Targeted Testing of Dynamic Stochastic General Equilibrium Models

Zhongjun Qu*
Boston University

Denis Tkachenko†
National University of Singapore

March 17, 2023

Abstract

This paper introduces targeted tests for assessing the specifications of Dynamic Stochastic General Equilibrium (DSGE) models, focusing on specific aspects such as a model's steady state properties, overall dynamic properties, and properties in selected frequency bands, such as business cycle frequencies. These proposed tests can also assist in identifying variables that are most affected by misspecification, while addressing issues of indeterminacy and weak identification. Results show that a small-scale DSGE model is rejected over the period of 1960-2007, indicating issues related to inflation dynamics and comovements between variables over business cycle frequencies, but is not rejected in subsamples when a regime change is allowed in 1979. The Smets-Wouters model is not rejected over the same period. Additionally, a medium-scale model with news shocks is rejected based on business cycle frequencies, and issues related to hours worked are reported. The proposed methods are applicable to Gaussian (factor-augmented) Vector Autoregressions.

Keywords: DSGE, misspecification, frequency domain methods, weak identification.
JEL classification: C3, C52, E1.

*Department of Economics, Boston University (qu@bu.edu).
†Department of Economics, National University of Singapore (ecstd@nus.edu.sg).

1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models are widely employed in academia and policymaking institutions such as central banks, due to their ability to provide a comprehensive framework for analyzing business cycles, understanding monetary and fiscal policies, and making forecasts. Despite significant modeling improvements over the past two decades, important misspecifications may still persist in various parts of these models. For instance, Schorfheide (2013) has pointed out that DSGE models have a tendency to underperform in capturing low-frequency fluctuations, which could lead to erroneous conclusions about the drivers of business cycles. As new models continue to be proposed, it is crucial for practitioners to have access to a robust set of statistical methods to diagnose the compatibility of DSGE models with the data and identify areas for further improvement.

This paper introduces a family of specification tests, developed in the frequency domain, for diagnosing misspecifications in DSGE models. The tests can examine a model's steady state properties, overall dynamic properties, and properties in selected frequency bands, such as business cycle frequencies. Importantly, the ability to focus on specific frequency bands means that the tests do not require the DSGE model to match the dynamic behavior of the data in every fine detail. The tests can also be applied to a subset of variables in the model, such as testing inflation and GDP dynamics without involving interest rates. This feature is helpful in indicating the variables most affected by misspecification, pointing to relevant directions for model improvement. We call the tests targeted tests because they can focus on testing various aspects of the model in addition to the full model.

The tests are based on weighted integrated periodograms. Bartlett (1955) and Grenander and Rosenblatt (1957) considered a periodogram-based approach to specification testing for univariate models. We apply this idea to a multivariate setting to examine the specification of DSGE models. We first introduce a Kolmogorov-Smirnov (KS) type tests for checking the dynamic specification of a DSGE model. The test can be motivated from a likelihood perspective, as it is related to the divergence between frequency domain Gaussian likelihood from its its expected value of no misspecification. The test is further generalized in various practical directions to examine: 1) the steady state properties; 2) selected frequency bands such as the business cycle frequencies; 3) the specification for a subset of observables; and 4) smooth weights over frequencies. We tabulate the relevant 10%, 5%, and 1% critical values. We also show that the tests are consistent against global alternatives, as long as the discrepancy is incorporated when computing the tests.

We propose a two-step procedure to implement the tests and address parameter uncertainty. In the first step, we obtain a set of plausible parameter values using an inference procedure. In the second step, we check the compatibility of these parameter values with the data using the proposed frequency domain tests, controlling the overall significance level with a Bonferroni adjustment. If all parameters are incompatible, we reject the model. Otherwise, the model is not rejected, and the surviving parameter values can be utilized to revisit the model's implications, including the robustness of its policy recommendations. Additionally, in the second step, the tests can be applied to selected frequency bands and subsets of variables to further understand the source of misspecification.

We obtain plausible parameter values for the first step in two ways. The first is by inverting a statistic that is robust to weak identification, given that weak identification of parameters is a prominent concern in these models, as highlighted by recent literature. For this purpose, we use the score test of Qu (2014), although the tests of Guerron-Quintana et al. (2013) and Andrews and Mikusheva (2015) may also be useful. This approach produces correct frequentist coverage under weak conditions and can easily incorporate parameter bounds. Nevertheless, the confidence sets are inefficient if other information, such as some parameters being strongly identified, is available. The second way is to use parameter values from the high-density region of a proper Bayesian posterior distribution. Although the resulting parameter values typically do not provide accurate frequentist coverage in this setting, considering them is useful as it allows us to examine whether and to what extent the findings regarding misspecification are sensitive to the selected parameter values. Moreover, considering them facilitates a conversion with the Bayesian DSGE literature as posterior distributions are routinely used to derive policy recommendations.

Our framework encompasses both determinacy and indeterminacy to allow wide empirical applicability. Lubik and Schorfheide (2004) have found that indeterminacy is a feature of US monetary policy practices during 1960-1979. Other related studies that examine monetary or fiscal policies include Leeper (1991), Clarida, Galí, and Gertler (2000), Benhabib, Schmitt-Grohé, and Uribe (2001), Boivin and Giannoni (2006), Benati and Surico (2009), Mavroeidis (2010), Cochrane (2011, 2014), and Leeper et.al (2017). In our empirical applications, we use the tests to compare a model's fit under different policy regimes during the same sample period, and to evaluate its fit over different subsample periods after introducing a change in policy regimes. A similar analysis from a Bayesian perspective is conducted in Lubik and Schorfheide (2004).

Our methods contribute to existing approaches for diagnosing DSGE models. In this literature, a common approach is to compare models based on marginal likelihoods or forecasting accuracy. A model is preferred over another if its marginal likelihood value or forecasting accuracy is significantly higher. This approach is informative for ranking models, but it does not reveal whether the preferred model is compatible with the data. An alternative approach is to use a structural VAR (SVAR) as a benchmark and compare the impulse responses of the DSGE model with those of the SVAR; see, e.g., Christiano et al. (2005). However, this approach is not always feasible, as the identification conditions for impulse responses may not be compatible between the DSGE and SVAR models, as discussed in Del Negro et al. (2007). Recent studies have investigated alternative methods to overcome these limitations. Del Negro et al. (2007)

developed a hybrid DSGE-VAR model as an encompassing model to evaluate DSGE models. The posterior distributions of the model are informative about the areas of the model that are subject to misspecification, and Del Negro and Schorfheide (2009) subsequently applied this framework to assess the DSGE model's policy predictions. More recently, Inoue et al. (2020) introduced stochastic processes as specification errors into a DSGE model and evaluated the improvement in model fit through forecasting error decompositions and marginal likelihood comparisons. Similar to our goal, they aim to understand which aspects of the model are most affected by misspecification. Our methods differ from the above in several aspects: we adopt a frequentist approach, our methods do not involve a parametric reference model, we emphasize the examination of subsets of variables and across frequency bands, and we address weak identification. Evaluating the specifications of DSGE models is challenging, and we hope that our methods provide a useful alternative for this purpose.

Our analysis builds on the literature that evaluates DSGE models or rational expectations models in the frequency domain, where important studies include Watson (1993), King and Watson (1996), and Diebold et al. (1998). Watson (1993) recommended using model and data spectra plots as diagnostic tools. Diebold et al. (1998) emphasized the importance of accounting for estimation uncertainty and examining model fit across frequencies and explored the advantages of taking a graphical approach. Our new statistic extends this line of inquiry to the current generation of DSGE models using a new test along with alternative approaches to address parameter uncertainty.

We evaluate the size and power properties of the tests using the small-scale model from Lubik and Schorfheide (2004) as the data generating process. We consider both determinacy and indeterminacy. The results indicate that the tests have excellent size properties in typical sample sizes after applying a simple model-dependent prewhitening filter to flatten the spectrum near the zero frequency. We also examine the tests' power in various scenarios, including testing the full model and a subset of variables, as well as testing a subset of frequencies.

We then move on to the empirical applications where we examine three DSGE models: the small-scale model by Lubik and Schorfheide (2004), and two medium-scale models - the Smets and Wouters (2007) model and the news shocks model of Schmitt-Grohé and Uribe (2012). The Lubik and Schorfheide (2004) model contrasts determinacy and indeterminacy within a small-scale framework. The Smets and Wouters (2007) model is a benchmark medium-scale New Keynesian model. This model extends the standard New Keynesian model by incorporating additional frictions and real rigidities and allows us to examine how model specification improves compared to the baseline small-scale model. Additionally, the Schmitt-Grohé and Uribe (2012) model provides an opportunity to evaluate whether the proposed information structure generates dynamics that fit the data adequately, as well as how the proposed structure compares to the standard structure assumed in the small- and medium-scale models.

Our results indicate that the small-scale DSGE model is rejected at the 10% significance level for both determinacy and indeterminacy specifications over the period of 1960-2007, based on both the full spectrum analysis and business cycle frequency analysis. Further analysis reveals that misspecification is evident in most segments of the model, particularly in the inflation dynamics and comovements between variables. The imaginary parts of the cross-spectrum, which indicate the leads and lags relationships between variables, deviate the most from the data. These conclusions are reinforced by using the MCMC draws from posterior distributions. Splitting the full sample using 1979:II, we find that the model is no longer rejected if indeterminacy is used for the first subsample and determinacy for the second, which supports Lubik and Schorfheide's (2004) conclusion that U.S. monetary policy post-1982 is consistent with determinacy, while the pre-Volcker period exhibits greater uncertainty. We also analyze the change in data dynamics, particularly the cross-spectrum, that brings the data closer to the model.

For the Smets and Wouters (2007) model, we find that it is not rejected at either full spectrum or business cycle frequencies at the 10% significance level using the full sample 1960:I-2007:IV. However, further examination reveals that, similar to the small scale case, one of the main sources of tension between the model and the data are inflation and its comovements with other variables, especially with the interest rate. The results obtained using posterior distribution draws are qualitatively similar. Moving on to the news shocks model, our analysis shows that it is rejected, and the results suggest issues related to hours worked. We are currently conducting further analysis based on the posterior distribution for this model.

The paper is structured as follows. In Section 2, we explain how to compute a DSGE model's spectrum allowing for both determinacy and indetermancy, highlighting the issue of weak parameter identification. Section 3 introduces the specification tests, highlights the likelihood perspective, and characterizes their asymptotic properties under the null and alternative hypotheses. This section also describes a model-dependent prewhitening filter that improves the tests' finite sample properties under the null hypothesis. In this section, the parameter values are assumed known, and the issue of parameter uncertainty is addressed in Section 4. Section 5 provides calibrated simulations for finite sample properties. Section 6 presents three empirical applications, and Section 7 provides concluding remarks. The online appendix includes proofs of the results, details on the empirical applications, and additional tables and figures that complement the main analysis.

2 The spectrum of a DSGE model

In this section, we describe the spectrum of a log linearized DSGE model to provide a basis for our analysis. Consider a DSGE model log linearized around its steady state (Sims, 2002):

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

where S_t is a vector that includes endogenous variables, conditional expectations, and variables from exogenous shock processes if they are serially correlated. The vector ε_t contains serially uncorrelated structural shocks and η_t contains expectation errors. The elements of Γ_0, Γ_1, Ψ and Π are functions of structural parameters of the model. Depending on Γ_0 and Γ_1, the system can have none, a unique, or multiple stable solutions (indeterminacy). Under indeterminacy, the structural parameters alone do not uniquely determine the dynamics of the model. The above formulation is sufficiently flexible and it allows for medium scale model such as Smets and Wouters (2007) and Schmitt-Grohé and Uribe (2012).

Lubik and Schorfheide (2003) show that the full set of solutions to the model is representable as

$$S_t = \Phi_1 S_{t-1} + \Phi_\varepsilon \varepsilon_t + \Phi_\epsilon \epsilon_t, \tag{2}$$

or equivalently,

$$S_t = (1 - \Phi_1 L)^{-1} [\Phi_\varepsilon \ \Phi_\epsilon] \begin{bmatrix} \varepsilon_t \\ \epsilon_t \end{bmatrix},$$

where L is the lag operator. In (1), Φ_1, Φ_ε and Φ_ϵ depend only on Γ_0, Γ_1, Ψ and Π, therefore, are functions of the structural parameters only. The term ϵ_t contains the sunspot shocks. The DSGE model alone imposes few restrictions on ϵ_t, i.e., it needs to be a martingale difference, so that $E_t \epsilon_{t+1} = 0$, but it can be arbitrarily contemporaneously correlated with the fundamental shocks ε_t. Intuitively, the properties of ϵ_t depend on how agents form their expectations, which is not fully revealed by the model under indeterminacy. To reflect this, Qu and Tkachenko (2017) adopted the following parameterization that expresses ϵ_t as an orthogonal projection onto ε_t and a residual term:

$$\epsilon_t = M \varepsilon_t + \widetilde{\epsilon}_t,$$

where M is a matrix of constants and $\widetilde{\epsilon}_t$ is now uncorrelated with ε_t, with $Var(\widetilde{\epsilon}_t) = \Sigma_\epsilon$. Let θ^D be a p-by-1 vector consisting of all the structural parameters in (1). Let θ^U be a q-by-1 vector consisting of the sunspot parameters $\theta^U = (\text{vec}(\Sigma_\epsilon)', \text{vec}(M)')'$. We define an augmented parameter vector as follows:

$$\theta = \begin{bmatrix} \theta^D \\ \theta^U \end{bmatrix}.$$

This augmented parameter vector uniquely determines the dynamics of the model.

In practice, the estimation is typically based on a subset of S_t or some linear transformations involving its current and lagged values. To be consistent with this, we use a matrix $A(L)$ of finite order lag polynomials to specify the observables and define

$$Y_t(\theta) = A(L) S_t = H(L; \theta)(\varepsilon_t' \ \epsilon_t')',$$

where

$$H(L;\theta) = A(L)(1-\Phi_1 L)^{-1}[\Phi_\varepsilon\ \Phi_\epsilon].$$

Then, the spectral density of $Y_t(\theta)$ is given by

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

where * denotes the conjugate transpose and

$$\Sigma(\theta) = \begin{pmatrix} I & 0 \\ M & I \end{pmatrix} \begin{pmatrix} \Sigma_\varepsilon & 0 \\ 0 & \Sigma_\epsilon \end{pmatrix} \begin{pmatrix} I & 0 \\ M & I \end{pmatrix}'.$$

We let $\{Y_t\}$ denote a stochastic process whose spectral density is given by $f_{\theta_0}(\omega)$ for $\omega \in [-\pi,\pi]$ under the null hypothesis of correct model specification. The process $\{Y_t\}$ is usually assumed to be stationary after model-dependent detrending operations. But its population mean, which relates to the model's steady state, is typically nonzero. To capture this, we let $\mu(\theta_0)$ denote the mean of $\{Y_t\}$ implied by the model and write

$$Y_t = \mu(\theta_0) + Y_t(\theta_0).$$

This formulation provides a system of equations that relate the data (i.e., the left hand side) to the model (i.e., the right hand side). The model is correctly specified if these equations hold at some θ_0 for all t. Additionally, the model's steady-state properties are correctly specified if the mean of Y_t is equal to $\mu(\theta_0)$. Its overall dynamic properties are correctly specified if the spectral density of Y_t matches that of $Y_t(\theta_0)$. Moreover, the model is correctly specified over a frequency band if the two spectral densities agree with each other over this band. These properties allow us to develop tests for various aspects of the DSGE model in a unified framework.

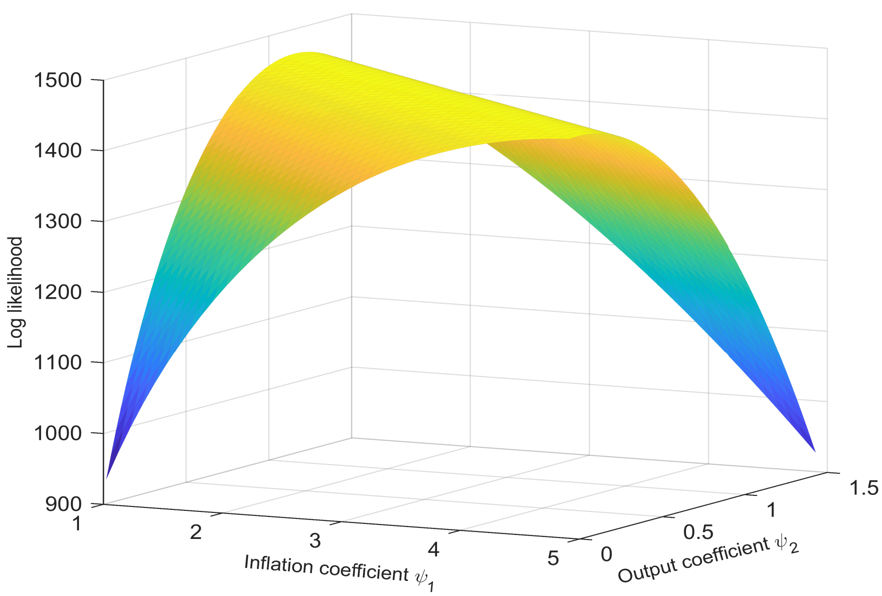
Before proceeding further, we present two examples to illustrate the interplay between parameter identification, equilibrium indeterminacy, and model specification. These examples help to motivate the need to allow for indeterminacy and weak identification when conducting model specification analysis.

Example 1 *Consider $\mathbb{E}_t x_{t+1} = \alpha x_t$, where x_t is a scalar random variable (e.g., inflation), and α is a structural parameter. Assume $\lim_{k\to\infty}\mathbb{E}_t x_{t+k}$ is finite. Depending on the value of α, two regimes arise. If $|\alpha|>1$, we can solve the model forward to obtain $x_t = \lim_{k\to\infty}(1/\alpha)^k \mathbb{E}_t x_{t+k} = 0$. In this regime, the model has a unique equilibrium, but α is not identified, i.e., it is impossible to determine its value uniquely, even with infinite sample size. If $|\alpha| \le 1$, solving the model forward is no longer informative; however, we can define a sunspot shock $\epsilon_{t+1} = x_{t+1} - \mathbb{E}_t x_{t+1}$ and solve the model backward to obtain*

$$x_{t+1} = \alpha x_t + \epsilon_{t+1},\ \textit{with}\ \mathbb{E}_t\epsilon_{t+1} = 0.$$

In this regime, the model exhibits indeterminacy (multiple equilibria) because x_t displays stochastic fluctuations in the absence of any shocks to fundamentals. Furthermore, if ϵ_t is stationary with a positive variance, then α is identified because x_t follows an AR(1) process. This simple example highlights three messages that are applicable to general DSGE models: 1) parameter identification failure can occur within DSGE models; 2) identification properties can differ across regimes; and 3) a model's dynamic properties can differ significantly across regimes. Therefore, alternative regimes may yield different fits to empirical data. Recent DSGE literature has documented these properties in various contexts; see, for example, Canova and Sala (2009), Qu and Tkachenko (2017), and Lubik and Schorfheide (2004) on these three issues, respectively.*

Figure 1: Log likelihood surface with respect to Taylor rule coefficients.



Example 2 *An and Schorfheide (2007) used the following model to demonstrate the Bayesian analysis of DSGE models, where y_t, π_t, and r_t are log deviations of output, inflation, and interest rate from their steady states:*

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

where $g_t = \rho_g g_{t-1} + \varepsilon_{gt}, z_t = \rho_z z_{t-1} + \varepsilon_{zt}, \varepsilon_{rt} \overset{iid}{\sim} N(0, \sigma_r^2), \varepsilon_{gt} \overset{iid}{\sim} N(0, \sigma_g^2)$, and $\varepsilon_{zt} \overset{iid}{\sim} N(0, \sigma_z^2)$, and the three shocks are mutually independent. This model exhibits both determinate and indeterminate regimes as in the previous example. For the determinate regime, it has been shown that the *four parameters in the Taylor equation (ρ_r, ψ_1, ψ_2, and σ_r) are not separately identified, meaning that their values can be moved along a path without altering the model's dynamic properties (Qu and Tkachenko, 2012). To illustrate the effect of this property on inference, we plot the log likelihood surface for a simulated sample of size 1000 with respect to ψ_1 and ψ_2 while fixing all other parameters, including ρ_r and σ_r, at their posterior mean values reported in An and Schorfheide (2007). In a regular inference problem, we would expect the log likelihood to resemble a dome uniquely peaked at the maximum likelihood estimate (MLE). Instead, the surface here displays a ridge, which is virtually flat in a single direction. If all four parameters (ρ_r, ψ_1, ψ_2, and σ_r) are allowed to vary, then the likelihood surface will be entirely flat in one direction, which means that the standard MLE is inconsistent, and the information matrix is singular. Consequently, a specification test will likely have a nonstandard distribution if it relies on consistent estimation of model's parameters.*

These two examples highlight the importance of addressing weak identification and indeterminacy when conducting specification analysis for DSGE models.

3 Frequency-domain specification tests

In this section, we present the frequency domain tests and examine their asymptotic properties under null and alternative hypotheses. The value of θ_0 is assumed known. We will address the issue of parameter uncertainty in the next section, where we will perform the same tests using a range of plausible parameter values.

3.1 Proposed tests

We propose a family of tests based on integrated periodograms. The use of integrated periodograms for misspecification testing in the univariate case dates back to Grenander and Rosenblatt (1957) and Bartlett (1955). We extend this idea to the multivariate setting to test the specification of a DSGE model.

Suppose we have a sample of T observations: $\{Y_1, Y_2, ..., Y_T\}$, for which the Fourier frequencies are given by $\omega_j = 2\pi j/T$ ($j = 0, 1, ..., T-1$). The Fourier transform and periodogram are defined, respectively, as follows:

$$w_T(\omega_j) = (2\pi T)^{-1/2} \sum_{t=1}^{T} (Y_t - \mu(\theta_0)) \exp(-i\omega_j t),$$

and

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

where the asterisk denotes the conjugate transpose. The Fourier transform projects the process $Y_t - \mu(\theta_0)$ onto the frequency domain and preserves all the information contained in its mean and covariances. Since the Fourier transform of a constant equals zero at any nonzero frequency, we can express $w_T(\omega_j)$ as $(2\pi T)^{-1/2}\sum_{t=1}^{T} Y_t \exp(-i\omega_j t)$ for $j = 1, ..., T-1$.

We first consider testing whether the model's dynamic properties are correctly specified, which corresponds to the following null and alternative hypotheses

$$\begin{aligned} H_0 &: \text{ The spectral density of } Y_t \text{ equals } f_{\theta_0}(\omega) \text{ for any } \omega \in [-\pi, \pi], \\ H_1 &: \text{ The spectral density of } Y_t \text{ differs from } f_{\theta_0}(\omega) \text{ for some } \omega \in [-\pi, \pi]. \end{aligned}$$

Recall that if the model's dynamic properties are correctly specified, then $I_T(\omega_j) - f_{\theta_0}(\omega_j)$ are approximately uncorrelated with a zero mean. Based on this property, we propose the following Kolmogorov-Smirnov (KS) test for checking the model's dynamic specification:

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

The main component of the test is $I_T(\omega_j) - f_{\theta_0}(\omega_j)$, where $f_{\theta_0}(\omega_j)$ is the spectral density function implied by the model, and the division by $f_{\theta_0}(\omega_j)$ acts as a normalization to ensure that the test can be asymptotically pivotal. The norm $\|.\|_\infty$ represents the supremum norm, which is used to search for the strongest evidence against the null hypothesis, i.e., for a generic vector $z = (z_1, ..., z_k) \in \mathbb{C}^k$, $\|z\|_\infty = \max(|z_1|, ..., |z_k|)$. The zero frequency is excluded, therefore the test is invariant to the model's steady state properties. The test is straightforward to compute because $f_{\theta_0}(\omega_j)$ follows from solving the model, while $I_T(\omega_j)$ is the squared discrete Fourier transform of the observed data. No simulation is needed to implement the test.

An immediate generalization of $\mathcal{H}_{dT}(\theta_0)$ is to assign weights to different frequencies. This feature is useful because DSGE models are not intended to capture high frequency fluctuations in the data. Let $W(\omega_j)$ be a smooth scalar-valued function or an indicator function to select the target frequencies. We propose

$$\mathcal{H}_{dT}^{W}(\theta_0) = \sup_{r\in[0,1]} \left\| (T/2)^{-1/2} \sum_{j=1}^{[Tr/2]} W(\omega_j) \operatorname{vec}\left\{ f_{\theta_0}(\omega_j)^{-1/2} \left(I_T(\omega_j) - f_{\theta_0}(\omega_j) \right) f_{\theta_0}(\omega_j)^{-1/2} \right\} \right\|_\infty . \tag{4}$$

In subsequent analysis, we consider two specifications for $W(\omega_j)$. In the first case, $W(\omega)$ is equal to one for business cycle frequencies ($\pi/16 \le \omega \le \pi/3$ for quarterly data), and zero otherwise. In the second case, $W(\omega)$ is a linearly decreasing function that assigns lower weights to high frequencies ($W(\omega) = 1 - \omega/\pi$). Note that the variance of the partial sum in (4) grows linearly with r, and the latter $W(\omega)$ counterbalances this tendency by putting more emphasis on business cycle and lower frequencies. We provide critical values for these two cases and provide computer codes for other specifications. Furthermore, it is possible to specify $W(\omega)$ as a moving or expanding window to investigate rejections occurring as frequency components change. However, we do not delve into this aspect further due to space limitations.

Now we turn to the model's steady state properties. In this case, the null and alternative hypotheses are

$$\begin{aligned} H_0 &: \text{ The mean of } Y_t \text{ equals } \mu(\theta_0), \\ H_1 &: \text{ The mean of } Y_t \text{ differs from } \mu(\theta_0). \end{aligned}$$

Under the null hypothesis, the Fourier transform at the zero frequency, $w_T(0) = (2\pi T)^{-1/2} \sum_{t=1}^{T} (Y_t - \mu(\theta_0))$, is asymptotically normally distributed with a zero mean. The values at nonzero Fourier frequencies are irrelevant for testing $\mu(\theta_0)$ as they are invariant to a location shift. Hence, we obtain a test for the model's steady state properties as follows:

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

This test has been used previously to test for structural changes in the mean of an otherwise stationary time series. It still has power if $E(Y_t) = \mu(\theta_0)$ is violated only for a part of the sample. As in the previous test, a supremum operator is used to detect the strongest evidence against the null hypothesis.

Finally, we can combine the aforementioned statistics to obtain a joint test for the model's static and dynamic specifications:

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

Alternatively, using weights, the joint test can be expressed as:

$$\mathcal{H}_T^W(\theta_0) = \max \left(\mathcal{H}_{sT}(\theta_0), \mathcal{H}_{dT}^W(\theta_0) \right).$$

The above tests consider the full model. In practice, it is often important to analyze a subset of variables. For example, King and Watson (1996) compared three rational expectations models to capture the relationship between a real (GDP) and a nominal (interest rate) variable. This type of analysis is particularly useful when the full model is rejected, as it can help identify the source of the rejection. The generalization in our frequency-domain framework is simple. Let A be a variable selection matrix. For instance, if the model has three variables and the goal is to examine only the first two, set $A = [1, 0, 0; 0, 1, 0]$. To investigate the difference between the first and second variables, set $A = [1, -1, 0]$. To study the relationship between their growth rates, use $A = [1 - L, 0, 0; 0, 1 - L, 0]$, where L represents the lag operator. After defining A, construct the tests by replacing Y_t and $f_{\theta_0}(\omega)$ with AY_t and $Af_{\theta_0}(\omega)A'$, respectively. The remaining analysis remains unchanged.

In the empirical applications, it will be demonstrated how these tests can be employed to specifically evaluate a model's steady-state properties (using the $\mathcal{H}_{sT}$ test), overall dynamic properties (using $\mathcal{H}_{dT}^W$ with $W(\omega) = 1 - \omega/\pi)$), and properties within a selected frequency band (e.g., business cycle frequencies, using $\mathcal{H}_{dT}^W$ with $W(\omega)$ as an indicator function). Furthermore, the analysis will assess which variables are most influenced by misspecification using suitable A matrices.

3.1.1 A likelihood perspective

The tests can be derived from a likelihood perspective. Recall that the Whittle likelihood is a frequency domain approximation to the time domain Gaussian log likelihood function. For a DSGE model, the Whittle likelihood (using dynamic properties only and omitting an additive constant) has the expression

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

Suppose the true spectral density underlying the data is $f_0(\cdot)$, which may differ from the DSGE model-implied spectral density $f_\theta(\cdot)$. Then, the Whittle log likelihood computed using the same data but under the true spectral density is

$$L_{0,T} = -\frac{1}{2T}\sum_{j=1}^{T-1}\left[\log\det\left(f_0(\omega_j)\right) + \operatorname{tr}\left(f_0^{-1}(\omega_j)I_T(\omega_j)\right)\right],$$

whose expected value is

$$E(L_{0,T}) = -\frac{1}{2T}\sum_{j=1}^{T-1}\left[\log\det(f_0(\omega_j)) + n_Y\right] + o\left(T^{-1/2}\right),$$

where n_Y is the number of the observable and the $o\left(T^{-1/2}\right)$ term arises because $E(I_T(\omega_j)) - f_0(\omega_j) = o(T^{-1/2})$. These expressions imply that the divergence $T^{1/2}(L_T(\theta) - E(L_{0,T}))$ is equal to

$$\frac{1}{2T^{1/2}}\sum_{j=1}^{T-1}\operatorname{vec}\left\{f_\theta(\omega_j)^{-1/2}\left(f_\theta(\omega_j) - I_T(\omega_j)\right)f_\theta(\omega_j)^{-1/2}\right\}$$
$$+\frac{1}{2T^{1/2}}\sum_{j=1}^{T-1}[\log\det(f_0(\omega_j)) - \log\det(f_\theta(\omega_j))] + o(1),$$

where the second term is independent of the data, and the first term yields the $\mathcal{H}_{dT}(\theta_0)$ test upon taking the supremum. In summary, the $\mathcal{H}_{dT}(\theta_0)$ statistic is based on the distance between the log likelihood and its expected value under the true DGP, and the sup operator acts as a searching mechanism for the strongest evidence of model misspecification. This interpretation also offers an explanation for why the test is consistent against global alternatives.

3.1.2 Applications beyond DSGE models

The above tests are applicable to other dynamic linear models such as Gaussian Structural Vector Autoregressions and Factor Augmented VARs. We highlight these connections below.

Structural Vector Autoregressions. A typical SVAR model (Sims, 1980) is

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

where Y_t is an n-dimensional column vector, $\Phi_0, \Phi_1, ..., \Phi_p$ are coefficient matrices, and ε_t is an n-by-1 vector of serially uncorrelated structural disturbances with $Var(\varepsilon_t) = \Sigma$. We define $\theta = (\Phi_0, \Phi_1, ..., \Phi_p, \Sigma)$ and $\Pi(L;\theta) = \Phi_0 - \sum_{j=1}^{p} \Phi_j L^j$. The spectral density of Y_t is given by

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

For any given parameter value θ, the proposed tests can be constructed using Y_t and $f_\theta(\omega)$, or AY_t and $Af_\theta(\omega)A'$ for a subset of variables, where A is a selection matrix as defined previously. These tests can be used to evaluate the model's specification over a chosen band of frequencies and test low frequency hypotheses for both just-identified and over-identified models.

Factor Augmented VAR. A typical model (see, for example, Stock and Watson, 2005) is

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

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

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

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

where $H_1(L;\theta) = [I - D(L)L]^{-1} \lambda(L) [I - \Gamma(L)L]^{-1}$ and $H_2(L;\theta) = [I - D(L)L]^{-1}$. The spectral density of Y_t is thus given by

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

Stock and Watson (2005) discussed several identification strategies for this model, including (1) contemporaneous timing restrictions on the zero-order term in $H_1(L;\theta)$, (2) long-run restrictions on $H_1(1;\theta)$, (3) factor loading restrictions on $\lambda(L)$ and (4) Uhlig's (2005) sign restrictions on the coefficients of $H_1(L;\theta)$. Since these restrictions only affect $f_\theta(\omega)$, they are implementable in the frequency domain, and model diagnostics can be carried out under such restrictions. Importantly, our framework works directly with the structural parameter vector, so we can avoid making assumptions about the reduced form parameters. However, it is important to note that our framework assumes that the dimension of Y_t, n, is finite, which excludes the direct analysis of high-dimensional factor models. One potential generalization in that direction is to estimate the factors first and then treat them as part of the data, which can be implemented in a two-step procedure in the time domain (Stock and Watson, 2005). It would be interesting to explore these aspects in the frequency domain as well.

3.2 Asymptotic properties under null and alternative hypotheses

We require the following assumptions in order to study the tests' asymptotic properties under the null hypothesis of correct model specification.

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

Assumption 2. The model solution can be written as

$$Y_t(\theta) = H(L;\theta)u_t(\theta) \;\text{ with }\; H(L;\theta) = \textstyle\sum_{j=0}^{\infty} h_j(\theta)L^j,$$

where $u_t(\theta)$ is serially uncorrelated with a nonsingular covariance matrix denoted by $\Sigma(\theta)$.

Assumption 3. For all $\omega \in [-\pi, \pi]$ and $\theta \in \Theta$, there exist constants C_L and C_U such that: (i) the eigenvalues of $f_\theta(\omega)$ satisfy $C_L \leq \text{eig}(f_\theta(\omega)) \leq C_U$; (ii) $f_\theta(\omega)$ belongs to the Lipschitz class of order β (Lip(β)) with $\beta > 1/2$; (iii) $||\partial \operatorname{vec} f_\theta(\omega)/\partial\theta'|| \leq C_U$; (iv) $\partial \operatorname{vec} f_\theta(\omega)/\partial\theta$ belongs to Lip(β) with $\beta > 1/2$. (v) $||\partial\mu(\theta)/\partial\theta'|| \leq C_U$.

Assumption 4. $\{Y_t\}_{t=1}^{T}$ are multivariate normal random vectors.

Assumption 1 is standard. Assumption 2 states that the DSGE model is correctly specified. Assumptions 3(i)-(ii) are satisfied by stationary finite order vector autoregressive moving average processes with finite error covariance matrices, which are typically the forms of solutions in linearized DSGE models. Assumptions 3(iii)-(iv) hold when parameters enter smoothly into $f_\theta(\omega)$ and $\mu(\theta)$. Qu and Tkachenko (2017) used assumptions similar to 1-3 in their study of the identification properties of log linearized DSGE models. Assumption 4 assumes normality, a common specification in DSGE models. Without this assumption, the test distributions will depend on the third and fourth cumulants of the shocks $u_t(\theta)$.

To present the limiting distributions under the null hypothesis, let

$$\widetilde{B}(r) = (B_1(r) + iB_2(r))/\sqrt{2},$$

where i is the imaginary unit and $B_1(r)$ and $B_2(r)$ are two independent Wiener processes.

Theorem 1 *Suppose $\{Y_t\}_{t=1}^{T}$ is a covariance stationary process with spectral density $f_{\theta_0}(\omega)$ over $\omega \in [-\pi, \pi]$ (or over the frequencies selected by $W(\cdot)$ if $\mathcal{H}_{dT}^{W}$ is computed). Under Assumptions 1-4, and assuming $E(Y_t) = \mu(\theta_0)$ only when testing the steady-state, we have*

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

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

3. *$\mathcal{H}_{sT}(\theta_0) \Rightarrow \sup_{r\in[0,1]} \|G_s(r)\|_\infty$, where $G_s(r)$ is an n_Y vector of independent Wiener processes.*

4. $\mathcal{H}_T(\theta_0) \Rightarrow \max\left(\sup_{r\in[0,1]} \|G_d(r)\|_\infty, \sup_{r\in[0,1]} \|G_s(r)\|_\infty\right)$*, where the elements of* $[G_d(r), G_s(r)]$ *are mutually independent.*

The first two results are not dependent on the correct specification of the steady-state. The elements of $G_d(r)$ can be divided into two parts: the first n_Y elements represent the diagonal elements in $I_T(.)$, and the remaining $n_Y(n_Y-1)/2$ elements represent the off-diagonal elements. The limiting distributions of $\mathcal{H}_{dT}(\theta_0)$, $\mathcal{H}_{sT}(\theta_0)$ and $\mathcal{H}_T(\theta_0)$ are pivotal, meaning that they depend only on the number of variables being tested (n_Y) and can be easily simulated. Table 1 presents the 10%, 5%, and 1% critical values for the number of variables n_Y ranging from 1 to 10. When a subset of model variables are tested, n_Y refers to the dimension of AY_t. Critical values for $\mathcal{H}_{dT}^W$ are specific to $W(\omega)$, and they need to be tabulated on a case-by-case basis. In the table, we provide the values for two cases: a smooth function that underweights high frequencies and an indicator function that only selects business cycle frequencies. We provide computer code for simulating other cases since only the modification of $W(\omega)$ is needed.

Table 1: Critical Values of the Specification Tests

Test	Size	Number of Observables Tested (n_Y)									
		1	2	3	4	5	6	7	8	9	10
Full Spectrum	10	1.946	2.261	2.423	2.529	2.615	2.677	2.735	2.779	2.816	2.856
$\mathcal{H}_{dT}(\theta_0)$	5	2.231	2.510	2.649	2.748	2.832	2.885	2.930	2.976	3.010	3.045
	1	2.804	3.015	3.143	3.219	3.287	3.339	3.373	3.403	3.440	3.475
Full Spectrum	10	0.867	0.981	1.040	1.081	1.110	1.136	1.155	1.172	1.187	1.201
Weighted by	5	0.977	1.077	1.128	1.167	1.196	1.216	1.235	1.251	1.264	1.277
$1-\omega/\pi$	1	1.190	1.271	1.311	1.347	1.372	1.387	1.403	1.418	1.428	1.447
Business Cycle	10	1.004	1.166	1.253	1.310	1.351	1.384	1.415	1.436	1.459	1.478
Frequencies	5	1.151	1.295	1.370	1.424	1.460	1.491	1.519	1.540	1.558	1.579
	1	1.446	1.559	1.619	1.668	1.705	1.724	1.756	1.766	1.787	1.804
Steady State	10	1.944	2.219	2.357	2.466	2.540	2.605	2.659	2.703	2.744	2.771
$\mathcal{H}_{sT}(\theta_0)$	5	2.224	2.480	2.614	2.709	2.778	2.845	2.894	2.937	2.968	2.989
	1	2.794	3.017	3.117	3.199	3.275	3.310	3.370	3.403	3.433	3.456
Steady State and	10	2.218	2.485	2.624	2.722	2.799	2.858	2.907	2.953	2.983	3.012
Full Spectrum	5	2.478	2.729	2.851	2.944	3.014	3.060	3.114	3.153	3.186	3.214
$\mathcal{H}_T(\theta_0)$	1	3.012	3.217	3.326	3.395	3.460	3.502	3.552	3.591	3.609	3.640

Note. The critical values are obtained through simulation using a sample size of 1000 and 100,000 replications. n_Y denotes the number of variables being tested.

The tests have power against local alternatives of order $T^{-1/2}$. The next result shows that they are consistent against covariance stationary global alternatives.

Theorem 2 *Suppose* $\{Y_t\}$ *is a covariance stationary process with* $\mathbb{E}Y_t = \mu_0$ *and spectral density* $f_0(\omega)$ *that satisfy Assumptions 1-4. Let* $\mu(\theta_0)$ *and* $f_{\theta_0}(\omega)$ *be the mean and spectral density of*

$\{Y_t\}$ *implied by the DSGE model, satisfying Assumptions 1-4. Let* $\delta > 0$ *be an arbitrary constant independent of T. Then:*

1. $\mathcal{H}_{dT}(\theta_0) \to \infty$ *if* $\|f_0(\omega) - f_{\theta_0}(\omega)\| > \delta$ *for some* $\omega \in [0, \pi]$.

2. $\mathcal{H}_{dT}^{W}(\theta_0) \to \infty$ *if* $\|f_0(\omega) - f_{\theta_0}(\omega)\| > \delta$ *for some* ω *with* $W(\omega) = 1$.

3. $\mathcal{H}_{sT}(\theta_0) \to \infty$ *if* $\|\mu_0 - \mu(\theta_0)\| > \delta$.

4. $\mathcal{H}_{T}(\theta_0) \to \infty$ *if* $\|f_0(\omega) - f_{\theta_0}(\omega)\| > \delta$ *for some* $\omega \in [0, \pi]$ *or* $\|\mu_0 - \mu(\theta_0)\| > \delta$.

The power of $\mathcal{H}_{dT}$ increases to 1 when the model and data spectra differ significantly over a set of frequencies. Similarly, the power of $\mathcal{H}_{dT}^{W}$ approaches 1 when there is a significant difference in spectra within the frequencies selected by $W(\omega)$. The $\mathcal{H}_{sT}$ test is consistent for misspecification in the mean, while the $\mathcal{H}_{T}$ test combines information from both the mean and spectrum and is consistent in detecting misspecification when either of them differs significantly.

3.2.1 Improving finite sample properties with prewhitening

Periodograms are known to exhibit downward biases near the zero frequency for persistent time series, which can affect our tests. The fully parametric nature of a DSGE model provides a simple solution to address this issue. By using the model to simulate a long time series at θ_0 (denoted as $Y_t(\theta_0)$, where $t = 1, ..., \bar{T}$), we can construct a filter to prewhiten the empirical data, flattening the spectral density near the zero frequency and significantly reducing the bias even in small samples. It is crucial to use the DSGE model, rather than the actual empirical data, to compute the filter, which ensures that the filtering procedure does not influence the test distributions.

The steps of the filtering operation are as follows: (i) Simulate a long time series using the DSGE model at θ_0, denoted by $Y_t(\theta_0)$ for $t = 1, ..., \bar{T}$. (ii) Estimate a VAR(1) model using the simulated data: $Y_t(\theta_0) = BY_t(\theta_0) + e_t$. (iii) Use the estimated value of B, denoted as $\hat{B}$, and the actual empirical data to compute the filtered data $(I - \hat{B}L)Y_t$ for $t = 1, ..., T$. (iv) Compute the tests using the filtered data by replacing Y_t and $f_{\theta_0}(\omega)$ with $(I - \hat{B}L)Y_t$ and $(I - \hat{B}\exp(-i\omega))f_{\theta_0}(\omega)(I - \hat{B}\exp(-i\omega))^*$, respectively. To test a subset of variables, we estimate a VAR(1) for $AY_t(\theta_0)$ and apply the filtering to AY_t. Other aspects of the procedure remain the same. The asymptotic null distributions of the tests remain unchanged if $\bar{T}$ is sufficiently large (i.e., $\bar{T}/T \to \infty$) because the estimation uncertainty of the filter is negligible. The tests remain consistent against global alternatives. We use this prewhitening procedure when presenting our results for both simulations and empirical applications.

4 Accounting for parameter uncertainty

In the previous section, we considered a prespecified θ_0. In this section, we develop a testing procedure that accounts for parameter uncertainty and addresses the issue of weak identification. The main idea is to implement the test in two steps with a Bonferroni adjustment. First, we obtain an identification-robust confidence set for θ at the α_S level. This is achieved by inverting (i.e., sampling) the identification-robust score test proposed by Qu (2014), although other statistics such as those proposed by Guerron-Quintana et al. (2013) and Andrews and Mikusheva (2015) can also be used. Next, we compute the specification test at the $\alpha_{\mathcal{H}} = (\alpha - \alpha_S)$ level using all values in this confidence set. The overall significance level of the resulting test is ensured to be no more than α percent. If the model is not rejected, the parameter values that pass the specification checks are obtained. These values are valuable as they can be used to revisit the model's implications and provide more robust policy recommendations. We begin by describing Qu's (2014) test.

4.1 The score test of Qu (2014)

The test is based on the Whittle likelihood. For any θ_0, the score function of the Whittle likelihood is equal to

$$\begin{aligned} D_T(\theta_0) &= \frac{1}{2\sqrt{T}} \sum_{j=0}^{T-1} \left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'} \right)^* \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j) \right) \operatorname{vec}\left(I_T(\omega_j) - f_{\theta_0}(\omega_j) \right) \\ &\quad + \frac{1}{2\pi\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \mu(\theta_0)'}{\partial \theta} f_{\theta_0}^{-1}(0)(Y_t - \mu(\theta_0)), \end{aligned} \tag{5}$$

A consistent approximation to the information matrix under normality is given by

$$\begin{aligned} M_T(\theta_0) &= \frac{1}{2T} \sum_{j=0}^{T-1} \left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'} \right)^* \left(f_{\theta_0}^{-1}(\omega_j)' \otimes f_{\theta_0}^{-1}(\omega_j) \right) \frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'} \\ &\quad + \frac{1}{2\pi} \frac{\partial \mu(\theta_0)'}{\partial \theta} f_{\theta_0}^{-1}(0) \frac{\partial \mu(\theta_0)}{\partial \theta'}. \end{aligned} \tag{6}$$

Using this notation, the score test for testing the null hypothesis of $\theta = \theta_0$ can be expressed as

$$S_T(\theta_0) = D_T(\theta_0)' M_T^+(\theta_0) D_T(\theta_0) \tag{7}$$

where $M_T^+(\theta)$ denotes the Moore-Penrose generalized inverse of $M_T(\theta)$. The generalized inverse is needed because $M_T(\theta)$ is not full rank if some parameters are unidentified. Qu (2014) showed that $S_T(\theta_0)$ is related to a linear multivariate regression. The dependent variables are related to $\operatorname{vec}(I_T(\omega_j) - f_{\theta_0}(\omega_j))$, and the regressors to $\frac{\partial \operatorname{vec} f_{\theta_0}(\omega_j)}{\partial \theta'}$. The rank of the regressors matrix is always bounded from above by the dimension of the structural parameter vector, irrespective of the strength of identification. When some parameters are unidentified, some columns of the regressor matrix become linearly dependent. Consequently, the dependent variables are projected onto a lower-dimensional space, resulting in a smaller test value. Formally, for any $0 \leq c < \infty$,

$$\lim_{T\to\infty} \Pr\left(S_T(\theta_0) \leq c\right) \to \Pr\left(\chi^2_{q-q_1} \leq c\right) \leq \Pr\left(\chi^2_q \leq c\right),$$

where $q = \dim(\theta_0)$ and q_1 is the number of unidentified parameter directions (i.e., the number of zero eigenvalues of $M_T(\theta_0)$). If the number of unidentified parameters is unknown, using the critical value of a χ^2_q distribution results in conservative inference.

4.2 Implementation

We explain the implementation of the tests using $\mathcal{H}_{dT}$ as an example. Suppose the desired significance level is α%. We first define positive constants α_S and $\alpha_{\mathcal{H}}$ such that $\alpha = \alpha_S + \alpha_{\mathcal{H}}$. Next, we invert (i.e., sample) the $S_T(\theta_0)$ test to obtain an $(1-\alpha_S)$% confidence set for the parameters, denoted by $C_\theta\left(1-\alpha_S\right)$. Then, we calculate

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

and reject the null hypothesis if it exceeds the $\alpha_{\mathcal{H}}$% critical value of $\sup_{r\in[0,1]} \left\|G_d(r)\right\|_\infty$. This two-step procedure first identifies a set of plausible parameter values and then assesses whether the most favorable model among them is supported by the data. In our implementation, for a 10% test, we set $\alpha_{\mathcal{H}} = \alpha_S = 5$, although other combinations might yield better power properties depending on the model. We leave such a power analyses for further work.

To determine the confidence set $C_\theta\left(1-\alpha_S\right)$, a grid search is not feasible for even small-scale DSGE models. Instead, we employ the Metropolis algorithm as described in Qu (2014). This algorithm uses Metropolis steps to generate frequent draws from areas in the parameter space where the values of $S_T(\theta)$ are low, and infrequent draws where $S_T(\theta)$ are high, creating an adaptive grid that is dense in important regions and sparse in unimportant areas. As in Qu (2014), we adjust the algorithm to account for any potential ridges or local minima in the surface of $S_T(\theta)$. The algorithm uses different proposal distributions to generate new parameter values. Specifically, the new draw is written as $\theta^* = \theta^{(j)} + \varepsilon$, where the first distribution gives $\varepsilon \sim N(0, M_T(\theta^{(j)}))$ with c being a tuning constant, and the second gives $\varepsilon = cV_T(\theta^{(j)})$ or $-cV_T(\theta^{(j)})$ with $V_T(\theta^{(j)})$ being the eigenvector corresponding to the smallest eigenvalue of $M_T(\theta^{(j)})$. These two distributions produce draws that move both across and along the ridges of $S_T(\theta)$. The tuning parameter c is also allowed to take on multiple values to avoid getting stuck in a small neighborhood around a local minimum. Finally, multiple Markov chains are run with different initial values, and the confidence set is obtained by merging the accepted values from all chains. This set can then be approximated using the values of θ for which $S_T(\theta)$ does not exceed the critical value of the Chi-square distribution.

We have chosen to obtain the testing result in two steps, rather than computing the infimum of the specification test over the entire parameter space, for two reasons. The first reason is computational: it becomes challenging to calculate the infimum when the number of parameters is high, such as when there are more than ten parameters. The second reason is from a modeling perspective: researchers often have interests in both inference and model testing. Our procedure provides both results: the score test provides confidence sets for the parameters, allowing for weak identification, while the specification test subsequently checks whether any of them are compatible with the data. The parameters that survive the testing can then be further analyzed, for example, by plotting their impulse responses, to gain insights into the model implications that are not rejected by the data. This approach separates specification testing from parameter estimation, enabling the assessment of any given set of parameter values with the data.

In our empirical applications, we also apply the specification tests to Bayesian posterior distributions of DSGE models. We first obtain the posterior distribution under informative priors and then apply our test to the parameters within the credible region. This analysis provides additional insights into the results obtained from the frequentist two-step procedure. If both methods lead to a rejection or acceptance of the model, we have greater confidence in the conclusion. If the posterior distribution results in a rejection while the frequentist two-step procedure does not, we can further investigate the surviving parameter values to assess their economic interpretations.

5 Simulations of size and power properties

We evaluate the size and power properties of our proposed specification tests using the model from Lubik and Schorfheide (2004) with empirically calibrated parameter values. The model is

$$
\begin{aligned}
y_t &= E_t y_{t+1} - \tau(r_t - E_t\pi_{t+1}) + g_t, \qquad (8)\\
\pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - z_t),\\
r_t &= \rho_r r_{t-1} + (1-\rho_r)\psi_1\pi_t + (1-\rho_r)\psi_2(y_t - z_t) + \varepsilon_{rt},\\
g_t &= \rho_g g_{t-1} + \varepsilon_{gt}, \quad z_t = \rho_z z_{t-1} + \varepsilon_{zt},
\end{aligned}
$$

where y_t, π_t, and r_t are log deviations of output, inflation and nominal interest rate from their steady states, respectively. The shocks ε_{rt}, ε_{gt}, and ε_{gt} are independently and identically distributed as $N(0,\sigma_r^2), N(0,\sigma_g^2)$, and $N(0,\sigma_z^2)$, respectively, and ε_{gt} and ε_{zt} are cross correlated with correlation coefficient ρ_{gz}. The observables are log levels of output, inflation and interest rate (both annualized), which are represented as $Y_t = (0,\, \pi^*,\, \pi^* + r^*)' + (y_t,\, 4\pi_t,\, 4r_t)'$, where the output is detrended and π^* and r^* are annualized steady-state rates of inflation and real interest rate with $\beta = (1 + r^*/100)^{-1/4}$. Under determinacy, the parameters and their values (posterior mean estimates from Lubik and Schorfheide, 2004) are

$$\begin{aligned}\theta &= (\tau, \beta, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r, \sigma_g, \sigma_z, \rho_{gz}, \pi^*)' \\ &= (0.54, 0.992, 0.58, 2.19, 0.30, 0.84, 0.83, 0.85, 0.18, 0.18, 0.64, 0.36, 3.43)'.\end{aligned}$$

Under indeterminacy, the sunspot parameters are added, so

$$\begin{aligned}\theta &= (\tau, \beta, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r, \sigma_g, \sigma_z, \rho_{gz}, M_{r\epsilon}, M_{g\epsilon}, M_{z\epsilon}, \sigma_\epsilon, \pi^*)' \\ &= (0.69, 0.997, 0.77, 0.77, 0.17, 0.60, 0.68, 0.82, 0.23, 0.27, 1.13, 0.14, -0.68, 1.74, -0.69, 0.20, 4.28).\end{aligned}$$

Lubik and Schorfheide (2004) transformed the model's solution to ensure that its impulse responses are continuous at the boundary between the determinacy and indeterminacy regions. The transformation applied is $S_t = \Theta_1 S_{t-1} + \widetilde{\Theta}_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t$ with $\widetilde{\Phi}_\varepsilon = \Phi_\varepsilon + \Phi_\epsilon(\Phi_\epsilon' \Phi_\epsilon)^{-1}\Phi_\epsilon'(\Phi_\varepsilon^b - \Phi_\varepsilon)$, where Φ_ε^b is the counterpart of Φ_ε with ψ_1 replaced by $\widetilde{\psi}_1 = 1 - (\beta\psi_2/\kappa)(1/\beta - 1)$. We apply the same transformation in order to be consistent with their analysis. Finally, the sunspot shock ϵ_t and the sunspot parameter are specified in the same way as described previously.

5.1 Size properties

We examine the size properties of the tests under both determinate and indeterminate regimes of the DGP, considering tests of the full model as well as subsets of one or two observables. We consider tests based on the business cycle frequencies only, the full spectrum, the full spectrum with the weighting function $W(\omega) = 1 - \omega/\pi$ and, finally, the mean and spectrum, with all test statistics computed with the prewhitening procedure described in Section 3. The sample sizes are chosen to reflect those frequently encountered in practice when working with DSGE models. All simulations presented in this subsection are conducted over 5000 replications.

Table 2 shows the results of testing all three observables. In the determinate case, we can see that, at the 10% nominal level, the tests using business cycle frequencies and weighted full spectrum have size close to the nominal level for all sample sizes. The full spectrum and mean and spectrum tests tend to be slightly conservative for smaller sample sizes, but improve with larger sample sizes. At the 5% nominal level, the mean and spectrum and full spectrum based tests have size close to nominal, while the business cycle frequencies and weighted full spectrum based tests have slight upward size distortions at 80 and 160 observations. This is improved for larger samples. In the indeterminate DGP, all tests generally perform well at the 10% nominal level, with mild downward and upward size distortions seen in the business cycle and full spectrum weighted tests, respectively, at T=80. At the 5% nominal level, small upward size distortions are seen in the business cycle frequencies and weighted full spectrum based tests for smaller samples, while the size is well-controlled for the rest. Overall, the tests tend to have correct size across various sample sizes and policy regimes of the DGP when all observables are tested.

Table 2: Rejection frequencies under the null hypothesis (10%, testing all variables)

Level	T	BC frequencies	Full spectrum	Weighted spectrum	Mean and spectrum
			Determinacy		
10%	80	0.099	0.085	0.105	0.085
	160	0.098	0.083	0.100	0.088
	240	0.090	0.087	0.098	0.101
	320	0.096	0.085	0.090	0.096
5%	80	0.061	0.045	0.070	0.048
	160	0.063	0.044	0.065	0.046
	240	0.046	0.044	0.057	0.049
	320	0.052	0.042	0.053	0.049
			Indeterminacy		
10%	80	0.106	0.088	0.122	0.108
	160	0.102	0.097	0.111	0.102
	240	0.104	0.092	0.109	0.099
	320	0.104	0.091	0.095	0.098
5%	80	0.072	0.049	0.081	0.059
	160	0.058	0.047	0.070	0.050
	240	0.059	0.048	0.064	0.051
	320	0.060	0.044	0.057	0.047

Note. T: sample size; all tests are computed with prewhitening. π, y, and r denote inflation, output, and interest rate, respectively.

We next analyze the results presented in Table A1, which shows the empirical size of tests conducted on pairs of observables. The results indicate that the size of the tests is well-controlled across sample sizes and policy regimes when testing pairs of observables. At a 10% nominal level, most tests tend to be slightly below the nominal size for sample sizes less than 320 under the determinate regime, and to a lesser extent under the indeterminate regime. At a 5% nominal level, the tests based on the weighted full spectrum tend to produce the closest empirical sizes to the nominal level (5%) across all sample sizes. The other testing procedures maintain empirical sizes that are well within 1.5% of the nominal level for all pairs of observables and sample sizes.

Finally, we turn to Table A2, which contains empirical sizes for testing a single observable. We can see that, in general, the tests tend to be slightly conservative at both nominal levels, particularly so for tests based on the business cycle frequencies. The size also depends on the observable tested and hence on the nature of the DGP. For example, tests of inflation tend to be very well sized compared with the rest, particularly when considering testing based on the mean and spectrum and full spectrum.

In conclusion, our examination of empirical sizes for tests of the full model and subsets of observables suggests that the proposed tests exhibit good size control across various sample sizes and regimes of the DGP. However, it is important to emphasize that prewhitening the data before performing the tests is a critical step in maintaining accurate size control, as substantial distortions were observed when using unfiltered data. For example, without prewhitening, the rejection rates corresponding to the first row of Table 2 were 0.183, 0.285, 0.217, and 0.268, respectively, which all exceeded 20%. The fact that a DSGE model is fully parametric enables us to stabilize the size properties without introducing additional estimation uncertainty.

5.2 Power properties

We next examine the empirical power properties of the proposed tests. To do so, we compute the size-adjusted power of the tests at a 10% nominal significance level, against alternatives that perturb a random element of the parameter vector by a fixed percentage (we consider 20% and 40%). The tests are computed using pre-whitening, as in the previous section.

Table 3 displays the results of testing all observables. The tests that use the full spectrum and the combination of the spectrum and mean have the highest rejection rates. The rejection frequencies for the mean and spectrum tend to be similar or lower than the full spectrum, likely due to the presence of an extra parameter in the steady state. In the determinate case, tests based on the business cycle frequencies have a power of 52-62% compared to the full spectrum, and tests based on the weighted full spectrum achieve 62-67% of the full spectrum's power when the parameter is perturbed by 20%. These ranges become 47%-75% and 56-84%, respectively, when the perturbation is 40%. In the indeterminate case, the rejection frequencies are often lower because of the additional four parameters in the model. Otherwise, the patterns are similar to the determinate case: the business cycle frequencies and the weighted full spectrum tests have 60-70% and 68-82% of the full spectrum's power, respectively, when the random parameter is perturbed by 20%. These ranges become 63%-79% and 70-89% when the perturbation is 40%. The reason why the tests based on weighting functions appear to be more powerful (relative to the full spectrum based tests) in this scenario is likely because the spectral densities of observables under indeterminacy have much more mass at lower and business cycle frequencies. Since these frequencies receive higher weight, the business cycle and weighted full spectrum based tests become relatively more informative compared to the determinate case.

Tables A3 and A4 in the Appendix contain the results summarizing the empirical power of the tests involving pairs of observables and individual observables, respectively. Under determinacy, the inflation and output pair has rejection frequencies only slightly below those of the full model. The inflation and interest rate pair produces similar or lower power across the board, while the output and interest rate pair consistently has lower power than the other two for all tests. Under indeterminacy, the best performing pairings under 20% perturbation, opposite to the determinacy case, are now those of inflation and output with interest rate. When the perturbation is increased to 40%, the differences between rejection frequencies among the pairs become much more muted.

Turning to tests using individual observables, Table A4 shows that such tests can be infor-

Table 3: Rejection frequencies under the alternative hypothesis (10%, testing all variables)

T	BC frequencies	Full spectrum	Weighted spectrum	Mean and spectrum
		Determinacy		
		Perturb a random element of θ by 20%		
80	0.210	0.341	0.228	0.305
160	0.266	0.506	0.312	0.425
240	0.330	0.627	0.413	0.520
320	0.384	0.691	0.458	0.598
		Perturb a random element of θ by 40%		
80	0.290	0.618	0.345	0.544
160	0.459	0.790	0.536	0.700
240	0.566	0.833	0.666	0.752
320	0.653	0.872	0.729	0.780
		Indeterminacy		
		Perturb a random element of θ by 20%		
80	0.232	0.332	0.271	0.317
160	0.299	0.436	0.324	0.394
240	0.332	0.525	0.358	0.448
320	0.348	0.576	0.416	0.497
		Perturb a random element of θ by 40%		
80	0.320	0.510	0.359	0.480
160	0.403	0.644	0.481	0.579
240	0.502	0.688	0.571	0.635
320	0.577	0.731	0.648	0.671

Note. T: sample size; all tests are computed with prewhitening. π, y, and r denote inflation, output, and interest rate, respectively.

mative as the rejection frequencies show nontrivial power even at small sample sizes. Under determinacy, it seems the most informative tests obtain when looking at inflation, while the lowest rejection frequencies consistently pertain to output. Under indeterminacy, interest rate produces the highest rejection frequencies, followed somewhat closely by inflation, while tests based on output have markedly lower rejection frequencies.

In summary, the results above show that the proposed tests have nontrivial power in empirically relevant sample sizes, that the tests using weighting functions, such as business cycle frequency indicator or a smoothing function de-emphasizing higher frequencies, can be informative, and that the tests still have power when considering subsets of observables. The power properties can depend on the structure of the DGP, e.g., business cycle based tests could be more informative if the model implies that the spectral densities of observables have a substantive mass in that band.

6 Empirical applications (in progress)

In this section, we examine three DSGE models: a small-scale model of Lubik and Schorfheide (2004), previously considered in Section 5, and two medium-scale models: the Smets and Wouters (2007) model and the news shocks model of Schmitt-Grohé and Uribe (2012). The Lubik and Schorfheide (2004) model is a popular choice for contrasting determinacy and indeterminacy within a small-scale framework. The Smets and Wouters (2007) model is a benchmark medium-scale New Keynesian model in academia and central banks. This model extends the standard New Keynesian model by incorporating additional frictions and real rigidities, allowing us to examine how model specification improves compared to the baseline small-scale model. The Schmitt-Grohé and Uribe (2012) model provides an opportunity to evaluate whether the proposed information structure generates dynamics that fit the data adequately.

6.1 The small-scale model

We perform specification testing on the model described in Section 5 at the 10% level under both determinacy and indeterminacy. The data are linearly detrended US log GDP, and annualized inflation and interest rates for the period 1960:I-2007:IV. We do not use Hordick-Prescott filter to avoid potential filtering-induced discrepancies near the zero frequency. We consider this full sample period as a starting point, and to evaluate to what extent the results are driven by potential differences in monetary policy regimes over time, we also consider two subsamples: the pre-Volcker period (1960:I-1979:II) and the post-Volcker period (1979:III-2007:IV), which are associated with indeterminate and determinate policy regimes, respectively, see Clarida, Gali and Gertler (2000) and Lubik and Schorfheide (2004). The partitioning of the sample using 1979:II is the same as in Clarida, Gali and Gertler (2000).

We obtain results using the two-step procedure outlined in Subsection 4.2. First, a 95% confidence set is obtained by inverting the score test of Qu (2014) in (7). The modified Metropolis algorithm from Qu (2014) is used to generate 100 Markov chains from different initial values, each of which is run until 1000 draws are accepted and subsequently merged. Then, the specification tests based on the weighted full spectrum, business cycle frequencies and mean and full spectrum are conducted at the 5% level for each parameter vector in the confidence set to examine the specification of the full model, each observable separately and their pairs. The test statistic, critical value, and the percentage of draws rejected are reported for each case. All specification test statistics are computed with prewhitening and the test statistics based on the full spectrum are computed as $\mathcal{H}_{dT}^{W}(\theta_0)$ in (4) with the weight function $W(\omega) = 1 - \omega/\pi$ that puts less weight on higher frequencies that DSGE models are not designed to capture. Note that the test statistic for the mean and spectrum case is $\mathcal{H}_{T}(\theta_0)$, i.e., it does not apply the weight function. Otherwise, the steady state test $\mathcal{H}_{sT}(\theta_0)$ would dominate due to the critical values for the static test being much higher, essentially reducing the test to focusing on the steady state only. Finally, to relate our results to Bayesian DSGE literature, we obtain 200k draws from the posterior distribution with the likelihood under informative priors, drop 5% of the draws corresponding to the lowest density regions, and apply our specification tests to these parameters.

Below, for each case, we first summarize the main findings and then provide numerical details.

Result 1 *When using the full sample 1960:I-2007:IV, the model is rejected at the 10% level based on both the full spectrum analysis and business cycle frequency analysis for both determinacy and indeterminacy specifications. Subsequent analysis reveals that misspecification impacts most segments of the model, and, in particular, inflation dynamics and its comovements with GDP are incompatible with the data at these parameter values over business cycle frequencies. Using the MCMC draws from posterior distributions reinforces these conclusions.*

Table 4 presents the 95% confidence intervals for the 1960:I-2007:IV sample based on the mean and full spectrum for the determinacy specification. The intervals are most informative

Table 4: 95% confidence intervals, 1960:I-2007:IV, determinacy

θ	Parameter	Bounds	CI
τ	intertemporal substitution elasticity	[0.10, 1.00]	[0.10, 0.99]
β	discount factor	[0.98, 0.999]	[0.984, 0.999]
κ	Phillips curve slope	[0.01, 2.00]	[0.01, 1.998]
ψ_1	inflation target	[1.01, 3.00]	[1.01, 2.56]
ψ_2	output target	[0.01, 5.00]	[0.01, 4.99]
ρ_r	interest rate smoothing	[0.10, 0.90]	[0.68, 0.90]
ρ_g	exogenous spending AR	[0.10, 0.98]	[0.88, 0.98]
ρ_z	technology shock AR	[0.10, 0.98]	[0.92, 0.98]
σ_r	monetary policy shock SD	[0.01, 3.00]	[0.20, 0.40]
σ_g	exogenous spending SD	[0.01, 3.00]	[0.03, 0.15]
σ_z	technology shock SD	[0.01, 3.00]	[0.64, 2.09]
ρ_{gz}	exogenous spending-technology CORR	[-0.90, -0.90]	[-0.45, 0.90]
π^*	steady state inflation	[2.00, 8.00]	[2.00, 7.95]

Note. Values are based on the mean and full spectrum. Column 2: parameter interpretation. Column 3: bounds for permissible parameter values. Columns 4: Confidence intervals, obtained by sampling the score test and applying projections.

about the autoregressive coefficients and standard deviations for the exogenous shock processes, with the standard deviation of the technological shock having a relatively wider interval. Thus, the confidence set includes models with widely varying behavioral and policy parameters, but relatively restricted exogenous shock behavior.

Examining the specification test results in Table 5, the first notable conclusion is that 100% of all parameter values in the confidence set are rejected when considering either the weighted full spectrum or only business cycle frequencies. Therefore, the null of correct model specification is rejected at the 10% significance level in both cases. In order to pinpoint the sources of this misspecification, further tests based on individual and pairs of observables can be considered. Based on the full spectrum, all parameter draws are rejected for inflation and the GDP/inflation pair and less than 1% of draws survive for the GDP/interest rate pair. The results at business cycle frequencies are qualitatively similar, with only 1% of draws surviving for inflation and less than 1% surviving for the two pairs.

Table 5: Specification test results, 1960:I-2007:IV, determinacy

	Weighted spectrum			BC frequencies			Mean and spectrum		
	Test	CV	Rej.	Test	CV	Rej.	Test	CV	Rej.
Full model	1.239	1.128	100	1.779	1.370	100	2.614	2.851	99.97
GDP	0.619	0.977	93	0.452	1.151	82	1.715	2.478	30
Inflation	0.964	0.977	100	0.710	1.151	99	2.402	2.478	99.98
Interest rate	0.320	0.977	31	0.306	1.151	6	0.807	2.478	65
GDP-Inflation	1.140	1.077	100	1.086	1.295	99.95	2.286	2.729	99
GDP-Interest rate	0.953	1.077	99.91	1.090	1.295	99.69	1.894	2.729	98
Inflation-Interest rate	0.877	1.077	85	0.979	1.295	67	2.654	2.729	99.99

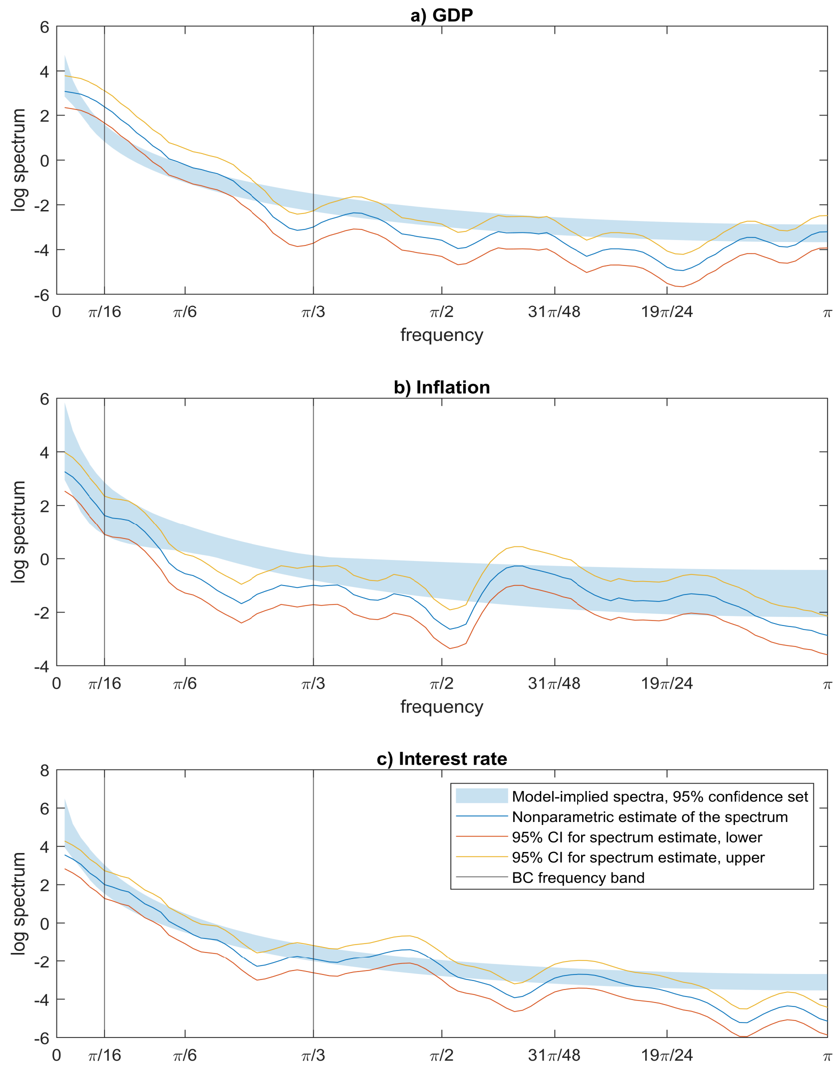
Note. The significance level is 10%. Test: the spcification test value; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test.

To help visualize these results, Figure 2 contrasts nonparametrically estimated log spectral densities against the model implied log spectra using the parameter values of the 95% confidence set, while Figures 3 and 4 plot real and imaginary parts of cross-periodograms against their counterparts implied by the confidence set, respectively. The real part of the cross-spectrum at frequency ω, also called the cospectrum, shows which portion of the covariance is due to cycles at that frequency. The imaginary part of the cross-spectrum, also called the quadrature spectrum, indicates whether one series leads or lags the other series, as determined by its sign relative to the real part (i.e., a positive sign in both cases indicates that the first series leads the second series).

Upon examination of the figures, it is clear that there is a significant lack of overlap between the nonparametric 95% confidence interval and the model-implied spectral density for inflation within the business cycle frequency band. Specifically, the model overpredicts the distribution of the variation at these frequencies, which explains why both full spectrum and business cycle frequency-based specification tests reject the model when only inflation is considered. In comparison, the fit for GDP within the business cycle frequencies is slightly better, while the agreement for interest rates is much higher, consistent with the results in the table, where only 6% of the draws are rejected.

Figures 3 and 4 illustrate the estimated and model-implied cross-spectra for all pairs at

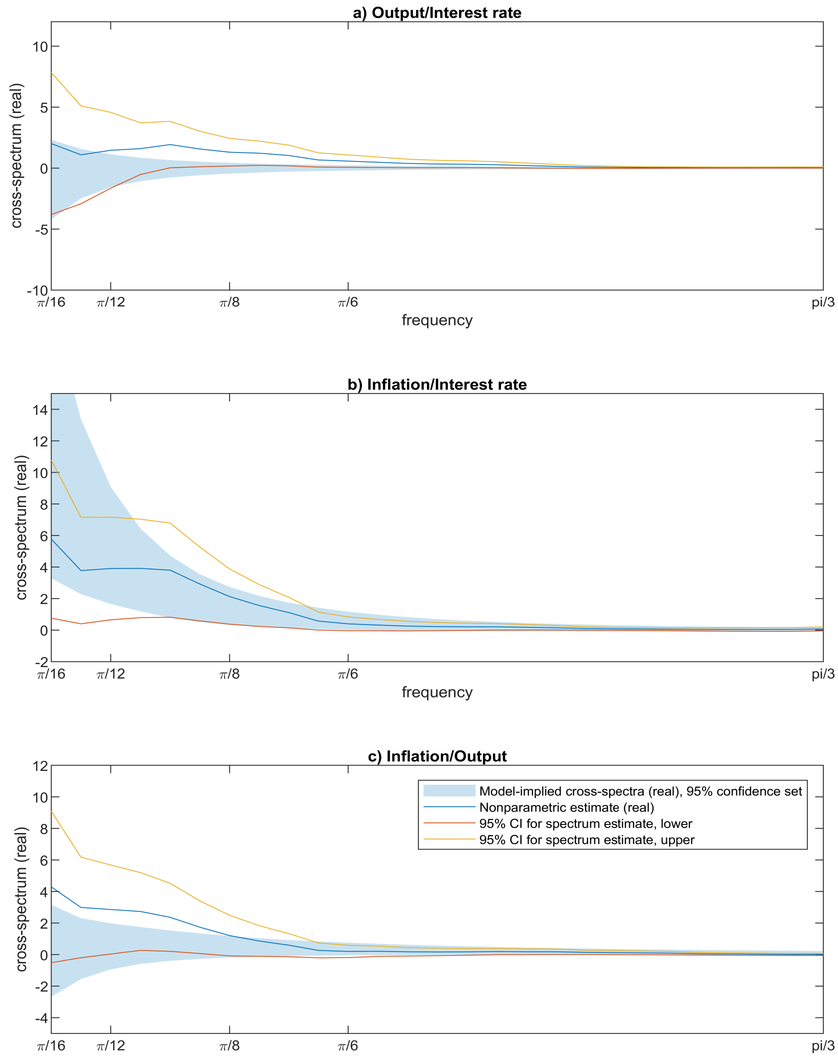
Figure 2: Log spectra under determinacy, 1960-2007.



business cycle frequencies. Visible discrepancies occur between the inflation/output and output/interest rate pairs. For inflation/output, the data suggest that output leads inflation at business cycle frequencies (Figure 4(c)), whereas the model is ambiguous in this regard. For output/interest, the data suggest stronger out-of-phase comovement between these two variables than what is implied by the model (Figure 4(a)). Thus, the model struggles to capture dynamic correlations between two pairs of the three observables, suggesting limitations in its ability to explain the data. Finally, for the inflation/interest rate pair, the fit is better, with both the model and the data indicating that inflation leads interest rates. Nevertheless, the model still tends to signal a stronger relationship than in the data.

Next, model specification for the full sample under indeterminacy is examined. The relevant 95% confidence intervals are reported in Table 6. It can be seen that the intervals, again, are

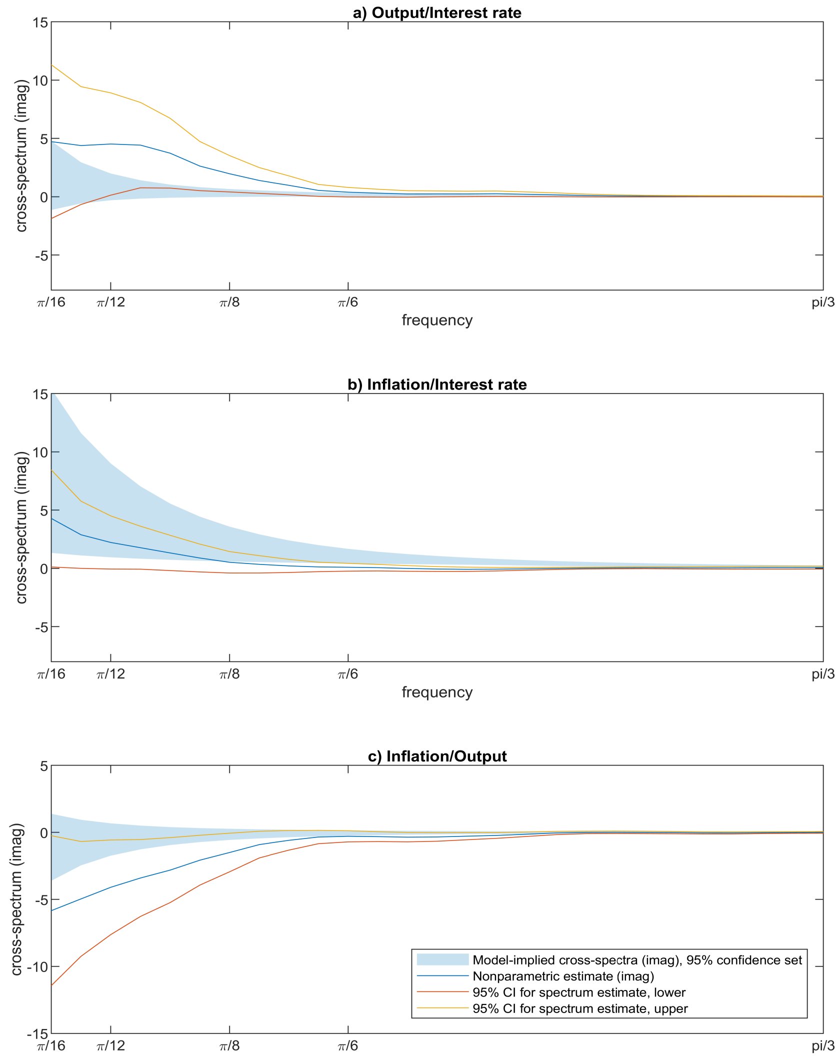
Figure 3: Cross spectra (real part) under determinacy, 1960-2007.



most informative about the parameters governing the dynamics of exogenous shocks. Compared to the determinacy case, all these intervals are wider. Different from the determinacy case, the interval for the slope of the Phillips curve is much tighter. Among the sunspot parameters, only the interval for the standard deviation of the sunspot shock is somewhat informative.

Proceeding to specification testing, results in Table 7 show that 100% and 99.92% of parameter values are rejected based on the weighted full spectrum and business cycle band, respectively – this conclusion is qualitatively similar to that under determinacy. It is apparent from Table 7 that comovements between variables are identified as a primary source of misspecification at both frequencies. Interest rate dynamics remain the model component that is most compatible with the data, particularly at business cycle frequencies, where 97% of draws cannot be rejected. Figures A1, A2, and A3, in the appendix show plots of estimated versus model implied log spectra

Figure 4: Cross spectra (imaginary part) under determinacy, 1960-2007.



and real and imaginary parts of estimated versus model implied cross-spectra, respectively. The plots are qualitatively similar to those under determinacy. The main difference is that the model-implied confidence bands tend to be significantly wider, consistent with the wider confidence intervals in Table 6. These results show that the model fit has not fundamentally changed by switching from determinacy to indeterminacy for the entire sample period.

We also consider parameter values from a posterior distribution using priors from Lubik and Schorfheide (2004), and retain all posterior draws that are in the 95% highest density region. Specification test results under determinacy and indeterminacy can be found in Tables A5 and A6 in the appendix, respectively. The overall conclusion remains unchanged: the full model is rejected based on the weighted full spectrum in both cases, with comovements between variables identified as a primary source of misspecification. However, a difference from the previous results

Table 6: 95% confidence intervals, 1960:1-2007:IV, indeterminacy

θ	Parameter	Bounds	CI
τ	intertemporal substitution elasticity	[0.10, 1.00]	[0.10, 0.9999]
β	discount factor	[0.98, 0.999]	[0.989, 0.999]
κ	Phillips curve slope	[0.01, 2.00]	[0.01, 0.458]
ψ_1	inflation target	[0.01, 0.99]	[0.01, 0.952]
ψ_2	output target	[0.01, 5.00]	[0.14, 4.999]
ρ_r	interest rate smoothing	[0.10, 0.90]	[0.10, 0.90]
ρ_g	exogenous spending AR	[0.10, 0.98]	[0.72, 0.97]
ρ_z	technology shock AR	[0.10, 0.98]	[0.88. 0.98]
σ_r	monetary policy shock SD	[0.01, 3.00]	[0.01, 0.39]
σ_g	exogenous spending SD	[0.01, 3.00]	[0.04, 0.68]
σ_z	technology shock SD	[0.01, 3.00]	[0.65, 2.69]
ρ_{gz}	exogenous spending-technology CORR	[-0.90, -0.90]	[-0.90, 0.83]
$M_{r\epsilon}$	sunspot-monetary coeff	[-3.00, 3.00]	[-2.956, 2.999]
$M_{g\epsilon}$	sunspot-exogenous spending coeff	[-3.00, 3.00]	[-3.00, 2.48]
$M_{z\epsilon}$	sunspot-technology coeff	[-3.00, 3.00]	[-0.54, 0.96]
σ_ϵ	sunspot shock SD	[0.01, 3.00]	[0.01, 0.84]
π^*	steady state inflation	[2.00, 8.00]	[2.00, 8.00]

Note. Values are based on the mean and full spectrum. Column 2: parameter interpretation. Column 3: bounds for permissible parameter values. Columns 4: Confidence intervals, obtained by sampling the score test and applying projections.

Table 7: Specification test results, 1960:I-2007:IV, indeterminacy

	Weighted spectrum			BC frequencies			Mean and Spectrum		
	Test	CV	Rej.	Test	CV	Rej.	Test	CV	Rej.
Full model	1.137	1.128	100	1.277	1.370	99.92	2.504	2.851	98
GDP	0.471	0.977	97	0.302	1.151	16	1.359	2.478	41
Inflation	1.101	0.977	100	1.190	1.151	100	2.534	2.478	100
Interest rate	0.324	0.977	16	0.293	1.151	3	0.727	2.478	21
GDP-Inflation	1.214	1.077	100	1.300	1.295	100	2.556	2.729	99.67
GDP-Interest rate	0.662	1.077	95	0.799	1.295	86	1.521	2.729	35
Inflation-Interest rate	1.106	1.077	100	0.957	1.295	99.88	2.556	2.729	99.85

Note. The significance level is 10%. Test: the spcification test value; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test.

is that inflation dynamics are no longer rejected when considered individually under determinacy, with 38% and 44% of draws surviving based on full spectrum and business cycle frequencies, respectively. Under indeterminacy, inflation dynamics have 4% and 19% of draws surviving in the respective cases. This finding suggests that incorporating an informative prior can alter the fit of the model in some dimensions, but it does not make the model fully compatible with the data. Diagnostic plots for log spectra and cross-spectra can be found in Figures A4-A9 in the

Appendix. As before, the imaginary parts of the output/interest and inflation/output pairs show the largest discrepancies, with their nonparametric counterparts remaining largely outside the 95% model-implied confidence sets. The model-implied confidence sets are smaller than their frequentist counterparts, reflecting the effect of the prior.

The analysis has shown that both determinacy and indeterminacy specifications are rejected based on the weighted full spectrum test when applied to the 1960-2007 period. It has also revealed that dynamic correlations, particularly the leads and lags relationships, are often at odds with data. Moreover, the analysis suggests that examining the imaginary part of the spectra is informative, particularly for understanding the leads and lags relationships. Lubik and Schorfheide (2004) have convincingly demonstrated that a determinate regime fits better for the post-Volcker subsample, while an indeterminate regime fits the data better for the earlier subsample with heightened parameter uncertainty. Therefore, it is appropriate to examine the model specification that matches the two monetary policy regimes to their respective subsamples. Our findings for the 1979-2007 period are summarized below:

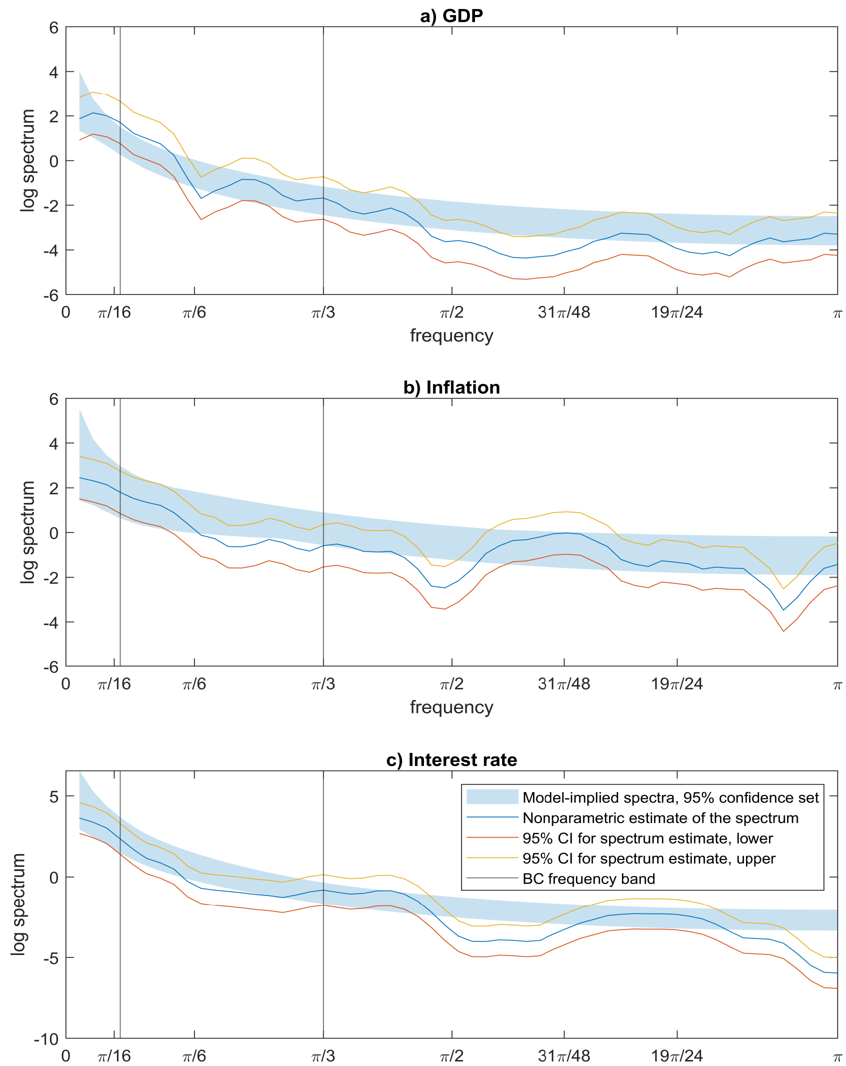
Result 2 *For the subsample 1979:III-2007:IV under determinacy, the model is no longer rejected when using both weighted full spectrum and using business cycle frequencies. Further analysis revealed that the model's dynamic properties remain similar to the full-sample determinacy estimates. However, in this part of the data, the comovements between variables are stronger between inflation and interest rate and weaker otherwise than the measurements using the full sample, resulting in a closer match between the model and the data. Using the MCMC draws from posterior distributions delivers qualitatively similar conclusions.*

The 95% confidence intervals for this case are collected in Table A7. Qualitatively, their pattern is similar to the full sample case considered earlier under determinacy - the most informative intervals pertain to autoregressive coefficients and standard deviations of the shock processes. Most of those intervals are moderately wider, owing most likely to smaller sample size.

Turning to the specification tests in Table A8, we observe that the tests reject only 15% and 39% of draws for the weighted full spectrum and business cycle frequencies, respectively. Examining the results for individual variables and their comovements, it becomes clear that the model specification substantially improves, except for inflation dynamics and its comovement with GDP, which have the highest proportions of rejected draws based on the full spectrum (97% and 86%, respectively) that remain high at business cycle frequencies (92% and 76%, respectively).

Figures A10-A12 show diagnostic spectral plots for this case. The model-implied sets of log spectra now visibly overlap more with the 95% confidence intervals of their nonparametrically estimated counterparts. For the cross spectrum, the data and model-implied spectra of inflation and interest rates agree well with each other because the contemporaneous comovement

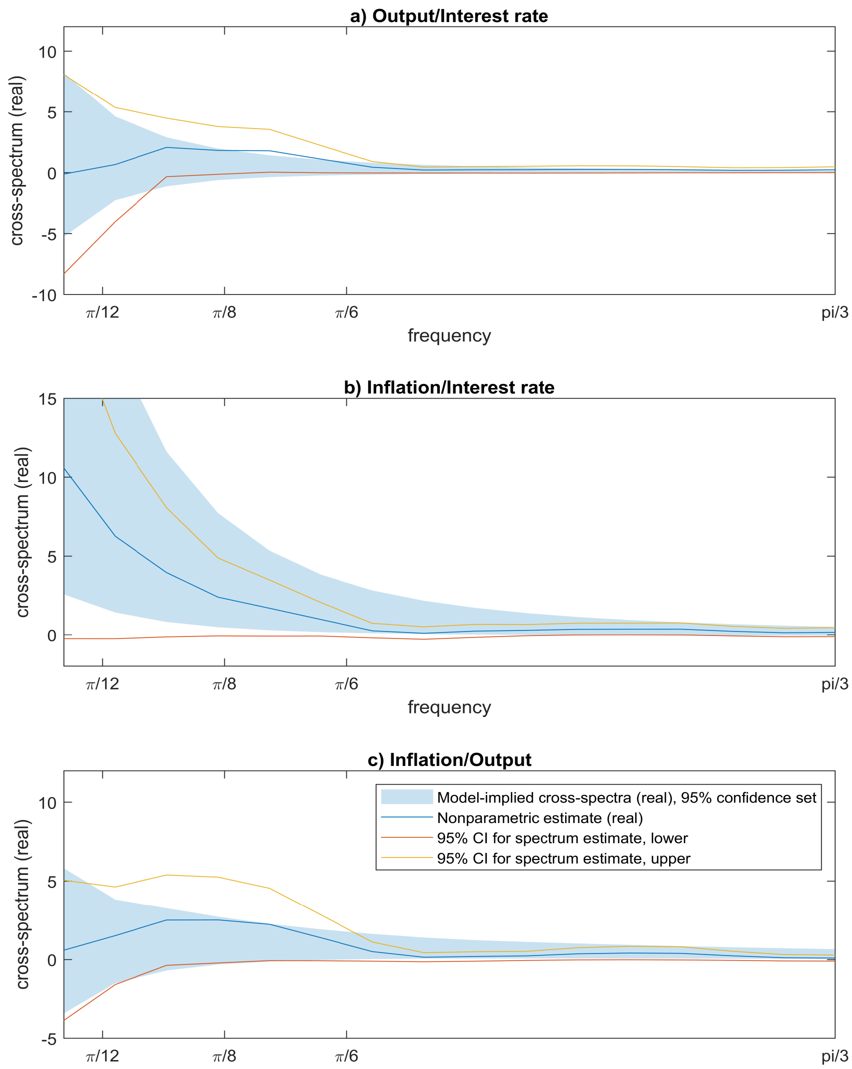
Figure 5: Log spectra under determinacy, 1979-2007.



between these two variables (i.e., the real part of the cross-spectrum) is stronger in this part of the sample. Additionally, the out-of-phase comovement between output and interest in the data is weaker in this part of the sample, bringing the model closer to mirroring the data. Finally, for the inflation/output pair, the real parts of the cross spectrum are closer, however, the lead-lag relationship in the data remains in tension with the model, in the sense that the nonparametric estimate partly lies outside of the model-implied confidence band, consistent with the test rejecting most of the draws for this pair.

Table A9 contains the test results using the Bayesian posterior draws. We find qualitatively similar overall conclusions: the model is not rejected using both full spectrum and business cycle frequencies, The comovement between inflation and interest rate is well-captured by the model, and inflation/GDP comovement constitutes a major tension between the model and data, with

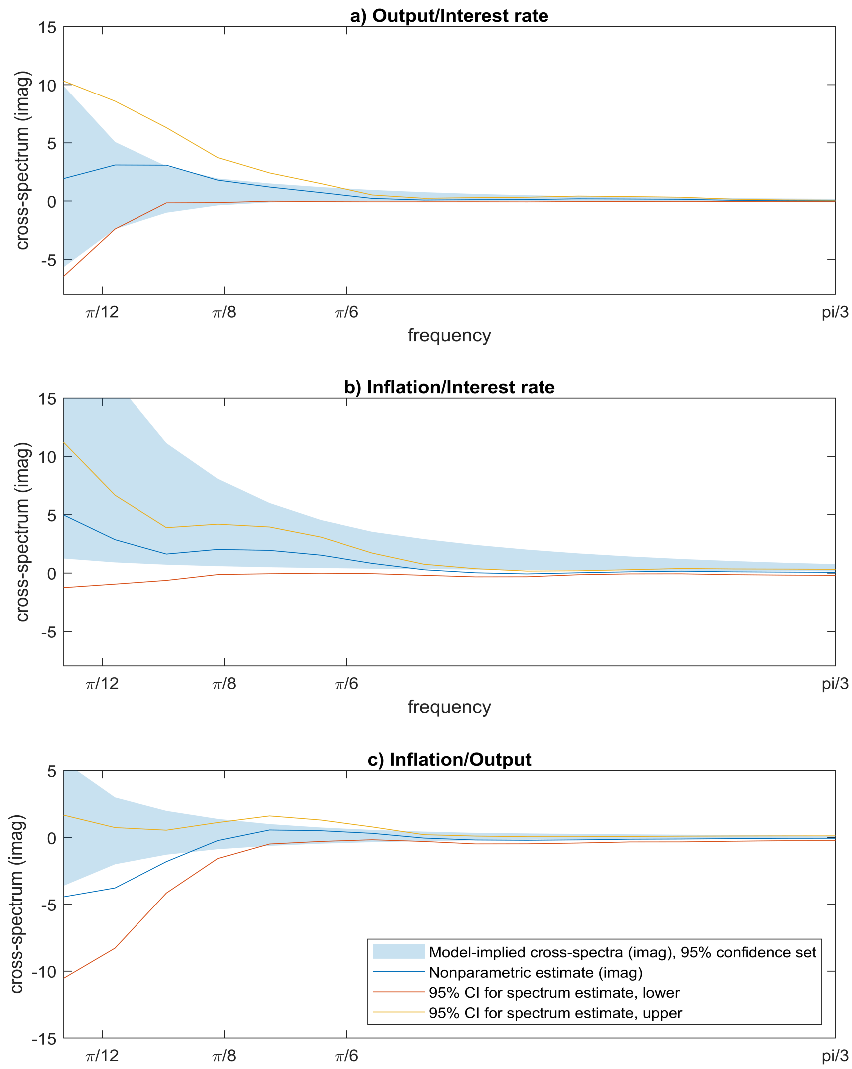
Figure 6: Cross spectra (real part) under determinacy, 1979-2007.



78% of posterior draws rejected based on the weighted full spectrum. The spectra are displayed in Figures A10-A12. The marginal spectra results are comparable to their frequentist counterparts, while the model implied intervals for the cross-spectra are significantly narrower, reflecting the effect of the prior. Next, we turn to the 1960-1979 period.

Result 3 *When examining the subsample 1960:I-1979:II under indeterminacy, the model is not rejected. An important factor contributing to the improved fit is the wider model-implied confidence intervals, which can encompass richer patterns in the data. Meanwhile, using MCMC draws from posterior distributions yields much higher percentage of rejected draws, specifically 99.99% for the full-model weighted full spectrum case. This may reflect the effect of the prior, coupled with a small sample size. Therefore, the use of frequentist confidence sets yields a more positive evaluation of the model specification for this sample period than using the Bayesian posterior*

Figure 7: Cross spectra (imaginary part) under determinacy, 1979-2007.



distribution.

The 95% confidence intervals for this case are presented in Table A10. Compared to the full sample indeterminacy case, the pattern of the most informative intervals remains the same (slope of the Phillips curve, exogenous shock parameters and the sunspot shock standard deviation), however the intervals become substantially longer, most likely due to reduced sample size. Table A11 contains the specification test results, which show that the model is not rejected at the 10% level based on both the weighted full spectrum (55% of draws rejected) and the business cycle band (44% of draws rejected). Examining tests for specification of separate aspects of the model, it can be seen the agreement of the individual variables and their comovements is improved compared to the full sample case, with wide model-implied confidence intervals, especially for cross spectra. Inflation dynamics and its comovement with GDP still have the highest proportions of rejected draws using full spectrum (61% and 64%, repsectively), however, these fall when only business cycle band is used, to 11% and 49%, respectively. The plots in Figures A13-A15 further corroborate these findings.

Table A12 and Figures A16-A18 provide results of using the Bayesian posterior distribution. The credible sets for the spectra plots are substantially narrower than the frequentist case, consistent with the high rejection rates of the tests.

We note that the data suggest that interest rates lead inflation over business cycle frequencies (Figure A18(b)), a departure from the lagging pattern observed during the 1979-2007 period (Figure A12(b)). The model tends to indicate the opposite relationship for this period, regardless of whether frequentist or Bayesian confidence sets are used. This tension is insufficient to cause a rejection of the model.

In summary, we have applied our diagnostic methods to a small-scale DSGE model and provided a detailed illustration of the results. The first part of our findings reaffirms Lubik and Schorfheide's (2004) results obtained from a Bayesian perspective. Specifically, our tests support that the post-Volcker sample is consistent with a determinate monetary regime, while the earlier sample is consistent with an indeterminacy regime, albeit with heightened parameter uncertainty. Furthermore, our results show that the commovement between variables, particularly the lead and lag relationships indicated by the imaginary part of the spectra, often constitutes a significant source of tension between model and data. Comparing the model and data spectra along with their confidence intervals offers a useful way to gain insights into the model's specification. Finally, for this model, both frequentist and Bayesian confidence sets yield qualitatively similar conclusions. The most significant difference occurs for the pre-Volcker sample case, where the use of frequentist confidence sets leads to a more positive evaluation of the model specification compared to using the Bayesian posterior distribution.

6.2 The Smets and Wouters model

The next model we consider is the medium-scale model of Smets and Wouters (2007), which is described briefly in the Supplementary Appendix. This model includes 36 free parameters and is estimated on seven observables: consumption growth, investment growth, output growth, labor hours, inflation, wage growth, and interest rate. Unlike the previous subsection, the variables including inflation and interest rate are not annualized here, as in the original analysis.

We perform specification testing at the 10% level using the same two-step procedure. Given the model's larger dimension, the confidence set is formed via merging output from 100 Markov chains each of which produces 5000 accepted draws. The specification testing proceeds in the same way as in the previous subsection. To relate to the Bayesian literature, the specification tests are also run on 0.5 million draws from the posterior distribution. In order to evaluate the contrast in specification for common observables such as inflation and interest rate, the application first considers US data for the same sample period of 1960:I-2007:IV as in the previous subsection. Subsequently, the tests are repeated on the subsample 1965:I-2004:IV, which corresponds to the original Smets and Wouters (2007) data.

Result 4 *The model is not rejected at either full spectrum or business cycle frequencies at the 10% significance level using the full sample 1960:I-2007:IV. This contrasts with the case of the small scale model examined earlier. Further examination reveals that, similar to the small scale case, one of the main sources of tension between the model and the data are inflation and its comovements with other variables, especially with interest rate. Using draws from the posterior distribution produces qualitatively similar results. Regardless of which set of draws is used, over 60% and 80% of the draws are rejected based on the weighted spectrum and business cycle frequencies, respectively, indicating significant room for model improvement.*

Table 8: Specification test results for the SW model, 1960:I-2007:IV

	(a) Frequentist set			(b) Posterior distribution		
	Test	CV	Rej.	Test	CV	Rej.
Full model, weighted spectrum	0.852	1.235	83	0.878	1.235	67
Full model, BC frequencies	1.10	1.519	95	1.128	1.519	81
GDP growth	0.287	0.977	60	0.317	0.977	20
Inflation	0.378	0.977	93	0.398	0.977	58
Interest rate	0.323	0.977	32	0.372	0.977	30
GDP-Inflation	0.583	1.077	92	0.597	1.077	26
GDP-Interest rate	0.620	1.077	68	0.673	1.077	55
Inflation-Interest rate	0.499	1.077	90	0.552	1.077	81

Note. Each row represents a set of variables tested. The significance level is 10% for each case. Test: the specification test value, based on the weighted full spectrum unless indicated otherwise; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test. For (a) the parameter values are obtained by sampling the score test, and for (b) they are values from 95% high density region of the posterior distribution using SW's prior.

Table 8 presents the results of specification tests using the 1960:I-2007:IV sample. Panel (a) uses parameter values from the score test and (b) uses values from the 95% highest density region of the posterior distribution under Smets and Wouter's prior. In (a), the null of a correct specification cannot be rejected at the 10% level using either the weighted full spectrum or business cycle frequencies. Nonetheless, since the majority of the draws are rejected, additional testing for variable subsets is conducted to reveal the dimensions along which the model and the data are most at odds with each other. In this regard, the table shows that 10% or less of draws survive when testing inflation by itself or in pairs with all other observables, indicating challenges for the model in capturing inflation dynamics and its comovements with other variables, similar to the small scale model for this sample. In (b), the rejection rates are generally lower than in (a), but the inflation variable again tends to have higher rejection frequencies. Notably, the inflation-interest pair is rejected over 80% of the time, the highest among all pairwise testing results. Further spectra diagnostics are in progress to understand these results.

We next turn to Smets and Wouters' original sample period.

Result 5 *When the original Smets and Wouters' (2007) sample period is considered, the model is, again, not rejected at 10% level using either the weighted full spectrum or business cycle frequencies. Overall, the rates of rejection are lower than the full sample case, while the specification of inflation and its comovements with other observables produce most tension with the data. The analysis is repeated using the Bayesian version of the confidence set, and qualitatively similar conclusions are reached.*

Table 9: Specification test results for the SW model, 1965:I-2004:IV

	Frequentist set			Posterior distribution		
	Test	CV	Rej.	Test	CV	Rej.
Full model	0.868	1.235	65	0.901	1.235	89
Full model, BC frequencies	1.019	1.519	85	1.048	1.519	71
GDP growth	0.227	0.977	20	0.194	0.977	0.3
Inflation	0.323	0.977	77	0.371	0.977	42
Interest rate	0.294	0.977	8	0.331	0.977	47
GDP-Inflation	0.503	1.077	67	0.516	1.077	10
GDP-Interest rate	0.561	1.077	26	0.626	1.077	52
Inflation-Interest rate	0.470	1.077	66	0.628	1.077	88

Note. Each row represents a set of variables tested. The significance level is 10% for each case. Test: the specification test value, based on the weighted full spectrum unless indicated otherwise; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test. For (a) the parameter values are obtained by sampling the score test, and for (b) they are values from 95% high density region of the posterior distribution using SW's prior.

Table 9 presents the specification test results for the sample period of 1965:I-2004:IV. It shows that the correct specification null cannot be rejected at the 10% level for all the tests considered. The rejection frequencies overall are lower than in the full sample case. Using the Bayesian credible set yields similar conclusions. The inflation and interest rate pair remains the most rejected case among all pair-wise cases. Further spectra diagnostics are in progress to dissect these results.

6.3 A model with news shocks

We study the model of Schmitt-Grohé and Uribe (2012). The model features anticipated shocks, the quantitative importance of which is actively investigated in the literature; see Milani and Treadwell (2012), Christiano et al. (2014), and Forni et al. (2017), among others. We outline the model in the Online Appendix. There are seven exogenous shocks in the model, and all of them are assumed to have anticipated components. They are: 1) the stationary neutral productivity shock z_t, 2) the nonstationary neutral productivity shock X_t, 3) the stationary investment-specific productivity shock z_t^i, 4) the nonstationary investment-specific productivity shock A_t, 5) the government spending shock G_t, 6) the wage markup shock μ_t, 7) the preference shock ζ_t. The shocks X_t and A_t are made stationary using growth rates, with the respective variables being $\mu_t^x = X_t/X_{t-1}$ and $\mu_t^a = A_t/A_{t-1}$. G_t is detrended to form $g_t \equiv G_t/X_t^G$, where $X_t^G = (X_{t-1}^G)^{\rho_{xg}}(X_{t-1}A_{t-1}^{\alpha_K/(\alpha_K-1)})^{1-\rho_{xg}}$ is the trend in government spending. All seven processes x_t ($x = \{z, \mu^x, z^i, \mu^a, g, \mu, \zeta\}$) are assumed to follow $\ln(x_t/x) = \rho_x \ln(x_{t-1}/x) + \varepsilon_{x,t}$ with $\varepsilon_{x,t} = \varepsilon_{x,t}^0 + \varepsilon_{x,t-4}^4 + \varepsilon_{x,t-8}^8$, where x denote the steady state values of the variables and $\varepsilon_{x,t}^j$ are i.i.d. $N(0, \sigma_x^j)$. The total number of shocks is 21.

After log linearization, Schmitt-Grohé and Uribe (2012) estimate the model on seven demeaned observables: real GDP growth, real consumption growth, real investment growth, labor hours, real government spending growth, TFP growth, and relative price of investment growth; we use the same set of variables over the same sample period. The full vector of structural parameters is given by:

$$\begin{aligned}\phi \;=\; & [\theta, \gamma, \kappa, \delta_2, b, \rho_{xg}, \rho_{\mu^a}, \rho_{\mu^x}, \rho_{z^i}, \rho_z, \rho_\mu, \rho_g, \rho_\zeta, \sigma_{\mu^a}^0, \sigma_{\mu^a}^4, \sigma_{\mu^a}^8, \sigma_{\mu^x}^0, \sigma_{\mu^x}^4, \sigma_{\mu^x}^8, \sigma_{z^i}^0, \sigma_{z^i}^4, \sigma_{z^i}^8, \\ & \sigma_z^0, \sigma_z^4, \sigma_z^8, \sigma_\mu^0, \sigma_\mu^4, \sigma_\mu^8, \sigma_g^0, \sigma_g^4, \sigma_g^8, \sigma_\zeta^0, \sigma_\zeta^4, \sigma_\zeta^8, \sigma_{g^y}^{me}].\end{aligned}$$

Similar to the previous subsection, we use the two-step procedure to perform specification testing at the 10% level. The data sample used corresponds to that considered in Schmitt-Grohé and Uribe (2012) : 1955:Q2-2006:Q4.

Result 6 *The model is rejected at the business cycle frequencies and by the weighted full spectrum test at the 10% significance level using the original Schmitt-Grohé and Uribe (2012) sample. This is in contrast with the case of the medium scale model of Smets and Wouters (2007), where neither frequency range produced a rejection. Further examination reveals that the main source of incompatibility between the model and the data is the per capita labor hours and its comovements with all the other observables.*

Table 10 presents the specification test results for the weighted full spectrum case and the business cycle frequency case for the full model. The results show that 100% of the draws are rejected when the full model is considered. Testing individual observables reveals that only 2% of the draws survive for labor hours growth, while the other observables generate much lower rejection frequencies. Testing pairs shows that all the comovements between labor hours and other observables are at odds with the data, with less than 1% of the draws surviving for its comovements with GDP, consumption, and investment growth. These findings can be related to those of Schmitt-Grohé and Uribe (2012), whose Table III presented the model's predictions

Table 10: Specification test results for the SGU model, 1955:II-2006:IV

	Robust inference		
	Test	CV	Rej.
Full model, weighted spectrum	1.254	1.235	100
Full model, BC frequencies	1.646	1.519	100
GDP growth	0.231	0.977	8
Consumption growth	0.243	0.977	39
Investment growth	0.230	0.977	4
Labor hours growth	0.512	0.977	98
Government spending growth	0.246	0.977	18
TFP growth	0.299	0.977	71
Rel. price of investment growth	0.337	0.977	12
GDP-Consumption	0.672	1.077	52
GDP-Investment	0.753	1.077	51
GDP-Labor hours	1.105	1.077	100
GDP-Gov. spending	0.433	1.077	16
GDP-TFP	0.698	1.077	52
GDP-Rel. price of Investment	0.527	1.077	9
Consumption-Investment	0.515	1.077	26
Consumption-Labor hours	0.886	1.077	99.77
Consumption-Gov. spending	0.407	1.077	35
Consumption-TFP	0.521	1.077	25
Consumption-Rel. price of Investment	0.632	1.077	33
Investment-Labor hours	1.074	1.077	100
Investment-Gov. spending	0.425	1.077	11
Investment-TFP	0.780	1.077	56
Investment-Rel. price of Investment	0.391	1.077	8
Labor hours-Gov. spending	0.559	1.077	98.84
Labor hours-TFP	0.645	1.077	99
Labor hours-Rel. price of Investment	0.529	1.077	98
Gov. spending-TFP	0.577	1.077	54
Gov. spending-Rel. price of Investment	0.409	1.077	15
TFP-Rel. price of Investment	0.472	1.077	62

Note. All results use the weighted full spectrum unless indicated otherwise. Each row represents a set of variables tested. The significance level is at 10% for each case. Test: the specification test value; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test. The parameter values are obtained by sampling the score test.

regarding standard deviations, correlations with output growth, and serial correlations of the seven observables. Their results indicated that the model's predicted second moments were overall similar to empirical second moments, but notable discrepancies were found in the serial correlation of the growth rate of hours and, to a lesser extent, in the correlation of hours and output. Our results show that these discrepancies are large enough from a statistical testing perspective to cause a clear rejection of the model.

7 Conclusion

This paper introduced a framework for evaluating the specification of log linearized DSGE models. The approach involves proposing a set of plausible parameter values and testing them for compatibility with data, which can separately examine a model's steady-state properties, overall dynamic properties, and behavior in selected frequency bands. We presented three empirical applications to small and medium DSGE models. Our approach has general applicability and can be used to evaluate new DSGE models as they are proposed, which could prove valuable for researchers looking to assess the suitability of these models for various applications.

References

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

Andrews, I., and A. Mikusheva (2015): "Maximum Likelihood Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 6, 123–152.

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

Benati, L., and P. Surico (2009): "VAR Analysis and the Great Moderation," *American Economic Review*, 99, 1636–1652.

Benhabib, J., S. Schmitt-Grohé, and M. Uribe (2001): "Monetary Policy and Multiple Equilibria," *American Economic Review*, 91, 167–186.

Boivin, J., and M. P. Giannoni (2006): "Has Monetary Policy Become More Effective?," *The Review of Economics and Statistics*, 88, 445–462.

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

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

Christiano, L. J., R. Motto, and M. Rostagno (2014): "Risk Shocks," *American Economic Review*, 104, 27–65.

Clarida, R., J. Galí, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics*, 115, 147–180.

Cochrane, J. H. (2011): "Determinacy and Identification with Taylor Rules," *Journal of Political Economy*, 119, 565–615.

——— (2017): "The New-Keynesian Liquidity Trap," *Journal of Monetary Economics*, 92, 47–63.

Del Negro, M., and F. Schorfheide (2009): "Monetary Policy Analysis with Potentially Misspecified Models," *American Economic Review*, 99, 1415–1450.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): "On the Fit of New Keynesian Models," *Journal of Business & Economic Statistics*, 25, 123–143.

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

Forni, M., L. Gambetti, M. Lippi, and L. Sala (2017): "Noisy News in Business Cycles," *American Economic Journal: Macroeconomics*, 9, 122–152.

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

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): "Frequentist Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 4, 197–229.

Inoue, A., C.-H. Kuo, and B. Rossi (2020): "Identifying the Sources of Model Misspecification," *Journal of Monetary Economics*, 110, 1–18.

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

Leeper, E. M., N. Traum, and T. B. Walker (2017): "Clearing Up the Fiscal Multiplier Morass," *American Economic Review*, 107, 2409–2454.

Lubik, T., and F. Schorfheide (2003): "Computing Sunspot Equilibria in Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 28, 273–285.

——— (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94, 190–217.

Mavroeidis, S. (2010): "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence," *American Economic Review*, 100, 491–503.

Milani, F., and J. Treadwell (2012): "The Effects of Monetary Policy "News" and "Surprises"," *Journal of Money, Credit and Banking*, 44, 1667–1692.

Qu, Z. (2014): "Inference in Dynamic Stochastic General Equilibrium Models with Possible Weak Identification," *Quantitative Economics*, 5, 457–494.

Qu, Z., and D. Tkachenko (2017): "Global Identification in DSGE Models Allowing for Indeterminacy," *Review of Economic Studies*, 84, 1306–1345.

Schmitt-Grohé, S., and M. Uribe (2012): "What's News in Business Cycles," *Econometrica*, 80, 2733–2764.

Schorfheide, F. (2013): "Estimation and Evaluation of DSGE Models: Progress and Challenges," in *Advances in Economics and Econometrics*, ed. by D. Acemoglu, M. Arellano, and E. Deckel, pp. 184–230. Cambridge University Press.

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

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

Stock, J. H., and M. W. Watson (2005): "Implications of Dynamic Factor Models for VAR Analysis," NBER Working Papers 11467, National Bureau of Economic Research.

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

Supplementary Appendix: Proofs, Model Details, and Additional Tables and Figures

A.1 Proofs of results in the paper

Proof of Theorem 1. Consider $\mathcal{H}_{dT}(\theta_0)$. We first prove finite dimensional convergence and then verify tightness.[1] The term inside the norm can be equivalently represented as

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

with

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

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

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

where the last equality follows from

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

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

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

Further, because

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

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

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

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

Apply Assumption 2, we have

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

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

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

$$\max_{\omega_j\in[0,\pi]} \mathbb{E}\left|R_{kl}(\omega_j)\right|^2 = O(T^{-1}). \tag{A.3}$$

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

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

with

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

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

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

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

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

which implies

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

because of the Cauchy-Schwarz inequality. Thus,

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

where the first inequality is because $|R_{kl}(\omega_j)|$ are nonnegative and the second is due to the Chebyshev inequality. Apply (A.3), the numerator in the preceding display is of order O(T). Therefore, the whole term is of order O(1), which can be made small by choosing a large C. The above results imply the tightness.

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

Proof of Theorem 2. Consider $\mathcal{H}_{dT}(\theta_0)$. Let $\mathcal{H}_{dT}(\theta_0;r)$ denote $\mathcal{H}_{dT}(\theta_0)$ before taking the supremum, i.e.,

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

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

$$
\begin{aligned}
& (T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0; r) \\
= & \left\| \frac{2}{T} \sum_{j=1}^{[Tr/2]} \operatorname{vec}\left\{ f_{\theta_0}(\omega_j)^{-1/2} \left(f_0(\omega_j) - f_{\theta_0}(\omega_j) \right) f_{\theta_0}(\omega_j)^{-1/2} \right\} \right\|_\infty + o_p(1) \\
= & \left\| \frac{1}{\pi} \int_0^{\pi r} \operatorname{vec}\left(f_{\theta_0}(\omega)^{-1/2} \left(f_0(\omega) - f_{\theta_0}(\omega) \right) f_{\theta_0}(\omega)^{-1/2} \right) d\omega \right\|_\infty + o_p(1), \qquad \text{(A.4)}
\end{aligned}
$$

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

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

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

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

Consider (A.4) with $r = \omega_U/\pi$,

$$
\begin{aligned}
& (T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0; \omega_U/\pi) \\
= & \left\| \frac{1}{\pi} \int_0^{\omega_U} \operatorname{vec}\left(f_{\theta_0}(\omega)^{-1/2} \left(f_0(\omega) - f_{\theta_0}(\omega) \right) f_{\theta_0}(\omega)^{-1/2} \right) d\omega \right\|_\infty + o_p(1) \\
\geq & \left| \frac{1}{\pi} \int_0^{\omega_U} \zeta_{\theta_0}(\omega) d\omega \right| + o_p(1) \\
\geq & \frac{1}{\pi} \int_{\omega_L}^{\omega_U} \zeta_{\theta_0}(\omega) d\omega - \left| \frac{1}{\pi} \int_0^{\omega_L} \zeta_{\theta_0}(\omega) d\omega + \right| + o_p(1) \\
\geq & \frac{C\delta}{2\pi} (\omega_U - \omega_L) - \left| \frac{1}{\pi} \int_0^{\omega_L} \zeta_{\theta_0}(\omega) d\omega \right| + o_p(1) \\
\geq & \frac{C\delta}{4\pi} (\omega_U - \omega_L) - \left| \frac{1}{\pi} \int_0^{\omega_L} \zeta_{\theta_0}(\omega) d\omega \right|, \qquad \text{(A.5)}
\end{aligned}
$$

where the first inequality uses the definition of $\zeta_{\theta_0}(\omega)$ and the property of the supremum norm, the second uses the triangle inequality, the third is because $\zeta_{\theta_0}(\omega)$ is greater than C$\delta/2$ over the interval and the last is because $\frac{C\delta}{4\pi}(\omega_U - \omega_L)$ is positive thus dominating the $o_p(1)$ term. We now apply (A.5) to find a lower bound for $(T/2)^{-1/2}\mathcal{H}_{dT}(\theta_0)$. There are only two possibilities:

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

In Case 1,

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

where the equality uses the definition of $\mathcal{H}_{dT}(\theta_0)$, the first inequality is because the supremum norm must be no less than any of its admissible values and the second inequality is because (A.5) and the definition of Case 1. In Case 2,

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

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

Now consider the order of $\mathcal{H}_{sT}(\theta_0)$ under global alternatives.

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

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

A.2 Smets and Wouters (2007) model equations

The vector of observable variables includes output growth (Δy_t), consumption growth (Δc_t), investment growth (Δi_t), wage growth (Δw_t), labor hours (l_t), inflation (π_t) and the interest rate (r_t). As in Smets and Wouters (2007), five parameters are fixed as follows: $\epsilon_p = \epsilon_w = 10, \delta = 0.025, g_y = 0.18, \phi_w = 1.50$. The analysis allows the remaining 36 structural parameters to vary. They are ordered as

$$\begin{aligned} \theta^D = \; & (\rho_{ga}, \mu_w, \mu_p, \alpha, \psi, \varphi, \sigma_c, \lambda, \phi_p, \iota_w, \xi_w, \iota_p, \xi_p, \sigma_l, r_\pi, r_{\Delta y}, r_y, \rho, \rho_a, \rho_b, \rho_g, \\ & \rho_i, \rho_r, \rho_p, \rho_w, \sigma_a, \sigma_b, \sigma_g, \sigma_i, \sigma_r, \sigma_p, \sigma_w, \overline{\gamma}, 100(1/\beta - 1), \overline{\pi}, \overline{l})'. \end{aligned}$$

Below is the log linearized system consistent with Smets and Wouters' (2007) code.

The aggregate resource constraint: It satisfies

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

Output (y_t) is composed of consumption (c_t), investment (i_t), capital utilization costs as a function of the capital utilization rate (z_t), and exogenous spending (ε_t^g). The latter follows an AR(1) model with an i.i.d. Normal error term (η_t^g), and is also affected by the productivity shock (η_t^a) as follows:

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

The coefficients c_y, i_y and z_y are functions of the steady state spending-output ratio (g_y), steady state output growth (γ), capital depreciation (δ), household discount factor (β), intertemporal elasticity of substitution (σ_c), fixed costs in production (ϕ_p), and share of capital in production (α): $i_y = (\gamma - 1 + \delta)k_y$, $c_y = 1 - g_y - i_y$, and $z_y = R_*^k k_y$. Here, k_y is the steady state capital-output ratio, and R_*^k is the steady state rental rate of capital: $k_y = \phi_p \left(L_*/k_*\right)^{\alpha-1} = \phi_p \left[((1-\alpha)/\alpha) \left(R_*^k/w_*\right)\right]^{\alpha-1}$ with $w_* = \left(\alpha^\alpha (1-\alpha)^{(1-\alpha)}/[\phi_p \left(R_*^k\right)^\alpha]\right)^{1/(1-\alpha)}$, and $R_*^k = \beta^{-1}\gamma^{\sigma_c} - (1-\delta)$.

Households: The consumption Euler equation is

$$c_t = c_1 c_{t-1} + (1-c_1)E_t c_{t+1} + c_2(l_t - E_t l_{t+1}) - c_3(r_t - E_t \pi_{t+1}) - \varepsilon_t^b. \tag{A.6}$$

where l_t is hours worked, r_t is the nominal interest rate, and π_t is inflation. The disturbance ε_t^b follows

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

The relationships of the coefficients in (A.6) to the habit persistence (λ), steady state labor market mark-up (ϕ_w), and other structural parameters highlighted above are

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

where

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

where R_*^k and k_y are defined as above, and $c_y = 1 - g_y - (\gamma - 1 + \delta)k_y$.

The dynamics of households' investment are given by

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

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

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

The coefficients satisfy

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

where φ is the steady state elasticity of the capital adjustment cost function. The corresponding arbitrage equation for the value of capital is given by

$$q_t = q_1 E_t q_{t+1} + (1-q_1) E_t r_{t+1}^k - (r_t - E_t \pi_{t+1}) - \frac{1}{c_3}\varepsilon_t^b\ , \tag{A.7}$$

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

Final and intermediate goods market: The aggregate production function is

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

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

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

The current capital service usage (k_t^s) is a function of capital installed in the previous period (k_{t-1}) and the degree of capital utilization (z_t):

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

Furthermore, the capital utilization is a positive fraction of the rental rate of capital (r_t^k):

$$z_t = z_1 r_t^k, \quad \text{where } z_1 = (1-\psi)/\psi,$$

and ψ is a positive function of the elasticity of the capital utilization adjustment cost function and normalized to be between zero and one. The accumulation of installed capital (k_t) satisfies

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

where ε_t^i is the investment specific technology process as defined before, and k_1 and k_2 satisfy

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

The price mark-up satisfies

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

where w_t is the real wage. The New Keynesian Phillips curve is

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

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

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

The MA(1) term is intended to pick up some of the high frequency fluctuations in prices. The Phillips curve coefficients depend on price indexation (ι_p) and stickiness (ξ_p), the curvature of the goods market Kimball aggregator (ϵ_p), and other structural parameters:

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

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

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

Labor market: The wage mark-up is

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

where σ_l is the elasticity of labor supply. Real wage w_t adjusts slowly according to

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

where the coefficients are functions of wage indexation (ι_w) and stickiness (ξ_w) parameters, and the curvature of the labor market Kimball aggregator (ϵ_w):

$$\begin{aligned} w_1 &= \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}, w_2 = \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}, w_3 = \frac{\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}, \\ w_4 &= \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} \frac{\left(1-\beta\gamma^{(1-\sigma_c)}\xi_w\right)(1-\xi_w)}{\xi_w\left((\phi_w - 1)\,\epsilon_w + 1\right)}. \end{aligned}$$

The wage mark-up disturbance follows an ARMA(1,1) process:

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

Monetary policy: The empirical monetary policy reaction function is

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

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

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

The variable y_t^* stands for the time-varying optimal output level that is the result of a flexible price-wage economy. Since the equations for the flexible price-wage economy are essentially the same as above, but with the variables μ_t^p and μ_t^w set to zero, we omit the details.

A.3 Outline of the Schmitt-Grohé and Uribe (2012) model

The economy is populated with agents maximizing lifetime utility $E_0 \sum_{t=0}^{\infty} \beta^t \zeta_t U(V_t)$, where ζ_t is an exogenous preference shock, and $U(V_t) = (V_t^{1-\sigma} - 1)/(1-\sigma)$ with $V_t = C_t - bC_{t-1} - \psi h_t^\theta S_t$, where $S_t = (C_t - bC_{t-1})^\gamma S_{t-1}^{1-\gamma}$ so that consumer preferences are defined over V_t which represents a bundle of consumption (C_t), labor (h_t) and an additional variable S_t. Jaimovich and Rebelo (2009) found that this form of preferences, together with other real rigidities, is key for generating aggregate comovement in response to news about fundamental shocks. Households own physical capital stock K_t, which evolves according to $K_t = (1-\delta(u_t))K_{t-1} + z_t^I I_t \left[1 - S\left(I_t/I_{t-1}\right)\right]$, where I_t is gross investment and u_t measures capacity utilization, so that the effective amount of capital supplied to firms is $u_t K_{t-1}$. The depreciation rate $\delta(u_t)$ satisfies $\delta(u_t) = \delta_0 + \delta_1(u_t - 1) + (\delta_2/2)(u_t - 1)^2$. The investment adjustment cost function $S(\cdot)$, due to Christiano et al. (2005), is given by $S(x) = (\kappa/2)(x - \mu^i)^2$, where μ^i is the steady state growth rate of investment. Finally, the stationary exogenous shock z_t^I affects the technology transforming investment goods into capital goods.

The production function is of Cobb-Douglas form:

$$Y_t = z_t(u_t K_{t-1})^{\alpha_k}(X_t h_t)^{\alpha_h}(X_t L)^{1-\alpha_k-\alpha_h}, \tag{A.8}$$

where Y_t is output, z_t is an exogenous productivity shock, X_t is a nonstationary labor-augmenting productivity shock, and L is a fixed factor of production. The capital and labor shares satisfy $\alpha_k, \alpha_h \in (0,1), \alpha_k + \alpha_h \leq 1$. The aggregate resource constraint is given by $Y_t = C_t + A_t I_t + G_t$, where G_t is government spending and A_t is a nonstationary shock to investment-specific technology.

The model features an imperfectly competitive labor market. The households supply labor to monopolistically competitive labor unions, which sell differentiated labor inputs to the final good producers. The elasticity of substitution between differentiated labor inputs is time-varying, with the wage markup denoted μ_t. In equilibrium, the wage rate paid by the union to its members is smaller than the wage rate firms pay to unions, and all unions charge the same wage rate.

A.4 Tables and figures

Table A1: Rejection frequencies under the null hypothesis (pairwise testing)

Level	T	BC frequencies			Full spectrum			Weighted spectrum			Mean and spectrum		
		Determinacy											
		(π,y)	(π,r)	(y,r)	(π,y)	(π,r)	(y,r)	(π,y)	(π,r)	(y,r)	(π,y)	(π,r)	(y,r)
10%	80	0.079	0.073	0.086	0.079	0.081	0.081	0.081	0.090	0.087	0.084	0.094	0.083
	160	0.085	0.079	0.092	0.087	0.082	0.090	0.082	0.089	0.088	0.081	0.091	0.083
	240	0.085	0.090	0.088	0.078	0.087	0.085	0.086	0.090	0.087	0.084	0.092	0.083
	320	0.092	0.094	0.082	0.093	0.092	0.085	0.090	0.095	0.093	0.089	0.094	0.093
5%	80	0.044	0.044	0.052	0.038	0.042	0.041	0.052	0.058	0.056	0.041	0.046	0.042
	160	0.047	0.041	0.055	0.046	0.042	0.044	0.046	0.053	0.051	0.041	0.049	0.042
	240	0.044	0.045	0.047	0.039	0.042	0.043	0.047	0.052	0.046	0.043	0.042	0.038
	320	0.044	0.050	0.044	0.048	0.044	0.041	0.045	0.050	0.047	0.049	0.048	0.042
		Indeterminacy											
10%	80	0.091	0.106	0.087	0.082	0.098	0.079	0.090	0.115	0.090	0.072	0.106	0.081
	160	0.086	0.098	0.089	0.089	0.093	0.090	0.085	0.096	0.088	0.083	0.115	0.079
	240	0.092	0.091	0.091	0.090	0.095	0.097	0.093	0.096	0.091	0.077	0.100	0.082
	320	0.097	0.098	0.086	0.092	0.095	0.094	0.078	0.097	0.091	0.086	0.102	0.086
5%	80	0.058	0.065	0.054	0.045	0.053	0.041	0.057	0.080	0.055	0.035	0.061	0.045
	160	0.047	0.056	0.051	0.043	0.052	0.047	0.052	0.057	0.051	0.043	0.061	0.040
	240	0.049	0.050	0.050	0.047	0.044	0.048	0.054	0.056	0.051	0.037	0.048	0.038
	320	0.049	0.055	0.045	0.044	0.045	0.046	0.041	0.051	0.051	0.041	0.053	0.045

Note. T: sample size; all tests are computed with prewhitening. π, y, and r denote inflation, output, and interest rate, respectively.

Table A2: Rejection frequencies under the null hypothesis (single variable testing)

Level	T	BC frequencies			Full spectrum			Weighted spectrum			Mean and spectrum		
		Determinacy											
		π	y	r	π	y	r	π	y	r	π	y	r
10%	80	0.068	0.066	0.068	0.078	0.080	0.084	0.066	0.067	0.072	0.096	0.074	0.066
	160	0.080	0.081	0.080	0.090	0.081	0.081	0.079	0.074	0.078	0.093	0.077	0.080
	240	0.083	0.081	0.092	0.096	0.091	0.094	0.081	0.087	0.082	0.086	0.085	0.085
	320	0.087	0.086	0.092	0.091	0.096	0.093	0.083	0.080	0.086	0.096	0.081	0.086
5%	80	0.035	0.039	0.038	0.042	0.038	0.039	0.037	0.038	0.044	0.048	0.039	0.034
	160	0.032	0.045	0.040	0.042	0.044	0.037	0.042	0.039	0.041	0.049	0.037	0.042
	240	0.045	0.040	0.050	0.047	0.044	0.046	0.042	0.050	0.043	0.046	0.042	0.047
	320	0.046	0.041	0.047	0.043	0.046	0.047	0.043	0.039	0.041	0.051	0.040	0.042
		Indeterminacy											
10%	80	0.074	0.066	0.074	0.091	0.080	0.088	0.079	0.069	0.065	0.075	0.083	0.067
	160	0.085	0.081	0.079	0.085	0.087	0.093	0.080	0.075	0.080	0.076	0.086	0.073
	240	0.082	0.086	0.091	0.094	0.092	0.098	0.091	0.080	0.086	0.081	0.089	0.088
	320	0.086	0.085	0.085	0.087	0.089	0.098	0.089	0.094	0.083	0.088	0.089	0.079
5%	80	0.042	0.035	0.043	0.044	0.038	0.045	0.047	0.037	0.039	0.037	0.046	0.036
	160	0.048	0.041	0.041	0.042	0.042	0.047	0.043	0.041	0.044	0.037	0.046	0.037
	240	0.040	0.042	0.049	0.051	0.044	0.050	0.044	0.041	0.047	0.039	0.041	0.046
	320	0.039	0.043	0.046	0.043	0.046	0.050	0.043	0.050	0.043	0.044	0.043	0.039

Note. T: sample size; all tests are computed with pre-whitening. π, y, and r denote inflation, output, and interest rate, respectively.

Table A3: Rejection frequencies under the alternative hypothesis (pairwise testing; 10%)

T	BC frequencies			Full spectrum			Weighted spectrum			Mean and spectrum		
	(π, y)	(π, r)	(y, r)	(π, y)	(π, r)	(y, r)	(π, y)	(π, r)	(y, r)	(π, y)	(π, r)	(y, r)
Determinacy												
Perturb a random element of θ by 20%												
80	0.184	0.218	0.212	0.353	0.346	0.283	0.213	0.235	0.219	0.281	0.268	0.253
160	0.256	0.269	0.252	0.492	0.469	0.381	0.304	0.307	0.302	0.427	0.380	0.327
240	0.331	0.330	0.316	0.571	0.550	0.484	0.427	0.372	0.369	0.505	0.444	0.410
320	0.361	0.365	0.344	0.631	0.599	0.498	0.474	0.419	0.400	0.563	0.510	0.453
Perturb a random element of θ by 40%												
80	0.255	0.283	0.276	0.600	0.553	0.456	0.319	0.316	0.318	0.512	0.466	0.396
160	0.445	0.411	0.409	0.737	0.679	0.576	0.531	0.465	0.473	0.658	0.596	0.528
240	0.555	0.494	0.481	0.800	0.724	0.626	0.646	0.561	0.504	0.704	0.644	0.582
320	0.617	0.555	0.500	0.834	0.759	0.658	0.700	0.605	0.549	0.742	0.675	0.592
Indeterminacy												
Perturb a random element of θ by 20%												
80	0.165	0.212	0.184	0.226	0.284	0.296	0.185	0.209	0.228	0.227	0.248	0.290
160	0.213	0.264	0.258	0.343	0.368	0.406	0.230	0.276	0.308	0.299	0.330	0.373
240	0.220	0.296	0.283	0.401	0.415	0.463	0.292	0.314	0.365	0.332	0.369	0.416
320	0.276	0.305	0.331	0.459	0.443	0.496	0.323	0.350	0.412	0.406	0.403	0.450
Perturb a random element of θ by 40%												
80	0.267	0.287	0.299	0.439	0.421	0.479	0.284	0.332	0.333	0.416	0.380	0.443
160	0.361	0.371	0.400	0.572	0.536	0.580	0.416	0.409	0.478	0.511	0.472	0.534
240	0.433	0.435	0.494	0.630	0.608	0.641	0.518	0.452	0.567	0.581	0.524	0.589
320	0.514	0.460	0.539	0.685	0.629	0.685	0.578	0.528	0.602	0.639	0.577	0.609

Note. T: sample size; all tests are computed with prewhitening. π, y, and r denote inflation, output, and interest rate, respectively.

Table A4: Rejection frequencies under the alternative hypothesis (single variable testing; 10%)

T	BC frequencies			Full spectrum			Weighted spectrum			Mean and spectrum		
	π	y	r	π	y	r	π	y	r	π	y	r
Determinacy												
Perturb a random element of θ by 20%												
80	0.204	0.121	0.235	0.334	0.196	0.268	0.224	0.162	0.240	0.296	0.180	0.247
160	0.253	0.167	0.271	0.432	0.271	0.333	0.308	0.209	0.305	0.358	0.223	0.303
240	0.318	0.182	0.305	0.506	0.327	0.378	0.370	0.235	0.309	0.432	0.285	0.347
320	0.349	0.191	0.319	0.550	0.343	0.430	0.409	0.252	0.338	0.485	0.309	0.372
Perturb a random element of θ by 40%												
80	0.282	0.164	0.272	0.547	0.283	0.372	0.355	0.174	0.305	0.456	0.244	0.316
160	0.398	0.222	0.333	0.647	0.381	0.443	0.479	0.268	0.375	0.566	0.334	0.401
240	0.467	0.247	0.381	0.703	0.423	0.507	0.553	0.287	0.416	0.623	0.391	0.449
320	0.520	0.291	0.398	0.747	0.462	0.518	0.614	0.323	0.458	0.654	0.391	0.467
Indeterminacy												
Perturb a random element of θ by 20%												
80	0.135	0.136	0.166	0.178	0.181	0.259	0.155	0.158	0.207	0.188	0.174	0.244
160	0.163	0.165	0.219	0.239	0.241	0.340	0.199	0.203	0.273	0.227	0.221	0.297
240	0.197	0.167	0.254	0.302	0.272	0.368	0.231	0.227	0.310	0.271	0.252	0.337
320	0.219	0.217	0.295	0.336	0.309	0.417	0.266	0.265	0.355	0.300	0.265	0.358
Perturb a random element of θ by 40%												
80	0.225	0.192	0.270	0.337	0.274	0.379	0.238	0.237	0.322	0.321	0.245	0.360
160	0.289	0.244	0.350	0.450	0.369	0.443	0.347	0.305	0.398	0.400	0.324	0.425
240	0.337	0.281	0.398	0.488	0.403	0.526	0.380	0.338	0.432	0.447	0.364	0.467
320	0.369	0.303	0.416	0.534	0.440	0.546	0.429	0.394	0.445	0.484	0.403	0.515

Note. T: sample size; all tests are computed with prewhitening. π, y, and r denote inflation, output, and interest rate, respectively.

Table A5: Test results for the LS model using posterior draws, 1960:I-2007:IV, determinacy.

	Weighted spectrum			BC frequencies			Mean and Spectrum		
	Test	CV	Rej.	Test	CV	Rej.	Test	CV	Rej.
Full model	1.214	1.128	100	1.629	1.370	100	2.306	2.851	99
GDP	0.875	0.977	99.98	0.442	1.151	50	2.068	2.478	73
Inflation	0.641	0.977	62	0.554	1.151	56	1.180	2.478	43
Interest rate	0.337	0.977	4	0.369	1.151	0.01	0.819	2.478	15
GDP-Inflation	0.963	1.077	99.91	0.927	1.295	75	1.988	2.729	58
GDP-Interest rate	0.902	1.077	99.92	0.957	1.295	98	2.132	2.729	79
Inflation-Interest rate	0.832	1.077	94	0.809	1.295	36	1.987	2.729	95

Note. The significance level is 10%. Test: the spcification test value; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test.

Table A6: Test results for the LS model using posterior draws, 1960:I-2007:IV, indeterminacy

	Weighted spectrum			BC frequencies			Mean and Spectrum		
	Test	CV	Rej.	Test	CV	Rej.	Test	CV	Rej.
Full model	1.694	1.128	100	2.073	1.370	100	3.448	2.851	100
GDP	0.359	0.977	90	0.280	1.151	83	0.875	2.478	18
Inflation	0.592	0.977	96	0.431	1.151	81	1.499	2.478	83
Interest rate	0.308	0.977	28	0.382	1.151	14	0.907	2.478	18
GDP-Inflation	0.909	1.077	99.61	0.953	1.295	95	1.854	2.729	79
GDP-Interest rate	1.439	1.077	100	1.675	1.295	100	3.051	2.729	100
Inflation-Interest rate	0.859	1.077	98	0.751	1.295	88	1.891	2.729	87

Note. The significance level is 10%. Test: the spcification test value; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test.

Table A7: 95% confidence intervals for the LS model, 1979-2007, determinacy.

θ	Parameter	Bounds	CI
τ	intertemporal substitution elasticity	[0.10, 1.00]	[0.10, 0.996]
β	discount factor	[0.98, 0.999]	[0.980, 0.999]
κ	Phillips curve slope	[0.01, 2.00]	[0.02, 1.999]
ψ_1	inflation target	[1.01, 3.00]	[1.010, 2.999]
ψ_2	output target	[0.01, 5.00]	[0.010, 4.999]
ρ_r	interest rate smoothing	[0.10, 0.90]	[0.62, 0.90]
ρ_g	exogenous spending AR	[0.10, 0.98]	[0.70, 0.98]
ρ_z	technology shock AR	[0.10, 0.98]	[0.85, 0.98]
σ_r	monetary policy shock SD	[0.01, 3.00]	[0.22, 0.51]
σ_g	exogenous spending SD	[0.01, 3.00]	[0.03, 0.38]
σ_z	technology shock SD	[0.01, 3.00]	[0.46, 1.47]
ρ_{gz}	exogenous spending-technology CORR	[-0.90, -0.90]	[-0.46, 0.90]
π^*	steady state inflation	[2.00, 8.00]	[2.00, 8.00]

Note. The results are obtained using the mean and the full spectrum.

Table A8: Specification test results, 1979:III-2007:IV, determinacy.

	Weighted spectrum			BC frequencies			Mean and Spectrum		
	Test	CV	Rej.	Test	CV	Rej.	Test	CV	Rej.
Full model	0.682	1.128	15	0.769	1.370	39	2.519	2.851	99.93
GDP	0.468	0.977	7	0.306	1.151	0.03	2.180	2.478	99.99
Inflation	0.500	0.977	97	0.381	1.151	92	2.021	2.478	84
Interest rate	0.286	0.977	54	0.350	1.151	82	0.596	2.478	60
GDP-Inflation	0.597	1.077	86	0.424	1.295	76	2.356	2.729	99.82
GDP-Interest rate	0.581	1.077	45	0.639	1.295	72	2.612	2.729	99.99
Inflation-Interest rate	0.558	1.077	42	0.640	1.295	70	2.027	2.729	92

Note. The significance level is 10%. Test: the spcification test value; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test.

Table A9: Test results for the LS model using posterior draws, 1979:III-2007:IV, determinacy.

	Weighted spectrum			BC frequencies			Mean and Spectrum		
	Test	CV	Rej.	Test	CV	Rej.	Test	CV	Rej.
Full model	0.794	1.128	86	0.809	1.370	12	1.866	2.851	62
GDP	0.563	0.977	89	0.336	1.151	7	1.432	2.478	29
Inflation	0.401	0.977	16	0.337	1.151	10	0.729	2.478	32
Interest rate	0.279	0.977	2	0.339	1.151	1.5	0.556	2.478	2
GDP-Inflation	0.572	1.077	78	0.348	1.295	5	1.290	2.729	12
GDP-Interest rate	0.603	1.077	78	0.634	1.295	2	1.558	2.729	31
Inflation-Interest rate	0.537	1.077	26	0.354	1.295	2	1.515	2.729	47

Note. The significance level is 10%. Test: the spcification test value; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test.

Table A10: 95% confidence intervals for the LS model, indeterminacy, 1960-1979.

θ	Parameter	Bounds	CI
τ	intertemporal substitution elasticity	[0.10, 1.00]	[0.10, 0.9999]
β	discount factor	[0.90, 0.999]	[0.984, 0.999]
κ	Phillips curve slope	[0.01, 2.00]	[0.01, 1.01]
ψ_1	inflation target	[0.01, 0.99]	[0.01, 0.989]
ψ_2	output target	[0.01, 5.00]	[0.01, 5.00]
ρ_r	interest rate smoothing	[0.10, 0.90]	[0.10, 0.90]
ρ_g	exogenous spending AR	[0.10, 0.98]	[0.26, 0.98]
ρ_z	technology shock AR	[0.10, 0.98]	[0.52, 0.98]
σ_r	monetary policy shock SD	[0.01, 3.00]	[0.01, 0.75]
σ_g	exogenous spending SD	[0.01, 3.00]	[0.03, 1.23]
σ_z	technology shock SD	[0.01, 3.00]	[0.42, 2.99]
ρ_{gz}	exogenous spending-technology CORR	[-0.90, -0.90]	[-0.90, 0.90]
$M_{r\epsilon}$	sunspot-monetary coeff	[-3.00, 3.00]	[-3.00, 3.00]
$M_{g\epsilon}$	sunspot-exogenous spending coeff	[-3.00, 3.00]	[-3.000, 2.999]
$M_{z\epsilon}$	sunspot-technology coeff	[-3.00, 3.00]	[-3.00, 1.66]
σ_ϵ	sunspot shock SD	[0.01, 3.00]	[0.01, 1.68]
π^*	steady state inflation	[2.00, 8.00]	[2.00, 8.00]

Note. The results are obtained using the mean and the full spectrum.

Table A11: Test results for the LS model, 1960:I-1979:II, indeterminacy.

	Weighted spectrum			BC frequencies			Mean and Spectrum		
	Test	CV	Rej.	Test	CV	Rej.	Test	CV	Rej.
Full model	0.716	1.128	55	0.712	1.370	44	1.725	2.851	54
GDP	0.306	0.977	46	0.210	1.151	47	0.786	2.478	18
Inflation	0.440	0.977	61	0.315	1.151	11	1.063	2.478	94
Interest rate	0.315	0.977	17	0.190	1.151	2	1.071	2.478	93
GDP-Inflation	0.749	1.077	64	0.746	1.295	49	1.471	2.729	84
GDP-Interest rate	0.581	1.077	57	0.644	1.295	55	1.536	2.729	91
Inflation-Interest rate	0.650	1.077	9	0.682	1.295	1	1.458	2.729	66

Note. The significance level is 10%. Test: the spcification test value; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test.

Table A12: Test results for the LS model using posterior draws, 1960:I-1979:II, indeterminacy.

	Weighted spectrum			BC frequencies			Mean and Spectrum		
	Test	CV	Rej.	Test	CV	Rej.	Test	CV	Rej.
Full model	1.085	1.128	99.99	1.127	1.370	99	2.317	2.851	99
GDP	0.303	0.977	60	0.231	1.151	48	0.791	2.478	5
Inflation	0.239	0.977	4	0.255	1.151	4	0.420	2.478	3
Interest rate	0.277	0.977	61	0.193	1.151	73	0.971	2.478	31
GDP-Inflation	0.466	1.077	65	0.639	1.295	71	0.961	2.729	12
GDP-Interest rate	0.714	1.077	99.66	1.038	1.295	99.68	1.496	2.729	81
Inflation-Interest rate	0.566	1.077	86	0.748	1.295	91	1.209	2.729	63

Note. The significance level is 10%. Test: the spcification test value; CV: critical value; Rej.: percentage of parameter draws rejected by the specification test.

Figure A1: Log spectra of the LS model under indeterminacy, 1960-2007.

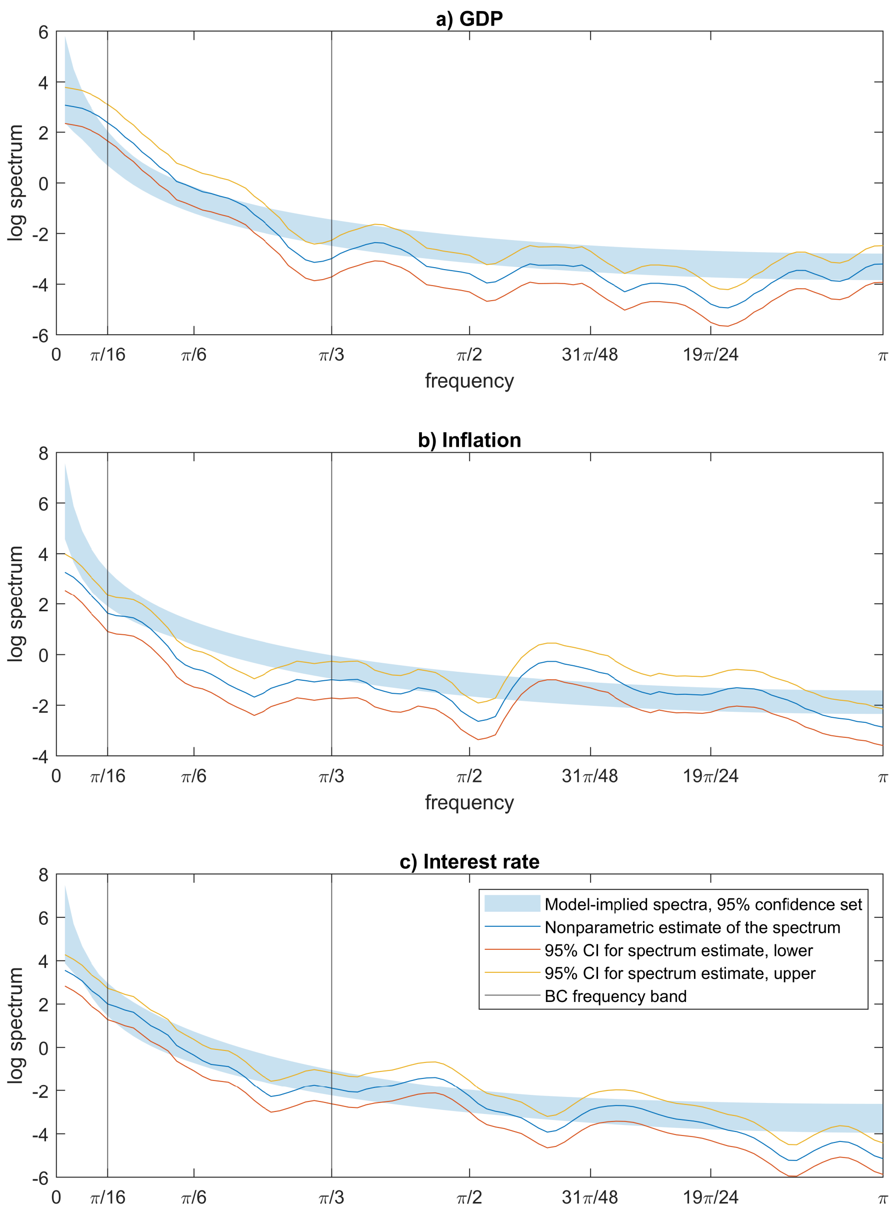


Figure A2: Cross spectra (real part) under indeterminacy, 1960-2007.

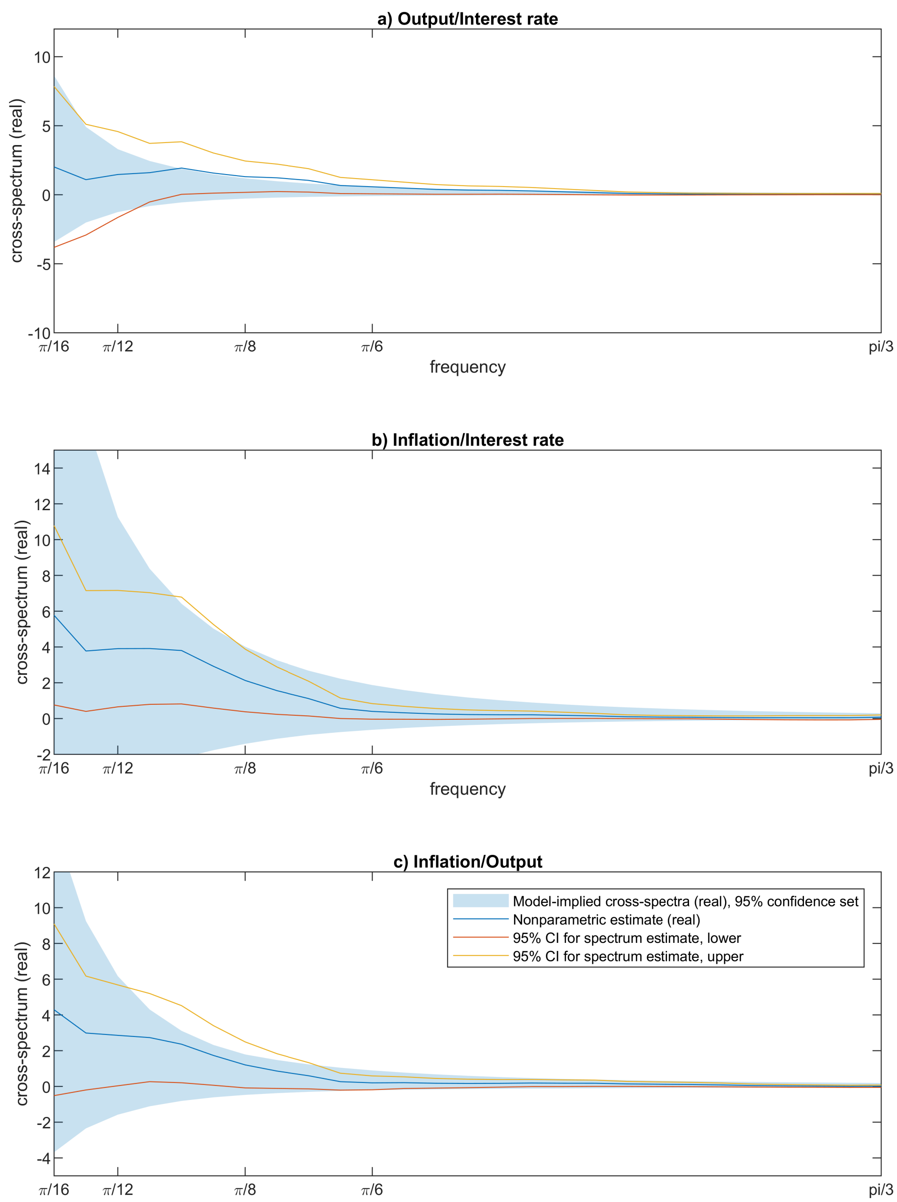


Figure A3: Cross spectra (imaginary part) under indeterminacy, 1960-2007.

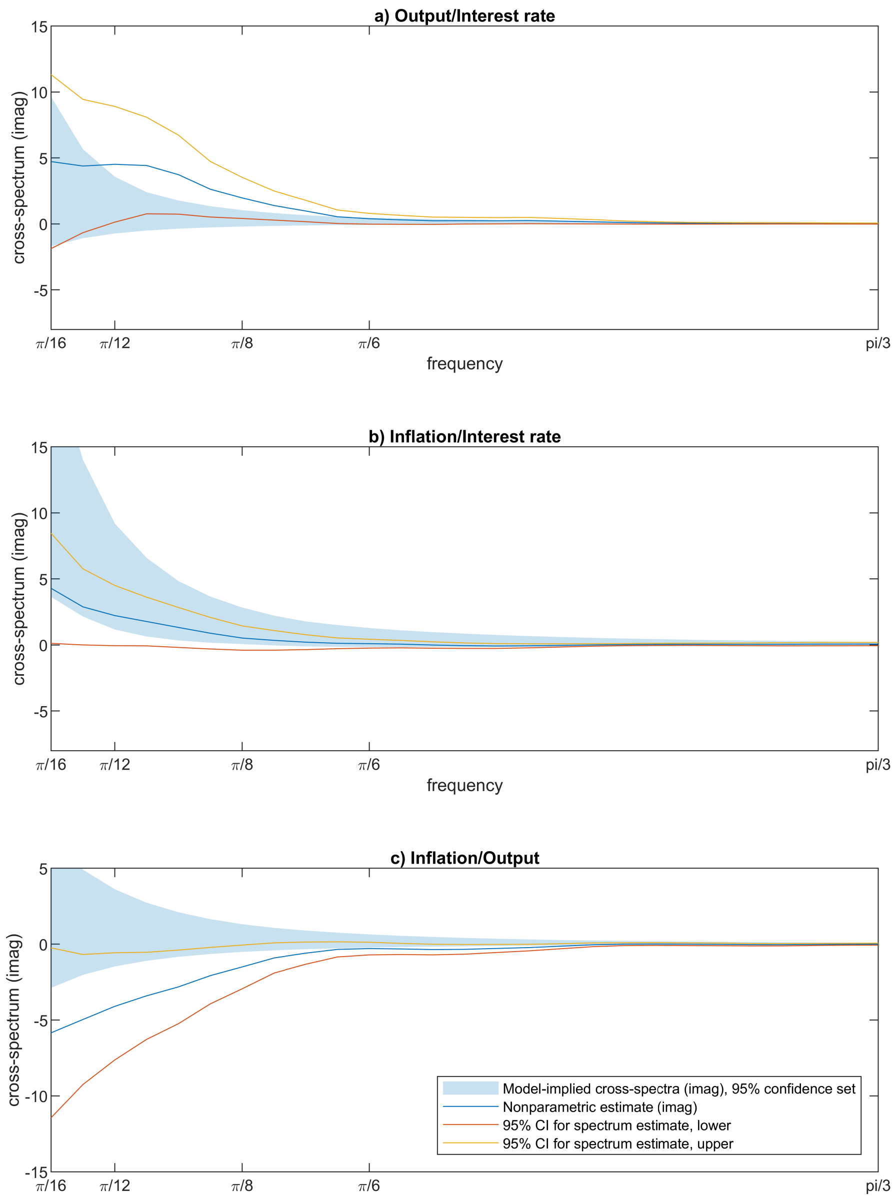


Figure A4: Log spectra using the posterior distribution under determinacy, 1960-2007.

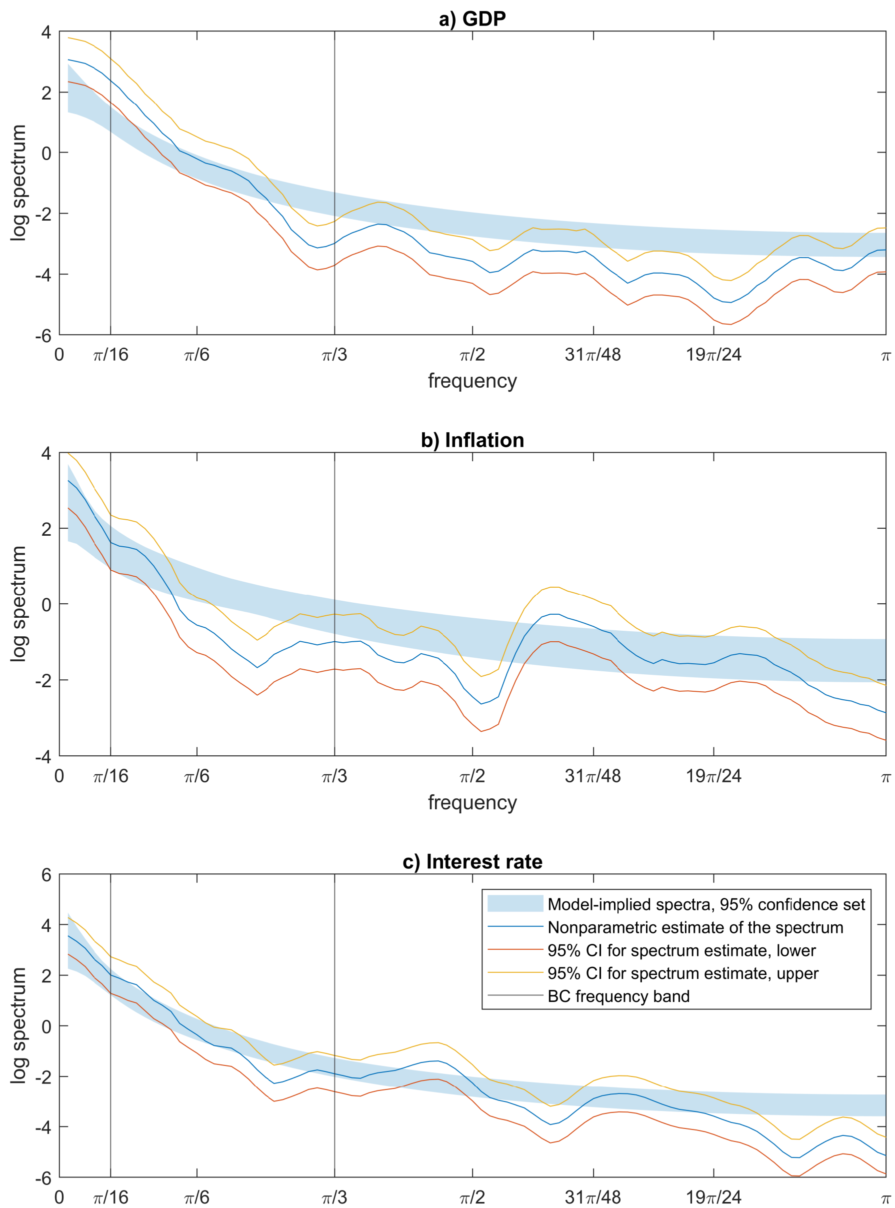


Figure A5: Cross spectra (real part) using the posterior under determinacy, 1960-2007.

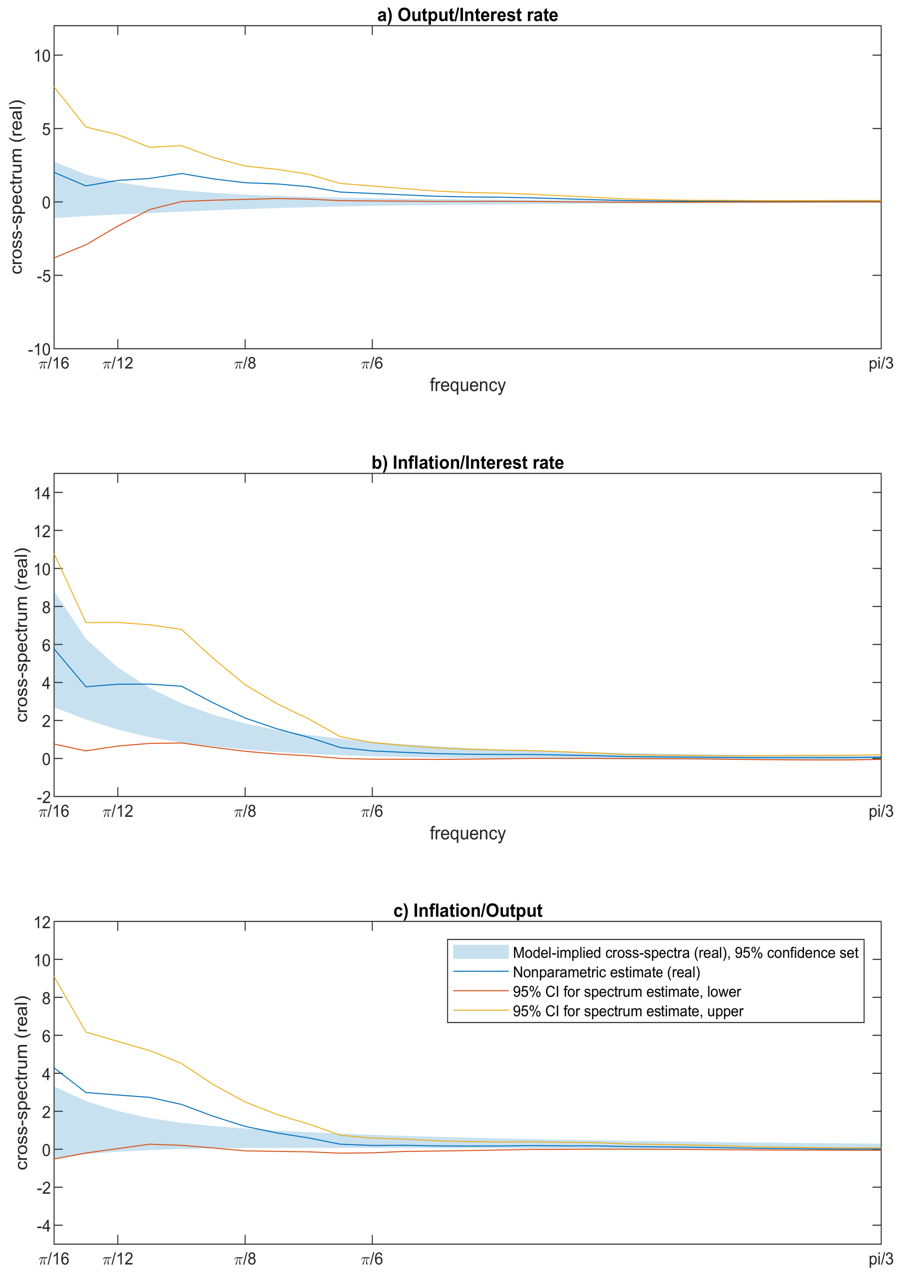


Figure A6: Cross spectra (imaginary part) using the posterior under determinacy, 1960-2007.

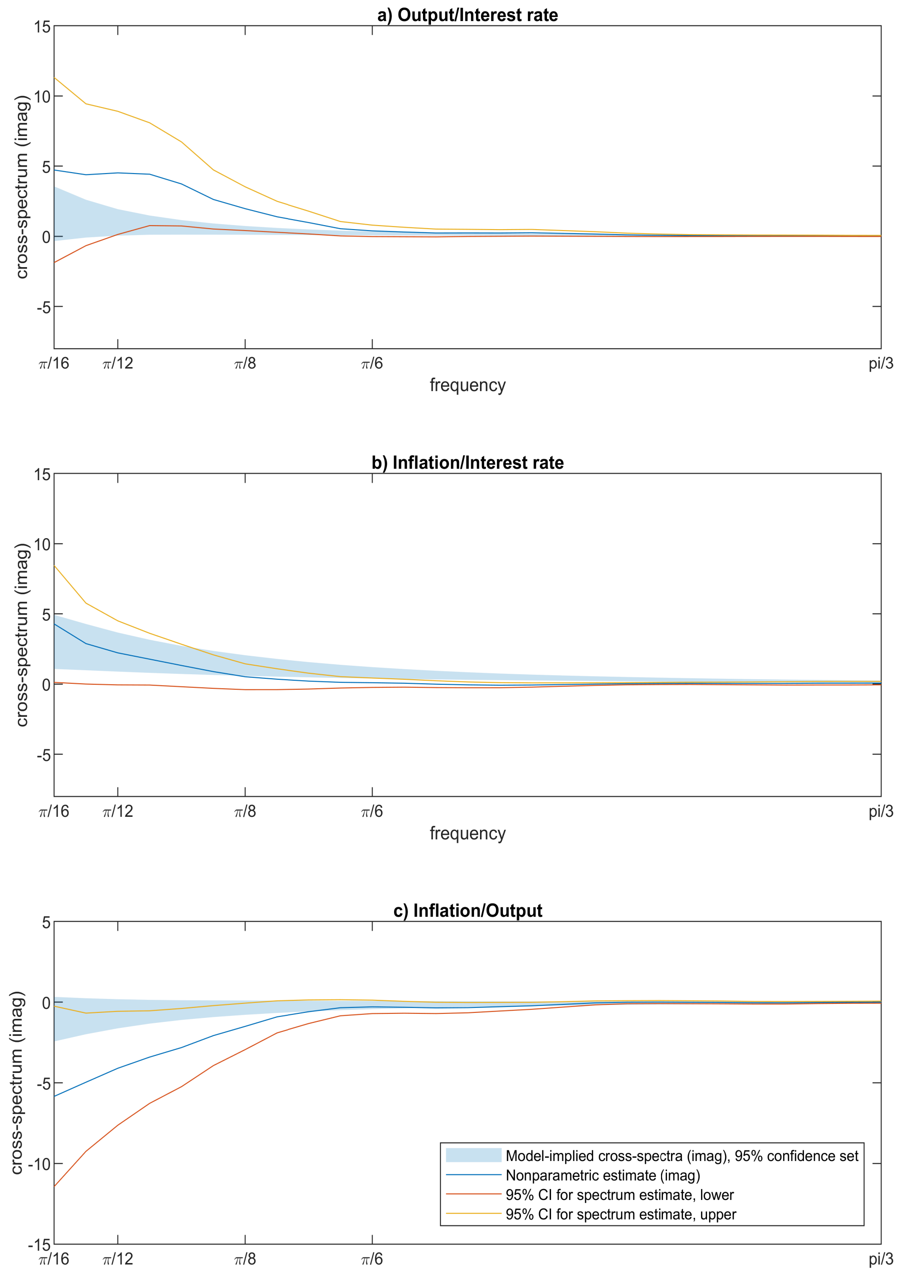


Figure A7: Log spectra using the posterior distribution under indeterminacy, 1960-2007.

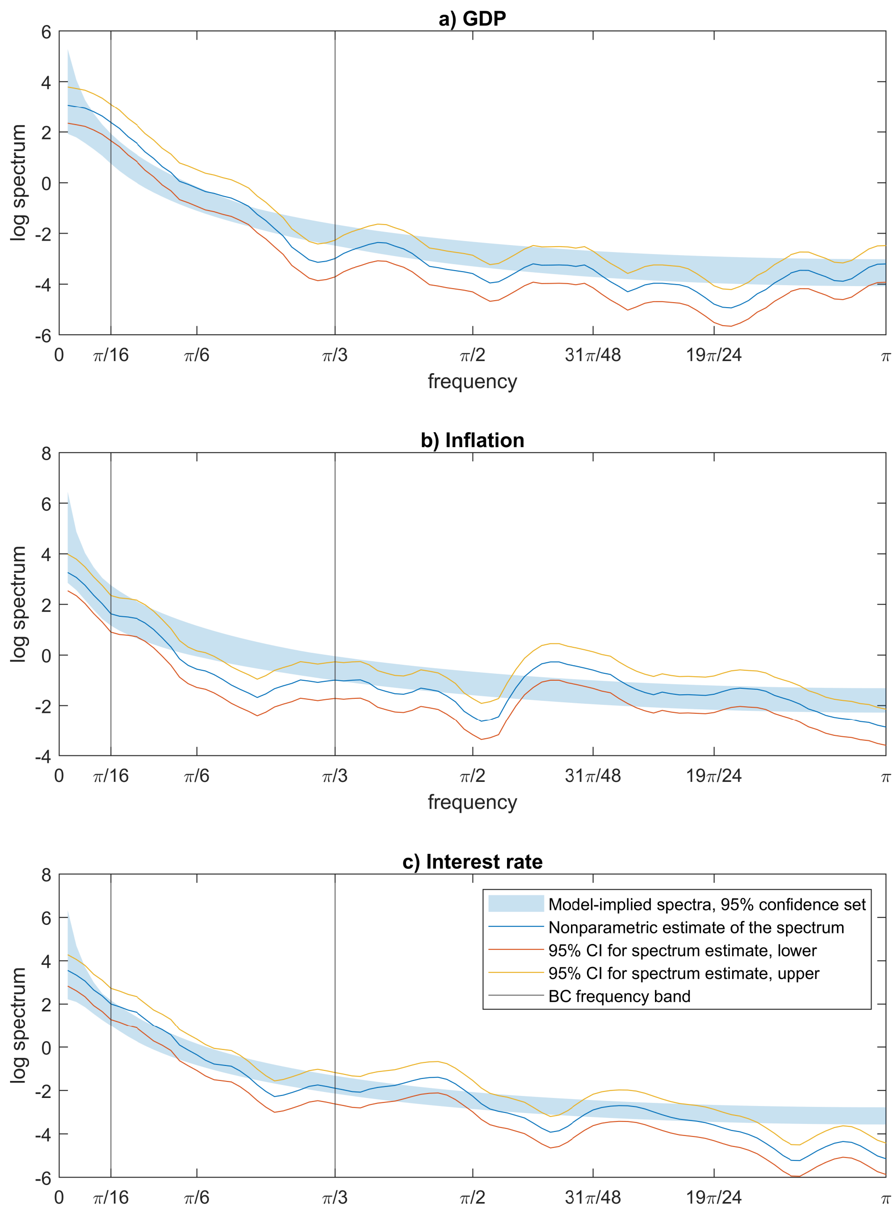


Figure A8: Cross spectra (real part) using the posterior under indeterminacy, 1960-2007.

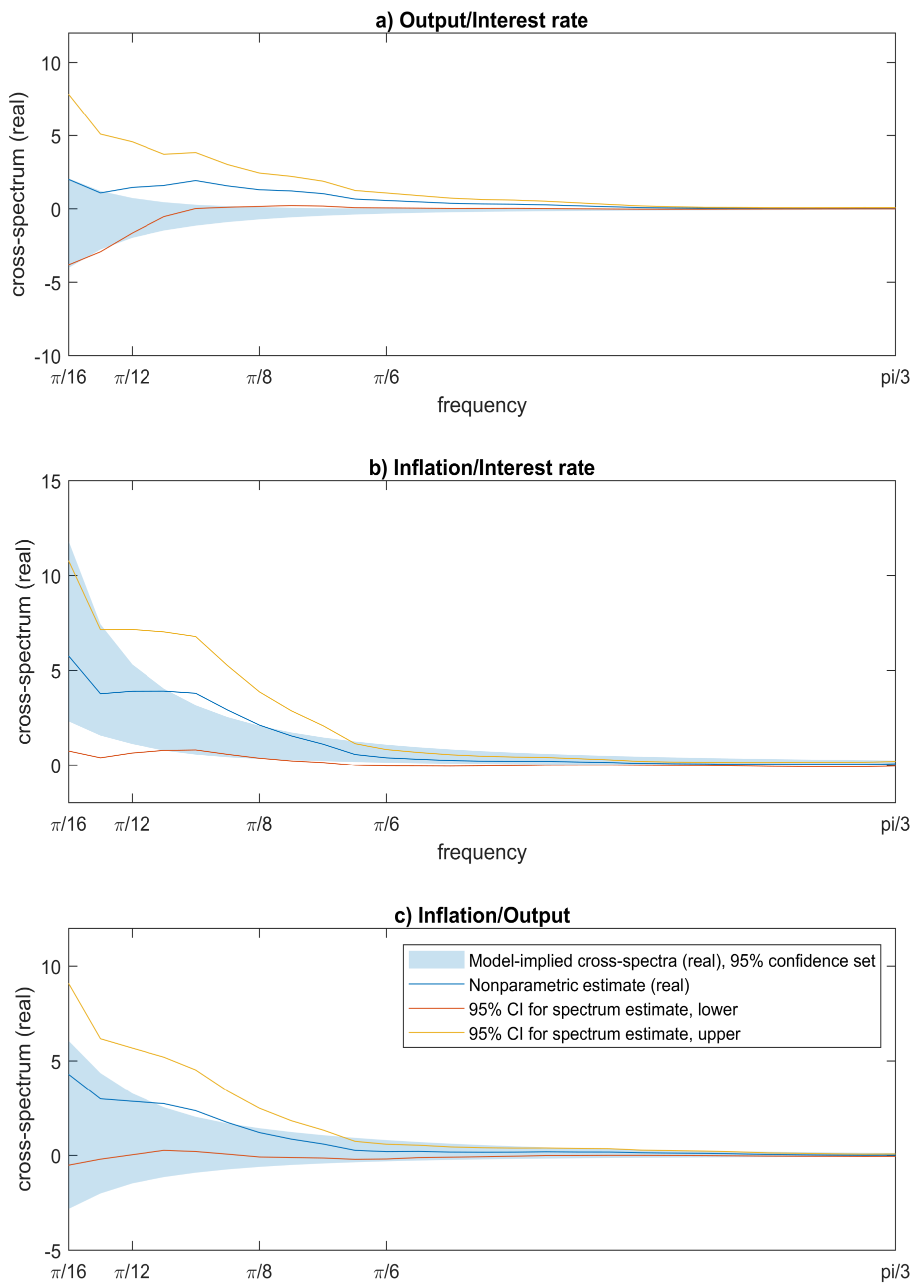


Figure A9: Cross spectra (imaginary part) using the posterior under indeterminacy, 1960-2007.

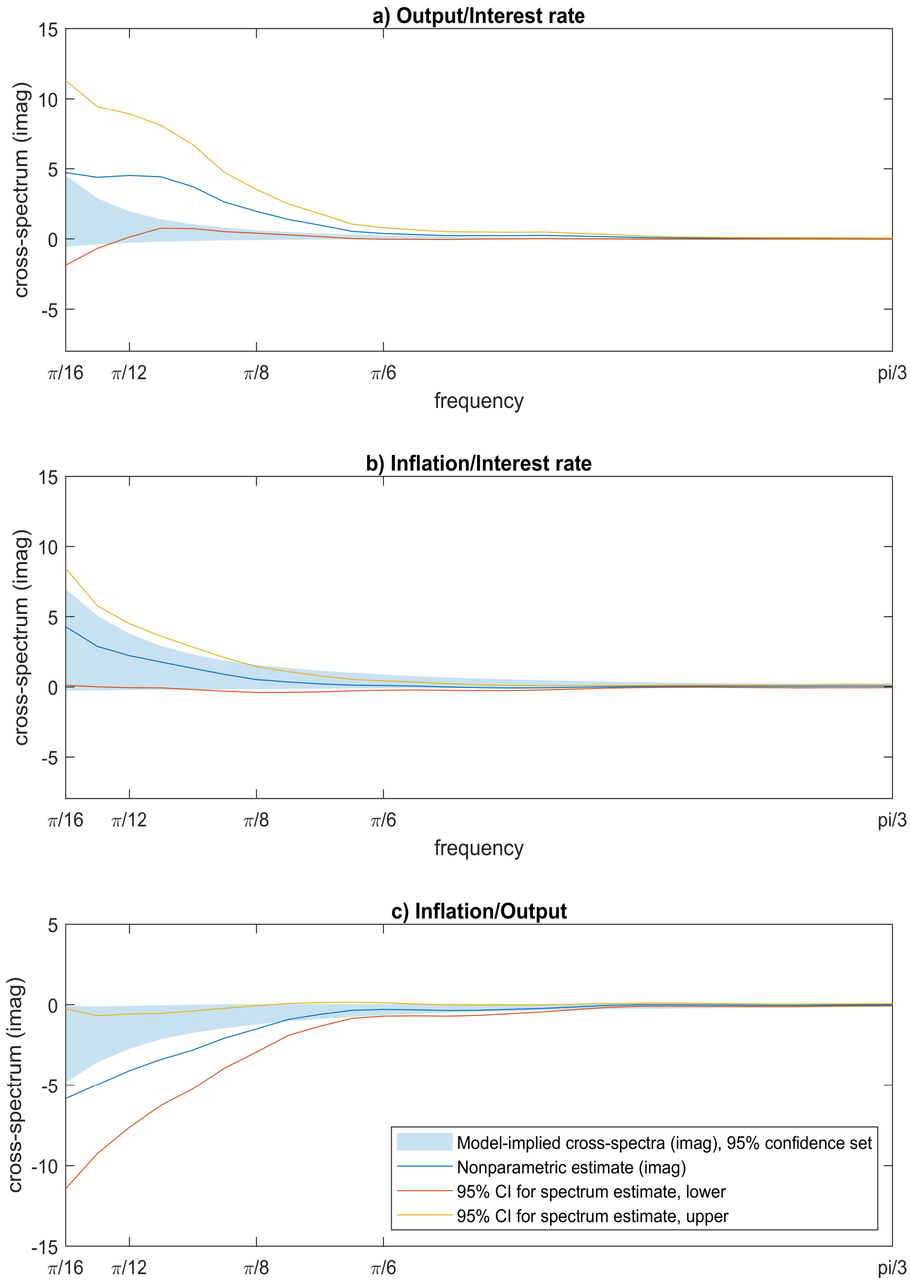


Figure A10: Log spectra using the posterior under determinacy, 1979-2007.

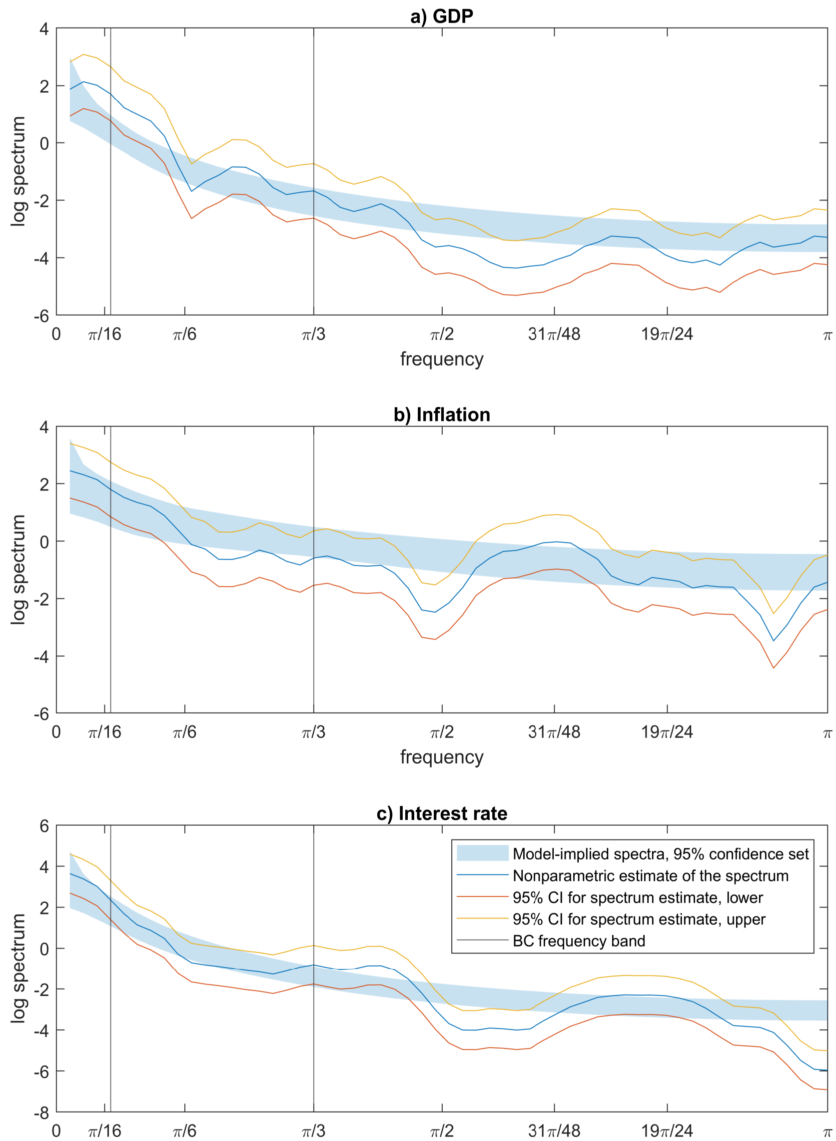


Figure A11: Cross spectra (real part) using the posterior under determinacy, 1979-2007.

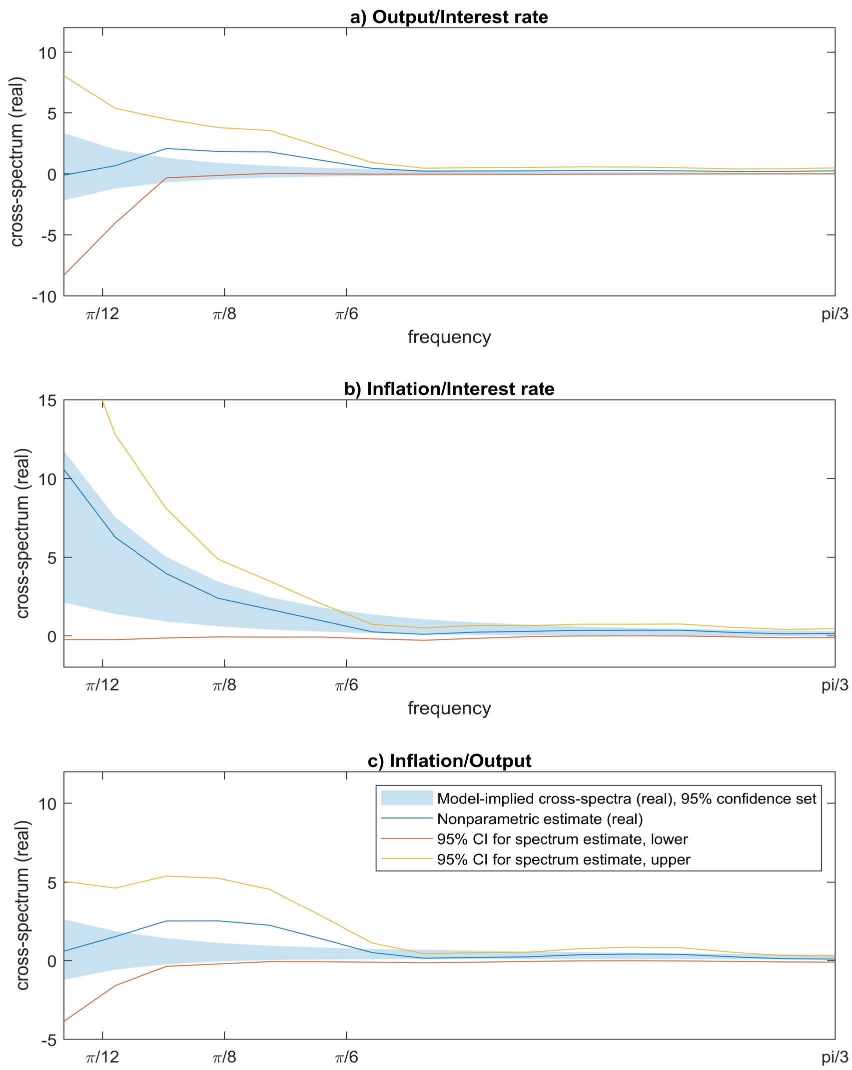


Figure A12: Cross spectra (imaginary part) using the posterior under determinacy, 1979-2007.

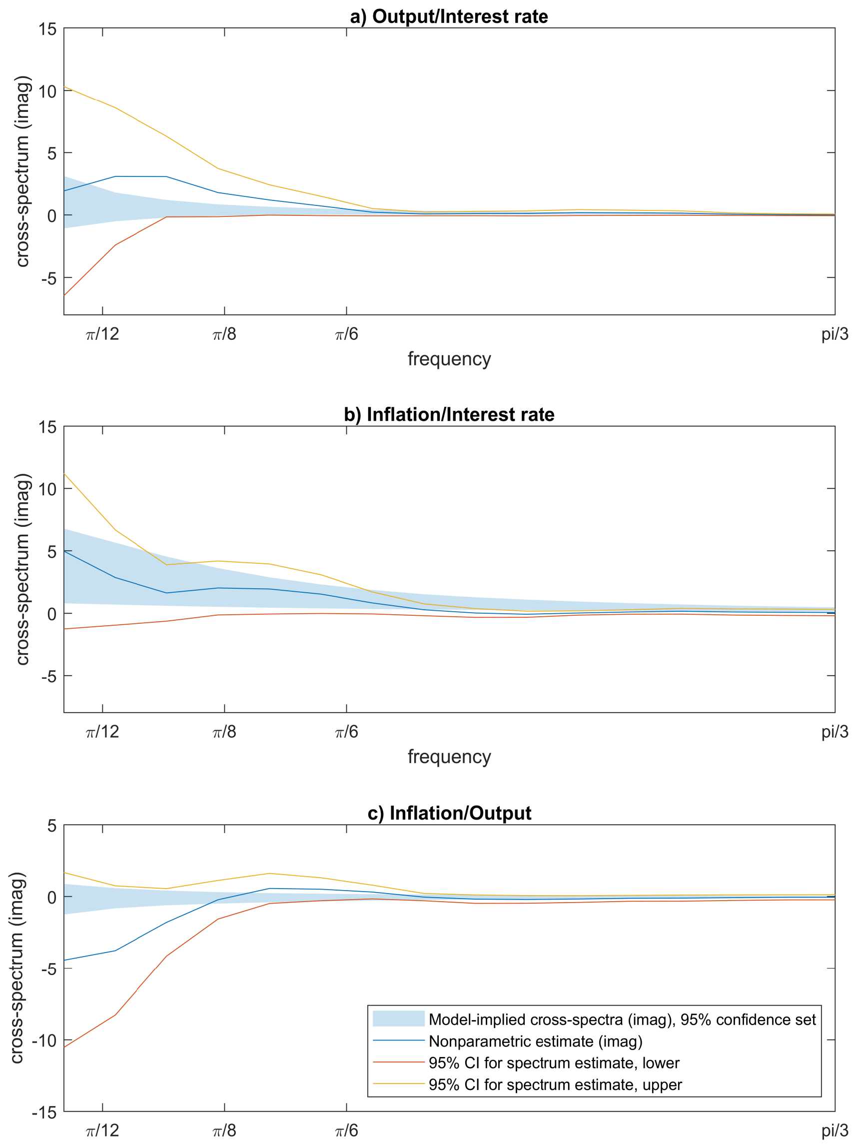


Figure A13: Log spectra under indeterminacy, 1960-1979.

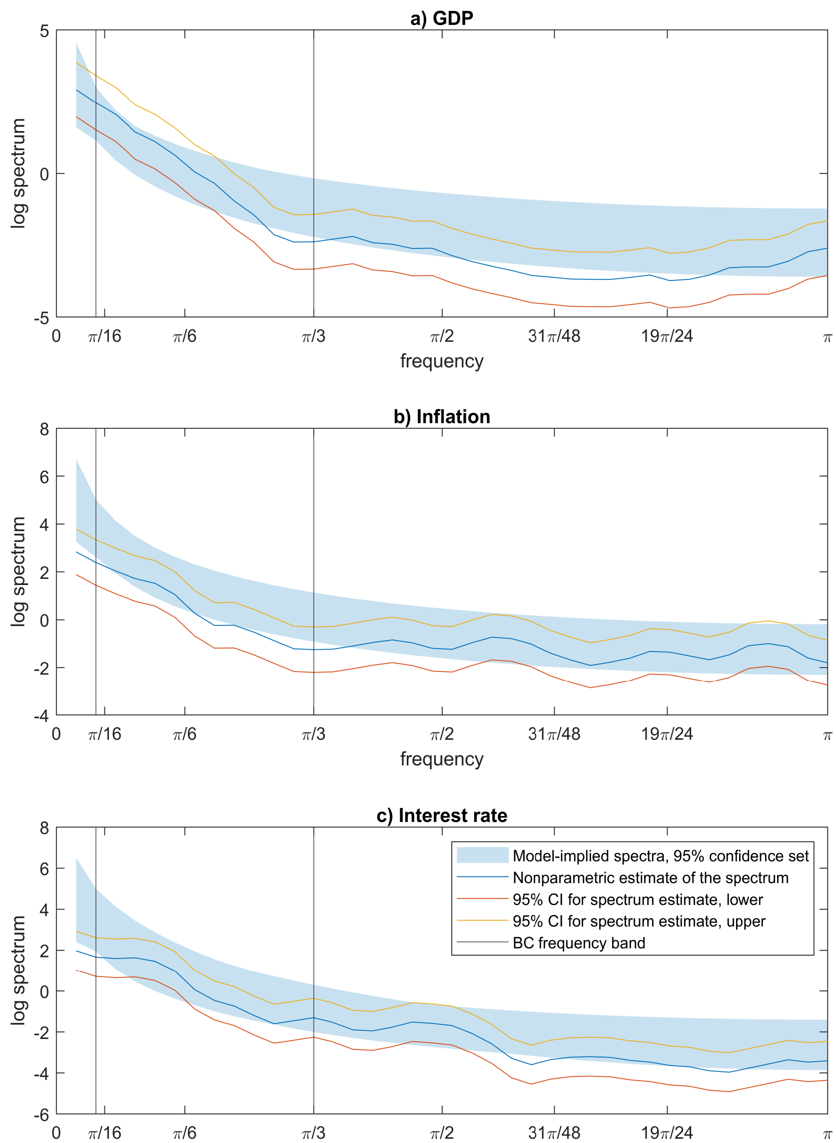


Figure A14: Cross spectra (real part) under indeterminacy, 1960-1979.

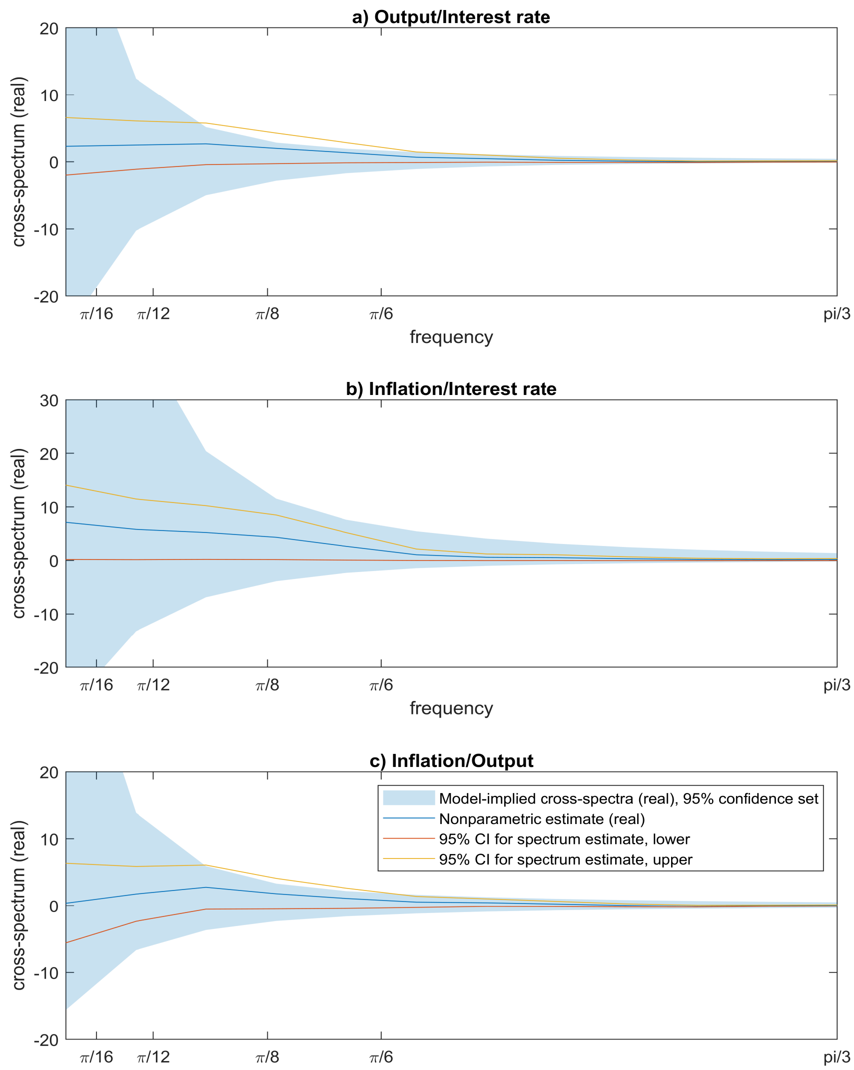


Figure A15: Cross spectra (imaginary part) under indeterminacy, 1960-1979.

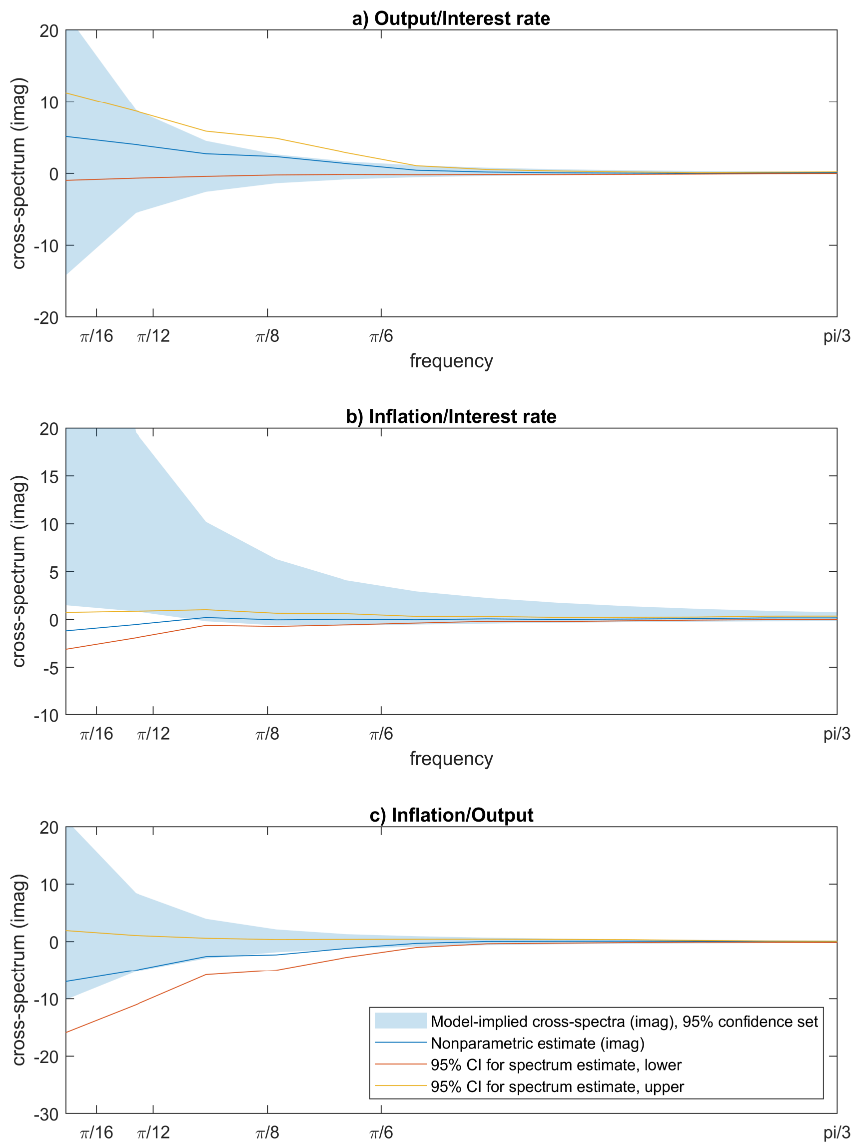


Figure A16: Log spectra using the posterior under indeterminacy, 1960-1979.

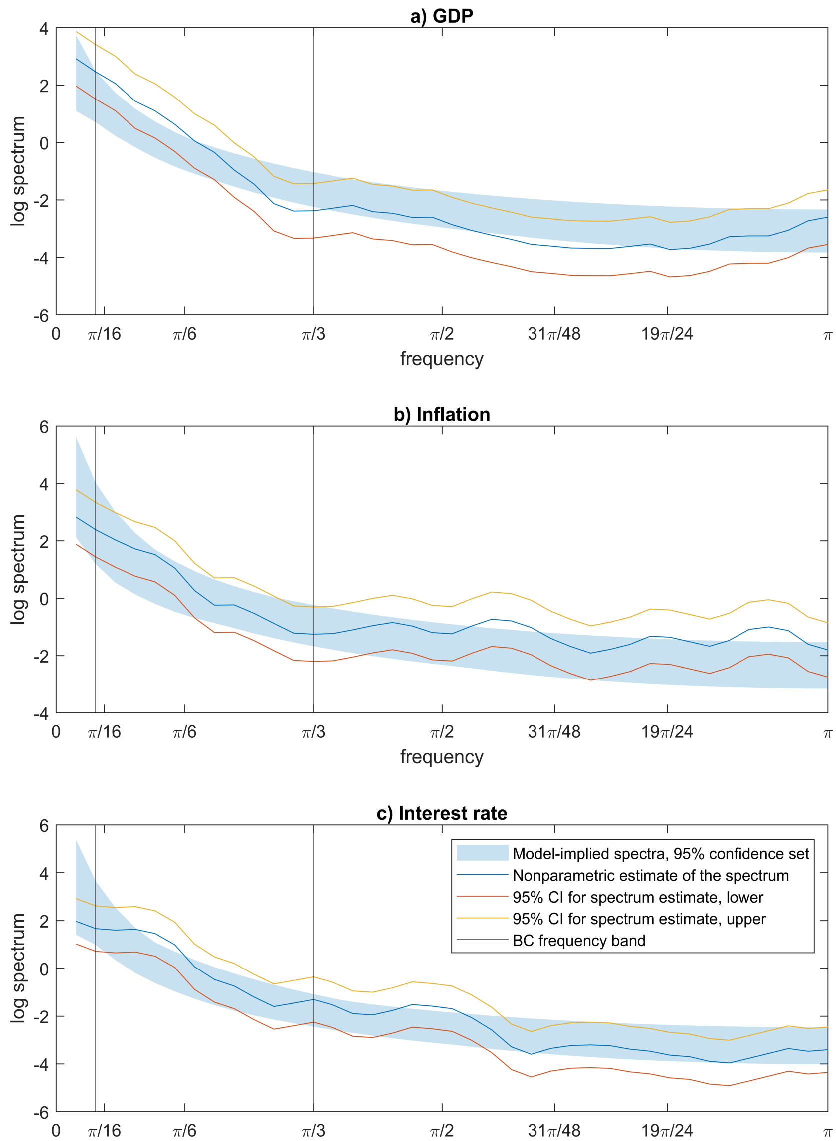


Figure A17: Cross spectra (real part) using the posterior under indeterminacy, 1960-1979.

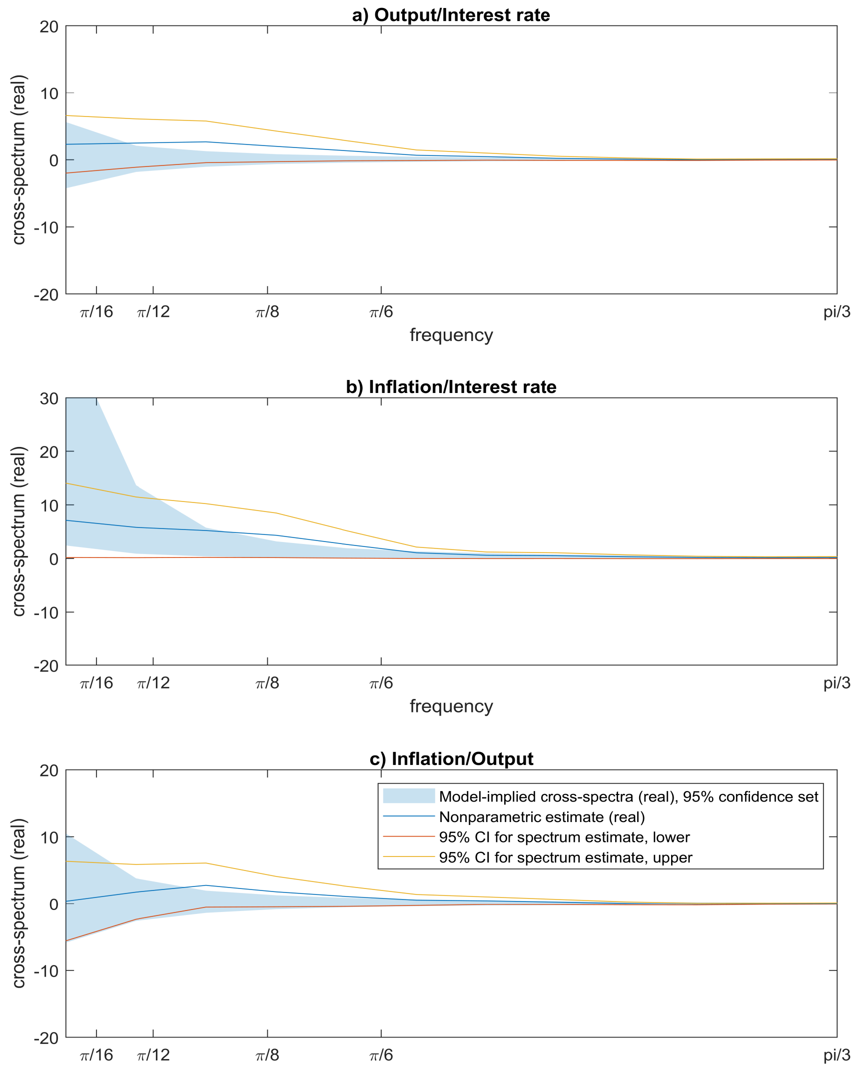


Figure A18: Cross spectra (imaginary part) using the posterior under indeterminacy, 1960-1979.