Inference and Specification Testing in DSGE Models with Possible Weak Identification

Zhongjun Qu*

Boston University

April 22, 2011; This version: November 25, 2011

Abstract

This paper considers inference and model diagnostics for log-linearized DSGE models allowing an unknown subset of parameters to be weakly (including un-) identified. The framework allows for latent state variables, measurement errors and also permits analysis using only part of the spectrum, say at the business cycle frequencies. The latter is important because DSGE models are often designed to explain business cycle movements, not very long-run or very short-run fluctuations. For inference, we first characterize weak identification from a frequency domain perspective and propose a score test for the structural parameters based on the frequency domain maximum likelihood. The construction heavily exploits the structures of the DSGE solution, the score function and the information matrix. In particular, we show that the test statistic can be represented as the explained sum of squares from a complex-valued multivariate linear regression, where weak identification surfaces as (imperfectly) multicollinear regressors. Then, we prove that asymptotically valid inference can be carried out by inverting this test statistic and using Chi-square critical values. Next, we suggest procedures to construct uniform confidence bands for the impulse response function, the time path of the variance decomposition, the individual spectrum and the absolute coherency. For model diagnostics, we propose a family of frequency domain misspecification tests that are robust to weak identification. They can be used to test for misspecification in the mean, in the spectrum as well as misspecification within a band of frequencies. A simulation experiment using a calibrated model suggests that the tests have adequate size even in relatively small samples. It also suggests that it is possible to have informative confidence sets in DSGE models with unidentified parameters, particularly regarding the impulse responses functions. Although the paper focuses on DSGE models, the methods developed are potentially applicable to other dynamic models with well defined spectra, such as the stationary (factor-augmented) structural vector autoregression.

Keywords: Business cycle, frequency domain, impulse response, inference, model diagnostics, rational expectations models, weak identification.

---

*Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (qu@bu.edu). I thank Mingli Chen, Adam McCloskey, Marcelo Moreira, Serena Ng, Pierre Perron, Frank Schorfheide, seminar participates at Queen's, NC State, MIT/Harvard, Columbia and BU for comments and suggestions. The simulation analysis builds upon the GAUSS code of An and Schorfheide (2007). I thank the authors for making their code available.

1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models play an important role in quantitative macroeconomics, being considered both in academia and by policy institutions such as central banks. Inference and specification analysis for such models are often conducted in the time domain from a Bayesian perspective. Comprehensive reviews can be found in An and Schorfheide (2007), Fernández-Villaverde (2010) and in the textbooks by Canova (2007) and DeJong and Dave (2007).

Frequentist inference and specification testing in DSGE models is challenging. The likelihood surface can be flat or display near ridges in a large portion of the parameter space (Canova and Sala, 2009), mirroring the weak identification problem studied in the IV and GMM literature (Staiger and Stock, 1997 and Stock and Wright, 2000). For example, Del Negro and Schorfheide (2008) considered a New Keynesian DSGE model and showed that the data provides similar support for a model with moderate price and trivial wage rigidity and one in which both rigidities are high. In the extreme case, varying the structural parameter vector in certain directions may leave the likelihood unchanged, leading to so-called lack of identification. Such an example is provided in Qu and Tkachenko (2011), concerning the parameters of a Taylor rule equation. The above features imply that the conventional approach to inference and specification testing, which relies on a $\sqrt{T}$-convergent, asymptotically normal estimator, can be very inadequate. Robust inferential procedures have been proposed recently. In particular, Guerron-Quintana, Inoue and Kilian (2010) exploited the role played by the reduced form parameters and obtained confidence sets by inverting the likelihood ratio and the Bayes factor. Dufour, Khalaf and Kichian (2010) considered inverting moment based tests. In both cases, the analysis is conducted from a time domain perspective. In a related literature, Iskrev (2010), Komunjer and Ng (2011) and Qu and Tkachenko (2011) proposed rank conditions for checking whether the structural parameters are identified from the population. They did not consider weak identification.

The first goal of the paper is to obtain identification robust confidence sets for the structural parameters and confidence bands for impulse response functions from a frequency domain perspective, using a maximum likelihood (Whittle, 1951) approach. Working in the frequency domain offers several desirable features. First, the structure of the frequency domain likelihood, as well as that of the score and the information matrix, permits a simple and transparent testing procedure robust to weak identification. This simplicity also applies in the presence of latent state variables and measurement errors. Second, it allows for inference on dynamic parameters without specifying the steady state parameters or demeaning the data. This is useful when the misspecification of the steady states is a concern. Third, the researcher can choose desired frequencies for inference. This is valuable because DSGE models are designed to explain medium-term business cycle fluctuations, not very long-run or very short-run fluctuations, see Del Negro, Diebold and Schorfheide (2008). Also, measurement error and unmodeled seasonality may have greater effects at some frequencies than others, as discussed in various contexts by Engle (1974), Altug (1989) and Hansen and Sargent (1993).

We start by characterizing weak identification from a frequency domain perspective. The characterizing condition (see Assumption W) involves the eigenvalues of the information matrix, which converge to zero as the sample size approaches infinity. A subset of eigenvalues is allowed to be exactly zero, thus permitting some parameters to be unidentified for any sample size. As commented in Rothenberg (1971), the information matrix describes the local curvature of the log-likelihood. Modeling eigenvalues as local to zero is therefore a natural way to obtain a curvature that is small in some directions even when the sample size is large. The condition can also be viewed as a generalization of Corollary 1 in Qu and Tkachenko (2011), which shows that the structural parameters are identified from the population (i.e., with an infinite sample size) if and only if the information matrix has full rank. There, the eigenvalues are not sample size-dependent. In the appendices of this paper, we further illustrate the condition along two dimensions. Appendix A relates the condition to those used by Staiger and Stock (1997), Stock and Wright (2000) and Kleibergen (2005) and shows that their characterizations can also be stated in terms of particular eigenvalues approaching zero. In Appendix B, we consider a two equation dynamic model and show how the conditions on parameters map into the orders of the relevant eigenvalues. Prior to our work, Guerron-Quintana, Inoue and Kilian (2010) also suggested using the local curvature of the likelihood to characterize weak identification. The key differences are that here we work in the frequency domain and that we make no identifying assumptions about the reduced form parameters.

We then propose a test statistic for the structural parameters that is a quadratic form in the score vector and the information matrix. Two features underlie its robustness to weak identification. Consider inference on dynamic parameters. First, the information matrix depends only on the spectral density of the observables and its derivatives with respect to the dynamic parameter vector. Since the solution to log-linearized DSGE models follows a vector linear process, these quantities are known from the model without looking at the data. Consequently, the score vector is the only random element in the test statistic for any sample size. Second, the test statistic is related to the explained sum of squares in a complex-valued multivariate linear regression, where the regressors are again non-random and are governed by the derivatives mentioned above. If weak identification is present, the derivatives will be small in some parameter directions, implying that the regressors will be imperfectly multicollinear. In the extreme case, when some parameters are unidentified, the derivatives are zero in some directions and we have perfect multicollinearity. Overall, irrespective of the strength of identification, the rank of the regressors matrix is always bounded by the dimension of the dynamic parameter vector. Consequently, the test statistic has a Chi-square limiting distribution whose degrees of freedom is bounded by the dimension of the dynamic parameter vector. When steady state parameters are included in the analysis, similar results hold except that the information matrix also depends on the derivatives of the mean, which can again be easily computed from the model. The limiting distribution is still a Chi-square distribution with an increased degrees of freedom to account for the steady state parameters.

A confidence set for the full structural parameter vector can be constructed by inverting the proposed test statistic. Confidence sets for subvectors can be obtained using the projection method. If the dimension of the parameter vector is low, such sets can be constructed in a straightforward manner. If the dimension is high, we suggest to use a Metropolis type algorithm. It does not require maximizing the likelihood function. The main computational cost comes from evaluating the spectral density and its first order derivative over different values of the parameter vector.

Impulse response functions play a central role in assessing the implications of a DSGE model. Building on the confidence set for the full parameter vector, we propose a confidence band that covers this function asymptotically with desired probability even under weak or lack of identification. To the author's knowledge, this is the first time such robust confidence band for impulse responses is proposed in the DSGE literature. The same idea can be applied to construct confidence bands for the time path of the variance decomposition, the individual spectrum and the absolute coherency. It can also be used to examine certain low frequency hypotheses such as those considered in Sargent and Surico (2011). These examples showcase the empirical importance of the joint confidence set, whose value is sometimes underappreciated in the frequentist literature.

The second goal of the paper is to provide a procedure for model diagnostics allowing for weak identification. Specifically, we first propose a family of Kolmogorov-Smirnov type tests by generalizing results of Grenander and Rosenblatt (1957) and Bartlett (1955) to a multivariate setting. The tests are asymptotically pivotal under the null hypothesis of correct model specification and diverge to infinity under global alternatives. Then, we introduce a simple method to account for parameter uncertainty. The resulting procedure can be used to test for misspecification in the mean, the spectrum, as well as misspecification within a band of frequencies. The latter feature allows us to test the model's business cycle implications without first filtering the data. This is the first specification test with such a feature in the DSGE literature. The work here is related to Watson (1993) and King and Watson (1996). The former paper suggested plotting the model and data spectra as one of the most informative diagnostics, while the latter paper compared the spectra of three quantitative rational expectations models with that of the data. The results obtained here can be used to carry out similar analyses for DSGE models, allowing for weakly identified parameters.

Although the paper focuses on DSGE models, the methods are potentially applicable to other models with well defined spectra. We illustrate this using two examples, namely the stationary structural vector autoregression (SVAR) and the low dimensional factor augmented SVAR. These models can be represented as vector moving average processes, thus falling under the theoretical framework used in the paper. Similarly to the DSGE case, the methods developed here can be used to construct identification robust confidence sets for the structural parameter vector, to study the impulse response function and to test the specification of the model at a particular band of frequencies. Note that we work directly with the structural parameters, thus can avoid making identifying assumptions about the reduced form parameter vector in the case of factor augmented SVAR.

We evaluate the finite sample properties of the proposed tests using a model studied in An and Schorfheide (2007). Three cases are considered: analysis based on (a) the business cycle frequencies, (b) the full spectrum and (c) the mean and the full spectrum. The tests show decent sizes in all three cases even with relatively small sample sizes. A power comparison suggests that Case (a) involves a loss of power when compared with (b) and (c). However, it can still be informative. In practice, this offers the researcher a choice. If one firmly believes that the model is reasonably specified at all frequencies, then the full spectrum should be used and the inference can be more efficient. If instead one suspects that modeling of the trend, or, more generally, of the very low frequency features is inconsistent with the data (for example, the data has a broken trend while the model has a linear trend), then using a subset of frequencies is more preferable.

We examine the confidence intervals and error bands for impulse responses using the same model and specifications. The results show that even unidentified parameters themselves can have tight confidence intervals. This appears to be a new finding in the DSGE literature. The confidence bands for impulse response functions can also be narrow with unidentified parameters. Intuitively, because observationally equivalent parameter values may generate the same set of response functions, uncertainty about the former does not necessarily translate into uncertainty about the latter.

From a methodological perspective, the current paper is related to recent work on inference under weak identification in a GMM setting. Kleibergen and Mavroeidis (2009a) provided a review of the related literature. The methods discussed there can be used to analyze a subset of equations/variables in a DSGE model while being agnostic about, and therefore not using information from, the rest of the equations/variables. The current paper is useful for studying the full system. The latter framework is necessary if the issues of interests are the impulse response function, the variance decomposition, forecasting, counterfactual policy analysis and other objects that require fully specifying the system dynamics. Thus, the methods complement each other. The paper is also related to a relatively small literature that exploits the advantage of estimation and diagnosis of dynamic equilibrium models in the spectral domain. Altug (1989) applied the frequency domain likelihood to estimate models with additive measurement errors. Hansen and Sargent (1993) considered the effect of seasonal adjustment on estimation. Diebold, Ohanian and Berkowitz (1998) discussed a general framework for loss-function based estimation and model evaluation. Christiano and Vigfusson (2003) applied frequency domain likelihood to study a model with time-to-plan in the investment technology. Del Negro, Diebold and Schorfheide (2008) emphasized that the misspecification of DSGE models is more prevalent at some frequencies than at others. They developed a framework in which DSGE models are used to derive restrictions for vector autoregressions, but only over selected frequencies of interest. To our knowledge, the current paper is the first to study identification robust inference and specification testing from the frequency domain perspective.

The paper is structured as follows. Section 2 sets the stage for the analysis by illustrating how to compute the spectrum of a DSGE model. It also discusses how to analyze models with latent state variables and measurement errors and how to base inference on a subset of frequencies rather than the full spectrum. Section 3 presents the theoretical framework and the assumptions. Section 4 characterizes weak identification from a frequency domain perspective. Section 5 proposes the score tests, examines their properties and discusses how to obtain confidence sets. It also considers uniform confidence bands for impulse response functions and some other objects. Section 6 proposes a family of frequency domain misspecification tests and studies their asymptotic properties. Section 7 illustrates the applicability of the methods to other dynamic models. Section 8 includes Monte Carlo experiments and Section 9 concludes. The paper contains three appendices with the proofs included in Appendix C.

The following notation is used. $|\cdot|$ returns the modulus of a complex number; the imaginary unit is denoted by i. $||x||$ is the Euclidean norm of a vector x, i.e., $||x|| = (|x_1|^2 + |x_2|^2 + ... + |x_n|^2)^{1/2}$. $||X||$ is the vector induced norm for a complex-valued matrix X. Its Moore-Penrose Pseudoinverse is denoted by X^+. X^* stands for its conjugate transpose, i.e., $X^* = \overline{X'}$, where X' gives the conventional transpose. If X is a square matrix, then $\dim(X)$ returns its row or column dimension and eig(X) returns its eigenvalues as a vector. If $f_\theta \in R^k$ is a differentiable function of $\theta \in R^p$, then $\partial f_{\theta_0}/\partial\theta'$ is a k-by-p matrix of partial derivatives evaluated at θ_0. "$\rightarrow^p$" and "$\rightarrow^d$" signify convergence in probability and in distribution. And $O_p(\cdot)$ and $o_p(\cdot)$ are the usual symbols for stochastic orders of magnitude.

2 Preliminaries: the spectrum of a log-linearized DSGE model

Suppose a discrete-time DSGE model has been log-linearized around the steady state. Assume that the solution exists and is unique. Then, it can be computed and represented in a variety of ways using the algorithms of Anderson and Moore (1985), Uhlig (1999), Klein (2000), King and Watson (2002) and Sims (2002). The methods proposed in this paper can work with any of these representations because they all permit relatively straightforward computation of the spectrum.

Given that the spectral density matrix plays an important role in the analysis, below we illustrate how to compute it from the output of Sims (2002) and Uhlig (1999). We also illustrate how to analyze models with latent state variables and/or measurement errors and how to approach inference and specification testing using a subset of frequencies rather than the full spectrum. As a matter of notation, let θ denote the dynamic parameter vector, i.e., the structural parameters entering the log-linearized DSGE system.

Computing the spectral density matrix using Sims' (2002) method. Sims considered the following representation for a log-linearized system:

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

where S_t is a vector that includes endogenous variables (both observed and latent), conditional expectation terms and processes of exogenous shocks if they are serially correlated, ϵ_t is a vector of serially uncorrelated structural shocks, and η_t is a vector of expectation errors. The elements of the coefficients matrices are known deterministic functions of θ. Models with more lags or with lagged expectations can be accommodated by expanding S_t accordingly. Then, under determinacy (Sims, 2002, p. 12), the solution is given by

$$S_t = \Phi_1(\theta)S_{t-1} + \Phi_0(\theta)\epsilon_t, \tag{1}$$

where the coefficients matrices $\Phi_1(\theta)$ and $\Phi_0(\theta)$ are again deterministic functions of θ. Since the solution is stable, the preceding display can be equivalently represented as

$$S_t = (I - \Phi_1(\theta)L)^{-1}\Phi_0(\theta)\epsilon_t.$$

Some elements of S_t, such as the conditional expectations, are unobservable to the econometrician and therefore can not be used directly for inference. Let $A(L)$ be a matrix of finite order lag polynomials to select the observable, i.e.,

$$Y_t^d = A(L)S_t = A(L)(I - \Phi_1(\theta)L)^{-1}\Phi_0(\theta)\epsilon_t. \tag{2}$$

Then, the spectral density of Y_t^d is given by

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

where $\Sigma(\theta) = Var(\epsilon_t)$ and

$$H(L;\theta) = A(L)(I - \Phi_1(\theta)L)^{-1}\Phi_0(\theta). \tag{4}$$

Remark 1 (Latent endogenous variable) *The selection matrix $A(L)$ permits substantial flexibility. To illustrate this, suppose S_t includes two 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 models with latent endogenous variables such as those in Ireland (2007) and Smets and Wouters (2003, 2007). In the former, the monetary policy rule depends on a latent inflation target, π_t^*, which appears as an element of S_t in (1). In the latter, S_t includes variables from a frictionless economy unobservable to the econometrician. In these two models, the spectral density matrix of the observables can be computed by simply assigning zero entries in $A(L)$ to exclude π_t^* or variables from the flexible price economy.*

Remark 2 (Measurement error) *Suppose the data are generated by $Y_t^d + \zeta_t$, where ζ_t is a vector of measurement errors independent of Y_t^d with spectral density $f_\theta^m(\omega)$. Then, the spectral density matrix of the observed random vector is*

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

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

Remark 3 (Inference and specification testing using a subset of frequencies) *We may be interested in conducting inference using a subset of frequencies, say those corresponding to business cycle fluctuations (with periods of 6–32 quarters, see King and Watson, 1996). Then, with quarterly data, we only need to compute (3) or (5) for*

$$\omega \in [\pi/16, \pi/3] \cup [5\pi/3, 31\pi/16].$$

Note that the second interval is present because the likelihood function depends on frequencies between $[\pi, 2\pi]$ *in addition to* $[0, \pi]$*. For notional purposes, we define an indictor function,* $W(\omega)$*, to select such frequencies. For example, to select business cycle frequencies, we simply let* $W(\omega) = 1$ *if* ω *belongs to the preceding set and 0 otherwise. The analysis then proceeds in the same way as for the full spectrum case.*

Computing the spectral density matrix using Uhlig's (1999) solution. Uhlig (1999) provided the following representation for the solution of a DSGE model:

$$\begin{aligned} x_t &= P(\theta)x_{t-1} + Q(\theta)z_t \\ y_t &= R(\theta)x_{t-1} + S(\theta)z_t \\ z_t &= N(\theta)z_{t-1} + \epsilon_t, \end{aligned} \tag{6}$$

where x_{t-1} is a vector of endogenous (state) variables, y_t is a vector of endogenous (jump) variables and z_t is a vector process of exogenous shocks. Let $S_t = (x_t', y_t')'$. Then, from (6):

$$S_t = C(L;\theta)\epsilon_t \quad \text{with } C(L;\theta) = \begin{pmatrix} I - P(\theta)L & 0 \\ -R(\theta)L & I \end{pmatrix}^{-1} \begin{pmatrix} Q(\theta) \\ S(\theta) \end{pmatrix} [I - N(\theta)L]^{-1}.$$

Let $A(L)$ be a selection matrix such that

$$Y_t^d = A(L)S_t = A(L)C(L;\theta)\epsilon_t. \tag{7}$$

Then, the spectral density of Y_t^d is as in (3) with

$$H(L;\theta) = A(L)C(L;\theta).$$

The computation of $f_\theta(\omega)$ is again straightforward given the solution of the model. The issues of measurement errors, latent state variables and inference using a subset of frequencies can be handled in the same way as before. The details are omitted.

3 Setup and assumptions

Let $\{Y_1, Y_2, ..., Y_T\}$ be a sample of random vectors. Define an augmented parameter vector

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

where θ still denotes the dynamic parameter vector, affecting the log-linearized solution and the measurement error process, and γ contains structural parameters affecting only the model's steady state. Throughout the paper, assume there exists some (not necessarily unique) parameter values θ_0 and $\bar{\theta}_0$, such that Y_t is related to the DSGE solution via

$$\textbf{DGP 1}: \;\; Y_t = \mu(\bar{\theta}_0) + Y_t^d(\theta_0),$$

or

$$\textbf{DGP 2}: \;\; Y_t = \mu(\bar{\theta}_0) + Y_t^d(\theta_0) + \zeta_t(\theta_0), \tag{8}$$

where $Y_t^d(\theta_0)$ denotes[1] the log-linearized solution (2) when $\theta = \theta_0$, $\mu(\bar{\theta}_0)$ is the mean of Y_t implied by the model's steady state and $\zeta_t(\theta_0)$ is a vector of measurement errors.

Assumption 1. $\theta_0 \in \Theta \subset \mathbb{R}^q$ and $\bar{\theta}_0 \in \bar{\Theta} \subset \mathbb{R}^{p+q}$ with Θ and $\bar{\Theta}$ being compact sets.

Assumption 2. The log-linearized solution is unique and is representable as

$$Y_t^d(\theta) = H(L;\theta)\epsilon_t(\theta) \quad \text{with} \quad H(L;\theta) = \sum_{j=0}^{\infty} h_j(\theta)L^j, \tag{9}$$

where $h_j(\theta)$ $(j = 0, ..., \infty)$ are real valued matrices and $\epsilon_t(\theta)$ are serially uncorrelated structural shocks with a nonsingular covariance matrix $\Sigma(\theta)$. In DGP 2, $\zeta_t(\theta)$ are serially uncorrelated with a nonsingular covariance matrix $\Sigma_\zeta(\theta)$. $\mathbb{E}(\zeta_t(\theta)\epsilon_s'(\theta)) = 0$ for all t and s.

Assumption 3. There exist finite positive constants C_L and C_U such that for all $\omega \in [-\pi, \pi]$ and all $\theta \in \Theta$:

(i) $C_L \leq \mathrm{eig}(f_\theta(\omega)) \leq C_U$;

(ii) the elements of $f_\theta(\omega)$ belong to the Lipschitz class of degree $\beta > 1/2$ with respect to ω;[2]

(iii) $||\partial \,\mathrm{vec}\, f_\theta(\omega)/\partial\theta'|| \leq C_U$ and the elements of $\partial \,\mathrm{vec}\, f_\theta(\omega)/\partial\theta$ belong to the Lipschitz class of degree $\beta > 1/2$ with respect to ω;

(iv) $||\partial\mu(\bar{\theta})/\partial\bar{\theta}'|| \leq C_U$ over $\bar{\theta} \in \bar{\Theta}$.

Assumption 4. $\{Y_t\}_{t=1}^T$ is a sequence of multivariate normal random vectors.

Assumption 1 imposes restrictions on the parameter space. The boundedness assumption is unrestrictive as economic theory often provides natural bounds on DSGE parameters. The requirement for closedness is to ensure that the procedure for computing the confidence sets, which involves searching over the parameter space, is well defined.

[1] From now on, we write $Y_t^d(\theta_0)$ instead of Y_t^d to emphasize its dependence on the parameter value.

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

Assumption 2 concerns the properties of the log-linearized solution. It is formulated to allow for different representations discussed in the previous section, c.f. (2) and (7). Note that the dimensions of the relevant variables and parameters are:

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

Under this Assumption, Y_t follows a vector moving average process. For DGP1, this is obvious. For DGP 2, simply define an augmented shock process $\epsilon_t^a(\theta) = (\epsilon_t(\theta)', \zeta_t(\theta)')'$ and let $H^a(L;\theta) = \sum_{j=0}^{\infty} h_j^a(\theta)L^j$ with $h_0^a(\theta)$=$[h_0(\theta), I_{n_Y}]$ and $h_j^a(\theta) = [h_j(\theta), 0_{n_Y}]$ for $j > 0$, where I_{n_Y} and 0_{n_Y} are n_Y dimensional identity and zero matrices respectively. This feature implies that we only need to prove results under DGP1. They will automatically hold under DGP 2 by replacing $H(L;\theta)$ with $H^a(L;\theta)$. This assumption can be further relaxed to allow $\zeta_t(\theta)$ to be serially correlated. The analysis is the same as long as it remains independent of $Y_t^d(\theta)$.

Assumption 3 imposes some restrictions on the mean and spectrum of Y_t. Part (i) assumes that the spectral density matrix is finite and positive definite. If unit roots are present in the DSGE model, then it requires that the series have been appropriately differenced prior to applying the methods. The positive definiteness requires augmenting the system with measurement errors if $n_\epsilon < n_Y$, i.e., DGP 2 applies. Parts (ii)-(iii) assume that the spectral density and its first derivatives are smooth in ω, which are commonly employed when estimating vector linear processes, see Dunsmuir and Hannan (1976, p.362). They can be verified under more primitive conditions. Specifically, Part (ii) is satisfied if $\sum_{j=0}^{\infty} j^\beta ||h_j(\theta)|| \leq \infty$ (see Hannan, 1970, p. 311-312). Applying Sims' representation (2) and assuming without loss of generality $A(L) = A$, then $h_j(\theta) = A\Phi_1^j(\theta)\Phi_0(\theta)$. Provided that all the eigenvalues of $\Phi_1(\theta)$ are inside the unit circle (i.e., no unit roots), $h_j(\theta)$ will be finite and decay exponentially and the condition is satisfied. Part (iii) is satisfied if $\sum_{j=0}^{\infty} j^\beta ||\partial \operatorname{vec} h_j(\theta)/\partial\theta'|| \leq \infty$, which holds if, in addition to the eigenvalues of $\Phi_1(\theta)$ being inside the unit circle, we have $||\partial \operatorname{vec} \Phi_1(\theta)/\partial\theta'|| \leq M$ and $||\partial \operatorname{vec} \Phi_0(\theta)/\partial\theta'|| \leq M$ for some $M > 0$ and all $\theta \in \Theta$.

Assumption 4 requires normality. Without it, the test statistics proposed later will depend on nuisance parameters. Note that this assumption is common in DSGE literature. Because the data are often measured at quarterly frequency, the effect of non-Gaussianity on inference is expected to be mild.

In what follows, we will consider inference on θ_0 based on the spectrum alone, and on $\bar{\theta}_0$ based jointly on the mean and spectrum.

4 Weak identification from a frequency domain perspective

Weak identification reflects both the model structure and the criterion function used for inference. The inference here is based on the frequency domain maximum likelihood. We therefore start with a brief review of the basic ideas underlying it.

4.1 The frequency domain maximum likelihood

Let ω_j denote the Fourier frequencies, i.e., $\omega_j = 2\pi j/T$ $(j = 1, 2, ..., T-1)$. The discrete Fourier transforms of Y_t are

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

Asymptotically, $w_T(\omega_j)$ has a complex-valued multivariate normal distribution with density (see Hannan (1970), p. 223-225)

$$\frac{1}{\pi^{n_Y} \det(f_{\theta_0}(\omega_j))} \exp\left[-\mathrm{tr}\left\{f_{\theta_0}^{-1}(\omega_j) w_T(\omega_j) w_T(\omega_j)^*\right\}\right].$$

Because $w_T(\omega_j)$ and $w_T(\omega_k)$ are asymptotically independent if $\omega_j \neq \omega_k$ and $\omega_j + \omega_k \neq 2\pi$, an approximate log-likelihood function for θ based on $\{Y_t\}_{t=1}^{T}$, up to a constant, is

$$-\sum_{j=1}^{T-1} \left[\log\det(f_\theta(\omega_j)) + \mathrm{tr}\left\{f_\theta^{-1}(\omega_j) I_T(\omega_j)\right\}\right], \tag{10}$$

where

$$I_T(\omega_j) = w_T(\omega_j) w_T(\omega_j)^*.$$

The criterion function (10) has been extensively studied in the statistics literature, mainly in the context of estimating (vector) linear processes, see Dunsmuir and Hannan (1976) and Dunsmuir (1979).

Let $W(\omega_j)$ be an indicator function to select the desired frequency components. In this paper, we consider the following generalized version of (10):

$$L_T(\theta) = -\sum_{j=1}^{T-1} W(\omega_j) \left[\log\det(f_\theta(\omega_j)) + \mathrm{tr}\left\{f_\theta^{-1}(\omega_j) I_T(\omega_j)\right\}\right]. \tag{11}$$

$L_T(\theta)$ allows us to conduct inference on θ without reference to γ. On the other hand, it is also simple to include steady state parameters into the analysis. Define

$$w_{\bar{\theta},T}(0) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \left(Y_t - \mu(\bar{\theta})\right) \text{ and } I_{\bar{\theta},T}(0) = w_{\bar{\theta},T}(0) w_{\bar{\theta},T}(0)^*.$$

Because $w_{\bar{\theta}_0,T}(0)$ is asymptotically distributed as $N(0, f_{\theta_0}(0))$ and independent of $w_T(\omega_j)$ ($j = 1, 2, ..., T-1$), we obtain the following approximate log-likelihood function for $\bar{\theta}$:

$$\bar{L}_T(\bar{\theta}) = L_T(\theta) - \left[\log\det(f_\theta(0)) + \operatorname{tr}\left\{f_\theta^{-1}(0) I_{\bar{\theta},T}(0)\right\}\right]. \tag{12}$$

Hansen and Sargent (1993) derived (12) as an approximation to the time domain Gaussian likelihood and used it to understand the effect of seasonal adjustment on estimation. Qu and Tkachenko (2011) established the asymptotic properties of its maximizer, under the assumption that the parameters are strongly identified. Neither paper allowed for weak identification.

4.2 Weak identification

We characterize weak identification from a frequency domain perspective. The characterizing conditions (Assumptions W and W2 below) are motivated by Rothenberg (1971) and Qu and Tkachenko (2011) and stated using the eigenvalues of the information matrix. We allow a subset of eigenvalues to converge to zero as $T \to \infty$, such that the local curvature of the likelihood remains small in some directions, even with a large sample size.

First, consider inference on θ based on the spectrum. The score function of (11), normalized by $2\pi T^{-1/2}$, is given by

$$D_T(\theta_0) = 2\pi T^{-1/2} \sum_{j=1}^{T-1} W(\omega_j) \left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'}\right)^* (f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j)) \operatorname{vec}(I_T(\omega_j) - f_{\theta_0}(\omega_j)). \tag{13}$$

The information matrix is

$$M_T(\theta_0) = 8\pi^2 T^{-1} \sum_{j=1}^{T-1} W(\omega_j) \left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'}\right)^* \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j)\right) \frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'}.$$

Here, the information matrix has a simple expression because, although $\{Y_t\}_{t=1}^T$ can have a complex dependence structure, their Fourier transforms are asymptotically independent with known variances (see Hannan, 1970, p. 249, Corollary 1). Because $M_T(\theta_0)$ is real, symmetric and positive semi-definite, its eigen-decomposition always exists:

$$M_T(\theta_0) = Q_T(\theta_0)\Lambda_T(\theta_0)Q_T(\theta_0)^{-1}, \tag{14}$$

where the columns of $Q_T(\theta_0)$ are the orthonormal eigenvectors and $\Lambda_T(\theta_0)$ contains the eigenvalues in a non-increasing order. Partition $\Lambda_T(\theta_0)$ as

$$\Lambda_T(\theta_0) = \begin{bmatrix} \Lambda_{1T}(\theta_0) & 0 & 0 \\ 0 & \Lambda_{2T}(\theta_0) & 0 \\ 0 & 0 & \Lambda_{3T}(\theta_0) \end{bmatrix},$$

where $\Lambda_{1T}(\theta_0)$, $\Lambda_{2T}(\theta_0)$ and $\Lambda_{3T}(\theta_0)$ are q_1, q_2 and q_3 dimensional diagonal matrices.

Assumption W. (i) The diagonal elements of $T\Lambda_{1T}(\theta_0)$ diverge to ∞; (ii) The diagonal elements of $T\Lambda_{2T}(\theta_0)$ converge to positive constants; (iii) $\Lambda_{3T}(\theta_0) = 0$ for any T; (iv) The elements of

$$\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'} Q_T(\theta_0)\Lambda_T^+(\theta_0)^{1/2}$$

are finite and belong to the Lipschitz class of degree $\beta > 1/2$ with respect to $\omega \in [-\pi, \pi]$, where $\Lambda_T^+(\theta_0)^{1/2}$ is the square root of the Moore-Penrose Pseudoinverse of $\Lambda_T(\theta_0)$.

W(i) to W(iii) are formulated to allow for different degrees of identification. $\Lambda_{1T}(\theta_0)$ corresponds to parameter directions that are strongly or semi-strongly identified (the latter notion follows Andrews and Cheng, 2011). $\Lambda_{2T}(\theta_0)$ corresponds to directions that are weakly identified, while $\Lambda_{3T}(\theta_0)$ corresponds to directions that are unidentified for any sample sizes. The condition is related to Corollary 1 in Qu and Tkachenko (2011), which shows that the structural parameters are identified from the population (i.e., with an infinite sample size) if and only if the information matrix has full rank.[3] Here, the eigenvalues are sample size-dependent, therefore identification is no longer a zero-one phenomenon.

W(i)-W(iii) are closely related to the characterizing conditions used in the IV and GMM literature (Staiger and Stock, 1997, Stock and Wright, 2000 and Kleibergen, 2005). We illustrate this along two dimensions. Appendix A shows that the latter conditions can also be stated in terms of the eigenvalues that measure the local curvature of the corresponding criterion functions. Appendix B considers a two-equation dynamic model in which the score and information matrix can be computed analytically. It shows that the conditions in Staiger and Stock (1997) translate into W(i) to W(iii). Prior to our work, Guerron-Quintana, Inoue and Kilian (2010) also suggested using the local curvature of the likelihood to characterize weak identification. The key differences are that here we work in the frequency domain and that we make no identifying assumptions about the reduced form parameters.

W(iv) strengthens Assumption 3(iii) by requiring sufficient smoothness of $\partial \operatorname{vec} f_{\theta_0}(\omega)/\partial\theta'$ in ω. Because $Q_T(\theta_0)$ is a rotation matrix[4], its effect is to map the row vectors of $\partial \operatorname{vec} f_{\theta_0}(\omega)/\partial\theta'$ into a new coordinate system common to all $\omega \in [-\pi, \pi]$. The assumption therefore requires $\partial \operatorname{vec} f_{\theta_0}(\omega)/\partial\theta'$ to be well behaved in this new coordinate system. If θ_0 is strongly identified, then $\Lambda_T(\theta_0)$ is finite and positive definite and the assumption is trivially satisfied. Under weak identifi-

[3] They allow $f_{\theta_0}(\omega)$ to be singular. In the nonsingular case, the above statement applies.

[4] A rotation matrix is an orthogonal matrix with determinant +1. $Q_T(\theta_0)$ is an orthogonal matrix by the symmetry of $M_T(\theta_0)$. Its eigenvectors can always be chosen such that its determinant equals +1 instead of -1.

cation, it is less transparent because some entries in $\Lambda_T^+(\theta_0)^{1/2}$ diverge to infinity. In Appendix B, we illustrate this assumption using a simple dynamic model and show that it is satisfied.

The above idea extends immediately to inference on $\bar{\theta}$ based on the mean and the spectrum. The score, normalized by $2\pi T^{-1/2}$, is given by

$$\bar{D}_T\left(\bar{\theta}_0\right) = 2\pi T^{-1/2}\sum_{j=0}^{T-1} W(\omega_j)\left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'}\right)^* \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j)\right) \operatorname{vec}\left(I_T\left(\omega_j\right) - f_{\theta_0}(\omega_j)\right)$$
$$+ 2T^{-1/2}\sum_{t=1}^{T} \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}} f_{\theta_0}^{-1}(0)(Y_t - \mu(\bar{\theta}_0)), \tag{15}$$

where the first summation starts at $j=0$ instead of $j=1$ and $I_T\left(0\right)=I_{\bar{\theta}_0,T}\left(0\right)$. The information matrix is

$$\begin{aligned}\bar{M}_T(\bar{\theta}_0) \quad &= \quad 8\pi^2 T^{-1}\sum_{j=0}^{T-1} W(\omega_j)\left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'}\right)^* \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j)\right) \frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'} \\ &\quad + 8\pi \frac{\partial \mu(\bar{\theta}_0)'}{\partial \bar{\theta}} f_{\theta_0}^{-1}(0) \frac{\partial \mu(\bar{\theta}_0)}{\partial \bar{\theta}'}.\end{aligned}$$

Write $\bar{M}_T(\bar{\theta}_0) = \bar{Q}_T(\bar{\theta}_0)\bar{\Lambda}_T(\bar{\theta}_0)\bar{Q}_T(\bar{\theta}_0)^{-1}$ and partition $\bar{\Lambda}_T(\bar{\theta}_0)$ as

$$\operatorname{Diag}(\bar{\Lambda}_T(\bar{\theta}_0)) = \operatorname{Diag}(\bar{\Lambda}_{1T}(\bar{\theta}_0), \bar{\Lambda}_{2T}(\bar{\theta}_0), \bar{\Lambda}_{3T}(\bar{\theta}_0)).$$

Assumption W2. (i) The diagonal elements of $T\bar{\Lambda}_{1T}(\bar{\theta}_0)$ diverge to ∞. (ii) The diagonal elements of $T\bar{\Lambda}_{2T}(\bar{\theta}_0)$ converge to positive constants. (iii) $\bar{\Lambda}_{3T}(\bar{\theta}_0) = 0$ for any T. (iv) The elements of

$$\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \bar{\theta}'} \bar{Q}_T(\bar{\theta}_0)\bar{\Lambda}_T^+(\bar{\theta}_0)^{1/2}$$

are finite and belong to the Lipschitz class of degree $\beta > 1/2$ with respect to $\omega \in [-\pi, \pi]$, where $\bar{\Lambda}_T^+(\bar{\theta}_0)^{1/2}$ denotes the square root of the Moore-Penrose Pseudoinverse of $\bar{\Lambda}_T(\bar{\theta}_0)$.

5 Confidence sets robust to weak identification

We first propose two score tests, illustrate their properties under weak identification and derive their limiting distributions.

5.1 Score tests based on the frequency domain likelihood

First, consider inference on θ using the spectrum. Define

$$S_T(\theta_0) = D_T\left(\theta_0\right)' M_T^+(\theta_0) D_T\left(\theta_0\right)$$

where $M_T^+(\theta)$ denotes the Moore-Penrose Pseudoinverse of $M_T(\theta)$. Next, consider inference on $\bar{\theta}$ using both the mean and the spectrum. Define

$$\bar{S}_T(\bar{\theta}_0) = \bar{D}_T\left(\bar{\theta}_0\right)' \bar{M}_T^+(\bar{\theta}_0)\bar{D}_T\left(\bar{\theta}_0\right),$$

where $\bar{M}_T^+(\bar{\theta})$ denotes the Moore-Penrose Pseudoinverse of $\bar{M}_T(\bar{\theta})$.

To better understand the properties of $S_T(\theta_0)$ and $\bar{S}_T(\bar{\theta}_0)$ under weak identification, we show that they can be given regression interpretations. Specifically, consider the following complex-valued multivariate linear regression:

$$\mathcal{Y}_j = X_j\beta + U_j, \quad (j = 0, 1, ..., T-1), \tag{16}$$

where $\mathcal{Y}_j$ is a vector and X_j is a matrix, whose values are specified below, β is an unknown parameter vector and U_j is a vector of regression errors. The least square estimator of β is

$$\hat{\beta} = \left(\sum_{j=0}^{T-1} X_j^* X_j\right)^+ \left(\sum_{j=0}^{T-1} X_j^* \mathcal{Y}_j\right)$$

and the explained sum of squares is

$$ESS = \sum_{j=0}^{T-1} \widehat{\mathcal{Y}}_j^* \widehat{\mathcal{Y}}_j = \left(\sum_{j=0}^{T-1} \mathcal{Y}_j^* X_j\right) \left(\sum_{j=0}^{T-1} X_j^* X_j\right)^+ \left(\sum_{j=0}^{T-1} X_j^* \mathcal{Y}_j\right). \tag{17}$$

When there is perfect multicollinearity, $\hat{\beta}$ is not unique but ESS will be. To establish the relation between ESS and the test statistics, we need the following notation. Let H be a generic positive definite Hermitian matrix (e.g., $H = f_{\theta_0}(\omega)$). Then, H admits the following eigen-decomposition (Horn and Johnson, 2005, Theorem 4.1.5): $H = U\Lambda U^*$, where Λ is a real-valued diagonal matrix and $UU^* = U^*U = I$. Define $H^{1/2} = U\Lambda^{1/2}U^*$ and $H^{-1/2} = U\Lambda^{-1/2}U^*$. Then both $H^{1/2}$ and $H^{-1/2}$ are Hermitian by construction. Note that $H' \otimes H$, $(H^{1/2})' \otimes H^{1/2}$ and $(H^{-1/2})' \otimes H^{-1/2}$ are also Hermitian (see Horn and Johnson, 2006, p. 243).

Lemma 1 *The statistics $S_T(\theta_0)$ and $\bar{S}_T(\bar{\theta}_0)$ are both equal to $(1/2)ESS$ defined in (17) if we set:*

1. *for $S_T(\theta_0)$,*

$$\begin{aligned} X_j &= W(\omega_j)\left(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j)\right) \frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'}, \\ \mathcal{Y}_j &= W(\omega_j)\left(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j)\right) \operatorname{vec}\left(I_T\left(\omega_j\right) - f_{\theta_0}(\omega_j)\right), \end{aligned} \tag{18}$$

for $j = 1, ..., T-1$, and $X_0 = 0, \mathcal{Y}_0 = 0$.

2. for $\bar{S}_T(\bar{\theta}_0)$,

$$X_j = \left[\begin{array}{c} W(\omega_j)\left(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j)\right) \frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \bar{\theta}'} \\ (\pi f_{\theta_0}(0))^{-1/2} \frac{\partial \mu(\bar{\theta}_0)}{\partial \bar{\theta}'} \end{array} \right], \tag{19}$$

$$\mathcal{Y}_j = \left[\begin{array}{c} W(\omega_j)\left(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j)\right) \operatorname{vec}\left(I_T(\omega_j) - f_{\theta_0}(\omega_j)\right) \\ (\pi f_{\theta_0}(0))^{-1/2} T^{-1} \sum_{t=1}^{T} \left(Y_t - \mu(\bar{\theta}_0)\right) \end{array} \right],$$

for $j = 0, 1, ..., T-1$, *where* $I_T(0) = I_{\bar{\theta}_0, T}(0)$.

Consider Lemma 1.1. X_j is an n_Y^2-by-q matrix. Its first component, $W(\omega_j)(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j))$, can be viewed as a scaling factor, which has full rank and is invariant to the strength of identification. The second component, $\partial \operatorname{vec} f_{\theta_0}(\omega_j)/\partial\theta'$, is also an n_Y^2-by-q matrix, whose property is closely linked to the strength of identification. Specifically, if some parameters are weakly identified, then $\partial \operatorname{vec} f_{\theta_0}(\omega_j)/\partial\theta'$ will be small in some directions. That is, by Assumption W there exists a q-by-1 vector $c(\theta_0)$ such that $[\partial \operatorname{vec} f_{\theta_0}(\omega_j)/\partial\theta']c(\theta_0) = O(T^{-1/2})$ for all $j = 1, ..., T-1$, implying there is (imperfect) multicollinearity in the explanatory variables in (16). In the extreme case, when some parameters are unidentified, $\partial \operatorname{vec} f_{\theta_0}(\omega_j)/\partial\theta'$ will equal zero in some directions and consequently (16) will have perfect multicollinearity. Meanwhile, $\mathcal{Y}_j$ asymptotically has mean zero with an identity covariance matrix and is uncorrelated with X_j because the latter is non-random. Therefore, the explained sum of squares is naturally expected to be related to a Chi-square limiting distribution, whose degrees of freedom is to be determined by the column rank of the regressors matrix, which can be smaller than q if some parameters are unidentified. The second result in Lemma 1 has a similar interpretation. Note that in both results, $S_T(\theta_0)$ and $\bar{S}_T(\bar{\theta}_0)$ equals $(1/2)ESS$ but not ESS. This is because of the conjugacy, i.e., for any $\omega \in [0, 2\pi]$, $f_{\theta_0}(2\pi - \omega) = \overline{f_{\theta_0}(\omega)}$ and $I_T(2\pi - \omega) = \overline{I_T(\omega)}$. Also note that X_j being non-random plays a central role in achieving the robustness to weak identification.

The insight that score tests can often be expressed using projected values from linear regressions dates back to Breusch and Pagan (1980), where the relationship was considered as a computational device. Under the above specifications of X_j and $\mathcal{Y}_j$, (16) is a complex-valued Gauss-Newton regression. Davidson and MacKinnon (1993, Chapter 6) provide a detailed discussion of such regressions applied to estimation and hypothesis testing. To the author's knowledge, the current paper is the first that uses such a relationship to understand testing procedures under weak identification.

Theorem 1 *Let Assumptions 1-4 hold and* χ_s^2 *be a Chi-square variable with* s *degrees of freedom.*

1. *Under Assumption W:*

$$\lim_{T\to\infty} \Pr\left(S_T(\theta_0) \leq c\right) \to \Pr\left(\chi_r^2 \leq c\right),$$

where $r = q - q_3$ *with* $q = \dim(\theta_0)$ *and* $q_3 = \dim(\Lambda_{3T}(\theta_0))$.

2. *Under Assumption W2:*

$$\lim_{T\to\infty} \Pr\left(\bar{S}_T(\bar{\theta}_0) \leq c\right) \to \Pr\left(\chi_r^2 \leq c\right),$$

where $r = p + q - \bar{q}_3$ *with* $p + q = \dim(\bar{\theta}_0)$ *and* $\bar{q}_3 = \dim(\bar{\Lambda}_{3T}(\bar{\theta}_0))$.

The above result suggests the following procedure for inference. Without loss of generality, we focus on the dynamic parameters.

- Apply an eigenvalue decomposition to $M_T(\theta_0)$ to determine the dimension of $\Lambda_{3T}(\theta_0)$. Due to numerical errors from solving the model and computing $\partial \operatorname{vec} f_{\theta_0}(\omega_j)/\partial\theta'$, the diagonal elements of $\Lambda_{3T}(\theta_0)$ will not be exactly zero. It is therefore necessary to set a tolerance level. We suggest using the MATLAB default: $\dim(M_T(\theta_0))\text{eps}(\|M_T(\theta_0)\|)$, where eps(.) returns the floating point precision of the corresponding matrix.

- Set the eigenvalues below the tolerance level to exact zeros and use the new $\Lambda_T(\theta_0)$ and the original $Q_T(\theta_0)$ to recompute $M_T(\theta_0)$; see (14). Use the new $M_T(\theta_0)$ and the original $D_T(\theta_0)$ to compute $S_T(\theta_0)$.

- Reject the null hypothesis of $\theta = \theta_0$ if the test statistic exceeds the critical value based on the χ_r^2 distribution.

Two issues require some attention. First, it is important to set the eigenvalues below the tolerance level to exact zero. This ensures that the rank of the regressors matrix in (16) will be exactly r. Otherwise, over-rejection may occur. Second, the choice of the tolerance level introduces some arbitrariness. One of the following three situations may occur: (1) The rank of $M_T(\theta_0)$ is correctly estimated. Then the test will achieve the desired size asymptotically. (2) The rank is overestimated, i.e., some zero eigenvalues are classified as non-zeros. Then the test can be conservative. (3) The rank is underestimated. Then, the proof of Theorem 1 implies that the test is still asymptotically correctly sized. Intuitively, this is because the test behaves asymptotically as the sum of independent χ_1^2 variables. Setting a non-zero eigenvalue to zero is equivalent to removing a variable from the sum, which still yields a Chi-square distribution except that the degree of freedom is reduced by one. Some power might be lost if the eigenvector corresponds to a major deviation from the null hypothesis. Such issues will be illustrated in the simulation section. Note that the above robustness to rank estimation is particularly desirable when the model's identification property depend on the value of of the structural parameters.

5.2 Confidence sets by inverting the score tests

Valid $\alpha\%$ confidence sets for θ and $\bar{\theta}$ can be obtained by inverting $S_T(\theta)$ and $\bar{S}_T(\bar{\theta})$, leading to

$$C_\theta(\alpha) = \left\{\theta \in \Theta : S_T(\theta) \leq \chi^2_{q-q_3}(\alpha)\right\}$$

and

$$C_{\bar{\theta}}(\alpha) = \left\{\bar{\theta} \in \bar{\Theta} : \bar{S}_T(\bar{\theta}) \leq \chi^2_{p+q-\bar{q}_3}(\alpha)\right\}$$

where $\chi^2_s(\alpha)$ denotes the α-th percentile of a Chi-square random variable with s degrees of freedom. Because the sets contain the minimizers of the likelihood function, they are always nonempty.

If the dimension of the parameter vector is small, then the above sets can be constructed in a straightforward manner. If the dimension is high, we suggest using a Metropolis type algorithm, motivated by the analysis in Chernozhukov and Hong (2003). We focus on the dynamic parameter vector case. Let $\pi(\theta)$ be a function that equals to one over Θ and zero otherwise. Then, as in Chernozhukov and Hong (2003), define

$$p_T(\theta) = \frac{\pi(\theta)\exp\left(-\frac{1}{2}S_T(\theta)\right)}{\int_\Theta \pi(\theta)\exp\left(-\frac{1}{2}S_T(\theta)\right)d\theta},$$

which is a proper distribution density over the parameters of interest. We can use random draws from $p_T(\theta)$ to estimate the level set determining $C_\theta(\alpha)$. The main steps are:

1. Choose a starting value $\theta^{(0)}$ and set $j = 0$.

2. Draw θ^* from some proposal distribution $q(.|\theta^{(j)})$.

3. Calculate the ratio
$$s = \min\left\{\frac{\pi(\theta^*)e^{-\frac{1}{2}S_T(\theta^*)}}{\pi(\theta^{(j)})e^{-\frac{1}{2}S_T(\theta^j)}}\frac{q(\theta^{(j)}|\theta^*)}{q(\theta^*|\theta^{(j)})}, 1\right\}$$
and set
$$\theta^{(j+1)} = \begin{cases} \theta^* & \text{with probability } s \\ \theta^{(j)} & \text{with probability } 1-s. \end{cases}$$

4. Increase j by 1 and then repeat Steps 2 and 3. Continue until $j = B$ with B being some large value.

5. Sort the draws according to the values of $S_T(\theta^{(j)})$. Keep the draws corresponding to $S_T(\theta^{(j)}) \leq \chi^2_{q-q_3}(\alpha)$. Use their closure as an estimate for the the level set determining $C_\theta(\alpha)$.

Confidence sets for parameter subvectors can be obtained using the projection method. That is, we use the first k Cartesian coordinates of the MCMC draws in Step 5 to form a confidence set for the first k parameters in θ. Such a method is implemented in Guerron-Quintana, Inoue and Kilian (2010). A thorough discussion of this method in the IV context can be found in Dufour and Taamouti (2005). Here we omit the details.

5.3 Uniform confidence bands for impulse response functions and other objects

The impulse response function measures the effect of a one standard deviation change in the impulse variable on the response variable over an extended horizon. It plays a central role in assessing the implications of a DSGE model. Below, we propose a confidence band that covers this function asymptotically with probability at least $\alpha\%$ even under weak identification. To the author's knowledge, this is the first time such a robust confidence band for impulse responses is proposed in the DSGE literature.

Without loss of generality, consider the impulse response function of the j-th variable in Y_t to the l-th orthogonal shock. This function, when evaluated at horizon k, equals to the (j,l)-th element of $IR(k,\theta) = h_k(\theta)\Sigma^{1/2}(\theta)$, where $h_k(\theta)$ is the k-th coefficients matrix in the vector moving average representation (9). This is easily computed using the output from Sims (2002), because (see (2); without loss of generality, assume $A(L) = A$)

$$IR(k,\theta) = A\Phi_1^k(\theta)\Phi_0(\theta)\Sigma^{1/2}(\theta).$$

The band can be obtained in three steps.

- Step 1. Apply the Metropolis-Hastings algorithm described above to construct $C_\theta(\alpha)$ (or $C_{\bar{\theta}}(\alpha)$ if the steady state parameters are also included in the analysis).

- Step 2. Compute the impulse response function using all parameter values in $C_\theta(\alpha)$. In practice, this step can be approximated using the MCMC draws from Step 1 satisfying $S_T(\theta) \leq \chi^2_{q-q_3}(\alpha)$.

- Step 3. Sort the resulting values at each horizon of interest. Use their maxima and minima to form a confidence band.

It follows directly from Step 1 that this band covers the impulse response function with probability at least $\alpha\%$ asymptotically, because $IR(k,\theta)$ is a deterministic function of θ and k. It is important to note that the band can be narrow even if some parameters are unidentified. This is because if two different parameter values produce the same spectral density over $\omega \in [-\pi,\pi]$ (therefore unidentified), they may also lead to the same set of impulse response functions.[5] This feature will be illustrated in the simulation section.

The same idea can be applied to construct confidence sets for other objects that are deterministic functions of the structural parameter vector. Below we discuss three examples. As a matter of notation, let $\mathbf{e}_j$ be the j-th column of an identity matrix whose dimension depends on the context.

Variance decomposition. Applying the vector moving average representation (9), the contribution of the l-th orthogonal shock to the j-th variable's k-step-ahead forecasting error variance is

$$\sum_{i=0}^{k-1}\left(\mathbf{e}_j' h_k(\theta)\,\Sigma^{1/2}(\theta)\,\mathbf{e}_l\right)^2.$$

The variance decomposition of the j-th variable at the horizon k is therefore

$$R_{jl}^2(k,\theta) = \frac{\sum_{i=0}^{k-1}\left(\mathbf{e}_j' h_k(\theta)\,\Sigma^{1/2}(\theta)\,\mathbf{e}_l\right)^2}{\sum_{l=1}^{n_\varepsilon}\sum_{i=0}^{k-1}\left(\mathbf{e}_j' h_k(\theta)\,\Sigma^{1/2}(\theta)\,\mathbf{e}_l\right)^2}.$$

This is straightforward to compute using the output of Sims (2002). Its confidence band can be obtained using the same three-step procedure outlined above, by replacing $IR(k,\theta)$ in the second step with $R_{jl}^2(k,\theta)$. Because $R_{jl}^2(k,\theta)$ is a deterministic function of θ, the band is uniform in k by construction.

Individual spectrum and coherency. The spectrum of the j-th variable in Y_t is given by $\mathbf{e}_j' f_\theta(\omega)\mathbf{e}_j$. The absolute coherency, which measures the strength of correlation between the j-th and l-th variable at a particular frequency ω, is given by

$$\frac{|\mathbf{e}_j' f_\theta(\omega)\mathbf{e}_l|}{\sqrt{\mathbf{e}_j' f_\theta(\omega)\mathbf{e}_j \mathbf{e}_l' f_\theta(\omega)\mathbf{e}_l}}.$$

It is useful to contrast the model implied confidence bands for these two quantities with some model free (i.e, nonparametric) estimates computed directly from the data. This can potentially reveal

[5] Having the same spectrum is necessary but not sufficient for having the same impulse response function. For example, for any noninvertible MA(1) process, there always exists an invertible MA(1) with the same spectrum but a different impulse response function.

the frequencies at which the model captures or misses important dynamic features in the data. Because the quantities are deterministic functions of θ, their confidence bands uniform in ω can again be computed using the three-step procedure outlined above.

Low frequency hypotheses. Lucas (1980) used the slopes of univariate regressions of moving averages of inflation (π_t) and interest rates (r_t) on money growth (Δm_t) to illustrate the two central implications of the quantity theory of money: that a given change in the rate of money growth induces (i) an equal change in the rate of price inflation and (ii) an equal change in nominal rates of interest. Whiteman (1984) observed that the slopes are related to the coherency between the respective variables at frequency zero. In our notation, the estimated slope approximates

$$\frac{\mathbf{e}_j' f_\theta(0)\mathbf{e}_l}{\mathbf{e}_j' f_\theta(0)\mathbf{e}_j} \tag{20}$$

where j corresponds to Δm_t and l is either π_t or r_t. Sargent and Surico (2011) used DSGE models to show that the slopes are policy dependent. They tacked this issue from a Bayesian perspective. The methods developed in this paper can be used to construct frequentist confidence intervals for (20), therefore to evaluate whether a unit slope is consistent with the model and the data.

6 Identification robust specification testing

We propose a family of tests based on integrated periodograms. In the univariate case, the idea of using integrated periodograms as the basis for misspecification testing can be traced back to Grenander and Rosenblatt (1957) and Bartlett (1955). Here, we generalize this idea to the multivariate setting to examine the specification of a DSGE model.

Suppose the DSGE model is correctly specified with spectral density $f_{\theta_0}(\omega)$ for $\omega \in [-\pi, \pi]$. Then, $\text{vec}\,(I_T(\omega_j) - f_{\theta_0}(\omega_j))$ is asymptotically independently distributed with mean zero for all $j = 1, ..., [T/2]$. This suggests the following Kolmogorov-Smirnov (KS) type test for checking dynamic specification:

$$\mathcal{H}_{dT}(\theta_0) = \sup_{r\in[0,1]} \left\| (T/2)^{-1/2} \sum_{j=1}^{[Tr/2]} \text{vec}\left\{ f_{\theta_0}(\omega_j)^{-1/2} \left(I_T(\omega_j) - f_{\theta_0}(\omega_j)\right) f_{\theta_0}(\omega_j)^{-1/2} \right\} \right\|_\infty ,$$

where $\|.\|_\infty$ is the supremum norm, i.e., for a generic vector $z = (z_1, ..., z_k) \in \mathbb{C}^k$,

$$\|z\|_\infty = \max\left(|z_1|, ..., |z_k|\right).$$

The zero frequency is excluded, therefore the test is invariant to the specification of the steady states. It does not use frequencies outside of $[0, \pi]$. This is because $f_{\theta_0}(2\pi - \omega) = \overline{f_{\theta_0}(\omega)}$ and

$I_T(2\pi-\omega) = \overline{I_T(\omega)}$ for $\pi < \omega < 2\pi$, therefore adding the latter frequencies does not bring new information.

The test can be generalized to examine the model specification over a particular frequency band:

$$\mathcal{H}_{dT}^{W}(\theta_0) = \sup_{r\in[0,1]} \left\| (T/2)^{-1/2} \sum_{j=1}^{[Tr/2]} W(\omega_j)\, \mathrm{vec}\left\{ f_{\theta_0}(\omega_j)^{-1/2} \left(I_T(\omega_j) - f_{\theta_0}(\omega_j) \right) f_{\theta_0}(\omega_j)^{-1/2} \right\} \right\|_{\infty},$$

where $W(\omega_j)$ is an indicator function to select the target frequencies. It can also be extended to construct a joint test for the specification of the mean and the spectrum:

$$\mathcal{H}_T(\bar{\theta}_0) = \max\left(\mathcal{H}_{sT}(\bar{\theta}_0), \bar{\mathcal{H}}_{dT}(\theta_0) \right),$$

where

$$\mathcal{H}_{sT}(\bar{\theta}_0) = \sup_{r\in[0,1]} \left\| \frac{1}{\sqrt{2\pi T}} f_{\theta_0}^{-1/2}(0) \sum_{j=1}^{[Tr]} \left(Y_t - \mu(\bar{\theta}_0) \right) \right\|_{\infty}$$

and $\bar{\mathcal{H}}_{dT}$ is the same as $\mathcal{H}_{dT}$ but with the summation starting at 0 instead of 1 to incorporate the information from the zero frequency.

To present the limiting distributions, let

$$\widetilde{B}(r) = \frac{1}{\sqrt{2}} \left(B_1(r) + iB_2(r) \right),$$

where $B_1(r)$ and $B_2(r)$ are two independent standard Brownian motion processes.

Theorem 2 *Suppose $\{Y_t\}_{t=1}^{T}$ satisfies $\mathbb{E}Y_t = \mu(\bar{\theta}_0)$ with spectral density $f_{\theta_0}(\omega)$ for $\omega \in [-\pi, \pi]$. Under Assumptions 1-4:*

1. *$\mathcal{H}_{dT}(\theta_0) \Rightarrow \sup_{r\in[0,1]} \|G_d(r)\|_{\infty}$, where $G_d(r)$ is an $n_Y(n_Y+1)/2$ vector of independent processes, whose first n_Y elements are standard Brownian motions and the last $n_Y(n_Y-1)/2$ elements are independent copies of $\widetilde{B}(r)$.*

2. *$\mathcal{H}_{dT}^{W}(\theta_0) \Rightarrow \sup_{r\in[0,1]} \left\| \int_0^1 W(r)\, dG_d(r) \right\|_{\infty}$, where $W(r)$ is the indicator function specified by the researcher.*

3. *$\mathcal{H}_{sT}(\bar{\theta}_0) \Rightarrow \sup_{r\in[0,1]} \|G_s(r)\|_{\infty}$, where $G_s(r)$ is an n_Y vector of independent standard Brownian motion processes.*

4. *$\mathcal{H}_T(\bar{\theta}_0) \Rightarrow \max\left(\sup_{r\in[0,1]} \|G_d(r)\|_{\infty}, \sup_{r\in[0,1]} \|G_s(r)\|_{\infty} \right)$ with the elements of $G_d(r)$ and $G_s(r)$ being mutually independent.*

The limiting distribution is invariant to the strength of identification. The first two results are independent of whether the steady states are correctly specified. In $G_d(r)$, the first n_Y elements correspond to the diagonal elements in $I_T(.)$ and the remaining $n_Y(n_Y-1)/2$ elements correspond to their off-diagonal elements. The limiting distributions of $\mathcal{H}_{dT}(\theta_0)$, $\mathcal{H}_{sT}(\bar{\theta}_0)$ and $\mathcal{H}_T(\bar{\theta}_0)$ are pivotal, depending only on n_Y and can be easily simulated. In Table 1, we provide 10%, 5% and 1% critical values for n_Y between 1 and 10. Critical values for $\mathcal{H}_{dT}^W$ depend also on $W(\omega)$, therefore need to be tabulated on a case by case basis. We will provide a GAUSS code for such a purpose.

The next result establishes the global power properties of the tests.

Theorem 3 *Suppose $\{Y_t\}$ is a weakly stationary process; $\mathbb{E}Y_t = \mu_0$ and its spectral density $f_0(\omega)$ satisfies Assumptions 1-4. Suppose $\mu(\bar{\theta}_0)$ and $f_{\theta_0}(\omega)$ are the mean and spectral density implied by the DSGE model, satisfying Assumptions 1-4. Let $\delta > 0$ be an arbitrary constant independent of T. Then:*

1. *$\mathcal{H}_{dT}(\theta_0) \to \infty$ if $\|f_0(\omega) - f_{\theta_0}(\omega)\| > \delta$ for some $\omega \in [0,\pi]$.*
2. *$\mathcal{H}_{dT}^W(\theta_0) \to \infty$ if $\|f_0(\omega) - f_{\theta_0}(\omega)\| > \delta$ for some ω with $W(\omega) = 1$.*
3. *$\mathcal{H}_{sT}(\bar{\theta}_0) \to \infty$ if $\|\mu_0 - \mu(\bar{\theta}_0)\| > \delta$.*
4. *$\mathcal{H}_T(\bar{\theta}_0) \to \infty$ if $\|f_0(\omega) - f_{\theta_0}(\omega)\| > \delta$ for some $\omega \in [0,\pi]$ or $\|\mu_0 - \mu(\bar{\theta}_0)\| > \delta$.*

The result implies that the $\mathcal{H}_{dT}$ test has power approaching one if there is a nonvanishing difference between the model and data spectra over an arbitrary band of frequencies. The same holds for $\mathcal{H}_{dT}^W$, provided that the difference exists within the frequencies selected by $W(\omega)$. The $\mathcal{H}_{sT}$ test can consistently detect misspecification in the mean. Finally, the $\mathcal{H}_T$ test pools information from the mean and the spectrum and is consistent if there is a nonvanishing difference in either of them.

When implementing the tests, we can use a Bonferroni procedure to acknowledge that θ_0 and $\bar{\theta}_0$ are unknown and may be weakly identified. For illustration, consider constructing the $\mathcal{H}_{dT}$ test with asymptotic size not exceeding α%. First, specify positive constants α_S and $\alpha_{\mathcal{H}}$ such that $\alpha = \alpha_S + \alpha_{\mathcal{H}}$. Then, apply $S_T(\theta_0)$ to obtain an α_S% confidence set $C_\theta(1-\alpha_S)$. Next, compute

$$\inf_{\theta \in C_\theta(1-\alpha_S)} \mathcal{H}_{dT}(\theta)$$

and reject the null hypothesis if it exceeds the $\alpha_{\mathcal{H}}$% critical value of $\sup_{r\in[0,1]} \|G_d(r)\|_\infty$. Intuitively, this procedure can be viewed as first finding a set of plausible models and then checking whether the most favorable model is supported by the data. From a computational point of view, if the dimension of θ_0 is low, then $\inf_{\theta \in C_\theta(1-\alpha_S)}$ can be approximated by using a grid over $C_\theta(1-\alpha_S)$. If the dimension is high, then it can be approximated by computing the infimum of $\mathcal{H}_{dT}$ over the MCMC draws obtained in the previous section. The same idea can be applied to the remaining three tests, except that when applying $\mathcal{H}_{sT}$ and $\mathcal{H}_T$, $S_T(\theta_0)$ needs to be replaced by $\bar{S}_T(\bar{\theta}_0)$ and the infimum replaced by $\inf_{\bar{\theta} \in C_{\bar{\theta}}(1-\alpha_S)}$.

Remark 4 *The above tests can be jointly applied to deliver informative results on the nature of the misspecification. We can first apply $\mathcal{H}_T$ to examine whether misspecification is present in either the mean or the spectrum. If a rejection occurs, then apply $\mathcal{H}_{dT}$ and $\mathcal{H}_{sT}$ to trace out the source of the misspecification. If indeed the spectrum is misspecified, then $\mathcal{H}_{dT}^W$ can be used to further examine the business cycle frequency. In particular, we may plot the processes inside the norm in $\mathcal{H}_{dT}^W$ as a function of r, which can potentially reveal the frequencies responsible for rejecting the model.*

Remark 5 *It is often important to analyze a subset of variables within the full DSGE system. For example, King and Watson (1996) compared three rational expectations models to capture the relationship between a real (GDP) and a nominal (interest rate) variable. In our framework, such analysis can be carried out by introducing a k-by-n_Y dimensional selection matrix $A_{\mathcal{H}}$ to select the relevant k variables. Then, the above tests can be constructed by replacing $f_{\theta_0}(\omega)$ and $\mu(\bar{\theta}_0)$ with $A_{\mathcal{H}} f_{\theta_0}(\omega) A'_{\mathcal{H}}$ and $A_{\mathcal{H}}\mu(\bar{\theta}_0)$, respectively. Theorem 2 then holds with n_Y replaced by k. Note that $A_{\mathcal{H}}$ plays a different role than the $A(L)$ matrix in (2). Specifically, $A(L)$ is determined by the variables observable to the econometrician and needs to be specified prior to estimation, while $A_{\mathcal{H}}$ is typically specified later to allow us to focus on some key variables in the model.*

7 Applicability of the results to other dynamic models

Although the paper focuses on DSGE models, the methods developed for confidence sets and model diagnostics are applicable to other dynamic models satisfying Assumptions 1-4, W and W2. Below we discuss two such models.

Structural Vector Regression (SVAR). The model is

$$\Phi_0 Y_t = \sum_{j=1}^{p} \Phi_j Y_{t-j} + \varepsilon_t,$$

where Y_t is an n-by-1 random vector, $\Phi_0, \Phi_1, ..., \Phi_p$ are coefficients matrices and ε_t is an n-by-1 vector of serially uncorrelated structural disturbances with $Var(\varepsilon_t) = \Sigma$. To relate the model to

Assumptions 1-4, let $\theta = (\Phi_0, \Phi_1, ..., \Phi_p, \Sigma)$ and

$$\Pi(L;\theta) = \Phi_0 - \sum_{j=1}^{p} \Phi_j L^j.$$

Assumption 1 holds if Σ is positive definite. Assumption 2 holds if all the roots of $|\Pi(z;\theta)| = 0$ are outside of the unit circle. There, $H(\theta; L) = \Pi(L;\theta)^{-1}$ and the spectral density of Y_t is given by

$$f_\theta(\omega) = \frac{1}{2\pi}\left[\Pi(\exp(-i\omega);\theta)^{-1}\right]\Sigma\left[\Pi(\exp(-i\omega);\theta)^{-1}\right]^*.$$

Assumption 3 is satisfied if Σ is positive definite and Assumption 4 holds if ε_t are i.i.d. Normally distributed.

Watson (2006) illustrated the cause and consequences of weak identification in the SVAR model. The methods proposed in this paper can be used to construct robust confidence set for θ, to study the impulse responses, to test low frequency hypotheses and to evaluate the specification of the model at a particular band of frequencies.

Factor Augmented VAR (FAVAR). The model is (same as Equations (20)-(21) in Stock and Watson, 2005, but with slightly different notation)

$$\begin{aligned} Y_t &= \lambda(L) f_t + D(L) Y_{t-1} + v_t, \\ f_t &= \Gamma(L) f_{t-1} + \zeta_t, \end{aligned}$$

where Y_t is an n-by-1 vector of observables, f_t comprises of the latent factors, ζ_t is a serially uncorrelated structural disturbance with $Var(\zeta_t) = I$, $Var(v_t) = \Sigma$ and $\mathbb{E}\zeta_t v_s' = 0$ for all t and s. $\lambda(L)$, $D(L)$ and $\Gamma(L)$ are matrix lag polynomials with $D(L)$ typically being diagonal.

The structural parameter vector θ consists of the elements in $\lambda(L)$, $D(L), \Gamma(L)$ and Σ. Under stationarity, Y_t has the following moving average representation:

$$Y_t = H_1(L;\theta)\zeta_t + H_2(L;\theta) v_t$$

where

$$\begin{aligned} H_1(L;\theta) &= [I - D(L)L]^{-1}\lambda(L)[I - \Gamma(L)L]^{-1}, \\ H_2(L;\theta) &= [I - D(L)L]^{-1}, \end{aligned}$$

Therefore, Y_t can be viewed as generated by (8) but with serially correlated measurement errors $H_2(L;\theta)v_t$. Its spectral density is given by

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

Assumptions 1-4 are expected to hold under mild conditions.

Stock and Watson (2005) discussed a variety of identification strategies, including using (1) contemporaneous timing restrictions, i.e., on the zero-order term in $H_1(L;\theta)$, (2) long-run restrictions, i.e., on $H_1(1;\theta)$, (3) factor loading restrictions, i.e., on $\lambda(L)$ and (4) Uhlig's (2005) sign restrictions on the coefficients of $H_1(L;\theta)$. Because these restrictions only concern $f_\theta(\omega)$, they are implementable in the frequency domain. Inference and model diagnostics can then be carried out under such restrictions. Note that because we work directly with the structural parameter vector, we can avoid making identifying assumptions about the reduced form parameters. Also note that our framework assumes n is finite, thus precluding the analysis of high-dimensional factor models. A potential generalization towards such a direction is to first estimate the factors and then treat them as part of the data, see Stock and Watson (2005) and the references therein for details regarding such a two step procedure implemented in the time domain. We intend to carry out a frequency domain analysis in further work.

8 Monte Carlo experiments

This section examines the finite sample properties of the following objects: the proposed tests, the confidence intervals and the confidence bands for the impulse response functions.

The model is taken from An and Schorfheide (2007). Its log-linearized solutions are:

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

where $\epsilon_{rt} \sim N(0,\sigma_r^2)$, $\epsilon_{gt} \sim N(0,\sigma_g^2)$, and $\epsilon_{zt} \sim N(0,\sigma_z^2)$ are serially and mutually independent shocks. The observables are GDP growth (YGR_t), inflation ($INFL_t$) and interest rate (INT_t):

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

where $\gamma^{(Q)} = 100(\gamma - 1)$, $\pi^{(A)} = 400(\overline{\pi} - 1)$, $r^{(A)} = 400(1/\beta - 1)$ with γ being a constant in the technological shock equation and $\overline{\pi}$ the steady state inflation rate. The full parameter vector is

$$\bar{\theta} = (\tau, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, 100\sigma_r, 100\sigma_g, 100\sigma_z, r^{(A)}, \pi^{(A)}, \gamma^{(Q)}).$$

The first 11 parameters are dynamic parameters ($r^{(A)}$ depends on the discount factor β which appears in the log-linearized equation). The remaining two are steady state parameters as they do not affect the log-linearized system. The parameter values are taken from the last column of Table 2 in An and Schorfheide (2007):

$$\bar{\theta}_0 = (2, 0.15, 1.5, 1.00, 0.60, 0.95, 0.65, 0.2, 0.8, 0.45, 0.40, 4.00, 0.50). \quad (21)$$

We consider the following three designs that correspond to different treatments of the mean and the spectrum:

Design 1 (BC frequencies). Inference on θ based on business cycle frequencies (i.e., periods of 6–32 quarters). With quarterly data, this amounts to considering $\omega \in [\pi/16, \pi/3] \cup [5\pi/3, 31\pi/16]$.

Design 2 (Full spectrum). Inference on θ based on the full spectrum.

Design 3 (Mean and full spectrum). Inference on $\bar{\theta}$ based on the mean and the full spectrum.

When implementing the tests, the model is solved using the GENSYS algorithm of Sims (2002). The derivatives of $\mu(\bar{\theta})$ with respect to $\bar{\theta}$ are computed analytically. The derivatives of $f_\theta(\omega)$ are computed using a simple two-point method with step size set to 10^{-6}. To evaluate the size and power, we consider four sample sizes typically encountered in practice with quarterly observations: $T = 80, 160, 240, 320$. In each case, we report rejection frequencies at the 10% and 5% nominal levels using 5000 replications.

We now illustrate the identification property of the model to facilitate a better understanding of the simulation results. When evaluated at (21), both $M_T(\theta_0)$ and $\bar{M}_T(\bar{\theta}_0)$ have one zero eigenvalue because the Taylor rule parameters $(\psi_1, \psi_2, \rho_r, \sigma_r)$ are not separately identifiable[6]. We apply the method of Qu and Tkachenko (2011) to trace out the non-identification curve along which the parameter values are observationally equivalent. The curve starts at θ_0 and is extended in both the positive and negative directions. In Direction 1, the curve is truncated before the output weight

[6] The identification issue associated with the Taylor rule parameters in this model was first discussed in Qu and Tkachenko (2011). Computing their rank condition in Theorem 1 at θ_0 returns an 11-by-11 matrix with rank 10 when the default tolerance level is used ($\dim(\cdot)\text{eps}(\|\cdot\|)$). The smallest eigenvalue equals 2.3×10^{-14}. It corresponds to the four parameters in the Taylor rule equation. The second smallest eigenvalue equals 5.4×10^{-8}. It corresponds to $r^{(A)}$ or equivalently the discount factor β. All the other eigenvalues are above 0.1. Computing the eigenvalues of $M_T(\theta_0)$ and $\bar{M}_T(\bar{\theta}_0)$ directly leads to the same conclusion concerning the rank deficiency. The latter result holds for all the four sample size considered and under all the three designs.

parameter ψ_2 turns negative. Along Direction 2, it reaches an indeterminacy region before any natural bounds on parameter values are violated, and is truncated at the last point that yields a determinate solution. Table 2 reports 10 evenly spaced points along each direction. Two patterns emerge. First, the non-identification curve occupies a substantial portion of the parameter space: the inflation weight parameter ψ_1 varies between 0.99 and 4.87, while the output weight ψ_2 varies between 0.00 and 1.15. Second, although the interest rate smoothing parameter ρ_r and standard deviation parameter σ_r are unidentified, the associated neighborhoods are relatively small. The value of ρ_r can only change between 0.58 and 0.60, while $100\sigma_r$ can only vary between 0.19 and 0.20. This feature suggests that the model can still be informative about ρ_r and σ_r even though they are unidentified.

8.1 Size in finite samples

The test statistics are constructed by setting the smallest eigenvalue in the information matrix to zero. The critical values of χ^2_{10} and χ^2_{12} are then used for $S_T(\theta_0)$ and $\bar{S}_T(\bar{\theta}_0)$ to determine whether a rejection occurs. The results are summarized in the first panel in Table 3. Under Design 1 (BC frequencies), the $S_T(\theta_0)$ test has very good size properties. Its rejection frequencies are close to the nominal levels even with $T = 80$. Under Designs 2 and 3, the $S_T(\theta_0)$ and $\bar{S}_T(\bar{\theta}_0)$ tests continue to have adequate sizes, although some mild over-rejection exists, with the maximum rejection frequencies being 14.4% and 9.5% at the 10% and 5% nominal levels respectively. The over-rejection appears to be because the spectral density is close to being singular near the zero frequency. Therefore, this is not a problem with the test statistics, but rather the specification of the model and how it is applied to the data. In such a context, the adequacy of the proposed procedures should be judged according to the results using the business cycle frequencies. The rejection frequencies of the $\mathcal{H}$ tests (including $\mathcal{H}^W_{dT}(\theta_0), \mathcal{H}_{dT}(\theta_0)$ and $\mathcal{H}_T(\bar{\theta}_0)$) are close to the nominal levels under all the three designs with all sample sizes, consistent with the theory that their limiting distributions are invariant to the strength of identification.

Since the above experiment considers a particular parameter value, it remains to verify whether the size is controlled in a more general situation. To analyze this, we draw parameter values from a prior distribution given in An and Schorfheide (2007, Table 2). In addition to requiring determinacy, the following bounds are also imposed on the permissible parameter values: $\tau \sim [\text{1e-5}, 5]$, $\kappa \sim [0, 1]$, $\psi_1 \sim [0, 5]$, $\psi_2 \sim [0, 2]$, $\rho_r \sim [0, 0.9]$, $\rho_g \sim [0, 0.99]$, $\rho_z \sim [0, 0.99]$, $100\sigma_r \sim [\text{1e-5}, 2]$, $100\sigma_g \sim [\text{1e-5}, 2]$, $100\sigma_z \sim [\text{1e-5}, 2]$, $r^{(A)} \sim [0, 5]$, $\pi^{(A)} \sim [0, 20]$, $\gamma^{(Q)} \sim [0, 5]$. The bounds are sufficiently wide to allow for estimates reported in the DSGE literature. To avoid confounding the results with the issue of near unit roots, the parameters ρ_r, ρ_g and ρ_z are fixed at their original values throughout the draws (their bounds are needed later when constructing confidence intervals). We generate 5000 draws satisfying the above requirements and compute the test statistics using them in the same way as in the previous case. The rejection frequencies are summarized in the second panel in Table 3. The $S_T(\theta_0)$ and $\bar{S}_T(\bar{\theta}_0)$ tests continue to show decent size properties, with the rejection frequencies varying between 8.6%-15.6% and 5.2%-10.2% at the 10% and 5% levels, respectively. The rejection frequencies of the $\mathcal{H}$ tests are between 7.9%-10.6% and 3.7-6.8% at the two levels. Overall, the results are similar to the previous case and are encouraging.

We now examine the size property when the rank of the information matrix is estimated with error. Table 4 presents results for the case where the rank is overestimated, i.e., all eigenvalues of the information matrix are treated as non-zeros when constructing the tests and the distributions of χ^2_{11} and χ^2_{13} are used for inference. As expected, the $S_T(\theta_0)$ and $\bar{S}_T(\bar{\theta}_0)$ tests reject less frequently compared with Table 3. The rates are between 6.3%-12.5% and 3.3%-8.2% at the two nominal levels. The $\mathcal{H}$ tests are not affected by the choice, thus are not reported. Table 5 presents results for the case where the rank is underestimated, i.e., two eigenvalues are classified as zeros when constructing the tests and the distributions of χ^2_9 and χ^2_{11} are used for inference. There, the size continues to be adequate: the overall rejection rates are between 7.5%-15.8% and 4.6%-10.8% at the two nominal levels. In the above, the rank of the information matrix stayed fixed across parameter values. Yet another alternative is to re-estimate its rank with the MATLAB default when each time a parameter is drawn, and then use this rank for computing the test statistics and for inference. This led to virtually the same results as in Table 3 and is omitted to save space. In summary, the size appears fairly robust to the classification of the zero eigenvalues.

To put the above results into perspective, we note that in this example, because the MLE from solving (11) (or (12)) is inconsistent (in the sense that it can converge to any point along the non-identification curve) and the information matrix is singular at θ_0, the distribution of the conventional Wald and likelihood ratio (LR) tests are highly nonstandard. Intuitively, for the Wald test, since the MLE $\hat{\theta}$ may lie anywhere close to the non-identification curve, $\sqrt{T}(\hat{\theta}-\theta_0)$ can be quite large, therefore the test can take on a large value if a finite covariance matrix is used to standardize the above difference. For the conventional LR test, since the likelihood surface displays a ridge along the non-identification curve, the standard quadratic approximation as in Van der Vaart (2000, p. 233) fails, and consequently the Chi-square approximation to the limiting distribution breaks down. The $S_T(\theta_0)$ and $\bar{S}_T(\bar{\theta}_0)$ tests do not require estimating the model, therefore circumvent the issue of parameter inconsistency. They exploit the projection relationship as shown in Lemma 1, and accordingly are also robust to the singularity of the information matrix.

8.2 Finite sample power

We perturb the individual element of θ_0 (or $\bar{\theta}_0$ in Design 3) given in (21) by a fixed percentage and then compute the rejection frequencies. Specifically, we take a uniform random draw from the index set $\{1, ..., 11\}$ (or $\{1, ..., 13\}$ in Design 3) and change the corresponding element of the parameter vector by κ% of its value (increasing or decreasing it with equal probability) without altering the others. This is repeated to generate 5000 parameter values yielding determinacy, which are then used to simulate 5000 processes and to compute the test statistics. The size-unadjusted rejection frequencies are reported in Table 6.

The first panel in Table 6 is for $\kappa = 20$. Using the business cycle frequencies, the $S_T(\theta_0)$ test achieves 49%-62% of the power attainable using the full spectrum. For the $\mathcal{H}$ tests, the corresponding ratios are between 58%-67%. The rejection frequencies under Designs 2 and 3 are similar. Note that because Design 3 involves more parameters, the power is not necessarily higher than under Design 2. Comparison between $S_T(\theta_0)$ and $\mathcal{H}$ shows that their powers are comparable. This is because by construction the model is correctly specified and all the tests act as statistics for parametric restrictions. In the presence of misspecification, their performance can be quite different. The second panel in Table 6 corresponds to $\kappa = 40$. There, the ratios are between 66%-78% and 71%-79% for the $S_T(\theta_0)$ and $\mathcal{H}$ tests, respectively. Designs 2 and 3 continue to show similar rejection frequencies.

We now examine the power property when the rank of the information matrix is estimated with error. Table 7 presents results for the case where all eigenvalues are treated as non-zeros. The powers of the $S_T(\theta_0)$ and $\bar{S}_T(\bar{\theta}_0)$ tests are now mildly lower compared with Table 6. The difference is between 0.01 and 0.05. This is consistent with the procedure being conservative. Table 8 presents results for the case where two eigenvalues are treated as zeros. There, the power is almost identical to Table 6, with the difference falling between -0.01 and 0.02. Intuitively, a small eigenvalue implies that the likelihood surface lacks curvature in the corresponding direction. Unsurprisingly, little power is lost by treating the curvature as exactly zero.

We note two things. First, the results from Designs 1 and 2 suggest that the tests using only business cycle frequencies can still be informative. This issue will be further illustrated in the next two subsections. Second, although the above analysis allows us to compare power across different designs, it is not ideal because the alternatives are limited to some particular parameter directions. If the alternative parameter values were instead on the non-identification curve, then the power of the tests would be the same as their size. We now turn to the confidence intervals and impulse responses to further study the informativeness of the different procedures.

8.3 Confidence intervals for structural parameters

We consider the length of 90% confidence intervals when $\bar{\theta}$ is given by (21) and $T = 240$. The Metropolis-Hastings algorithm is used to compute them. Specifically, a Markov chain is started with a random draw from the prior distribution given in An and Schorfheide (2007, Table 2). As the chain proceeds, the draws corresponding to $S_T(\theta)$ (or $\bar{S}_T(\bar{\theta})$ in Design 3) below the asymptotic critical value are kept. The determinacy requirement and the parameter bounds are also imposed. The chain continues until 2,000 valid draws are stored. A confidence interval is then formed by merging 5 independent chains and applying the projection method. This procedure is repeated for 100 simulated series and the results are summarized in Table 9.

The results from using business cycle frequencies are summarized in the fourth column in Table 9. The first value in each row is the median length of the confidence intervals. The remaining two values are the medians of their lower and upper limits. Three observations emerge. First, the intervals contain little information about the risk-aversion parameter τ, the policy rule coefficients ψ_1 and ψ_2 and the discount parameter $r^{(A)}$ (or equivalently β). This is consistent with findings reported elsewhere in the literature. For example, An and Schorfheide (2007, P.133-134) documented similar results about τ, ψ_1 and ψ_2 from a Bayesian perspective. It is also well known that β is difficult to estimate with data on aggregate quantities. Second, the intervals reveal limited information about the interest rate smoothing parameter ρ_r and the Phillips-curve slope coefficient κ. They are narrower than the parameter bounds, but are still wide. Third, the intervals related to the exogenous disturbances $(\rho_g, \rho_z, 100\sigma_r, 100\sigma_g$ and $100\sigma_z)$ are relatively tight and informative. This is again consistent with the findings of An and Schorfheide (2007, P.133-134).

The next column contains the results from using the full spectrum. The intervals for τ, ψ_1, ψ_2 and $r^{(A)}$ are little changed; the others narrow substantially. The efficiency gain from using the full spectrum is clearly parameter specific. When the steady state parameters are included in the analysis (Column 6 in Table 9), the intervals remain roughly the same except for $r^{(A)}$. The latter narrows because $r^{(A)}$ is tied to the steady state of the interest rate.

The above analysis suggests two conclusions. First, it is possible to have informative confidence intervals in DSGE models with unidentified parameters. In fact, even unidentified parameters themselves can have tight and valid confidence intervals; see the results for ρ_r and σ_r in Columns 5 and 6. This is a strong finding which has not been previously documented in the DSGE literature.

Second, business cycle based inference can be informative, see the results for $\rho_g, \rho_z, 100\sigma_r, 100\sigma_g$ and $100\sigma_z$ in Column 4. Meanwhile, moving from the business cycle frequencies to the full spectrum can bring substantial gain in efficiency. In practice, this offers the researcher a choice. If one firmly believes that the model is reasonably 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 is inconsistent with the data (for example, the data has a broken trend while the model has a linear trend), then using a subset of frequencies is more preferable.

8.4 Confidence bands for impulse responses

We illustrate the properties of the 90% uniform confidence bands using a simulated process with $\bar{\theta}$ defined by (21) and $T = 240$. The maximum horizon equals 20. The bands are computed using merged outcomes from 20 independent Markov chains with each chain producing 2000 valid draws.

Figure 1 presents results using only business cycle frequencies. In each plot, the shaded area is the 90% uniform confidence band. The solid line is the true impulse response function. The confidence bands are in general fairly wide but can be informative. They reveal that the three shocks have significant immediate effects on all the variables, with the exception of ϵ_{gt} on inflation and interest rate, which are identically zero dictated by the structure of the model. They correctly estimate the signs of the responses and are indicative of the possible magnitudes. Figure 2 presents results using the full spectrum. All bands narrow substantially and are now fairly informative. This is an interesting finding, given that the model has unidentified parameters. Figure 3 contains results using the mean and the spectrum. The bands are similar to those in Figure 2. Note that they are not necessarily narrower than those in Figure 2, because additional parameters are present.

In summary, it is possible to have informative interval estimates of the impulse response functions in DSGE models with unidentified parameters. Intuitively, because observationally equivalent parameter values may correspond to the same response functions, uncertainty about parameter values does not necessarily translate into uncertainty about the latter. For a further illustration, we computed the impulse response functions using the 20 points reported in Table 1. The maximum difference between them is of order 1e$-$7. This confirms that in this example, the parameters on the identification curve do deliver the same impulse responses.

9 Conclusion

This paper has developed asymptotically valid confidence sets for parameters in log-linearized DSGE models allowing an unknown subset to be weakly (including un-) identified. It also developed uniform confidence bands for impulse response functions and other objects that are functions of the structural parameters. In addition, it proposed a family of frequency domain misspecification tests that are robust to weak identification. They can be used to test for misspecification in the mean, in the spectrum as well as misspecification within a band of frequencies. The framework is fairly general, permitting latent endogenous variables, measurement errors and also inference using only part of the spectrum. The simulation experiment using a calibrated model suggests that the tests have decent size in relatively small samples. It also suggests that it is possible to obtain informative results in DSGE models with unidentified parameters.

Joint confidence sets are sometimes considered as not useful in the frequentist literature because they can be quite conservative about individual parameters. This paper suggests that this need not be the case. They can be useful for a wide range of purposes, including (1) constructing uniform confidence bands for the impulse response functions, the time path of the variance decomposition, the individual spectrum and absolute coherency, (2) examining certain low frequency hypotheses, and (3) testing the specification of the model. Parameters in DSGE models are often highly correlated. This can be seen from the non-identification curve reported in Table 2, and is also emphasized in the literature, for example by Del Negro and Schorfheide (2008). Such dependence is captured by joint confidence sets, but not by individual confidence intervals. It is therefore desirable to develop methods that can facilitate the visualization and characterization of such sets in a high dimensional setting. We view this as a challenging task that deserves further investigation.

References

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

Anderson, G., and G. Moore (1985): "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models," *Economics Letters*, 17, 247-252.

Andrews, D.W.K., and X. Cheng (2011): "Estimation and Inference with Weak, Semi-strong, and Strong Identification", Cowles Foundation Discussion Paper No. 1773R.

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

Bartlett, M.S. (1955): *An Introduction to Stochastic Processes with Special Reference to Methods and Applications.* Cambridge University Press, London.

Billingsley, P. (1968): *Convergence of Probability Measures.* Wiley, New York.

Breusch, T.S. and A.R. Pagan (1980): "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics," *The Review of Economic Studies*, 47, 239-253.

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

Canova, F. (2007): *Methods for Applied Macroeconomic Research*, Princeton University Press.

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

Chernozhukov, V., and H. Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.

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

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

Davidson, R., and J.G. MacKinnon (1993): *Estimation and Inference in Econometrics,* Oxford University Press.

Davis, C., and W. Kahan (1970): "The Rotation of Eigenvectors by a Perturbation: III," *SIAM Journal on Numerical Analysis,* 7, 1–46.

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

Del Negro, M., F.X. Diebold., and F. Schorfheide (2008): "Priors from Frequency Domain Dummy Observations," Working Paper, Federal Reserve Bank of New York.

DeJong, D.N., and C. Dave (2007): *Structural Macroeconometrics*, Princeton University Press.

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

Dufour, J.M., L. Khalaf., and M. Kichian (2010) "Structural Multi-equation Macroeconomic Models: Identification-robust Estimation and Fit", Working Paper, McGill University.

Dufour, J.M., and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments." *Econometrica,* 73, 1351–1365.

Dunsmuir, W. (1979): "A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Application to Models for a Signal Observed with Noise," *The Annals of Statistics*, 7, 490–506.

Dunsmuir, W., and E. J. Hannan (1976): "Vector Linear Time Series Models," *Advances in Applied Probability*, 8, 339–364.

Engle, R. F. (1974): "Band Spectrum Regression," *International Economic Review*, 15, 1–11.

Fernández-Villaverde, J. (2010): "The Econometrics of DSGE Models," *Journal of the Spanish Economic Association*, 1, 3–49.

Grenander, U., and M. Rosenblatt (1957): *Statistical Analysis of Stationary Time Series.* Wiley, New York.

Guerron-Quintana, P., A. Inoue., and L. Kilian (2010): "Frequentist Inference in Weakly Identified DSGE Models," Working Papers, Federal Reserve Bank of Philadelphia.

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

———(1976): "The Asymptotic Distribution of Serial Covariances," *Annals of Statistics,* 4, 396-399.

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

Horn, R.A., and C.R. Johnson (2005): *Matrix Analysis*, Cambridge University Press.

———(2006): *Topics in Matrix Analysis*, Cambridge University Press.

Ireland, P.N. (2007): "Changes in the Federal Reserve's Inflation Target: Causes and Consequences," *Journal of Money, Credit, and Banking*, 8, 1851-1882.

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

King, R. G., and M.W. Watson (1996): "Money, Prices, Interest Rates and the Business Cycle," *The Review of Economics and Statistics*, 78, 35–53.

——— (2002): "System Reduction and Solution Algorithms for Singular Linear Difference Systems under Rational Expectations," *Computational Economics*, 20, 57-86.

Kleibergen, F. (2001): "Testing Subsets of Structural Parameters in the IV Regression Model," *Review of Economics and Statistics*, 86, 418-423

———(2005): "Testing Parameters in GMM without Assuming that They are Identified, *Econometrica*, 73, 1103-1124.

Kleibergen, F., and S. Mavroeidis (2009a): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve", *Journal of Business and Economic Statistics* 27, 293-311.

———(2009b): "Inference on Subsets of Parameters in GMM without Assuming Identification", Working Paper, Brown University.

Klein, P. (2000): "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Journal of Economic Dynamics and Control*, 4, 257-271.

Komunjer, I., and S. Ng (2011): "Dynamic Identification of DSGE Models," forthcoming in *Econometrica.*

Lucas, R.E., Jr. (1980): "Two Illustrations of the Quantity Theory of Money." *American Economic Review*, 70: 1005–1014.

Magnus, J. R., and H. Neudecker, H. (2002): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: John Wiley.

Qu, Z. and D. Tkachenko (2011): "Identification and Frequency Domain QML Estimation of Linearized DSGE Models," forthcoming in *Quantitative Economics.*

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

Sargent, T.J. and P. Surico (2011): "Two Illustrations of the Quantity Theory of Money: Breakdowns and Revivals," *American Economic Review*, 101, 109–128.

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

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

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

Staiger, D., and J.H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.

Stock, J.H., and J.H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.

Stock, J.H., and M.W. Watson (2005): "Implications of Dynamic Factor Models for VAR Analysis," working paper, Harvard University.

Taylor, J.B. (1993): "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy, 39*, 195-214.

Uhlig, H. (1999): "A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily," in *Computational Methods for the Study of Dynamic Economies*, ed. by R. Marimon, and A. Scott, 30-61. Oxford University Press.

——— (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics,* 52, 381-419.

van der Vaart, A.W. (2000): *Asymptotic Statistics*. Cambridge.

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

———(2006): Comment on Christiano, Eichenbaum and Vigfusson: "Assessing Structural VARs", *NBER Macroeconomics Annual*, 21, 97-102.

Whiteman, C.H. (1984): "Lucas on the Quantity Theory: Hypothesis Testing without Theory," *American Economic Review*, 74, 742–749.

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

Zygmund, A.G. (1959): *Trigonometric Series*, Cambridge University Press.

Appendix A.
Eigenvalue conditions corresponding to other characterizations of weak identification

We illustrate that the characterizing conditions for weak identification used in the IV and GMM literature can be stated using the curvatures of the criterion functions used for inference as in Assumptions W and W2.

Linear IV (Staiger and Stock, 1997). Consider the model $y = Y\beta + u, Y = Z\Pi + v$, where y and Y are $T \times 1$ vectors, Z is a $T \times K$ matrix of instruments and u and v are $T \times 1$ vectors of disturbances with $\mathbb{E}uu' = \sigma_u^2 I_T$. The objective function is $Q(\beta) = (y - Y\beta)' P_Z (y - Y\beta)$. Its first order derivative, normalized by $T^{-1/2}$, equals

$$D_T(\beta_0) = -2T^{-1/2} u' Z\Pi - 2T^{-1/2} (u'Z) (Z'Z)^{-1} (Z'v).$$

If β_0 is strongly identified, i.e., Π is nonzero and independent of T, then the first term in $D_T(\beta_0)$ is of exact order $O_p(1)$ and the second is $O_p(T^{-1/2})$. Therefore,

$$\lim_{T\to\infty} \mathbb{E}\left(D_T(\beta_0) D_T(\beta_0)'\right) = \lim_{T\to\infty} 4T^{-1}\mathbb{E}\left(\Pi' Z' Z \Pi\right)\sigma_u^2,$$

consistent with the order of $\Lambda_{1T}(\theta_0)$ in Assumption W. If β_0 is weakly identified, i.e., $\Pi = T^{-1/2}C$, then $D_T(\beta_0)$ is of exact order $O_p(T^{-1/2})$. Therefore the eigenvalue of $\mathbb{E}(D_T(\beta_0) D_T(\beta_0)')$ is of order $O(T^{-1})$, consistent with the order of $\Lambda_{2T}(\theta_0)$ in Assumption W.

Weak identification in a CU-GMM setting (Kleibergen, 2005). Consider inference based on the moment restriction $\mathbb{E}\phi_t(\theta_0) = 0$ with $\theta_0 \in R^m$. Without loss of generality, assume $\phi_t(\theta_0)$ is serially uncorrelated. Let $f_T(\theta) = T^{-1/2}\sum_{t=1}^T \phi_t(\theta)$. Then the CU-GMM criterion function is given by $Q_T(\theta) = f_T(\theta)' \hat{V}_{ff}(\theta)^{-1} f_T(\theta)$, where $\hat{V}_{ff}(\theta) \to^p V_{ff}(\theta) = \lim_{T\to\infty}\text{Var}(f_T(\theta))$. Define

$$\frac{\partial f_T(\theta_0)}{\partial \theta'} = q_T(\theta_0) = (q_{1,T}(\theta_0), ..., q_{m,T}(\theta_0)).$$

Kleibergen (2005) characterized the strength of identification using the order of $\mathbb{E}q_T(\theta_0)$. Under strong identification, $T^{-1/2}\mathbb{E}q_T(\theta_0)$ has a fixed full rank value while under weak identification $T^{-1/2}\mathbb{E}q_T(\theta_0) = T^{-1/2}C$. We have

$$\begin{aligned} D_T(\theta_0)' &= 2T^{-1/2} f_T(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1}\left(\hat{R}_T(\theta_0) - \mathbb{E}q_T(\theta_0)\right) && \text{(a)} \\ &\quad + 2T^{-1/2} f_T(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1}\mathbb{E}q_T(\theta_0), && \text{(b)} \end{aligned}$$

where the j-th column of $\hat{R}_T(\theta_0)$ equals $q_{j,T}(\theta_0) - \hat{V}_{\theta f,j}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0)$, i.e., the residual from projecting $q_{j,T}(\theta_0)$ onto $f_T(\theta_0)$; $\hat{V}_{\theta f,j}(\theta_0)$ is the sample covariance between $f_T(\theta_0)$ and $q_{j,T}(\theta_0)$. Thus

$$\mathbb{E}\left(D_T(\theta_0) D_T(\theta_0)'\right) = \mathbb{E}(a'a) + \mathbb{E}(b'b) + \mathbb{E}(a'b) + \mathbb{E}(b'a).$$

The first term $\mathbb{E}(a'a)$ is of order $O\left(T^{-1}\right)$ irrespective of the strength of identification. The second term $\mathbb{E}(b'b)$ is of exact order $O(1)$ under strong and $O(T^{-1})$ under weak identification, respectively.

The order of $\mathbb{E}(b'b) + \mathbb{E}(a'b)$ is always lower than that of $\mathbb{E}(b'b)$. Therefore, the eigenvalues of $\mathbb{E}(D_T(\theta_0) D_T(\theta_0)')$ are $O(1)$ under strong identification and $O(T^{-1})$ under weak identification, consistent with Assumption W in the paper.

Weak identification under a GMM setting (Stock and Wright, 2000). Consider the same setup as in the CU-GMM case, but with inference based on the following GMM criterion function

$$Q_T(\theta) = f_T(\theta)' W_T f_T(\theta),$$

where W_T is some consistent estimate of the optimal weighting matrix that is, without loss of generality, assumed to be non-random. Then,

$$\begin{aligned} D_T(\theta_0)' &= 2T^{-1/2} f_T(\theta_0)' W_T \left(q_T(\theta_0) - \hat{R}_T(\theta_0) \right) && \text{(c)} \\ & \quad + 2T^{-1/2} f_T(\theta_0)' W_T \left(\hat{R}_T(\theta_0) - \mathbb{E} q_T(\theta_0) \right) && \text{(d)} \\ & \quad + 2T^{-1/2} f_T(\theta_0)' W_T \mathbb{E} q_T(\theta_0). && \text{(e)} \end{aligned}$$

Simple algebra shows that the leading term in $\mathbb{E}(D_T(\theta_0) D_T(\theta_0)')$ is

$$\mathbb{E}(c'c) + \mathbb{E}(c'e) + \mathbb{E}(e'c) + \mathbb{E}(d'd) + \mathbb{E}(e'e).$$

The first four terms are always of order $O(T^{-1})$ irrespective of the strength of identification. The last term converges to a positive definite matrix under strong identification. Therefore, all the eigenvalues of $\mathbb{E}(D_T(\theta_0) D_T(\theta_0)')$ are of order $O(1)$ under strong identification.

Under weak identification, Stock and Wright (2000) assumed that θ admits a partition: $\theta = (\alpha', \beta')'$, such that α is weakly identified while β is strongly identified. Specifically, let $m_T(\alpha, \beta) = \mathbb{E} f_T(\alpha, \beta)$ and write $m_T(\alpha, \beta) = m_T(\alpha_0, \beta_0) + T^{-1/2} m_{1T}(\alpha, \beta) + m_{2T}(\beta)$ with $m_{1T}(\alpha, \beta) = T^{1/2}(m_T(\alpha, \beta) - m_T(\alpha_0, \beta))$ and $m_{2T}(\beta) = m_T(\alpha_0, \beta) - m_T(\alpha_0, \beta_0)$. Stock and Wright (2000) assumed (c.f. Assumption C in their paper)

$$m_{1T}(\alpha, \beta) \to m_1(\alpha, \beta) \text{ and } m_{2T}(\beta) \to m_2(\beta).$$

Let $C_T = \text{Diag}(T^{1/2} I_{\dim(\alpha)}, I_{\dim(\beta)})$, then

$$\begin{aligned} & C_T \mathbb{E}(e'e) C_T \\ = \; & 4T^{-1} C_T \mathbb{E} q_T(\theta_0)' W_T \mathbb{E}\left(f_T(\theta_0) f_T(\theta_0)' \right) W_T \mathbb{E} q_T(\theta_0) C_T \\ = \; & 4T^{-1} C_T \mathbb{E} q_T(\theta_0)' (W_T V_{ff}(\theta_0) W_T) \mathbb{E} q_T(\theta_0) C_T \\ \to \; & 4 \begin{bmatrix} \frac{\partial m_1(\alpha_0, \beta_0)}{\partial \alpha'} & \frac{\partial m_2(\beta_0)}{\partial \beta'} \end{bmatrix}' V_{ff}^{-1}(\theta_0) \begin{bmatrix} \frac{\partial m_1(\alpha_0, \beta_0)}{\partial \alpha'} & \frac{\partial m_2(\beta_0)}{\partial \beta'} \end{bmatrix}. \end{aligned}$$

The limit is a positive definite matrix. Therefore, in large samples, $\mathbb{E}(e'e)$ has $\dim(\beta)$ eigenvalues that are $O(1)$ and $\dim(\alpha)$ eigenvalues of order $O(T^{-1})$. So is $\mathbb{E}(D_T(\theta_0) D_T(\theta_0)')$.

Appendix B. Weak identification in a two-equation dynamic model

We consider a model whose frequency domain properties can be studied analytically to illustrate Assumption W. It consists of two equations:

$$\begin{aligned} r_t &= \gamma y_t + \beta \pi_t + u_t, \\ \pi_t &= \rho \pi_{t-1} + v_t, \end{aligned}$$

with

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim N \left(0, \begin{array}{cc} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{array} \right),$$

where the first equation is a monetary policy rule (Taylor, 1993) with y_t and π_t being deviations of GDP and inflation from their targets and the second equation describes the inflation dynamics. The error terms are serially uncorrelated but can be mutually correlated. Consequently, π_t can be correlated with u_t. The parameter of interest is β. To simplify the derivation, we assume that the relevant steady state parameters, as well as ρ, γ and σ_v^2, are known. The unknown parameter vector is therefore $\theta = (\beta, \sigma_u^2, \sigma_{uv})$.

Rewrite the model as

$$\begin{aligned} \widetilde{r}_t &= \beta \pi_t + u_t, \\ \pi_t &= \rho \pi_{t-1} + v_t \end{aligned} \tag{B.1}$$

with $\widetilde{r}_t = r_t - \gamma y_t$. It can then be viewed as a dynamic version of the limited information simultaneous equation model, in which π_t is the endogenous explanatory and π_{t-1} is the instrument. The parameter β is weakly identified if ρ is small. Intuitively, because there is little persistence in π_t, it is difficult to differentiate between systematic policy responses ($\beta\pi_t$) and random disturbances (u_t). Geometrically, it is possible to move θ along a certain direction such that the likelihood surface changes little. In the extreme case with $\rho = 0$, β becomes unidentified. Then, there exists a path along which the likelihood is completely flat (It turns out changing θ in the direction given by $(1, -2\sigma_{uv}, -\sigma_v^2)$ yields such a non-identification curve). Note that Kleibergen and Mavroeidis (2009b) considered a very similar model, where the issue of interest is to conduct inference in a GMM setting under potential weak identification.

We let $\rho = T^{-1/2}c$ with $c > 0$; other parameter values are independent of T. Let $W(\omega) = 1$ for all $\omega \in [-\pi, \pi]$. The following Lemma corresponds to Assumption W in the paper.

Lemma B.1 *Let θ_0 denote the true value of $\theta = (\beta, \sigma_u^2, \sigma_{uv})$, then $M_T(\theta_0)$ satisfies:*

1. *It has two positive eigenvalues λ_{1T} and λ_{2T} satisfying $T\lambda_{1T} \to \infty$ and $T\lambda_{2T} \to \infty$.*
2. *The smallest eigenvalue λ_{3T} satisfies*

$$T\lambda_{3T} \to \frac{16\pi^2 \sigma_v^4 c^2}{(1 + \sigma_v^4 + 4\sigma_{uv}^2)(\sigma_v^2 \sigma_u^2 - \sigma_{uv}^2)}.$$

3. The elements of

$$\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'} Q_T(\theta_0) \Lambda_T(\theta_0)^{-1/2}$$

are bounded and Lipschitz continuous in ω.

Note that Lemma B.1.1 corresponds to Assumption W(i), while B.1.2 corresponds to W(ii). B.1.3 is a stronger result than W(iv).

Proof of Lemma B.1. Rewrite (B.1) as

$$\begin{bmatrix} 1 & -\beta \\ 0 & 1-\rho L \end{bmatrix} \begin{pmatrix} \widetilde{r}_t \\ \pi_t \end{pmatrix} = \begin{pmatrix} u_t \\ v_t \end{pmatrix}.$$

Let $f_{\theta_0}(\omega)$ denote the spectral density of $(\widetilde{r}_t, \pi_t)$, then

$$f_{\theta_0}^{-1}(\omega) = \frac{2\pi}{\sigma_u^2\sigma_v^2 - \sigma_{uv}^2} \begin{bmatrix} \sigma_v^2 & -\left(1-\rho e^{-i\omega}\right)\sigma_{uv} - \beta\sigma_v^2 \\ -\left(1-\rho e^{i\omega}\right)\sigma_{uv} - \beta\sigma_v^2 & \left|1-\rho e^{-i\omega}\right|^2 \sigma_u^2 + 2\beta\sigma_{uv}\left(1-\rho\cos(\omega)\right) + \beta^2\sigma_v^2 \end{bmatrix}$$

and

$$\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'} = \frac{1}{2\pi\left|1-\rho e^{-i\omega}\right|^2} \begin{bmatrix} 2\sigma_{uv}\left(1-\rho\cos(\omega)\right) + 2\beta\sigma_v^2 & \left|1-\rho e^{-i\omega}\right|^2 & 2\beta\left(1-\rho\cos(\omega)\right) \\ \sigma_v^2 & 0 & 1-\rho e^{i\omega} \\ \sigma_v^2 & 0 & 1-\rho e^{-i\omega} \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus,

$$\begin{aligned} &\left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'}\right)^* \left(f_{\theta_0}^{-1}(\omega)' \otimes f_{\theta_0}^{-1}(\omega)\right) \frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'} \\ = \;& \frac{1}{\sigma_v^2\sigma_u^2 - \sigma_{uv}^2} \begin{bmatrix} \frac{2\sigma_v^4}{1+\rho^2-2\rho\cos(\omega)} & 0 & \frac{2\sigma_v^2(1-\rho\cos(\omega))}{1+\rho^2-2\rho\cos(\omega)} \\ 0 & \frac{\sigma_v^4}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2} & -\frac{2\sigma_v^2\sigma_{uv}}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2} \\ \frac{2\sigma_v^2(1-\rho\cos(\omega))}{1+\rho^2-2\rho\cos(\omega)} & -\frac{2\sigma_v^2\sigma_{uv}}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2} & \frac{2\left(\sigma_v^2\sigma_u^2+\sigma_{uv}^2\right)}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2} \end{bmatrix} \end{aligned}$$

Note that only the (1,1)-th, (1,3)-th and (3,1)-th elements of the matrix depend on ρ and ω. Taking the average for these terms, we have

$$\frac{1}{T}\sum_{j=1}^{T} \frac{2\sigma_v^4}{1+\rho^2-2\rho\cos(\omega_j)} = \frac{2\sigma_v^4}{1-\rho^2} + O(T^{-2}), \tag{B.2}$$

$$\frac{1}{T}\sum_{j=1}^{T} \frac{2\sigma_v^2\left(1-\rho\cos(\omega_j)\right)}{1+\rho^2-2\rho\cos(\omega_j)} = 2\sigma_v^2 + O(T^{-2}). \tag{B.3}$$

The result (B.2) can be proved using the Euler–Maclaurin formula. The latter states that for a generic function $g(x)$, the following relation between a summation and the corresponding integral holds:

$$\sum_{j=0}^{T} g(j) = \int_0^T g(x)dx + \frac{1}{2}(g(0) + g(T)) + \frac{1}{12}\left(g'(T) - g'(0)\right) + R,$$

where $g'(x)$ is the derivative of $g(x)$ and R is a remainder term satisfying $|R| \leq 1/(2\pi^2) \int_0^T |g''(x)|\, dx$. To obtain (B.2), let

$$g(j) = \frac{2\sigma_v^4}{1 + \rho^2 - 2\rho\cos(\omega_j)}.$$

Then, $g(0) + g(T) = 2g(0)$ and $g'(T) - g'(0) = 0$ because $g(j)$ is a periodic function with period T; $|R| = O(T^{-2})$ after some straightforward algebra. Therefore,

$$T^{-1}\sum_{j=1}^{T} g(j) = T^{-1}\int_0^T g(x)dx + O(T^{-2}) = \frac{1}{2\pi}\int_0^{2\pi} \frac{2\sigma_v^4}{1 + \rho^2 - 2\rho\cos(\omega)}d\omega + O(T^{-2}) = \frac{2\sigma_v^4}{1-\rho^2} + O(T^{-2}).$$

The term (B.3) can be analyzed in the same way. Thus,

$$\frac{1}{T}\sum_{j=1}^{T}\left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial\theta'}\right)^* (f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j))\frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial\theta'} \tag{B.4}$$

$$= \frac{1}{\sigma_v^2\sigma_u^2 - \sigma_{uv}^2}\begin{bmatrix} \frac{2\sigma_v^4}{1-\rho^2} & 0 & 2\sigma_v^2 \\ 0 & \frac{\sigma_v^4}{\sigma_v^2\sigma_u^2 - \sigma_{uv}^2} & -\frac{2\sigma_v^2\sigma_{uv}}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2} \\ 2\sigma_v^2 & -\frac{2\sigma_v^2\sigma_{uv}}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2} & \frac{2(\sigma_v^2\sigma_u^2+\sigma_{uv}^2)}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2}\end{bmatrix} + O(T^{-2}).$$

Because $\rho = T^{-1/2}c$ with $c > 0$, the right hand side can be written as

$$\frac{1}{\sigma_v^2\sigma_u^2 - \sigma_{uv}^2}\begin{bmatrix} 2\sigma_v^4 & 0 & 2\sigma_v^2 \\ 0 & \frac{\sigma_v^4}{\sigma_v^2\sigma_u^2 - \sigma_{uv}^2} & -\frac{2\sigma_v^2\sigma_{uv}}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2} \\ 2\sigma_v^2 & -\frac{2\sigma_v^2\sigma_{uv}}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2} & \frac{2(\sigma_v^2\sigma_u^2+\sigma_{uv}^2)}{\sigma_v^2\sigma_u^2-\sigma_{uv}^2}\end{bmatrix} \tag{I}$$

$$+\frac{1}{T}\left(\frac{1}{\sigma_v^2\sigma_u^2 - \sigma_{uv}^2}\right)\begin{bmatrix} \frac{2c^2\sigma_v^4}{1-\rho^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + O(T^{-2}) \tag{II}$$

Term (I) is independent of ρ and ω. Its rank is two, therefore having one zero eigenvalue and two strictly positive eigenvalues independent of T. The eigenvector corresponding to the zero eigenvalue is given by

$$\Gamma = \left(\frac{1}{\sqrt{1+\sigma_v^4+4\sigma_{uv}^2}}, \frac{-2\sigma_{uv}}{\sqrt{1+\sigma_v^4+4\sigma_{uv}^2}}, \frac{-\sigma_v^2}{\sqrt{1+\sigma_v^4+4\sigma_{uv}^2}}\right) \tag{B.5}$$

Term (II) has eigenvalues of order $O(T^{-1})$. Apply the matrix perturbation theory (e.g., Theorem 6.3.5 in Horn and Johnson, 2005, applied with A=Term (I) and E=Term (II)), (B.4) has two strictly positive eigenvalues and a third eigenvalue of order $O(T^{-1})$. This proves Lemma B.1.1.

We use the method of undetermined coefficients to prove Lemma B.1.2. The characteristic polynomial of (B.4), omitting the remainder term of order $O(T^{-2})$, is

$$x^3 - \frac{1}{(1-\rho^2)\left(\sigma_v^2\sigma_u^2 - \sigma_{uv}^2\right)^2}\left[\left(1-\rho^2\right)\left(\sigma_v^4 + 2\sigma_v^2\sigma_u^2 + 2\sigma_{uv}^2\right) + 2\sigma_v^4\left(\sigma_v^2\sigma_u^2 - \sigma_{uv}^2\right)\right]x^2$$
$$+\frac{2\sigma_v^4}{(1-\rho^2)\left(\sigma_v^2\sigma_u^2 - \sigma_{uv}^2\right)^3}\left[1 + \sigma_v^4 + 4\sigma_{uv}^2 - \rho^2(1 - 2\sigma_v^2\sigma_u^2 + 2\sigma_{uv}^2)\right]x - \frac{4\rho^2\sigma_v^8}{(1-\rho^2)\left(\sigma_v^2\sigma_u^2 - \sigma_{uv}^2\right)^4} = 0.$$

Let $x = b\rho^2 + o(\rho^2)$, then

$$x = \frac{2\sigma_v^4}{\left(1 + \sigma_v^4 + 4\sigma_{uv}^2\right)\left(\sigma_v^2\sigma_u^2 - \sigma_{uv}^2\right)}\rho^2 + o(\rho^2).$$

Multiplying x by $8\pi^2$ leads to the desired result.

Consider Lemma B.1.3 and let $Q_{jT}(\theta_0)$ denote the j-th column of $Q_T(\theta_0)$. It suffices to show that

$$\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'} Q_{jT}(\theta_0)\lambda_{jT}^{-1/2} \quad (j = 1, 2, 3)$$

satisfy the stated properties in the Lemma. This is obvious when $j = 1$ and 2, because λ_{1T} and λ_{2T} are strictly positive and finite in large samples, $Q_{1T}(\theta_0)$ and $Q_{2T}(\theta_0)$ are orthonormal vectors and the elements of $\partial \operatorname{vec} f_{\theta_0}(\omega)/\partial\theta'$ are bounded and Lipschitz continuous. For $j = 3$, rewrite the preceding display as

$$\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'}\Gamma'\lambda_{3T}^{-1/2} + \frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'}\left(Q_{3T}(\theta_0) - \Gamma'\right)\lambda_{3T}^{-1/2}$$

where Γ is defined by (B.5). The first term in the preceding display equals

$$\frac{c}{\sqrt{1 + \sigma_v^4 + 4\sigma_{uv}^2}\sqrt{T\lambda_{3T}}\left|1 - \rho e^{-i\omega}\right|^2}\begin{bmatrix} -2\sigma_{uv}\rho + 2\sigma_{uv}\cos(\omega) + 2\beta\sigma_v^2\cos(\omega) \\ \sigma_v^2 e^{i\omega} \\ \sigma_v^2 e^{-i\omega} \\ 0 \end{bmatrix},$$

which is finite and has finite derivatives with respect to ω for large T, therefore is lipschitz continuous. For the second term, because Γ is the eigenvector corresponding to the zero eigenvalue of Term (I), the result of Davis and Kahan (1970, p.11) implies $||Q_{3T}(\theta_0) - \Gamma'|| = O(T^{-1})$. Therefore,

$$\left\|\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'}\left(Q_{3T}(\theta_0) - \Gamma'\right)\lambda_{3T}^{-1/2}\right\| \leq \left\|\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'}\right\| \left\|Q_{3T}(\theta_0) - \Gamma'\right\| \left\|\lambda_{3T}^{-1/2}\right\| = O(T^{-1/2}),$$

which is asymptotically negligible. Finally, note that in (B.4), the summation is from $j = 1$ to $j = T$, while in Assumption W it is from $j = 1$ to $j = T - 1$. This difference does not affect the result. ■

Appendix C. Proofs of the main results

The following Lemma is needed for proving Theorem 1.

Lemma C.1 *Suppose Assumptions 1-4 and W hold. Let* $\Lambda_T^c(\theta_0)$ *denote the upper-left non-zero corner of* $\Lambda_T(\theta_0)$ *(i.e., the submatrix containing* $\Lambda_{1T}(\theta_0)$ *and* $\Lambda_{2T}(\theta_0)$*) and let* $Q_T^c(\theta_0)$ *be the corresponding orthonormal eigenvectors. Define*

$$\xi_T = 2\pi T^{-1/2} \sum_{j=1}^{T-1} \phi_T(\omega_j)^* \operatorname{vec}\left(I_T\left(\omega_j\right) - f_{\theta_0}(\omega_j)\right)$$

with

$$\phi_T(\omega) = W(\omega)\left(f_{\theta_0}^{-1}(\omega)' \otimes f_{\theta_0}^{-1}(\omega)\right)\left\{\left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta'}\right) Q_T^c(\theta_0)\Lambda_T^c(\theta_0)^{-1/2}\right\}.$$

Then

$$\xi_T \to^d N(0, \mathbb{I}_{q_1+q_2}),$$

where $\mathbb{I}_{q_1+q_2}$ *is a* (q_1+q_2)*-dimensional identity matrix.*

Proof of Lemma C.1. The proof uses similar arguments as in Dunsmuir (1979). The differences are that we allow for weak identification and selecting a subset of frequencies using $W(\omega)$. The proof consists of two steps. Step 1 proves asymptotic normality and Step 2 verifies that the limiting covariance matrix is an identity matrix.

Step 1. Rewrite ξ_T as

$$\xi_T = 2\pi T^{-1/2} \sum_{j=1}^{T-1} \phi_T(\omega_j)^* \operatorname{vec}\left(I_T\left(\omega_j\right) - \mathbb{E}I_T\left(\omega_j\right)\right) \tag{C.1}$$

$$+2\pi T^{-1/2} \sum_{j=1}^{T-1} \phi_T(\omega_j)^* \operatorname{vec}\left(\mathbb{E}I_T\left(\omega_j\right) - f_{\theta_0}(\omega_j)\right). \tag{C.2}$$

We now show that (C.2) is asymptotically negligible. As in Zygmund (1959), $\mathbb{E}I_T(\omega)$ can be expressed as the $(T-1)$-th order Cesaro sum of the Fourier series for $f_{\theta_0}(\omega)$:

$$\mathbb{E}I_T(\omega) = \sum_{s=-T+1}^{T-1}\left(1 - \frac{|s|}{T}\right)\Gamma(s)\exp\left(-is\omega\right)$$

where $\Gamma(s)$ are the Fourier coefficients (in this case the auto covariances of Y_t divided by 2π):

$$\Gamma(s) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{\theta_0}(\omega)\exp\left(is\omega\right)d\omega.$$

Using the property of the Cesaro sum and that $f_{\theta_0}(\omega)$ belongs to the Lipschitz class of degree β with respect to ω, we have (Hannan, 1970, p. 513)

$$\sup_{\omega\in[-\pi,\pi]} \left\|\operatorname{vec}\left(\mathbb{E}I_T(\omega) - f_{\theta_0}(\omega)\right)\right\| = O(T^{-\beta}).$$

The term (C.2) is therefore bounded by

$$2\pi T^{1/2}\sup_{\omega\in[-\pi,\pi]}\|\phi_T(\omega)\|\sup_{\omega\in[-\pi,\pi]}\left\|\mathrm{vec}\left(\mathbb{E}I_T\left(\omega\right)-f_{\theta_0}(\omega)\right)\right\|=O\left(T^{-\beta+1/2}\right)=o(1),$$

where the first equality is because $\phi_T(\omega)$ is finite by Assumption W and the last follows because $\beta>1/2$. Thus, to derive the limiting distribution of ξ_T, it suffices to consider (C.1) only.

The elements of $\phi_T(\omega)$ are deterministic functions of ω. Let $\phi_{T,M}(\omega)$ denote the $(M-1)$-th order Cesaro sum of the Fourier series for $\phi_T(\omega)$, i.e.,

$$\phi_{T,M}(\omega)=\sum_{s=-M+1}^{M-1}\left(1-\frac{|s|}{M}\right)\eta_T(s)\exp\left(-is\omega\right)$$

with

$$\eta_T(s)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_T(\omega)\exp\left(is\omega\right)d\omega.$$

Then,

$$\begin{aligned}(\mathrm{C.1}) &= 2\pi T^{-1/2}\sum_{j=1}^{T-1}\phi_{T,M}(\omega_j)^*\,\mathrm{vec}\left(I_T\left(\omega_j\right)-\mathbb{E}I_T\left(\omega_j\right)\right) \\ &\quad +2\pi T^{-1/2}\sum_{j=1}^{T-1}\left(\phi_T(\omega_j)-\phi_{T,M}(\omega_j)\right)^*\mathrm{vec}\left(I_T\left(\omega_j\right)-\mathbb{E}I_T\left(\omega_j\right)\right).\end{aligned} \tag{C.3}$$

We now show the second term on the right hand side is asymptotically negligible. Because of conjugacy, it suffices to prove

$$2\pi T^{-1/2}\sum_{j=1}^{[T/2]}\left(\phi_T(\omega_j)-\phi_{T,M}(\omega_j)\right)^*\mathrm{vec}\left(I_T\left(\omega_j\right)-\mathbb{E}I_T\left(\omega_j\right)\right)=o_p\left(1\right). \tag{C.4}$$

Here, some complication arises because $\phi_T(\omega)$ has a finite number of discontinuities within $[0,\pi]$ due to the presence of $W(\omega)$. This implies $\phi_T(\omega_j)-\phi_{T,M}(\omega_j)$ does not converge uniformly to zero over $[0,\pi]$ (the Gibbs phenomenon). However, results in Hannan (1970, p. 506-507) imply that $\phi_{T,M}(\omega)$ converges uniformly to $\phi_T(\omega)$ over all closed intervals excluding the jumps. At the jumps, the approximation errors are bounded. Assuming the jumps occur at $\widetilde{\omega}^k$ $(k=1,...,K)$, then, for any $\varepsilon>0$ there exist finite constants $M>0$ and $C>0$ independent of T, such that

$$\left\|\phi_{T,M}(\omega)-\phi_T(\omega)\right\|\leq C\qquad\text{if } \omega\in I_1\equiv\cup_{k=1}^K[\widetilde{\omega}^k-\varepsilon,\widetilde{\omega}^k+\varepsilon]$$

and

$$\left\|\phi_{T,M}(\omega)-\phi_T(\omega)\right\|\leq\varepsilon\qquad\text{if } \omega\in[0,\pi]\text{ but }\omega\notin I_1.$$

Apply the above partition and decompose the left side in (C.4) into

$$2\pi T^{-1/2}\sum_{j=1}^{[T/2]}1\left(\omega_j\in I_1\right)\left(\phi_T(\omega_j)-\phi_{T,M}(\omega_j)\right)^*\mathrm{vec}\left(I_T\left(\omega_j\right)-EI_T\left(\omega_j\right)\right) \tag{T1}$$

$$+2\pi T^{-1/2}\sum_{j=1}^{[T/2]}1\left(\omega_j\notin I_1\right)\left(\phi_T(\omega_j)-\phi_{T,M}(\omega_j)\right)^*\mathrm{vec}\left(I_T\left(\omega_j\right)-EI_T\left(\omega_j\right)\right) \quad (\mathrm{T2}),$$

The variance of the above quantity equals

$$Var(\mathrm{T1}) + Var(\mathrm{T2}) + \mathbb{E}(\mathrm{T1T2}^*) + \mathbb{E}(\mathrm{T2T1}^*). \tag{C.5}$$

To analyze its order, we use Theorem 11.7.1 in Brockwell and Davis (1991), i.e., for any ω_j and ω_h in $[0, \pi]$,

$$\mathbb{E}\left\{\mathrm{vec}\left(I_T(\omega_j) - \mathbb{E}I_T(\omega_j)\right)\mathrm{vec}\left(I_T(\omega_h) - \mathbb{E}I_T(\omega_h)\right)^*\right\} = \begin{cases} O(1) & \text{if } h = j, \\ O(T^{-1}) & \text{otherwise.} \end{cases}$$

Apply this result to the first term in (C.5):

$$\begin{aligned}
& \|Var(\mathrm{T1})\| \\
\leq\ & 4\pi^2 C^2 T^{-1} \sum_{j=1}^{[T/2]} 1(\omega_j \in I_1) \|Var\{\mathrm{vec}(I_T(\omega_j) - \mathbb{E}I_T(\omega_j))\}\| \\
& + 8\pi^2 T^{-1} C^2 \sum_{j=1}^{[T/2]} \sum_{h=1, h\neq j}^{[T/2]} 1(\omega_j \in I_1) \|\mathbb{E}\{\mathrm{vec}(I_T(\omega_j) - \mathbb{E}I_T(\omega_j))\mathrm{vec}(I_T(\omega_h) - \mathbb{E}I_T(\omega_h))^*\}\| \\
\leq\ & 4\pi^2 T^{-1} C^2 \sum_{j=1}^{[T/2]} 1(\omega_j \in I_1) D + 8\pi^2 C^2 T^{-2} \sum_{j=1}^{[T/2]} \sum_{h=1}^{[T/2]} 1(\omega_j \in I_1) D \\
\leq\ & 12\pi^2 C^2 D K \varepsilon,
\end{aligned}$$

where D is some finite constant. Because the length of the interval I_1 can be made arbitrarily small by choosing a small ε and a large M, we have $Var(\mathrm{T1}) = o(1)$. Similar arguments can be applied to the second term in (C.5):

$$\begin{aligned}
& \|Var(\mathrm{T2})\| \\
\leq\ & 4\pi^2 T^{-1} \varepsilon^2 \sum_{j=1}^{[T/2]} 1(\omega_j \notin I_1) \|Var\{\mathrm{vec}(I_T(\omega_j) - \mathbb{E}I_T(\omega_j))\}\| \\
& + 8\pi^2 T^{-1} \varepsilon^2 \sum_{j=1}^{[T/2]} \sum_{h=1, h\neq j}^{[T/2]} 1(\omega_j \notin I_1) \|\mathbb{E}\{\mathrm{vec}(I_T(\omega_j) - \mathbb{E}I_T(\omega_j))\mathrm{vec}(I_T(\omega_h) - \mathbb{E}I_T(\omega_h))^*\}\| \\
\leq\ & 4\pi^2 T^{-1} \varepsilon^2 \sum_{j=1}^{[T/2]} 1(\omega_j \notin I_1) D + 8\pi^2 \varepsilon^2 T^{-2} \sum_{j=1}^{[T/2]} \sum_{h=1}^{[T/2]} 1(\omega_j \notin I_1) D \\
\leq\ & 12\pi^3 D \varepsilon^2,
\end{aligned}$$

which can again be made small by choosing a small ε and a large M. Thus, $Var(\mathrm{T2}) = o(1)$. Similar arguments can be applied to the remaining two terms in (C.5), leading to $\mathbb{E}(\mathrm{T1T2}^*) = o(1)$ and $\mathbb{E}(\mathrm{T2T1}^*) = o(1)$. Combining the above results, we have proved (C.4).

It remains to analyze the first term on the right hand side in (C.3). Apply the definition of $\phi_{T,M}(\omega_j)$:

$$
\begin{aligned}
& 2\pi T^{-1/2}\sum_{j=1}^{T-1}\phi_{T,M}(\omega_j)^*\operatorname{vec}\left(I_T\left(\omega_j\right)-\mathbb{E}I_T\left(\omega_j\right)\right) \\
= & \ 2\pi T^{-1/2}\sum_{s=-M+1}^{M-1}\left(1-\frac{|s|}{M}\right)\eta_T(s)^*\left\{\sum_{j=1}^{T-1}\operatorname{vec}\left(I_T\left(\omega_j\right)-\mathbb{E}I_T\left(\omega_j\right)\right)\exp\left(is\omega_j\right)\right\} \\
= & \sum_{s=-M-1}^{M-1}\left(1-\frac{|s|}{M}\right)\eta_T(s)^*\left\{\sqrt{T}\operatorname{vec}\left(\hat{\Gamma}(s)-\mathbb{E}\hat{\Gamma}(s)\right)\right\} \qquad \text{(T3)} \\
& +\sum_{s=-M-1}^{M-1}\left(1-\frac{|s|}{M}\right)\eta_T(s)^*\left\{\sqrt{T}\operatorname{vec}\left(\hat{\Gamma}(s-T)-\mathbb{E}\hat{\Gamma}(s-T)\right)\right\}, \qquad \text{(T4)}
\end{aligned}
$$

where the last equality uses $\sum_{s=1}^{T}\exp\left(-is\omega_j\right)=0$ unless $s=kT$ $(k=0,\pm1,...)$. In the above, $\hat{\Gamma}(s)$ is defined as

$$
\begin{aligned}
\hat{\Gamma}(s) &= T^{-1}\sum_{t=1}^{T-s}\left(Y_{t+s}-\mu(\theta_0)\right)\left(Y_t-\mu(\theta_0)\right)' \quad \text{if } 0\le s\le T-1, \\
\hat{\Gamma}(s) &= \hat{\Gamma}(-s)' \quad \text{if } -T+1\le s\le 0.
\end{aligned}
$$

Term (T4) converges in probability to zero. This is because M is finite, $\eta_T(s)^*$ is uniformly bounded and $\sqrt{T}\operatorname{vec}(\hat{\Gamma}(s-T)-\mathbb{E}\hat{\Gamma}(s-T))\to^p 0$ for each $|s|<M$ by the definition of $\hat{\Gamma}(s-T)$ (note that the summation in the definition of $\hat{\Gamma}(s-T)$ involves at most M terms). In (T3), $\sqrt{T}\operatorname{vec}(\hat{\Gamma}(s)-\mathbb{E}\hat{\Gamma}(s))$ satisfies a central limit theorem for each $|s|\le M$, see Hannan (1976). Thus, (T3) also converges to a vector of normal random variables because M is finite. Therefore, ξ_T has a multivariate normal limiting distribution.

Step 2. We now examine the covariance matrix of (T3) by analyzing its elements separately. Note that (T3) is a q-by-1 vector. Apply the definition of $\eta_T(s)$ and use the relationship between the vec and the trace operator, its l-th element can be written as

$$
\xi_{TM}(l)=\frac{1}{2\pi}\sum_{s=-M-1}^{M-1}\left(1-\frac{|s|}{M}\right)\int_{-\pi}^{\pi}W(\omega)\operatorname{tr}\left\{B(l,\omega)\sqrt{T}(\hat{\Gamma}(s)-\mathbb{E}\hat{\Gamma}(s))\right\}\exp\left(-is\omega\right)d\omega, \qquad \text{(C.6)}
$$

where

$$
B(l,\omega)=f_{\theta_0}^{-1}(\omega)\left(\sum_{k=1}^{q}\frac{\partial f_{\theta_0}(\omega)}{\partial\theta_k}\left[Q_T^c(\theta_0)\Lambda_T^c(\theta_0)^{-1/2}\right]_{kl}\right)f_{\theta_0}^{-1}(\omega)
$$

with $[.]_{kl}$ denoting the (k,l)-th element of the matrix inside the bracket. Because $B(l,\omega)$ is an n_Y-by-n_Y matrix, (C.6) can be further rewritten as

$$
\xi_{TM}(l)=\frac{1}{2\pi}\sum_{s=-M-1}^{M-1}\left(1-\frac{|s|}{M}\right)\int_{-\pi}^{\pi}W(\omega)\left\{\sum_{a,b=1}^{n_Y}B_{ba}(l,\omega)\sqrt{T}\left(\hat{\Gamma}_{ab}(s)-\mathbb{E}\hat{\Gamma}_{ab}(s)\right)\right\}\exp\left(-is\omega\right)d\omega,
$$

where $B_{ba}(l,\omega)$ and $\hat{\Gamma}_{ab}(s)$ are the (b,a)-th and (a,b)-th element of the corresponding matrix. Therefore,

$$\begin{aligned} & Cov(\xi_{TM}(l), \xi_{TM}(k)) \\ = \; & \frac{1}{4\pi^2} \sum_{a,b,c,d=1}^{n_Y} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W(r) B_{ba}(l,r) W(\lambda) B_{dc}(k,\lambda)^* \sum_{s,h=-M+1}^{M-1} \left(1 - \frac{|s|}{M}\right)\left(1 - \frac{|h|}{M}\right) \\ & \times \mathbb{E}\left\{ T \left(\hat{\Gamma}_{ab}(s) - \mathbb{E}\hat{\Gamma}_{ab}(s)\right)\left(\hat{\Gamma}_{cd}(h) - \mathbb{E}\hat{\Gamma}_{cd}(h)\right)\right\} \exp\left(-isr\right) dr \exp\left(ih\lambda\right) d\lambda \qquad \text{(C.7)} \end{aligned}$$

The only random elements in (C.7) are the autocovariances, which satisfy (see eq. (3) in p.397 in Hannan, 1976)

$$\begin{aligned} & \mathbb{E}\left\{ T \left(\hat{\Gamma}_{ab}(s) - \mathbb{E}\hat{\Gamma}_{ab}(s)\right)\left(\hat{\Gamma}_{cd}(h) - \mathbb{E}\hat{\Gamma}_{cd}(h)\right)\right\} \\ \to \; & 2\pi \int_{-\pi}^{\pi} f_{ac}(\omega) \overline{f_{bd}(\omega)} \exp\left(-i(h-s)\omega\right) d\omega \qquad \text{(T5)} \\ & +2\pi \int_{-\pi}^{\pi} f_{ad}(\omega) \overline{f_{bc}(\omega)} \exp\left(i(s+h)\omega\right) d\omega, \qquad \text{(T6)} \end{aligned}$$

where $f_{ac}(\omega)$ stands for the (a,c)-th element of $f_{\theta_0}(\omega)$. We now approximate the covariances in (C.7) using (T5)+(T6). Applying (T5) leads to the following term

$$\begin{aligned} & 2\pi \int_{-\pi}^{\pi} \sum_{a,b,c,d=1}^{q} f_{ac}(\omega) \overline{f_{bd}(\omega)} \left(\int_{-\pi}^{\pi} W(r) B_{ba}(l,r) \left\{ \frac{1}{2\pi} \sum_{s=-M+1}^{M-1} \left(1 - \frac{|s|}{M}\right) \exp\left(-is(r-\omega)\right) \right\} dr \right) \\ & \times \left(\int_{-\pi}^{\pi} W(\lambda) B_{dc}(k,\lambda)^* \left\{ \frac{1}{2\pi} \sum_{h=-M+1}^{M-1} \left(1 - \frac{|h|}{M}\right) \exp\left(-ih(\omega-\lambda)\right) \right\} d\lambda \right) d\omega. \end{aligned}$$

Because the two terms inside the two curly brackets are Fejér's kernels, we have

$$\begin{aligned} \int_{-\pi}^{\pi} W(r) B_{ba}(l,r) \left\{ \frac{1}{2\pi} \sum_{s=-M+1}^{M-1} \left(1 - \frac{|s|}{M}\right) \exp\left(-is(r-\omega)\right) \right\} dr \quad & \to \quad W(\omega) B_{ba}(l,\omega) \\ \int_{-\pi}^{\pi} W(\lambda) B_{dc}(k,\lambda)^* \left\{ \frac{1}{2\pi} \sum_{h=-M+1}^{M-1} \left(1 - \frac{|h|}{M}\right) \exp\left(-ih(\omega-\lambda)\right) \right\} d\lambda \quad & \to \quad W(\omega) B_{dc}(k,\omega)^* \end{aligned}$$

uniformly over all closed intervals excluding the jumps. At the jumps, again the approximation error is finite, therefore does not interfere with the limiting results. The effect of applying (T6) can be analyzed similarly. Combining the two results, we have

$$
\begin{aligned}
& Cov(\xi_{TM}(l)\xi_{TM}(k)) \\
\to\ & 2\pi \int_{-\pi}^{\pi} \sum_{a,b,c,d=1}^{n_Y} W(\omega) \left\{ f_{ac}(\omega)\overline{f_{bd}(\omega)} B_{ba}(l,\omega) B_{dc}(k,\omega)^* + f_{ad}(\omega)\overline{f_{bc}(\omega)} B_{ba}(l,\omega) B_{dc}(k,-\omega)^* \right\} d\omega \\
=\ & 4\pi \int_{-\pi}^{\pi} W(\omega) \mathrm{tr} \left\{ f_{\theta_0}(\omega) B(k,\omega) f_{\theta_0}(\omega) B(l,\omega) \right\} d\omega \\
=\ & 4\pi \int_{-\pi}^{\pi} W(\omega) \,\mathrm{vec}\left(B(l,\omega)\right)^* \left(f_{\theta_0}(\omega)' \otimes f_{\theta_0}(\omega) \right) \mathrm{vec}\left(B(k,\omega)\right) d\omega \\
=\ & \frac{8\pi^2}{T} \sum_{j=1}^{T-1} \mathrm{vec}\left(B(l,\omega_j)\right)^* \left(f_{\theta_0}(\omega_j)' \otimes f_{\theta_0}(\omega_j) \right) \mathrm{vec}\left(B(k,\omega_j)\right) + o(1),
\end{aligned}
$$

where the first equality uses $B_{dc}(k,\omega)^* = B_{cd}(k,\omega)$, $\overline{f_{bd}(\omega)} = f_{db}(\omega)$, $\overline{f_{bc}(\omega)} = f_{cb}(\omega)$, $B_{dc}(k,-\omega)^* = B_{dc}(k,\omega)$, the second equality uses the relation between the trace and the vec operator, and the last equality is because the summand belongs to the Lipschitz class of degree $\beta > 1/2$. In matrix notation, the above result can be equivalently stated as

$$
Var\left(\xi_{TM}\right) = \Lambda_T^c(\theta_0)^{-1/2} Q_T^c(\theta_0)' M_T\left(\theta_0\right) Q_T^c(\theta_0) \Lambda_T^c(\theta_0)^{-1/2} = \mathbb{I}_{q_1+q_2},
$$

where the last equality uses the definition of $Q_T^c(\theta_0)$ and $\Lambda_T^c(\theta_0)$. ■

Proof of Theorem 1. Consider the first result. Because $M_T(\theta_0)$ is real and positive semi-definite, by the property of the Moore-Penrose Pseudoinverse (p.35 in Magnus and Neudecker, 2002), we have

$$
M_T^+(\theta_0) = Q_T^c(\theta_0) \Lambda_T^c(\theta_0)^{-1} Q_T^c(\theta_0)',
$$

where $\Lambda_T^c(\theta_0)$ and $Q_T^c(\theta_0)$ are defined as in the previous Lemma. Note that columns of $Q_T^c(\theta_0)$ are mutually orthogonal. Thus,

$$
\begin{aligned}
S_T(\theta_0) &= D_T\left(\theta_0\right)' Q_T^c(\theta_0) \Lambda_T^c(\theta_0)^{-1} Q_T^c(\theta_0)' D_T\left(\theta_0\right) \\
&= \left[\Lambda_T^c(\theta_0)^{-1/2} Q_T^c(\theta_0)' D_T\left(\theta_0\right) \right]' \left[\Lambda_T^c(\theta_0)^{-1/2} Q_T^c(\theta_0)' D_T\left(\theta_0\right) \right].
\end{aligned}
$$

Let

$$
\xi_T = \Lambda_T^c(\theta_0)^{-1/2} Q_T^c(\theta_0)' D_T\left(\theta_0\right).
$$

From the previous Lemma, $\xi_T \to^d N(0, \mathbb{I}_{q_1+q_2})$. Thus, $S_T(\theta_0) \to^d \chi^2_{q_1+q_2}$. The second result can be proved in the same way. The details are omitted. ■

Proof of Theorem 2. Consider $\mathcal{H}_{dT}(\theta_0)$. We first prove finite dimensional convergence and then verify tightness.[7] The term inside the norm can be equivalently represented as

$$
\Psi_T(r) = \left(\frac{T}{2}\right)^{-1/2} \sum_{j=1}^{[Tr/2]} \varphi_T\left(\omega_j\right)
$$

[7]See P. 37 in Billingsley (1968) for the definition of tightness.

with

$$\varphi_T(\omega) = \left(f_{\theta_0}^{-1/2}(\omega)' \otimes f_{\theta_0}^{-1/2}(\omega) \right) \text{vec}\left(I_T(\omega) - f_{\theta_0}(\omega) \right).$$

For a fixed r, the asymptotic normality of $\Psi_T(r)$ follows directly from Lemma C.1, by replacing $\phi_T(\omega)$ with $(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j))$. We now verify that the covariance matrix of $\Psi_T(r)$ has the desired structure. Note that $\varphi_T(\omega_j)$ are asymptotically independent in j, having zero mean and satisfying

$$\mathbb{E}\left(\varphi_T(\omega_j)\varphi_T(\omega_j)^*\right) = \mathbb{I}_{n_Y^2} + O(T^{-1/2}),$$

where the last equality follows from

$$\mathbb{E}\left(\text{vec}\left\{I_T(\omega_j) - f_{\theta_0}(\omega_j)\right\} \text{vec}\left\{I_T(\omega_j) - f_{\theta_0}(\omega_j)\right\}^*\right) = f_{\theta_0}(\omega_j)' \otimes f_{\theta_0}(\omega_j) + O(T^{-1/2}).$$

Therefore, for any fixed $r \in [0, 1]$,

$$\mathbb{E}\left(\Psi_T(r)\Psi_T(r)^*\right) = r\mathbb{I}_{n_Y^2} + O(T^{-1/2}). \tag{C.8}$$

Further, because

$$f_{\theta_0}(\omega_j)^{-1/2}\left(I_T(\omega_j) - f_{\theta_0}(\omega_j)\right) f_{\theta_0}(\omega_j)^{-1/2} \tag{C.9}$$

is a Hermitian matrix, the elements of $\Psi_T(r)$ take particular forms. The element is real valued if it corresponds to a diagonal entry in (C.9) and is complex valued otherwise. For a closer look, we consider the special case with $n_Y = 2$. Then, $\Psi_T(r)$ takes the form $(a_{11}, a_{21} + ib_{21}, a_{21} - ib_{21}, a_{22})'$, where a_{11}, a_{21}, b_{21} and a_{22} real numbers. Because of (C.8), we must have a_{11} and a_{22} converging to $N(0, 1)$ random variables and a_{21} and b_{21} converging to independent $N(0, 1/2)$ random variables. The case with a general n_Y follows similarly. Thus, we have established the finite dimensional convergence.

We now verify tightness, i.e., to prove that for any $\varepsilon > 0$, there exists constants C and T_0, such that

$$P\left(\mathcal{H}_{dT}(\theta_0) > C\right) \leq \varepsilon \text{ for all } T > T_0.$$

Apply Assumption 2, we have

$$I_T(\omega_j) - f_{\theta_0}(\omega_j) = H(\exp(-i\omega_j); \theta_0) I_\epsilon(\omega_j) H(\exp(-i\omega_j); \theta_0)^* - f_{\theta_0}(\omega_j) + R(\omega_j),$$

where $I_\epsilon(\omega_j)$ denotes the periodogram $\epsilon_t(\theta_0)$ at the frequency ω_j and $R(\omega_j)$ is a remainder term. Let $R_{kl}(\omega_j)$ denote the (k,l)-th element of $R(\omega_j)$, then Proposition 11.7.4 in Brockwell and Davis (1991, p. 445-446) implies

$$\max_{\omega_j \in [0,\pi]} \mathbb{E}\left|R_{kl}(\omega_j)\right|^2 = O(T^{-1}). \tag{C.10}$$

Applying the above decomposition, $\Psi_T(r)$ can be written as

$$\Psi_T(r) = \Psi_{T,1}(r) + \Psi_{T,2}(r)$$

with

$$\Psi_{T,1}(r) = \left(\frac{T}{2}\right)^{-1/2} \sum_{j=1}^{[Tr/2]} \left(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j)\right) \mathrm{vec}\left\{H(e^{-i\omega_j};\theta_0) I_\epsilon(\omega_j) H(e^{-i\omega_j};\theta_0)^* - f_{\theta_0}(\omega_j)\right\}$$

$$\Psi_{T,2}(r) = \left(\frac{T}{2}\right)^{-1/2} \sum_{j=1}^{[Tr/2]} \left(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j)\right) \mathrm{vec}\left(R(\omega_j)\right).$$

We now analyze the two terms separately. The summands of $\Psi_{T,1}(r)$ form a sequence of martingale differences. Apply a standard functional central limit theorem, we have

$$P\left(\sup_{r\in[0,1]} \|\Psi_{T,1}(r)\|_\infty > \frac{C}{2}\right) \le \varepsilon \text{ for some C and all } T > T_0.$$

For $\Psi_{T,2}(r)$, because of Assumption 3, there exists a finite constant D>0 such that

$$\left\|\mathrm{vec}\left(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j)\right)\right\|_\infty < D,$$

which implies

$$\|\Psi_{T,2}(r)\|_\infty \le \left(\frac{T}{2}\right)^{-1/2} D \sum_{k,l=1}^{n_Y} \sum_{j=1}^{[Tr/2]} |R_{kl}(\omega_j)|$$

because of the Cauchy-Schwarz inequality. Thus,

$$\begin{aligned} P\left(\sup_{r\in[0,1]} \|\Psi_{T,2}(r)\|_\infty > \frac{C}{2}\right) &\le P\left((T/2)^{-1/2} D \sum_{k,l=1}^{n_Y} \sum_{j=1}^{T/2} |R_{kl}(\omega_j)| > \frac{C}{2}\right) \\ &\le 16T^{-1}D^2 \frac{\sum_{k,l,u,v=1}^{n_Y} \sum_{j,h=1}^{T/2} \mathbb{E}\left(|R_{kl}(\omega_j)|\,|R_{uv}(\omega_h)|\right)}{C^2}, \end{aligned}$$

where the first inequality is because $|R_{kl}(\omega_j)|$ are nonnegative and the second is due to the Chebyshev inequality. Apply (C.10), the numerator in the preceding display is of order O(T). Therefore, the whole term is of order O(1), which can be made small by choosing a large C. The above results imply the tightness.

The second result follows from the same argument. The third result follows from a standard functional central limit theorem. For the fourth result, the independence between $G_d(r)$ and $G_s(r)$ is implied by the Normality (Assumption 4). The proof is complete. ■

Proof of Theorem 3. Consider $\mathcal{H}_{dT}(\theta_0)$. Let $\mathcal{H}_{dT}(\theta_0; r)$ denote $\mathcal{H}_{dT}(\theta_0)$ before taking the supremum, i.e.,

$$\mathcal{H}_{dT}(\theta_0; r) = \left\|(T/2)^{-1/2} \sum_{j=1}^{[Tr/2]} \mathrm{vec}\left\{f_{\theta_0}(\omega_j)^{-1/2}\left(I_T(\omega_j) - f_{\theta_0}(\omega_j)\right) f_{\theta_0}(\omega_j)^{-1/2}\right\}\right\|_\infty.$$

Then, for any fixed $r \in [0,1]$,

$$\begin{aligned}
& (T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0; r) \\
= & \left\| \frac{2}{T} \sum_{j=1}^{[Tr/2]} \text{vec}\left\{ f_{\theta_0}(\omega_j)^{-1/2} \left(f_0(\omega_j) - f_{\theta_0}(\omega_j) \right) f_{\theta_0}(\omega_j)^{-1/2} \right\} \right\|_\infty + o_p(1) \\
= & \left\| \frac{1}{\pi} \int_0^{\pi r} \text{vec}\left(f_{\theta_0}(\omega)^{-1/2} \left(f_0(\omega) - f_{\theta_0}(\omega) \right) f_{\theta_0}(\omega)^{-1/2} \right) d\omega \right\|_\infty + o_p(1), \qquad \text{(C.11)}
\end{aligned}$$

where the first equality is because of the law of large numbers and the second is due to the smoothness of the functions in ω.

Because $\|f_0(\omega) - f_{\theta_0}(\omega)\| > \delta$ for some ω, there exists a constant $C > 0$ such that

$$\left\| \text{vec}\left(f_{\theta_0}(\omega)^{-1/2} \left(f_0(\omega) - f_{\theta_0}(\omega) \right) f_{\theta_0}(\omega)^{-1/2} \right) \right\|_\infty > C\delta$$

holds for the same ω because of the positive-definiteness of $f_{\theta_0}(\omega)^{-1/2}$. By the property of the supremum norm, one of the elements of $\text{vec}\left(f_{\theta_0}(\omega)^{-1/2} \left(f_0(\omega) - f_{\theta_0}(\omega) \right) f_{\theta_0}(\omega)^{-1/2} \right)$ must have a modulus greater than $C\delta$. Without loss of generality, assume it is the first element and denote it by $\zeta_{\theta_0}(\omega)$. Then, because of the continuity in ω, there is an interval with positive radius on which $\zeta_{\theta_0}(\omega) > C\delta/2$. Denote this interval by $[\omega_L, \omega_U]$.

Consider (C.11) with $r = \omega_U/\pi$,

$$\begin{aligned}
& (T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0; \omega_U/\pi) \\
= & \left\| \frac{1}{\pi} \int_0^{\omega_U} \text{vec}\left(f_{\theta_0}(\omega)^{-1/2} \left(f_0(\omega) - f_{\theta_0}(\omega) \right) f_{\theta_0}(\omega)^{-1/2} \right) d\omega \right\|_\infty + o_p(1) \\
\geq & \left| \frac{1}{\pi} \int_0^{\omega_U} \zeta_{\theta_0}(\omega) d\omega \right| + o_p(1) \\
\geq & \frac{1}{\pi} \int_{\omega_L}^{\omega_U} \zeta_{\theta_0}(\omega) d\omega - \left| \frac{1}{\pi} \int_0^{\omega_L} \zeta_{\theta_0}(\omega) d\omega + \right| + o_p(1) \\
\geq & \frac{C\delta}{2\pi} (\omega_U - \omega_L) - \left| \frac{1}{\pi} \int_0^{\omega_L} \zeta_{\theta_0}(\omega) d\omega \right| + o_p(1) \\
\geq & \frac{C\delta}{4\pi} (\omega_U - \omega_L) - \left| \frac{1}{\pi} \int_0^{\omega_L} \zeta_{\theta_0}(\omega) d\omega \right|, \qquad \text{(C.12)}
\end{aligned}$$

where the first inequality uses the definition of $\zeta_{\theta_0}(\omega)$ and the property of the supremum norm, the second uses the triangle inequality, the third is because $\zeta_{\theta_0}(\omega)$ is greater than C$\delta/2$ over the interval and the last is because $\frac{C\delta}{4\pi}(\omega_U - \omega_L)$ is positive thus dominating the $o_p(1)$ term. We now apply (C.12) to find a lower bound for $(T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0)$. There are only two possibilities:

$$\begin{aligned}
\text{Case 1:} \quad & \left| \frac{1}{\pi} \int_0^{\omega_L} \zeta_{\theta_0}(\omega) d\omega \right| < \frac{C\delta}{8\pi} (\omega_U - \omega_L) \\
\text{Case 2:} \quad & \left| \frac{1}{\pi} \int_0^{\omega_L} \zeta_{\theta_0}(\omega) d\omega \right| \geq \frac{C\delta}{8\pi} (\omega_U - \omega_L).
\end{aligned}$$

In Case 1,

$$(T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0) = (T/2)^{-1/2} \sup_{r\in[0,1]} \mathcal{H}_{dT}(\theta_0; r) \geq (T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0; \omega_U/\pi) \geq \frac{C\delta}{8\pi}(\omega_U - \omega_L) > 0,$$

where the equality uses the definition of $\mathcal{H}_{dT}(\theta_0)$, the first inequality is because the supremum norm must be no less than any of its admissible values and the second inequality is because (C.12) and the definition of Case 1. In Case 2,

$$(T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0) = (T/2)^{-1/2} \sup_{r\in[0,1]} \mathcal{H}_{dT}(\theta_0; r) \geq (T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0; \omega_L/\pi) \geq \frac{C\delta}{8\pi}(\omega_U - \omega_L) > 0,$$

where the second inequality uses the definition of Case 2. Therefore, in both cases, $\mathcal{H}_{dT}(\theta_0) \to^p \infty$.

Now consider the order of $\mathcal{H}_{sT}(\bar{\theta}_0)$ under global alternatives.

$$\begin{aligned}
T^{-1/2}\mathcal{H}_{sT}(\bar{\theta}_0) &\geq \left\| f_{\theta_0}^{-1/2}(0) T^{-1} \sum_{j=1}^{T} (Y_t - \mu(\bar{\theta}_0)) \right\|_\infty \\
&\to^p \left\| f_{\theta_0}^{-1/2}(0)\left(\mu_0 - \mu(\bar{\theta}_0)\right) \right\|_\infty \\
&\geq \sqrt{n_Y^{-1} \left\| f_{\theta_0}^{-1/2}(0)\left(\mu_0 - \mu(\bar{\theta}_0)\right) \right\|^2} \\
&= \sqrt{n_Y^{-1} \left\| \left(\mu_0 - \mu(\bar{\theta}_0)\right)' f_{\theta_0}^{-1}(0)\left(\mu_0 - \mu(\bar{\theta}_0)\right) \right\|^2} \\
&> C,
\end{aligned}$$

where C is a positive constant and the last inequality follows because $f_{\theta_0}^{-1}(0)$ is positive definite. Therefore, $\mathcal{H}_{sT}(\bar{\theta}_0) \to^p \infty$. The property of $\mathcal{H}_T(\bar{\theta}_0)$ follows by combining the results for $\mathcal{H}_{sT}(\bar{\theta}_0)$ and $\mathcal{H}_{dT}(\bar{\theta}_0)$. ■

Table 1. Critical values for the test statistics

Test	Size	Dimension (n_Y)									
	(%)	1	2	3	4	5	6	7	8	9	10
	10	1.947	2.263	2.425	2.528	2.605	2.686	2.736	2.779	2.808	2.856
$\mathcal{H}_{dT}(\theta_0)$	5	2.217	2.509	2.654	2.747	2.813	2.897	2.944	2.977	3.000	3.054
	1	2.782	3.005	3.137	3.202	3.267	3.342	3.400	3.423	3.426	3.485
	10	1.947	2.209	2.361	2.459	2.548	2.613	2.647	2.705	2.735	2.770
$\mathcal{H}_{sT}(\bar{\theta}_0)$	5	2.217	2.471	2.616	2.703	2.789	2.844	2.881	2.937	2.961	2.991
	1	2.782	3.034	3.120	3.186	3.267	3.316	3.354	3.395	3.414	3.439
	10	2.212	2.482	2.627	2.719	2.793	2.865	2.909	2.951	2.976	3.017
$\mathcal{H}_{T}(\bar{\theta}_0)$	5	2.471	2.717	2.856	2.934	3.011	3.073	3.120	3.156	3.173	3.213
	1	2.988	3.220	3.335	3.384	3.464	3.508	3.582	3.596	3.600	3.643

Table 2. Parameter values along the non-identification curve

	ψ_1	ψ_2	ρ_r	$100\sigma_r$
θ_0	1.5	1.0	0.6	0.2
		Direction 1		
θ_1	1.836868445	0.900004035	0.598458759	0.199486255
θ_2	2.173736829	0.800008051	0.596905592	0.198968539
θ_3	2.510605180	0.700012143	0.595340369	0.198446796
θ_4	2.847473509	0.600016358	0.593762961	0.197920981
θ_5	3.184341764	0.500020520	0.592173193	0.197391053
θ_6	3.521209942	0.400024625	0.590570930	0.196856970
θ_7	3.858078165	0.300029089	0.588956076	0.196318654
θ_8	4.194946193	0.200033105	0.587328358	0.195776097
θ_9	4.531814220	0.100037328	0.585687731	0.195229219
θ_{10}	4.868682201	0.000041617	0.584034010	0.194677969
		Direction 2		
θ_{-1}	1.449287583	1.015053453	0.600230997	0.200077000
θ_{-2}	1.398575164	1.030106903	0.600461720	0.200153908
θ_{-3}	1.347862753	1.045160386	0.600692186	0.200230727
θ_{-4}	1.297150322	1.060213806	0.600922373	0.200307461
θ_{-5}	1.246437899	1.075267255	0.601152303	0.200384106
θ_{-6}	1.195725490	1.090320753	0.601381980	0.200460662
θ_{-7}	1.145013063	1.105374198	0.601611380	0.200537132
θ_{-8}	1.094300631	1.120427628	0.601840515	0.200613515
θ_{-9}	1.043588191	1.135481038	0.602069376	0.200689808
θ_{-10}	0.992875774	1.150534530	0.602297996	0.200766012

Note. θ_j represent equally spaced points taken from the non-identification curve extended from θ_0. Along Direction 1, the curve is truncated at the closest point to zero where ψ_2 is still positive. Along Direction 2, the curve is truncated at the last point yielding a determinate solution.

Table 3. Rejection frequencies under the null hypothesis

		BC frequencies		Full spectrum		Mean and full spectrum	
Level	T	$S_T(\theta_0)$	$\mathcal{H}^W_{dT}(\theta_0)$	$S_T(\theta_0)$	$\mathcal{H}_{dT}(\theta_0)$	$\bar{S}_T(\bar{\theta}_0)$	$\mathcal{H}_T(\bar{\theta}_0)$
			$\bar{\theta}_0$ taken from Table 2 in AS (2007)				
10%	80	0.080	0.096	0.125	0.092	0.135	0.078
	160	0.083	0.102	0.115	0.093	0.127	0.076
	240	0.091	0.099	0.138	0.103	0.140	0.079
	320	0.097	0.090	0.143	0.094	0.144	0.078
5%	80	0.058	0.067	0.086	0.055	0.095	0.051
	160	0.049	0.062	0.073	0.049	0.078	0.038
	240	0.049	0.056	0.073	0.050	0.078	0.036
	320	0.050	0.051	0.087	0.052	0.085	0.037
			$\bar{\theta}_0$ drawn from a prior distribution				
10%	80	0.086	0.102	0.135	0.099	0.156	0.100
	160	0.086	0.106	0.122	0.103	0.136	0.090
	240	0.101	0.099	0.142	0.106	0.145	0.085
	320	0.099	0.095	0.145	0.099	0.149	0.079
5%	80	0.056	0.067	0.088	0.063	0.102	0.061
	160	0.056	0.068	0.080	0.057	0.086	0.047
	240	0.052	0.053	0.083	0.052	0.086	0.043
	320	0.054	0.059	0.083	0.051	0.084	0.037

Note. The first panel: $\bar{\theta}_0$ is taken from the last column of Table 2 in An and Schorfheide (2007). The second panel: $\bar{\theta}_0$ is randomly drawn from the prior distribution specified in Table 2 in An and Schorfheide (2007); the values of ρ_r, ρ_g, ρ_z are fixed at their original values.

Table 4. Null rejection frequencies under alternative computations of the test statsitics

(All eigenvalues are treated as non-zeros)

Level	T	BC frequencies $S_T(\theta_0)$	Full spectrum $S_T(\theta_0)$	Mean and full spectrum $\bar{S}_T(\bar{\theta}_0)$
		$\bar{\theta}_0$ taken from Table 2 in AS (2007)		
10%	80	0.063	0.097	0.109
	160	0.065	0.085	0.100
	240	0.069	0.107	0.111
	320	0.067	0.108	0.113
5%	80	0.039	0.061	0.079
	160	0.033	0.055	0.061
	240	0.038	0.066	0.060
	320	0.039	0.068	0.067
		$\bar{\theta}_0$ drawn from a prior distribution		
10%	80	0.069	0.109	0.125
	160	0.066	0.094	0.108
	240	0.072	0.110	0.116
	320	0.070	0.116	0.123
5%	80	0.045	0.071	0.082
	160	0.037	0.059	0.069
	240	0.042	0.069	0.066
	320	0.040	0.072	0.067

Note. See Table 3.

Table 5. Null rejection frequencies under alternative computations of the test statsitics
(The two smallest eigenvalues are set to exact zeros)

		BC frequencies	Full spectrum	Mean and full spectrum
Level	T	$S_T(\theta_0)$	$S_T(\theta_0)$	$\bar{S}_T(\bar{\theta}_0)$
		$\bar{\theta}_0$ taken from Table 2 in AS (2007)		
10%	80	0.075	0.115	0.140
	160	0.089	0.121	0.115
	240	0.097	0.131	0.130
	320	0.092	0.137	0.142
5%	80	0.046	0.071	0.093
	160	0.051	0.072	0.070
	240	0.060	0.083	0.080
	320	0.053	0.084	0.090
		$\bar{\theta}_0$ drawn from a prior distribution		
10%	80	0.078	0.128	0.158
	160	0.096	0.131	0.126
	240	0.098	0.136	0.140
	320	0.093	0.138	0.146
5%	80	0.048	0.082	0.108
	160	0.055	0.078	0.074
	240	0.059	0.085	0.086
	320	0.053	0.086	0.091

Note. See Table 3.

Table 6. Rejection frequencies under the alternative hypothesis

		BC frequencies		Full spectrum		Mean and full spectrum	
Level	T	$S_T(\theta_0)$	$\mathcal{H}^W_{dT}(\theta_0)$	$S_T(\theta_0)$	$\mathcal{H}_{dT}(\theta_0)$	$\bar{S}_T(\bar{\theta}_0)$	$\mathcal{H}_T(\bar{\theta}_0)$
		Randomly perturb the elements of θ_0 (or $\bar{\theta}_0$) by 20%					
10%	80	0.230	0.231	0.444	0.381	0.467	0.353
	160	0.367	0.363	0.592	0.540	0.593	0.516
	240	0.404	0.399	0.672	0.629	0.708	0.632
	320	0.436	0.436	0.729	0.676	0.764	0.694
5%	80	0.190	0.180	0.388	0.308	0.409	0.279
	160	0.309	0.289	0.528	0.474	0.527	0.448
	240	0.346	0.334	0.610	0.560	0.652	0.578
	320	0.409	0.369	0.695	0.622	0.708	0.646
		Randomly perturb the elements of θ_0 (or $\bar{\theta}_0$) by 40%					
10%	80	0.463	0.458	0.663	0.634	0.709	0.657
	160	0.564	0.594	0.839	0.790	0.865	0.792
	240	0.623	0.641	0.873	0.850	0.918	0.865
	320	0.694	0.689	0.889	0.871	0.934	0.900
5%	80	0.429	0.422	0.593	0.577	0.655	0.605
	160	0.528	0.546	0.796	0.749	0.820	0.748
	240	0.565	0.583	0.852	0.816	0.895	0.834
	320	0.619	0.633	0.874	0.847	0.918	0.870

Note. $\bar{\theta}_0$ is taken from the last column of Table 2 in An and Schorfheide (2007).

Table 7. Power under alternative computations of the test statistics

(All eigenvalues are treated as non-zeros)

		BC frequencies	Full spectrum	Mean and full spectrum
Level	T	$S_T(\theta_0)$	$S_T(\theta_0)$	$\bar{S}_T(\bar{\theta}_0)$
		Randomly perturb the elements of $\bar{\theta}_0$ by 20%		
10%	80	0.203	0.411	0.436
	160	0.332	0.556	0.557
	240	0.367	0.637	0.679
	320	0.409	0.695	0.737
5%	80	0.170	0.358	0.382
	160	0.278	0.497	0.495
	240	0.323	0.578	0.619
	320	0.358	0.641	0.684
		Randomly perturb the elements of $\bar{\theta}_0$ by 40%		
10%	80	0.442	0.621	0.682
	160	0.542	0.816	0.842
	240	0.593	0.861	0.909
	320	0.649	0.880	0.926
5%	80	0.416	0.566	0.634
	160	0.508	0.769	0.798
	240	0.538	0.843	0.884
	320	0.580	0.867	0.907

Note. See Table 6

Table 8. Power under alternative computations of the test statistics

(The two smallest eigenvalues are set to exact zeros)

Level	T	BC frequencies $S_T(\theta_0)$	Full spectrum $S_T(\theta_0)$	Mean and full spectrum $\bar{S}_T(\bar{\theta}_0)$
		Randomly perturb the elements of $\bar{\theta}_0$ by 20%		
10%	80	0.231	0.446	0.467
	160	0.367	0.599	0.600
	240	0.410	0.682	0.716
	320	0.439	0.732	0.774
5%	80	0.187	0.387	0.406
	160	0.313	0.534	0.530
	240	0.355	0.622	0.657
	320	0.389	0.682	0.716
		Randomly perturb the elements of $\bar{\theta}_0$ by 40%		
10%	80	0.464	0.673	0.719
	160	0.572	0.842	0.870
	240	0.635	0.877	0.918
	320	0.708	0.889	0.935
5%	80	0.431	0.604	0.664
	160	0.529	0.803	0.828
	240	0.575	0.856	0.899
	320	0.633	0.875	0.919

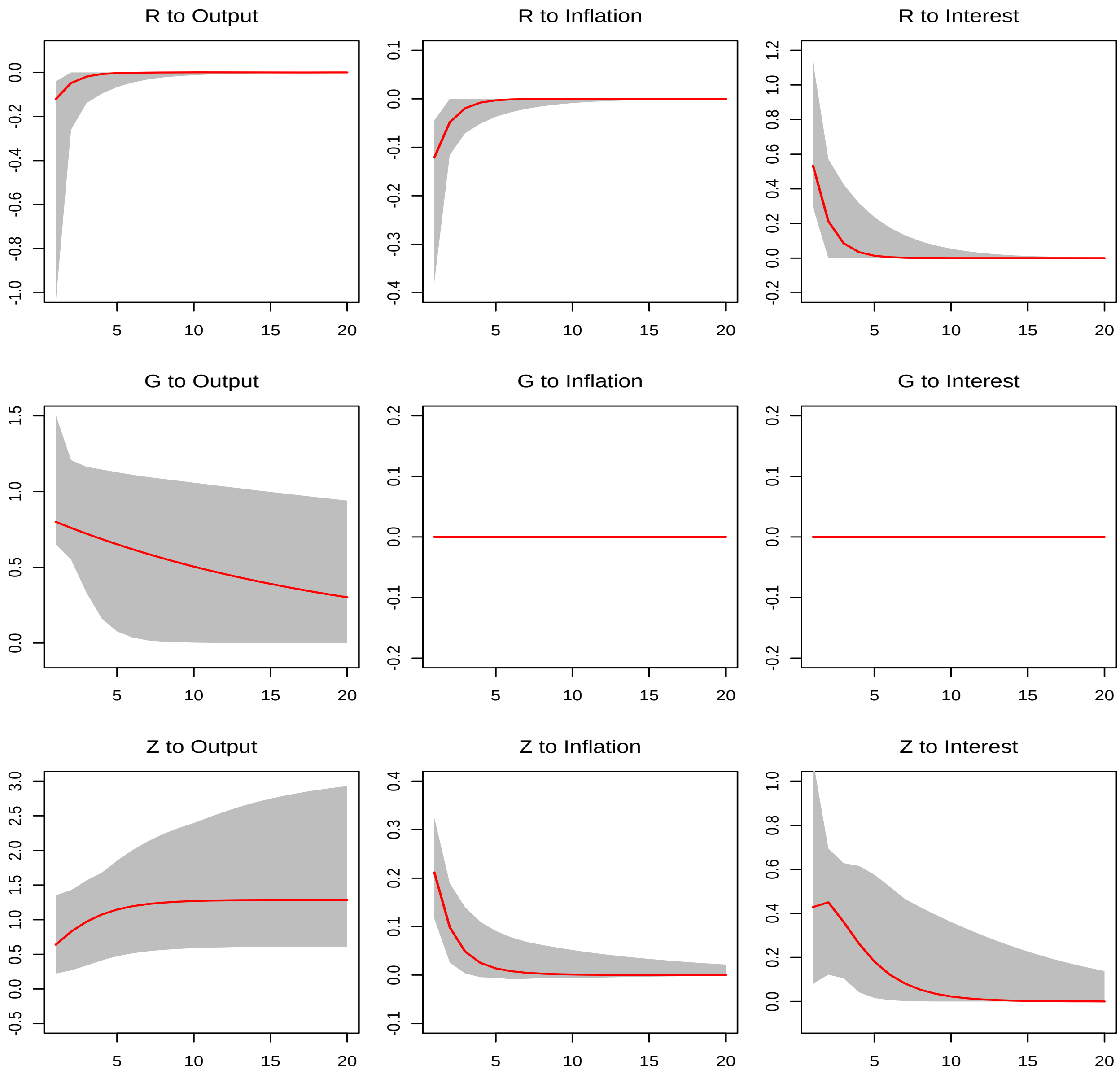
Note. See Table 6

Table 9. Lengths of the 90% confidence intervals

Parameter	$\bar{\theta}_0$	Bounds	BC frequencies	Full spectrum	Mean and full spectrum
τ	2	[1e-5, 5]	4.92, [0.05, 5.00]	4.17, [0.74, 4.99]	4.10, [0.85, 5.00]
κ	0.15	[0, 1]	0.63, [0.03, 0.66]	0.26, [0.08, 0.33]	0.27, [0.08, 0.34]
ψ_1	1.5	[0, 5]	4.16, [0.84, 5.00]	4.11, [0.89, 5.00]	4.03, [0.96, 5.00]
ψ_2	1.00	[0, 2]	2.00, [0.00, 2.00]	2.00, [0.00, 2.00]	2.00, [0.00, 2.00]
ρ_r	0.60	[0, 0.9]	0.74, [0.11, 0.89]	0.32, [0.46, 0.78]	0.34, [0.46, 0.79]
ρ_g	0.95	[0, 0.99]	0.44, [0.55, 0.99]	0.11, [0.88, 0.99]	0.10, [0.89, 0.99]
ρ_z	0.65	[0, 0.99]	0.49, [0.43, 0.93]	0.25, [0.56, 0.81]	0.26, [0.57, 0.82]
$100\sigma_r$	0.2	[1e-5, 2]	0.34, [0.14, 0.48]	0.08, [0.17, 0.26]	0.08, [0.17, 0.25]
$100\sigma_g$	0.8	[1e-5, 2]	0.66, [0.66, 1.32]	0.24, [0.71, 0.94]	0.24, [0.71, 0.94]
$100\sigma_z$	0.45	[1e-5, 2]	0.63, [0.23, 0.87]	0.30, [0.32, 0.63]	0.31, [0.32, 0.64]
$r^{(A)}$	0.4	[0, 5]	5.00, [0.00, 5.00]	5.00, [0.00, 5.00]	2.16, [0.00, 2.16]
$\pi^{(A)}$	4.00	[0, 20]	–	–	2.13, [3.23, 5.32]
$\gamma^{(Q)}$	0.50	[0, 5]	–	–	0.76, [0.00, 0.76]
Coverage	–	–	0.96	0.98	0.98

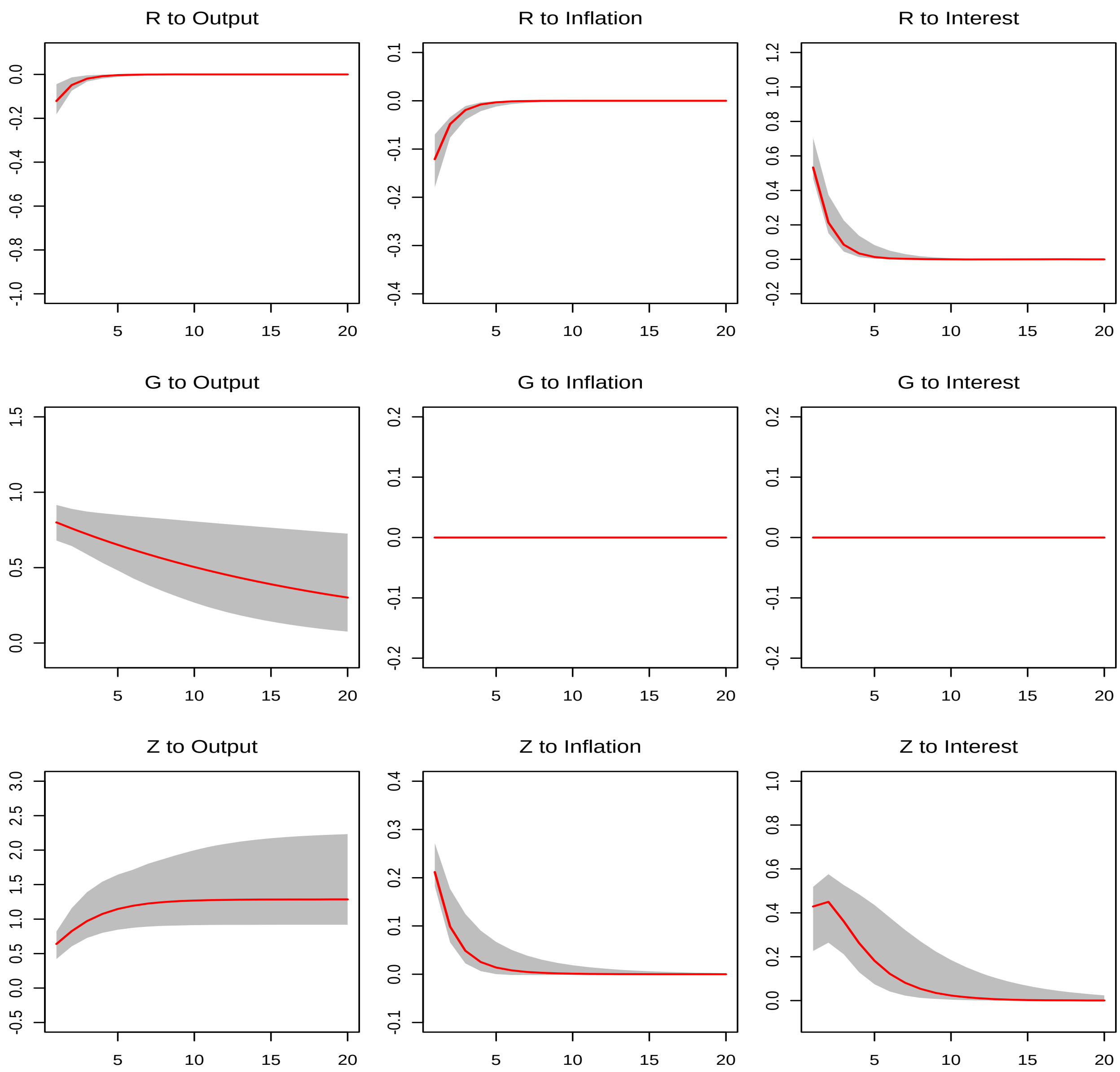
Note. The sample size is 240. Column 2: true parameter values. Column 3: bounds for permissible parameter values. Columns 4 to 6: lengths of the confidence intervals over 100 replications. In each cell, the first value is the median length of the intervals. The remaining two values are the medians of their lower and upper limits. The last row gives the frequencies that the confidence set contains the true parameter vector.

Figure 1. Uniform confidence bands for impulse response functions (90%, BC frequencies)



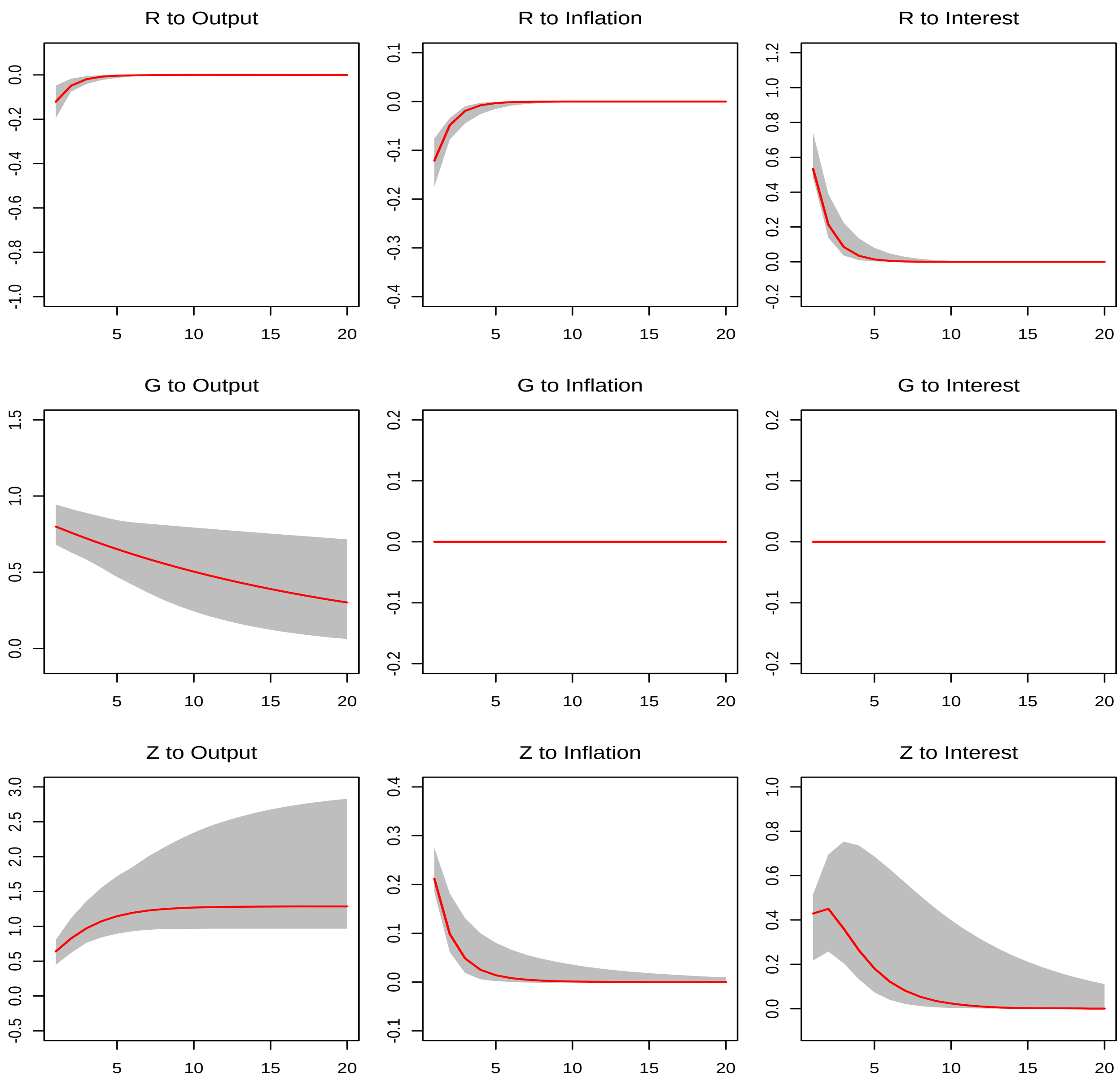
Note. R, G and Z: shocks to monetary policy, exogenous spending and technology. Gray area: the uniform band. Solid line: the true impulse response. Y-axis: percent. X-axis: horizon.

Figure 2. Uniform confidence bands for impulse response functions (90%, the full spectrum)



Note. R, G and Z: shocks to monetary policy, exogenous spending and technology. Gray area: the uniform band. Solid line: the true impulse response. Y-axis: percent. X-axis: horizon.

Figure 3. Uniform confidence bands for impulse response functions (90%, the mean and full spectrum)



Note. R, G and Z: shocks to monetary policy, exogenous spending and technology. Gray area: the uniform band. Solid line: the true impulse response. Y-axis: percent. X-axis: horizon.