

Inference in DSGE Models with Possible Weak Identification*

Zhongjun Qu[†]

Boston University

April 22, 2011; This version: September 18, 2013

Abstract

This paper considers inference in log-linearized DSGE models with weakly (including un-) identified parameters. The framework allows for analysis using only part of the spectrum, say at the business cycle frequencies. First, we characterize weak identification from a frequency domain perspective and propose a score test for the structural parameter vector based on the frequency domain approximation to the Gaussian likelihood. The construction heavily exploits the structures of the DSGE solution, the score function and the information matrix. In particular, we show the test statistic can be represented as the explained sum of squares from a complex-valued Gauss-Newton regression, where weak identification surfaces as (imperfect) multicollinearity. Second, we prove that asymptotically valid confidence sets can be obtained by inverting this test statistic and using Chi-square critical values. Third, we provide 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. Finally, a simulation experiment suggests that the test has adequate size even with relatively small sample sizes. It also suggests 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 are applicable to other dynamic models with well-defined spectra, such as stationary (factor-augmented) vector autoregressions.

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

JEL Classification: C12, C32, E1, E3.

*I would like to thank the editor Frank Schorfheide and three anonymous referees for detailed and valuable suggestions. I also thank Mingli Chen, Adam McCloskey, Marcelo Moreira, Serena Ng, Pierre Perron, seminar participants at BC, Columbia, MIT/Harvard, NC State, Queen's, Rutgers, UCSD and Yale for useful 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. A previous version of this paper circulated under the title "Inference and Specification Testing in DSGE Models with Possible Weak Identification".

[†]Department of Economics, Boston University, Boston, MA, 02215 (qu@bu.edu).

1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models play an important role in quantitative macroeconomics. Frequentist inference in such models is challenging. The likelihood surface can be flat or display near ridges over 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 (2012), concerning the parameters in a Taylor rule equation. The above features imply that the conventional framework for conducting inference, which relies on a \sqrt{T} -convergent, asymptotically normal estimator, can be very inadequate.

Recently, several studies have considered developing inferential procedures robust to weak identification. Guerron-Quintana, Inoue and Kilian (2013) obtained confidence sets by inverting the likelihood ratio test and the Bayes factor. Dufour, Khalaf and Kichian (2013) suggested inverting moment based tests. In on-going work, Andrews and Mikusheva (2013) study two LM tests (differing in how the information matrix is calculated) from a time domain perspective. An comparison with the latter is included in Section 7. In a related literature, Iskrev (2010), Komunjer and Ng (2011) and Qu and Tkachenko (2012) proposed rank conditions for local identification. They did not consider weak identification.

This paper develops 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 has two advantages. First, the information matrix is particularly simple to calculate in frequency domain. This leads to a simple and transparent test with a regression interpretation. Second, the researcher can choose desired frequencies for inference. This is valuable because DSGE models are designed to explain business cycle movements, not very long-run or very short-run fluctuations; see Del Negro, Diebold and Schorfheide (2008). The latter flexibility is difficult, if at all possible, to achieve in a time domain framework. Because of these two features, we obtain a unique test that is robust to both weak identification and frequency specific misspecifications.

We start by characterizing weak identification from a frequency domain perspective. The char-

acterizing condition involves the eigenvalues of the information matrix, some of which converge to zero as the sample size approaches infinity. A subset of eigenvalues is allowed to be exactly zero, permitting some parameters to be unidentified for any sample size. The condition is motivated by Rothenberg (1971) and can be viewed as a generalization of Corollary 1 in Qu and Tkachenko (2012). The latter shows that the parameters are identified from the population if and only if the information matrix (expressed as a function of the spectral density) has full rank.

We then propose a score test for the structural parameter vector. Two features underlie its robustness to weak identification. First, its normalization matrix (i.e., the information matrix) is computed directly from the model's solution by exploiting its vector linear structure. In fact, if this matrix was estimated from a finite sample then it would lead to size distortions because its dimension is typically high relative to the sample size. Second, the test statistic is related to the explained sum of squares in a complex-valued multivariate Gauss-Newton regression, where the regressors are non-random and are governed by the derivatives mentioned above. Irrespective of the strength of identification, the rank of the regressors matrix is always bounded by the dimension of the structural parameter vector. This provides intuition why the test statistic has a Chi-square limiting distribution with the degrees of freedom bounded by the dimension of the same vector; see Section 5. This is the only test with such a regression structure in the weak identification literature.

A confidence set for the structural parameter vector can be obtained by inverting this statistic. For implementation, we suggest a Metropolis algorithm. It mainly involves solving the model and computing the spectral density and its first order derivatives at different parameter values.

Impulse response function plays a central role for assessing the implications of a DSGE model. Building on the confidence set for the structural parameter vector, we propose a confidence band that covers this function with desired probability (i.e., a uniform band) even under weak identification. This is done by considering the envelope of the impulse response functions associated with all the parameter values contained in the joint confidence set. The same idea can be applied to obtain confidence bands for the time path of the variance decomposition, the individual spectrum and the absolute coherency. It can also be used to study certain low frequency hypotheses as those 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.

We evaluate the finite sample properties of the proposed procedures using a model studied in An and Schorfheide (2007). The test shows good size properties even with relatively small sample sizes. The results show that even unidentified parameters 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 tight 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. Canova and Sala (2009) also observed that wide confidence intervals for parameters can be accompanied by narrow bands for impulse responses for the minimum distance estimator they consider.

We also consider the test's properties when low frequency misspecifications are present (caused by an unmodeled structural change in the steady state growth rate and/or a smoothly varying inflation target). The results show that using the business cycle frequencies can lead to valid inference even when using the full spectrum erroneously rejects the null hypothesis 100% of the time. In practice, this offers researchers a choice. If the model is well specified at all frequencies, then all the frequencies should be used and the inference will be more precise. If the model is misspecified over some frequencies, then using parts of the spectrum is preferable.

This paper contributes to the literature that analyzes dynamic equilibrium models from a frequency domain perspective. 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 parameter 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. Tkachenko and Qu (2012) and Sala (2013) analyzed medium-scale DSGE models in the frequency domain. The current paper is the first to study identification robust inference from the frequency domain perspective.

The paper is structured as follows. Section 2 illustrates how to compute the spectral density. Section 3 presents the framework and the assumptions. Section 4 characterizes weak identification from a frequency domain perspective. Section 5 proposes the score test and studies its asymptotic properties. Section 6 discusses how to obtain robust confidence sets. It also considers uniform confidence bands for impulse response functions and some other objects. Section 7 considers an illustrative model and Section 8 concludes.

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

Suppose a DSGE model has been log-linearized around the steady state. Assume it has a unique stable solution. Then, the solution can be computed and represented in a variety of ways using the algorithms of Uhlig (1999), Klein (2000), King and Watson (2002) and Sims (2002). The methods developed in this paper can work with any of these representations. Given that the spectral density plays a central role in the analysis, below we illustrate how to compute it from Sims (2002).

Let θ denote the structural parameter vector. Sims (2002) 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 $\Gamma_0(\theta)$, $\Gamma_1(\theta)$, $\Psi(\theta)$ and $\Pi(\theta)$ are coefficients matrices, S_t includes endogenous variables (both observed and latent), conditional expectations and exogenous shocks (if they are serially correlated), ϵ_t is a vector of serially uncorrelated structural disturbances and η_t contains expectation errors. Under determinacy, its solution can be represented as $S_t = \Phi_1(\theta)S_{t-1} + \Phi_0(\theta)\epsilon_t$, or equivalently

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

Let $A(L)$ be a matrix of finite-order lag polynomials to specify the observables:

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

Its spectral density is given by

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

where the superscript “*” stands for the conjugate transpose, $\Sigma(\theta) = \text{Var}(\epsilon_t)$ and

$$H(L; \theta) = A(L)(I - \Phi_1(\theta)L)^{-1}\Phi_0(\theta). \quad (3)$$

As in the time domain, the above framework can easily handle models with latent endogenous variables and measurement errors. In the former, we simply assign zero entries in $A(L)$ to exclude the latent variables. For the latter, suppose $\zeta_t(\theta)$ are serially uncorrelated measurement errors independent of $Y_t^d(\theta)$ with covariance $\Sigma_\zeta(\theta)$. Then, the spectral density of $Y_t^d(\theta) + \zeta_t(\theta)$ is given by

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

where $H(\cdot)$ is defined by (3).

3 Setup and assumptions

Let $\{Y_t\}_{t=1}^T$ be a sample of random vectors. Assume there exists some (not necessarily unique) θ_0 such that

$$Y_t = \mu(\theta_0) + Y_t^d(\theta_0),$$

where $Y_t^d(\theta_0)$ denotes the solution (1) when $\theta = \theta_0$ and $\mu(\theta_0)$ is the mean of Y_t implied by the model's steady state. We require four assumptions. Let $\|x\|$ be the Euclidean norm of a vector x and $\|X\|$ be the vector induced norm of a matrix X . For a square matrix, let $\text{eig}(X)$ denote its eigenvalues as a vector. For a differentiable function $f_\theta \in \mathbb{R}^k$ of $\theta \in \mathbb{R}^q$, let $\partial f_{\theta_0}/\partial \theta'$ be the k -by- q matrix of partial derivatives evaluated at θ_0 .

Assumption 1. $\theta_0 \in \Theta \subset \mathbb{R}^q$ with Θ being compact.

Assumption 2. The solution is unique and is representable as

$$Y_t^d(\theta) = H(L; \theta)\epsilon_t(\theta) \quad \text{with} \quad H(L; \theta) = \sum_{j=0}^{\infty} h_j(\theta)L^j, \quad (4)$$

where $h_j(\theta)$ ($j = 0, \dots, \infty$) are real valued matrices and $\epsilon_t(\theta)$ are serially uncorrelated structural disturbances with a nonsingular covariance matrix $\Sigma(\theta)$.

Assumption 3. There exist $0 < C_L \leq C_U < \infty$ such that for all $\omega \in [-\pi, \pi]$ and all $\theta \in \Theta$:

- (i) $C_L \leq \text{eig}(f_\theta(\omega)) \leq C_U$;
- (ii) the elements of $f_\theta(\omega)$ belong to $\text{Lip}(\beta)$ with $\beta > \frac{1}{2}$ with respect to ω ;¹
- (iii) $\left\| \frac{\partial \text{vec } f_\theta(\omega)}{\partial \theta'} \right\| \leq C_U$ and the elements of $\frac{\partial \text{vec } f_\theta(\omega)}{\partial \theta'}$ belong to $\text{Lip}(\beta)$ with $\beta > \frac{1}{2}$ with respect to ω ;
- (iv) $\|\partial \mu(\theta)/\partial \theta'\| \leq C_U$.

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.

In Assumption 2, the dimensions of the variables and parameters are

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

¹Let $g(\omega)$ be a scalar valued function defined on an interval B . We say g belongs to $\text{Lip}(\beta)$ if there exists a finite constant M such that $|g(\omega_1) - g(\omega_2)| \leq M|\omega_1 - \omega_2|^\beta$ for all $\omega_1, \omega_2 \in B$.

This assumption allows for noninvertibility and is weaker than assuming $Y_t^d(\theta)$ follows a VAR.

Assumption 2 encompasses models with measurement errors. To see this, suppose we observe $Y_t = \mu(\theta_0) + Y_t^d(\theta_0) + \zeta_t(\theta_0)$, where $\zeta_t(\theta)$ are serially uncorrelated measurement errors with a nonsingular covariance matrix $\Sigma_\zeta(\theta)$ and $\mathbb{E}(\zeta_t(\theta)\epsilon'_s(\theta)) = 0$ for all t and s . Define $\epsilon_t^a(\theta) = (\epsilon_t(\theta)', \zeta_t(\theta)')$ and $H^a(L; \theta) = \sum_{j=0}^{\infty} h_j^a(\theta)L^j$ with $h_0^a(\theta) = [h_0(\theta), I_{n_Y}]$ and $h_j^a(\theta) = [h_j(\theta), 0_{n_Y}]$ for $j > 0$. Then, $Y_t - \mu(\theta_0)$ satisfies (4) with $H(L; \theta)\epsilon_t(\theta)$ replaced by $H^a(L; \theta)\epsilon_t^a(\theta)$. Therefore, the subsequent results apply automatically to models with measurement errors.

Assumption 3(i) requires the spectral density matrix being finite and nonsingular. If unit roots are present in the DSGE model, then it requires appropriately differencing the series prior to applying the methods. 3(ii) and 3(iii) assume the spectral density and its first derivatives are smooth in ω . They can be verified under more primitive conditions. Specifically, 3(ii) is satisfied if $\sum_{j=0}^{\infty} j^\beta \|h_j(\theta)\| \leq \infty$ (Hannan, 1970, p. 311-312). The latter holds because (1) implies $h_j(\theta)$ decays exponentially. Assumption 3(iii) is satisfied if $\sum_{j=0}^{\infty} j^\beta \|\partial \text{vec } h_j(\theta)/\partial \theta'\| \leq \infty$, which holds if $\|\partial \text{vec } \Phi_1(\theta)/\partial \theta'\| \leq M$ and $\|\partial \text{vec } \Phi_0(\theta)/\partial \theta'\| \leq M$ for some $M > 0$ and all $\theta \in \Theta$.

Assumption 4 requires normality. If it is violated, then the distribution of the proposed test will depend on nuisance parameters. This is because non-Gaussian features such as skewness or excess kurtosis will alter the information matrix. The possibility of relaxing of this assumption is discussed in Section 5. Note that this assumption is common in the DSGE literature.

In the next three sections, we consider inference on θ_0 based on the mean and the spectrum. The inference on the dynamic parameters (i.e., excluding the parameters affecting only the mean of Y_t) based on the full or parts of the spectrum is treated as a special case.

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 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., $\omega_j = 2\pi j/T$ ($j = 1, 2, \dots, T-1$). The discrete Fourier transforms and periodograms of Y_t at such frequencies are

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

At the zero frequency, let

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

Under Assumption 2, asymptotically $w_T(\omega_j)$ ($j = 1, 2, \dots, T-1$) have complex-valued multivariate normal distributions with densities (Hannan, 1970, p. 223-225)

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

while $w_{\theta,T}(0) \sim N(0, f_{\theta_0}(0))$. Because the Fourier transforms are asymptotically independent for $\omega_j + \omega_k \neq 2\pi$, an approximate log-likelihood function for θ , up a constant addition, is given by

$$-\frac{1}{2} \sum_{j=1}^{T-1} [\log \det (f_{\theta}(\omega_j)) + \text{tr} \{ f_{\theta}^{-1}(\omega_j) I_T(\omega_j) \}] - \frac{1}{2} [\log \det (f_{\theta}(0)) + \text{tr} \{ f_{\theta}^{-1}(0) I_{\theta,T}(0) \}]. \quad (5)$$

Hansen and Sargent (1993) originally derived (5) as an approximation to the time domain Gaussian likelihood and used it to understand the effect of seasonal adjustment on parameter estimation.

In this paper, we consider the following generalized version of (5):

$$\begin{aligned} L_T(\theta) &= -\frac{1}{2} \sum_{j=1}^{T-1} W(\omega_j) [\log \det (f_{\theta}(\omega_j)) + \text{tr} \{ f_{\theta}^{-1}(\omega_j) I_T(\omega_j) \}] \\ &\quad - \frac{1}{2} W(0) [\log \det (f_{\theta}(0)) + \text{tr} \{ f_{\theta}^{-1}(0) I_{\theta,T}(0) \}], \end{aligned} \quad (6)$$

where $W(\omega)$ is an indicator function that selects the desired frequencies for inference. In particular, to conduct inference using the second but not the first order properties, we set $W(0) = 0$ and $W(\omega_j) = 1$ for all $\omega_j \neq 0$. To conduct inference using only the business cycle frequencies (with periods of 6–32 quarters, see King and Watson, 1996), with quarterly observations we set $W(\omega_j) = 1$ if $\omega_j \in [\pi/16, \pi/3] \cup [5\pi/3, 31\pi/16]$ and 0 otherwise. The latter allows us to assess the model's business cycle implications without first filtering the data. The above flexibility is difficult, if at all possible, to achieve in the time domain.

4.2 Weak identification

This section characterizes weak identification from a frequency domain perspective. The characterizing condition is motivated by Rothenberg (1971) and Qu and Tkachenko (2012) and stated using the eigenvalues of the information matrix. Some eigenvalues approach zero as $T \rightarrow \infty$, such that the local curvature of the likelihood remains small in some directions in the presence of a large sample size.

Let the superscript “ r ” denote the transpose without taking the conjugate. The score of (6) is

$$D_T(\theta_0) = \frac{1}{2\sqrt{T}} \sum_{j=0}^{T-1} W(\omega_j) \left(\frac{\partial \text{vec } f_{\theta_0}(\omega_j)}{\partial \theta'} \right)^* \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j) \right) \text{vec} (I_T(\omega_j) - f_{\theta_0}(\omega_j)) \\ + \frac{1}{2\pi\sqrt{T}} W(0) \sum_{t=1}^T \frac{\partial \mu(\theta_0)'}{\partial \theta} f_{\theta_0}^{-1}(0) (Y_t - \mu(\theta_0)), \quad (7)$$

where $I_T(0) = I_{\theta_0, T}(0)$. Under normality, the information matrix is

$$M_T(\theta_0) = \frac{1}{2T} \sum_{j=0}^{T-1} W(\omega_j) \left(\frac{\partial \text{vec } f_{\theta_0}(\omega_j)}{\partial \theta'} \right)^* \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j) \right) \frac{\partial \text{vec } f_{\theta_0}(\omega_j)}{\partial \theta'} \\ + \frac{1}{2\pi} W(0) \frac{\partial \mu(\theta_0)'}{\partial \theta} f_{\theta_0}^{-1}(0) \frac{\partial \mu(\theta_0)}{\partial \theta'}. \quad (8)$$

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 mean zero and known variances. Expressions (7) and (8) further simplify if the goal is to conduct inference on dynamic parameters based on the full spectrum or the business cycle frequencies. The involved derivatives only need to be taken with respect to the dynamic parameters. The second term in the expressions will no longer be present. Also, the summations should start at $j = 1$ instead of $j = 0$.

Because $M_T(\theta_0)$ is real and symmetric, its eigen-decomposition always exists:

$$M_T(\theta_0) = Q_T(\theta_0) \Lambda_T(\theta_0) Q_T(\theta_0)^{-1}, \quad (9)$$

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, respectively.

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

$$[\partial \text{vec } f_{\theta_0}(\omega) / \partial \theta'] Q_T(\theta_0) \Lambda_T^+(\theta_0)^{1/2} \quad (10)$$

are finite and belong to $\text{Lip}(\beta)$ with $\beta > 1/2$ with respect to ω , where $\Lambda_T^+(\theta_0)^{1/2}$ is the square root of the Moore-Penrose Pseudoinverse of $\Lambda_T(\theta_0)$.

W(i)-W(iii) 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, 2012). $\Lambda_{2T}(\theta_0)$ imply directions that are weakly identified, while $\Lambda_{3T}(\theta_0)$ corresponds to directions along which the parameter values are observationally equivalent for any sample size. The condition is related to Corollary 1 in Qu and Tkachenko (2012), which shows θ_0 is locally identified if and only if the information matrix has full rank. (They allowed $f_{\theta_0}(\omega)$ to be singular. In the nonsingular special case, the above statement applies.) Here, the eigenvalues are sample size-dependent, therefore identification is no longer a zero-one phenomenon.

W(i)-W(iii) are related to the characterizing conditions in the IV and GMM literature (Staiger and Stock, 1997, Stock and Wright, 2000 and Kleibergen, 2005). This is illustrated along two dimensions in the on-line appendix. First, it is shown that the latter conditions can also be stated using the eigenvalues that measure the local curvature of the criterion functions. Next, it is shown using a two-equation model that the conditions in Staiger and Stock (1997) translate into W(i) to W(iii). Prior to our work, Guerron-Quintana, Inoue and Kilian (2013) also suggested using the local curvature of the likelihood to characterize weak identification. The key differences are that 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 $\partial \text{vec } f_{\theta_0}(\omega)/\partial \theta'$ in ω . The effect of $Q_T(\theta_0)$ is to map the row vectors of $\partial \text{vec } f_{\theta_0}(\omega)/\partial \theta'$ into a new coordinate system common to all $\omega \in [-\pi, \pi]$. The assumption thus requires $\partial \text{vec } f_{\theta_0}(\omega)/\partial \theta'$ to be well behaved in this new coordinate system. If θ_0 is strongly identified, the assumption is trivially satisfied. Under weak identification, it is less transparent because some entries in $\Lambda_T^+(\theta_0)^{1/2}$ diverge to infinity. In the on-line appendix, it is shown that this assumption is satisfied in the simple dynamic model. In more general models, when a formal justification is not possible, the assumption can still be inspected using a graphical procedure. Specifically, because the matrix (10) is non-random and is fully determined by the DSGE model, its components can be plotted as a function of the frequency index under any given θ . Although this is not a formal test, it can be quite informative about the smoothness and magnitudes of the elements of (10). This point will be further illustrated in Section 7.

5 A frequency domain score test

This section proposes a score test and studies its properties under weak identification. It also discusses its flexibility for allowing for low frequency misspecifications.

Define

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

where $M_T^+(\theta)$ denotes the Moore-Penrose Pseudoinverse of $M_T(\theta)$.

To better understand the properties of $S_T(\theta_0)$ under weak identification, we show that it can be given a regression interpretation. Consider the following complex-valued multivariate regression:

$$\mathcal{Y}_j = X_j \beta + U_j, \quad (j = 0, 1, \dots, T-1), \quad (11)$$

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 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

$$ESS = \sum_{j=0}^{T-1} \hat{\mathcal{Y}}_j^* \hat{\mathcal{Y}}_j = \left(\sum_{j=0}^{T-1} \mathcal{Y}_j^* X_j \right) \left(\sum_{j=0}^{T-1} X_j^* X_j \right)^+ \left(\sum_{j=0}^{T-1} X_j^* \mathcal{Y}_j \right). \quad (12)$$

To establish the relation between ESS and $S_T(\theta_0)$, we need the following notation. Let H be a positive definite Hermitian matrix (e.g., $H = f_{\theta_0}(\omega)$). Then, H has 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 U is unitary. Define $H^{1/2} = U \Lambda^{1/2} U^*$ and $H^{-1/2} = U \Lambda^{-1/2} U^*$. Then, $H^{1/2}$, $H^{-1/2}$, $H' \otimes H$, $(H^{1/2})' \otimes H^{1/2}$ and $(H^{-1/2})' \otimes H^{-1/2}$ are Hermitian (Horn and Johnson, 2006, p. 243).

Lemma 1 *Under Assumption 3, we have $S_T(\theta_0) = (1/2)ESS$ with*

$$\begin{aligned} X_j &= \begin{bmatrix} W(\omega_j) \left(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j) \right) \frac{\partial \text{vec } f_{\theta_0}(\omega_j)}{\partial \theta'} \\ W(0) (\pi f_{\theta_0}(0))^{-1/2} \frac{\partial \mu(\theta_0)}{\partial \theta'} \end{bmatrix}, \\ \mathcal{Y}_j &= \begin{bmatrix} W(\omega_j) \left(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j) \right) \text{vec} (I_T(\omega_j) - f_{\theta_0}(\omega_j)) \\ W(0) (\pi f_{\theta_0}(0))^{-1/2} T^{-1} \sum_{t=1}^T (Y_t - \mu(\theta_0)) \end{bmatrix}, \end{aligned} \quad (13)$$

for $j = 0, 1, \dots, T-1$, where $I_T(0) = I_{\theta_0, T}(0)$.

In the Lemma, the matrix X_j is nonrandom and of dimension $(n_Y^2 + n_Y) \times q$. Its components $(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j))$ and $(\pi f_{\theta_0}(0))^{-1/2}$ are scaling factors. They are invariant to the strength of identification. The identification strength is embedded in $\partial \text{vec } f_{\theta_0}(\omega_j) / \partial \theta'$ and $\partial \mu(\theta_0) / \partial \theta'$. If

some parameters are weakly identified, then by Assumption W there exists a vector $c(\theta_0)$ such that $[\partial\mu(\theta_0)/\partial\theta'] c(\theta_0) = O(T^{-1/2})$ and $[\partial \text{vec } f_{\theta_0}(\omega_j)/\partial\theta'] c(\theta_0) = O(T^{-1/2})$ for all $j = 0, \dots, T-1$. The columns of X_j are thus (imperfectly) multicollinear. In the extreme case when some parameters are unidentified, we have $[\partial\mu(\theta_0)/\partial\theta'] c(\theta_0) = 0$ and $[\partial \text{vec } f_{\theta_0}(\omega_j)/\partial\theta'] c(\theta_0) = 0$ for all $j = 0, \dots, T-1$. Consequently, X_j exhibits perfect multicollinearity.

Asymptotically, \mathcal{Y}_j has mean zero with an identity covariance matrix. It is uncorrelated with X_j because the latter is nonrandom. Therefore, the explained sum of squares ESS is naturally expected to be related to a Chi-square limiting distribution with the degrees of freedom determined by the column rank of X_j , which can be smaller than q if some parameters are unidentified. Here, $S_T(\theta_0)$ equals $(1/2)ESS$ but not ESS because $I(2\pi - \omega) = \overline{I(\omega)}$. Note that X_j being nonrandom plays an important role in achieving the robustness to weak identification.

The insight that score tests can in some occasions 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 , (11) is a complex-valued Gauss-Newton regression. Davidson and MacKinnon (1993, Chapter 6) provided a detailed discussion of Gauss-Newton regressions applied to estimation and hypothesis testing. The current paper is the first that uses such a relationship to understand testing procedures under weak identification.

Theorem 1 *Let Assumptions 1-4 and W hold. Then,*

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

where χ_r^2 is a Chi-square variable with r degrees of freedom, $r = q - q_3$, $q = \dim(\theta_0)$ and $q_3 = \dim(\Lambda_{3T}(\theta_0))$.

Normality is a key assumption for Theorem 1 to hold. If it is violated, the distribution of $S_T(\theta_0)$ will depend on nuisance parameters. This is because the variance of the score $D_T(\theta_0)$ is no longer $M_T(\theta_0)$, but rather dependent on the third and fourth moments of the structural shocks. Specifically, its (j, l) -th element equals

$$\begin{aligned} [M_T(\theta_0)]_{jl} + \left(\frac{1}{4\pi}\right)^2 \sum_{a,b,c,d=1}^{n_\epsilon} \kappa_{abcd} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial\theta_j} H(\omega) d\omega \right]_{ab} \\ \times \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial\theta_l} H(\omega) d\omega \right]_{cd} + \frac{1}{8\pi^2} (A_{jl} + A_{lj}) \end{aligned} \quad (14)$$

where $[\cdot]_{ab}$ denotes the (a,b) -th element of the matrix, $[\cdot]_c$ denotes the c -th element of the vector, $M_T(\theta_0)$ is given by (8), κ_{abcd} is the fourth cross cumulant of $\epsilon_{ta}, \epsilon_{tb}, \epsilon_{tc}$ and ϵ_{td} , $H(\omega) =$

$H(\exp(-i\omega); \theta_0) = \sum_{j=0}^{\infty} h_j(\theta_0) \exp(-i\omega j)$, $H^*(\omega)$ is its conjugate transpose, $\partial f_{\theta_0}^{-1}(\omega)/\partial \theta_j$ is the derivative of $f_{\theta_0}^{-1}(\omega)$ with respect to the j -th element of θ and

$$A_{jl} = \sum_{a,b,c=1}^{n_\epsilon} \xi_{abc} \left\{ \int_{-\pi}^{\pi} W(\omega) \left[H^*(\omega) \frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_j} H(\omega) \right]_{ab} d\omega \right\} \times \left[W(0) \frac{\partial \mu(\theta_0)'}{\partial \theta_l} f_{\theta_0}^{-1}(0) H(0) \right]_c$$

with $\xi_{abc} = \mathbb{E}(\epsilon_{ta}\epsilon_{tb}\epsilon_{tc})$. Note that the term $(A_{jl} + A_{lj})$ is absent when the inference concerns only the dynamic parameters.

If the DSGE model fully specifies the distributions of the shocks, then κ_{abcd} and ξ_{abc} can be calculated as functions of the structural parameters. For example, if the shocks follow t distributions, then κ_{abcd} can be expressed as a function of the degrees of freedom parameters and $A_{jl} = A_{lj} = 0$. In such a situation, a modified test statistic can be constructed as

$$\tilde{S}_T(\theta_0) = D_T(\theta_0)' \tilde{M}_T^+(\theta_0) D_T(\theta_0), \quad (15)$$

where $\tilde{M}_T(\theta_0)$ is given by (14). Its limiting distribution is the same and can be established using similar arguments as in Theorem 1. Its size and power properties are illustrated in Section 7.

5.1 Implementing the test

Theorem 1 suggests the following procedure for inference.

- Apply an eigenvalue decomposition to $M_T(\theta_0)$ to determine $\dim(\Lambda_{3T}(\theta_0))$. We suggest using the MATLAB default tolerance level: $tol = \dim(M_T(\theta_0)) \text{eps}(\|M_T(\theta_0)\|)$, where $\text{eps}(\|M_T(\theta_0)\|)$ equals the machine precision (2^{-52}) times the maximum eigenvalue of $M_T(\theta_0)$.
- Set the eigenvalues below tol 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 (9). Use this $M_T(\theta_0)$ and the original $D_T(\theta_0)$ to compute $S_T(\theta_0)$.
- Reject the null hypothesis of $\theta = \theta_0$ if $S_T(\theta_0)$ exceeds the critical value of the χ_r^2 distribution.

It is desirable to let the tolerance level depend on $M_T(\theta_0)$, say its maximum eigenvalue or the average of its eigenvalues. Also, it is important to set the eigenvalues below tol to exact zero. This ensures that the column rank of X_j will be exactly r . Otherwise, over-rejection may occur when the χ_r^2 distribution is used for inference. The above procedure exploits the feature that $M_T(\theta_0)$ is nonrandom. Without this feature, the eigenvalues would be sample dependent and the rank estimation in general would not work.

The procedure involves choosing a tolerance level for deciding the rank of $M_T(\theta_0)$. This introduces some arbitrariness. The following two step method, due to Qu and Tkachenko (2012, p.120), can be used to reduce the arbitrariness and to improve the robustness of the rank estimator.

Step 1. Compute the ranks of $M_T(\theta_0)$ using a range of tolerance levels. Locate the outcomes with the smallest rank.

Step 2. Derive the non-identification curves conditioning on the smallest rank reported. (Under correct rank estimation, the curves consist of parameter values observationally equivalent to θ_0 , i.e., with $\mu(\theta) = \mu(\theta_0)$ and $f_\theta(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$. See Section 3.1 in Qu and Tkachenko, 2012, for the definition and computation of non-identification curves.) Compute $\mu(\theta)$ and $f_\theta(\omega)$ using the values on the curves and measure their differences from $\mu(\theta_0)$ and $f_{\theta_0}(\omega)$ over $\omega \in [-\pi, \pi]$.

The purpose of Step 1 is to avoid over-estimating the rank. It may result in under-estimation, which is further addressed in Step 2. The idea is, if this has occurred, then some curves reported in Step 2 will in fact correspond to parameter subsets that are locally identifiable. Consequently, noticeable discrepancies should emerge as we move along such curves away from θ_0 . There, the discrepancies can be interpreted in light of the magnitude of $f_{\theta_0}(\omega)$ and the distance of θ from θ_0 . This is typically more straightforward than interpreting the magnitudes of the eigenvalues of $M_T(\theta_0)$. In the application in Section 7.1, we consider tolerance levels between $10^{-10} \times tol$ and $10^{10} \times tol$. The results show consistently that the Taylor rule parameters are not identified.

In some occasions, the result might be inclusive even after implementing the above method. Then, we can set the degrees of freedom to $\dim(\theta_0)$. This leads to a conservative test without affecting its asymptotic validity because of the regression interpretation (11).

As an additional feature, the test continues to have correct size if the rank is under-estimated. 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. This does not alter the 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 are illustrated in Section 7 (see Tables 3, 4, 6 and 7).

5.2 Robustness to low frequency misspecification

DSGE models are often designed to explain business cycle movements, not very long run or very short run fluctuations. At the latter frequencies, such models can be severely misspecified. Schorfheide (2010) emphasized that "many time series exhibit low frequency behavior that is difficult, if not impossible, to reconcile with the model being estimated. This low frequency misspeci-

fication contaminates the estimation of shocks and thereby inference about the sources of business cycle". Therefore, it is valuable to be able to conduct inference excluding the very low frequencies, and more generally to compare inferential results when different sets of frequencies are allowed to enter. The proposed procedure allows for such investigations through the specification of $W(\omega)$. Later in Section 7.7, we explicitly consider two types of low frequency misspecifications. In one case, the growth rate of productivity is affected by a structural change at an unknown date. In the other, the inflation target has an unmodeled smoothly varying trend. The results show that using the business cycle frequencies can lead to valid inference even when using the full spectrum erroneously rejects the null hypothesis 100% of the time.

Recently, several studies have analyzed the effect of low frequency misspecifications on parameter estimation from a frequency domain perspective. Perron and Qu (2010) considered a weakly stationary process (e.g., ARMA models) affected by occasional level shifts. They showed in Proposition 3 that the level shift component affects the periodogram only up to $j = O(T^{1/2})$. Thus, out of a total of $T - 1$ Fourier frequencies, only an asymptotically negligible fraction of $O(T^{-1/2})$ is distorted. Qu (2011, Lemma 1) obtained a similar result for a stationary process affected by a smoothly varying trend. These results suggest that consistent parameter estimation is possible under such misspecifications by judiciously excluding a number of frequency components. McCloskey and Hill (2013) obtained such estimators for ARMA, GARCH and stochastic volatility models. Tkachenko and Qu (2012) analyzed Smets and Wouters' (2007) model using only the business cycle frequencies and compared with results obtained using the full spectrum. They found notably different parameter values and impulse response functions. The work here further develops this literature by simultaneously allowing for the selection of frequencies and weak identification.

6 Confidence sets robust to weak identification

Because of the duality between confidence sets and hypothesis tests, a valid confidence set for θ can be constructed by inverting $S_T(\theta)$. Specifically, applying Theorem 1, an asymptotically valid $100(1 - \alpha)\%$ confidence set is given by

$$C_\theta(1 - \alpha) = \{\theta \in \Theta : S_T(\theta) \leq \chi_{q-q_3}^2(1 - \alpha)\}$$

where $\chi_s^2(1 - \alpha)$ denotes the $100(1 - \alpha)$ -th percentile of a Chi-square variable with s degrees of freedom. Because this set contains the minimizers of the likelihood function, it is always nonempty.

To obtain this set, a direct grid search is computationally infeasible even for small scale DSGE models. We suggest using a Metropolis algorithm. The idea is to use Metropolis steps to generate

frequent draws from regions of Θ where the values of $S_T(\theta)$ are small and infrequent draws where $S_T(\theta)$ are large. This delivers a grid over Θ that adapts to the shape of $S_T(\theta)$, being dense at the desired areas (i.e., where $S_T(\theta)$ takes values near or below $\chi_{q-q_3}^2(1-\alpha)$) and sparse at the unimportant areas (i.e., where $S_T(\theta)$ is far above $\chi_{q-q_3}^2(1-\alpha)$). The confidence set can then be approximated using the values of θ for which $S_T(\theta)$ do not exceed $\chi_{q-q_3}^2(1-\alpha)$.

Let $\pi(\theta)$ be an indicator that equals 1 over Θ and 0 otherwise. Because of Assumption 1, $\pi(\theta)$ acts as a uniform prior with a compact support. This ensures that the resulting quasi-posterior will be proper even if $S_T(\theta)$ is flat. The basic steps for constructing the confidence set are:

1. Choose a starting value $\theta^{(0)}$ and set $j = 0$.
2. Draw θ^* from some proposal distribution $q(\cdot|\theta^{(j)})$.
3. Calculate the ratio

$$s = \min \left\{ \frac{\pi(\theta^*)e^{-\frac{1}{2}S_T(\theta^*)} q(\theta^{(j)}|\theta^*)}{\pi(\theta^{(j)})e^{-\frac{1}{2}S_T(\theta^{(j)})} q(\theta^*|\theta^{(j)})}, 1 \right\}$$

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 $j = B$ for some large B .
5. Sort the draws according to the values of $S_T(\theta)$ and keep those satisfying $S_T(\theta^{(j)}) \leq \chi_{q-q_3}^2(1-\alpha)$ ($j = 1, \dots, B$). Use the envelope of these draws to form the confidence set.

The above procedure is motivated by Chernozhukov and Hong (2003) in which MCMC is used as a computational device for classical estimation. Because the goal here is not to obtain a point estimate for θ , the assumptions for asymptotic distributions (Assumptions 2-4 in Chernozhukov and Hong, 2003) are not required.

Steps 1 to 5 cover only the basic aspects. In practice, it is important to fine tune them to better account for the potential ridges or local minima in the surface of $S_T(\theta)$. For example, we have incorporated the following elements when analyzing the model of An and Schorfheide (2007) in Section 7. First, in Steps 2 and 3, two different proposal distributions are applied iteratively to generate new parameter values. More specifically, write the new draw in Step 2 as $\theta^* = \theta^{(j)} + \varepsilon$. The first distribution gives $\varepsilon \sim N(0, M_T(\theta^{(j)}))$ with c being a tuning constant, while the second gives $\varepsilon = cV_T(\theta^{(j)})$ or $-cV_T(\theta^{(j)})$ with $V_T(\theta^{(j)})$ being the eigenvector corresponding to the smallest eigenvalue of $M_T(\theta^{(j)})$. These two distributions produce draws that travel across and along the

ridges of $S_T(\theta)$. Second, for each proposal distribution, we let the tuning parameter c take on multiple values. This prevents the sampling process from getting locked in some small neighborhood of a local minimum. Finally, multiple Markov Chains are run with different initial values. The confidence set is then obtained by merging the accepted values from all the chains.

Once the joint confidence set is obtained, confidence sets for parameter subvectors can be obtained using the projection method, i.e., 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 (2013). A discussion of this method in the IV context can be found in Dufour and Taamouti (2005).

6.1 Extensions

Below, we propose procedures to construct uniform confidence bands for the impulse response function and some other objects.

Let $IR_{jl}(k, \theta_0)$ be the impulse response of the j -th variable in Y_t to the l -th orthogonal shock in ϵ_t at the horizon k when the true parameter value is θ_0 . The next definition specifies a uniform confidence band for $IR_{jl}(k, \theta_0)$.

Definition 1 *Let $C_{IR}(k; T)$ be a confidence band for $IR_{jl}(k, \theta_0)$ indexed by $k \in [0, \infty)$. We say it is uniform at the level $1 - \alpha$ if*

$$\liminf_{T \rightarrow \infty} \Pr(IR_{jl}(k, \theta_0) \in C_{IR}(k; T) \text{ for all } k \in [0, \infty)) \geq 1 - \alpha$$

The band can be constructed by considering the envelope of the impulse response functions associated with all the parameter values contained in the joint confidence set obtained in the previous subsection. 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 (4). This is easily computed using the output from Sims (2002), because (see (1); 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 $C_\theta(1 - \alpha)$.

- Step 2. Compute the impulse response function using all parameter values in $C_\theta(1 - \alpha)$. This step can be approximated using the MCMC draws from Step 1 satisfying $S_T(\theta) \leq \chi_{q-q_3}^2(1 - \alpha)$.
- Step 3. Sort the resulting values at each horizon of interest. Use their maxima and minima to form a confidence band.

Because $IR(k, \theta)$ is a deterministic function of θ and k , this band covers the impulse response function with probability at least $(1 - \alpha)$ asymptotically. (A proof for this claim is provided in the on-line appendix.) 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.² This feature will be illustrated in Section 7.

The same idea can be applied to construct confidence sets for the time path of the variance decomposition, or other objects that are deterministic functions of the structural parameter vector. Below we discuss two such 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.

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 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

²Having 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.

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

$$\frac{\mathbf{e}'_j f_\theta(0) \mathbf{e}_l}{\mathbf{e}'_j f_\theta(0) \mathbf{e}_j} \quad (16)$$

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 tackled this issue from a Bayesian perspective. The methods developed in this paper can be used to construct frequentist confidence intervals for (16), therefore to evaluate whether a unit slope is consistent with the model and the data.

7 Finite sample properties

This section first examines the finite sample properties of the following objects: the proposed tests, the confidence intervals and the confidence bands for the impulse response functions. Then, it considers the performance of the tests under non-Gaussian innovations and low frequency misspecifications.

The model is taken from An and Schorfheide (2007):

$$\begin{aligned} y_t &= \mathbb{E}_t y_{t+1} + g_t - \mathbb{E}_t g_{t+1} - \frac{1}{\tau} (r_t - \mathbb{E}_t \pi_{t+1} - \mathbb{E}_t z_{t+1}) \\ \pi_t &= \beta \mathbb{E}_t \pi_{t+1} + \kappa (y_t - g_t) \\ r_t &= \rho_r r_{t-1} + (1 - \rho_r) \psi_1 \pi_t + (1 - \rho_r) \psi_2 (y_t - g_t) + \epsilon_{rt} \\ g_t &= \rho_g g_{t-1} + \epsilon_{gt} \\ z_t &= \rho_z z_{t-1} + \epsilon_{zt}, \end{aligned} \quad (17)$$

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):

$$\begin{aligned} YGR_t &= \gamma^{(Q)} + 100(y_t - y_{t-1} + z_t) \\ INFL_t &= \pi^{(A)} + 400\pi_t \\ INT_t &= \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400r_t \end{aligned} \quad (18)$$

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

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

The first 11 parameters are dynamic parameters ($r^{(A)}$ depends on β which appears in the log-linearized equations). The parameter values are taken from Table 2 in An and Schorfheide (2007):

$$\theta_0 = (2, 0.15, 1.5, 1.00, 0.60, 0.95, 0.65, 0.2, 0.8, 0.45, 0.40, 4.00, 0.50). \quad (19)$$

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

Design 1 (BC frequencies). Inference on the 11 dynamic parameters based on business cycle frequencies (i.e., periods of 6–32 quarters): $\omega \in [\pi/16, \pi/3] \cup [5\pi/3, 31\pi/16]$.

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

Design 3 (Mean and full spectrum). Inference on θ based on the mean and the full spectrum.

When implementing the tests, $\partial\mu(\theta)/\partial\theta'$ is computed analytically while $\partial \text{vec } f_\theta(\omega)/\partial\theta'$ is computed using a two-point method with step size 10^{-6} . We consider four empirically relevant sample sizes to evaluate the size and power properties: $T = 80, 160, 240, 320$. In each case, we report rejection frequencies based on 5000 replications.

7.1 The model's identification properties

This subsection illustrates the model's identification properties for a better understanding of the simulation results. We focus on Design 2 with $T = 80$, although the findings are quite similar under Designs 2 and 3. The MATLAB default tolerance level yields $\text{rank}(M_T(\theta_0)) = 10 < 11$. The method of Qu and Tkachenko (2012, Section 3.1) shows there exists a unique non-identification curve generated by $(\psi_1, \psi_2, \rho_r, \sigma_r)$. The curve extends in both the positive and negative directions around θ_0 . In Direction 1, it is truncated before ψ_2 turns negative. Along Direction 2, it reaches an indeterminacy region before any natural bounds are violated, and is truncated at the last point that yields a determinate solution. Table 1 reports 10 evenly spaced points along each direction. Two interesting patterns emerge. First, for ψ_1 and ψ_2 , the curve extends over a fairly large neighborhood: ψ_1 varies between 0.99 and 4.87, while ψ_2 between 0.00 and 1.15. Second, the corresponding neighborhoods for ρ_r and σ_r are relatively small: ρ_r can only change between 0.58 and 0.60, while $100\sigma_r$ between 0.19 and 0.20. The latter feature suggests that the data can still be informative about ρ_r and σ_r even though they are not separately identifiable from ψ_1 and ψ_2 .

The rank estimate is insensitive to the sample size considered. The smallest eigenvalue equals 4.4E-15, 4.8E-15, 4.8E-15 and 5.4E-15 when $T = 80, 160, 240$ and 320. They are well below the default tolerance level 3.6E-13. This insensitivity follows because the summands in $M_T(\theta_0)$ are de-

terministic and smooth functions of ω . This differs from usual score tests in which the normalization matrices are sample dependent.

The rank result is reconfirmed after applying the two step procedure discussed in Section 5.1. Specifically, in Step One, the minimal rank equals 9 when considering a wide range of tolerance levels: $10^j \times \dim(M_T(\theta_0))\text{eps}(\|M_T(\theta_0)\|)$ ($j = -10, -9, \dots, 10$). Two curves are obtained in Step 2, generated by $(\psi_1, \psi_2, \rho_r, \sigma_r)$ and $r^{(A)}$ (or equivalently β). The values on the first curve (i.e., those in Table 1) produce essentially the same spectral density as $f_{\theta_0}(\omega)$ with the maximum absolute difference being 1.4E-7. This confirms that $\psi_1, \psi_2, \rho_r, \sigma_r$ are not separately identifiable. In contrast, noticeable discrepancies emerge when increasing the value of $r^{(A)}$ from 0.4. When the value reaches 4.0 (i.e., when β changes from 0.999 to 0.990), the maximum difference between $f_{\theta}(\omega)$ and $f_{\theta_0}(\omega)$ reaches 0.002. Further increases lead to greater differences. This confirms that $r^{(A)}$ is locally identified, but only weakly, from the second order properties of the observables.

The reduced form parameters in this model are not immune to identification problems. Specifically, a minimal state space representation of the model's solution evaluated at θ_0 is given by

$$\begin{aligned} \begin{pmatrix} z_{t+1} \\ g_{t+1} \\ r_{t+1} \end{pmatrix} &= \underbrace{\begin{pmatrix} 0.65 & 0 & 0 \\ 0 & 0.95 & 0 \\ 0.1548 & 0 & 0.4 \end{pmatrix}}_A \begin{pmatrix} z_t \\ g_t \\ r_t \end{pmatrix} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.2382 & 0 & 0.6667 \end{pmatrix}}_B \underbrace{\begin{pmatrix} \epsilon_{zt+1} \\ \epsilon_{gt+1} \\ \epsilon_{rt+1} \end{pmatrix}}_{\epsilon_{t+1}}, \\ \begin{pmatrix} r_{t+1} \\ y_{t+1} \\ \pi_{t+1} \end{pmatrix} &= \underbrace{\begin{pmatrix} 0.1548 & 0 & 0.4 \\ 0.2724 & 0.95 & -0.3637 \\ 0.0764 & 0 & -0.0909 \end{pmatrix}}_C \begin{pmatrix} z_t \\ g_t \\ r_t \end{pmatrix} + \underbrace{\begin{pmatrix} 0.2382 & 0 & 0.6667 \\ 0.4191 & 1 & -0.6061 \\ 0.1176 & 0 & -0.1514 \end{pmatrix}}_D \begin{pmatrix} \epsilon_{zt+1} \\ \epsilon_{gt+1} \\ \epsilon_{rt+1} \end{pmatrix}. \end{aligned}$$

The covariance of ϵ_{t+1} , $\Sigma(\theta_0)$, is not separately identifiable from B and D without further normalizations. We therefore fix $\Sigma(\theta_0)$ and treat the non-zero elements in A, B, C and D as reduced form parameters. Denote the collection of these parameters as ϕ . Write $\phi = \phi_0$ when $\theta = \theta_0$.

The rank condition in Qu and Tkachenko (2012, Theorem 1) shows that ϕ_0 is unidentifiable based on the second order properties of $\{r_t, y_t, \pi_t\}$. The criterion function $G(\phi_0)$ has four zero eigenvalues, suggesting there are multiple sources contributing to the identification failure. In fact, for any invertible lower triangular 3-by-3 matrix U with $U_{21} = U_{32} = 0$, replacing A, B and C by UAU^{-1}, UB and CU^{-1} leaves the dynamic properties of $\{r_t, y_t, \pi_t\}$ unchanged. This result confirms that the reduced form parameters in a minimal state space representation can be unidentifiable even after fixing the covariance matrix of the structural shocks and imposing the knowledge about the

zero entries. Gueron-Quintana, Inoue and Kilian (2013) found similar results in their applications. They suggested deducting the number of free parameters in U from the degrees of freedom of the likelihood ratio test. This identification feature also poses computational challenges because the likelihood is flat in multiple dimensions and it has a continuum of global maximizers.

7.2 Size in finite samples

The test statistics are constructed by setting the smallest eigenvalue in the information matrix to zero. The χ_{10}^2 (for dynamic parameters) and χ_{12}^2 (for the full parameter vector) distributions are used to determine whether rejection occurs. The results are summarized in the first panel of Table 2. The rejection rates are overall close to the nominal levels. Some mild over-rejections persist under Designs 2 and 3, with the maximum rejection frequencies being 9.5% and 14.4% at the 5% and 10% nominal levels, respectively. This appears to be because the spectral density is close to being singular near the zero frequency. When the lowest frequency component is excluded, the rejection frequencies under Design 2 decrease to 7.3% and 11.9% at the 5% and 10% nominal levels when $T = 320$. Therefore, the over rejection 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 above experiment considers a particular parameter value. It remains to verify whether the size is controlled in a more general situation. 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: $\tau \sim [1e-5, 5]$, $\kappa \sim [0, 1]$, $\psi_1 \sim [0, 5]$, $\psi_2 \sim [0, 2]$, $\rho_r \sim [0, 0.9]$, $\rho_g \sim [0, 0.99]$, $\rho_z \sim [0, 0.99]$, $100\sigma_r \sim [1e-5, 2]$, $100\sigma_g \sim [1e-5, 2]$, $100\sigma_z \sim [1e-5, 2]$, $r^{(A)} \sim [0, 5]$, $\pi^{(A)} \sim [0, 20]$, $\gamma^{(Q)} \sim [0, 5]$. 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). The rejection frequencies are summarized in the second panel in Table 2. The results are quite similar to the previous case.

We now purposely misspecify the rank of the information matrix and examine the tests' size properties. In Table 3, all eigenvalues of the information matrix are treated as non-zeros and the distributions of χ_{11}^2 and χ_{13}^2 are used for inference. As expected, the tests reject less frequently compared with Table 2. The rates are between 4.5%-8.2% and 7.9%-12.5% at the two nominal levels. Table 4 reports the results when two eigenvalues are classified as zeros and the distributions of χ_9^2 and χ_{11}^2 are used for inference. There, the size continues to be adequate: the overall rejection

rates are between 6.6%-10.8% and 10.8%-15.8% 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 2 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 model the distributions of the conventional Wald and likelihood ratio (LR) tests are highly nonstandard. For the Wald test, because the MLE $\hat{\theta}$ can lie close to a point distant from θ_0 on the non-identification curve even in large samples, $\sqrt{T}(\hat{\theta} - \theta_0)$ can take on a very large value. Consequently, the test can diverge if a finite covariance matrix is used to standardize the above difference. For the conventional LR test, because the likelihood surface displays a ridge along the non-identification curve, the standard quadratic approximation is no longer adequate. Consequently, the Chi-square approximation to the limiting distribution also breaks down. In contrast, The distribution of the $S_T(\theta)$ test is not established under an expansion around a point estimate, therefore the inference is not affected by the above non standard features. Instead, the key assumption validating the asymptotic distribution is Assumption W(iv). In this model, $[\partial \text{vec } f_{\theta_0}(\omega)/\partial \theta'] Q_T(\theta_0) \Lambda_T^+(\theta_0)^{1/2}$ is a n_Y^2 -by- q matrix. The elements in its first row are plotted in Figure 1 as a function of ω to illustrate their magnitudes and smoothness in ω . The remaining elements exhibit similar features and are omitted to save space. The figure supports Assumption W(iv).

7.3 Finite sample power

We perturb the individual element of θ_0 (only the dynamic parameters in Designs 1 and 2) given in (19) by a fixed percentage and then compute the rejection frequencies. Specifically, we take a uniform random draw from the index set $\{1, \dots, 13\}$ (or $\{1, \dots, 11\}$ in Designs 1 and 2) and change the corresponding element of the parameter vector by $\kappa\%$ 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-adjusted rejection frequencies are reported in Table 5.

The first panel is for $\kappa = 20$. Using the business cycle frequencies, the $S_T(\theta_0)$ test achieves 51.1%-61.4% of the power attainable using the full spectrum. The rejection rates under Designs 2 and 3 are similar. Because Design 3 involves more parameters, the power is not necessarily higher than under Design 2. The second panel corresponds to $\kappa = 40$. There, the ratios are between

48.5%-86.1%. Designs 2 and 3 continue to show similar rejection frequencies.

We misspecify the rank of the information matrix and compute the size adjust rejection frequencies. In Table 6, all eigenvalues are treated as non-zeros. In Table 7, the two smallest eigenvalues are treated as zeros. In both tables, the power is very similar to that in Table 5. Intuitively, a small eigenvalue implies that the likelihood surface lacks curvature in the corresponding direction. There is little to lose, or even something to gain, from treating it as zero.

In summary, the results from Designs 1 and 2 suggest that the tests using only business cycle frequencies can be informative. Meanwhile, 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. The informativeness of the different procedures will be further studied in the next two subsections through the lenses of confidence intervals and impulse responses. Prior to that, we first report a comparison with Andrews and Mikusheva's (2013) tests using the current model.

Andrews and Mikusheva (2013) studied two LM tests, LM_o and LM_e , from a time domain perspective. Given that they can not be applied to Designs 1 and 2, below we only consider their finite sample properties under Design 3. Panel (a) in Table 8 summarizes their rejection frequencies under the null hypothesis. The two tests perform quite differently. The LM_o test exhibits substantial size distortions, while the LM_e test performs similarly to $S_T(\theta_0)$. The difference follows from the handling of the information matrix. The LM_o test tries to estimate it from a finite sample. This involves estimating $q(q + 1)/2$ unknown parameters with q being the dimension of θ_0 . The noise in this estimate can significantly affect the test's size, especially under weak identification. The LM_e test, similarly as $S_T(\theta_0)$, computes the information matrix directly from the model, therefore is unaffected by this problem. In the frequency domain, there is a simple formula available for computing the information matrix, while in the time domain one needs to use some tailored procedures or simulations. Andrews and Mikusheva (2013) suggest using the method of Iskrev (2008), along with the MATLAB toolbox E4. Panel (b) in Table 8 summarizes their rejection frequencies under the alternative hypothesis. Overall, the values are similar to those reported in the last two columns in Tables 5. Specifically, the differences between LM_e and $S_T(\theta_0)$ are within in 0.02 to 0.09, and LM_o and $S_T(\theta_0)$ within -0.17 to 0.10. These results are consistent with those reported in Andrews and Mikusheva (2013) using a different model.

7.4 Confidence intervals for structural parameters

Table 9 summarizes the length of the 90% confidence intervals when θ equals (19) and $T = 240$. Column 4 corresponds to the business cycle case. First, the intervals reveal little information about τ , ψ_1 , ψ_2 and $r^{(A)}$ (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 ρ_r and κ . Third, the intervals related to the exogenous disturbances ($\rho_g, \rho_z, 100\sigma_r, 100\sigma_g$ and $100\sigma_z$) are relatively informative. This is again consistent with the findings of An and Schorfheide (2007, P.133-134).

Column 5 corresponds to the full spectrum case. 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. Column 6 incorporates the steady state parameters. There, the intervals remain roughly the same except for $r^{(A)}$. The latter interval narrows because $r^{(A)}$ is tied to the steady state of the interest rate.

The above comparison shows that inference using only business cycle frequencies can be informative (see $\rho_g, \rho_z, 100\sigma_r, 100\sigma_g$ and $100\sigma_z$), while using the full spectrum can bring substantial gain in efficiency. In practice, this offers researchers a choice. If the model is reasonably specified at all frequencies, then the full spectrum should be used and the inference will be more precise. If the model is misspecified over some frequencies, then using parts of the spectrum is preferable.

Importantly, the results suggest that it is possible to have informative confidence intervals in DSGE models with unidentified parameters. Further, even unidentified parameters themselves can have tight confidence intervals; see ρ_r and σ_r in Columns 5 and 6. To see why the latter has happened in this model, note that ρ_r and σ_r are elements of the Taylor rule parameters $(\psi_1, \psi_2, \rho_r, \sigma_r)$ which lie on a non-identification curve depicted in Table 1. The likelihood function is completely flat along this curve but has curvature in all other directions. For ρ_r and σ_r , the curve occupies only a relatively small neighborhood. Consequently the effect of the identification failure is relatively mild. The confidence intervals can still be tight. More generally, such a feature can arise if identification failure involves multiple parameters but the non-identified directions are limited relative to the number of such parameters.

7.5 Confidence bands for impulse responses

We illustrate the properties of the 90% uniform confidence bands using a simulated process with θ_0 equal to (19) 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 2 contains confidence bands 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 bands are in general fairly wide but can be informative. They show 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 3 corresponds to 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 4 contains results using the mean and the spectrum. The bands are similar to those in Figure 3. They are not necessarily narrower than those in Figure 3 because additional parameters are present.

Therefore, it is possible to have informative interval estimates of the impulse response functions in DSGE models with unidentified parameters. 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 $1\text{E}-7$. This confirms that in this model, the parameters on the identification curve do deliver the same impulse responses.

7.6 Non-Gaussian innovations

This subsection studies the finite sample properties of the modified test (15) when the shocks follow Student-t distributions. Specifically, we let

$$\epsilon_{rt} \stackrel{i.i.d.}{\sim} t(0, \sigma_r^2, \lambda_r), \quad \epsilon_{gt} \stackrel{i.i.d.}{\sim} t(0, \sigma_g^2, \lambda_g), \quad \epsilon_{zt} \stackrel{i.i.d.}{\sim} t(0, \sigma_z^2, \lambda_z),$$

where $t(0, \sigma^2, \lambda)$ denotes a Student-t distribution with location parameter 0, scale parameter σ^2 and λ degrees of freedom. The values λ_r , λ_g and λ_z are taken from the last three columns in Table 2 in Cúrdia, Del Negro and Greenwald (2013). They are treated as additional structural parameters when constructing the tests. All other specifications are the same as in (17).

The rejection frequencies are summarized in Table 10. Panel (a) corresponds to the null hypothesis. The values are fairly close to the nominal levels. Panel (b) corresponds to the alternative hypothesis. There, the power is overall lower than that reported in Table 5, however the overall pattern is similar. In particular, using the business cycle frequencies, the test achieves 65.0% – 75.0% and 59.8% – 84.9% of the power attainable using the full spectrum when $\kappa = 20$ and 40.

7.7 Robustness to low frequency misspecification

This subsection studies the tests' size properties when the data exhibit certain unmodeled low frequency variations. First, we suppose the growth rate of technology, and therefore the means of YGR_t and INT_t , are affected by a structural change at T_b :

$$\gamma^{(Q)}(t) = \gamma^{(Q)} + \delta * 1(t \geq T_b). \quad (20)$$

Second, we suppose the inflation target, and therefore the means of $INFL_t$ and INT_t , are time varying:

$$\pi^{(A)}(t) = \pi^{(A)} + \xi(t). \quad (21)$$

In both cases, the log linearized solution still satisfies (17) if we abstract from the effect of learning and assume that the change in the inflation target is sufficiently smooth. Misspecification arises only when relating the observables to their log deviations using the time invariant relationship (18). Frequency domain methods provide a simple way to account for such misspecifications without requiring specifying the location of the change or the time path of the policy target. This is possible because such variations mainly affect the very low frequencies, which can be simply excluded or downweighted when conducting inference. Note that the above two misspecifications are different from Cogley (2001). In the latter, the degree of integration (i.e., unit root versus stationarity) for the technology process is misspecified. This affects the model's dynamics at all frequencies and, as shown in Cogley (2001), removing the low frequencies offers little help.

In order to make the analysis empirically relevant, we calibrate the values of δ , T_b and $\xi(t)$ using U.S. quarterly time series over the period 1947:Q1–2012:Q3 (263 observations). For δ and T_b , we regress GDP growth rates (Series GDPC1 obtained from the Saint Louis Fed Website) on a constant and a break in the intercept. The sum of squared residuals is minimized at $T_b = 105$, which corresponds to 1973:Q1. The estimated break magnitude equals -0.23 . Because the series in the model and the actual data have different standard deviations, we multiply -0.23 by their ratio to make the magnitude more comparable. Such calculations lead to $\delta = -0.26$. To obtain $\xi(t)$, we apply local regression to the inflation series (CPIAUCSL_PCH), using the LOESS command in R with the bandwidth parameter set to $0.5T$. We then set $\xi(t)$ to be the fitted smooth curve adjusted by the relative standard deviations. The resulting $\gamma^{(Q)}(t)$ and $\xi(t)$ each contain 263 values. The first 80, 160 and 240 values are used to simulate samples of corresponding sizes and the 263 values are linearly interpolated to generate samples of 320 observations.

Table 11 reports the rejection frequencies when the test statistics are constructed using business cycle frequencies only (BC), business cycle and all the higher frequencies (BC+High) and the

full spectrum (Full). Panel (a) corresponds to (20). The test based on business cycle frequencies performs quite well for all the sample sizes considered. Its rejection rates are only mildly inflated, falling between 6.7%–8.4% and 11.8%–13.1% at 5% and 10% nominal levels, respectively. In contrast, the test using the full spectrum suffers from serious size distortions. Its rejection frequencies reach 62.5% and 70.1% when $T = 320$. The test using all but the very low frequencies (BC+High) performs similarly to using only the business cycle frequencies. This confirms that the difference between BC and Full is indeed due to the low frequencies. Panel (b) corresponds to (21). The BC based test continues to perform well even when the full spectrum based test rejects close to 100% of the time. Panel (c) corresponds to the situation where both (20) and (21) are present. There, the difference is even more pronounced. Overall, the results suggest that substantial robustness can be achieved by excluding a relative small number of low frequency components.

8 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. 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 sizes in relatively small samples. It also suggests that it is possible to obtain informative results in DSGE models with unidentified parameters. Although the paper has focused on DSGE models, the methods developed are applicable to other dynamic models satisfying Assumptions 1-4 and W such as the FAVAR.

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 and (2) examining certain low frequency hypotheses. Parameters in DSGE models are often highly correlated. This can be seen from the non-identification curve reported in Table 1, 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.
- Andrews, D.W.K., and X. Cheng (2012): “Estimation and Inference with Weak, Semi-strong, and Strong Identification,” *Econometrica*, 80, 2153–2211.
- Andrews, I. and A. Mikusheva (2013): “Maximum Likelihood Inference in Weakly Identified DSGE Models,” Working paper, MIT.
- Altug, S. (1989): “Time-to-Build and Aggregate Fluctuations: Some New Evidence,” *International Economic Review*, 30, 889–920.
- 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., 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., and R. J. Vigfusson (2003): “Maximum Likelihood in the Frequency Domain: the Importance of Time-to-plan,” *Journal of Monetary Economics*, 50, 789–815.
- Cogley, T. (2001): “Estimating and Testing Rational Expectations Models when the Trend Specification is Uncertain,” *Journal of Economic Dynamics and Control*, 5, 1485–1525.
- Cúrdia, V., M. Del Negro., and D.L. Greenwald (2013): “Rare shocks, Great Recessions,” Working Paper, Federal Reserve Bank of San Francisco.
- Davidson, R., and J.G. MacKinnon (1993): *Estimation and Inference in Econometrics*, Oxford University Press.
- 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.
- 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 (2013) “Identification-robust Analysis of DSGE and Structural Macroeconomic Models”, *Journal of Monetary Economics*, 60, 340–350.

- 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.
- Guerron-Quintana, P., A. Inoue., and L. Kilian (2013): “Frequentist Inference in Weakly Identified DSGE Models,” *Quantitative Economics*, 4, 197-229.
- 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.
- Iskrev, N. (2008): “Evaluating the Information Matrix in Linearized DSGE Models,” *Economics Letters*, 99, 607-610.
- (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. (2005): “Testing Parameters in GMM without Assuming that They are Identified,” *Econometrica*, 73, 1103-1124.
- 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 Dynamic Stochastic General Equilibrium Models,” *Econometrica*, 79, 1995–2032.
- Lucas, R.E., Jr. (1980): “Two Illustrations of the Quantity Theory of Money,” *American Economic Review*, 70, 1005–1014.
- McCloskey, A., and J.B. Hill (2013): “Parameter Estimation Robust to Low-Frequency Contamination,” Working paper, Brown University.
- Magnus, J. R., and H. Neudecker, H. (2002): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: John Wiley.

- Perron, P., and Z. Qu (2010): “Long-memory and Level Shifts in the Volatility of Stock Market Return Indices,” *Journal of Business & Economics Statistics* 28, 275-290.
- Qu, Z. (2011): “A Test against Spurious Long Memory,” *Journal of Business & Economics Statistics*, 29, 423–438.
- Qu, Z., and D. Tkachenko (2012): “Identification and Frequency Domain Quasi-maximum Likelihood Estimation of Linearized Dynamic Stochastic General Equilibrium Models,” *Quantitative Economics* 3, 95-132.
- Rothenberg, T. J. (1971): “Identification in Parametric Models,” *Econometrica*, 39, 577–591.
- Sala, L. (2013): “DSGE Models in the Frequency Domain,” Working paper, Università Bocconi.
- Sargent, T.J., and P. Surico (2011): “Two Illustrations of the Quantity Theory of Money: Breakdowns and Revivals,” *American Economic Review*, 101, 109–128.
- Schorfheide, F. (2010): “Estimation and Evaluation of DSGE Models: Progress and Challenges,” Forthcoming in *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, Cambridge University Press.
- Sims, C. (2002): “Solving Linear Rational Expectations Models,” *Computational Economics*, 20, 1-20.
- Smets, F., and R. Wouters (2007): “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,” *The American Economic Review*, 97, 586-606.
- 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.
- Tkachenko, D., and Z. Qu (2012): “Frequency Domain Analysis of Medium Scale DSGE Models with Application to Smets and Wouters (2007),” *Advances in Econometrics: DSGE Models in Macroeconomics – Estimation, Evaluation and New Developments*, 319-385.
- 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.
- 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.

Appendix

The following Lemma is needed for proving Theorem 1. Its proof, along with some other intermediate results, are in the online appendix.

Lemma A.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*

$$\xi_{1T} = \frac{1}{2\sqrt{T}} \sum_{j=1}^{T-1} \phi_T(\omega_j)^* \text{vec}(I_T(\omega_j) - f_{\theta_0}(\omega_j)), \quad \xi_{2T} = \frac{1}{2\pi\sqrt{T}} \sum_{t=1}^T \psi_T^*(Y_t - \mu(\theta_0)),$$

where

$$\begin{aligned} \phi_T(\omega) &= W(\omega) \left(f_{\theta_0}^{-1}(\omega)' \otimes f_{\theta_0}^{-1}(\omega) \right) \left\{ \left(\frac{\partial \text{vec} f_{\theta_0}(\omega)}{\partial \theta'} \right) Q_T^c(\theta_0) \Lambda_T^c(\theta_0)^{-1/2} \right\}, \\ \psi_T &= W(0) f_{\theta_0}^{-1}(0) \frac{\partial \mu(\theta_0)}{\partial \theta'} Q_T^c(\theta_0) \Lambda_T^c(\theta_0)^{-1/2}. \end{aligned}$$

Then

$$\xi_{1T} + \xi_{2T} \rightarrow^d N(0, \mathbb{I}_{q_1+q_2}),$$

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

Proof of 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),

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

where $\Lambda_T^c(\theta_0)$ and $Q_T^c(\theta_0)$ are defined as in the previous Lemma. Thus,

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

Let $\xi_T = \Lambda_T^c(\theta_0)^{-1/2} Q_T^c(\theta_0)' D_T(\theta_0)$. From the previous Lemma, $\xi_T \rightarrow^d N(0, \mathbb{I}_{q_1+q_2})$. This implies $S_T(\theta_0) \rightarrow^d \chi_{q_1+q_2}^2$. ■

Table 1. 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 2. Rejection frequencies under the null hypothesis

T	BC frequencies		Full spectrum		Mean and full spectrum	
	5%	10%	5%	10%	5%	10%
θ_0 taken from Table 2 in AS (2007)						
80	0.084	0.130	0.086	0.125	0.095	0.135
160	0.073	0.117	0.073	0.115	0.078	0.127
240	0.065	0.109	0.073	0.138	0.078	0.140
320	0.060	0.108	0.087	0.143	0.085	0.144
θ_0 drawn from a prior distribution						
80	0.089	0.135	0.088	0.135	0.102	0.156
160	0.076	0.122	0.080	0.122	0.086	0.136
240	0.067	0.114	0.083	0.142	0.086	0.145
320	0.065	0.111	0.083	0.145	0.084	0.149

Note. The first panel: θ_0 is taken from the last column of Table 2 in An and Schorfheide (2007). The second panel: θ_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 3. Null rejection frequencies under alternative computations of the test statistics
(All eigenvalues are treated as non-zeros)

T	BC frequencies		Full spectrum		Mean and full spectrum	
	5%	10%	5%	10%	5%	10%
θ_0 taken from Table 2 in AS (2007)						
80	0.067	0.102	0.061	0.097	0.079	0.109
160	0.055	0.089	0.055	0.085	0.061	0.100
240	0.051	0.085	0.066	0.107	0.060	0.111
320	0.045	0.079	0.068	0.108	0.067	0.113
θ_0 drawn from a prior distribution						
80	0.072	0.107	0.071	0.109	0.082	0.125
160	0.059	0.093	0.059	0.094	0.069	0.108
240	0.054	0.089	0.069	0.110	0.066	0.116
320	0.047	0.083	0.072	0.116	0.067	0.123

Note. See Table 2.

Table 4. Null rejection frequencies under alternative computations of the test statistics
(The two smallest eigenvalues are treated as zeros)

T	BC frequencies		Full spectrum		Mean and full spectrum	
	5%	10%	5%	10%	5%	10%
θ_0 taken from Table 2 in AS (2007)						
80	0.084	0.130	0.071	0.115	0.093	0.140
160	0.071	0.116	0.072	0.121	0.070	0.115
240	0.068	0.112	0.083	0.131	0.080	0.130
320	0.082	0.108	0.084	0.137	0.090	0.142
θ_0 drawn from a prior distribution						
80	0.090	0.135	0.082	0.128	0.108	0.158
160	0.076	0.118	0.078	0.131	0.074	0.126
240	0.071	0.116	0.085	0.136	0.086	0.140
320	0.066	0.114	0.086	0.138	0.091	0.146

Note. See Table 2.

Table 5. Size adjusted power

T	BC frequencies		Full spectrum		Mean and full spectrum	
	5%	10%	5%	10%	5%	10%
Randomly perturb the elements of θ_0 by 20%						
80	0.166	0.224	0.290	0.374	0.282	0.384
160	0.230	0.319	0.418	0.519	0.451	0.548
240	0.279	0.356	0.546	0.637	0.593	0.675
320	0.337	0.414	0.632	0.697	0.661	0.721
Randomly perturb the elements of θ_0 by 40%						
80	0.288	0.360	0.593	0.689	0.637	0.730
160	0.445	0.536	0.747	0.782	0.785	0.821
240	0.556	0.660	0.791	0.831	0.835	0.874
320	0.671	0.729	0.815	0.847	0.859	0.894

Note. θ_0 is taken from the last column of Table 2 in An and Schorfheide (2007).

Table 6. Size adjusted power under alternative computations of the test statistics
(All eigenvalues are treated as non-zeros)

T	BC frequencies		Full spectrum		Mean and full spectrum	
	5%	10%	5%	10%	5%	10%
Randomly perturb the elements of θ_0 by 20%						
80	0.157	0.223	0.283	0.368	0.287	0.387
160	0.223	0.309	0.418	0.519	0.446	0.546
240	0.301	0.381	0.541	0.634	0.572	0.666
320	0.340	0.416	0.610	0.687	0.643	0.716
Randomly perturb the elements of θ_0 by 40%						
80	0.293	0.365	0.607	0.698	0.625	0.729
160	0.439	0.533	0.750	0.787	0.791	0.822
240	0.571	0.664	0.780	0.822	0.817	0.859
320	0.669	0.720	0.802	0.835	0.860	0.902

Note. See Table 5.

Table 7. Size adjusted power under alternative computations of the test statistics
(The two smallest eigenvalues are set to exact zeros)

T	BC frequencies		Full spectrum		Mean and full spectrum	
	5%	10%	5%	10%	5%	10%
Randomly perturb the elements of θ_0 by 20%						
80	0.186	0.253	0.328	0.416	0.322	0.411
160	0.254	0.350	0.460	0.573	0.496	0.600
240	0.318	0.414	0.579	0.666	0.599	0.689
320	0.371	0.467	0.629	0.705	0.693	0.765
Randomly perturb the elements of θ_0 by 40%						
80	0.328	0.413	0.658	0.737	0.670	0.752
160	0.474	0.577	0.752	0.795	0.791	0.827
240	0.612	0.712	0.803	0.837	0.829	0.875
320	0.699	0.756	0.812	0.853	0.865	0.905

Note. See Table 5.

Table 8. Andrews and Mikusheva's (2012) tests

T	LM_e		LM_o	
	5%	10%	5%	10%
(a) Size				
80	0.079	0.121	0.198	0.289
160	0.070	0.113	0.138	0.211
240	0.071	0.118	0.115	0.182
320	0.069	0.114	0.098	0.162
(b) Size adjusted power				
Randomly perturb the elements of θ_0 by 20%				
80	0.329	0.413	0.305	0.376
160	0.474	0.580	0.482	0.559
240	0.642	0.726	0.577	0.668
320	0.751	0.805	0.686	0.761
Randomly perturb the elements of θ_0 by 40%				
80	0.660	0.751	0.514	0.581
160	0.823	0.855	0.708	0.797
240	0.887	0.916	0.901	0.944
320	0.929	0.948	0.955	0.967

Note. θ_0 is taken from the last column of Table 2 in An and Schorfheide (2007). The tests are computed by setting the smallest eigenvalue of the information matrix to zero.

Table 9. Lengths of the 90% confidence intervals

Parameter	θ_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.

Table 10. Finite sample properties with Student-t innovations

T	BC frequencies		Full spectrum		Mean and full spectrum	
	5%	10%	5%	10%	5%	10%
(a) Size						
$df = (8.2, 11.4, 7.5)$						
80	0.061	0.084	0.061	0.088	0.070	0.100
160	0.040	0.062	0.046	0.073	0.052	0.078
240	0.034	0.058	0.050	0.077	0.054	0.084
320	0.030	0.053	0.046	0.075	0.052	0.078
$df = (5.9, 8.3, 5.5)$						
80	0.070	0.096	0.077	0.104	0.084	0.115
160	0.048	0.071	0.062	0.087	0.067	0.096
240	0.043	0.067	0.062	0.094	0.070	0.102
320	0.038	0.061	0.066	0.097	0.067	0.097
$df = (8.1, 7.6, 5.6)$						
80	0.068	0.092	0.073	0.102	0.081	0.112
160	0.045	0.068	0.057	0.083	0.063	0.091
240	0.040	0.065	0.060	0.090	0.065	0.096
320	0.035	0.060	0.058	0.089	0.061	0.091
(b) Size adjusted power						
$df = (8.1, 7.6, 5.6)$; randomly perturb the elements of θ_0 by 20%						
80	0.124	0.192	0.179	0.256	0.168	0.252
160	0.165	0.246	0.254	0.346	0.289	0.371
240	0.206	0.286	0.315	0.438	0.348	0.443
320	0.272	0.349	0.393	0.498	0.430	0.548
$df = (8.1, 7.6, 5.6)$; randomly perturb the elements of θ_0 by 40%						
80	0.207	0.276	0.328	0.447	0.392	0.488
160	0.331	0.415	0.554	0.629	0.598	0.662
240	0.428	0.524	0.631	0.686	0.668	0.727
320	0.520	0.600	0.652	0.707	0.693	0.745

Note. The degrees of freedom parameters are taken from Cúrdia, Del Negro and Greenwald (2013); the other parameters are from the last column of Table 2 in An and Schorfheide (2007).

Table 11. Robustness to low frequency misspecifications

T	(a) Structural change in technology growth			(b) Smoothly varying inflation target			(c) Combined changes			
	BC	BC+High	Full	BC	BC+High	Full	BC	BC+High	Full	
5%	80	0.084	0.076	0.079	0.086	0.080	0.124	0.086	0.080	0.124
	160	0.082	0.087	0.319	0.069	0.071	0.968	0.081	0.087	0.943
	240	0.070	0.079	0.527	0.063	0.068	0.990	0.073	0.080	0.988
	320	0.067	0.072	0.625	0.062	0.067	1.000	0.070	0.074	1.000
10%	80	0.128	0.123	0.123	0.129	0.126	0.179	0.129	0.126	0.179
	160	0.131	0.139	0.412	0.113	0.122	0.979	0.131	0.140	0.958
	240	0.119	0.132	0.615	0.108	0.122	0.993	0.122	0.135	0.992
	320	0.118	0.121	0.701	0.109	0.114	1.000	0.118	0.129	1.000

Note. θ_0 is taken from the last column of Table 2 in An and Schorfheide (2007). BC, BC+High and Full correspond to inference using business cycle frequencies, business cycle and higher frequencies, and the full spectrum.

Figure 1. A graphical illustration of Assumption W(iv)

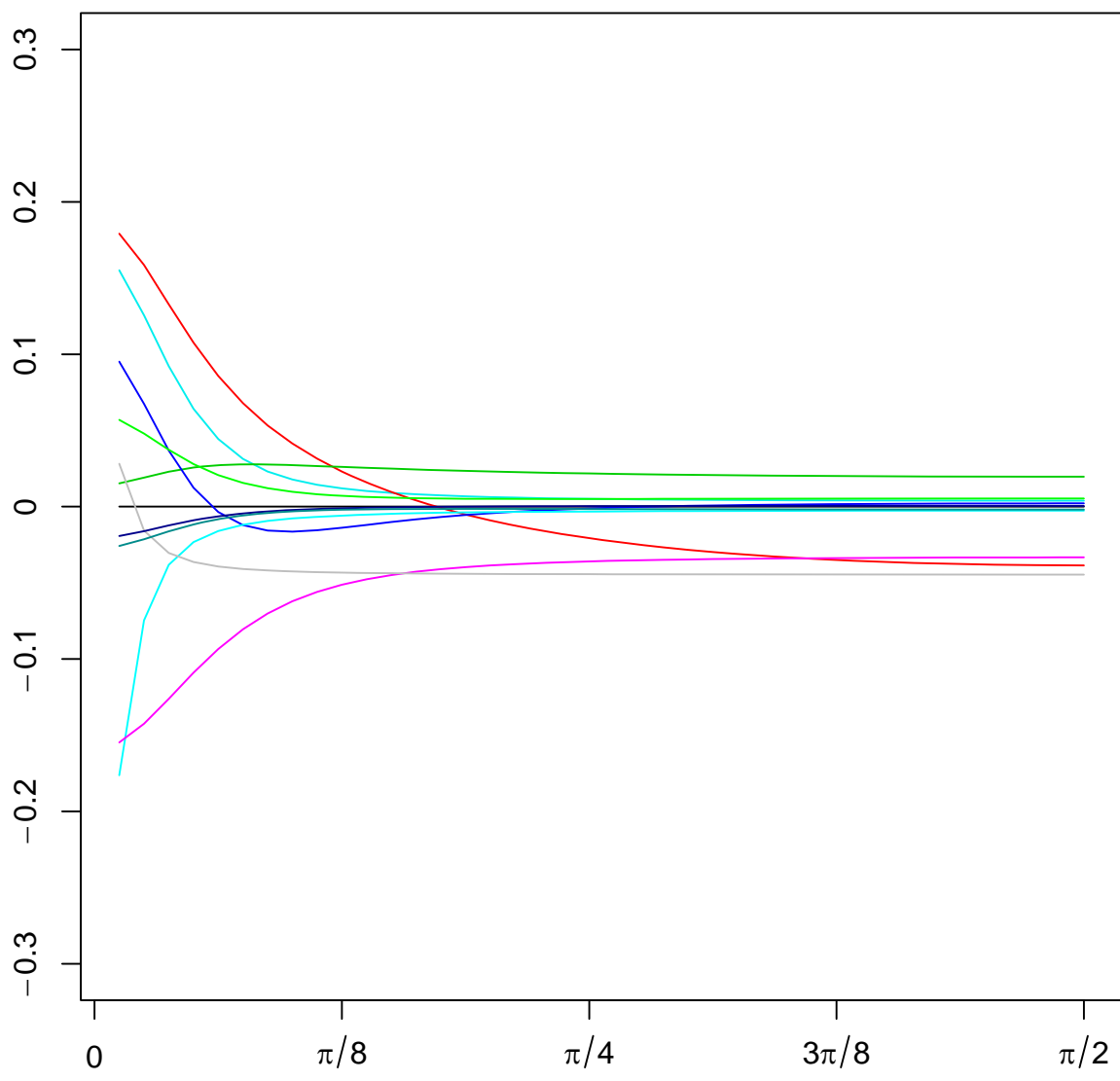
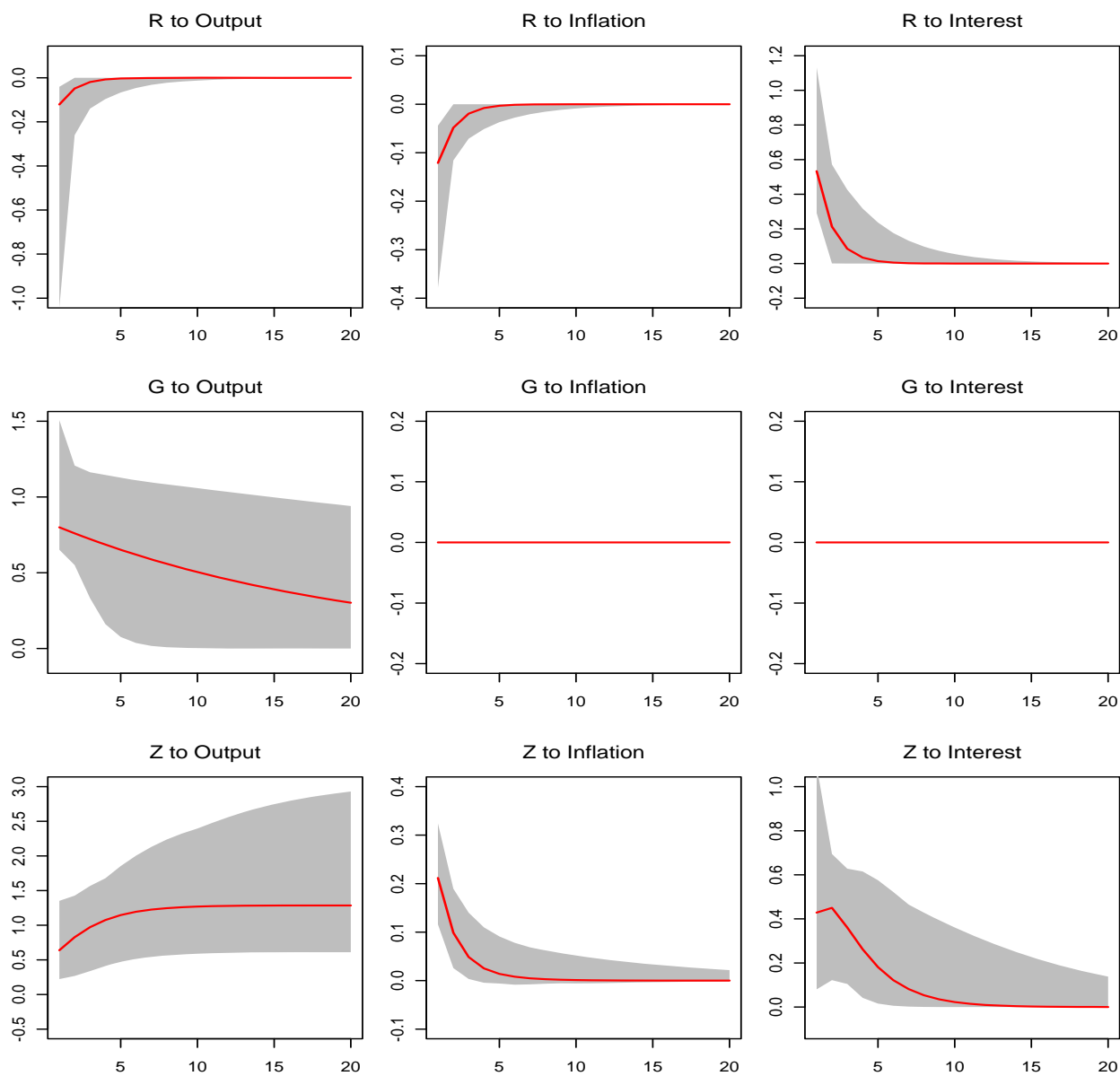
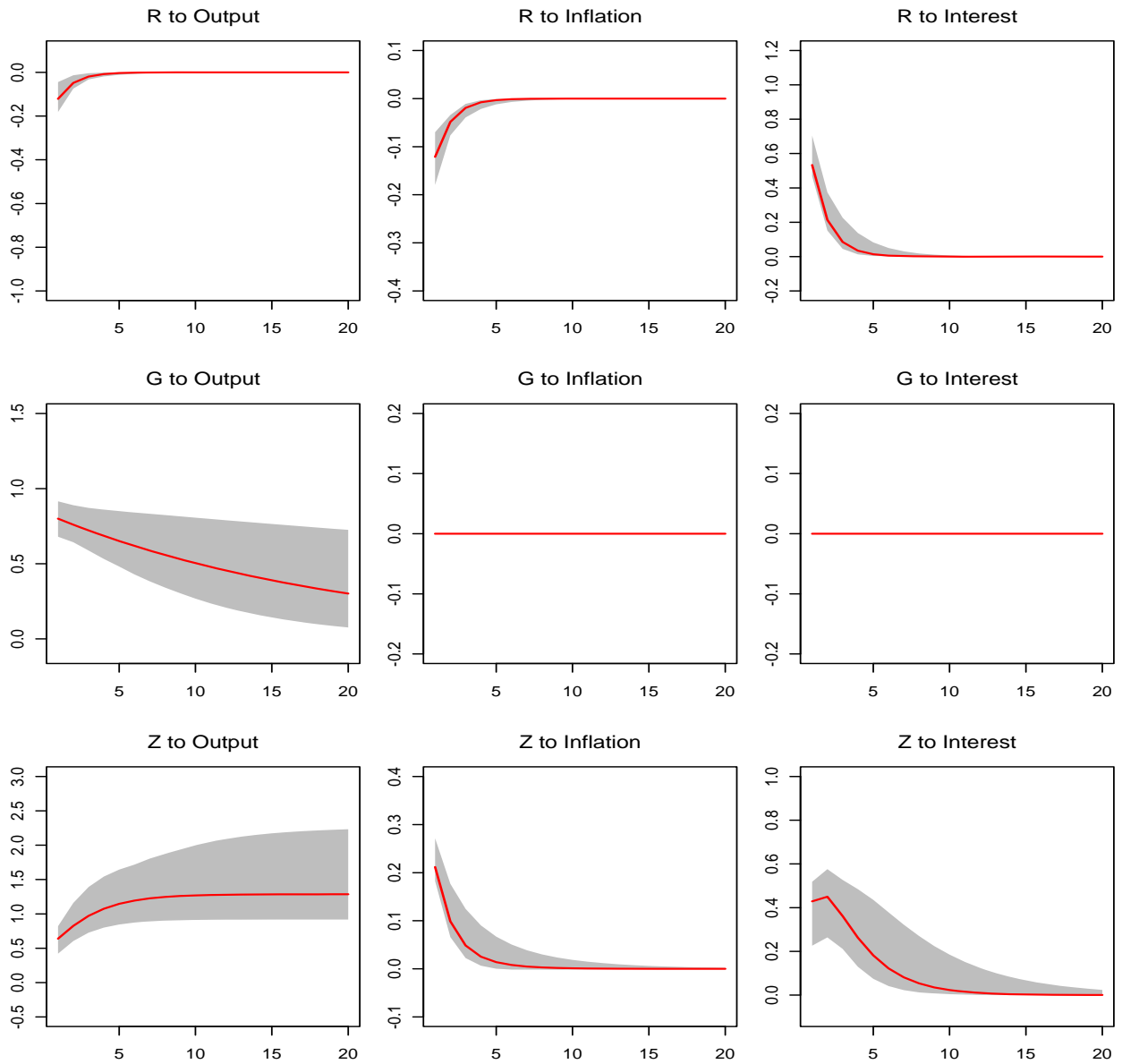


Figure 2. Uniform confidence bands for impulse response functions
(90%, BC frequencies)



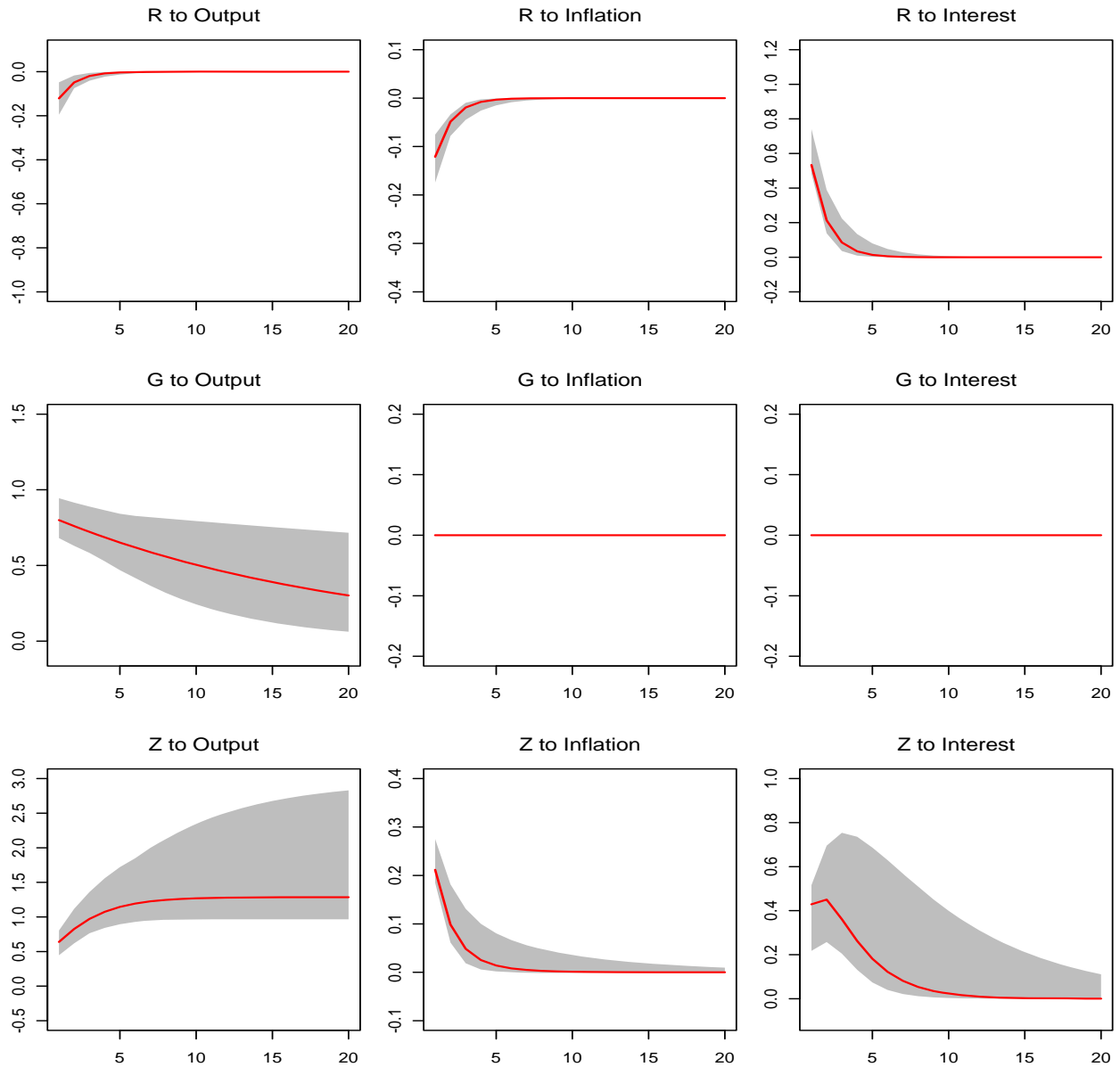
Note. R, G and Z: shocks to monetary policy, exogenous spending and technology. Gray area: the uniform band. Solid line: the true impulse response. Y-axis: percent. X-axis: horizon.

Figure 3. 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 4. 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.

Online Appendix

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

Step 1. First consider ξ_{1T} . Rewrite it as

$$\xi_{1T} = \frac{1}{2\sqrt{T}} \sum_{j=1}^{T-1} \phi_T(\omega_j)^* \text{vec}(I_T(\omega_j) - \mathbb{E}I_T(\omega_j)) \quad (\text{B.1})$$

$$+ \frac{1}{2\sqrt{T}} \sum_{j=1}^{T-1} \phi_T(\omega_j)^* \text{vec}(\mathbb{E}I_T(\omega_j) - f_{\theta_0}(\omega_j)). \quad (\text{B.2})$$

The term (B.2) is asymptotically negligible. Specifically, $\mathbb{E}I_T(\omega)$ can be expressed as $\mathbb{E}I_T(\omega) = \sum_{s=-T+1}^{T-1} (1 - |s|/T) \Gamma(s) \exp(-is\omega)$ with $\Gamma(s) = (1/(2\pi)) \int_{-\pi}^{\pi} f_{\theta_0}(\omega) \exp(is\omega) d\omega$. 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) $\sup_{\omega \in [-\pi, \pi]} \|\text{vec}(\mathbb{E}I_T(\omega) - f_{\theta_0}(\omega))\| = O(T^{-\beta})$. The term (B.2) is therefore bounded by

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

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_{1T} , it suffices to consider (B.1) only.

Let $\phi_{TM}(\omega)$ denote the $(M-1)$ -th order Cesaro sum of the Fourier series for $\phi_T(\omega)$:

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

with $\eta_T(s) = 1/(2\pi) \int_{-\pi}^{\pi} \phi_T(\omega) \exp(is\omega) d\omega$. Then,

$$\begin{aligned} (\text{B.1}) &= \frac{1}{2\sqrt{T}} \sum_{j=1}^{T-1} \phi_{TM}(\omega_j)^* \text{vec}(I_T(\omega_j) - \mathbb{E}I_T(\omega_j)) \\ &\quad + \frac{1}{2\sqrt{T}} \sum_{j=1}^{T-1} (\phi_T(\omega_j) - \phi_{TM}(\omega_j))^* \text{vec}(I_T(\omega_j) - \mathbb{E}I_T(\omega_j)). \end{aligned} \quad (\text{B.3})$$

The second term will be asymptotically negligible if, because of conjugacy,

$$\frac{1}{2\sqrt{T}} \sum_{j=1}^{\lfloor T/2 \rfloor} (\phi_T(\omega_j) - \phi_{TM}(\omega_j))^* \text{vec}(I_T(\omega_j) - \mathbb{E}I_T(\omega_j)) = o_p(1). \quad (\text{B.4})$$

Establishing this result faces some difficulty because $\phi_T(\omega)$ has a finite number of discontinuities within $[0, \pi]$ due to the presence of $W(\omega)$, implying $\phi_T(\omega_j) - \phi_{TM}(\omega_j)$ does not converge uniformly

to zero over $[0, \pi]$ (the Gibbs phenomenon). However, results in Hannan (1970, p. 506-507) imply that $\phi_{TM}(\omega)$ converges uniformly to $\phi_T(\omega)$ over all closed intervals excluding the jumps. At the jumps, the approximation errors remain bounded. Assume the jumps occur at $\tilde{\omega}^k$ ($k = 1, \dots, K$). Then, for any $\varepsilon > 0$ there exist finite constants $M > 0$ and $C > 0$ independent of T , such that

$$\|\phi_{TM}(\omega) - \phi_T(\omega)\| \leq \begin{cases} C & \text{if } \omega \in I_1 \equiv \cup_{k=1}^K [\tilde{\omega}^k - \varepsilon, \tilde{\omega}^k + \varepsilon] \\ \varepsilon & \text{if } \omega \in [0, \pi] \text{ but } \omega \notin I_1. \end{cases}$$

Apply the above partition, (B.4) can be decomposed into

$$\frac{1}{2\sqrt{T}} \sum_{j=1}^{[T/2]} 1(\omega_j \in I_1) (\phi_T(\omega_j) - \phi_{TM}(\omega_j))^* \text{vec}(I_T(\omega_j) - EI_T(\omega_j)) \quad (\text{T1})$$

$$+ \frac{1}{2\sqrt{T}} \sum_{j=1}^{[T/2]} 1(\omega_j \notin I_1) (\phi_T(\omega_j) - \phi_{TM}(\omega_j))^* \text{vec}(I_T(\omega_j) - EI_T(\omega_j)) \quad (\text{T2}).$$

For the first term:

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

where D is some finite constant and the second inequality follows from Theorem 11.7.1 in Brockwell and Davis (1991), i.e., for any ω_j and ω_h in $[0, \pi]$,

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

Because the length of I_1 can be made arbitrarily small by choosing a small ε and a large M , we have $Var(\text{T1}) = o(1)$. Similar arguments can be applied to (T2):

$$\|Var(\text{T2})\| \leq T^{-1}\varepsilon^2 \sum_{j=1}^{[T/2]} 1(\omega_j \notin I_1) D + \varepsilon^2 T^{-2} \sum_{j=1}^{[T/2]} \sum_{h=1}^{[T/2]} 1(\omega_j \notin I_1) D \leq 2D\varepsilon^2,$$

which can again be made small by choosing a small ε and a large M . Thus, $Var(\text{T2}) = o(1)$. Combining the above results, we have proved (B.4).

It remains to analyze the first term in (B.3). Apply the definition of $\phi_{TM}(\omega_j)$:

$$\begin{aligned} & \frac{1}{2\sqrt{T}} \sum_{j=1}^{T-1} \phi_{TM}(\omega_j)^* \text{vec} (I_T(\omega_j) - \mathbb{E}I_T(\omega_j)) \\ &= \frac{1}{4\pi} \sum_{s=-M-1}^{M-1} \left(1 - \frac{|s|}{M}\right) \eta_T(s)^* \left\{ \sqrt{T} \text{vec} \left(\hat{\Gamma}(s) - \mathbb{E}\hat{\Gamma}(s) \right) \right\} \end{aligned} \quad (\text{T3})$$

$$+ \frac{1}{4\pi} \sum_{s=-M-1}^{M-1} \left(1 - \frac{|s|}{M}\right) \eta_T(s)^* \left\{ \sqrt{T} \text{vec} \left(\hat{\Gamma}(s-T) - \mathbb{E}\hat{\Gamma}(s-T) \right) \right\}, \quad (\text{T4})$$

where the last equality uses $\sum_{s=1}^T \exp(-is\omega_j) = 0$ unless $s = kT$ ($k = 0, \pm 1, \dots$) and

$$\hat{\Gamma}(s) = \begin{cases} T^{-1} \sum_{t=1}^{T-s} (Y_{t+s} - \mu(\theta_0)) (Y_t - \mu(\theta_0))' & \text{if } 0 \leq s \leq T-1 \\ \hat{\Gamma}(-s)' & \text{if } -T+1 \leq s \leq 0. \end{cases}$$

Term (T4) converges in probability to zero. This is because M is finite, $\eta_T(s)^*$ is uniformly bounded and $\sqrt{T} \text{vec}(\hat{\Gamma}(s-T) - \mathbb{E}\hat{\Gamma}(s-T)) \rightarrow^p 0$ 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), $\sqrt{T} \text{vec}(\hat{\Gamma}(s) - \mathbb{E}\hat{\Gamma}(s))$ satisfies a central limit theorem for each $|s| \leq M$, see Hannan (1976). Thus, (T3) converges to a vector of normal random variables because M is finite. Therefore, ξ_{1T} has a multivariate normal limiting distribution. For ξ_{2T} , ψ_T is finite because of Assumption W. Its asymptotic normality then follows from the central limit theorem.

Step 2. For ξ_{1T} it suffices to examine the covariance matrix of (T3). 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

$$\xi_{TM}(l) = \frac{1}{8\pi^2} \sum_{s=-M-1}^{M-1} \left(1 - \frac{|s|}{M}\right) \int_{-\pi}^{\pi} W(\omega) \text{tr} \left\{ B(l, \omega) \sqrt{T} (\hat{\Gamma}(s) - \mathbb{E}\hat{\Gamma}(s)) \right\} \exp(-is\omega) d\omega, \quad (\text{B.5})$$

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 $[\cdot]_{kl}$ denoting the (k, l) -th element of the matrix inside the bracket. Because $B(l, \omega)$ is an n_Y -by- n_Y matrix, (B.5) can be further rewritten as

$$\xi_{TM}(l) = \frac{1}{8\pi^2} \sum_{s=-M-1}^{M-1} \left(1 - \frac{|s|}{M}\right) \int_{-\pi}^{\pi} W(\omega) \left\{ \sum_{a,b=1}^{n_Y} B_{ba}(l, \omega) \sqrt{T} \left(\hat{\Gamma}_{ab}(s) - \mathbb{E}\hat{\Gamma}_{ab}(s) \right) \right\} \exp(-is\omega) d\omega,$$

where $B_{ba}(l, \omega)$ is the (b, a) -th element of the corresponding matrix. Therefore,

$$\begin{aligned} & \text{Cov}(\xi_{TM}(l), \xi_{TM}(k)) \\ &= \frac{1}{64\pi^4} \sum_{a,b,c,d=1}^{n_Y} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W(r) B_{ba}(l, r) W(\lambda) B_{dc}(k, \lambda)^* \sum_{s,h=-M+1}^{M-1} \left(1 - \frac{|s|}{M}\right) \left(1 - \frac{|h|}{M}\right) \\ & \quad \times \mathbb{E} \left\{ T \left(\hat{\Gamma}_{ab}(s) - \mathbb{E}\hat{\Gamma}_{ab}(s) \right) \left(\hat{\Gamma}_{cd}(h) - \mathbb{E}\hat{\Gamma}_{cd}(h) \right) \right\} \exp(-isr) dr \exp(ih\lambda) d\lambda \end{aligned} \quad (\text{B.6})$$

The only random elements in (B.6) are the sample autocovariances, which satisfy (see eq. (3) in p.397 in Hannan, 1976)

$$\begin{aligned} & \mathbb{E} \left\{ T \left(\hat{\Gamma}_{ab}(s) - \mathbb{E} \hat{\Gamma}_{ab}(s) \right) \left(\hat{\Gamma}_{cd}(h) - \mathbb{E} \hat{\Gamma}_{cd}(h) \right) \right\} \\ & \rightarrow 2\pi \int_{-\pi}^{\pi} f_{ac}(\omega) \overline{f_{bd}(\omega)} \exp(-i(h-s)\omega) d\omega \quad (\text{T5}) \end{aligned}$$

$$+ 2\pi \int_{-\pi}^{\pi} f_{ad}(\omega) \overline{f_{bc}(\omega)} \exp(i(s+h)\omega) d\omega, \quad (\text{T6})$$

where $f_{ac}(\omega)$ stands for the (a,c)-th element of $f_{\theta_0}(\omega)$. Applying (T5) to (B.6) leads to

$$\begin{aligned} & \frac{1}{8\pi} \int_{-\pi}^{\pi} \sum_{a,b,c,d=1}^q f_{ac}(\omega) \overline{f_{bd}(\omega)} \left(\int_{-\pi}^{\pi} W(r) B_{ba}(l, r) \left\{ \frac{1}{2\pi} \sum_{s=-M+1}^{M-1} \left(1 - \frac{|s|}{M} \right) \exp(-is(r-\omega)) \right\} dr \right) \\ & \times \left(\int_{-\pi}^{\pi} W(\lambda) B_{dc}(k, \lambda)^* \left\{ \frac{1}{2\pi} \sum_{h=-M+1}^{M-1} \left(1 - \frac{|h|}{M} \right) \exp(-ih(\omega-\lambda)) \right\} d\lambda \right) d\omega. \end{aligned}$$

The two terms inside the two curly brackets are Fejér's kernels. Therefore,

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

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

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

where the first equality uses $B_{dc}(k, \omega)^* = B_{cd}(k, \omega)$, $\overline{f_{bd}(\omega)} = f_{db}(\omega)$, $\overline{f_{bc}(\omega)} = f_{cb}(\omega)$, $B_{dc}(k, -\omega)^* = B_{dc}(k, \omega)$ and the last equality follows because the summand belongs to $\text{Lip}(\beta)$ with $\beta > 1/2$. In matrix notation, the above result can be stated as

$$\begin{aligned} & Var(\xi_{1T}) \\ & = \Lambda_T^c(\theta_0)^{-1/2} Q_T^c(\theta_0)' \\ & \times \left\{ \frac{1}{2T} \sum_{j=0}^{T-1} W(\omega_j) \left(\frac{\partial \text{vec} f_{\theta_0}(\omega_j)}{\partial \theta'} \right)^* (f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j)) \frac{\partial \text{vec} f_{\theta_0}(\omega_j)}{\partial \theta'} \right\} Q_T^c(\theta_0) \Lambda_T^c(\theta_0)^{-1/2} + o(1). \end{aligned}$$

Now consider ξ_{2T} . It is asymptotically independent of ξ_{1T} , satisfying

$$\text{Var}(\xi_{2T}) \rightarrow \frac{1}{2\pi} \Lambda_T^c(\theta_0)^{-1/2} Q_T^c(\theta_0)' \left\{ W(0) \frac{\partial \mu(\theta_0)'}{\partial \theta} f_{\theta_0}^{-1}(0) \frac{\partial \mu(\theta_0)}{\partial \theta'} \right\} Q_T^c(\theta_0) \Lambda_T^c(\theta_0)^{-1/2}.$$

Therefore, $\text{Var}(\xi_{1T} + \xi_{2T}) = \Lambda_T^c(\theta_0)^{-1/2} Q_T^c(\theta_0)' M_T(\theta_0) Q_T^c(\theta_0) \Lambda_T^c(\theta_0)^{-1/2} + o(1) \rightarrow \mathbb{I}_{q_1+q_2}$, where the last equality uses the definition of $Q_T^c(\theta_0)$ and $\Lambda_T^c(\theta_0)$. ■

Additional Proof. This proof shows that the confidence band covers the impulse response function with probability at least $(1 - \alpha)$ asymptotically. Let $C_\theta(1 - \alpha)$ denote the $(1 - \alpha)$ confidence set for θ obtained by inverting $S_T(\theta)$ and C_{IR} the confidence band for the impulse response function obtained from Steps 1 to 3. By construction, if $\theta_0 \in C_\theta(1 - \alpha)$, then $IR(\theta_0) \in C_{IR}$. Thus, if $IR(\theta_0) \notin C_{IR}$, then $\theta_0 \notin C_\theta(1 - \alpha)$. Equivalently, $\Pr(IR(\theta_0) \notin C_{IR}) \leq \Pr(\theta_0 \notin C_\theta(1 - \alpha))$. As $T \rightarrow \infty$, $\Pr(\theta_0 \notin C_\theta(1 - \alpha)) \rightarrow \alpha$. Therefore, $\lim_{T \rightarrow \infty} \Pr(IR(\theta_0) \notin C_{IR}) \leq \alpha$. ■

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 Assumption W.

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,

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

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 $\hat{V}_{ff}(\theta) \xrightarrow{p} V_{ff}(\theta) = \lim_{T \rightarrow \infty} \text{Var}(f_T(\theta))$. Define

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

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

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

$$+ 2T^{-1/2}f_T(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1}\mathbb{E}q_T(\theta_0), \quad (\text{b})$$

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

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

The first term $\mathbb{E}(a'a)$ is of order $O(T^{-1})$ 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

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

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

$$D_T(\theta_0)' = 2T^{-1/2}f_T(\theta_0)'W_T\left(q_T(\theta_0) - \hat{R}_T(\theta_0)\right) \quad (\text{c})$$

$$+ 2T^{-1/2}f_T(\theta_0)'W_T\left(\hat{R}_T(\theta_0) - \mathbb{E}q_T(\theta_0)\right) \quad (\text{d})$$

$$+ 2T^{-1/2}f_T(\theta_0)'W_T\mathbb{E}q_T(\theta_0). \quad (\text{e})$$

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)

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

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

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

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 $\mathbb{E}(D_T(\theta_0) D_T(\theta_0)')$.

Weak identification in a two-equation model

The model 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 $\text{var}(u_t) = \sigma_u^2$, $\text{var}(v_t) = \sigma_v^2$, $\text{cov}(u_t, v_t) = \sigma_{uv}$ and $E u_t u_s = E v_t v_s = E u_t v_s = 0$ for all $t \neq s$. 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 parameter of interest is β . To simplify the derivation, we assume 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

$$\begin{aligned}
\tilde{r}_t &= \beta \pi_t + u_t, \\
\pi_t &= \rho \pi_{t-1} + v_t
\end{aligned} \tag{B.7}$$

with $\tilde{r}_t = r_t - \gamma y_t$. 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).

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

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 $T\lambda_{1T} \rightarrow \infty$ and $T\lambda_{2T} \rightarrow \infty$.*
2. *The smallest eigenvalue λ_{3T} satisfies*

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

3. The elements of

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

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).