Inference in dynamic stochastic general equilibrium models with possible weak identification

Zhongjun Qu
Department of Economics, Boston University

This paper considers inference in log-linearized dynamic stochastic general equilibrium (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 that 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-squared 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 response 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.

1. Introduction

Dynamic stochastic general equilibrium (DSGE) models play an important role in quantitative macroeconomics. Frequentist inference in such models is challenging. The like-

lihood 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 instrumental variable (IV) and generalized method of moments (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 provide 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 ongoing work, Andrews and Mikusheva (2013) studied two Lagrange multiplier (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 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 the 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 characterizing 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 nonrandom 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 for why the test statistic has a chi-squared 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, Oha-
nian, and Berkowitz (1998) discussed a general framework for loss function based esti-
mation and model evaluation. Christiano and Vigfusson (2003) applied frequency do-
main 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 frame-
work 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 character-
izes weak identification from a frequency domain perspective. Section 5 proposes the
score test and studies its asymptotic properties. Section 6 discusses how to obtain ro-
bust confidence sets. It also considers uniform confidence bands for impulse response
functions and some other objects. Section 7 considers an illustrative model and Sec-
tion 8 concludes. Additional information is provided in supplementary files on the jour-
nal website, http://qeconomics.org/supp/287/supplement.pdf and http://qeconomics.
org/supp/287/code_and_data.zip.

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 vari-
ety 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 repre-
sentations. 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. For a log-linearized system, Sims (2002)
considered the representation

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

where $\Gamma_0(\theta)$, $\Gamma_1(\theta)$, $\Psi(\theta)$, and $\Pi(\theta)$ are coefficient 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,

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

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

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

Its spectral density is given by

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

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

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

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

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

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 let $\|X\|$ 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 $f_\theta \in R^k$ of $\theta \in R^q$, let $\partial f_{\theta_0}/\partial \theta'$ be the k-by-q matrix of partial derivatives evaluated at θ_0.

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

Assumption 2. *The solution is unique and is representable as*

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

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

Assumption 3. *There exists $0 < C_L \le C_U < \infty$ such that for all $\omega \in [-\pi, \pi]$ and all $\theta \in \Theta$, the following statements hold:*

(i) *We have $C_L \le \text{eig}(f_\theta(\omega)) \le C_U$.*

(ii) *The elements of $f_\theta(\omega)$ belong to* Lip(β) *with $\beta > \frac{1}{2}$ with respect to ω.*[1]

(iii) *We have $\|\frac{\partial \text{vec} f_\theta(\omega)}{\partial \theta'}\| \leq C_U$ and the elements of $\frac{\partial \text{vec} f_\theta(\omega)}{\partial \theta'}$ belong to* Lip(β) *with $\beta > \frac{1}{2}$ with respect to ω.*

(iv) *We have $\|\partial \mu(\theta)/\partial \theta'\| \leq C_U$.*

ASSUMPTION 4. *The observable random variables $\{Y_t\}_{t=1}^T$ form 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

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

This assumption allows for noninvertibility and is weaker than assuming $Y_t^d(\theta)$ follows a vector autoregression (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 to be finite and nonsingular. If unit roots are present in the DSGE model, then it requires appropriately differencing the series prior to applying the methods. Assumptions 3(ii) and (iii) assume the spectral density and its first derivatives are smooth in ω. They can be verified under more primitive conditions. Specifically, Assumption 3(ii) is satisfied if $\sum_{j=0}^\infty j^\beta \|h_j(\theta)\| \leq \infty$ (Hannan (1970, pp. 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 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 that affect only the mean of Y_t) based on the full or parts of the spectrum is treated as a special case.

---

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

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, that is, $\omega_j = 2\pi j/T$ ($j = 1, 2, \ldots, T - 1$). 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 $w_T(\omega_j)$ ($j = 1, 2, \ldots, T - 1$) have complex-valued multivariate normal distributions with densities (Hannan (1970, pp. 223–225))

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

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 to a constant addition, is given by

$$-\frac{1}{2} \sum_{j=1}^{T-1} \big[\log \det(f_\theta(\omega_j)) + \operatorname{tr}\{f_\theta^{-1}(\omega_j)I_T(\omega_j)\}\big]$$

$$-\frac{1}{2}\big[\log \det(f_\theta(0)) + \operatorname{tr}\{f_\theta^{-1}(0)I_{\theta,T}(0)\}\big]. \tag{5}$$

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 generalized version of (5),

$$L_T(\theta) = -\frac{1}{2} \sum_{j=1}^{T-1} W(\omega_j)\big[\log \det(f_\theta(\omega_j)) + \operatorname{tr}\{f_\theta^{-1}(\omega_j)I_T(\omega_j)\}\big]$$

$$-\frac{1}{2}W(0)\big[\log \det(f_\theta(0)) + \operatorname{tr}\{f_\theta^{-1}(0)I_{\theta,T}(0)\}\big], \tag{6}$$

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 is stated using the eigenvalues of the information matrix. Some eigenvalues approach zero as $T \to \infty$, such that the local curvature of the likelihood remains small in some directions in the presence of a large sample size.

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

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

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

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

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 eigendecomposition always exists,

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

where the columns of $Q_T(\theta_0)$ are the orthonormal eigenvectors and $\Lambda_T(\theta_0)$ contains the eigenvalues in a nonincreasing 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) *We have* $\Lambda_{3T}(\theta_0) = 0$ *for any* T. (iv) *The elements of*

$$\left[\partial \operatorname{vec} f_{\theta_0}(\omega)/\partial \theta'\right] Q_T(\theta_0) \Lambda_T^+(\theta_0)^{1/2} \tag{10}$$

are finite and belong to $\operatorname{Lip}(\beta)$ *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* $\Lambda_T(\theta_0)$.

Assumptions W(i)–(iii) allow for different degrees of identification. The component $\Lambda_{1T}(\theta_0)$ corresponds to parameter directions that are strongly or semistrongly identified (the latter notion follows Andrews and Cheng (2012)). The component $\Lambda_{2T}(\theta_0)$ implies 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 that 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.

Assumptions W(i)–(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 a supplementary file on the journal website, http://qeconomics.org/supp/287/supplement.pdf. 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 Assumption W(i)–(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.

Assumption W(iv) strengthens Assumption 3(iii) by requiring sufficient smoothness of $\partial \operatorname{vec} f_{\theta_0}(\omega)/\partial \theta'$ in ω. The effect of $Q_T(\theta_0)$ is to map the row vectors of $\partial \operatorname{vec} f_{\theta_0}(\omega)/\partial \theta'$ into a new coordinate system common to all $\omega \in [-\pi, \pi]$. The assumption thus requires $\partial \operatorname{vec} f_{\theta_0}(\omega)/\partial \theta'$ to be well behaved in this new coordinate system. If θ_0 is strongly identi-

fied, 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 supplementary files, 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 nonrandom 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 complex-valued multivariate regression

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

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 squares 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 (ESS) is

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

To establish the relationship 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 eigendecomposition (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* $S_T(\theta_0) = (1/2)\text{ESS}$ *with*

$$
X_j = \begin{bmatrix} W(\omega_j)\big(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j)\big)\dfrac{\partial \text{vec}\, f_{\theta_0}(\omega_j)}{\partial \theta'} \\ W(0)\big(\pi f_{\theta_0}(0)\big)^{-1/2}\dfrac{\partial \mu(\theta_0)}{\partial \theta'} \end{bmatrix},
$$

$$
\mathcal{Y}_j = \begin{bmatrix} W(\omega_j)\big(f_{\theta_0}^{-1/2}(\omega_j)' \otimes f_{\theta_0}^{-1/2}(\omega_j)\big)\text{vec}\big(I_T(\omega_j) - f_{\theta_0}(\omega_j)\big) \\ W(0)\big(\pi f_{\theta_0}(0)\big)^{-1/2} T^{-1}\displaystyle\sum_{t=1}^{T}\big(Y_t - \mu(\theta_0)\big) \end{bmatrix}
$$

(13)

for $j = 0, 1, \ldots, T-1$, *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 $j = 0, \ldots, T-1$. 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 $j = 0, \ldots, T-1$. 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-squared 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 $(1/2)\text{ESS}$ 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, on 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 to use such a relationship to understand testing procedures under weak identification.

THEOREM 1. *Let Assumptions* 1–4 *and* W *hold. Then*

$$
\lim_{T \to \infty} \Pr\big(S_T(\theta_0) \le c\big) \to \Pr\big(\chi_r^2 \le c\big),
$$

where χ_r^2 *is a chi-squared 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 is dependent on the third and fourth moments of the structural shocks. Specifically, its (j, l)th element equals

$$
\left[M_T(\theta_0)\right]_{jl} + \left(\frac{1}{4\pi}\right)^2 \sum_{a,b,c,d=1}^{n_\epsilon} \kappa_{abcd}
$$

$$
\times \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)H^*(\omega)\frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_j}H(\omega)\,dw\right]_{ab} \tag{14}
$$

$$
\times \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)H^*(\omega)\frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial \theta_l}H(\omega)\,dw\right]_{cd} + \frac{1}{8\pi^2}(A_{jl} + A_{lj}),
$$

where $[\cdot]_{ab}$ denotes the (a, b)th element of the matrix, $[\cdot]_c$ denotes the cth 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 jth 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

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

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, overrejection 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 nonidentification curves by conditioning on the smallest rank reported. (Under correct rank estimation, the curves consist of parameter values that are 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 nonidentification 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 overestimating the rank. It may result in underestimation, which is further addressed in Step 2. The idea is that 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.

On 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 underestimated. Intuitively, this is because the test behaves asymptotically as the sum of independent χ_1^2 variables. Setting a nonzero eigenvalue to zero is equivalent to removing a variable from the sum. This does not alter the chi-squared distribution except that the degrees of freedom is reduced by 1. Some power might be lost if the eigenvector cor-

responds 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 (2013) emphasized that "many time series exhibit low frequency behavior that is difficult, if not impossible, to reconcile with the model being estimated. This low frequency misspecification contaminates the estimation of shocks and thereby inference about the sources of business cycle." Therefore, it is valuable to be able to conduct inference by 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., autoregressive moving average (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, generalized autoregressive conditional heteroscedasticity, 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

$$C_\theta(1 - \alpha) = \left\{\theta \in \Theta : S_T(\theta) \le \chi^2_{q-q_3}(1 - \alpha)\right\},$$

where $\chi_s^2(1-\alpha)$ denotes the $100(1-\alpha)$th percentile of a chi-squared 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 quasiposterior will be proper even if $S_T(\theta)$ is flat. The basic steps for constructing the confidence set are the following.

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

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

STEP 3. Calculate the ratio

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

and set

$$\theta^{(j+1)} = \begin{cases} \theta^* & \text{with probability } s, \\ \theta^{(j)} & \text{with probability } 1-s. \end{cases}$$

STEP 4. Increase j by 1 and then repeat Steps 2 and 3. Continue until $j = B$ for some large B.

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

The above procedure is motivated by Chernozhukov and Hong (2003) in which Markov chain Monte Carlo (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 1–5 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 incorporate the following elements when we analyze 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 draw gives $\varepsilon = cV_T(\theta^{(j)})$ or $-cV_T(\theta^{(j)})$ with $V_T(\theta^{(j)})$ being the eigenvector that corresponds 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, that is, we use the first k Cartesian coordinates of the MCMC draws in Step 5 to form a confidence set for the first k parameters in θ. Such a method is implemented in Guerron-Quintana, Inoue, and Kilian (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 $\mathrm{IR}_{jl}(k, \theta_0)$ be the impulse response of the jth variable in Y_t to the lth 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 $\mathrm{IR}_{jl}(k, \theta_0)$.

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

$$\lim_{T \to \infty} \inf \Pr\big(\mathrm{IR}_{jl}(k, \theta_0) \in C_{\mathrm{IR}}(k; T) \text{ for all } k \in [0, \infty)\big) \geq 1 - \alpha.$$

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 jth variable in Y_t to the lth orthogonal shock. This function, when evaluated at horizon k, equals the (j, l)th element of $\mathrm{IR}(k, \theta) = h_k(\theta)\Sigma^{1/2}(\theta)$, where $h_k(\theta)$ is the kth coefficient 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 $A(L) = A$)

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

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 that satisfy $S_T(\theta) \leq \chi^2_{q-q_3}(1 - \alpha)$.

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

Because IR(k, θ) 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 supplementary file.) 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 jth column of an identity matrix whose dimension depends on the context.

Individual spectrum and coherency   The spectrum of the jth 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 jth and lth variables 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 of interest. Whiteman (1984) observed that the slopes are related to the coherency between the respective variables at frequency zero. In our notation, the estimated slope approximates

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

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

---

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

a Bayesian perspective. The methods developed in this paper can be used to construct frequentist confidence intervals for (16), and, 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 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, taken from An and Schorfheide (2007), is

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

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 gross domestic product (GDP) growth (YGR_t), inflation (INFL_t) and interest rate (INT_t),

$$
\begin{aligned}
\mathrm{YGR}_t &= \gamma^{(Q)} + 100(y_t - y_{t-1} + z_t), \\
\mathrm{INFL}_t &= \pi^{(A)} + 400\pi_t, \\
\mathrm{INT}_t &= \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400 r_t,
\end{aligned}
\tag{18}
$$

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

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

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):

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

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

Design 1 (Business cycle frequencies). Inference on the 11 dynamic parameters based on business cycle (BC) 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 $\partial\,\mathrm{vec}\,f_\theta(\omega)/\partial\theta'$ 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 1 and 3. The MATLAB default tolerance level yields $\mathrm{rank}(M_T(\theta_0)) = 10 < 11$. The method of Qu and Tkachenko (2012, Section 3.1) shows that there exists a unique nonidentification 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 varies 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 $100\sigma_r$ changes 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 deterministic 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 1, the minimal rank equals 9 when considering a wide range of tolerance levels: $10^j \times \dim(M_T(\theta_0))\,\mathrm{eps}(\|M_T(\theta_0)\|)$ $(j = -10, -9, \ldots, 10)$. 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 ψ_1, ψ_2, ρ_r, and σ_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.

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.

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

$$
\begin{pmatrix} z_{t+1} \\ g_{t+1} \\ r_{t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 0.65 & 0 & 0 \\ 0 & 0.95 & 0 \\ 0.1548 & 0 & 0.4 \end{pmatrix}}_{A} \begin{pmatrix} z_t \\ g_t \\ r_t \end{pmatrix} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.2382 & 0 & 0.6667 \end{pmatrix}}_{B} \underbrace{\begin{pmatrix} \epsilon_{zt+1} \\ \epsilon_{gt+1} \\ \epsilon_{rt+1} \end{pmatrix}}_{\epsilon_{t+1}},
$$

$$
\begin{pmatrix} r_{t+1} \\ y_{t+1} \\ \pi_{t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 0.1548 & 0 & 0.4 \\ 0.2724 & 0.95 & -0.3637 \\ 0.0764 & 0 & -0.0909 \end{pmatrix}}_{C} \begin{pmatrix} z_t \\ g_t \\ r_t \end{pmatrix}
$$

$$
+ \underbrace{\begin{pmatrix} 0.2382 & 0 & 0.6667 \\ 0.4191 & 1 & -0.6061 \\ 0.1176 & 0 & -0.1514 \end{pmatrix}}_{D} \begin{pmatrix} \epsilon_{zt+1} \\ \epsilon_{gt+1} \\ \epsilon_{rt+1} \end{pmatrix}.
$$

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 nonzero 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 that contribute 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. Guerron-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^2_{10} (for dynamic parameters) and χ^2_{12} (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 overrejections 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 overrejection 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: $\tau \sim [1\mathrm{E}{-}5, 5]$, $\kappa \sim [0, 1]$, $\psi_1 \sim [0, 5]$, $\psi_2 \sim [0, 2]$, $\rho_r \sim [0, 0.9]$, $\rho_g \sim [0, 0.99]$, $\rho_z \sim [0, 0.99]$, $100\sigma_r \sim [1\mathrm{E}{-}5, 2]$, $100\sigma_g \sim [1\mathrm{E}{-}5, 2]$, $100\sigma_z \sim [1\mathrm{E}{-}5, 2]$, $r^{(A)} \sim [0, 5]$, $\pi^{(A)} \sim [0, 20]$, and $\gamma^{(Q)} \sim [0, 5]$. 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.

TABLE 2. Rejection frequencies under the null hypothesis.

	BC Frequencies		Full Spectrum		Mean and Full Spectrum	
T	5%	10%	5%	10%	5%	10%
	θ_0 taken from Table 2 in An and Schorfheide (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 specified in the last column of that table. The second panel: the prior distribution is specified in the third and fourth columns of that table; 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 nonzeros.)

	BC Frequencies		Full Spectrum		Mean and Full Spectrum	
T	5%	10%	5%	10%	5%	10%
	θ_0 taken from Table 2 in An and Schorfheide (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.

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 nonzeros, 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 and 8.2% and 7.9 and 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 and 10.8% and 10.8 and 15.8% at the two nominal levels. In the above discussion, the rank of the information matrix stayed fixed across parameter values. Yet another alternative

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 An and Schorfheide (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.

is to reestimate its rank with the MATLAB default 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 maximum likelihood estimator (MLE) $\hat{\theta}$ can lie close to a point distant from θ_0 on the nonidentification 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 nonidentification curve, the standard quadratic approximation is no longer adequate. Consequently, the chi-squared 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 nonstandard features. Instead, the key assumption that validates the asymptotic distribution is Assumption W(iv). In this model, $[\partial \operatorname{vec} f_{\theta_0}(\omega)/\partial\theta']Q_T(\theta_0)\Lambda_T^+(\theta_0)^{1/2}$ is an 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. Figure 1 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 $\{1, \ldots, 13\}$ (or $\{1, \ldots, 11\}$ 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
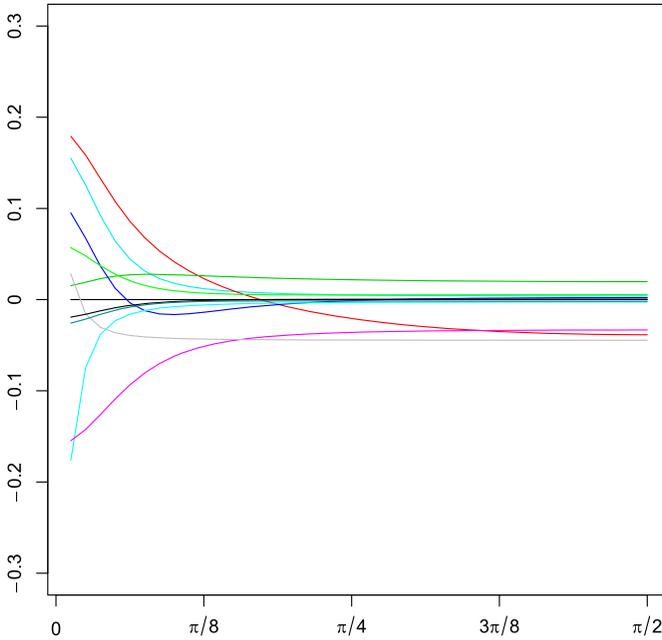
FIGURE 1. A graphical illustration of Assumption W(iv).

TABLE 5. Size adjusted power.

	BC Frequencies		Full Spectrum		Mean and Full Spectrum	
T	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).

others. This is repeated to generate 5000 parameter values that yield 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

TABLE 6. Size adjusted power under alternative computations of the test statistics. (All eigenvalues are treated as nonzeros.)

	BC Frequencies		Full Spectrum		Mean and Full Spectrum	
T	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.)

	BC Frequencies		Full Spectrum		Mean and Full Spectrum	
T	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.

power is not necessarily higher than under Design 2. The second panel corresponds to $\kappa = 40$. There the ratios are between 48.5 and 86.1%. Designs 2 and 3 continue to show similar rejection frequencies.

We misspecify the rank of the information matrix and compute the size-adjusted rejection frequencies. In Table 6, all eigenvalues are treated as nonzeros. 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.

TABLE 8. Andrews and Mikusheva's (2013) 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.

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 nonidentification 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 cannot 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, and, 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) suggested

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 Table 5. Specifically, the differences between LM_e and $S_T(\theta_0)$ are within 0.02–0.09, and between LM_o and $S_T(\theta_0)$ are within -0.17–0.10. 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 $T = 240$. The fourth column 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, pp. 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 (ρ_g, ρ_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, pp. 133–134).

The fifth column 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. The sixth column incorporates the

TABLE 9. Lengths of the 90% confidence intervals.

Parameter (1)	θ_0 (2)	Bounds (3)	BC Frequencies (4)	Full Spectrum (5)	Mean and Full Spectrum (6)
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.

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 ρ_g, ρ_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. Furthermore, even unidentified parameters themselves can have tight confidence intervals; see ρ_r and σ_r in the fifth and sixth columns. To see why the latter has happened in this model, note that ρ_r and σ_r are elements of the Taylor rule parameters (ψ_1, ψ_2, ρ_r, σ_r) that lie on a nonidentification 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 nonidentified 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 as 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
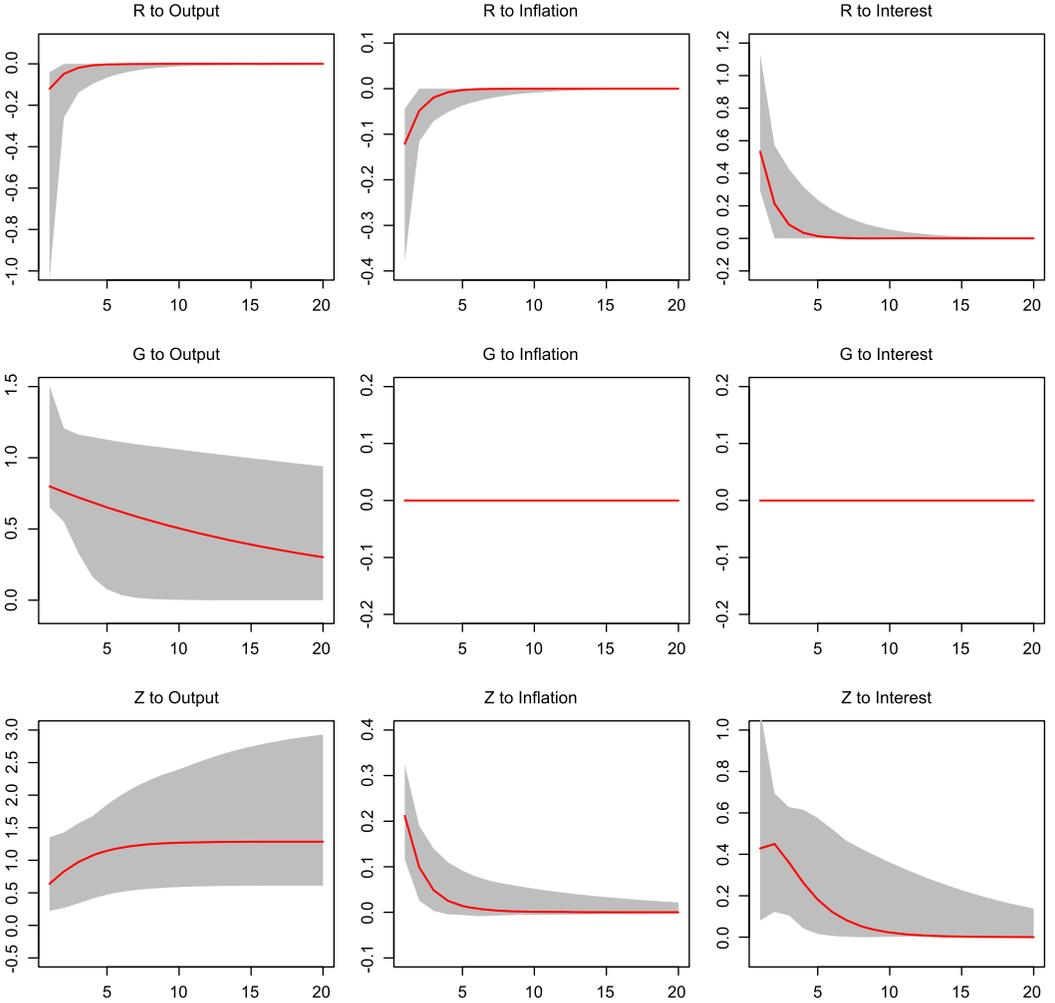
FIGURE 2. Uniform confidence bands for impulse response functions (90%, BC frequencies). *Note*: R, G, and Z: shocks to monetary policy, exogenous spending and technology. Gray area: the uniform band. Solid line: the true impulse response. Y-axis: percent. X-axis: horizon.

the latter. For a further illustration, we computed the impulse response functions using the 20 points reported in Table 1. The maximum difference between them is of order 1E−7. This confirms that in this 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

$$\epsilon_{rt} \stackrel{\text{i.i.d.}}{\sim} t\big(0, \sigma_r^2, \lambda_r\big), \qquad \epsilon_{gt} \stackrel{\text{i.i.d.}}{\sim} t\big(0, \sigma_g^2, \lambda_g\big), \qquad \epsilon_{zt} \stackrel{\text{i.i.d.}}{\sim} t\big(0, \sigma_z^2, \lambda_z\big),$$
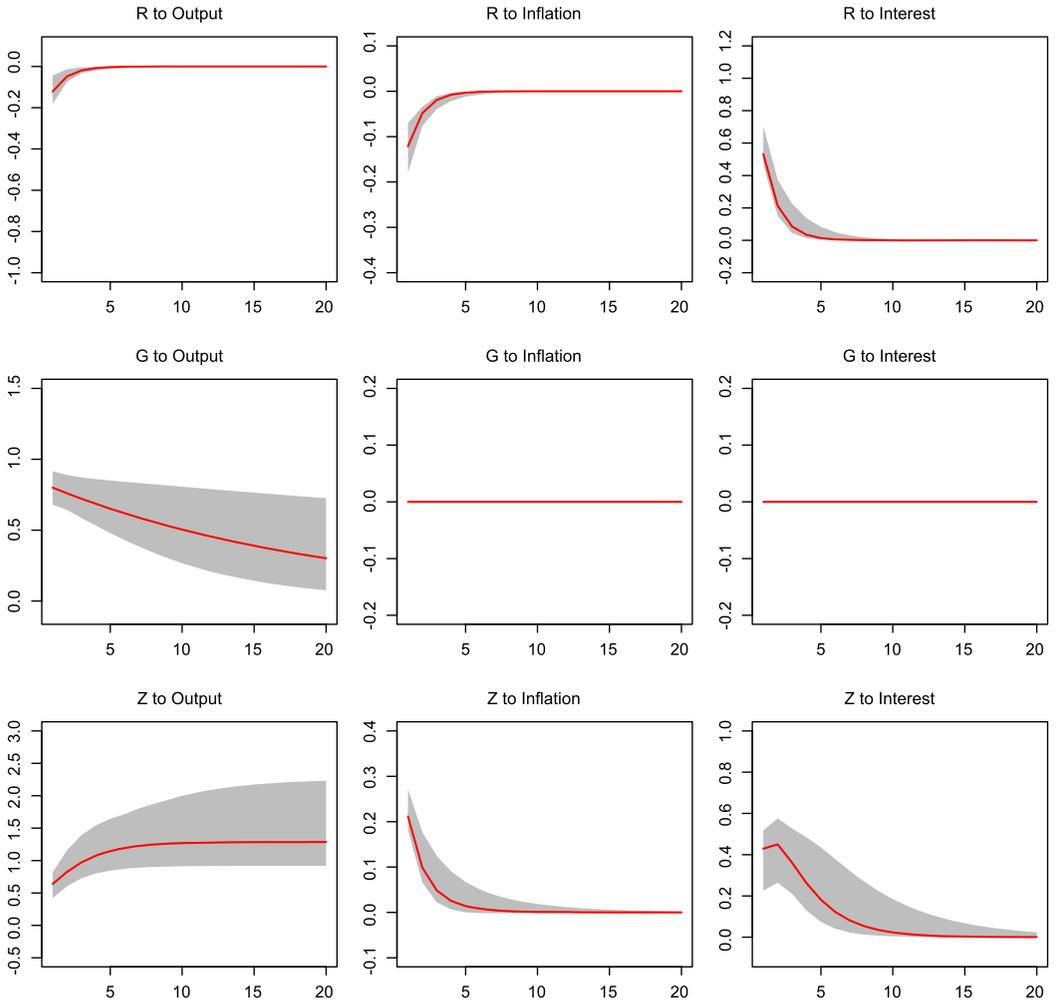
FIGURE 3. Uniform confidence bands for impulse response functions (90%, the full spectrum). *Note*: R, G, and Z: shocks to monetary policy, exogenous spending and technology. Gray area: the uniform band. Solid line: the true impulse response. Y-axis: percent. X-axis: horizon.

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.
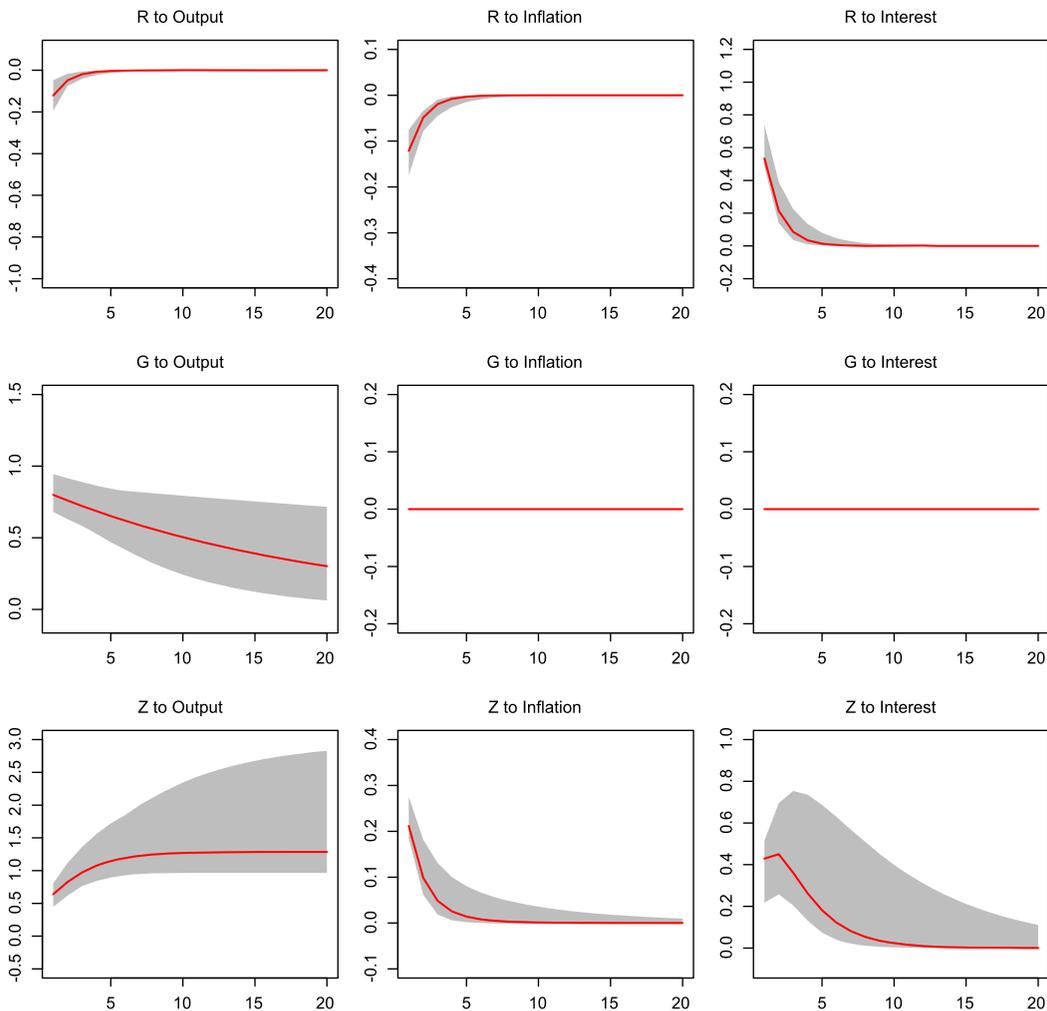
FIGURE 4. Uniform confidence bands for impulse response functions (90%, the mean and full spectrum). *Note*: R, G, and Z: shocks to monetary policy, exogenous spending and technology. Gray area: the uniform band. Solid line: the true impulse response. Y-axis: percent. X-axis: horizon.

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:

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

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

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

TABLE 10. Finite sample properties with Student-t innovations.

	BC Frequencies		Full Spectrum		Mean and Full Spectrum	
T	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).

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 simply can be 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.

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.

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 St. Louis Federal Reserve 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). Columns A correspond 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 and 8.4% and 11.8 and 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. Columns B correspond to (21). The BC based test continues to perform well even when the full spectrum based test rejects close to 100% of the time. Columns C correspond 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 relatively 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 factor augmented vector autoregression (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 (i) constructing uniform confidence bands for the impulse response functions, the time path of the variance decomposition, the individual spectrum, and absolute coherency, and (ii) examining certain low frequency hypotheses. Parameters in DSGE models are often highly correlated. This can be seen from the nonidentification 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.

Appendix A

The following lemma is needed for proving Theorem 1. Its proof, along with some other intermediate results, are in the supplementary file.

Lemma 2. *Suppose Assumptions 1–4 and W hold. Let $\Lambda_T^c(\theta_0)$ denote the upper-left nonzero 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*

$$\xi_{1T} = \frac{1}{2\sqrt{T}} \sum_{j=1}^{T-1} \phi_T(\omega_j)^* \operatorname{vec}(I_T(\omega_j) - f_{\theta_0}(\omega_j)),$$

$$\xi_{2T} = \frac{1}{2\pi\sqrt{T}} \sum_{t=1}^{T} \psi_T^*(Y_t - \mu(\theta_0)),$$

where

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

$$\psi_T = W(0) f_{\theta_0}^{-1}(0) \frac{\partial \mu(\theta_0)}{\partial \theta'} Q_T^c(\theta_0) \Lambda_T^c(\theta_0)^{-1/2}.$$

Then

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

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 semidefinite, 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,

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

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

REFERENCES

Altug, S. (1989), "Time-to-build and aggregate fluctuations: Some new evidence." *International Economic Review*, 30, 889–920. [459]

An, S. and F. Schorfheide (2007), "Bayesian analysis of DSGE models." *Econometric Reviews*, 26, 113–172. [457, 459, 471, 474, 477, 478, 479, 480, 482, 483, 488, 489]

Andrews, D. W. K. and X. Cheng (2012), "Estimation and inference with weak, semi-strong, and strong identification." *Econometrica*, 80, 2153–2211. [465]

Andrews, I. and A. Mikusheva (2013), "Maximum likelihood inference in weakly identified DSGE models." Working paper, MIT. [458, 482, 483]

Breusch, T. S. and A. R. Pagan (1980), "The Lagrange multiplier test and its applications to model specification in econometrics." *Review of Economic Studies*, 47, 239–253. [467]

Canova, F. and L. Sala (2009), "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics*, 56, 431–449. [458, 459]

Chernozhukov, V. and H. Hong (2003), "An MCMC approach to classical estimation." *Journal of Econometrics*, 115, 293–346. [471]

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. [460]

Cogley, T. (2001), "Estimating and testing rational expectations models when the trend specification is uncertain." *Journal of Economic Dynamics and Control*, 5, 1485–1525. [488]

Cúrdia, V., M. Del Negro, and D. L. Greenwald (2013), "Rare shocks, great recessions." Working paper, Federal Reserve Bank of San Francisco. [486, 488]

Davidson, R. and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*. Oxford University Press, London. [467]

Del Negro, M., F. X. Diebold, and F. Schorfheide (2008), "Priors from frequency domain dummy observations." Working paper, Federal Reserve Bank of New York. [458, 460]

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. [458, 490]

Diebold, F. X., L. E. Ohanian, and J. Berkowitz (1998), "Dynamic equilibrium economies: A framework for comparing models and data." *Review of Economic Studies*, 65, 433–451. [460]

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. [458]

Dufour, J. M. and M. Taamouti (2005), "Projection-based statistical inference in linear structural models with possibly weak instruments." *Econometrica*, 73, 1351–1365. [472]

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013), "Frequentist inference in weakly identified dynamic stochastic general equilibrium models." *Quantitative Economics*, 4, 197–229. [458, 465, 472, 477]

Hannan, E. J. (1970), *Multiple Time Series*. Wiley, New York. [462, 463]

Hansen, L. P. and T. J. Sargent (1993), "Seasonality and approximation errors in rational expectations models." *Journal of Econometrics*, 55, 21–55. [459, 463]

Horn, R. A. and C. R. Johnson (2005), *Matrix Analysis*. Cambridge University Press, Cambridge. [466]

Horn, R. A. and C. R. Johnson (2006), *Topics in Matrix Analysis*. Cambridge University Press, Cambridge. [466]

Iskrev, N. (2008), "Evaluating the information matrix in linearized DSGE models." *Economics Letters*, 99, 607–610. [483]

Iskrev, N. (2010), "Local identification in DSGE models." *Journal of Monetary Economics*, 57, 189–202. [458]

King, R. G. and M. W. Watson (1996), "Money, prices, interest rates and the business cycle." *Review of Economics and Statistics*, 78, 35–53. [464]

King, R. G. and M. W. Watson (2002), "System reduction and solution algorithms for singular linear difference systems under rational expectations." *Computational Economics*, 20, 57–86. [460]

Kleibergen, F. (2005), "Testing parameters in GMM without assuming that they are identified." *Econometrica*, 73, 1103–1123. [465]

Klein, P. (2000), "Using the generalized Schur form to solve a multivariate linear rational expectations model." *Journal of Economic Dynamics and Control*, 24, 1405–1423. [460]

Komunjer, I. and S. Ng (2011), "Dynamic identification of dynamic stochastic general equilibrium models." *Econometrica*, 79, 1995–2032. [458]

Lucas, R. E., Jr. (1980), "Two illustrations of the quantity theory of money." *American Economic Review*, 70, 1005–1014. [473]

Magnus, J. R. and H. Neudecker (2002), *Matrix Differential Calculus With Applications in Statistics and Econometrics*. New York, Wiley. [491]

McCloskey, A. and J. B. Hill (2013), "Parameter estimation robust to low-frequency contamination." Working paper, Brown University. [470]

Perron, P. and Z. Qu (2010), "Long-memory and level shifts in the volatility of stock market return indices." *Journal of Business & Economic Statistics*, 28, 275–290. [470]

Qu, Z. (2011), "A test against spurious long memory." *Journal of Business & Economic Statistics*, 29, 423–438. [470]

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. [458, 464, 465, 469, 475, 477]

Rothenberg, T. J. (1971), "Identification in parametric models." *Econometrica*, 39, 577–591. [458, 464]

Sala, L. (2013), "DSGE models in the frequency domain." Working paper, Universitá Bocconi. [460]

Sargent, T. J. and P. Surico (2011), "Two illustrations of the quantity theory of money: Breakdowns and revivals." *American Economic Review*, 101, 109–128. [459, 473]

Schorfheide, F. (2013), "Estimation and evaluation of DSGE models: Progress and challenges." In *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, 184–230, Cambridge University Press, Cambridge. [470]

Sims, C. (2002), "Solving linear rational expectations models." *Computational Economics*, 20, 1–20. [460, 472]

Smets, F. and R. Wouters (2007), "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American Economic Review*, 97, 586–606. [470]

Staiger, D. and J. H. Stock (1997), "Instrumental variables regression with weak instruments." *Econometrica*, 65, 557–586. [458, 465]

Stock, J. H. and J. H. Wright (2000), "GMM with weak identification." *Econometrica*, 68, 1055–1096. [458, 465]

Tkachenko, D. and Z. Qu (2012), "Frequency domain analysis of medium scale DSGE models with application to Smets and Wouters (2007)." In *DSGE Models in Macroeconomics: Estimation, Evaluation and New Developments*, Advances in Econometrics, Vol. 28, 319–385, Emerald, Bingley. [460, 470]

Uhlig, H. (1999), "A toolkit for analyzing nonlinear dynamic stochastic models easily." In *Computational Methods for the Study of Dynamic Economies* (R. Marimon and A. Scott, eds.), 30–61, Oxford University Press, London. [460]

Whiteman, C. H. (1984), "Lucas on the quantity theory: Hypothesis testing without theory." *American Economic Review*, 74, 742–749. [473]

Whittle, P. (1951), *Hypothesis Testing in Time Series Analysis.* Thesis, Uppsala University, Almqvist and Wiksell, Uppsala; Hafner, New York. [458]