrease from 0.24 and 0.24 to 0.08 and 0.13, respectively. The parameter differences outlined above are significant in the sense that their 90%

probability intervals do not overlap. For the remaining three shock processes, exogenous spending, monetary policy and price mark-up, the magnitudes of the AR and MA coefficients either remain the same or show a small decrease, while the standard deviations of these shocks become smaller. Other notable differences in estimated parameters include the adjustment cost elasticity (φ), which goes down to 3.03 from 5.76, the degree of price indexation (ι_p), which increases from 0.21 to 0.61, and the coefficient on the lagged interest rate (ρ), which goes down from 0.85 to 0.76. These results imply that the model estimated using business cycle frequencies will potentially deliver different impulse responses from those obtained using the full spectrum. We explore this issue in the next section.

2.5 Impulse response analysis

Motivated by the differences found between parameter estimates obtained using the full spectrum and business cycle frequencies, we estimate the impulse response functions of the seven observables to the shocks for the two cases. Figures 2.2 through 2.8 report the posterior means, along with the 90% posterior intervals for horizons of up to 20 quarters. Each figure corresponds to a single observable. One notable difference between the responses of nearly all of the variables to a risk premium shock is that the impulse responses obtained using business cycle frequencies display a hump shaped dynamic, as opposed to an almost monotonic decay of those obtained using the full spectrum, as well as those in SW (2007). One exception is wage, where the impulse response with the full spectrum is itself somewhat hump shaped, but still the hump shaped pattern of the business cycle impulse response is much more pronounced. In all other cases it appears that the effects of both exogenous spending and investment shocks are in general significantly less pronounced when business cycle frequencies are used for estimation, perhaps with the exception of an investment

shock to inflation and an exogenous spending shock to consumption and wage, for which the differences are not as clear cut. The effect of a wage mark-up shock dies out faster for all variables if estimated using business cycle frequencies. Its effects are also significantly less pronounced after about 5 quarters for consumption and wage, after 10 quarters for output and labor hours, and for the whole 20 quarters for inflation and interest rate. It is interesting to note that the business cycle impulse response of investment to this shock is more pronounced initially for about five quarters, but then goes to zero faster after about 14 quarters. The monetary policy shock also has smaller impact and goes to zero faster. Little difference can be observed when considering the responses to the price mark-up shock, as the two posterior intervals mostly overlap for the whole 20 quarters. However, responses become less pronounced and decay faster for consumption after roughly 10 quarters, and for output and labor hours after 15 quarters. The responses to the productivity shock are also very similar to the full spectrum case, except for the cases of output, consumption and wage, for which the response is lower and decaying faster in the case of business cycle frequencies.

It is important to ask whether the difference is due to the impact of the prior, which has a greater effect in the business cycle frequency case as some information from the data is discarded. We address this as follows. First, we compute the value of the log likelihood constructed using the business cycle frequencies, but at the parameter values estimated using the full spectrum. Then, we compare this value with the same likelihood function computed using the estimates from business cycles. The results are reported in Table 2.8. If the difference were in fact driven by the prior, then the latter would be smaller or of similar magnitude to the former. The result suggests otherwise. Similarly, we evaluate the log likelihood function constructed using the full spectrum at the business cycle estimates

and compare with that at full spectrum estimates. The difference is even more pronounced. Overall, the result suggests that estimates obtained from business cycle frequencies do a good job at matching these frequencies, but are at odds with other frequencies, in this case the very low frequencies.

Since the above analysis omits frequencies from both sides of the business cycle frequency band, it leaves unclear which components are driving the difference. To investigate this, we consider estimation omitting only frequencies below the business cycle band. Figures 2.9 to 2.15 contain the impulse responses for this case. The estimates from the business cycle case are also included to ease the comparison. The figures show that results are overall similar to those using business cycle frequencies. Therefore, most of the differences observed between the impulse responses computed using the full spectrum estimates and those using business cycle frequencies can be attributed to the omission of the frequencies below the business cycle band. There are a few deviations from this pattern. The hump shaped responses of all seven variables to the risk premium shock observed in business cycle results are no longer present. The same can be noted about the initial few quarters of responses of inflation to the productivity and the price mark-up shocks, as well as of wage to the price mark-up shock.

2.6 Model diagnostics from a frequency domain perspective

King and Watson (1996) compared the spectra of three quantitative rational expectations models with that of the data. The models were calibrated and of small scale. Below, we carry out similar analysis for the medium scale DSGE model considered here. The goal of the analysis is two-fold. First, we examine whether the model captures the variability of and the comovements between relevant macroeconomic variables. Second, we compare

the model spectrum estimated using all frequencies with that using only business cycle frequencies. The latter will highlight the potential value from using a subset of frequencies in estimation.

We obtain nonparametric estimate of the spectral density by smoothing the periodograms using demeaned data. Suppose Y_t contains only one variable. Then, the estimator is given by

$$\widehat{f}(\omega_j) = \sum_{|k| \leq m} \mathcal{W}_T(k) I_T(\omega_{j+k}) \quad \text{for } j \geq 1$$

and

$$\widehat{f}(0) = \mathcal{W}_T(0) I_T(\omega_1) + 2 \sum_{k=1}^{m} \mathcal{W}_T(k) I_T(\omega_{j+k}),$$

where m is a positive integer, $\mathcal{W}_T(k)$ is a weight function satisfying $\mathcal{W}_T(k) = \mathcal{W}_T(-k)$, $\mathcal{W}_T(k) \geq 0 \ \forall \ k$, $\sum_{|k| \leq m} \mathcal{W}_T(k) = 1$ and $I_T(\omega_j)$ is the periodogram of the data. The estimator is consistent under mild conditions (see Brockwell and Davis (1991) for a rigorous treatment) and the asymptotic 95% confidence intervals for the estimates of the log of spectral density are given by

$$\log(\widehat{f}(\omega_j)) \pm 1.96 \left(\sum_{|k| \leq m} \mathcal{W}_T(k)^2 \right)^{1/2}.$$

We apply the same type of estimator to obtain absolute coherency between pairs of variables. Let Y_t be a bivariate demeaned time series. The spectral density matrix is estimated in the same way as above but with $I_T(\omega_{j+k})$ being a 2×2 matrix. Let $\widehat{f}_{hk}(\omega_j)$ denote the (h, k)-th element of $\widehat{f}(\omega)$, then the absolute coherency estimate ($|\widehat{\Re}_{12}(\omega_j)|$) between Y_{1t} and Y_{2t} is

$$|\widehat{\Re}_{12}(\omega_j)| = [\widehat{c}_{12}^2(\omega_j) + \widehat{q}_{12}^2(\omega_j)]^{1/2} / [\widehat{f}_{11}(\omega_j) \widehat{f}_{22}(\omega_j)]^{1/2},$$

where

$$\widehat{c}_{12}(\omega_j) \;=\; [\widehat{f}_{12}(\omega_j) + \widehat{f}_{21}(\omega_j)]/2,$$

$$\widehat{q}_{12}(\omega_j) \;=\; i[\widehat{f}_{12}(\omega_j) - \widehat{f}_{21}(\omega_j)]/2.$$

The approximate 95% confidence bounds can be computed as follows

$$|\widehat{\Re}_{12}(\omega_j)| \;\pm\; 1.96(1 - |\widehat{\Re}_{12}(\omega_j)|^2)\left(\sum_{|k|\le m} \mathcal{W}_T(k)^2\right)^{1/2}/\sqrt{2}.$$

In applications, the choice of $\mathcal{W}_T(k)$ depends on the characteristics of the data series at hand. It is possible and sometimes advantageous to use different weighting functions for estimation of different elements of the spectral density matrix due to potentially different features of the time series (see Ch. 9 in Priestley (1981) for a discussion). In our case, we apply the same weight function in all estimations, with $m = 4$ and the weights given by $\left\{\frac{1}{21}, \frac{2}{21}, \frac{3}{21}, \frac{3}{21}, \frac{3}{21}, \frac{3}{21}, \frac{3}{21}, \frac{2}{21}, \frac{1}{21}\right\}$, which is obtained by the successive application of two Daniell filters with weights given by $\left\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right\}$ and $\left\{\frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}\right\}$. This choice of $\mathcal{W}_T(k)$ produces spectra estimates that are not as rough as the raw periodogram, and in the meantime do not appear oversmoothed.

Figure 2.16 plots the log spectra of the seven variables. Three results are reported in each sub-figure. First, we report the nonparametric estimates of the spectrum of the demeaned data series along with the pointwise 95% confidence intervals. They are used as a benchmark to assess the model's ability in capturing these key features. The solid curve is the spectrum implied by the model with parameters estimated using the full spectrum. The dashed line is the model spectrum with business cycle based estimates. Two patterns emerge. First, the solid curve captures the overall shape of the data spectrum, although there are noticeable departures which often occur inside of the business cycle frequencies. It

should be noted that for the growth series (sub-figures i-iii,vi), the model implies that their spectral density at frequency zero is zero (as the figure reports log spectra, the frequency zero is omitted from the figures). This is inconsistent with the data spectra, which are positive at the origin. When frequencies very near zero are included in the estimation, the model will try to reduce such a departure by having very persistent estimates. This potentially affects the other frequencies, which partly explains why the full spectrum based estimates do not capture the slope of the spectrum very well inside of the business cycle frequencies. When using only business cycle frequencies for estimation, such a tension is absent and the estimates do a better job at matching variations at these frequencies. The lines never fall substantially outside of the confidence bands. However, the departures from the data spectrum can be substantial outside of the business cycle frequencies. In practice, this offers the researcher a choice. If one firmly believes that the DSGE model is well specified at all frequencies, then, they should all enter the estimation and the estimates will be more efficient. If one suspects that the modeling of the trend, or, more generally, of the very low frequency behavior in the model is inconsistent with the data (for example, the data has a broken trend while the model has a linear trend), then the subset based approach may be a more robust choice.

Figures 2.17 to 2.19 report the absolute coherency between the seven variables. Notice that their values can be interpreted as a measure of strength of correlation at a particular frequency. Both the business cycle and the full spectrum based estimates achieve something at capturing their overall magnitudes, with the exception of the comovements between interest rate and other four variables (consumption growth, investment growth, output growth, and labor hours). In the latter case, the two estimates are close and are consistently below the nonparametric estimates. This unanimous finding suggests a dimension along

which the model can be further improved. For the other cases, the business cycle based estimates typically do a better job at the intended frequencies. They largely stay within the confidence intervals, and are better at capturing the peaks of the coherency, while the full spectrum based estimates miss them in the majority of cases.

In summary, the DSGE model does a reasonable job at matching the spectra of individual time series and the absolute coherency implied by the data. The subset based estimates offer the flexibility to focus on a particular frequency band and to achieve a better fit at such frequencies. In practice, both analyses can be carried out, allowing us to assess to what extent the results are driven by the very low frequencies.

2.7 Conclusion and discussion

This chapter has considered identification, estimation and inference in medium scale DSGE models using SW (2007) as an illustrative example. A key element in the analysis is that we can focus on part of the spectrum. For identification, we derived the nonidentification curve to reveal which and how many parameters need to be fixed to achieve local identification. For estimation and inference, we compared estimates obtained using the full spectrum with those using only business cycle frequencies and reported notably different parameter values and impulse response functions. Further analysis shows that the differences are mainly due to the frequencies below the business cycle frequency band. We have also considered model diagnostics by contrasting the model based and the nonparametrically estimated spectra as well as examining the absolute coherency. The result suggests that SW (2007) does a reasonable job at matching these two features observed in the data, with the exception of the comovements between interest rate and other four variables (consumption growth, investment growth, output growth, and labor hours). The subset based estimates, due to

their ability to focus on a particular frequency band, achieve a better fit at such frequencies.

From a methodological perspective, the results contribute to the relatively sparse literature that exploits the advantage of model estimation and diagnostics using a subset of frequencies. Engle (1974) is a seminal contribution. It proposed band spectrum regression as a way to allow for frequency specific misspecification, seasonality, measurement errors, and to obtain better understanding of some common time domain procedures such as applying a moving average filter. Sims (1993) and Hansen and Sargent (1993) considered the effect of removing or downweighting seasonal frequencies in estimating rational expectations models. Diebold, Ohanian and Berkowitz (1998) discussed a general framework for loss function based estimation and model evaluation. In a different context, McCloskey (2010) considered parameter estimation in ARMA, GARCH and stochastic volatility models robust to low frequency contamination caused by a changing mean or misspecified trend. Qu and Tkachenko (2012) provided a comprehensive treatment of the theoretical and computational aspects of the frequency domain quasi-likelihood applied to DSGE models. By working through a concrete example, this chapter demonstrates that such an approach is applicable to medium scale DSGE models and that it offers substantial depth and flexibility when compared with time domain methods. We intend to apply the methodology to a relatively broad class of DSGE models and hope to report results in the near future.

2.8   Mathematical appendix 2

The numbering of the equations below corresponds to SW (2007). Subscript star denotes steady state values. Note that some parameters are expressed as functions of the structural parameters. We highlight such relationships when relevant.

2.8.1   The sticky price-wage economy

1. The resource constraint is

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

Output (y_t) divides into consumption (c_t), investment (i_t), capital utilization costs as a function of the capital utilization rate (z_t), and exogenous spending (ε_t^g). The coefficients c_y, i_y and z_y are functions of the following structural parameters.

$$g_y, \gamma, \delta, \beta, \sigma_c, \phi_p, \alpha.$$

Their relationship to the coefficients above is:

$$
\begin{aligned}
i_y &= (\gamma - 1 + \delta)k_y, \\
c_y &= 1 - g_y - i_y, \\
z_y &= R_*^k k_y,
\end{aligned}
$$

where k_y is the steady state capital-output ratio, and R_*^k is the steady state rental rate of capital (see the Appendix to SW (2007)):

$$k_y = \phi_p \left(L_*/k_*\right)^{\alpha-1} = \phi_p \left[((1-\alpha)/\alpha)\left(R_*^k/w_*\right)\right]^{\alpha-1},$$

with

$$w_* = \left(\frac{\alpha^\alpha(1-\alpha)^{(1-\alpha)}}{\phi_p\left(R_*^k\right)^\alpha}\right)^{1/(1-\alpha)} \text{ and } R_*^k = \beta^{-1}\gamma^{\sigma_c} - (1-\delta).$$

2. The dynamics of consumption follow from the consumption Euler equation

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

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

Here the basic parameters are

$$\lambda, \gamma, \sigma_c, \phi_w, \alpha, \delta$$

and some parameters in the resource constraint (1). Their relationship to the coefficients above is:

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

where $w_*^h L_*/c_*$ are related to the steady state and are given by

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

where R_*^k and k_y are defined as above, and

$$c_y = 1 - g_y - (\gamma - 1 + \delta)k_y.$$

It seems that

$$\phi_w = \lambda_w \text{ instead of } 1 + \lambda_w.$$

In the code, $c_3 \varepsilon_t^b$ is redefined as ε_t^b, that is

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

Therefore the equation 4 (below) is also redefined accordingly.

3. The investment Euler equation is given by

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

The basic parameters are

$$\beta, \gamma, \sigma_c, \varphi.$$

i_1 and i_2 are related to them as

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

4. The value of capital is given by

$$q_t = q_1 E_t q_{t+1} + (1 - q_1) E_t r_{t+1}^k - \left(r_t - E_t \pi_{t+1} + \varepsilon_t^b\right).$$

The basic parameters are

$$\beta, \gamma, \sigma_c, \delta,$$

and parameters determining R_*^k. Their relationship to the coefficient above is:

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

Note that the code is programmed using

$$q_1 = \frac{1 - \delta}{R_*^k + 1 - \delta}.$$

Because of the redefinition of ε_t^b, this equation appears as

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

where

$$c_3 = \frac{1 - \lambda/\gamma}{(1 + \lambda/\gamma)\sigma_c}.$$

5. The aggregate production function is

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

The basic parameters are

$$\phi_p, \alpha.$$

6. Current capital service use is a function of capital installed in the previous period and the degree of capital utilization

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

7. Degree of capital utilization is a positive fraction of the rental rate of capital

$$z_t = z_1 r_t^k,$$

where

$$z_1 = \frac{1 - \psi}{\psi}.$$

The basic parameter is ψ.

8. Households rent capital services to firms and decide how much capital to accumulate given the capital adjustment cost they face. The accumulation of installed capital is given by

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

where

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

9. Price mark-up or the real marginal cost is

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

10. The New Keynesian Phillips curve is given by

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

where

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

Besides the basic parameters defined above, we have in addition

$$\iota_p, \xi_p, \varepsilon_p.$$

11. The rental rate of capital is

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

Note that in the original paper k_t instead of k_t^s shows up. It is likely a typo, as in their Dynare code SW (2007) have k_t^s.

12. Labor is differentiated by a union. The wage mark-up is

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

A new basic parameter is

$$\sigma_l.$$

13. Real wage adjusts slowly due to the rigidity

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

where

$$w_1 = \frac{1}{1 + \beta\gamma^{(1-\sigma_c)}}, w_2 = \frac{1 + \beta\gamma^{(1-\sigma_c)}\iota_w}{1 + \beta\gamma^{(1-\sigma_c)}}, w_3 = \frac{\iota_w}{1 + \beta\gamma^{(1-\sigma_c)}},$$

$$w_4 = \frac{1}{1 + \beta\gamma^{(1-\sigma_c)}}\frac{\left(1 - \beta\gamma^{(1-\sigma_c)}\xi_w\right)\left(1 - \xi_w\right)}{\xi_w\left((\phi_w - 1)\varepsilon_w + 1\right)}.$$

New basic parameters are

$$\iota_w, \xi_w, \phi_w, \varepsilon_w.$$

14. The empirical monetary policy reaction function is

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

The new basic parameters are

$$\rho, r_\pi, r_Y, r_{\Delta y}.$$

The shocks are (all AR and MA coefficients are basic parameters):

15.

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

16.

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

17.

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

18.

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

19.

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

20.

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

21.

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

2.8.2 The flexible price-wage economy

For the flexible price-wage economy, the equations are essentially the same as above, but with the variables μ_t^p and μ_t^w set to zero. The shock processes are also the same, thus we do not repeat them here.

1. The resource constraint:

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

2. The dynamics of consumption follow from the consumption Euler equation

$$c_t^* = c_1 c_{t-1}^* + (1 - c_1) E_t c_{t+1}^* + c_2 (l_t^* - E_t l_{t+1}^*) - c_3 (r_t^* - 0) - \varepsilon_t^b.$$

Note that the expected inflation is zero because the price adjusts instantaneously.

3. The dynamics of investment come from the investment Euler equation

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

4. The corresponding arbitrage equation for the value of capital is given by

$$q_t^* = q_1 E_t q_{t+1}^* + (1 - q_1) E_t r_{t+1}^{*k} - (r_t^* - 0) - \frac{1}{c_3} \varepsilon_t^b.$$

The expected inflation is zero for the same reason as above.

5. The aggregate production function is

$$y_t^* = \phi_p \left(\alpha k_t^{*s} + (1 - \alpha) l_t^* + \varepsilon_t^a \right).$$

6. Current capital service use is a function of capital installed in the previous period and the degree of capital utilization

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

7. The degree of capital utilization is a positive fraction of the rental rate of capital

$$z_t^* = z_1 r_t^{*k}.$$

8. The accumulation of installed capital is

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

9. Because $\mu_t^p = 0$ and the relationship with rigidity is: $\mu_t^p = \alpha \left(k_t^s - l_t \right) + \varepsilon_t^a - w_t$, we have

$$0 = \alpha \left(k_t^s - l_t \right) + \varepsilon_t^a - w_t$$

or, equivalently,

$$\alpha r_t^{*k} + (1 - \alpha) w_t^* = \varepsilon_t^a.$$

There is no New Keynesian Phillips curve as price adjusts instantaneously.

10. The rental rate of capital is

$$r_t^{*k} = - (k_t^{*s} - l_t^*) + w_t^*.$$

11. The wage mark-up is now $\mu_t^w = 0$. Therefore,

$$0 = w_t^* - \left(\sigma_l l_t^* + \frac{1}{1 - \lambda} \left(c_t^* - \lambda c_{t-1}^* \right) \right),$$

or

$$w_t^* = \left(\sigma_l l_t^* + \frac{1}{1 - \lambda} \left(c_t^* - \lambda c_{t-1}^* \right) \right).$$

2.9 Supplementary materials appendix 2

Table 2.1: Parameter values and the corresponding two smallest eigenvalues along the nonidentification curve

	φ	λ	γ	β	δ	λ_1	λ_2
θ_0	5.740000	0.710000	1.004300	0.998400	0.025000	1.80E-10	0.392865
			Panel (a).	Direction 1			
θ_1	12.417476	0.482721	0.682812	0.862248	0.337109	1.96E-14	0.808082
θ_2	19.113813	0.389080	0.550356	0.794406	0.465700	4.57E-14	1.210705
θ_3	25.812574	0.334809	0.473589	0.750327	0.540228	3.01E-14	1.599268
θ_4	32.512006	0.298325	0.42198	0.718141	0.590328	5.53E-15	1.975594
θ_5	39.211698	0.271647	0.384246	0.693026	0.626964	3.26E-15	2.341212
θ_6	45.911511	0.25105	0.35511	0.672563	0.655256	8.58E-15	2.697239
θ_7	52.611389	0.234516	0.331724	0.655380	0.677954	1.04E-14	3.044732
θ_8	59.311305	0.220873	0.312427	0.640622	0.696688	4.10E-15	3.384357
θ_9	66.011244	0.209364	0.296147	0.627727	0.712493	5.40E-15	3.716722
θ_{10}	72.711198	0.199485	0.282174	0.616303	0.726059	9.96E-16	4.042423
			Panel (b).	Direction 2			
θ_{-1}	5.735346	0.710288	1.004707	0.998556	0.024605	5.27E-12	0.392485
θ_{-2}	5.730692	0.710576	1.005115	0.998711	0.024209	3.00E-12	0.392186
θ_{-3}	5.726038	0.710865	1.005523	0.998865	0.023812	2.95E-11	0.391895
θ_{-4}	5.721384	0.711154	1.005933	0.999019	0.023415	3.93E-11	0.391616
θ_{-5}	5.716730	0.711444	1.006342	0.999173	0.023018	9.91E-11	0.391323
θ_{-6}	5.712077	0.711732	1.006752	0.999328	0.022620	1.12E-10	0.391078
θ_{-7}	5.707423	0.712023	1.007162	0.999483	0.022221	8.78E-11	0.390749
θ_{-8}	5.702770	0.712314	1.007573	0.999638	0.021823	8.39E-11	0.390467
θ_{-9}	5.698117	0.712605	1.007984	0.999793	0.021423	1.97E-10	0.390278
θ_{-10}	5.693464	0.712896	1.008396	0.999948	0.021024	1.13E-10	0.389814

Note. θ_j represent equally spaced points taken from the nonidentification curve extended from θ_0 for 670,000 steps in direction 1, and for 472 steps in direction 2. λ_1 and λ_2 represent the smallest and the second smallest eigenvalues of $G(\theta_i)^s$. The step size for computing the curve is 10^{-4}. Along direction 1, the curve is truncated at the point where β is closest to 1, as it is the discount factor. Results are rounded to the nearest sixth digit to the right of decimal.

Table 2.2: Deviations of spectra across frequencies (direction 1)

	8 largest deviations across frequencies and elements in descending order							
	1	2	3	4	5	6	7	8
	Maximum absolute deviations across frequencies							
θ_1	8.99E-05	2.98E-05	1.24E-05	1.24E-05	1.09E-05	1.09E-05	9.24E-06	9.24E-06
θ_2	1.17E-04	3.88E-05	1.61E-05	1.61E-05	1.42E-05	1.42E-05	1.20E-05	1.20E-05
θ_3	1.31E-04	4.31E-05	1.79E-05	1.79E-05	1.59E-05	1.59E-05	1.33E-05	1.33E-05
θ_4	1.38E-04	4.57E-05	1.90E-05	1.90E-05	1.68E-05	1.68E-05	1.41E-05	1.41E-05
θ_5	1.43E-04	4.74E-05	1.97E-05	1.97E-05	1.74E-05	1.74E-05	1.46E-05	1.46E-05
θ_6	1.47E-04	4.85E-05	2.02E-05	2.02E-05	1.78E-05	1.78E-05	1.50E-05	1.50E-05
θ_7	1.49E-04	4.94E-05	2.05E-05	2.05E-05	1.81E-05	1.81E-05	1.53E-05	1.53E-05
θ_8	1.51E-04	5.01E-05	2.08E-05	2.08E-05	1.83E-05	1.83E-05	1.55E-05	1.55E-05
θ_9	1.52E-04	5.06E-05	2.10E-05	2.10E-05	1.84E-05	1.84E-05	1.56E-05	1.56E-05
θ_{10}	1.53E-04	5.10E-05	2.12E-05	2.12E-05	1.86E-05	1.86E-05	1.58E-05	1.58E-05
	Maximum absolute deviations across frequencies in relative form							
θ_1	7.81E-06	5.33E-06	4.60E-06	4.34E-06	4.19E-06	3.73E-06	3.73E-06	3.34E-06
θ_2	1.02E-05	6.93E-06	5.98E-06	5.66E-06	5.47E-06	4.86E-06	4.86E-06	4.39E-06
θ_3	1.13E-05	7.73E-06	6.70E-06	6.32E-06	6.13E-06	5.41E-06	5.41E-06	4.89E-06
θ_4	1.20E-05	8.18E-06	7.07E-06	6.69E-06	6.49E-06	5.73E-06	5.73E-06	5.18E-06
θ_5	1.24E-05	8.48E-06	7.31E-06	6.93E-06	6.71E-06	5.93E-06	5.93E-06	5.37E-06
θ_6	1.27E-05	8.68E-06	7.50E-06	7.09E-06	6.86E-06	6.07E-06	6.07E-06	5.50E-06
θ_7	1.29E-05	8.82E-06	7.63E-06	7.21E-06	6.96E-06	6.18E-06	6.18E-06	5.59E-06
θ_8	1.31E-05	8.91E-06	7.71E-06	7.28E-06	7.07E-06	6.25E-06	6.25E-06	5.61E-06
θ_9	1.33E-05	8.98E-06	7.79E-06	7.34E-06	7.12E-06	6.31E-06	6.31E-06	5.66E-06
θ_{10}	1.34E-05	9.05E-06	7.85E-06	7.40E-06	7.18E-06	6.36E-06	6.36E-06	5.71E-06
	Maximum relative deviations across frequencies							
θ_1	5.94E-05	5.94E-05	2.67E-05	2.67E-05	1.52E-05	1.52E-05	1.37E-05	1.37E-05
θ_2	7.75E-05	7.75E-05	3.49E-05	3.49E-05	1.99E-05	1.99E-05	1.79E-05	1.79E-05
θ_3	8.65E-05	8.65E-05	3.91E-05	3.91E-05	2.23E-05	2.23E-05	2.00E-05	2.00E-05
θ_4	9.16E-05	9.16E-05	4.14E-05	4.14E-05	2.36E-05	2.36E-05	2.12E-05	2.12E-05
θ_5	9.48E-05	9.48E-05	4.28E-05	4.28E-05	2.44E-05	2.44E-05	2.20E-05	2.20E-05
θ_6	9.71E-05	9.71E-05	4.38E-05	4.38E-05	2.49E-05	2.49E-05	2.24E-05	2.24E-05
θ_7	9.88E-05	9.88E-05	4.45E-05	4.45E-05	2.53E-05	2.53E-05	2.28E-05	2.28E-05
θ_8	9.99E-05	9.99E-05	4.49E-05	4.49E-05	2.55E-05	2.55E-05	2.30E-05	2.30E-05
θ_9	1.01E-04	1.01E-04	4.54E-05	4.54E-05	2.56E-05	2.56E-05	2.32E-05	2.32E-05
θ_{10}	1.02E-04	1.02E-04	4.58E-05	4.58E-05	2.58E-05	2.58E-05	2.34E-05	2.34E-05

Note. θ_1 to θ_{10} are as defined in Table 2.1. We report 8 largest deviations across 49 elements of each $G(\theta_i)^s$ computed at 5,000 frequencies to conserve space.

Table 2.3: Deviations of spectra across frequencies (direction 2)

	8 largest deviations across frequencies and elements in descending order							
	1	2	3	4	5	6	7	8
	Maximum absolute deviations across frequencies							
θ_{-1}	1.59E-07	3.14E-08	3.14E-08	3.09E-08	1.77E-08	1.77E-08	1.65E-08	1.65E-08
θ_{-2}	2.38E-07	6.50E-08	4.93E-08	4.93E-08	2.33E-08	2.33E-08	2.27E-08	2.27E-08
θ_{-3}	3.54E-07	1.14E-07	5.52E-08	5.52E-08	3.80E-08	3.80E-08	2.59E-08	2.59E-08
θ_{-4}	5.88E-07	1.68E-07	8.09E-08	8.09E-08	7.20E-08	7.20E-08	5.26E-08	5.26E-08
θ_{-5}	8.55E-07	2.42E-07	1.12E-07	1.12E-07	1.07E-07	1.07E-07	8.22E-08	8.22E-08
θ_{-6}	1.08E-06	3.11E-07	1.34E-07	1.34E-07	1.24E-07	1.24E-07	8.90E-08	8.90E-08
θ_{-7}	1.32E-06	3.76E-07	1.82E-07	1.82E-07	1.54E-07	1.54E-07	1.39E-07	1.39E-07
θ_{-8}	1.40E-06	4.11E-07	1.83E-07	1.83E-07	1.62E-07	1.62E-07	1.30E-07	1.30E-07
θ_{-9}	1.44E-06	4.42E-07	1.80E-07	1.80E-07	1.62E-07	1.62E-07	1.18E-07	1.18E-07
θ_{-10}	1.47E-06	4.57E-07	1.80E-07	1.80E-07	1.71E-07	1.71E-07	1.17E-07	1.17E-07
	Maximum absolute deviations across frequencies in relative form							
θ_{-1}	2.24E-08	1.54E-08	1.36E-08	1.16E-08	1.07E-08	9.44E-09	7.97E-09	7.78E-09
θ_{-2}	4.52E-08	3.14E-08	2.63E-08	1.82E-08	1.65E-08	1.56E-08	1.51E-08	1.51E-08
θ_{-3}	4.03E-08	2.96E-08	2.82E-08	2.75E-08	1.81E-08	1.48E-08	1.45E-08	1.45E-08
θ_{-4}	4.37E-08	4.35E-08	3.55E-08	3.55E-08	3.00E-08	2.52E-08	2.30E-08	2.30E-08
θ_{-5}	1.07E-07	6.35E-08	4.97E-08	4.44E-08	3.98E-08	3.56E-08	3.56E-08	3.09E-08
θ_{-6}	1.50E-07	8.22E-08	5.96E-08	5.90E-08	5.21E-08	4.62E-08	4.62E-08	4.11E-08
θ_{-7}	1.69E-07	9.95E-08	7.35E-08	7.21E-08	5.73E-08	5.73E-08	5.08E-08	4.82E-08
θ_{-8}	1.81E-07	1.08E-07	7.71E-08	7.59E-08	6.21E-08	6.02E-08	6.02E-08	5.22E-08
θ_{-9}	1.87E-07	1.17E-07	7.91E-08	7.73E-08	7.12E-08	6.13E-08	6.13E-08	5.62E-08
θ_{-10}	1.91E-07	1.20E-07	8.17E-08	7.94E-08	7.69E-08	6.10E-08	6.10E-08	5.76E-08
	Maximum relative deviations across frequencies							
θ_{-1}	8.38E-08	8.38E-08	6.39E-08	6.39E-08	5.12E-08	5.12E-08	3.22E-08	3.22E-08
θ_{-2}	2.51E-07	2.51E-07	1.38E-07	1.38E-07	1.23E-07	1.23E-07	5.72E-08	5.72E-08
θ_{-3}	3.32E-07	3.32E-07	1.68E-07	1.68E-07	1.12E-07	1.12E-07	7.00E-08	7.00E-08
θ_{-4}	3.76E-07	3.76E-07	1.89E-07	1.89E-07	1.39E-07	1.39E-07	1.02E-07	1.02E-07
θ_{-5}	4.58E-07	4.58E-07	2.23E-07	2.23E-07	1.64E-07	1.64E-07	1.42E-07	1.42E-07
θ_{-6}	6.72E-07	6.72E-07	3.34E-07	3.34E-07	2.28E-07	2.28E-07	1.93E-07	1.93E-07
θ_{-7}	6.52E-07	6.52E-07	3.07E-07	3.07E-07	2.63E-07	2.63E-07	2.14E-07	2.14E-07
θ_{-8}	8.18E-07	8.18E-07	3.95E-07	3.95E-07	2.78E-07	2.78E-07	2.38E-07	2.38E-07
θ_{-9}	9.84E-07	9.84E-07	4.79E-07	4.79E-07	2.88E-07	2.88E-07	2.55E-07	2.55E-07
θ_{-10}	1.06E-06	1.06E-06	5.19E-07	5.19E-07	2.97E-07	2.97E-07	2.62E-07	2.62E-07

Note. θ_{-1} to θ_{-10} are as defined in Table 2.1. We report 8 largest deviations across 49 elements of each $G(\theta_i)^s$ computed at 5,000 frequencies to conserve space.

Table 2.4: Rank sensitivity analysis

	Differentiation step size $\times$ θ_0							
	1E-02	1E-03	1E-04	1E-05	1E-06	1E-07	1E-08	1E-09
TOL								
				Rank of $G(\theta_0)$				
1E-03	37	36	36	36	36	36	36	36
1E-04	37	37	37	36	36	36	36	36
1E-05	37	37	37	36	36	36	36	36
1E-06	37	37	37	36	36	36	36	36
1E-07	38	37	37	37	36	36	36	37
1E-08	39	37	37	37	36	36	37	37
1E-09	39	38	38	37	37	36	37	37
1E-10	39	39	39	37	37	37	37	39
Default	39	38	37	37	37	36	37	37

Note. TOL refers to the tolerance level used to determine the rank. Default refers to the MATLAB default tolerance level.

Table 2.5: Prior distribution of the parameters

	Distribution	Mean	St. Dev.
ρ_{ga}	Normal	0.50	0.25
μ_w	Beta	0.50	0.20
μ_p	Beta	0.50	0.20
α	Normal	0.30	0.05
ψ	Beta	0.50	0.15
φ	Normal	4.00	1.50
σ_c	Normal	1.50	0.38
λ	Beta	0.70	0.10
ϕ_p	Normal	1.25	0.13
ι_w	Beta	0.50	0.15
ξ_w	Beta	0.50	0.10
ι_p	Beta	0.50	0.15
ξ_p	Beta	0.50	0.10
σ_l	Normal	2.00	0.75
r_π	Normal	1.50	0.25
$r_{\Delta y}$	Normal	0.13	0.05
r_y	Normal	0.13	0.05
ρ	Beta	0.75	0.10
ρ_a	Beta	0.50	0.20
ρ_b	Beta	0.50	0.20
ρ_g	Beta	0.50	0.20
ρ_i	Beta	0.50	0.20
ρ_r	Beta	0.50	0.20
ρ_p	Beta	0.50	0.20
ρ_w	Beta	0.50	0.20
σ_a	Invgamma	0.10	2.00
σ_b	Invgamma	0.10	2.00
σ_g	Invgamma	0.10	2.00
σ_i	Invgamma	0.10	2.00
σ_r	Invgamma	0.10	2.00
σ_p	Invgamma	0.10	2.00
σ_w	Invgamma	0.10	2.00
$\bar{\gamma}$	Normal	0.40	0.10
$100(\beta^{-1} - 1)$	Gamma	0.25	0.10
$\bar{\pi}$	Gamma	0.62	0.10
$\bar{l}$	Normal	0.00	2.00

Note. Prior distributions are taken from SW(2007) Dynare code.

Table 2.6: Posterior distribution of the parameters

	Full Spectrum and mean				SW(2007) Tables 1 A,B			
	Mode	Mean	5%	95%	Mode	Mean	5%	95%
ρ_{ga}	0.48	0.48	0.38	0.58	0.52	0.52	0.37	0.66
μ_w	0.94	0.92	0.88	0.96	0.88	0.84	0.75	0.93
μ_p	0.68	0.66	0.51	0.78	0.74	0.69	0.54	0.85
α	0.20	0.20	0.18	0.22	0.19	0.19	0.16	0.21
ψ	0.72	0.70	0.56	0.83	0.54	0.54	0.36	0.72
φ	5.47	5.72	4.26	7.41	5.48	5.74	3.97	7.42
σ_c	1.83	1.81	1.56	2.08	1.39	1.38	1.16	1.59
λ	0.64	0.65	0.59	0.71	0.71	0.71	0.64	0.78
ϕ_p	1.60	1.61	1.50	1.71	1.61	1.60	1.48	1.73
ι_w	0.55	0.54	0.37	0.72	0.59	0.58	0.38	0.78
ξ_w	0.84	0.82	0.76	0.87	0.73	0.70	0.60	0.81
ι_p	0.19	0.21	0.10	0.33	0.22	0.24	0.10	0.38
ξ_p	0.66	0.66	0.60	0.72	0.65	0.66	0.56	0.74
σ_l	2.16	2.05	1.22	2.98	1.92	1.83	0.91	2.78
r_π	2.18	2.20	1.95	2.47	2.03	2.04	1.74	2.33
$r_{\Delta y}$	0.24	0.25	0.21	0.28	0.22	0.22	0.18	0.27
r_y	0.13	0.13	0.10	0.17	0.08	0.08	0.05	0.12
ρ	0.85	0.85	0.82	0.87	0.81	0.81	0.77	0.85
ρ_a	0.98	0.98	0.98	0.99	0.95	0.95	0.94	0.97
ρ_b	0.19	0.21	0.11	0.31	0.18	0.22	0.07	0.36
ρ_g	0.93	0.92	0.89	0.95	0.97	0.97	0.96	0.99
ρ_i	0.71	0.71	0.64	0.78	0.71	0.71	0.61	0.80
ρ_r	0.08	0.10	0.03	0.17	0.12	0.15	0.04	0.24
ρ_p	0.86	0.85	0.78	0.91	0.90	0.89	0.80	0.96
ρ_w	0.97	0.96	0.94	0.98	0.97	0.96	0.94	0.99
σ_a	0.47	0.48	0.44	0.51	0.45	0.45	0.41	0.50
σ_b	0.24	0.24	0.21	0.27	0.24	0.23	0.19	0.27
σ_g	0.50	0.51	0.47	0.54	0.52	0.53	0.48	0.58
σ_i	0.47	0.47	0.42	0.53	0.45	0.45	0.37	0.53
σ_r	0.23	0.24	0.22	0.25	0.24	0.24	0.22	0.27
σ_p	0.14	0.14	0.12	0.17	0.14	0.14	0.11	0.16
σ_w	0.25	0.25	0.22	0.27	0.24	0.24	0.20	0.28
$\bar{\gamma}$	0.27	0.27	0.17	0.36	0.43	0.43	0.40	0.45
$100(\beta^{-1} - 1)$	0.17	0.19	0.09	0.32	0.16	0.16	0.07	0.26
$\bar{\pi}$	0.71	0.73	0.56	0.91	0.81	0.78	0.61	0.96
$\bar{l}$	0.52	0.41	-0.90	1.76	-0.1	0.53	-1.3	2.32

Note: 5% and 95% columns refer to the 5th and 95th percentiles of the distribution of RWM draws.

**Table 2.7:** Posterior distribution of the dynamic parameters

	Full Spectrum				Business cycle			
	Mode	Mean	5%	95%	Mode	Mean	5%	95%
ρ_{ga}	0.48	0.47	0.38	0.57	0.24	0.24	0.11	0.37
μ_w	0.94	0.92	0.88	0.96	0.28	0.32	0.11	0.58
μ_p	0.68	0.67	0.53	0.78	0.65	0.55	0.24	0.77
α	0.21	0.21	0.18	0.23	0.18	0.19	0.16	0.21
ψ	0.70	0.68	0.54	0.82	0.52	0.56	0.34	0.77
φ	5.52	5.76	4.32	7.39	2.55	3.03	2.15	4.37
σ_c	1.90	1.88	1.61	2.16	1.31	1.50	1.18	1.95
λ	0.64	0.64	0.58	0.70	0.58	0.55	0.45	0.66
ϕ_p	1.61	1.61	1.51	1.72	1.43	1.46	1.34	1.59
ι_w	0.55	0.55	0.37	0.72	0.58	0.56	0.33	0.79
ξ_w	0.84	0.82	0.76	0.87	0.81	0.80	0.73	0.86
ι_p	0.19	0.21	0.10	0.33	0.66	0.61	0.35	0.83
ξ_p	0.66	0.66	0.60	0.71	0.70	0.69	0.62	0.76
σ_l	2.05	1.97	1.14	2.88	2.66	2.51	1.53	3.53
r_π	2.18	2.20	1.95	2.46	2.11	2.10	1.82	2.40
$r_{\Delta y}$	0.24	0.25	0.21	0.28	0.21	0.22	0.18	0.26
r_y	0.13	0.13	0.10	0.17	0.15	0.15	0.10	0.20
ρ	0.85	0.85	0.82	0.87	0.77	0.76	0.71	0.81
ρ_a	0.98	0.98	0.97	0.99	0.82	0.84	0.70	0.94
ρ_b	0.19	0.21	0.11	0.31	0.81	0.75	0.60	0.87
ρ_g	0.92	0.92	0.89	0.95	0.90	0.89	0.83	0.95
ρ_i	0.72	0.72	0.65	0.79	0.70	0.67	0.53	0.79
ρ_r	0.08	0.09	0.03	0.17	0.35	0.34	0.13	0.55
ρ_p	0.86	0.86	0.79	0.91	0.80	0.75	0.48	0.91
ρ_w	0.97	0.96	0.93	0.98	0.57	0.56	0.37	0.73
σ_a	0.47	0.48	0.44	0.51	0.47	0.48	0.42	0.55
σ_b	0.24	0.24	0.21	0.27	0.07	0.08	0.06	0.11
σ_g	0.50	0.51	0.47	0.54	0.35	0.36	0.32	0.41
σ_i	0.47	0.47	0.42	0.52	0.33	0.38	0.27	0.53
σ_r	0.23	0.24	0.22	0.25	0.12	0.13	0.10	0.16
σ_p	0.14	0.14	0.12	0.17	0.08	0.08	0.06	0.12
σ_w	0.25	0.25	0.23	0.27	0.16	0.19	0.12	0.29
$\bar{\gamma}$	0.40	0.41	0.25	0.57	0.39	0.40	0.23	0.56
$100(\beta^{-1} - 1)$	0.22	0.26	0.12	0.44	0.23	0.28	0.13	0.47

Note: 5% and 95% columns refer to the 5th and 95th percentiles of the distribution of the RWM draws.

Table 2.8: Log likelihood and log posterior values at posterior modes

| | **Posterior Mode** | | | |
	SW(2007)	Full Spectrum	Full Spectrum and Mean	BC Frequencies
Log likelihood				
Full Spectrum	2390.46	**2440.24**	2440.18	1150.83
Full Spectrum and Mean	2351.66	n/a	**2388.28**	n/a
BC Frequencies	511.74	523.42	523.71	**577.72**
	SW(2007)	Full Spectrum	Full Spectrum and Mean	BC Frequencies
Log posterior				
Full Spectrum	2375.75	**2418.07**	2416.88	1153.54
Full Spectrum and Mean	2368.27	n/a	**2412.28**	n/a
BC Frequencies	497.03	501.25	500.40	**580.43**

Note. Entries in the table correspond to the log likelihoods/log posteriors, as specified by row labels, evaluated at different posterior modes, which were computed by maximizing the log posterior specified by column labels. For example, the upper left corner gives the value of the log likelihood constructed using Fourier frequencies between $2\pi/T$ and $2\pi(T-1)/T$ with the parameter value set to the posterior mode of SW(2007).
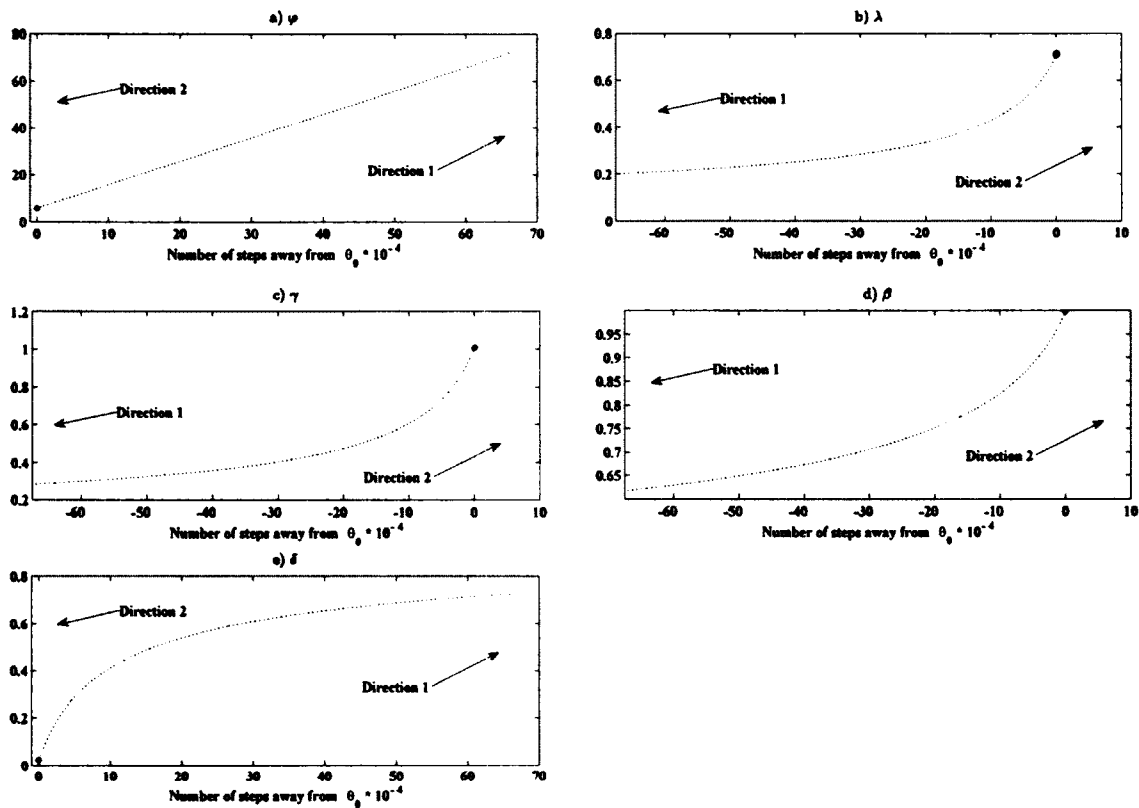
Note. The nonidentification curve is given by $\partial\theta(v)/\partial v = c(\theta)$, $\theta(0) = \theta_0$, where $c(\theta)$ is the eigenvector corresponding to the only zero eigenvalue of $G(\theta)$. The curve is computed recursively using the Euler method, so that $\theta(v_{j+1}) = \theta(v_j) + c(\theta(v_j))h$, where h is the step size, fixed at 1e-04. $(\varphi, \lambda, \gamma, \beta, \delta)$ 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 discount rate greater than 1 contradicts economic theory, direction 2 is truncated at the last point where β is below 1. The curve is extended for 670,000 steps in direction 1. Since there are only 472 steps in direction 2, the respective curve appears as a bold dot on the sub-figures.

Figure 2·1: The nonidentification curve $(\varphi, \lambda, \gamma, \beta, \delta)$
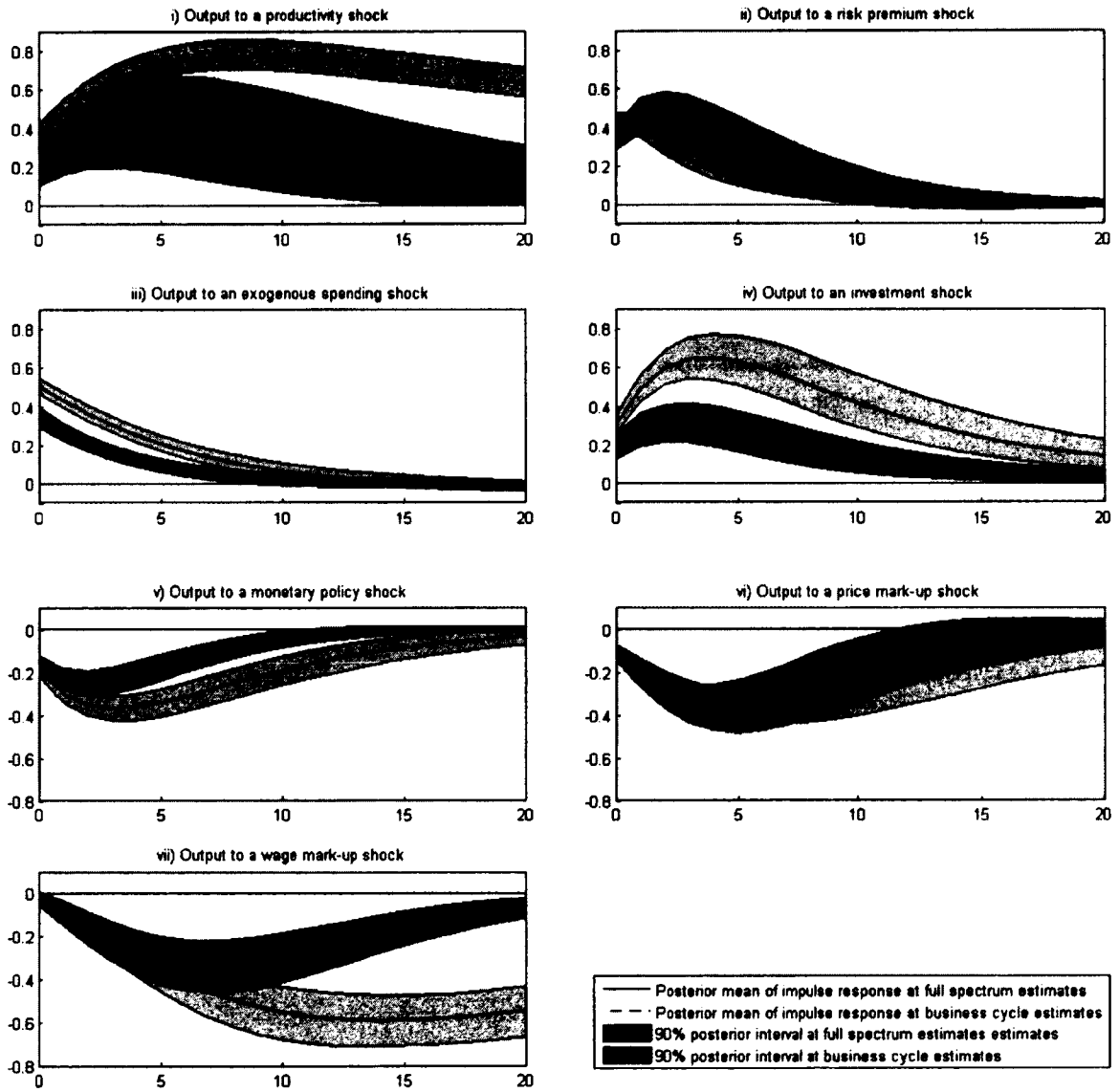
Figure 2·2: The estimated impulse responses of output to shocks
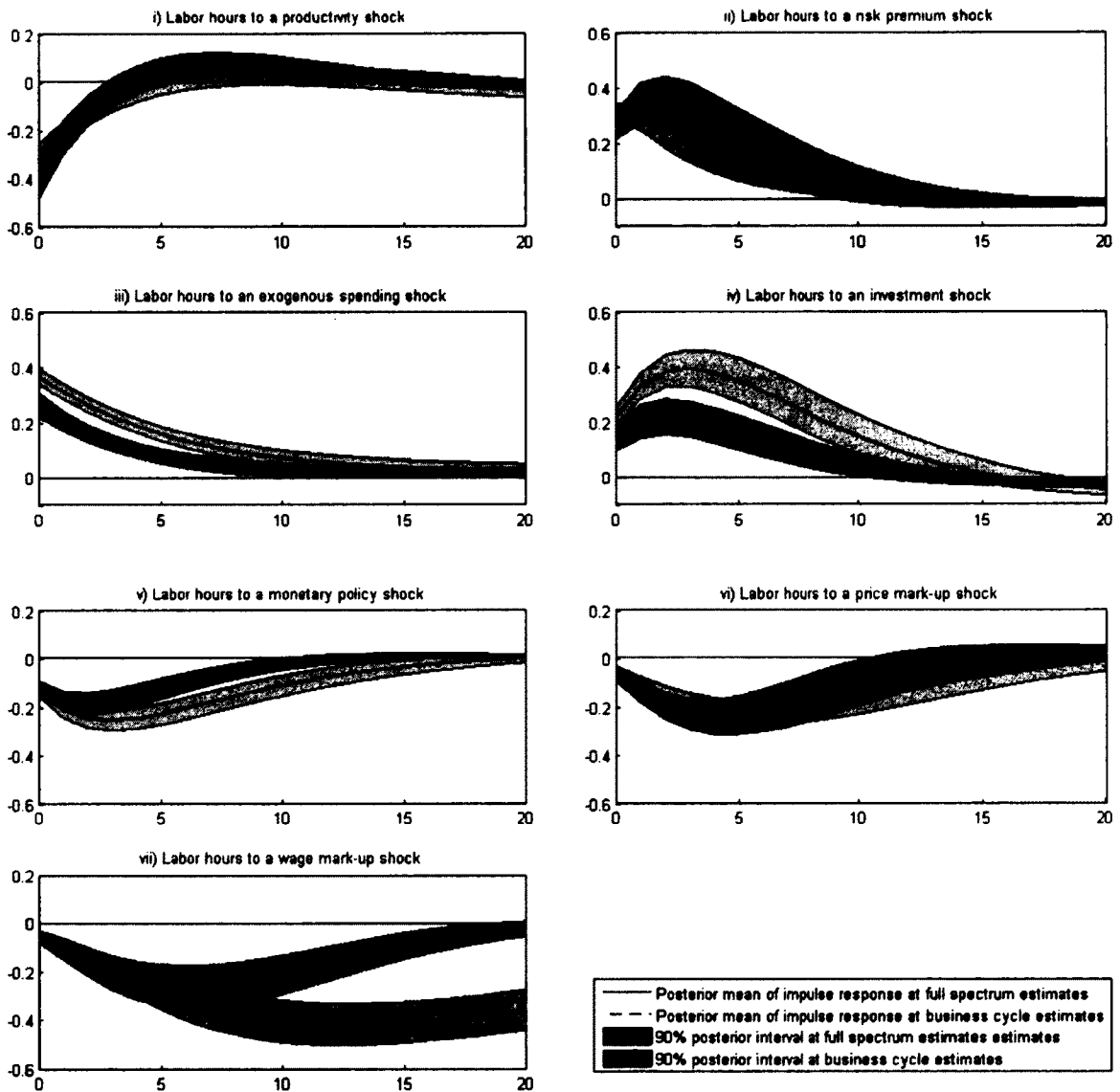
Figure 2·3: The estimated impulse responses of labor hours to shocks
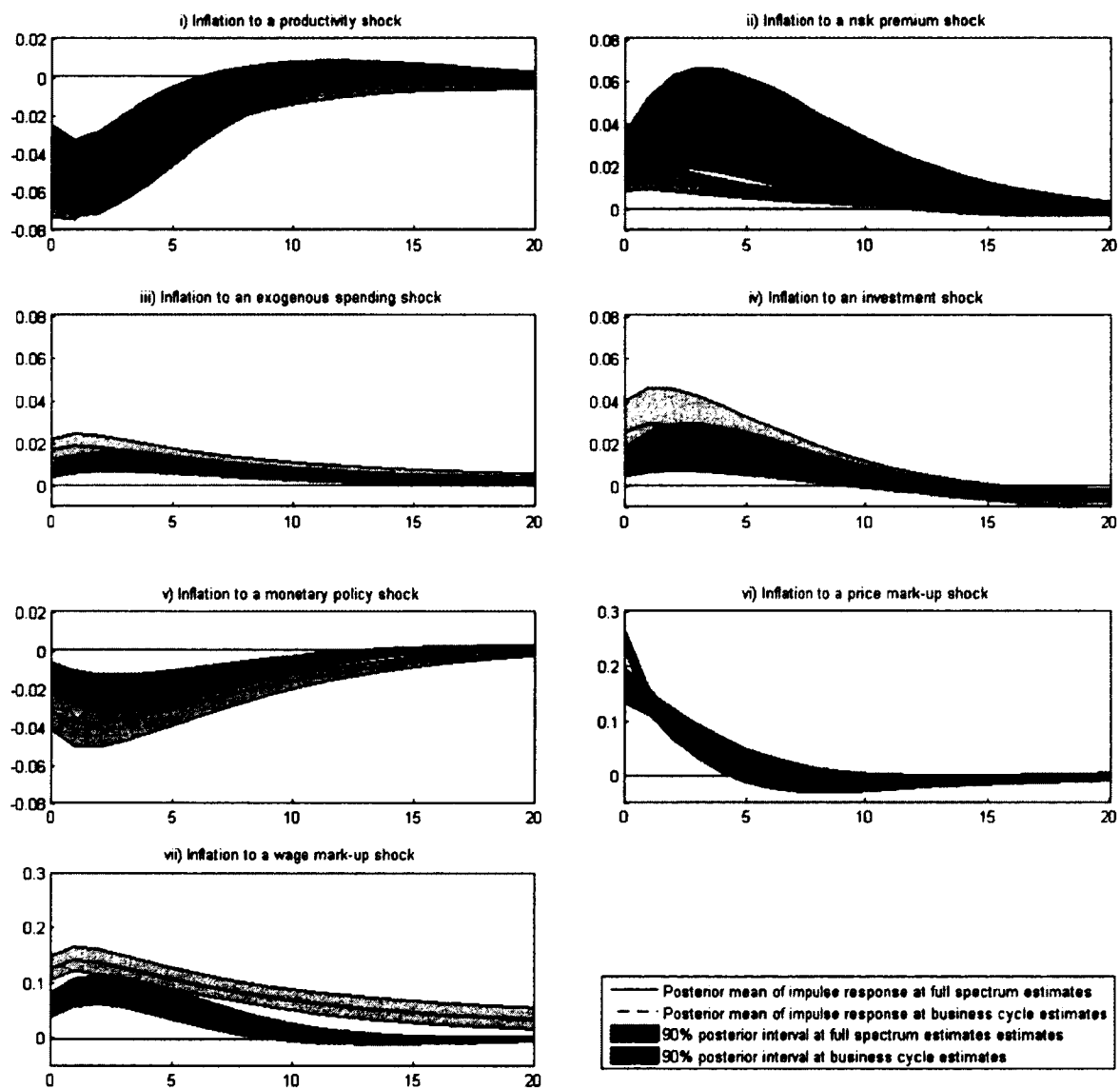
Figure 2·4: The estimated impulse responses of inflation to shocks
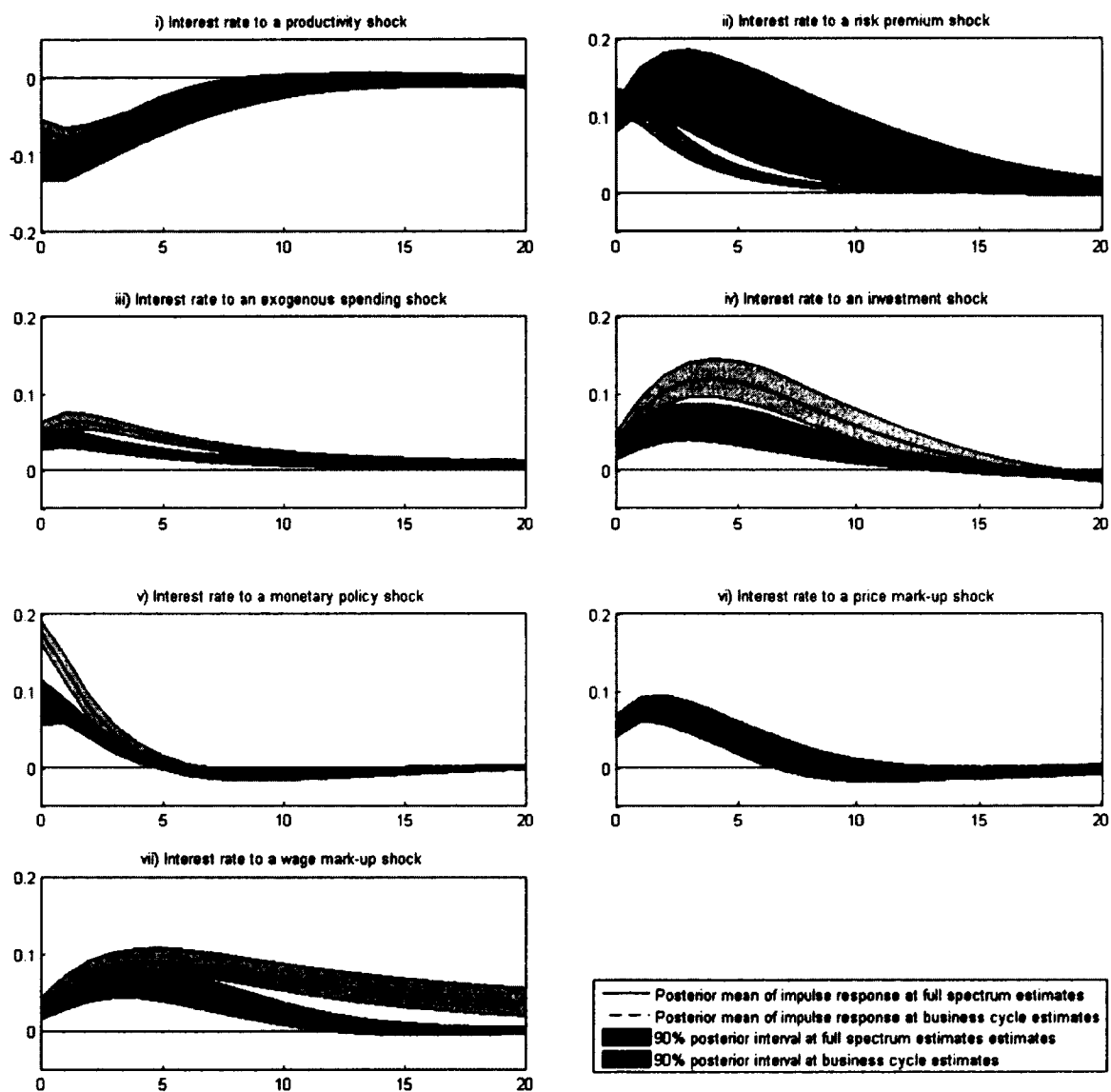
Figure 2·5: The estimated impulse responses of interest rate to shocks
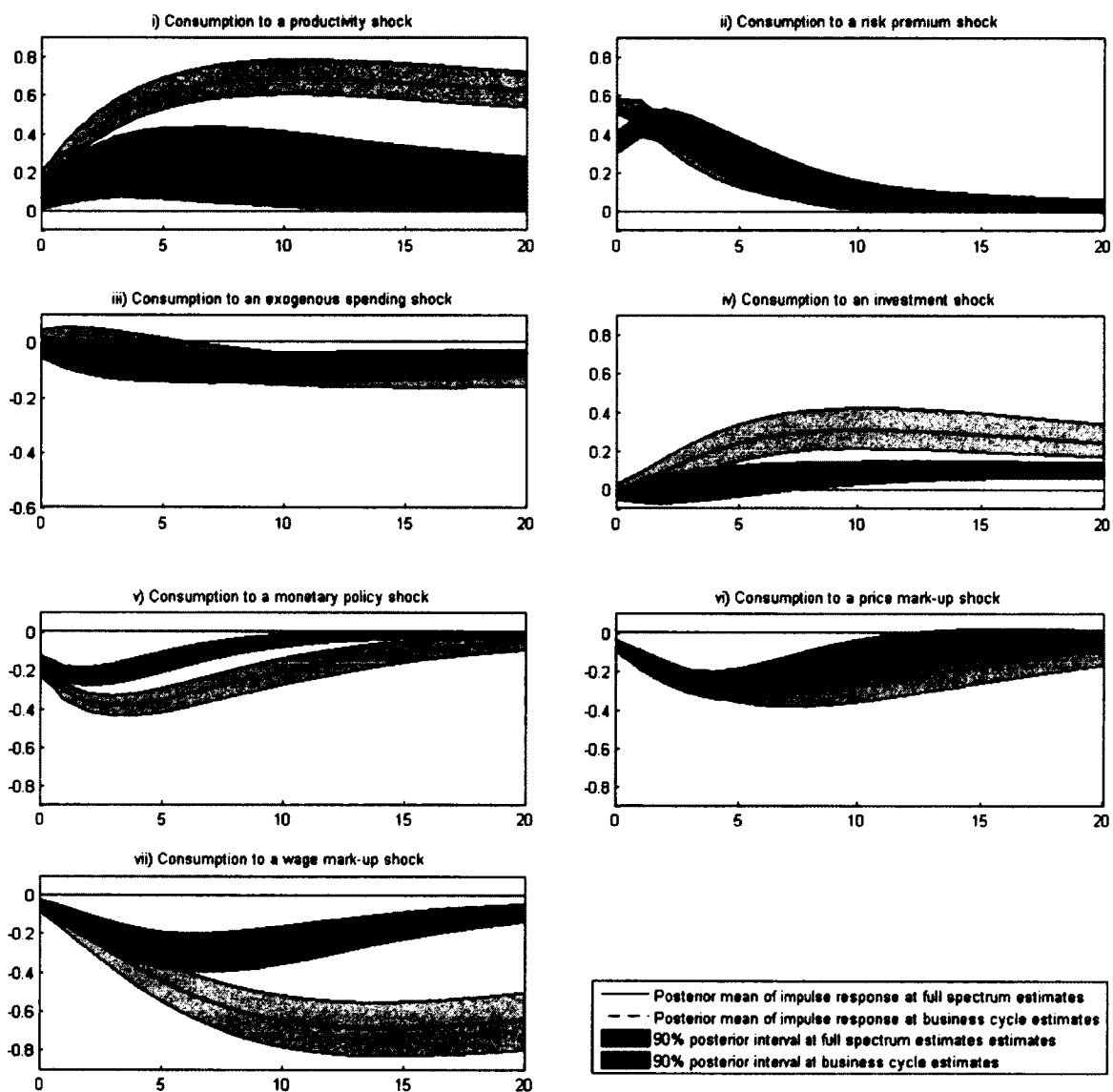
Figure 2·6: The estimated impulse responses of consumption to shocks
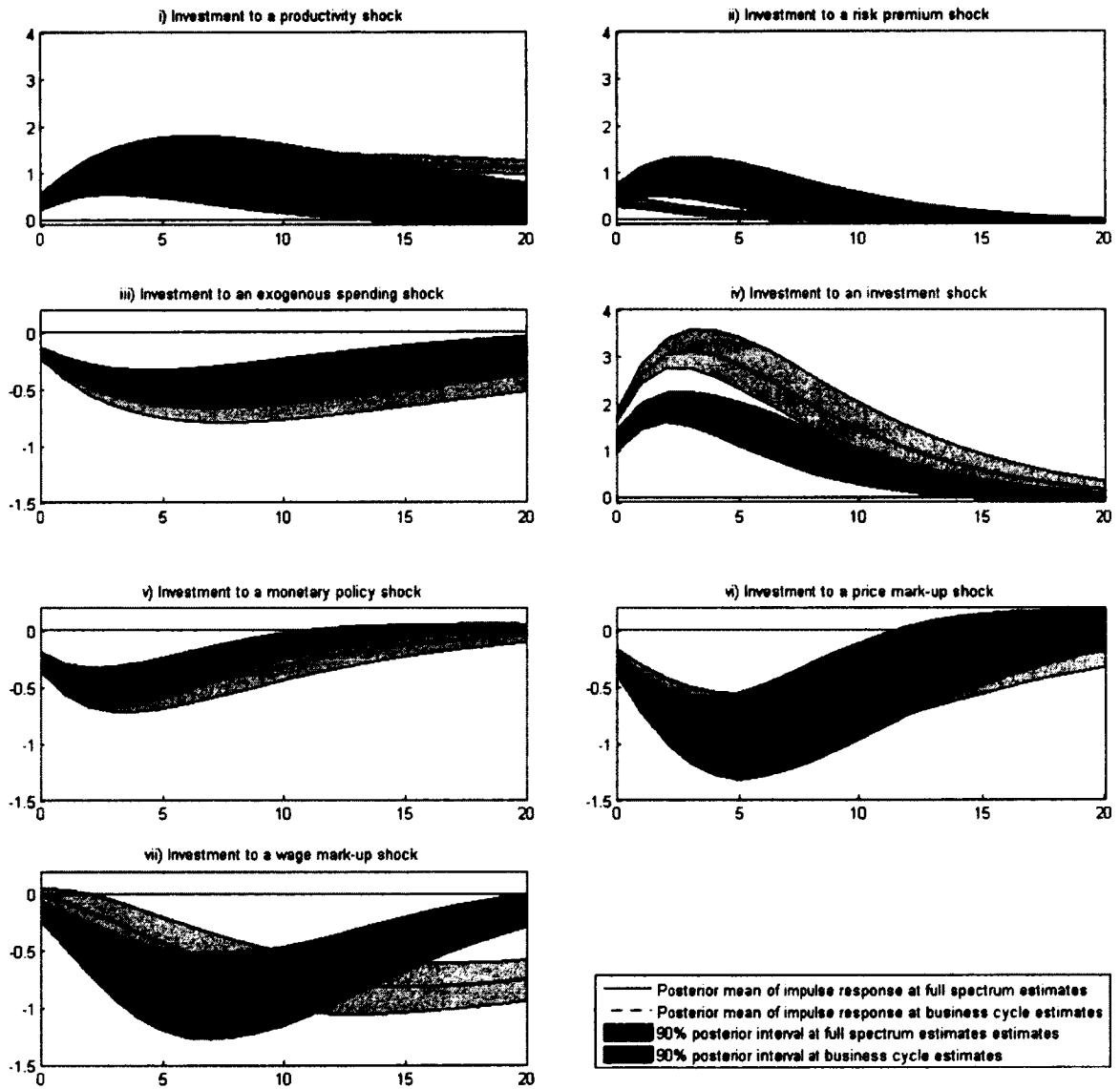
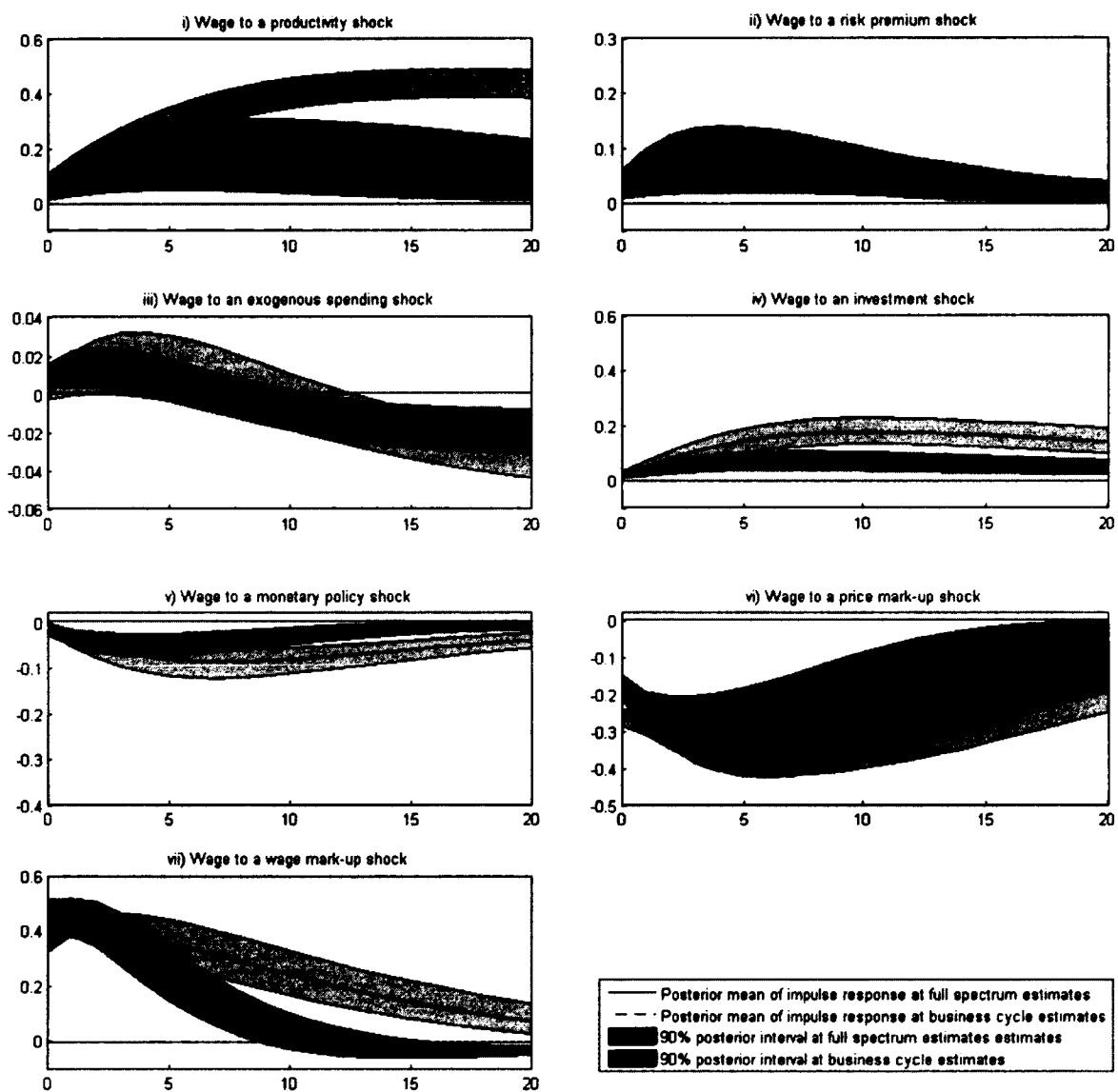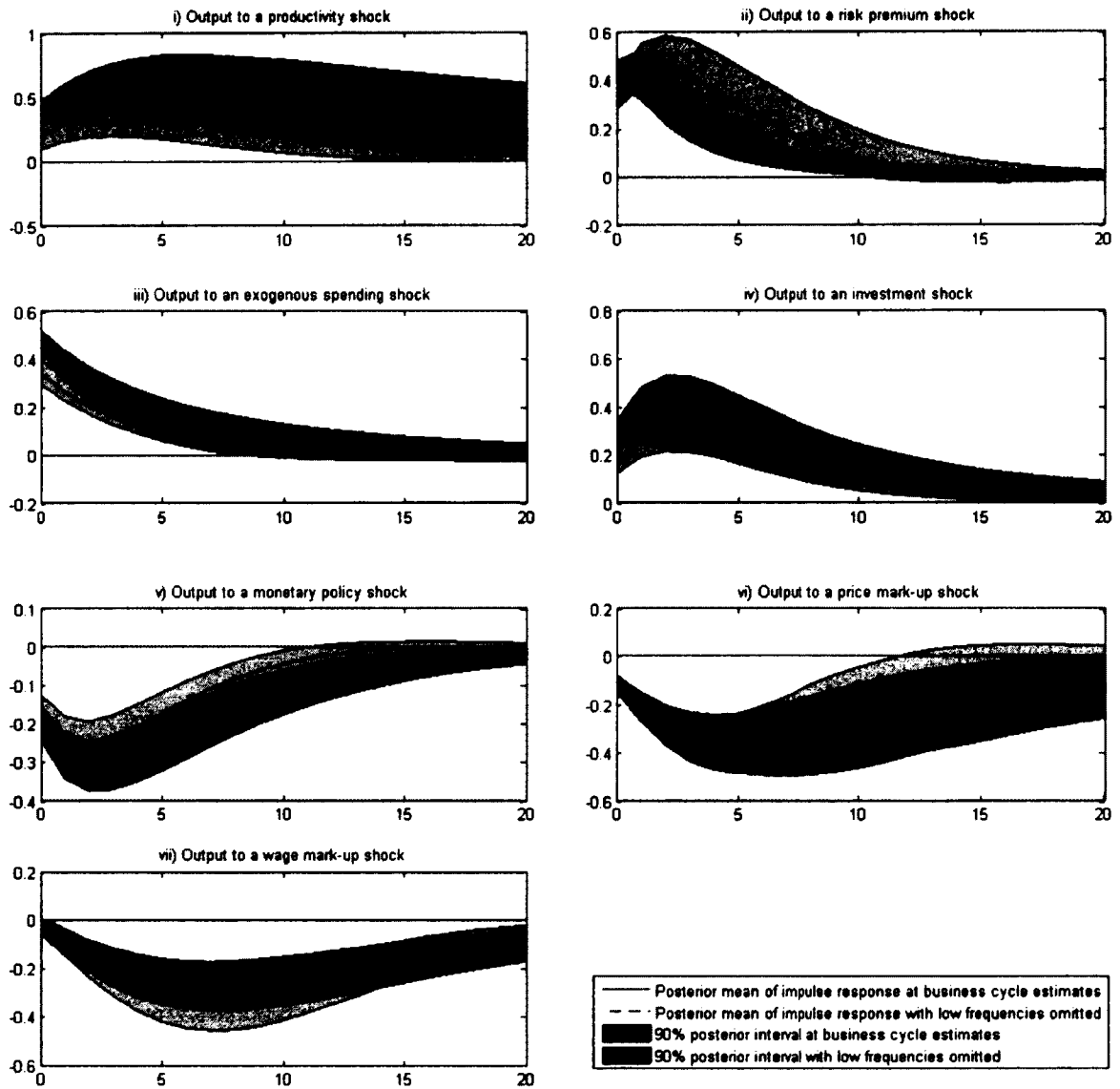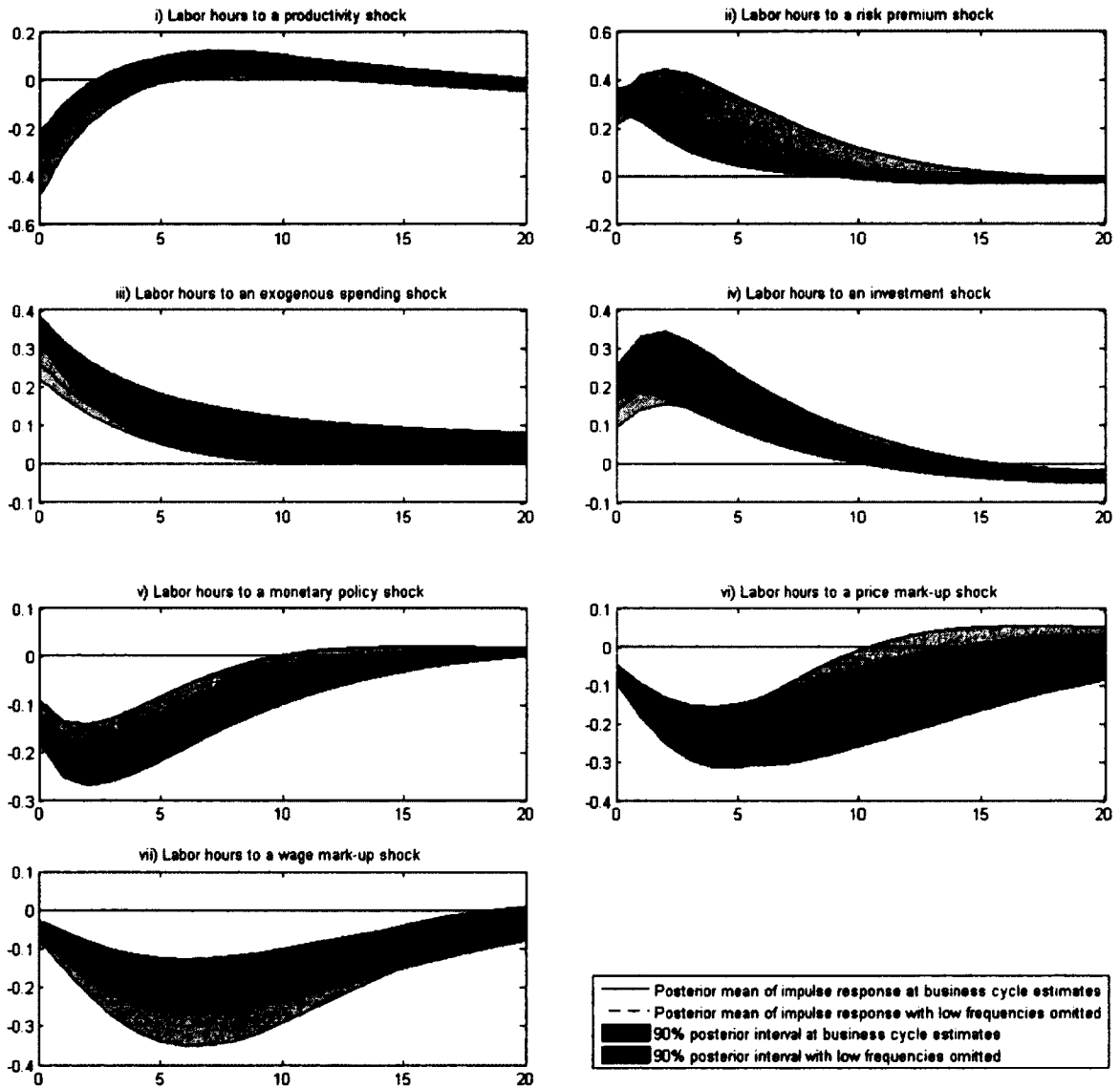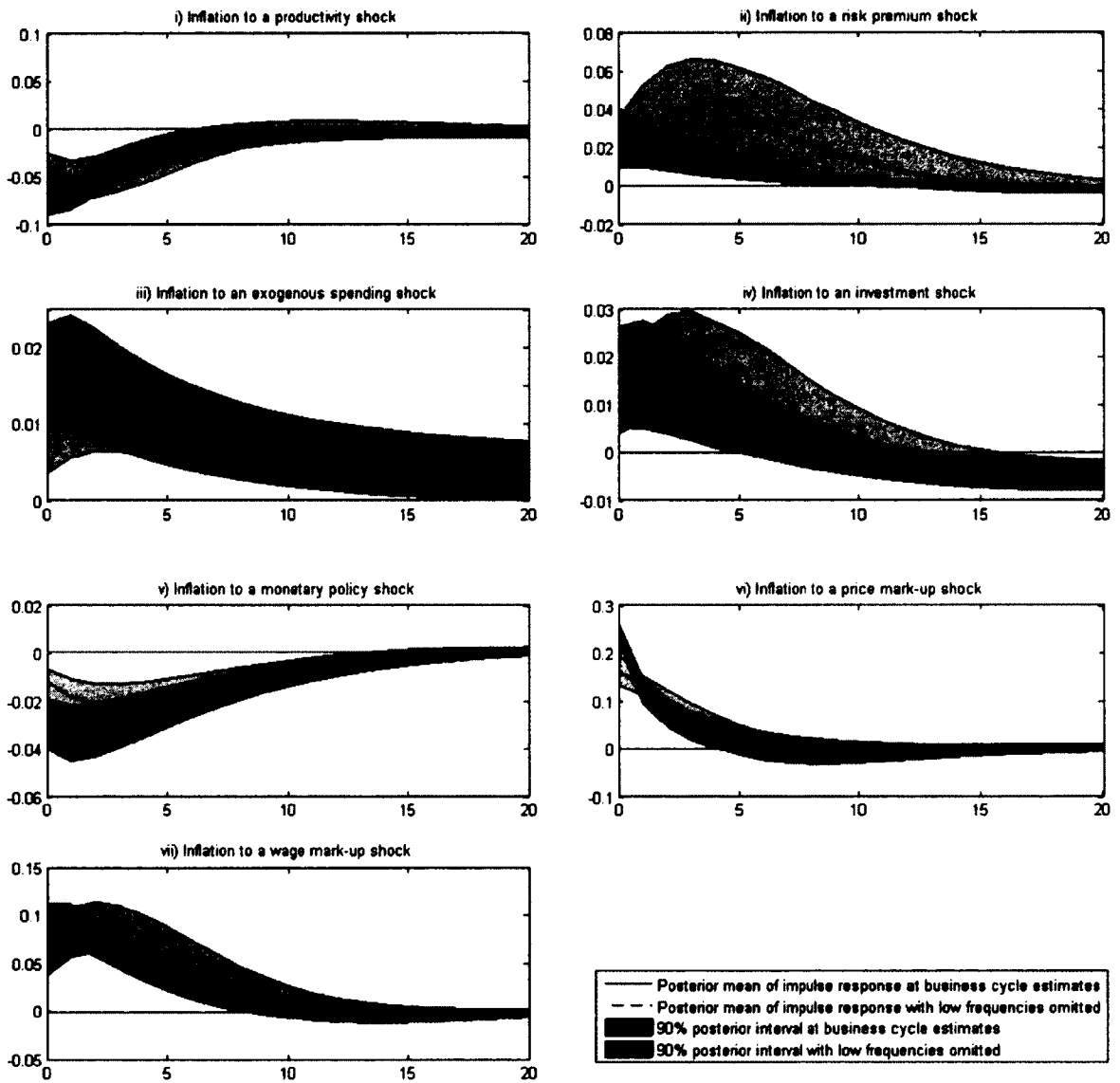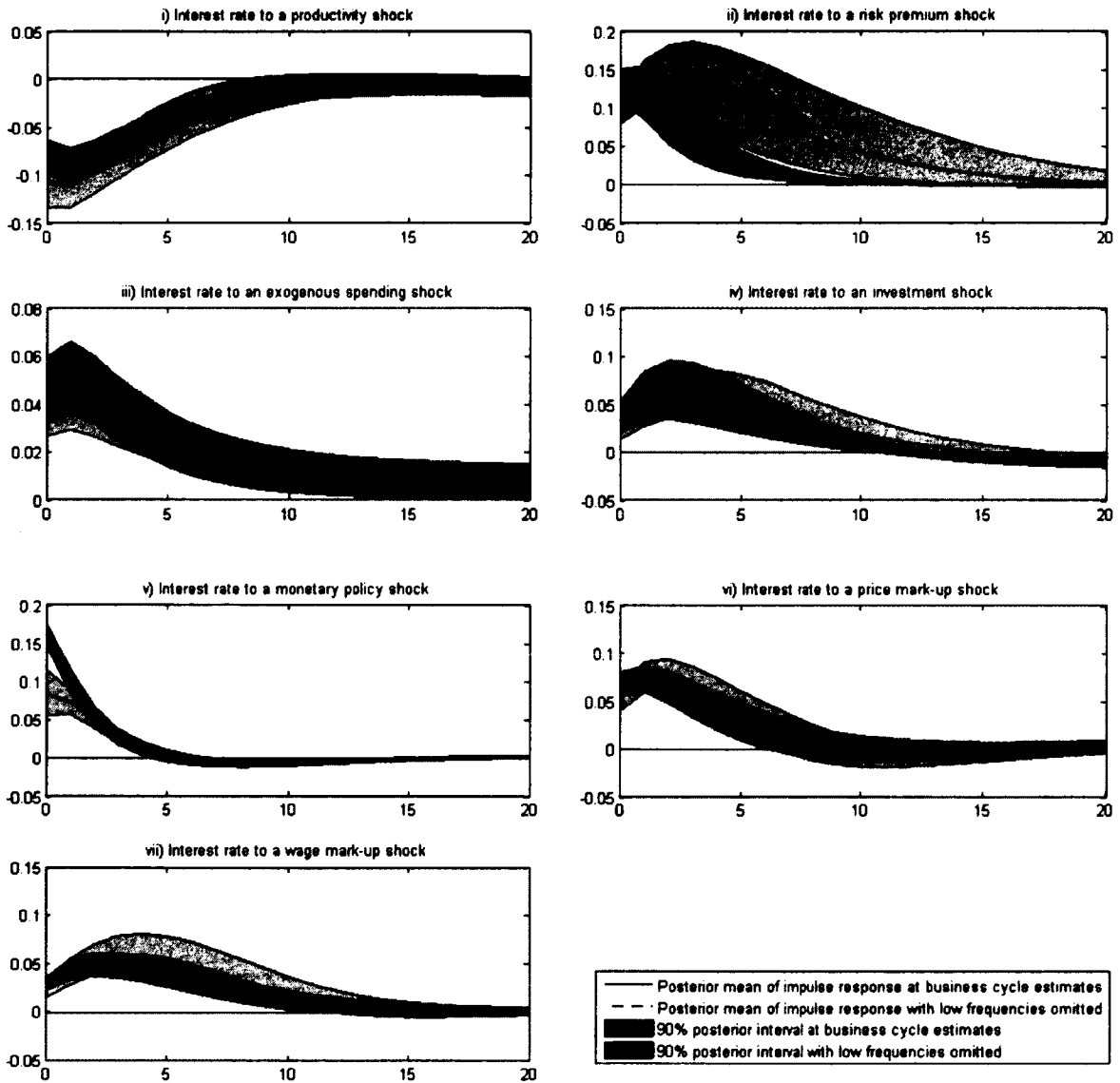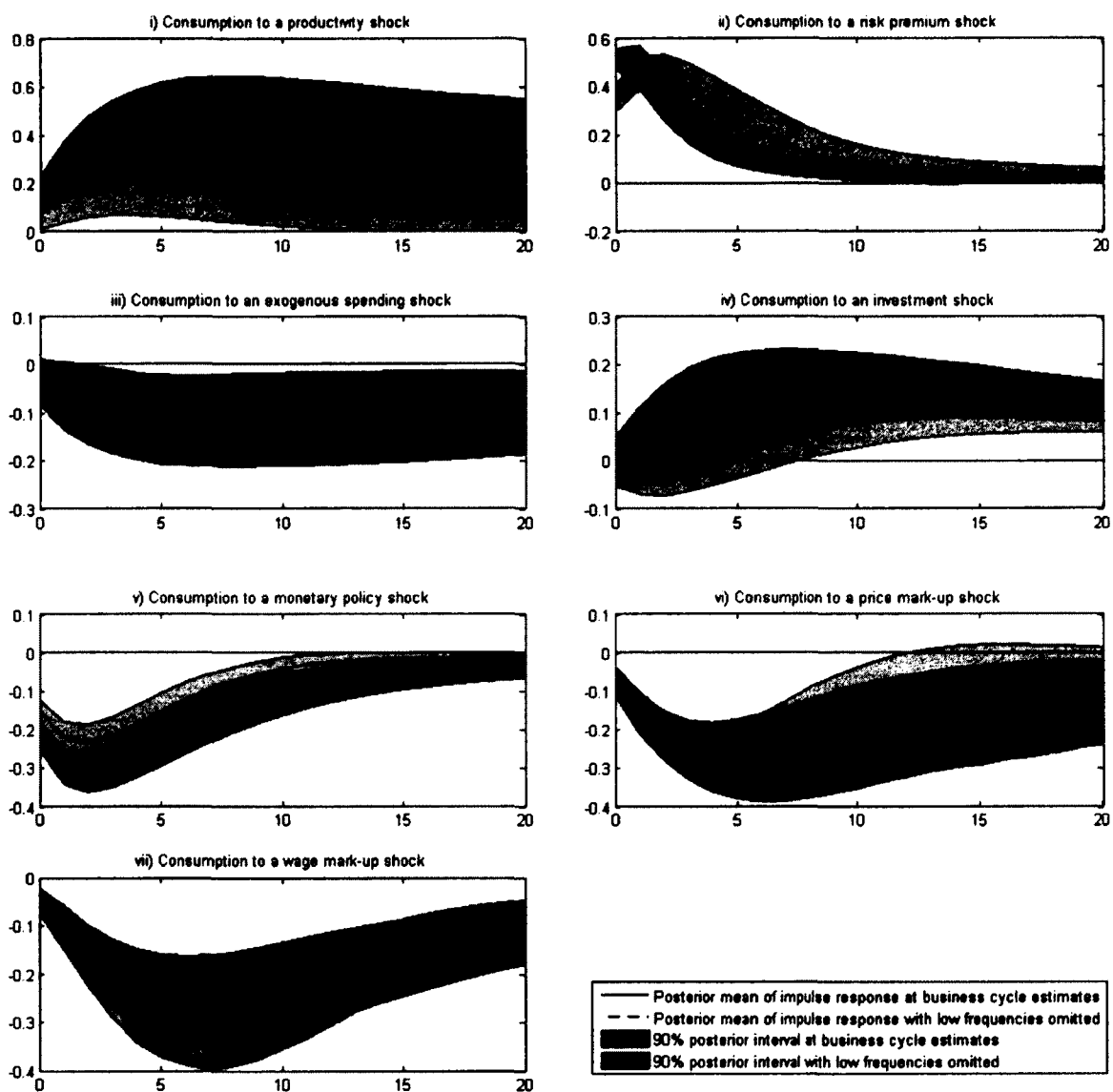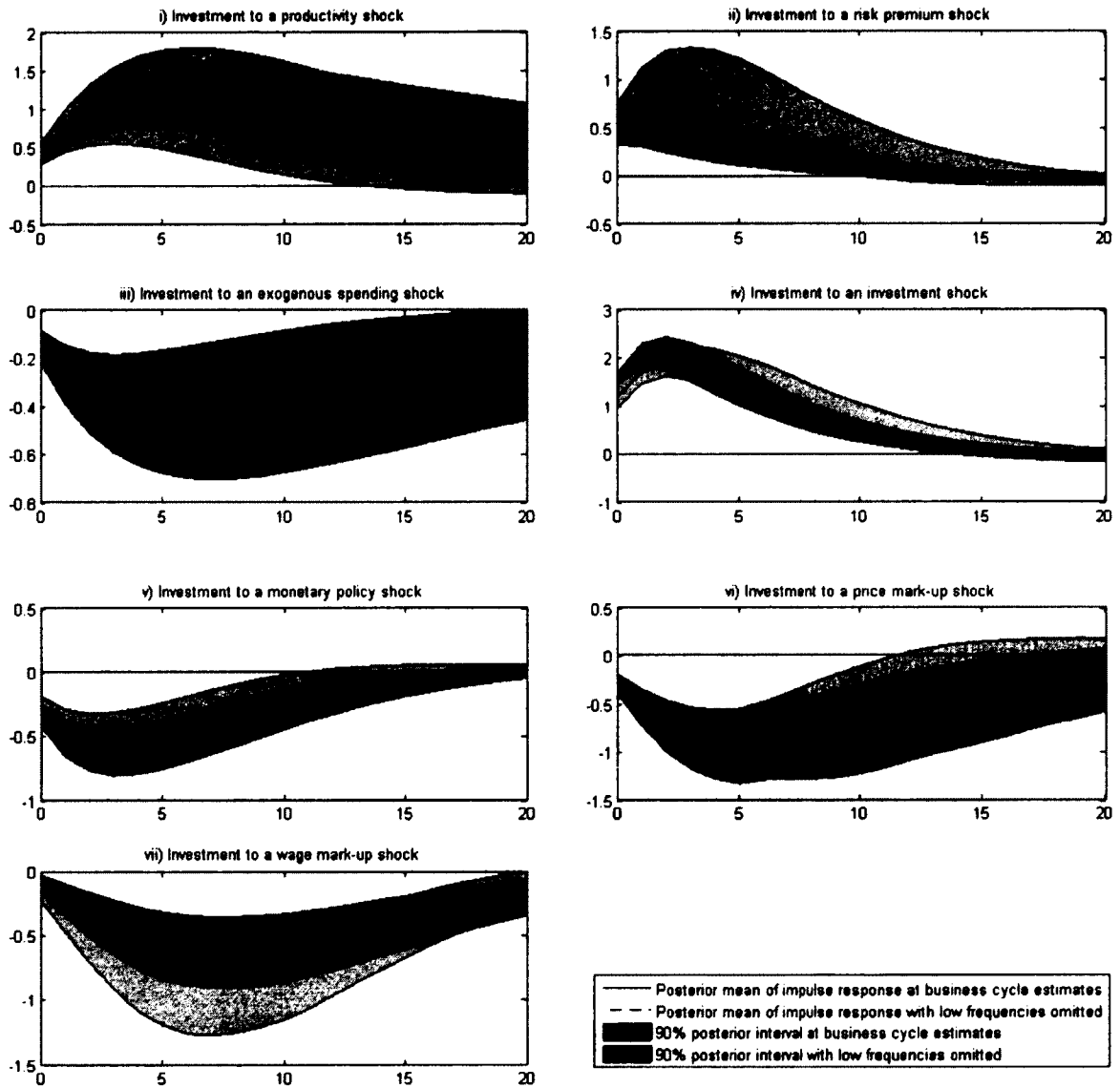**Figure 2·7:** The estimated impulse responses of investment to shocks

Figure 2·8: The estimated impulse responses of wage to shocks

Figure 2·9: The estimated impulse responses of output to shocks

Figure 2·10: The estimated impulse responses of labor hours to shocks

Figure 2·11: The estimated impulse responses of inflation to shocks

i) Interest rate to a productivity shock

ii) Interest rate to a risk premium shock

iii) Interest rate to an exogenous spending shock

iv) Interest rate to an investment shock

v) Interest rate to a monetary policy shock

vi) Interest rate to a price mark-up shock

vii) Interest rate to a wage mark-up shock

Posterior mean of impulse response at business cycle estimates
Posterior mean of impulse response with low frequencies omitted
90% posterior interval at business cycle estimates
90% posterior interval with low frequencies omitted

Figure 2·12: The estimated impulse responses of interest rate to shocks

Figure 2·13: The estimated impulse responses of consumption to shocks

Figure 2·14: The estimated impulse responses of investment to shocks
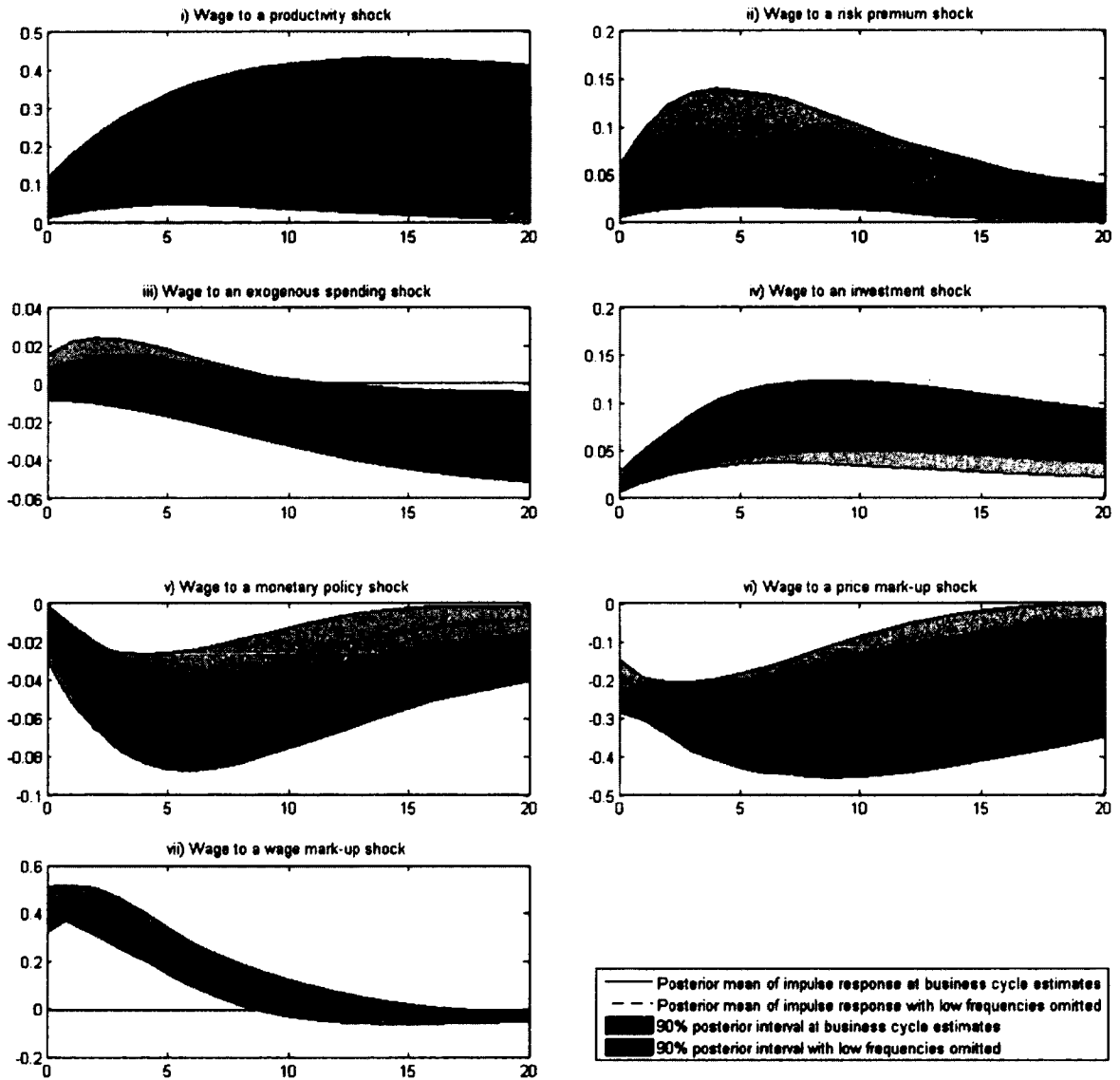
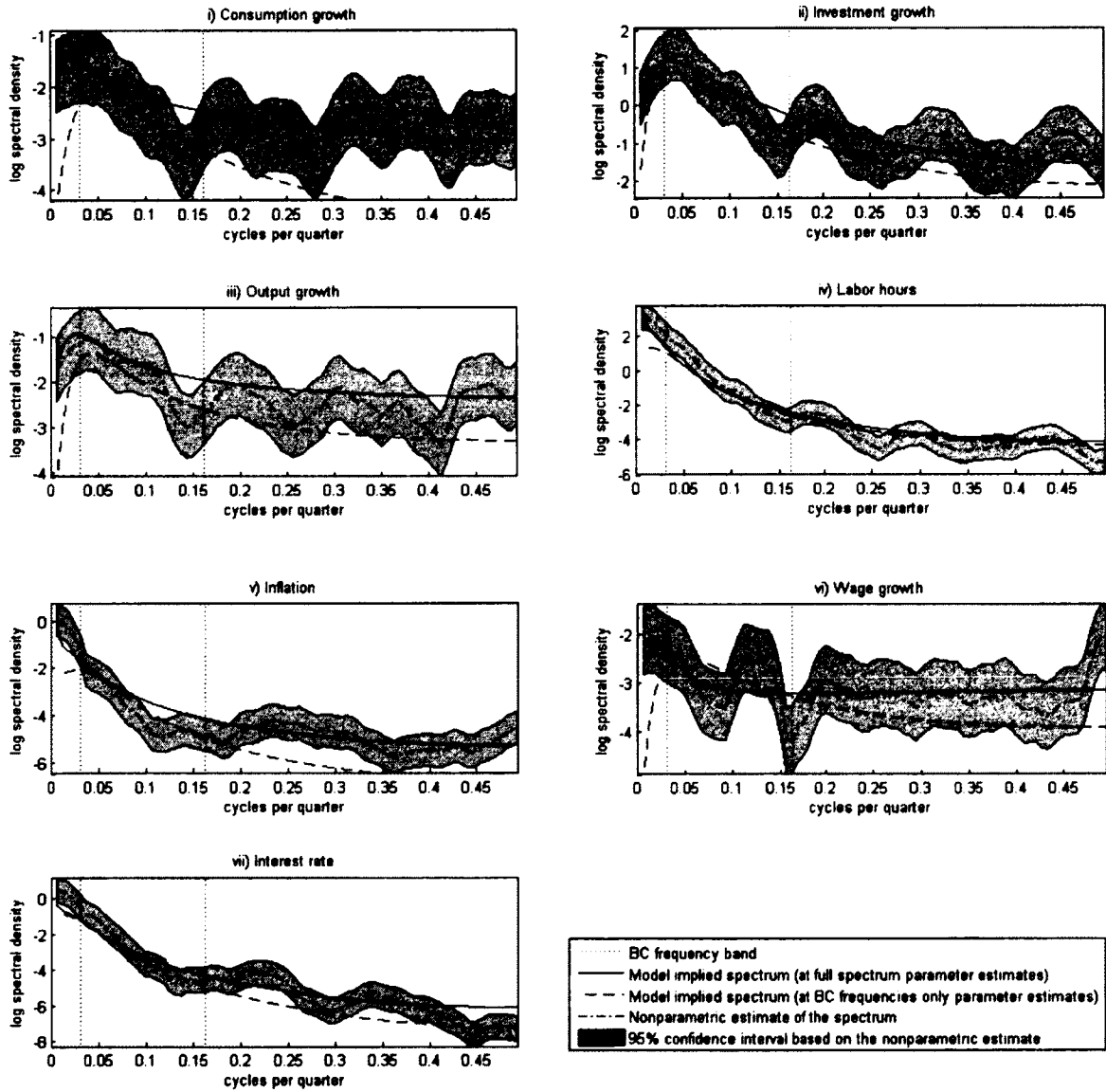**Figure 2·15:** The estimated impulse responses of wage to shocks

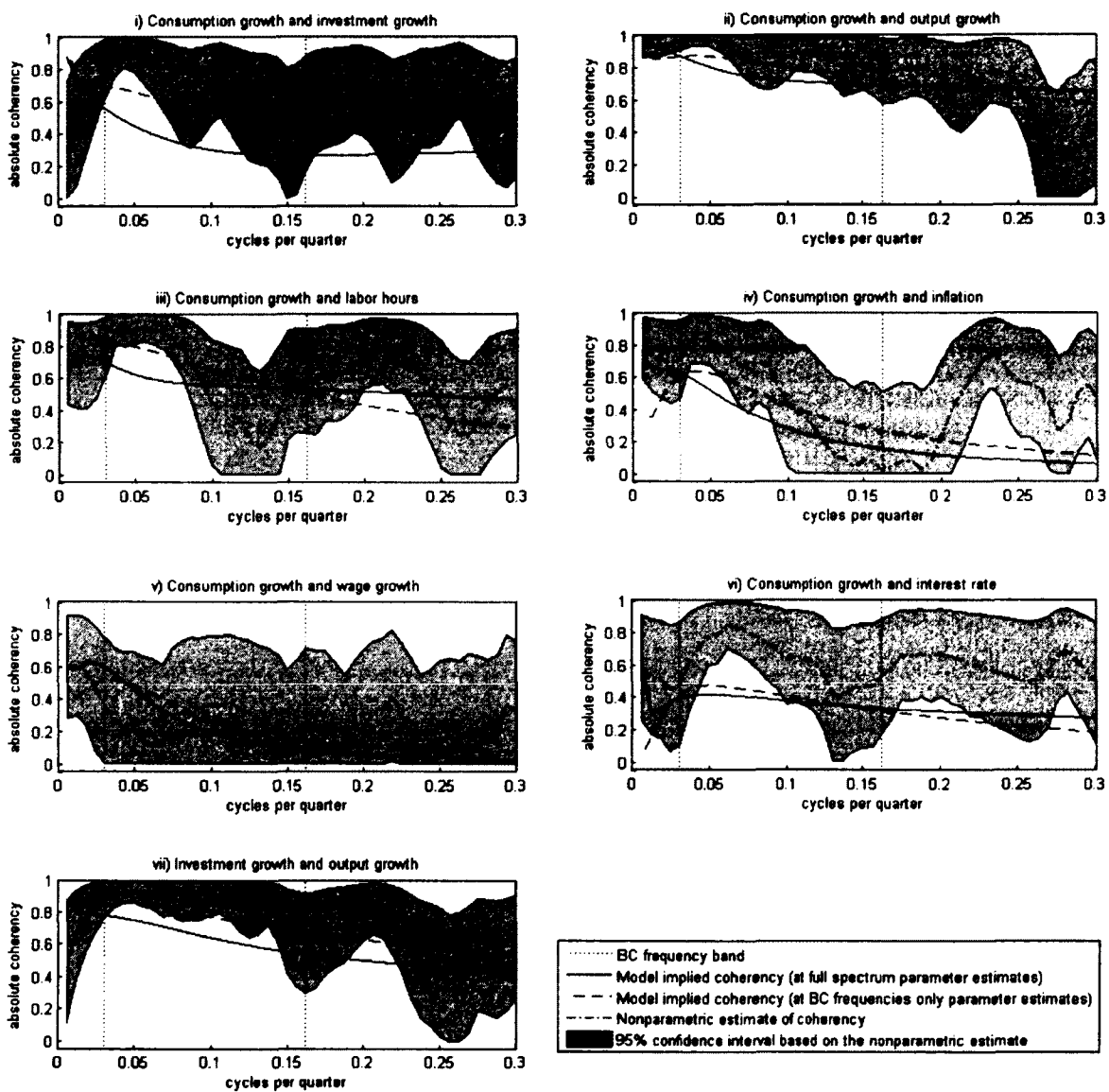**Figure 2·16:** Model implied and nonparametrically estimated log spectra of observables in the model

Figure 2·17: Model implied and nonparametrically estimated coherency between observables in the model
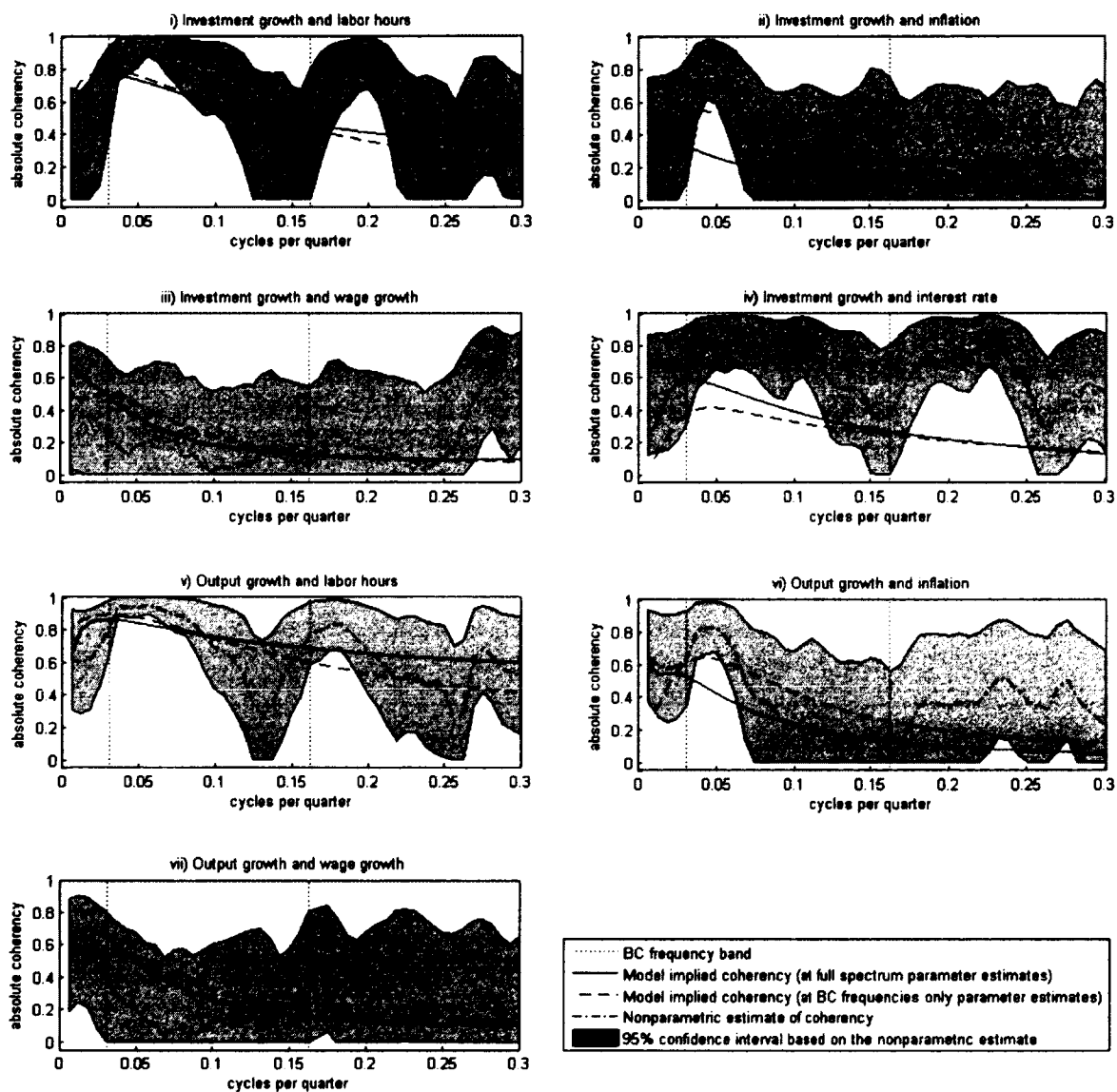
Figure 2·18: Model implied and nonparametrically estimated coherency between observables in the model
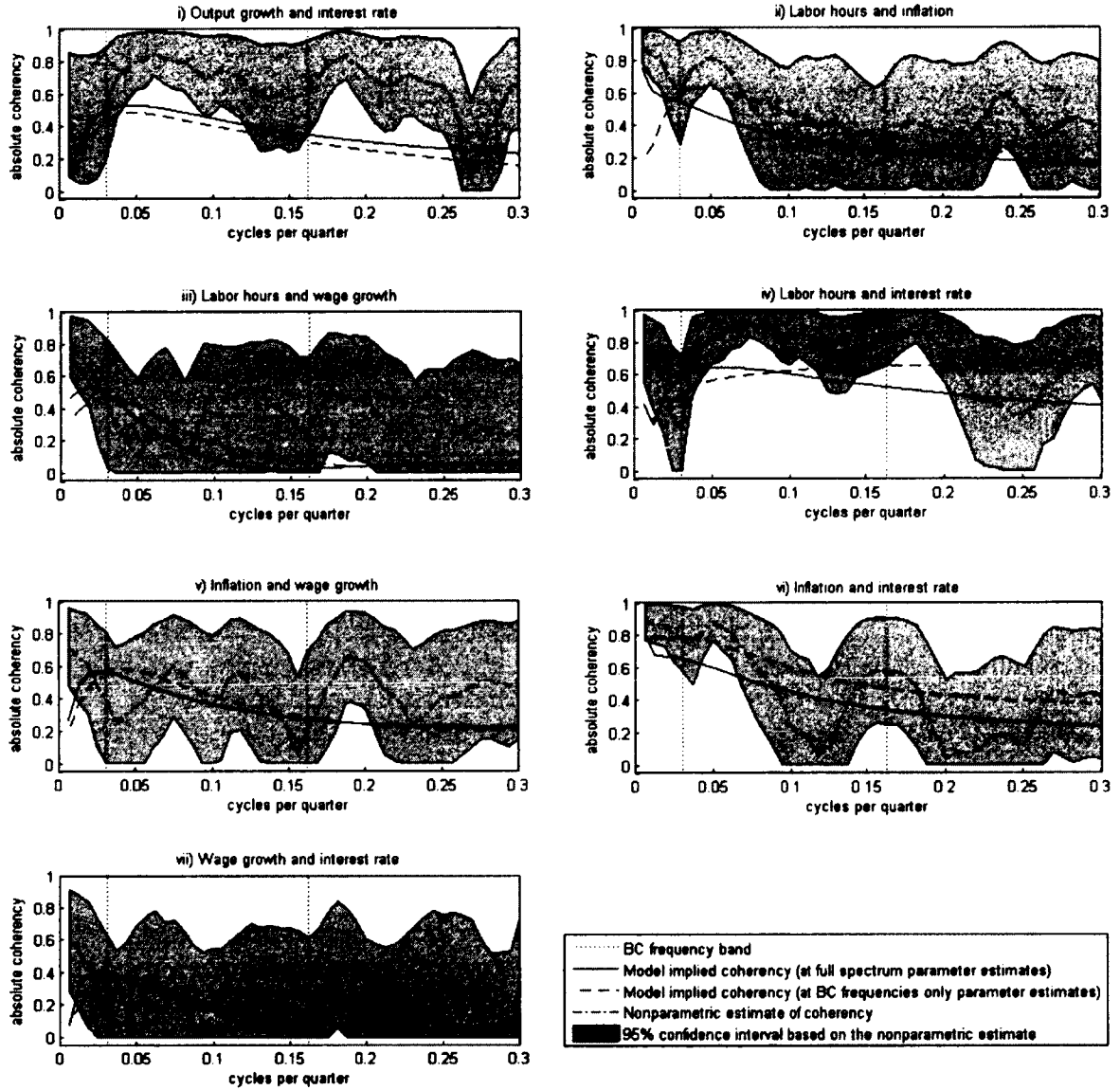
Figure 2·19: Model implied and nonparametrically estimated coherency between observables in the model

Chapter 3

Frequency Domain QML Volatility Estimation with Noisy High Frequency Data

3.1 Introduction

Integrated volatility (IV) of a financial asset is one of the key quantities in modern finance. The ability to obtain accurate estimates of intraday volatility is crucial in the areas of derivatives pricing, volatility forecasting, and evaluation of volatility models. It has been argued that daily volatility measures constructed from high frequency data capture more information and using them delivers better results in the above areas. However, despite the ever increasing availability of high frequency data, the issues pertaining to microstructure effects prevent researchers from using all of the available observations. The microstructure noise is inherent in the data due to various trading frictions, data recording errors, sampling methodology (i.e., using transactions or quote data), and becomes more severe at higher sampling frequencies. The common practice in financial econometrics literature is to aggregate data to lower sampling frequencies of 5 to 30 minutes in order to reduce the influence of microstructure noise, which results in discarding most observations in the process. Aït-Sahalia et al. (2005) have shown that if the microstructure noise is accounted for in estimation, then it is optimal to sample as often as possible. These considerations have motivated the growing literature on integrated volatility estimators robust to microstructure noise. Most of such suggested estimators are nonparametric. Specifically, one approach

consists of bias corrected subsampling and averaging over different time scales. The seminal contribution here is Zhang et al. (2005), who introduced the Two-Scale Realized Volatility (TSRV) estimator, which is consistent at the rate of $N^{1/6}$ in the presence of i.i.d. noise. Later, in the same setting, Zhang (2006) suggested a more involved Multi-Scale Realized Volatility (MSRV) estimator with an improved convergence rate of $N^{1/4}$, which was shown by Gloter and Jacod (2001) to be the optimal rate for this problem. More recently, both TSRV and MSRV have been modified by Aït-Sahalia et al. (2011) to be robust to serially dependent noise. Another class of estimators is based on weighting autocovariances and realized variances. The first estimator of this type was introduced by Zhou (1996) and later extended by Hansen and Lunde (2006) to accommodate stochastic volatility and serially dependent noise. Although unbiased, these estimators are inconsistent. Barndorff-Nielsen et al. (2008) introduced realized kernels and demonstrated that for certain choices of weight functions their estimators achieve the optimal convergence rate and can be asymptotically equivalent or even more efficient than TSRV and MSRV. However, these type of estimators involve choices of tuning parameters, such as the number of subsamples to average over or the bandwidth in the case of realized kernels.

The parametric approach, namely, the maximum likelihood estimator (MLE) in this setting has enjoyed less attention, but important contributions have been made recently. The simulation studies of Gatheral and Oomen (2010) and Aït-Sahalia and Yu (2009) have shown that the MLE estimator introduced in Aït-Sahalia et al. (2005) in the constant volatility setting performs well when applied to data generated from stochastic volatility models. In a recent paper Xiu (2010) has formalized the parametric approach by showing that, when viewed as a quasi-estimator that misspecifies spot volatility to be constant, the MLE of Aït-Sahalia et al. (2005) achieves consistency at the optimal rate and has a mixed

normal asymptotic distribution. Furthermore, he established that the quasi-maximum likelihood (QML) estimator is asymptotically equivalent to the optimal realized kernel with an implicitly specified bandwidth, and performs better than alternative realized kernels in finite samples.

While Xiu (2010) has shown that the time domain quasi-maximum likelihood (TDQML) estimator works well in the case of the i.i.d. microstructure noise, the likelihood function and the asypmtotic properties of the estimator become difficult to analyze once serially dependent noise is considered. Even considering the case with i.i.d. noise requires a cumbersome change of variables to represent the returns as an MA(1) process in order to both obtain theoretical results and perform the computations in practice. Xiu (2010) made a heuristic argument for combining subsampling with the QML estimator assuming i.i.d. noise when a finite order moving average (MA) noise process is suspected. However, this approach does not allow for extension to autoregressive (AR) or autoregressive moving average (ARMA) noise specifications, which may be empirically relevant, as argued, among others, by Engle and Sun (2007) and Aït-Sahalia et al. (2011) . Furthermore, it is not clear whether this time domain estimator can be extended to accommodate dependence between the noise and the efficient price process. This motivates us to tackle this problem from the frequency domain perspective, using the quasi-maximum likelihood estimation procedure that was first proposed by Whittle (1951). Intuitively, the desirable properties of the time domain estimator should be preserved by its frequency domain version as well. In addition, it should be more tractable analytically, allowing for more flexibility in specifying the time series properties of the noise process, as well as afford a potential possibility to incorporate endogenous noise by including a cross-spectrum term into the quasi-likelihood.

Although the literature on volatility estimation is vast, relatively few papers consid-

ered the problem from the frequency domain perspective. Malliavin and Mancino (2002) introduced the Fourier estimator, which is based on the truncated Dirichlet kernel. Mancino and Sanfelici (2008) developed a variant of this estimator using the Fejér kernel, and Malliavin and Mancino (2009) extended it to the multivariate setting and provided the optimal number of Fourier coefficients to minimize the mean squared error (MSE) of the estimator. However, these estimators are rarely used in the literature and Gatheral and Oomen (2010) found that they are dominated in finite samples by the time domain QML and realized kernel estimators. The work closest to the current chapter is the article by Olhede et al. (2009), who use the type of FDQML considered in this chapter in order to compute weights for their shrinkage estimator of the integrated volatility that takes the form of the sum of weighted periodograms of contaminated log returns.

We conduct our analysis under the same assumption on the log price process as in Xiu (2010), namely that it follows a Brownian motion with stochastic volatility that is a positive and locally bounded Brownian semimartingale. This specification of the stochastic volatility is quite general and encompasses most continuous time financial models (e.g., see Jacod (2008), hypothesis $(L-s)$). We also follow Xiu (2010) in omitting the drift term in our specification in order to simplify the algebra as he has argued that the effects of the drift are asymptotically negligible. The data is assumed to be equally spaced in time and sampled at a very high frequency. The asymptotic results are thus considered within the infill asymptotics framework where the number of observations within a fixed time interval, e.g., one day, goes to infinity.

There are two conceptual complications for FDQML estimation that arise in this framework. First, as the volatility of the log return process is stochastic, it is nonstationary and thus its spectral density in the traditional sense is not defined. This presents a problem,

because the Whittle likelihood that is maximized to obtain FDQML estimates requires specification of the spectral density of the underlying data generating process. Second, the integrated volatility that we are trying to estimate is a random object itself. We suggest to deal with the first issue by replacing the spectral density of the log returns by the variance of their discrete Fourier transform. The second issue can be circumvented by using the concept of stable convergence, which is used extensively in the finance literature, to take expectations and derive the asymptotic distribution conditional on the realization of the integrated volatility on a given day.

We begin by considering the case without microstructure noise present. In this case, the FDQML estimator reduces to the well known realized variance (RV) estimator, whose properties were studied, among others, by Barndorff-Nielsen and Shephard (2002). This is a very intuitive result, as RV is known to be consistent and efficient in this case. However, it was demonstrated by Zhang et al. (2005) that it consistently estimates the variance of the noise instead of IV when i.i.d. microstructure noise is present, a problem that motivated the search for alternative estimators. We proceed to include the i.i.d. noise into our specification, which amounts to adding a term corresponding to the spectral density of the first differenced noise into the Whittle likelihood. The closed form solution in this case is not available and hence the maximization is done numerically. In this case, the simulation results suggest that the asymptotic properties of the estimator are the same as in Xiu (2010). Finally, we suggest a more general estimation approach that admits microstructure noise that follows a linear stationary process. Berk (1974) has shown that the spectral density of a linear process can be consistently estimated by fitting a finite order AR model, with the order of the approximation growing with sample size, while Shibata (1981) suggested that using the Akaike's (1973) information criterion (AIC) for lag

selection in such a procedure is asymptotically efficient. We therefore consider estimating integrated volatility by FDQML while specifying the spectrum of the noise as that of an AR process, with the order chosen using AIC. We illustrate the application of the method via simulation, by considering two empirically relevant microstructure noise specifications given by the AR(1) and ARMA(1,1) processes. Our simulations show that the proposed FDQML performs adequately well, while the TDQML estimator for these processes is very difficult to specify and may not be feasible in practice.

The rest of this chapter is organized as follows. Section 3.2 introduces the high frequency data setting that we are going to work with. Section 3.3 puts forward the FDQML estimator, whose conjectured asymptotic properties are discussed in Section 3.4. Section 3.5 presents the findings of the simulation study, and, finally, Section 3.6 concludes.

3.2 Setup

Throughout the chapter, we work with the following setup. The latent efficient log price process is given by

$$dX_t = \sigma_t dW_t,$$

where $X_0 = 0$, W_t is a Brownian motion, and the stochastic volatility process σ_t is assumed to be a positive and locally bounded Brownian semimartingale. This assumption on σ_t is quite general, allowing for almost any existing continuous time stochastic volatility model (see Hypothesis $(L - s)$ in Jacod (2008) for more details). The object of estimation, given by

$$IV_{(0,T)} = \int_0^T \sigma_t^2 dt,$$

is the integrated volatility of the above process over some fixed interval $[0, T]$, which can be thought of as a trading day for most empirical applications. Here we assume that the

observed log prices are regularly spaced with $N + 1$ observations per day, and indexed by $\tau_k = k\Delta$, $k = 0, 1, 2..., N$. The observations lie within the interval $[0, T]$, so $T = N\Delta$. Since T is assumed to be fixed, the asymptotics are considered with $N \to \infty$ and $\Delta \to 0$ simultaneously, which is a standard framework in the literature. To shorten notation, denote the log return between periods τ_i and τ_{i-1} by Y_i.

When noise is present, the econometrician observes the contaminated process

$$\widetilde{X}_{\tau_i} = X_{\tau_i} + U_{\tau_i},$$

where $\{U_{\tau_i}\}$ is the microstructure noise process independent of $\{X_t\}$. In general, $\{U_{\tau_i}\}$ can be assumed to be a stationary ARMA(P,Q) process

$$A(L)U_{\tau_i} = B(L)\varepsilon_{\tau_i},$$

where L denotes the lag operator, $A(L) = (1 - \sum_{k=1}^{p} \phi_k L^k)$, $B(L) = (1 + \sum_{j=1}^{q} \eta_j L^j)$, and ε_{τ_i} is assumed to be i.i.d. with mean zero and variance a^2.

3.3 FDQML estimator

Let the Fourier frequencies be denoted by $\lambda_j = 2\pi j/N$, $j = 1, 2, ..., N - 1)$. Let $\omega_Y(\lambda_j)$ denote the discrete Fourier transforms given by

$$\omega_Y(\lambda_j) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} Y_k \exp(-i\lambda_j k), \quad j = 1, 2, ..., N - 1.$$

Since the volatility of the log price increment process is stochastic, it is not second order stationary and its spectral density is not defined in the traditional sense. Instead, we will work with the covariance of its discrete Fourier transform defined by

$$f_X(\lambda_j, \lambda_k) = E(\omega_X(\lambda_j)\omega_X(\lambda_k)^*) - E(\omega_X(\lambda_j))E(\omega_X(\lambda_k)^*),$$

where the asterisks denote conjugate transpose, and we will denote the variance (i.e., when $j = k$) simply by $f_X(\lambda_j)$ to cut down on notation. The microstructure noise process is stationary by assumption, so we can directly compute the spectral density of its first difference as

$$f_U(\lambda_j) = |1 - \exp(-i\lambda_j)|^2 \frac{a^2 |B(\exp(-i\lambda_j))|^2}{|A(\exp(-i\lambda_j))|^2}.$$

Since $IV_{(0,T)}$ is a random quantity itself, stable convergence arguments are used for analysis. Hence, in the rest of the chapter, all expectations are taken conditional on a particular realization of $IV_{(0,T)}$. Using this, we obtain

$$
\begin{aligned}
f_X(\lambda_j) &= E(\omega_X(\lambda_j)\omega_X(\lambda_j)^*) = \frac{1}{N} E\left(\sum_{k=1}^{N}(X_{\tau_k} - X_{\tau_{k-1}})^2\right) = \\
&= \frac{1}{N} \int_0^T E(\sigma_t^2)dt = \frac{1}{N} \int_0^T \sigma_t^2 dt.
\end{aligned}
$$

Hence, the contribution of each frequency to the integrated volatility is approximately the same. This resemblance to the spectral properties of white noise motivates us to consider FDQML estimation of $IV_{(0,T)}$ by purposely misspecifying the spectral density of the latent log return as if its volatility were constant. In order to remove dependence on N in the estimated parameter, we define

$$\bar{\sigma}^2 \equiv \frac{1}{T} \int_0^T \sigma_t^2 dt$$

to be the parameter of interest, which implies $f_X(\lambda_j) = \bar{\sigma}^2 \Delta$.

Following Whittle (1951) and using $(f_X(\lambda_j) + f_U(\lambda_j))$ as a proxy for the spectral density of the contaminated log returns, the Whittle likelihood for our problem is given,

up to a constant multiplication, by

$$L(\theta) = \sum_{j=1}^{N-1} \log(f_X(\lambda_j, \theta) + f_U(\lambda_j, \theta)) + \frac{I_Y(\lambda_j)}{(f_X(\lambda_j, \theta) + f_U(\lambda_j, \theta))},$$

where θ is the vector of estimated parameters ($\bar{\sigma}^2$, a^2 and any additional AR or MA coefficients of the noise process), and $I_Y(\lambda_j) = \omega_Y(\lambda_j)\omega_Y(\lambda_j)^*$ is the periodogram. The estimates $\hat{\theta}$ are obtained by minimizing $L(\theta)$. Next, we describe several empirically relevant microstructure noise specifications and discuss statistical properties of the FDQML estimator in each case.

3.4 Statistical properties of FDQML

3.4.1 Baseline case: no noise

First, it instructive to consider the simplest case where the microstructure noise is absent, i.e., $A(L) = B(L) = 1$ and $a^2 = 0$. The Whittle likelihood in this case is given by

$$L(\theta) = \sum_{j=1}^{N-1} \log(\bar{\sigma}^2 \Delta) + \frac{I_Y(\lambda_j)}{\bar{\sigma}^2 \Delta},$$

with $\theta = \bar{\sigma}^2$. Taking first order condition and setting it equal to zero yields the estimator

$$\hat{\bar{\sigma}}^2 = \frac{1}{(N-1)\Delta} \sum_{j=1}^{N-1} I_Y(\lambda_j).$$

Using the fact that the sum of periodograms across the Fourier frequencies is equal to the sum of squared returns (see p. 332 in Brockwell and Davis (2006)), we can write

$$\lim_{N \to \infty} \hat{\bar{\sigma}}^2 = \frac{1}{T} \sum_{j=1}^{N} Y_i^2,$$

which coincides with the traditional realized variance (RV) estimator. This is not entirely surprising, as Xiu (2010) obtained RV as a time domain QML estimator in the absence of

noise as well. It is well known that RV is both consistent and efficient in the absence of noise. Hence, there are no gains or losses compared to the standard estimator associated with using FDQML in the absence of noise.

3.4.2 i.i.d. microstructure noise

We next consider a very popular case in the literature where microstructure noise is i.i.d. $(0, a^2)$. The log returns now become

$$Y_i = \widetilde{X}_{\tau_i} - \widetilde{X}_{\tau_{i-1}} = X_{\tau_i} - X_{\tau_{i-1}} + U_{\tau_i} - U_{\tau_{i-1}},$$

and have an MA(1) structure with variance of Y_i given by $\int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt + 2a^2$ and the first order autocovariance equal to $(-a^2)$. The spectral density of the first differenced noise is simply

$$f_U(\lambda_j) = a^2 |1 - \exp(-i\lambda_j)|^2 = 2a^2(1 - \cos(\lambda_j)).$$

The Whittle likelihood in this case becomes

$$L(\theta) = \sum_{j=1}^{N-1} \log(\overline{\sigma}^2 \Delta + 2a^2(1 - \cos(\lambda_j))) + \frac{I_Y(\lambda_j)}{(\overline{\sigma}^2 \Delta + 2a^2(1 - \cos(\lambda_j)))}, \qquad (3.1)$$

with $\theta = (\overline{\sigma}^2, a^2)'$. There are no closed form expressions for $\widehat{\theta}$, so the maximization has to be performed numerically. Given that FDQML and TDQML are based on the same principle, and the quasi-likelihood (3.1) is the frequency domain approximation to the time domain Gaussian quasi-likelihood used in Xiu (2010), we conjecture that the two estimators share the same asymptotic properties in this case. These are summarized in Conjectures 1 and 2 below.

Conjecture 3.1. *Given the log price process defined in Section 3.2 and U_{τ_i} that is i.i.d.* $(0, a^2)$, *the FDQML estimator $\widehat{\theta} = (\widehat{\overline{\sigma}}^2, \widehat{a}^2)'$ satisfies:* $\widehat{\overline{\sigma}}^2 - \frac{1}{T} \int_0^T \sigma_t^2 dt \xrightarrow{P} 0$ *and* $\widehat{a}^2 - a^2 \xrightarrow{P} 0.$

Conjecture 3.2. *Given the log price process defined in Section 3.2 and U_{τ_i} that is i.i.d.* $(0, a^2)$*, the asymptotic distribution of the FDQML estimator* $\widehat{\theta} = (\widehat{\widetilde{\sigma}}^2, \widehat{a}^2)'$ *is given by*

$$
\begin{pmatrix} N^{1/4} \left(\widehat{\widetilde{\sigma}}^2 - \frac{1}{T} \int\limits_0^T \sigma_t^2 dt \right) \\ N^{1/2} \left(\widehat{a}^2 - a^2 \right) \end{pmatrix} \overset{\mathcal{L}_X}{\longrightarrow}
$$

$$
MN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \dfrac{5a \int\limits_0^T \sigma_t^4 dt}{T \left(\int\limits_0^T \sigma_t^2 dt \right)^{1/2}} + \dfrac{3a \left(\int\limits_0^T \sigma_t^2 dt \right)^{3/2}}{T^2} & 0 \\ 0 & 2a^4 + cum_4[U] \end{pmatrix} \right),
$$

where $\overset{\mathcal{L}_X}{\longrightarrow}$ *denotes stable convergence in* $\sigma(X)$*,* $cum_4[U]$ *denotes the fourth order cumulant of the noise process, and* MN *denotes mixed normal distribution.*

Unfortunately, the asymptotic distribution depends on the integrated quarticity, which is not straightforward to estimate in this setting. One feasible solution is to use the consistent preaveraging estimator of the integrated quarticity as in Jacod et al. (2009) in order to obtain confidence intervals based on the limiting distribution above.

3.4.3 Microstructure noise as a stationary linear process

Several authors have argued that the assumption of i.i.d. microstructure noise is overly simplistic. Hansen and Lunde (2006), Aït-Sahalia et al. (2011) and Engle and Sun (2007), among others, have concluded, based on both theoretical and empirical evidence, that microstructure noise demonstrates serial dependence and is likely correlated with the efficient price process. The above papers put forward different specifications for the independent noise component, based on empirical evidence from assets with different characteristics, that include MA, AR, and ARMA processes. It is therefore desirable for the practitioner to use an estimator that is agnostic to the underlying microstructure noise process.

Shibata (1981) has shown that, under some conditions, fitting an AR model with the

order selected using AIC produces an asymptotically optimal estimate of the spectral density of a stationary linear process representable as AR(∞). We suggest to incorporate a similar idea to obtain an FDQML estimation procedure that allows for the microstructure noise to follow an arbitrary stationary linear process. Specifically, the estimator solves

$$\min_{\bar{\sigma}^2, a^2, [\phi_k]_1^P, P} \left[2 \left(\sum_{j=1}^{N-1} \log(\bar{\sigma}^2 \Delta + 2a^2(1 - \cos(\lambda_j))) + \frac{I_Y(\lambda_j)}{(\bar{\sigma}^2 \Delta + f_U(\lambda_j))} \right) + 2(P+2) \right],$$

where $f_U(\lambda_j)$ is given by

$$f_U(\lambda_j) = \frac{2a^2(1 - \cos(\lambda_j))}{\left| 1 - \sum\limits_{k=1}^{P} \phi_k(\exp(-i\lambda_j))^k \right|^2}.$$

We conjecture that such an estimator should retain the desirable properties (consistency and asymptotic mixed normality), while allowing for a broad family of the underlying noise processes. The next section investigates these conjectures via simulations.

3.5   Simulation study

In this section we conduct Monte Carlo simulations to evaluate the performance of the FDQML estimator in the various settings described in the previous section. We generate samples of one day length, so that, in annual units, $T = 1/252$. The true data generating process for stochastic volatility follows the Heston (1993) model with the Cox-Ingersoll-Ross (CIR) volatility process (see Cox et al. (1985)):

$$dX_t = \sigma_t dW_{1t}$$

$$d\sigma_t^2 = \kappa(v - \sigma_t^2)dt + s\sigma_t dW_{2t},$$

where W_{1t} and W_{2t} are independent Brownian motions. The parameters of the model are chosen to be empirically relevant and are based on a similar simulation setup in Aït-Sahalia and Yu (2009). Specifically, we choose $v = 0.1$, which in this model is the unconditional mean of σ_t^2. The mean reversion parameter κ is set to 5, the volatility of volatility s is set to 0.5. The initial value σ_0^2 is drawn from the CIR stationary distribution, which is Gamma$(2\kappa v/s^2, s^2/2\kappa)$. The total number of simulated samples is 10,000.

First, we consider the case where the price process is contaminated by the i.i.d. Gaussian noise component with the standard deviation (a) of 0.1%. We report the results for both FDQML and TDQML estimators in order to ascertain that their asymptotic properties are the same. Table 3.1 contains the summary statistics for the bias $\left(\widehat{\sigma^2} - \frac{1}{T}\int_0^T \sigma_t^2 dt\right)$ of the QML estimators at sampling frequencies ranging from 1 second to 3 minutes, where the latter term is evaluated using the discrete integral approximation

$$\frac{1}{N}\sum_{t=1}^{N}\sigma_t^2.$$

We can see that the two estimators produce very similar results, although FDQML still has a small bias even at the highest sampling frequency, whereas TDQML is unbiased. However, the bias is still negligible as it corresponds to about 0.5% of the mean of the integrated volatilities in this study. Standard deviations and root mean square errors (RMSE) are very similar, with TDQML having a 1-2% edge across all sampling frequencies. Figure 3.1 plots the distribution of the bias of the FDQML integrated volatility estimates, standardized using the variance expression in Conjecture 3.2, against the standard normal distribution. We see that it gets very close to normal as the sampling frequency increases. It can be seen from Figure 3. that the distribution of standardized estimates of the noise variance converges to normal much faster, as implied by theory. We do not show the corresponding

distributions for TDQML as they are virtually identical and would not be distinguishable on the graphs. Overall, the results provide support for Conjectures 3.1 and 3.2, namely, that in the case of i.i.d. microstructure noise the FDQML estimator shares the same asymptotic properties with the TDQML of Xiu (2010).

Second, we specify the microstructure noise as an AR(1) process with the coefficient of -0.7, keeping the standard deviation of the shock at 0.1%. This simulation design is motivated by the empirical findings of Aït-Sahalia et al. (2011), who estimate that log prices of some highly liquid stocks like Microsoft or Intel are characterized by negative AR noise. The results are reported in Table 3.2. We do not include the results for TDQML, as it is very difficult to specify the criterion function for this case, while the misspecified TDQML assuming i.i.d. noise performs very poorly. We observe that the bias increases dramatically for lower sampling frequencies (about 6% of the mean value of integrated volatility) compared to the i.i.d. case, but decreases substantially as the interval between observations shrinks , down to about 0.7% of the average true value. The RMSE decreases with the increase in the sample size, in line with the results observed for the i.i.d. noise case. The distribution of the integrated volatility estimate appears to follow a mixture of normals. However, standardizing the bias via a naive replacement of a^2 in the distribution in Conjecture 3.2 by the variance of the AR process confirms that the result derived under the i.i.d. noise assumption does not carry over to the model under consideration. The estimates of both the autoregressive coefficient and the variance of the shock converge fast to a normal distribution, similarly to the case with i.i.d. noise. Overall, we see that FDQML performs reasonably well under this simulation design, while using TDQML is not feasible due to analytical intractability.

Finally, we follow Engle and Sun (2007) by letting the noise component follow an

ARMA(1,1) process, with the AR and MA coefficients of 0.5 and 0.1 respectively and the same shock variance as in the previous case. We estimate integrated volatility by approximating the spectral density of the first differenced noise with that of a first differenced AR process with the order chosen by AIC. Table 3.3 contains the simulation results. The resulting order of the AR approximation chosen for all sampling frequencies was 1, perhaps due to the fact that the MA coefficient of the noise process is relatively small. Overall, the bias and RMSE performance is very similar to the case with AR(1) noise, with a slight edge in 1 second and 5 second frequencies. This demonstrates that, even when approximating an unknown stationary linear noise process with a finite order autoregression, the FDQML estimator delivers satisfactory results.

3.6 Conclusion

In this chapter we have suggested to use the FDQML estimator for integrated volatility estimation in the presence of market microstructure noise. The proposed estimator coincides with RV in the absence of noise, and we conjecture that it possesses the same asymptotic properties as its time domain counterpart studied in Xiu (2010) when i.i.d. noise is considered. Furthermore, we propose extending our estimator to accommodate microstructure noise that follows a stationary linear process by approximating its spectral density with an autoregression of finite order chosen by AIC. The simulation study appears to confirm our former conjecture and shows that the latter estimation procedure performs relatively well.

The above findings pave the road to deriving relevant theoretical results characterizing asymptotic properties of the FDQML estimator under the considered assumptions on the microstructure noise. More importantly, the frequency domain approach appears well suited for tackling the issue of endogenous noise. Such a modification will amount to

including an extra term in the likelihood that corresponds to the cross-spectrum of the latent price process and noise. Another important challenge is that of consistent estimation of the integrated quarticity within the FDQML framework for construction of confidence intervals.

3.7 Supplementary materials appendix 3

Table 3.1: Summary statistics for the bias of IV estimates under i.i.d. noise

	1 sec	5 sec	10 sec	20 sec	30 sec	1 min	3min
			FDQML				
Mean	-0.0509	-0.0507	-0.0478	-0.0474	-0.0493	-0.0436	-0.0460
Std. Dev.	0.5493	0.8317	1.0055	1.2488	1.4296	1.8106	2.6853
RMSE	0.5516	0.8332	1.0066	1.2496	1.4303	1.8111	2.6856
			TDQML				
Mean	0.0001	-0.0003	0.0020	0.0029	0.0002	0.0068	0.0014
Std. Dev.	0.5431	0.8265	0.9999	1.2403	1.4177	1.7955	2.6537
RMSE	0.5431	0.8264	0.9999	1.2402	1.4176	1.7954	2.6536

Note. Statistics computed for the bias of the indicated integrated volatility estimates and multiplied by 100.

Table 3.2: Summary statistics for the bias of IV estimates under AR(1) noise

	1 sec	5 sec	10 sec	20 sec	30 sec	1 min	3min
Mean	0.0717	0.0693	0.0379	0.0152	-0.0012	-0.1269	-0.6064
Std. Dev.	0.4386	0.9620	1.1109	1.4251	1.6585	2.2276	3.7207
RMSE	0.4444	0.9644	1.1115	1.4251	1.6585	2.2311	3.7696

Note. Statistics computed for the bias of the FDQML integrated volatility estimates and multiplied by 100.

Table 3.3: Summary statistics for the bias of IV estimates under ARMA(1,1) noise

	1 sec	5 sec	10 sec	20 sec	30 sec	1 min	3min
Mean	0.0561	0.0409	0.0379	0.0152	-0.0012	-0.1304	-1.2679
Std. Dev.	0.3981	0.8928	1.1109	1.4251	1.6585	2.2202	3.9699
RMSE	0.4020	0.8937	1.1115	1.4251	1.6585	2.2239	4.1672

Note. Statistics computed for the bias of the FDQML integrated volatility estimates and multiplied by 100.
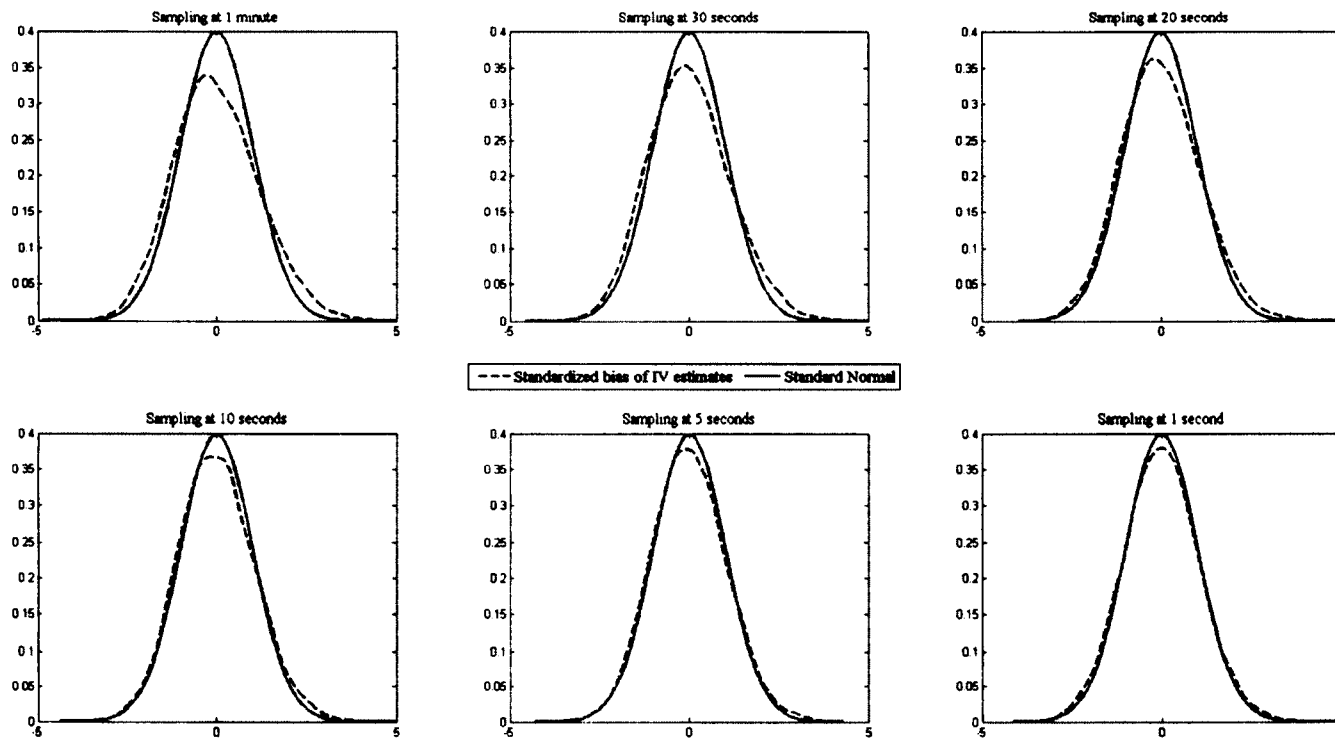
Figure 3·1: Distribution of the standardized bias of FDQML IV estimates
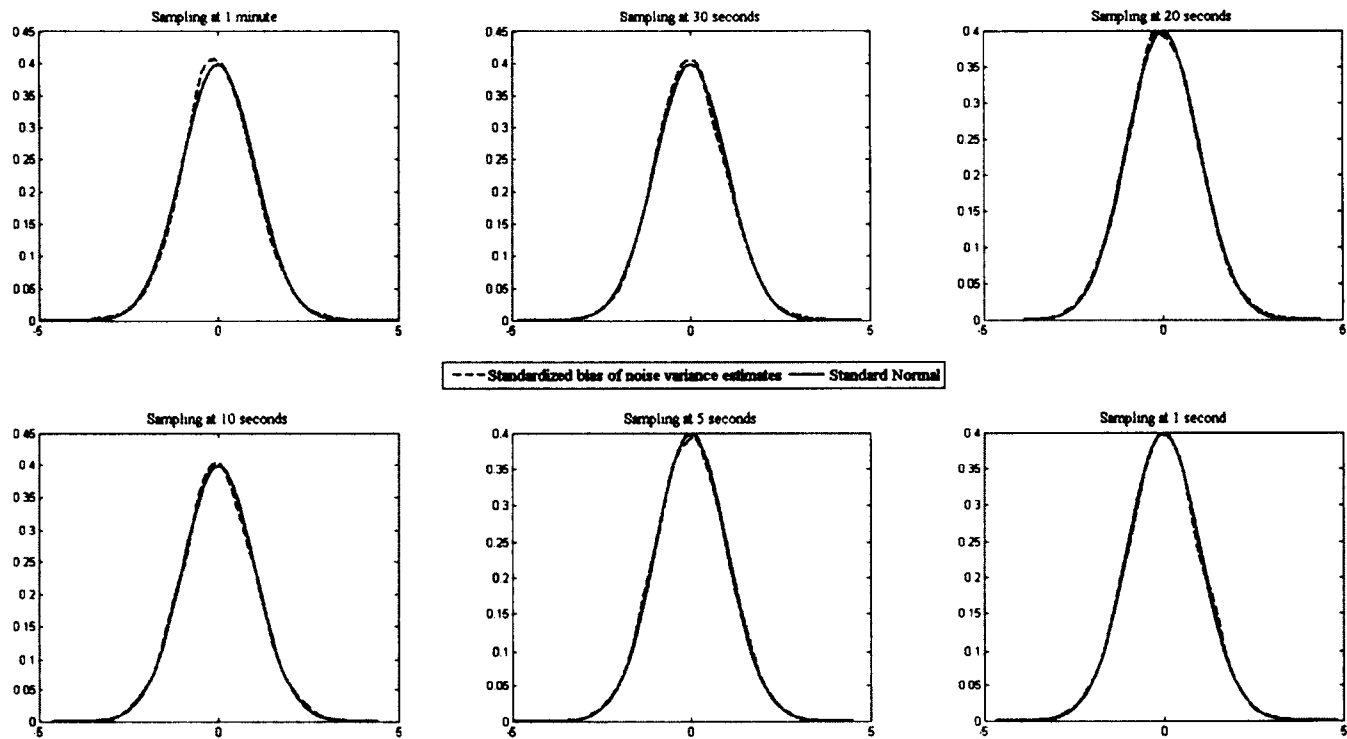
Figure 3·2: Distribution of the standardized bias of FDQML noise variance estimates

References

Y. Aït-Sahalia, P. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18:351–416, 2005.

Y. Aït-Sahalia, P. Mykland, and L. Zhang. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160:160–175, 2011.

Y. Aït-Sahalia and J. Yu. High frequency market microstructure noise estimates and liquidity measures. *The Annals of Applied Statistics*, 3:422–457, 2009.

H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, editors, *2nd International Symposium on Information Theory*, pages 267–281. Akademiai Kiado, 1973.

S. Altug. Time-to-build and aggregate fluctuations: Some new evidence. *International Economic Review*, 30:889–920, 1989.

S. An and F. Schorfheide. Bayesian analysis of DSGE models. *Econometric Reviews*, 26:113–172, 2007.

D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858, 1991.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Designing realized

kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76:1481–1536, 2008.

O. E. Barndorff-Nielsen and N. Shephard. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17:457–477, 2002.

K. N. Berk. Consistent autoregressive spectral estimates. *Annals of Statistics*, 2:489–502, 1974.

J. Berkowitz. Generalized spectral estimation of the consumption-based asset pricing model. *Journal of Econometrics*, 104:269–288, 2001.

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

P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. New York: Springer-Verlag, 2 edition, 1991.

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

V. Chernozhukov and H. Hong. An MCMC approach to classical estimation. *Journal of Econometrics*, 115:293–346, 2003.

L. J. Christiano, M. Eichenbaum, and D. Marshall. The permanent income hypothesis revisited. *Econometrica*, 59:397–423, 1991.

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

A. Consolo, C. A. Favero, and A. Paccagnini. On the statistical identification of DSGE models. *Journal of Econometrics*, 150:99–115, 2009.

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

W. Dunsmuir. 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, 1979.

W. Dunsmuir and E. J. Hannan. Vector linear time series models. *Advances in Applied Probability*, 8:339–364, 1976.

R. F. Engle. Band spectrum regression. *International Economic Review*, 15:1–11, 1974.

R. F. Engle and Z. Sun. When is noise not noise - a microstructure estimate of realized volatility. NYU Working Paper No FIN-07-047, 2007.

J. Fernández-Villaverde. The econometrics of DSGE models. *Journal of the Spanish Economic Association*, 1:3–49, 2010.

F. M. Fisher. *The Identification Problem in Economics*. New York: McGraw-Hill, 1966.

M. Fukaç, D. F. Waggoner, and T. Zha. Local and global identification of DSGE models: A simultaneous-equation approach. Working paper, 2007.

J. Gatheral and R. C. A. Oomen. Zero-intelligence realized variance estimation. *Finance and Stochastics*, 14:249–283, 2010.

A. Gloter and J. Jacod. Diffusions with measurement error: I. local asymptotic normality. *European Series in Applied and Industrial Mathematics*, 5:225–242, 2001.

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

L. P. Hansen and T. J. Sargent. Seasonality and approximation errors in rational expectations models. *Journal of Econometrics*, 55:21–55, 1993.

P. R. Hansen and A. Lunde. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24:127–161, 2006.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6:327–343, 1993.

Y. Hosoya and M. Taniguchi. A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, 10:132–153, 1982.

N. Iskrev. Local identification in DSGE models. *Journal of Monetary Economics*, 57:189–202, 2010.

R. G. King and M. W. Watson. Money, prices, interest rates and the business cycle. *The Review of Economics and Statistics*, 78:35–53, 1996.

I. Komunjer and S. Ng. Dynamic identification of DSGE models. *Econometrica*, 79:1995–2032, 2011.

S. G. Krantz and H. R. Parks. *The Implicit Function Theorem: History, Theory, and Applications*. Boston: Birkhäuser, 2002.

E. M. Leeper and C. A. Sims. Toward a modern macroeconomic model usable for policy analysis. In F. Stanley and J. J. Rotemberg, editors, *NBER Macroeconomics Annual*, pages 81–118. MIT Press, Cambridge, MA, 1994.

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, 2 edition, 1999.

P. Malliavin and M. Mancino. A Fourier transform method for nonparametric estimation of multivariate volatility. *Annals of Statistics*, 37:1983–2010, 2009.

P. Malliavin and M.E. Mancino. Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics*, 6:49–61, 2002.

M.E. Mancino and S. Sanfelici. Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Computational Statistics and Data Analysis*, 52:2966–2989, 2008.

A. McCloskey. Parameter estimation robust to low-frequency contamination with applications to ARMA, GARCH and stochastic volatility models. Working paper, Boston University, 2010.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

S. C. Olhede, A. M. Sykulski, and G. A. Pavliotis. Frequency domain estimation of integrated volatility for Itô processes in the presence of market-microstructure noise. *Multiscale Modeling and Simulation*, 8:393–427, 2009.

C. Otrok. On measuring the welfare cost of business cycles. *Journal of Monetary Economics*, 47:61–92, 2001.

M. B. Priestley. *Spectral Analysis and Time Series*, volume 1. New York: Academic Press, 1981.

Z. Qu. Inference and specification testing in DSGE models with possible weak identification. Working paper, Department of Economics, Boston University, 2011.

Z. Qu and D. Tkachenko. Identification and frequency domain QML estimation of linearized DSGE models. *Quantitative Economics*, 3:95–132, 2012.

G. Roberts, A. Gelman, and W. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, 7:110–120, 1997.

T. J. Rothenberg. Identification in parametric models. *Econometrica*, 39:577–591, 1971.

F. Schorfheide. Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics*, 15:645–670, 2000.

F. Schorfheide. Estimation and evaluation of DSGE models: Progress and challenges. NBER Working Paper 16781, 2011.

R. Shibata. An optimal autoregressive spectral estimate. *Annals of Statistics*, 9:300–306, 1981.

C. A. Sims. Rational expectations modeling with seasonally adjusted data. *Journal of Econometrics*, 55:9–19, 1993.

C. A. Sims. Solving linear rational expectations models. *Computational Economics*, 20:1–20, 2002.

F. Smets and R. Wouters. Shocks and frictions in US business cycles: A Bayesian DSGE approach. *The American Economic Review*, 97:586–606, 2007.

M. Taniguchi. On estimation of parameters of Gaussian stationary processes. *Journal of Applied Probability*, 16:575–591, 1979.

H. Uhlig. A toolkit for analyzing nonlinear dynamic stochastic models easily. In R. Marimon and A. Scott, editors, *Computational Methods for the Study of Dynamic Economies*. Oxford University Press, 1999.

A. Wald. Note on the identification of economic relations. In *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10. New York: John Wiley, 1950.

M. W. Watson. Measures of fit for calibrated models. *Journal of Political Economy*, 101:1011–1041, 1993.

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

D. Xiu. Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*, 159:235–250, 2010.

L. Zhang. Efficient estimation of stochastic volatility using noisy observations: a multiscale approach. *Bernoulli*, 12:1019–1043, 2006.

L. Zhang, P. Mykland, and Y. Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association*, 100:1394–1411, 2005.

B. Zhou. High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14:45–52, 1996.

Curriculum Vitae

DENIS TKACHENKO

Education

Ph.D., Economics, Boston University, 2012

Dissertation Title: *Frequency domain analysis of DSGE and stochastic volatility models*

Principal Advisor: Zhongjun Qu

M.Sc., Economics, Trinity College Dublin, 2006

B.A. (*first class honors*), Economic and Social Studies, Trinity College Dublin, 2005

Fields of Interest

Econometrics, Macroeconometrics, Financial Econometrics

Teaching Experience

Teaching Fellow, Advanced Econometrics II (Ph.D. Level), Boston University, Fall 2009, Fall 2010

Instructor, Introduction to Econometrics, Boston University, Spring 2010

Teaching Fellow, Introduction to Macroeconomics, Boston University, Fall 2008, Spring 2009

Fellowships, Honors and Awards

Special Research Fellowship, Boston University, Spring 2011, Fall 2012

Summer Research Grant, Boston University, 2010, 2011

Dean's Fellowship, Boston University, 2007-2012

Teaching Fellowship, Boston University, Fall 2008-Fall 2010

Postgraduate Research Award, Trinity College Dublin, 2006

Terence Gorman Prize, Trinity College Dublin, 2005

Foundation Scholarhip, Trinity College Dublin, 2003

Publications

"Identification and frequency domain QML estimation of linearized DSGE models"

(with Zhongjun Qu), *Quantitative Economics*, 3, 95-132, 2012.

Working papers

"Frequency domain analysis of medium scale DSGE models with application to Smets

and Wouters (2007)" (with Zhongjun Qu), January 2012, submitted.

"Frequency domain QML volatility estimation with noisy high frequency data",

March 2012

Work in Progress

"Realized volatility forecasting with filtered high frequency data, March 2012"

Invited Seminar and Conference Presentations

10th Annual *Advances in Econometrics* Conference "DSGE Models in Macroeconomics:

Estimation, Evaluation and New Developments", SMU, Dallas, November 2011

Western Economic Association International (WEAI) Graduate Student Workshop,

San-Diego, June 2011

Royal Economic Society (RES) Conference, London, April 2011

Computer Skills: MATLAB, Stata, R, EViews, LATEX, Scientific Workplace, Microsoft Office

Citizenship: Russia