Supplementary Appendix: Additional Results and Proofs

This appendix is structured as follows. Section S.1 includes additional theoretical results on local identification. Section S.2 considers Non-Gaussianity. Section S.3 illustrates the applicability of the methods to other vector linear processes using the (factor augmented) VARMA model as an example. Section S.4 contains additional results related to the An and Schorfheide (2007) model. Section S.5 applies the methods to the Lubik and Schorfheide (2004) model. Section S.6 contains the equations of the Smets and Wouters (2007) model and additional empirical results. Section S.7 contains additional proofs related to results in the main paper: proof of Theorem 1, followed by two ancillary lemmas that are needed for proving Corollaries 5 and 6. Section S.8 contains proofs of the results in this appendix. Some tables with parameter values, KL and empirical distances are included at the very end.

S.1 Additional conditions for local identification

The next two results generalize Corollaries 1 and 2 in Qu and Tkachenko (2012) to provide identification conditions based on the mean and the spectrum and a subset of frequencies. Let $\mu(\theta)$ denote the mean of $Y_t(\theta)$. Let $W(\omega)$ be an indicator function symmetric about zero to select the desired frequencies used for the identification.

Definition S.1 *The parameter vector θ_0 is said to be locally identifiable from the first and second order properties of $\{Y_t\}$ at a point θ_0 if there exists an open neighborhood of θ_0 in which $\mu(\theta_1) = \mu(\theta_0)$ and $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$ implies $\theta_1 = \theta_0$.*

Corollary S.1 *Assume $\mu(\theta)$ is continuously differentiable over an open neighborhood of θ_0. Let Assumptions 1, 2 and 3 hold, but with $G(\theta)$ in Assumption 3 replaced by*

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

Then, θ is locally identifiable from the first and second order properties of $\{Y_t\}$ at a point θ_0 if and only if $\bar{G}(\theta_0)$ is nonsingular.

Corollary S.2 *Let Assumptions 1-3 hold, but with $G(\theta)$ in Assumption 3 replaced by*

$$G^W(\theta) = \int_{-\pi}^{\pi} W(\omega) \left(\frac{\partial \operatorname{vec} f_\theta(\omega)}{\partial \theta'} \right)^* \left(\frac{\partial \operatorname{vec} f_\theta(\omega)}{\partial \theta'} \right) d\omega.$$

Then, θ is locally identifiable at θ_0 from the second order properties of $\{Y_t\}$ through the frequencies specified by $W(\omega)$ if and only if $G^W(\theta_0)$ is nonsingular.

We now consider the identification of θ^D without making statements about that of θ^U. Intuitively, if θ^D is locally identifiable, then it is potentially possible to determine the parameters describing technology and preferences, even though those governing the equilibrium beliefs may be unidentifiable.

Definition S.2 *The structural parameter vector θ^D is said to be locally partially identifiable from the second order properties of $\{Y_t\}$ at θ_0 if there exists an open neighborhood of θ_0 in which $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ for all $\omega \in [-\pi, \pi]$ necessarily implies $\theta_1^D = \theta_0^D$.*[2]

Corollary S.3 *Under Assumptions 1-3, θ^D is locally partially identifiable from the second order properties of $\{Y_t\}$ at θ_0 if and only if $G(\theta_0)$ and*

$$G^a(\theta_0) = \begin{bmatrix} G(\theta_0) \\ \partial \theta_0^D / \partial \theta' \end{bmatrix}$$

have the same rank.

[2]Note that, as in Rothenberg (1971, footnote p.586), the definition does not exclude there being two points satisfying $f_{\theta_1}(\omega) = f_{\theta_0}(\omega)$ and having the θ^D arbitrarily close in the sense of $\left\|\theta_0^D - \theta_1^D\right\| / \left\|\theta_0 - \theta_1\right\|$ being arbitrarily small.

This result generalizes the first result of Corollary 3 in Qu and Tkachenko (2012). It can also be used to check the local identification of a further subset of θ^D without making statements about the rest of θ, simply by replacing θ_0^D with the corresponding parameter subset of interest.

Next, we consider the identification of θ^D conditional on $\theta^U = \theta_0^U$. Intuitively, if θ^D is locally conditionally identifiable, then it is potentially possible to pin down the parameters describing technology and preferences once we select a mechanism for equilibrium belief formation.

Definition S.3 *The structural parameter vector θ^D is said to be locally conditionally identifiable from the second order properties of $\{Y_t\}$ at $\theta = \theta_0$ if there exists an open neighborhood of θ_0^D in which $f_{(\theta_1^D,\theta_0^U)}(\omega) = f_{(\theta_0^D,\theta_0^U)}(\omega)$ for all $\omega \in [-\pi, \pi]$ necessarily implies $\theta_1^D = \theta_0^D$.*

Corollary S.4 *Under Assumptions 1-3, θ^D is locally conditionally identifiable from the second order properties of $\{Y_t\}$ at θ_0 if and only if*

$$G^D(\theta_0) = \int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta^{D\prime}}\right)^* \left(\frac{\partial \operatorname{vec} f_{\theta_0}(\omega)}{\partial \theta^{D\prime}}\right) d\omega$$

is nonsingular.

This result generalizes the first result of Corollary 4 in Qu and Tkachenko (2012). It can also be used to check the local identification of a further subset of θ^D conditioning on the rest of θ, simply by replacing θ^D with the corresponding parameter subset of interest. The above result follows immediately from Theorem 1. Therefore, the proof is omitted. The above two results can also be applied with both mean and spectrum and with a subset of frequencies selected by $W(\omega)$.

In practice, these results can be applied sequentially. First, Theorem 1 can be applied to check identification based on the full spectrum. If the parameters are identified, then Corollary S.2 can be used to verify whether a subset of frequencies is sufficient for the identification. If the identification fails, then Corollary S.1 can signify whether the information from the steady state can improve the identification. In addition, Corollaries S.3 and S.4 can be used to find out which parameters are responsible for the lack of identification. One can refer to Sections 3.1 and 3.2 in Qu and Tkachenko (2012) for details on how such a procedure is carried out to pinpoint an identification failure due to the parameters in the Taylor rule equation. Because of the developments here, the same procedure can now be carried out under indeterminacy.

S.2 Non-Gaussianity

When the condition $\operatorname{cum}(\widetilde{\varepsilon}_{ta}, \widetilde{\varepsilon}_{sb}, \widetilde{\varepsilon}_{uc}, \widetilde{\varepsilon}_{vd}) = 0$ is relaxed, the variances in Theorem 3 will depend on the joint fourth cumulant of the shocks. Also, using only the second order properties no longer reflects the highest distinguishing power possible. Below we analyze two situations.

Suppose we still intend to compare the models' second order properties. This can be the case if the benchmark model is Gaussian, which performs well in matching the data's second order properties, and we wish to see whether such a feature is largely preserved by the non-Gaussian structure. Then, the measure $p_{fh}(\theta_0, \phi_0, \alpha, T)$ can still be used, but the asymptotic variances $V_{fh}(\theta_0, \phi_0)$ and $V_{hf}(\phi_0, \theta_0)$ will need to be modified, whose formulas are given below.

Corollary S.5 *Let Assumptions 1, 2, 4 and 5 hold for both $f_\theta(\omega)$ and $h_\phi(\omega)$, but with the cumulant condition in Assumption 5 replaced by* $\mathrm{cum}\left(\widetilde{\varepsilon}_{ta}, \widetilde{\varepsilon}_{sb}, \widetilde{\varepsilon}_{uc}, \widetilde{\varepsilon}_{vd}\right)$ *equal to κ_{abcd} if $t = s = u = v$ and 0 otherwise. Then, the results in Theorem 3 still hold after redefining $V_{fh}(\theta_0, \phi_0)$ and $V_{hf}(\phi_0, \theta_0)$ as*

$$
\begin{aligned}
V_{fh}(\theta_0, \phi_0) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \mathrm{tr}\left\{ \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \left[I - f_{\theta_0}(\omega) h_{\phi_0}^{-1}(\omega) \right] \right\} d\omega + M_{fh}(\theta_0, \phi_0), \\
V_{hf}(\phi_0, \theta_0) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \mathrm{tr}\left\{ \left[I - h_{\phi_0}(\omega) f_{\theta_0}^{-1}(\omega) \right] \left[I - h_{\phi_0}(\omega) f_{\theta_0}^{-1}(\omega) \right] \right\} d\omega + M_{hf}(\phi_0, \theta_0),
\end{aligned}
$$

where $M_{fh}(\theta_0, \phi_0)$ equals

$$
\tfrac{1}{64\pi^4} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd} \left[\int_{-\pi}^{\pi} H^*(\omega) (f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega)) H(\omega) d\omega \right]_{ab} \left[\int_{-\pi}^{\pi} H^*(\omega) (f_{\theta_0}^{-1}(\omega) - h_{\phi_0}^{-1}(\omega)) H(\omega) d\omega \right]_{cd}.
$$

In the above $[.]_{ab}$ denotes the (a,b)-th element of the matrix, κ_{abcd} denotes the joint cumulant in the null model, $H(\omega)$ equals $H(\exp(-i\omega); \theta_0)$ in (3), and $H^*(\omega)$ is its conjugate transpose. The term $M_{hf}(\phi_0, \theta_0)$ satisfies the same expression, but with κ_{abcd} now denoting the joint cumulant in the alternative model and $H(\omega)$ determined by the lag polynomial operator in $h_{\phi_0}(\omega)$.

Now, suppose we wish to go beyond the second order properties. Then, the power of the resulting likelihood ratio test will in general need to be computed with simulations. One possible procedure is as follows. Let $L_f(\theta_0)$ and $L_h(\phi_0)$ denote the possibly non-Gaussian log likelihoods associated with the benchmark structure f and the alternative structure h. First, generate time series of length T under the structure f and repeat to obtain the empirical distribution of $L_h(\phi_0) - L_f(\theta_0)$. Find the critical value corresponding to its $100(1-\alpha)$th percentile. Next, generate time series of length T under the structure h and again obtain the empirical distribution. Finally, locate the percentile of the latter distribution that corresponds to the critical value. One minus this percentile gives the power of the test, or the empirical distance measure incorporating the higher order properties.

The simulation aspect prevents the comparison between a large number of models. Nevertheless, one practical implementation can be as follows. Suppose the benchmark structure is Gaussian, and we are interested in models that are close to it in second order properties but differ in higher orders. Then, we can first use $KL_{fh}(\theta_0, \phi)$ and $p_{fh}(\theta_0, \phi, \alpha, T)$ to narrow down the set of alternative models to compare with. For example, if the alternative models have Student-t innovations, this involves narrowing down the degrees of freedom parameter to make the second order properties close to the Gaussian model. If the remaining models form a relatively small set, simulation based comparisons can then be carried out. This implementation reflects that, in practice, a researcher is rarely interested only in a model's high order properties, but rather whether they offer improvements. Therefore, being able to permit both analyses can be viewed as desirable.

S.3 Application to (factor augmented) VARMA models

This section illustrates the applicability of the results to other vector linear processes. Consider

$$
\begin{aligned}
A(L)Y_t &= \lambda(L) f_t + B(L)\varepsilon_t, \\
f_t &= \Gamma(L) f_{t-1} + \zeta_t,
\end{aligned}
$$

where Y_t is an n_Y-by-1 vector of observables, f_t contains the latent factors, ε_t and ζ_t are two sequences of serially uncorrelated structural shocks satisfying $Var(\varepsilon_t) = \Sigma, Var(\zeta_t) = I$ and $E\varepsilon_t\zeta_s' = 0$ for all t and s. $A(L), B(L)$, $\lambda(L)$ and $\Gamma(L)$ are finite order matrix lag polynomials. If $\lambda(L) = 0$, then the model reduces to a VARMA model. If $B(L) = I$, it becomes a factor augmented VAR. If $\lambda(L) = 0$ and $B(L) = I$, then it is simply a structural VAR.

Assume all the roots of $|A(z)| = 0$ lie outside of the unit circle. Let θ denote the vector of the structural parameters in the model. Then, Y_t has the following vector moving average representation

$$Y_t = H_1(L;\theta)\varepsilon_t + H_2(L;\theta)\zeta_t, \tag{S.1}$$

where

$$H_1(L;\theta) = A(L)^{-1}B(L), \quad H_2(L;\theta) = A(L)^{-1}\lambda(L)[I - \Gamma(L)L]^{-1}.$$

Its spectral density equals

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

Some of the identification restrictions, such as the diagonality of Σ or factor loading restrictions on elements of $\lambda(L)$ can be directly incorporated, while other forms of restrictions, such as the long run restrictions, can be written as constraints on θ. Let $\psi(\theta) = 0$ be the collection of such constraints one wishes to impose. Theorem 2 can then be used for checking global identification. There, global identification at θ_0 holds if and only if

$$KL(\theta_0, \theta_1) > 0$$

for all $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$ satisfying $\psi(\theta_1) = 0$. The procedures in Section 5 are also applicable. For example, we can contrast models with and without factor augmentations, models with different identification restrictions, or comparing a (factor augmented) VARMA with a DSGE model.

S.4 An and Schorfheide (2007)

S.4.1 A contrast between determinacy and indeterminacy

We draw parameter values from the posterior distribution in Table 1 and check the local identification at each point. The following parameter bounds are used: $\tau \in [0.01, 10]$, $\beta \in [0.9, 0.999]$, $\kappa \in [0.01, 5]$, $\psi_1 \in [0.01, 0.9]$, $\psi_2 \in [0.01, 5]$, $\rho_r \in [0.1, 0.99]$, $\rho_g \in [0.1, 0.99]$, $\rho_z \in [0.1, 0.99]$, $\sigma_r \in [0.01, 3]$, $\sigma_g \in [0.01, 3]$, $\sigma_z \in [0.01, 3]$, $M_{r\epsilon} \in [-3, 3]$, $M_{g\epsilon} \in [-3, 3]$, $M_{z\epsilon} \in [-3, 3]$, $\sigma_\epsilon \in [0.001, 3]$. Out of the 4000 draws, the smallest eigenvalues are consistently above the tolerance levels (i.e., implying the parameters are locally identified), except for two cases. The two cases involve ρ_r equal to 0.983 and 0.987, which are close to the boundary value of 0.99. Further, for these two points, the absolute and relative deviation measures increase noticeably along the curves, all exceeding E-02 when $\|\theta - \theta_0\|$ reaches 0.050 and 0.709 respectively. For these values the empirical distance measures for T=1000 equal 0.0524 and 0.0625. This suggests that the model is in fact identified at these two parameter values.

We also draw parameter values under determinacy from the posterior distribution in Table 1 and apply Theorem 1 to each point. The same parameter bounds are used, except $\psi_1 \in [1.1, 5]$ and the elements in θ^U are no longer present. Out of the 4000 draws, there are 3998 cases with the smallest eigenvalue below the default tolerance level (i.e., signaling identification failure). Even in the remaining two cases, the eigenvalues are very small, both being of order E-10, and barely exceed the tolerance levels. Also, in both cases, the values of the two measures along the curves (20) remain negligible, the largest still being of order E-05 after $\|\theta^D - \theta_0^D\|$ reaches 1. This shows that the two parameter values are also locally unidentified. In addition, in all cases considered, the lack of identification is caused by the four parameters in the Taylor rule.

S.4.2 Simulations

The empirical distance measures are derived using asymptotic arguments. It is important to examine whether they closely track the power in finite samples. To this end, we simulate the distribution of the exact time domain test (16) under the null hypothesis, obtain the 5% critical value, and then compare it with the simulated distribution under the alternative hypothesis to obtain the rejection frequency. The number of replications is set to 5E5 in both simulations.

The exact powers corresponding to the nine columns of Table 2 are as follows. They are uniformly close to the empirical distances reported in Table 3. Specifically, the nine sets of values are (in the same order as the columns of Table 3): (0.0516, 0.0524, 0.0520, 0.0534); (0.0555, 0.0582, 0.0589, 0.0713); (0.0630, 0.0683, 0.0701, 0.1046); (0.0545, 0.0552, 0.0574, 0.0662); (0.0694, 0.0763, 0.0831, 0.1417); (0.0797, 0.0912, 0.0989, 0.2028); (0.0600, 0.0653, 0.0677, 0.0956); (0.0839, 0.0994, 0.1083, 0.2343); (0.1151, 0.1502, 0.1710, 0.4644).

Regarding computation, it takes approximately 24 hours to obtain one set of values on a Xeon E5-2665 8-core 2.4Ghz processor, as opposed to about one minute to obtain a column of Table 3 using the asymptotic approximation. We will revisit the same issue when analyzing the Smets and Wouters (2007) model.

S.4.3 Asymmetry

We compute the KL and the empirical distance measure using the values in Table 2 but with the roles of θ_0 and θ_1 switched. The resulting values are all fairly close to those reported in Table 3. This further implies that the extent of asymmetry is fairly small. Specifically, the values corresponding to the nine columns in Table 3 are (KL followed by the four empirical distances in each case): (5.19E-07, 0.0509, 0.0513, 0.0515, 0.0534); (1.57E-05, 0.0554, 0.0575, 0.0588, 0.0711); (7.56E-05, 0.0627, 0.0678, 0.0709, 0.1048); (1.04E-05, 0.0545, 0.0562, 0.0572, 0.0669); (1.64E-04, 0.0703, 0.0786, 0.0838, 0.1436); (3.33E-04, 0.0804, 0.0939, 0.1023, 0.2063); (5.53E-05, 0.0601, 0.0644, 0.0669, 0.0942); (4.26E-04, 0.0816, 0.0971, 0.1068, 0.2304); (1.22E-03, 0.1135, 0.1479, 0.1704, 0.4633).

We also carry out the same analysis using the parameter values in Table 4. The finding is very similar. Specifically, the KL and empirical distances corresponding to the four columns in Table 5 are: (3.46E-14, 0.0500, 0.0500, 0.0500, 0.0500); (4.49E-14, 0.0500, 0.0500, 0.0500, 0.0500);

(1.94E-05, 0.0560, 0.0584, 0.0598, 0.0738); (6.00E-05, 0.0605, 0.0650, 0.0677, 0.0965).

The above results imply that if we treated θ_1 as the benchmark and θ_0 as the alternative model, the conclusions regarding (near) observational equivalence would remain the same.

S.4.4 Further illustrations of the empirical distance measure

This subsection further illustrates the informativeness of the empirical distance measure by considering a range of models different from the original An and Schorfheide (2007) model. Throughout the analysis, the original model is regarded as the default specification and the significance level for the empirical distance measure is set to 0.05.

First, consider a situation where the alternative model is known to be difficult to distinguish from the default model even with a large sample size. Specifically, we take a parameter vector from the nonidentification curve computed previously for θ_0^D in Subsection 8.1.1, such that $\left\|\theta^D - \theta_0^D\right\| = 1$, but truncate the values of the non-identified parameters $(\psi_1, \psi_2, \rho_r, \sigma_r)$ to leave 2 decimal places: $(2.06, 1.23, 0.68, 0.24)$. As expected, the KL criterion is small (1.04E-04), and the value of the empirical distances equal 0.0662, 0.0725, 0.0763 and 0.1193 for T=80, 150, 200 and 1000. This exercise is interesting as it illustrates the magnitude of the empirical distance measure that one could expect when the models are nearly observationally equivalent.

Second, consider the case where all of the parameter values in θ^D are the same as in θ_0^D except for the discount factor β, which is now lowered to 0.9852, implying a change in the discount rate from 2% to 6% on an annual basis. The KL criterion equals 1.25E-05. The empirical distance measure equals 0.0535, 0.0553, 0.0564 and 0.0671 for the four sample sizes. These values are similar to those in the previous paragraph. This confirms the empirical fact that it is hard to estimate β to any precision using the dynamic properties of aggregate data on consumption and interest rates.

Third, consider changing the Taylor rule weight ψ_2 in θ_0^D to 1.23 while keeping all other parameters fixed at their original values. The KL criterion equals 0.0579. The empirical distance measure equals 0.8295, 0.9883, 0.9987, 1.0000 for the four sample sizes. This suggests that it is quite feasible to differentiate between the two models with commonly used sample sizes. This also provides a sharp contrast with the first situation. There, ψ_1 and ψ_2 were simultaneously changed to more distant values, but resulted in near observational equivalence.

S.4.5 Global identification from business cycle frequencies

First, we re-compute the empirical distance measures using only business cycle frequencies (i.e., with periods between 6 and 32 quarters). The values associated with the 9 parameter vectors in Table 2 are all above 0.0500 and increasing with the sample size. This shows that there is no observational equivalence over business cycle frequencies. Meanwhile, the feasibility of distinguishing the 9 values from θ_0 clearly deteriorates. The empirical distances equal (compared with the rows of Table 3): for T=80: $0.0505, 0.0523, 0.0547, 0.0531, 0.0635, 0.0702, 0.0595, 0.0696, 0.1034$; for T=1000: $0.0517, 0.0592, 0.0720, 0.0621, 0.1127, 0.1518, 0.0811, 0.1345, 0.3015$. Also, the distances between the output gap and growth rules are (compared with the last two columns of Table 5) 0.0540 and

0.0568 for T=80, and 0.0651 and 0.0753 for T=1000. Therefore, these rules are even closer to being observationally equivalent when only considering their implications at business cycle frequencies.

Next, we minimize the KL criterion using only business cycle frequencies and obtain the empirical distances. We find, interestingly, that the resulting parameter values are very similar to those obtained using the full spectrum. For illustration, the counterpart to the case c=0.1 in Table 2 equals: $(2.41, 0.994, 0.49, 0.62, 0.27, 0.87, 0.66, 0.61, 0.27, 0.58, 0.62, 0.52, -0.06, 0.26, 0.18)'$. The patterns of parameter changes in the rest of the cases are also similar to those in Table 2. Because of the closeness in parameter values, the resulting empirical distance measures are only slightly below those in the previous paragraph. The details are omitted.

S.5 Lubik and Schorfheide (2004)

The log linearized model is

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

where y_t denotes output, π_t is inflation, r_t is the nominal interest rate, g_t is government spending and z_t captures exogenous shifts of the marginal costs of production. The shocks satisfy $\varepsilon_{rt} \sim N(0, \sigma_r^2)$, $\varepsilon_{gt} \sim N(0, \sigma_g^2)$ and $\varepsilon_{zt} \sim N(0, \sigma_z^2)$. Among the three shocks, ε_{gt} and ε_{zt} are allowed to be correlated with correlation coefficient ρ_{gz}. The structural parameters are

$$\theta^D = (\tau, \beta, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r, \sigma_g, \sigma_z, \rho_{gz})'.$$

The state vector is $S_t = (\pi_t, y_t, r_t, g_t, z_t, E_t\pi_{t+1}, E_t y_{t+1})'$ and the observables are r_t, y_t and π_t.

Lubik and Schorfheide (2004) applied the following transformation to the model's solutions to ensure the impulse responses are continuous at the boundary between the determinacy and indeterminacy regions: $S_t = \Theta_1 S_{t-1} + \widetilde{\Theta}_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t$ with $\widetilde{\Theta}_\varepsilon = \Theta_\varepsilon + \Theta_\epsilon(\Theta_\epsilon'\Theta_\epsilon)^{-1}\Theta_\epsilon'(\Theta_\varepsilon^b - \Theta_\varepsilon)$, where Θ_1, Θ_ϵ and Θ_ε are given in this paper's Appendix A and Θ_ε^b is the counterpart of Θ_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 θ^U are specified in the same way as with the An and Schorfheide (2007) model.

S.5.1 Comparing determinacy and indeterminacy

First, consider local identification at the posterior mean reported in Lubik and Schorfheide (2004, Column 1 in Table 3):

$$\theta_0 = (\underbrace{0.69, 0.997, 0.77, 0.77, 0.17, 0.60, 0.68, 0.82, 0.23, 0.27, 1.13, 0.14}_{\theta^D}, \underbrace{-0.68, 1.74, -0.69, 0.20}_{\theta^U})'.$$

The smallest eigenvalue of $G(\theta_0)$ equals 6.2E-04, well above the default tolerance level of 1.9E-09. The absolute and relative deviation measures along the curve (20) corresponding to the smallest eigenvalue exceed E-03 after $\|\theta - \theta_0\|$ reaches 0.085. The empirical distance corresponding to this value for T=1000 equals 0.0585. This further confirms that the parameter vector is locally identified at θ_0.

Next, we draw parameter values from the posterior distribution. The following parameter bounds are imposed: $\tau \in [0.1, 1]$, $\beta \in [0.9, 0.999]$, $\kappa \in [0.01, 5]$, $\psi_1 \in [0.01, 0.9]$, $\psi_2 \in [0.01, 5]$, $\rho_r \in [0.1, 0.99]$, $\rho_g \in [0.1, 0.99]$, $\rho_z \in [0.1, 0.99]$, $\sigma_r \in [0.01, 3]$, $\sigma_g \in [0.01, 3]$, $\sigma_z \in [0.01, 3]$, $\rho_{gz} \in [-0.9, 0.9]$, $M_{r\epsilon} \in [-3, 3]$, $M_{g\epsilon} \in [-3, 3]$, $M_{z\epsilon} \in [-3, 3]$, $\sigma_\epsilon \in [0.01, 3]$. Out of 4000 draws, the smallest eigenvalues are above the default tolerance level for 3996 cases. For the remaining 4 cases, the deviation measures increase noticeably along the curve (20), with absolute deviations exceeding E-03 after $\|\theta - \theta_0\|$ reaches 0.08, and relative absolute deviations reaching 1E-03 when $\|\theta - \theta_0\|$ reaches 0.45. The results indicate that these 4 points are also locally identified. Therefore, the local identification property is not confined to the posterior mean, but rather is a generic feature.

Then, consider local identification properties under determinacy. The smallest eigenvalue of $G(\theta_0^D)$ equals 2.5E-06 at the following posterior mean reported in Lubik and Schorfheide (2004, Table 3):

$$\theta_0^D = (0.54, 0.992, 0.58, 2.19, 0.30, 0.84, 0.83, 0.85, 0.18, 0.18, 0.64, 0.36)'.$$

The Matlab default tolerance level equals 1.1E-11. The largest absolute and relative deviations along the curve exceed E-03 when $\|\theta - \theta_0\|$ reaches 1. The corresponding empirical distance for T=1000 equals 0.0818. Thus, θ is locally identified at θ_0. We also take random draws from the posterior distribution of Lubik and Schorfheide (2004). Theorem 1 is then applied to all the resulting values. Out of 4000 draws, 3997 cases have their smallest eigenvalues above the default tolerance level. For the remaining 3 cases, the corresponding eigenvectors point consistently to the weak identification of β. After fixing its value, the eigenvalues become clearly above the tolerance level. Therefore, like in the indeterminacy case discussed above, the local identification property of θ^D is again a generic feature.

In summary, Taylor rule parameters are locally identified in this model but not in that of An and Schorfheide (2007). These results provide strong evidence that parameter identification is a system property. The identification conclusions reached from discussing a particular equation without referring to its background system are often, at best, fragile.

S.5.2 Global identification

This subsection considers global identification at θ_0 (indeterminacy) and θ_0^D (determinacy).

Under indeterminacy, the parameters minimizing the KL criterion for $c = 0.1, 0.5, 1.0$ are reported in Table S9 and the corresponding KL values and empirical distances are reported in Table S10. They show a pattern similar to the case of An and Schorfheide (2007). On one hand, globally no parameter value is found to be observationally equivalent to θ_0. On the other hand, even with $c = 1.0$, there still exist models with dynamics that are empirically hard to distinguish from those at θ_0. In all three cases, the parameter that moves the most is $M_{g\epsilon}$. We also repeat the analysis with $M_{g\epsilon}$ fixed at its original value to examine whether the identification improves substantially. As shown in Panel (b) in Tables S9 and S10, when c is increased to 1.0, distinguishing between the models becomes empirically feasible. In addition, the parameters that change the most are: σ_ϵ for $c = 0.1$ and $M_{r\epsilon}$ for $c = 0.5$ and $c = 1.0$.

Under determinacy, the parameters minimizing the KL criterion are reported in Table S11 and the KL values and empirical distances in Table S12. The empirical distances are all above 0.0500 and grow with the sample size, showing that the model is globally identified at θ_0^D. However, their values are quite small, suggesting that the models are hard to distinguish empirically. For all three values of c, the parameters that shift the most are the weights ψ_2 and ψ_1. In fact, when fixing all parameters except ψ_2 and ψ_1 at their original values and repeating the minimization, the empirical distance values obtained are very similar. Therefore, the Taylor rule parameters are the main source behind the weak identification. In addition, we study the identification strength when ψ_2 is fixed at its original value. As shown in Panel (b) of Table S12, the identification improves substantially. Fixing ψ_1 instead of ψ_2 leads to a similar pattern of empirical distances.

Given that the Taylor rule parameters are locally identified under determinacy in this model but not in the model of An and Schorfheide (2007), it is useful to further examine the strength of this identification. To this end, we trace out the curve in (20) by varying only the four Taylor rule parameters and following the eigenvector corresponding to the smallest eigenvalue. Table S13 shows ten equally spaced points on the curve along each direction. In direction 1, the curve is terminated when $\left\|\theta^D - \theta_0^D\right\|$ exceeds 1. In direction 2, it stops before ψ_2 turns negative. The table reveals two interesting features. First, the parameters are weakly identified. As shown in the last two columns in the Table, the empirical distance is only 0.1583 with T=1000 when the Euclidean distance from θ_0^D reaches 1.0. Second, along the curve, ψ_1 and ψ_2 move substantially in opposite directions, while ρ_r and σ_r change very little. This suggests that the effect of decreasing (increasing) the weight on the inflation target is largely offset by increasing (decreasing) the weight on the output gap. This implies that, within this model, a hawkish policy (i.e., with $\psi_1 = 2.4840$) can have similar behavioral implications as a more dovish policy (ψ_1 =1.4830), depending on the value of the output gap policy parameter.

We have also considered the identification of the alternative monetary rules as we have done for An and Schorfheide (2007). The results are very similar. That is, there exists an expected inflation rule that is observationally equivalent to the original rule and an output gap rule that is nearly so. This holds under both determinacy and indeterminacy. The details are omitted to save space.

The similarities and differences in the identification properties of the two models can be summarized as follows. (1) Both models are globally identified at the posterior mean under indeterminacy. (2) Under determinacy, the model of Lubik and Schorfheide (2004) is globally identified at the posterior mean while that of An and Schorfheide (2007) is not locally identified. For the latter, the Taylor rule parameters are the source of the lack of identification, while for the former the same parameters lead to near observational equivalence. (3) Both models possess parameters that are weakly identified, although those parameters can differ across models. For both models and under both determinacy and indeterminacy, fixing a small number of parameters can lead to a substantial improvement in global identification.

S.6 Smets and Wouters (2007)

The vector of observable variables includes output (y_t), consumption (c_t), investment (i_t), wage (w_t), labor hours (l_t), inflation (π_t) and the interest rate (r_t). 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 34 structural parameters to vary. They are ordered as

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

The bounds imposed on the parameters throughout the analysis follow those used in Smets and Wouters (2007), except the lower bound for r_π is raised to 1.1, the upper bounds for the autoregressive coefficients of the seven shocks are lowered to 0.99, and those for the moving average coefficients are lowered to 0.90: $\rho_{ga} \in [0.01, 2]$, $\mu_w \in [0.01, 0.9]$, $\mu_p \in [0.01, 0.9]$, $\alpha \in [0.01, 1]$, $\psi \in [0.01, 1]$, $\varphi \in [2, 15]$, $\sigma_c \in [0.25, 3]$, $\lambda \in [0.001, 0.99]$, $\phi_p \in [1, 3]$, $\iota_w \in [0.01, 0.99]$, $\xi_w \in [0.3, 0.95]$, $\iota_p \in [0.01, 0.99]$, $\xi_p \in [0.5, 0.95]$, $\sigma_l \in [0.25, 10]$, $r_\pi \in [1.1, 3]$, $r_{\Delta y} \in [0.01, 0.5]$, $r_y \in [0.01, 0.5]$, $\rho \in [0.5, 0.975]$, $\rho_a \in [0.01, 0.99]$, $\rho_b \in [0.01, 0.99]$, $\rho_g \in [0.01, 0.99]$, $\rho_i \in [0.01, 0.99]$, $\rho_r \in [0.01, 0.99]$, $\rho_p \in [0.01, 0.99]$, $\rho_w \in [0.01, 0.99]$, $\sigma_a \in [0.01, 3]$, $\sigma_b \in [0.025, 5]$, $\sigma_g \in [0.01, 3]$, $\sigma_i \in [0.01, 3]$, $\sigma_r \in [0.01, 3]$, $\sigma_p \in [0.01, 3]$, $\sigma_w \in [0.01, 3]$, $\overline{\gamma} \in [0.1, 0.8], 100(\beta^{-1} - 1) \in [0.01, 2])'$.

Below is an outline of the log linearized system. They are 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: $k_y = \phi_p (L_*/k_*)^{\alpha - 1} = \phi_p \left[((1-\alpha)/\alpha) \left(R_*^k / w_*\right)\right]^{\alpha - 1}$ 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

$$c_t = c_1 c_{t-1} + (1 - c_1) E_t c_{t+1} + c_2 (l_t - E_t l_{t+1}) - c_3 (r_t - E_t \pi_{t+1}) - \varepsilon_t^b. \tag{S.2}$$

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 (S.2) 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

$$q_t = q_1 E_t q_{t+1} + (1-q_1) E_t r_{t+1}^k - (r_t - E_t \pi_{t+1}) - \frac{1}{c_3}\varepsilon_t^b\ , \tag{S.3}$$

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

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

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:

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

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

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

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.

S.6.1 Global identification under determinacy

The second and third verification methods in Section 7 give the following results. Corresponding to the three c values of panel (a), the two deviation measures equal [1.2060,0.0188], [5.8134,0.0877] and [11.0631,0.1657] and the maximum differences between the cdfs with T=1.0E5 increase from 0.1248 to 0.1368 and 0.1681. Therefore, these two methods also support that no observational equivalence is present.

We consider different neighborhood specifications to evaluate the results reported in Subsection 8.2.1. First, we replace $||\cdot||_\infty$ by the L_1 and L_2 norms. The results are reported in Tables S14, S15, S16 and S17. In both cases, the conclusions remain qualitatively the same. Specifically, no observational equivalence is found and φ and σ_l move the most in all but three cases. In the latter cases, they remain among the three parameters that move the most. Next, we consider relative differences, replacing $||\theta-\theta_0||_\infty$ with $||(\theta-\theta_0)./w(\theta_0)||_\infty$ with $w(\theta_0)$ being the lengths of the 90% credible sets reported in Smets and Wouters (2007, Tables 1A and 1B). When all the parameters are allowed to vary, $\overline{\gamma}$ (i.e., the trend growth rate) and $100(\beta^{-1}-1)$ (i.e., the discount rate) are found to move the most in relative terms. These two parameters affect both the dynamic and steady state properties, with the latter utilized when forming the credible sets in Smets and Wouters (2007), but not here when constructing the KL. In light of this, we fix these two parameters in the subsequent analysis. The results are reported in Tables S18 and S19. Again, no observational equivalence is found. Further, the parameters that move the most in relative terms are r_π followed by σ_l and φ. The appearance of r_π is new, while σ_l and φ are consistently found to move the most under absolute changes. The empirical distances are slightly above those in Table 8, with the highest being 0.3462 when T=150. In summary, the results are overall consistent with those reported in the paper. In addition, they reveal substantially lower distinguishability between parameter values than what one might conclude from reading only the posterior credible sets.

Next, as in Subsection 8.1.2, we group parameters into the subsets corresponding to monetary policy (sel=$(r_\pi, r_{\Delta y}, r_y, \rho, \rho_r, \sigma_r)$), exogenous shock processes (sel=$(\rho_{ga}, \mu_w, \mu_p,$ $\rho_a, \rho_b, \rho_g, \rho_i, \rho_r, \rho_p,$ $\rho_w, \sigma_a, \sigma_b, \sigma_g, \sigma_i, \sigma_r, \sigma_p, \sigma_w)$), and behavioral parameters (sel=$(\alpha,$ $\psi,$ $\varphi,$ $\sigma_c,$ $\lambda,$ $\phi_p,$ $\iota_w,$ $\xi_w,$ $\iota_p,$ $\xi_p,$ $\sigma_l,$ $\overline{\gamma},$ $100(1/\beta-1))$. We then conduct the identification analysis with the constraint $\{\theta : ||\theta(\text{sel})-\theta_0(\text{sel})||_\infty \geq c\}$ used throughout. The results are reported in Tables S20 and S21. First, the results for the behavioral parameters are identical to those in the first three columns of Tables 7 and 8 with the constraint binding for φ in all cases. Second, for the monetary policy parameters, the constraint is binding for r_π except for the case with $c = 1.0$, where increasing r_π by 1.0 would exceed the upper bound for this parameter. For the latter, the constraint binds for σ_r. It is notable that, compared to the case with $c = 0.5$, the empirical distances jump from 0.2432 to 1.0 for T=150 because r_π is forced to be within the bounds. Third, for the parameters related to the exogenous shock processes, constraint binds for μ_p in cases $c = 0.1, 0.5$, and for σ_i when $c = 1.0$, where the latter occurs due to μ_p not being able to increase beyond its upper bound. The empirical distances at T=150 for $c = 0.1, 0.5, 1.0$ equal $0.1413, 0.5760, 0.8392$. Overall, the shock parameters appear to be better identified compared with the other two subsets. Finally, consistent with the full vector case, no observational equivalence is detected.

S.6.2 Identification of policy rules

The second and third verification methods in Section 7 give the following results. Corresponding to the three columns of Table 9, the two deviation measures equal [53.16,0.4262], [148.3,1.4473] and [190.1,1.4917] and the maximum differences between the cdfs with T=1.0E5 equal 0.59, 1.00 and 1.00. These two methods support that no observational equivalence is present.

S.6.3 Simulations

We consider the parameters in the columns of Table 7. The simulation design is the same as for the An and Schorfheide (2007) model. The six sets of values are as follows (in the same order as in Table 8): (0.0546, 0.0554, 0.0562, 0.0640); (0.0711, 0.0808, 0.0851, 0.1497); (0.0945, 0.1156, 0.1284, 0.3125); (0.0562, 0.0609, 0.0617, 0.0806); (0.0890, 0.1093, 0.1227, 0.2881); (0.1347, 0.1873, 0.2176, 0.6170).

Regarding computation, it takes approximately 26 hours to obtain one set of values on a Xeon E5-2665 8-core 2.4Ghz processor. It takes about two minutes to obtain a column of Table 8.

S.6.4 Asymmetry

We compute the KL and the empirical distance measure using the values in Table 7 with the roles of θ_0 and θ_1 switched. The values corresponding to the six columns in Table 8 are (KL followed by the four empirical distances in each case): (8.15E-06, 0.0538, 0.0553, 0.0561, 0.0646); (1.86E-04, 0.0703, 0.0793, 0.0848, 0.1500); (6.66E-04, 0.0933, 0.1150, 0.1290, 0.3111); (2.87E-05, 0.0574, 0.0603, 0.0620, 0.0799); (5.88E-04, 0.0899, 0.1096, 0.1223, 0.2864); (1.89E-03, 0.1343, 0.1832, 0.2158, 0.6134). These values are all fairly close to those in Table 8.

We also carry out the same analysis using the parameter values in Table 10. The KL and empirical distances corresponding to the eight columns in Table 10 are (KL followed by empirical distances at T=80 and 150 as in Table 10): (0.6706, 1.0000, 1.0000); (0.3447, 1.0000, 1.0000); (0.0079, 0.2914, 0.4465); (0.0296, 0.6699, 0.8937); (0.1271, 0.9948, 1.0000); (0.1962, 0.9998, 1.0000); (0.0266, 0.6422, 0.8505); (0.1154, 0.9737, 0.9998). The KL values are close to those in Table 10 except for the first value. All the empirical distances are close to the corresponding values in Table 10.

Finally, we consider the parameter values related to the three alternative policy rules in Subsection 8.2.2. The KL and empirical distances are: (0.0084, 0.2813, 0.4414, 0.5398, 0.9914); (0.0530, 0.7721, 0.9513, 0.9854, 1.0000); (0.1667, 0.9990, 1.0000, 1.0000, 1.0000). The above values are close to those reported in Table 9.

S.6.5 Global identification from business cycle frequencies

This subsection carries out the analysis similar to that in Subsection S.4.5. The findings can be summarized as follows. First, the model is still globally identified at θ_0^D when only business cycle frequencies are considered. The respective parameter values minimizing the KL defined in (9) are broadly similar to those in Table 7. Second, distinguishing between alternative policy rules is still feasible at typical sample sizes. For T=80, the distances at the minimizers of (9) are $0.1356, 0.2596, 0.8305$ and for T=150, they grow to $0.1785, 0.3743, 0.9615$ (these values can be contrasted with the second and the third rows in Table 9). It is notable that the changes in empirical distances vary substantially across the policy rules. For the expected inflation and the output growth rules the empirical distances are more than halved compared to the full spectrum case. In contrast, the empirical distances for the output gap rule change relatively little.

Finally, we also repeated the analysis summarized in Table 10 using business cycle frequencies. The resulting empirical distances at the minimizers of (9) with T=80 in the same order as in the table are (these can be compared with the second row in Table 10): $0.9460, 0.9370, 0.1090, 0.2221, 0.5894, 0.7662, 0.2433, 0.6631$. Overall, the values imply that the statements made in the preceding subsection about the relative importance of various frictions also hold when confining attention to the business cycle frequencies. Also, as in the case with monetary rules, the changes in empirical distances vary substantially across frictions. The results for price and wage stickiness display very little change, while those for price and wage indexation as well as for the variable capital utilization fall dramatically to about a third of the respective values for the full spectrum. The empirical distances for the remaining real frictions decrease moderately. These findings are informative in view of the fact that the current generation of DSGE models is designed for business cycle movements, not fluctuations at very low or high frequencies.

S.7 Proofs related to results in the main paper

The proof for Theorem 1 is essentially the same as that for Theorem 1 in Qu and Tkachenko (2012, supplementary appendix). This is because, after the parameter augmentation, θ determines the second order properties of the process. We outline the main steps. Let

$$f_\theta(\omega)^R = \begin{bmatrix} \mathrm{Re}(f_\theta(\omega)) & \mathrm{Im}(f_\theta(\omega)) \\ -\mathrm{Im}(f_\theta(\omega)) & \mathrm{Re}(f_\theta(\omega)) \end{bmatrix}, \tag{S.4}$$

where Re() and Im() denote the real and the imaginary parts of a complex matrix. Because $f_\theta(\omega)$ is Hermitian, $f_\theta(\omega)^R$ is real and symmetric. Further, let $R(\omega;\theta) = \mathrm{vec}(f_\theta(\omega)^R)$. Because the correspondence between $f_\theta(\omega)$ and $R(\omega;\theta)$ is one to one, to prove the results it suffices to consider

$R(\omega;\theta)$. In addition, Lemma A1 in Qu and Tkachenko (2012, supplementary material) states that

$$\left(\frac{\partial \operatorname{vec} f_\theta(\omega)}{\partial\theta'}\right)^* \left(\frac{\partial \operatorname{vec} f_\theta(\omega)}{\partial\theta'}\right) = \frac{1}{2}\left(\frac{\partial R(\omega;\theta)}{\partial\theta'}\right)' \left(\frac{\partial R(\omega;\theta)}{\partial\theta'}\right). \tag{S.5}$$

This implies that the left hand side, therefore $G(\theta)$, is real, symmetric and positive semidefinite.
Proof of Theorem 1. The relationship (S.5) implies that $G(\theta)$ equals

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

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

Suppose θ_0 is *not* locally identified. Then, there exists an infinite sequence of vectors $\{\theta_k\}_{k=1}^{\infty}$ approaching θ_0 such that, for each k: $R(\omega;\theta_0) = R(\omega;\theta_k)$ for all $\omega\in[-\pi,\pi]$. For an arbitrary $\omega\in[-\pi,\pi]$, by the mean value theorem and the differentiability of $f_\theta(\omega)$ in θ,

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

where the subscript j denotes the j-th element of the vector and $\widetilde{\theta}(j,\omega)$ lies between θ_k and θ_0 and in general depends on both ω and j. Let $d_k = (\theta_k-\theta_0)/\left\|\theta_k-\theta_0\right\|$, then

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

The sequence $\{d_k\}$ lies on the unit sphere and therefore it has a convergent subsequence with a limit point d (note that d does not depend on j or ω). Assume $\{d_k\}$ itself is the convergent subsequence. As $\theta_k\to\theta_0$, d_k approaches d and

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

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

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

Because $d\neq 0$, $G(\theta_0)$ is singular.

To show the converse, suppose that $G(\theta)$ has a constant reduced rank in a neighborhood of θ_0 denoted by $\delta(\theta_0)$. Then, consider the characteristic vector $c(\theta)$ associated with one of the zero roots of $G(\theta)$. Because $G(\theta)c(\theta)=0$, we have

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

Because the integrand is continuous in ω and always nonnegative, it must equal 0 for $\omega\in[-\pi,\pi]$ and all $\theta\in\delta(\theta_0)$. This further implies

$$\frac{\partial R(\omega;\theta)}{\partial\theta'}c(\theta) = 0 \tag{S.6}$$

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

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

Then,

$$\frac{\partial R(\omega;\theta(v))}{\partial v} = \frac{\partial R(\omega;\theta(v))}{\partial \theta(v)'}\frac{\partial \theta(v)}{\partial v} = \frac{\partial R(\omega;\theta(v))}{\partial \theta(v)'}c(\theta) = 0 \tag{S.7}$$

for all $\omega \in [-\pi, \pi]$ and $0 \leq v \leq \bar{v}$, where the last equality uses (S.6). Thus, $R(\omega;\theta)$ is constant on the curve χ. This implies that $f_\theta(\omega)$ is constant on the same curve and that θ_0 is locally unidentifiable. This completes the proof.

The next two lemmas are needed for Corollaries 5 and 6. Note that the dimension of Ω_{θ_0} approaches infinity as $T \to \infty$, while that of $f_{\theta_0}(\omega)$ stays finite.

Lemma S.1 *Under the conditions stated in Corollary 5, we have, for all $1 \leq j, k, l \leq p+q$:*

$$(i) \quad : \quad T^{-1/2} Y' \frac{\partial \Omega_{\theta_0}^{-1}}{\partial \theta_j} Y = T^{-1/2} \sum_{i=1}^{T-1} \operatorname{tr}\left\{ I(\omega_i) \frac{\partial f_{\theta_0}^{-1}(\omega_i)}{\partial \theta_j} \right\} + o_p(1),$$

$$(ii) \quad : \quad T^{-1/2} \frac{\partial \log \det \Omega_{\theta_0}^{-1}}{\partial \theta_j} = T^{-1/2} \sum_{i=1}^{T-1} \frac{\partial \log \det f_{\theta_0}^{-1}(\omega_i)}{\partial \theta_j} + o(1),$$

$$(iii) \quad : \quad T^{-1} \operatorname{tr}\left(\frac{\partial \Omega_{\theta_0}}{\partial \theta_j} \frac{\partial \Omega_{\theta_0}^{-1}}{\partial \theta_k} \right) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{tr}\left(f_{\theta_0}^{-1}(\omega) \frac{\partial f_{\theta_0}(\omega)}{\partial \theta_j} f_{\theta_0}^{-1}(\omega) \frac{\partial f_{\theta_0}(\omega)}{\partial \theta_k} \right) d\omega + o(1),$$

$$(iv) \quad : \quad \sup_{\theta \in B(\theta_0)} \left| \frac{1}{T} \frac{\partial^3 \mathcal{L}(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \right| = O_p(1),$$

where the spectral density of Y can be either $f_{\theta_0}(\omega)$ or $f_{\theta_T}(\omega)$, and $B(\theta_0)$ denotes the open neighborhood of θ_0 specified in Corollary 5.

Proof of Lemma S.1. The results (i) and (ii) follow from arguments in Brockwell and Davis (1991, p.392-393), but applied to multivariate processes. We present the main steps. The proof of (iii) requires some new arguments. We give more details.

The proof of (i) consists of three steps. Steps 1 and 2 obtain approximations to the right and left hand sides of (i), while Step 3 bridges the approximations. *Step 1.* Let $q_m(\omega)$ be the m-th order Fourier series approximation to $f_{\theta_0}^{-1}(\omega)$:

$$q_m(\omega) = \sum_{|k| \leq m} b_k \exp(-i\omega k), \tag{S.8}$$

where

$$b_k = (2\pi)^{-1} \int_{-\pi}^{\pi} f_{\theta_0}^{-1}(\omega) \exp(i\omega k)\, d\omega. \tag{S.9}$$

We require m being of order $T^{1/5}$. Then, the approximation errors satisfy (see Display (10.8.42) in Brockwell and Davis (1991))

$$\left\|q_m(\omega)-f_{\theta_0}^{-1}(\omega)\right\|+\left\|\partial q_m(\omega)/\partial\theta_j-\partial f_{\theta_0}^{-1}(\omega)/\partial\theta_j\right\|=O(T^{-3/5}) \text{ uniformly over } \omega\in[-\pi,\pi]. \tag{S.10}$$

This implies

$$T^{-1/2}\sum_{i=1}^{T-1} I(\omega_i)\left[\frac{\partial f_{\theta_0}^{-1}(\omega_i)}{\partial\theta_j}-\frac{\partial q_m(\omega_i)}{\partial\theta_j}\right]=O_p(T^{-3/5+1/2})=o_p(1). \tag{S.11}$$

Step 2. Because $q_m(\omega)$ is the spectral density of a VMA(m) process, $q_m^{-1}(\omega)$ is the spectral density of a VAR(m) process. This VAR process is an m-th order approximation to $Y_t(\theta_0)$, while $q_m^{-1}(\omega)$ is the associated approximation to $f_{\theta_0}(\omega)$. Let H_m be the covariance matrix of this VAR process. Then, applying the arguments that lead to Displays (10.8.17) and (10.8.45) in Brockwell and Davis (1991, p.392-393), we have

$$T^{-1/2}Y'\left(\frac{\partial\Omega_{\theta_0}^{-1}}{\partial\theta_j}-\frac{\partial H_m^{-1}}{\partial\theta_j}\right)Y=T^{-1/2}Y'\left(H_m^{-1}\frac{\partial H_m}{\partial\theta_j}H_m^{-1}-\Omega_{\theta_0}^{-1}\frac{\partial\Omega_{\theta_0}}{\partial\theta_j}\Omega_{\theta_0}^{-1}\right)Y=o_p(1), \tag{S.12}$$

where the last equality follows from the triangle inequality, $||H_m^{-1}-\Omega_{\theta_0}^{-1}||=||\Omega_{\theta_0}^{-1}\left(H_m-\Omega_{\theta_0}\right)H_m^{-1}||=O(T^{-3/5})$ and $||\partial H_m/\partial\theta_j-\partial\Omega_{\theta_0}/\partial\theta_j||=O(T^{-3/5})$. *Step 3.* To bridge the two approximations in (S.11) and (S.12), we introduce a new matrix, $\widetilde{H}_m^{-1}$, which equals the covariance matrix of a VMA(m) process with spectral density $(4\pi^2)^{-1}q_m(\omega)$. The (j,k)-th n_Y-*by*-n_Y block of this matrix is given by

$$\widetilde{h}_{jk}=(4\pi^2)^{-1}\int_{-\pi}^{\pi}q_m(\omega)\exp\left(i(j-k)\omega\right)d\omega. \tag{S.13}$$

Then, we have

$$\begin{aligned} T^{-1/2}\sum_{i=1}^{T-1} I(\omega_i)\frac{\partial q_m(\omega_i)}{\partial\theta_j}-T^{-1/2}Y'\frac{\partial\widetilde{H}_m^{-1}}{\partial\theta_j}Y &= o_p(1), \\ T^{-1/2}Y'\left(\frac{\partial H_m^{-1}}{\partial\theta_j}-\frac{\partial\widetilde{H}_m^{-1}}{\partial\theta_j}\right)Y &= o_p(1), \end{aligned} \tag{S.14}$$

where the second equality follows from the arguments in Brockwell and Davis (1991, Displays (10.8.18)-(10.8.20)) and the continuous differentiability of $f_{\theta_0}(\omega)$ with respect to θ. The result (i) follows by combining (S.11), (S.12) and (S.14).

For the result (ii), note that

$$T^{-1/2}\operatorname{tr}\left\{\Omega_{\theta_0}\frac{\partial\Omega_{\theta_0}^{-1}}{\partial\theta_j}\right\}=T^{-1/2}E\left\{Y(\theta_0)'\frac{\partial\Omega_{\theta_0}^{-1}(\theta)}{\partial\theta_j}Y(\theta_0)\right\},$$

where $Y(\theta_0)=(Y_1(\theta_0)',...,Y_T(\theta_0)')'$. The quantity inside the expectation operator has the same structure as the left hand side of (i), therefore can be analyzed in the same way. The display therefore equals

$$T^{-1/2}\sum_{i=1}^{T-1} E\,\mathrm{tr}\left\{I\left(\omega_i\right)\frac{\partial f_{\theta_0}^{-1}(\omega_i)}{\partial\theta_j}\right\}+o\left(1\right)=T^{-1/2}\sum_{i=1}^{T-1}\mathrm{tr}\left\{f_{\theta_0}(\omega_i)\frac{\partial f_{\theta_0}^{-1}(\omega_i)}{\partial\theta_j}\right\}+o\left(1\right). \tag{S.15}$$

This proves (ii).

Now consider the result (iii):

$$\begin{aligned}
&T^{-1}\,\mathrm{tr}\left(\frac{\partial\Omega_{\theta_0}}{\partial\theta_j}\frac{\partial\Omega_{\theta_0}^{-1}}{\partial\theta_k}\right)\\
=\;&\frac{1}{2T}\,\mathrm{tr}\left(\left(\frac{\partial\Omega_{\theta_0}}{\partial\theta_j}-\frac{\partial H_m}{\partial\theta_j}\right)\frac{\partial\Omega_{\theta_0}^{-1}}{\partial\theta_k}\right)+\frac{1}{2T}\,\mathrm{tr}\left(\frac{\partial H_m}{\partial\theta_j}\left(\frac{\partial\Omega_{\theta_0}^{-1}}{\partial\theta_k}-\frac{\partial\widetilde{H}_m^{-1}}{\partial\theta_k}\right)\right)+\frac{1}{2T}\,\mathrm{tr}\left(\frac{\partial H_m}{\partial\theta_j}\frac{\partial\widetilde{H}_m^{-1}}{\partial\theta_k}\right).
\end{aligned}$$

By the Von Neumann's trace inequality, the first term on the right hand side satisfies

$$\frac{1}{2T}\left|\mathrm{tr}\left(\frac{\partial(\Omega_{\theta_0}-H_m)}{\partial\theta_j}\frac{\partial\Omega_{\theta_0}^{-1}}{\partial\theta_k}\right)\right|\leq\frac{1}{2T}\sum_{i=1}^{T}\beta_i\alpha_i=O_p(T^{-3/5}), \tag{S.16}$$

where $\{\beta_i\}_{i=1}^T$ and $\{\alpha_i\}_{i=1}^T$ are the singular values of the two components inside the trace operator in decreasing order and the last equality follows because $\beta_i=O_p(T^{-3/5})$ and $\alpha_i=O_p(1)$. The second term can be analyzed in the same way. To further analyze the third term, let h_{a-b} and $\widetilde{h}_{a-b}$ denote the (a,b)-th n_Y-by-n_Y blocks of H_m and $\widetilde{H}_m^{-1}$ respectively. Then:

$$\frac{1}{2T}\,\mathrm{tr}\left(\frac{\partial H_m}{\partial\theta_j}\frac{\partial\widetilde{H}_m^{-1}}{\partial\theta_k}\right)=\mathrm{tr}\left\{\frac{1}{2T}\sum_{a=1}^{T}\sum_{b=1}^{T}\frac{\partial h_{a-b}}{\partial\theta_j}\frac{\partial\widetilde{h}_{b-a}}{\partial\theta_k}\right\}.$$

The term inside the curly brackets satisfies

$$\begin{aligned}
&\frac{1}{2T}\sum_{a=1}^{T}\sum_{b=1}^{T}\frac{\partial h_{a-b}}{\partial\theta_j}\frac{\partial\widetilde{h}_{b-a}}{\partial\theta_k}\\
=\;&\frac{1}{2T}\sum_{a=1}^{T}\sum_{b=1}^{T}\left(\int_{-\pi}^{\pi}\frac{\partial q_m^{-1}(\omega)}{\partial\theta_j}\exp\left(i(a-b)\omega\right)d\omega\right)\frac{\partial\widetilde{h}_{b-a}}{\partial\theta_k}\\
=\;&\pi\int_{-\pi}^{\pi}\frac{\partial q_m^{-1}(\omega)}{\partial\theta_j}\left((2\pi T)^{-1}\sum_{b=1}^{T}\sum_{a=1}^{T}\frac{\partial\widetilde{h}_{b-a}}{\partial\theta_k}\exp\left(-i(b-a)\omega\right)\right)d\omega\\
=\;&\pi\int_{-\pi}^{\pi}\frac{\partial q_m^{-1}(\omega)}{\partial\theta_j}\frac{\partial\left((2\pi T)^{-1}\sum_{b=1}^{T}\sum_{a=1}^{T}\widetilde{h}_{b-a}\exp\left(-i(b-a)\omega\right)\right)}{\partial\theta_k}d\omega\\
=\;&\pi\int_{-\pi}^{\pi}\frac{\partial q_m^{-1}(\omega)}{\partial\theta_j}\frac{\partial\left((2\pi)^{-1}\sum_{s=-T+1}^{T-1}\left(1-|s|/T\right)\widetilde{h}_s\exp\left(-is\omega\right)\right)}{\partial\theta_k}d\omega\\
=\;&\pi\int_{-\pi}^{\pi}\frac{\partial q_m^{-1}(\omega)}{\partial\theta_j}\frac{\partial\left[(2\pi)^{-1}\sum_{s=-m}^{m}\left(1-|s|/T\right)\widetilde{h}_s\exp\left(-is\omega\right)\right]}{\partial\theta_k}d\omega\\
=\;&\pi\int_{-\pi}^{\pi}\frac{\partial q_m^{-1}(\omega)}{\partial\theta_j}\frac{\partial\left((4\pi^2)^{-1}q_m(\omega)\right)}{\partial\theta_k}d\omega+o(1),
\end{aligned}$$

where the first equality uses the definition of H_m, the second follows from exchanging the order of the summation and the integration, the third involves exchanging the order of summation and differentiation, the fourth follows from rearranging the terms, the fifth follows because $\widetilde{h}_s$ equals zero for $|s| > M$, and the last equality follows because the quantity inside the brackets is a summation under the triangular window with the bandwidth $m = O(T^{1/5})$. By (S.10), the leading term in the last line of the preceding display converges to

$$\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial f_{\theta_0}^{-1}(\omega)}{\partial\theta_j}\frac{\partial f_{\theta_0}(\omega)}{\partial\theta_k}d\omega = -\frac{1}{4\pi}\int_{-\pi}^{\pi} f_{\theta_0}^{-1}(\omega)\frac{\partial f_{\theta_0}(\omega)}{\partial\theta_j} f_{\theta_0}^{-1}(\omega)\frac{\partial f_{\theta_0}(\omega)}{\partial\theta_k}d\omega.$$

This proves (iii).

We now prove (iv). By the chain rule:

$$\begin{aligned}
\frac{1}{T}\frac{\partial^3\mathcal{L}(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} = \; & \frac{1}{2T}\operatorname{tr}\left(\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}\frac{\partial^2\Omega_\theta}{\partial\theta_j\partial\theta_k}\right) - \frac{1}{2T}\operatorname{tr}\left(\Omega_\theta^{-1}\frac{\partial^3\Omega_\theta}{\partial\theta_j\partial\theta_k\partial\theta_l}\right) \\
& -\frac{1}{2T}\operatorname{tr}\left(\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\right) + \frac{1}{2T}\operatorname{tr}\left(\Omega_\theta^{-1}\frac{\partial^2\Omega_\theta}{\partial\theta_k\partial\theta_l}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\right) \\
& -\frac{1}{2T}\operatorname{tr}\left(\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\right) + \frac{1}{2T}\operatorname{tr}\left(\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}\frac{\partial^2\Omega_\theta}{\partial\theta_j\partial\theta_l}\right) \\
& +\frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\Omega_\theta^{-1}Y - \frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial^2\Omega_\theta}{\partial\theta_k\partial\theta_l}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\Omega_\theta^{-1}Y \\
& +\frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\Omega_\theta^{-1}Y - \frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}\frac{\partial^2\Omega_\theta}{\partial\theta_j\partial\theta_l}\Omega_\theta^{-1}Y \\
& +\frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}Y - \frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}\frac{\partial^2\Omega_\theta}{\partial\theta_j\partial\theta_k}\Omega_\theta^{-1}Y \\
& +\frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial^3\Omega_\theta}{\partial\theta_j\partial\theta_k\partial\theta_l}\Omega_\theta^{-1}Y - \frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial^2\Omega_\theta}{\partial\theta_j\partial\theta_k}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}Y \\
& +\frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}Y - \frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial^2\Omega_\theta}{\partial\theta_j\partial\theta_l}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}Y \\
& +\frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}Y - \frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\Omega_\theta^{-1}\frac{\partial^2\Omega_\theta}{\partial\theta_k\partial\theta_l}\Omega_\theta^{-1}Y \\
& +\frac{1}{2T}Y'\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_j}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_k}\Omega_\theta^{-1}\frac{\partial\Omega_\theta}{\partial\theta_l}\Omega_\theta^{-1}Y.
\end{aligned} \tag{S.17}$$

The following three results hold and are needed for analyzing the terms on the right hand side. First, by the law of large numbers, we have

$$\frac{1}{T}\|Y\|^2 = O_p(1). \tag{S.18}$$

Second, by Lemma 3.1(ii) in Davies (1973), we have

$$\left\|\Omega_\theta^{-1}\right\| \le \frac{1}{2\pi}\sup_{\omega\in[-\pi,\pi]}\left\|f_\theta^{-1}(\omega)\right\|, \tag{S.19}$$

Third, by the argument in the proof of Lemma 3.1(i) in Davies (1973), we have

$$\begin{aligned}
\left\| \frac{\partial \Omega_\theta}{\partial \theta_j} \right\| &\leq 2\pi \sup_{\omega \in [-\pi,\pi]} \left\| \frac{\partial f_\theta(\omega)}{\partial \theta_j} \right\|, \\
\left\| \frac{\partial^2 \Omega_\theta}{\partial \theta_j \partial \theta_k} \right\| &\leq 2\pi \sup_{\omega \in [-\pi,\pi]} \left\| \frac{\partial^2 f_\theta(\omega)}{\partial \theta_j \partial \theta_k} \right\|, \\
\left\| \frac{\partial^2 \Omega_\theta}{\partial \theta_j \partial \theta_k \partial \theta_l} \right\| &\leq 2\pi \sup_{\omega \in [-\pi,\pi]} \left\| \frac{\partial^3 f_\theta(\omega)}{\partial \theta_j \partial \theta_k \partial \theta_l} \right\|.
\end{aligned} \tag{S.20}$$

The four expressions on the right hand side in (S.19) and (S.20) are all finite uniformly over $\theta \in B(\theta_0)$ under the assumptions stated in Corollary 5.

We now apply the three results to bound the terms in (S.17). The first term satisfies

$$\begin{aligned}
\frac{1}{T}\left| \mathrm{tr}\left(\Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_l} \Omega_\theta^{-1} \frac{\partial^2 \Omega_\theta}{\partial \theta_j \partial \theta_k} \right) \right| &\leq \frac{1}{\sqrt{T}} \left| \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_l} \right| \frac{1}{\sqrt{T}} \left| \Omega_\theta^{-1} \frac{\partial^2 \Omega_\theta}{\partial \theta_j \partial \theta_k} \right| \\
&\leq \frac{1}{\sqrt{T}} \left| \Omega_\theta^{-1} \right| \left\| \frac{\partial \Omega_\theta}{\partial \theta_l} \right\| \frac{1}{\sqrt{T}} \left| \Omega_\theta^{-1} \right| \left\| \frac{\partial^2 \Omega_\theta}{\partial \theta_j \partial \theta_k} \right\| \\
&\leq n_Y \left\| \Omega_\theta^{-1} \right\|^2 \left\| \frac{\partial \Omega_\theta}{\partial \theta_l} \right\| \left\| \frac{\partial^2 \Omega_\theta}{\partial \theta_j \partial \theta_k} \right\| \\
&= O(1),
\end{aligned}$$

where the three inequalities follow from (B.1) and the equality holds because of (S.19) and (S.20). The second to the sixth terms can be analyzed in the same way and are all $O(1)$. The seventh term satisfies

$$\begin{aligned}
\frac{1}{T}\left| Y' \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_l} \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_k} \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_j} \Omega_\theta^{-1} Y \right| &\leq \frac{1}{\sqrt{T}} |Y| \frac{1}{\sqrt{T}} \left| \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_l} \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_k} \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_j} \Omega_\theta^{-1} Y \right| \\
&\leq \frac{1}{T} \|Y\|^2 \left\| \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_l} \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_k} \Omega_\theta^{-1} \frac{\partial \Omega_\theta}{\partial \theta_j} \Omega_\theta^{-1} \right\| \\
&\leq \frac{1}{T} \|Y\|^2 \left\| \Omega_\theta^{-1} \right\|^4 \left\| \frac{\partial \Omega_\theta}{\partial \theta_l} \right\| \left\| \frac{\partial \Omega_\theta}{\partial \theta_k} \right\| \left\| \frac{\partial \Omega_\theta}{\partial \theta_j} \right\| \\
&= O_p(1),
\end{aligned}$$

where the inequalities again follow from (B.1) and the equality holds because of (S.19), (S.20) and (S.18). The remaining terms in (S.17) can be analyzed in the same way as this term and are all $O_p(1)$. The above results hold uniformly over $B(\theta_0)$ because the derivatives and the inverse of the spectral density matrix are finite for all θ belonging to this neighborhood.

Lemma S.2 *Under the conditions stated in Corollary 6, we have, for all* $1 \leq j, k \leq p+q$*:*

$$
\begin{aligned}
(i) &: \quad T^{-1/2}\,\mathrm{tr}\left\{\Omega_f^{-1}\Omega_\delta\right\} = \frac{T^{1/2}}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left\{f_{\theta_0}^{-1}(\omega)\delta(\omega)\right\}d\omega + o(1),\\
(ii) &: \quad T^{-1}\,\mathrm{tr}\left\{[\Omega_f^{-1}\Omega_\delta]^2\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left\{[f_{\theta_0}^{-1}(\omega)\delta(\omega)]^2\right\}d\omega + o(1),\\
(iii) &: \quad T^{-1/2}Y'\Omega_f^{-1}\Omega_\delta\Omega_f^{-1}Y = T^{-1/2}\sum_{j=1}^{T}\mathrm{tr}\left\{f_{\theta_0}^{-1}(\omega_j)\delta(\omega_j)f_{\theta_0}^{-1}(\omega_j)I\left(\omega_j\right)\right\},\\
(iv) &: \quad T^{-1}Y'[\Omega_f^{-1}\Omega_\delta]^2\Omega_f^{-1}Y = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left\{[f_{\theta_0}^{-1}(\omega)\delta(\omega)]^2\right\}d\omega + o(1),\\
(v) &: \quad T^{-1}Y'[\Omega_f^{-1}\Omega_\delta]^2(\Omega_h^{-1}-\Omega_f^{-1})Y = o_p(1),
\end{aligned}
$$

where $\delta(\omega), \Omega_f, \Omega_h$ *and* Ω_δ *are defined in (17), (18) and (B.21) and the spectral density of* Y *can be either* $f_{\theta_0}(\omega)$ *or* $h_{\phi_T}(\omega)$*.*

Proof of Lemma S.2. The proof of (i) is similar to that of Lemma S.1.(iii). However, because the trace operator is multiplied by $T^{-1/2}$ rather than T^{-1}, some arguments need to be strengthened. Let H_m^{-1} and $\widetilde{H}_m^{-1}$ have the same definitions as in Lemma S.1 (c.f., Steps 2 and 3 in its proof). Adding and subtracting terms:

$$
T^{-1/2}\,\mathrm{tr}\left\{\Omega_f^{-1}\Omega_\delta\right\} = T^{-1/2}\,\mathrm{tr}\left(\widetilde{H}_m^{-1}\Omega_\delta\right)+T^{-1/2}\,\mathrm{tr}\left(\left(\Omega_f^{-1}-H_m^{-1}\right)\Omega_\delta\right)+T^{-1/2}\,\mathrm{tr}\left(\left(H_m^{-1}-\widetilde{H}_m^{-1}\right)\Omega_\delta\right),
$$

The second and third terms on the right hand side are both of order $O(T^{-3/5+1/2}) = o(1)$ in light of (S.16). Therefore,

$$
T^{-1/2}\,\mathrm{tr}\left\{\Omega_f^{-1}\Omega_\delta\right\} = T^{-1/2}\,\mathrm{tr}\left(\widetilde{H}_m^{-1}\Omega_\delta\right) + o(1).
$$

To further analyze the leading term on the right side, let g_{a-b} and $\widetilde{h}_{a-b}$ denote the (a,b)-th n_Y-by-n_Y blocks of Ω_δ and $\widetilde{H}_m^{-1}$ respectively. Then,

$$
T^{-1/2}\,\mathrm{tr}\left(\widetilde{H}_m^{-1}\Omega_\delta\right) = \mathrm{tr}\left\{T^{-1/2}\sum_{a=1}^{T}\sum_{b=1}^{T}g_{a-b}\widetilde{h}_{b-a}\right\}.
$$

The term inside the curly brackets equals

$$
\begin{aligned}
& T^{-1/2}\sum_{a=1}^{T}\sum_{b=1}^{T}\left(\int_{-\pi}^{\pi}\delta(\omega)\exp\left(i(a-b)\omega\right)d\omega\right)\widetilde{h}_{b-a} \qquad \text{(S.21)}\\
= \;& T^{1/2}\int_{-\pi}^{\pi}\delta(\omega)\left(\sum_{k=-T+1}^{T-1}\left(1-|k|/T\right)\widetilde{h}_k\exp\left(-ik\omega\right)\right)d\omega.
\end{aligned}
$$

Applying the definition of b_k (see (S.9)), we have

$$
\left\|\widetilde{h}_k-(2\pi)^{-1}b_k\right\| = (4\pi^2)^{-1}\left\|\int_{-\pi}^{\pi}[q_m(\omega)-f_{\theta_0}^{-1}(\omega)]\exp\left(ik\omega\right)d\omega\right\| = O(T^{-3/5}).
$$

Consequently, the right hand side of (S.21) equals

$$T^{1/2}\int_{-\pi}^{\pi}\delta(\omega)\left((2\pi)^{-1}\sum_{k=-T+1}^{T-1}\left(1-|k|/T\right)b_k\exp\left(-ik\omega\right)\right)d\omega+o(1).$$

By Assumption 6, $b_k=O(k^{-4})$. The preceding display therefore equals

$$T^{1/2}\int_{-\pi}^{\pi}\delta(\omega)\left((2\pi)^{-1}\sum_{k=-\infty}^{\infty}b_k\exp\left(-ik\omega\right)\right)d\omega+o(1)=\frac{T^{1/2}}{2\pi}\int_{-\pi}^{\pi}\delta(\omega)f_{\theta_0}^{-1}(\omega)d\omega+o(1).$$

This proves Lemma S.2(i).

The proof of Lemma S.2(ii) relies on Lemma 3.1 in Davies (1973). Let $Q=\zeta\otimes I$, where ζ denotes a $T\times T$ matrix whose (j,k)-th element $(j,k=1,...,T)$ equals $\exp(2\pi i(j-1)(k-1)/T)/\sqrt{T}$ and I is the $n_Y\times n_Y$ identity matrix. Let P_f and P_δ be $n_YT\times n_YT$ dimensional block diagonal matrices whose j-th diagonal blocks $(j=1,...,T)$ are equal to $f_{\theta_0}(\omega_{j-1})$ and $\delta(\omega_{j-1})$ respectively. The following identity follows from properties of the trace operator:

$$T^{-1}\operatorname{tr}\left\{[\Omega_f^{-1}\Omega_\delta]^2\right\}=T^{-1}\operatorname{tr}\left\{(Q\Omega_f^{-1}Q^*)(Q\Omega_\delta\Omega_f^{-1}\Omega_\delta Q^*)\right\}. \tag{S.22}$$

Adding and subtracting terms, the right hand side can be rewritten as

$$T^{-1}\operatorname{tr}\left\{(Q\Omega_f^{-1}Q^*-P_f^{-1})(Q\Omega_\delta\Omega_f^{-1}\Omega_\delta Q^*)\right\}+T^{-1}\operatorname{tr}\left\{(Q\Omega_\delta Q^*)(Q\Omega_f^{-1}\Omega_\delta Q^*P_f^{-1})\right\}.$$

The first term is bounded by $\left|T^{-1/2}\left(Q\Omega_f^{-1}Q^*-P_f^{-1}\right)\right|\left\|\sqrt{n_Y}Q\Omega_\delta\Omega_f^{-1}\Omega_\delta Q^*\right\|$. The first norm is $o(1)$ by Lemma 3.1(iv) in Davies (1973), while the second is finite by Lemma 3.1(i)-(ii) in the same paper. Therefore, this term is negligible. Consequently, (S.22) equals

$$T^{-1}\operatorname{tr}\left\{(Q\Omega_\delta Q^*)(Q\Omega_f^{-1}\Omega_\delta Q^*P_f^{-1})\right\}+o(1). \tag{S.23}$$

Note that the leading term in (S.23) has the same structure as the right hand side of (S.22). Therefore, the same argument as between (S.22) and (S.23) can be applied. This process can be continued, leading to

$$\begin{aligned}T^{-1}\operatorname{tr}\left\{[\Omega_f^{-1}\Omega_\delta]^2\right\} &= T^{-1}\operatorname{tr}\left\{[P_f^{-1}P_\delta]^2\right\}+o(1)\\ &= T^{-1}\operatorname{tr}\left\{\sum_{j=0}^{T-1}[f_{\theta_0}^{-1}(\omega_j)\delta(\omega_j)]^2\right\}+o(1)\\ &\to \frac{1}{2\pi}\int_{-\pi}^{\pi}\operatorname{tr}\left\{[f_{\theta_0}^{-1}(\omega)\delta(\omega)]^2\right\}d\omega.\end{aligned}$$

This proves Lemma S.2(ii).

Consider Lemma S.2(iii). Let $\delta_m(\omega)$ be the m-th order Fourier series approximation to $\delta(\omega)$. Let $\widetilde{R}_m$ be an n_YT-dimensional square matrix whose (j,k)-th n_Y-*by*-n_Y block is given by

$$\widetilde{r}_{jk}=(4\pi^2)^{-1}\int_{-\pi}^{\pi}\delta_m(\omega)\exp\left(i(j-k)\omega\right)d\omega. \tag{S.24}$$

Continue to let m be of order $T^{1/5}$. Then, by Step 2 in the proof of Lemma S.1(i), we have

$$T^{-1/2}Y'(\Omega_f^{-1}\Omega_\delta\Omega_f^{-1})Y = T^{-1/2}Y'(\widetilde{H}_m^{-1}\widetilde{R}_m\widetilde{H}_m^{-1})Y + o_p(1)\,. \tag{S.25}$$

To relate the right hand side of (S.25) to the periodograms of Y, we need a further approximation to $\widetilde{H}_m^{-1}$. Let $\bar{H}_m^{-1}$ be a $n_Y T$-dimensional square matrix whose (j,k)-th n_Y-by-n_Y block is given by

$$\bar{h}_{jk} = (2\pi T)^{-1}\sum_{a=1}^{T} q_m(\omega_a)\exp\left(i(j-k)\omega_a\right) \text{ if } |j-k|\le m \text{ and } 0 \text{ otherwise.} \tag{S.26}$$

The difference between $\bar{h}_{jk}$ and $\widetilde{h}_{jk}$ (recall $\widetilde{h}_{jk}$ is the (j,k)-th block of $\widetilde{H}_m^{-1}$,see (S.13)) is that the latter is defined using an integral rather than an average. Because $|j-k| = O(T^{1/5})$, this difference is small: $||\bar{h}_{jk}-\widetilde{h}_{jk}|| \le CT^{-4/5}$ for some $C<\infty$ and for all j and k. Applying this result along with the argument in Brockwell and Davis (1991, display (10.8.16)), we have $||\widetilde{H}_m^{-1} - \bar{H}_m^{-1}|| = O(T^{-3/5})$. Similarly, let $\bar{R}_m$ be an $n_Y T$-dimensional square matrix whose (j,k)-th n_Y-*by*-n_Y block is given by

$$\bar{r}_{jk} = (2\pi T)^{-1}\sum_{a=1}^{T} \delta_m(\omega_a)\exp\left(i(j-k)\omega_a\right) \text{ if } |j-k|\le m \text{ and } 0 \text{ otherwise.} \tag{S.27}$$

Comparing this with (S.24) and applying the same argument as above, we have $||\widetilde{R}_m - \bar{R}_m||| = O(T^{-3/5})$. Consequently,

$$\begin{aligned} T^{-1/2}Y'(\widetilde{H}_m^{-1}\widetilde{R}_m\widetilde{H}_m^{-1})Y &= T^{-1/2}Y'(\bar{H}_m^{-1}\bar{R}_m\bar{H}_m^{-1})Y + O_p(T^{-1/10}) \\ &= T^{-1/2}\sum_{j,k,l,n=1}^{T} Y_j'\bar{h}_{jk}\bar{r}_{kl}\bar{h}_{ln}Y_n + O_p(T^{-1/10}). \end{aligned}$$

We further analyze the leading term on the right hand side. Expressing $\bar{h}_{jk}$ and $\bar{r}_{kl}$ using their spectral densities as in (S.26) and S.27), we can write this leading term as

$$\begin{aligned} &\frac{1}{4\pi^2T^{5/2}}\sum_{j,k,l,n,a,b=1}^{T} Y_j' q_m(\omega_a)\delta_m(\omega_b)\exp\left(i(j-k)\omega_a\right)\exp\left(i(k-l)\omega_b\right)\bar{h}_{ln}Y_n \\ =\ &\frac{1}{4\pi^2T^{5/2}}\sum_{j,l,n,a,b=1}^{T} Y_j' q_m(\omega_a)\delta_m(\omega_b)\bar{h}_{ln}Y_n\exp\left(ij\omega_a\right)\exp\left(-il\omega_b\right)\sum_{k=1}^{T}\exp\left(ik\left(\omega_b-\omega_a\right)\right). \end{aligned}$$

The summation over k equals zero unless $\omega_b = \omega_a$. The preceding display therefore simplifies to

$$\frac{1}{4\pi^2T^{3/2}}\sum_{j,l,n,a=1}^{T} Y_j q_m(\omega_a)\delta_m(\omega_a)\bar{h}_{ln}Y_n\exp\left(ij\omega_a\right)\exp\left(-il\omega_a\right).$$

Expressing $\bar{h}_{ln}$ using its spectral density and repeating the same argument, we can further write the preceding display as

$$
\begin{aligned}
& \frac{1}{2\pi T^{3/2}} \sum_{j,n,a=1}^{T} Y_j' q_m(\omega_a)\delta_m(\omega_a) q_m(\omega_a) Y_n \exp(ij\omega_a)\exp(-in\omega_a) \\
= \;& T^{-1/2}\sum_{a=1}^{T} \operatorname{tr}\left\{ q_m(\omega_a)\delta_m(\omega_a) q_m(\omega_a) \left[(2\pi T)^{-1} \sum_{j,n=1}^{T} Y_n Y_j' \exp(i(j-n)\omega_a) \right] \right\} \\
= \;& T^{-1/2}\sum_{a=1}^{T} \operatorname{tr}\left\{ q_m(\omega_a)\delta_m(\omega_a) q_m(\omega_a) I(\omega_a) \right\} \\
= \;& T^{-1/2}\sum_{j=1}^{T} \operatorname{tr}\left\{ f_{\theta_0}^{-1}(\omega_j)\delta(\omega_j) f_{\theta_0}^{-1}(\omega_j) I(\omega_j) \right\} + o_p(1),
\end{aligned}
$$

where the last equality applies (S.10). This proves Lemma S.2(iii).

Consider Lemma S.2(iv). By the proof of S.2(iii), we have

$$
T^{-1}Y'([\Omega_f^{-1}\Omega_\delta]^2\Omega_f^{-1})Y = T^{-1}\sum_{j=1}^{T} \operatorname{tr}\left\{ [f_{\theta_0}^{-1}(\omega_j)\delta(\omega_j)]^2 f_{\theta_0}^{-1}(\omega_j) I(\omega_j) \right\} + o_p(1).
$$

The result then follows by the law of large numbers. For the result S.2(v), note that

$$
T^{-1}Y'([\Omega_f^{-1}\Omega_\delta]^2(\Omega_h^{-1} - \Omega_f^{-1}))Y = -T^{-3/2}Y'([\Omega_f^{-1}\Omega_\delta]^3\Omega_h^{-1})Y.
$$

The right hand side is of order $O(T^{-1/2})$ by the proof of (iii). This proves Lemma S.2(v).

S.8 Proofs for results in this supplementary appendix

Proof of Corollary S.1. Let

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

then

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

Using this representation, the proof proceeds in the same way as in Theorem 1, with $R(\omega;\theta)$ replaced by $\bar{R}(\omega;\theta)$. The detail is omitted.

Proof of Corollary S.2. This follows immediately from the proof of Theorem 1 because $W(\omega)$ is nonnegative and the integrand of $G(\theta)$ is positive semidefinite.

Proof of Corollary S.3. Recall $\theta = (\theta^{D\prime}, \theta^{U\prime})'$. Suppose θ_0^D is *not* locally identified. Then, there exists an infinite sequence of vectors $\{\theta_k\}_{k=1}^{\infty}$ approaching θ_0 such that

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

By the definition of the partial identification, $\{\theta_k^D\}$ can be chosen such that $\|\theta_k^D - \theta_0^D\| / \|\theta_k - \theta_0\| > \varepsilon$ with ε being some arbitrarily small positive number. The values of θ_k^U can either change or stay fixed in this sequence; no restrictions are imposed on them besides those in the preceding display. As in the proof of Theorem 1, in the limit, we have

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

with $d^D \neq 0$ (where d^D is comprised of the elements in d that correspond to θ^D). Therefore, on one hand,

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

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

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

which implies

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

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

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

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

which implies for all $\omega \in [-\pi, \pi]$ and all $\theta \in \delta(\theta_0)$ (see arguments leading to (S.6))

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

but

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

where $c^D(\theta)$ denotes the elements in $c(\theta)$ that correspond to θ^D. Because $G(\theta)$ is continuous and has constant rank in $\delta(\theta_0)$, the vector $c(\theta)$ is continuous in $\delta(\theta_0)$. As in Theorem 1, consider the curve χ defined by the function $\theta(v)$ which solves for $0 \leq v \leq \bar{v}$ the differential equation

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

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

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

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

Proof of Corollary S.5. Relaxing the cumulant condition does not affect the asymptotic normality. Therefore, it suffices to verify that the asymptotic variances have the stated expressions. We only consider the null hypothesis as the proof under the alternative hypothesis is similar. Let $\phi(\omega) = f_{\theta_0}^{-1}(\omega_j) - h_{\phi_0}^{-1}(\omega_j)$. Then, (B.5) can be written as

$$\frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \operatorname{tr}\left\{\phi(\omega_j)(I(\omega_j) - f_{\theta_0}(\omega_j))\right\} = \sum_{k,l=1}^{n_Y} \left\{ \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \phi_{kl}(\omega_j)(I_{lk}(\omega_j) - f_{\theta_0 lk}(\omega_j)) \right\},$$

where $\phi_{kl}(\omega_j)$ is the (k,l)-th element of $\phi(\omega_j)$ and other quantities are defined analogously. Denote the quantity in the curly brackets by $A(k,l)$. Applying the same argument as in Proposition 10.8.5 in Brockwell and Davis (1991), we have

$$A(k,l) = \frac{1}{2T^{1/2}} \sum_{j=1}^{T-1} \sum_{a,b=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{kl}(\omega_j) H_{la}(\omega_j) \left(I_{ab}^{\widetilde{\varepsilon}}(\omega_j) - E I_{ab}^{\widetilde{\varepsilon}}(\omega_j) \right) H_{bk}^*(\omega_j) + o_p(1),$$

where $I_{ab}^{\widetilde{\varepsilon}}(\omega_j)$ is the (a,b)-th element of the periodogram of $\widetilde{\varepsilon}_t$ and $H_{bk}^*(\omega_j)$ the (b,k)-th element of $H^*(\omega_j)$. Note that $H_{bk}^*(\omega_j) = \overline{H_{kb}(\omega_j)}$. The covariance between $A(k,l)$ and $A(m,n)$ then equals

$$\begin{aligned} & \frac{1}{4T} \sum_{j,h=1}^{T-1} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{kl}(\omega_j) H_{la}(\omega_j) H_{bk}^*(\omega_j) \operatorname{Cov}(I_{ab}^{\widetilde{\varepsilon}}(\omega_j), I_{cd}^{\widetilde{\varepsilon}}(\omega_h)) \overline{\phi_{mn}(\omega_h) H_{nc}(\omega_h) H_{dm}^*(\omega_h)} + o(1) \\ = & \frac{1}{4T} \sum_{j,h=1}^{T-1} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{kl}(\omega_j) H_{la}(\omega_j) H_{bk}^*(\omega_j) \operatorname{Cov}(I_{ab}^{\widetilde{\varepsilon}}(\omega_j), I_{cd}^{\widetilde{\varepsilon}}(\omega_h)) \phi_{nm}(\omega_h) H_{cn}^*(\omega_h) H_{md}(\omega_h) + o(1). \end{aligned}$$

Proposition 11.7.3 in Brockwell and Davis (1991) shows for $0 < \omega_j, \omega_h < \pi$:

$$\operatorname{Cov}\left(I_{ab}^{\widetilde{\varepsilon}}(\omega_j), I_{cd}^{\widetilde{\varepsilon}}(\omega_h) \right) = \begin{cases} \frac{1}{4\pi^2 T} \kappa_{abcd} + \frac{1}{4\pi^2} \sigma_{ac} \sigma_{db} & \text{if } \omega_j = \omega_h, \\ \frac{1}{4\pi^2 T} \kappa_{abcd} & \text{if } \omega_j \neq \omega_h, \end{cases}$$

where σ_{ac} is the covariance between the a-th and the c-th elements of $\widetilde{\varepsilon}_t$. Applying this result, the preceding summation equals

$$\begin{aligned} & \frac{1}{8\pi^2 T} \sum_{j=1}^{T-1} \left\{ \sum_{a,c=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{kl}(\omega_j) H_{la}(\omega_j) \sigma_{ac} H_{cn}^*(\omega_j) \sum_{b,d=1}^{\dim(\widetilde{\varepsilon}_t)} \phi_{nm}(\omega_j) H_{md}(\omega_j) \sigma_{db} H_{bk}^*(\omega_j) \right\} \\ & + \frac{1}{16\pi^2 T^2} \sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd} \left[\sum_{j=1}^{T-1} H_{bk}^*(\omega_j) \phi_{kl}(\omega_j) H_{la}(\omega_j) \sum_{h=1}^{T-1} H_{cn}^*(\omega_h) \phi_{nm}(\omega_h) H_{md}(\omega_h) \right] + o(1). \end{aligned}$$

The first term converges to

$$\frac{1}{4\pi} \int_{-\pi}^{\pi} \phi_{kl}(\omega) f_{\theta_0 ln}(\omega) \phi_{nm}(\omega) f_{\theta_0 mk}(\omega) d\omega.$$

Upon taking summation over $1\leq k,l,m,n \leq n_Y$, it equals

$$
\begin{aligned}
& \frac{1}{4\pi}\int_{-\pi}^{\pi} \operatorname{tr}\left\{\left(f_{\theta_0}^{-1}(\omega)-h_{\phi_0}^{-1}(\omega)\right) f_{\theta_0}(\omega)\left(f_{\theta_0}^{-1}(\omega)-h_{\phi_0}^{-1}(\omega)\right) f_{\theta_0}(\omega)\right\} d\omega \\
= & \frac{1}{4\pi}\int_{-\pi}^{\pi} \operatorname{tr}\left\{\left[I-f_{\theta_0}(\omega)h_{\phi_0}^{-1}(\omega)\right]\left[I-f_{\theta_0}(\omega)h_{\phi_0}^{-1}(\omega)\right]\right\} d\omega.
\end{aligned}
$$

Meanwhile, the second term summed over $1\leq k,l,m,n \leq n_Y$ equals

$$
\begin{aligned}
& \frac{1}{16\pi^2T^2}\sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd}\left[\sum_{j=1}^{T-1}\sum_{k,l=1}^{n_Y} H_{bk}^*(\omega_j)\phi_{kl}(\omega_j)H_{la}(\omega_j)\sum_{h=1}^{T-1}\sum_{m,n=1}^{n_Y} H_{cn}^*(\omega_h)\phi_{nm}(\omega_h)H_{md}(\omega_h)\right] \\
\to \ & \frac{1}{16\pi^2}\sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} H^*(\omega)\left(f_{\theta_0}^{-1}(\omega)-h_{\phi_0}^{-1}(\omega)\right)H(\omega)d\omega\right]_{ba} \\
& \times\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} H^*(\omega)\left(f_{\theta_0}^{-1}(\omega)-h_{\phi_0}^{-1}(\omega)\right)H(\omega)d\omega\right]_{cd}.
\end{aligned}
$$

Because $\kappa_{abcd}=\kappa_{bacd}$, the right hand side can also be expressed as

$$
\begin{aligned}
& \frac{1}{16\pi^2}\sum_{a,b,c,d=1}^{\dim(\widetilde{\varepsilon}_t)} \kappa_{abcd}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} H^*(\omega)\left(f_{\theta_0}^{-1}(\omega)-h_{\phi_0}^{-1}(\omega)\right)H(\omega)d\omega\right]_{ab} \\
& \times\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} H^*(\omega)\left(f_{\theta_0}^{-1}(\omega)-h_{\phi_0}^{-1}(\omega)\right)H(\omega)d\omega\right]_{cd}.
\end{aligned}
$$

This completes the proof.

Table S1. Parameter values minimizing the KL criterion, AS (2007) model, L_1 norm

		(a) All parameters can vary			(b) τ fixed			(c) τ and ψ_2 fixed		
	θ_0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
τ	2.51	**2.46**	**2.24**	**3.08**	2.51	2.51	2.51	2.51	2.51	2.51
β	0.995	0.995	0.994	0.996	0.996	0.998	0.999	0.997	0.999	0.999
κ	0.49	0.49	0.50	0.50	0.48	0.45	0.43	0.47	0.41	0.38
ψ_1	0.63	0.63	0.60	0.69	0.64	0.57	0.49	0.63	0.62	0.61
ψ_2	0.23	0.25	0.34	0.02	**0.19**	**0.43**	**0.67**	0.23	0.23	0.23
ρ_r	0.87	0.87	0.88	0.86	0.87	0.88	0.90	0.87	0.88	0.88
ρ_g	0.66	0.66	0.66	0.66	0.66	0.66	0.65	0.66	0.65	0.65
ρ_z	0.60	0.61	0.62	0.58	0.60	0.60	0.60	0.60	0.61	0.61
σ_r	0.27	0.27	0.27	0.27	0.27	0.27	0.28	0.27	0.27	0.27
σ_g	0.58	0.58	0.58	0.57	0.58	0.59	0.60	0.58	0.60	0.62
σ_z	0.62	0.62	0.61	0.66	0.61	0.56	0.51	**0.59**	**0.48**	0.36
$M_{r\epsilon}$	0.53	0.52	0.51	0.57	0.53	0.50	0.48	0.52	0.49	0.47
$M_{g\epsilon}$	-0.06	-0.06	-0.05	-0.08	-0.06	-0.04	-0.03	-0.05	-0.03	-0.01
$M_{z\epsilon}$	0.26	0.26	0.26	0.27	0.27	0.31	0.35	0.28	0.39	**0.63**
σ_ϵ	0.19	0.184	0.179	0.186	0.185	0.182	0.177	0.184	0.170	0.116

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$ with θ_0 corresponding to the default specification. All values are rounded to the second decimal place except for β and σ_ϵ. The bold value signifies the parameter that moves the most.

Table S2. KL and empirical distances between θ_c and θ_0, AS (2007) model, L_1 norm

	(a) All parameters can vary			(b) τ fixed			(c) τ and ψ_2 fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	1.62E-07	4.38E-06	2.12E-05	3.12E-06	6.28E-05	1.68E-04	5.99E-06	1.27E-04	3.52E-04
T=80	0.0505	0.0528	0.0564	0.0523	0.0610	0.0685	0.0533	0.0658	0.0799
T=150	0.0507	0.0539	0.0588	0.0532	0.0656	0.0769	0.0546	0.0728	0.0937
T=200	0.0508	0.0545	0.0603	0.0537	0.0684	0.0820	0.0553	0.0770	0.1023
T=1000	0.0519	0.0605	0.0751	0.0587	0.0981	0.1418	0.0624	0.1255	0.2091

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$ with θ_c given in the columns of Table S1. The empirical distance measure equals $p_{ff}(\theta_0, \theta_c, 0.05, T)$, where T is specified in the last four rows of the Table.

Table S3. Parameter values minimizing the KL criterion, AS (2007) model, L_2 norm

		(a) All parameters can vary			(b) τ fixed			(c) τ and ψ_2 fixed		
	θ_0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
τ	2.51	**2.42**	**2.98**	**3.48**	2.51	2.51	2.51	2.51	2.51	2.51
β	0.995	0.995	0.996	0.994	0.994	0.996	0.999	0.998	0.999	0.999
κ	0.49	0.49	0.49	0.52	0.50	0.46	0.42	0.45	0.38	0.40
ψ_1	0.63	0.62	0.68	0.70	0.61	0.50	0.35	0.62	0.61	0.61
ψ_2	0.23	0.27	0.06	0.01	**0.33**	**0.71**	**1.17**	0.23	0.23	0.23
ρ_r	0.87	0.87	0.86	0.85	0.88	0.90	0.92	0.87	0.88	0.88
ρ_g	0.66	0.66	0.66	0.66	0.66	0.65	0.65	0.66	0.65	0.64
ρ_z	0.60	0.61	0.58	0.57	0.60	0.60	0.59	0.61	0.60	0.66
σ_r	0.27	0.27	0.27	0.26	0.27	0.28	0.28	0.27	0.27	0.27
σ_g	0.58	0.58	0.57	0.56	0.58	0.59	0.60	0.59	0.62	0.61
σ_z	0.62	0.62	0.65	0.69	0.63	0.57	0.50	**0.55**	0.35	0.23
$M_{r\epsilon}$	0.53	0.52	0.56	0.60	0.53	0.50	0.46	0.51	0.47	0.48
$M_{g\epsilon}$	-0.06	-0.06	-0.08	-0.10	-0.06	-0.05	-0.03	-0.05	-0.01	-0.01
$M_{z\epsilon}$	0.26	0.26	0.27	0.27	0.26	0.30	0.37	0.31	**0.65**	**1.15**
σ_ϵ	0.19	0.184	0.186	0.184	0.186	0.185	0.175	0.182	0.110	0.001

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$, where θ_0 corresponds to the default specification. All values are rounded to the second decimal place except for β and σ_ϵ. The bold value signifies the parameter that moves the most.

Table S4. KL and empirical distances between θ_c and θ_0, AS (2007) model, L_2 norm

	(a) All parameters can vary			(b) τ fixed			(c) τ and ψ_2 fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	4.47E-07	1.36E-05	6.97E-05	9.79E-06	1.54E-04	3.18E-04	3.02E-05	3.65E-04	9.61E-04
T=80	0.0509	0.0550	0.0616	0.0540	0.0673	0.0761	0.0577	0.0807	0.1059
T=150	0.0512	0.0569	0.0665	0.0557	0.0752	0.0888	0.0608	0.0948	0.1347
T=200	0.0514	0.0580	0.0694	0.0566	0.0800	0.0967	0.0626	0.1036	0.1534
T=1000	0.0532	0.0694	0.1014	0.0660	0.1362	0.1951	0.0811	0.2137	0.3980

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$ with θ_c given in the columns of Table S3. The empirical distance measure equals $p_{ff}(\theta_0, \theta_c, 0.05, T)$, where T is specified in the last four rows of the Table.

Table S5. Parameter values minimizing the KL criterion, AS (2007) model, weighted constraints

		(a) All parameters can vary			(b) β fixed			(c) β and ψ_2 fixed		
	θ_0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
τ	2.51	2.48	2.33	2.21	2.60	2.96	2.06	**2.67**	**3.31**	3.41
β	0.995	**0.994**	**0.992**	**0.988**	0.995	0.995	0.995	0.995	0.995	0.995
κ	0.49	0.49	0.51	0.54	0.49	0.49	0.47	0.49	0.50	0.44
ψ_1	0.63	0.63	0.62	0.61	0.64	0.69	0.50	0.63	0.64	0.64
ψ_2	0.23	0.25	0.32	0.38	**0.19**	**0.02**	**0.66**	0.23	0.23	0.23
ρ_r	0.87	0.87	0.87	0.88	0.87	0.86	0.90	0.87	0.87	0.87
ρ_g	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.65
ρ_z	0.60	0.61	0.62	0.63	0.60	0.58	0.62	0.59	0.57	0.57
σ_r	0.27	0.27	0.27	0.27	0.27	0.27	0.28	0.27	0.27	0.27
σ_g	0.58	0.58	0.58	0.57	0.57	0.57	0.59	0.57	0.56	0.58
σ_z	0.62	0.62	0.62	0.62	0.63	0.65	0.57	0.63	0.65	0.44
$M_{r\epsilon}$	0.53	0.53	0.52	0.52	0.54	0.57	0.47	0.54	0.57	0.55
$M_{g\epsilon}$	-0.06	-0.06	-0.06	-0.06	-0.06	-0.08	-0.04	-0.07	-0.09	-0.07
$M_{z\epsilon}$	0.26	0.26	0.25	0.25	0.26	0.27	0.28	0.26	0.29	0.61
σ_ϵ	0.19	0.185	0.181	0.175	0.186	0.187	0.177	0.187	0.183	**0.035**

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$, where θ_0 contains parameter values of the default specification. All values are rounded to the second decimal place except for β and σ_ϵ. The constraint is given by $\{\theta_c : |(\theta_c - \theta_0)./w(\theta_0)|_\infty \geq c\}$, where $w(\theta_0)$ contains the lengths of the 90% credible sets from Table 1. The bold value signifies the binding constraint.

Table S6. KL and empirical distances between θ_c and θ_0, AS (2007) model, weighted constraints

	(a) All parameters can vary			(b) β fixed			(c) β and ψ_2 fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	2.85E-07	6.18E-06	2.29E-05	7.47E-07	1.91E-05	9.90E-05	4.82E-06	1.01E-04	2.48E-04
T=80	0.0509	0.0542	0.0579	0.0512	0.0563	0.0635	0.0528	0.0638	0.0719
T=150	0.0512	0.0554	0.0605	0.0516	0.0586	0.0694	0.0540	0.0699	0.0825
T=200	0.0513	0.0562	0.0621	0.0519	0.0600	0.0731	0.0546	0.0736	0.0891
T=1000	0.0527	0.0634	0.0778	0.0542	0.0739	0.1136	0.0609	0.1148	0.1691

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$ with θ_c given in the columns of Table S5. The empirical distance measure equals $p_{ff}(\theta_0, \theta_c, 0.05, T)$, where T is specified in the last four rows of the Table.

Table S7. Parameter values minimizing the KL criterion, AS (2007) model, subset constraints

		(a) Monetary policy			(b) Shock processes			(c) Behavioral		
	θ_0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
τ	2.51	2.31	2.00	2.38	2.91	2.96	2.38	**2.41**	**3.01**	**3.51**
β	0.995	0.994	0.992	0.999	0.994	0.999	0.999	*0.995*	*0.996*	*0.994*
κ	0.49	0.50	0.49	0.42	0.52	0.41	0.39	*0.49*	*0.49*	*0.52*
ψ_1	0.63	*0.61*	*0.49*	*0.32*	0.67	0.63	0.38	0.62	0.68	0.70
ψ_2	0.23	**0.33**	**0.73**	**1.23**	0.11	0.20	0.96	0.27	0.06	0.01
ρ_r	0.87	*0.88*	*0.90*	*0.92*	0.86	0.87	0.92	0.87	0.86	0.85
ρ_g	0.66	0.66	0.66	0.65	*0.66*	*0.65*	*0.64*	0.66	0.66	0.66
ρ_z	0.60	0.61	0.63	0.60	*0.57*	*0.60*	*0.66*	0.61	0.58	0.57
σ_r	0.27	*0.27*	*0.28*	*0.28*	*0.27*	*0.27*	*0.28*	0.27	0.27	0.26
σ_g	0.58	0.58	0.59	0.60	*0.56*	*0.60*	*0.62*	0.58	0.57	0.56
σ_z	0.62	0.61	0.57	0.51	**0.72**	*0.35*	*0.20*	0.62	0.65	0.69
$M_{r\epsilon}$	0.53	0.51	0.47	0.45	*0.57*	*0.52*	*0.44*	0.52	0.57	0.60
$M_{g\epsilon}$	-0.06	-0.06	-0.04	-0.02	*-0.08*	*-0.04*	*0.01*	-0.06	-0.08	-0.10
$M_{z\epsilon}$	0.26	0.26	0.27	0.35	*0.23*	**0.76**	**1.26**	0.26	0.27	0.27
σ_ϵ	0.19	0.181	0.174	0.179	*0.190*	*0.001*	*0.001*	0.183	0.186	0.183

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$, where θ_0 contains parameter values of the default specification. All values are rounded to the second decimal place except for β and σ_ϵ. The bold value signifies the binding constraint. The italicized values signify that the parameter belongs to the constrained subset.

Table S8. KL and empirical distances between θ_c and θ_0, AS (2007) model, subset constraints

	(a) Monetary policy			(b) Exogenous shocks			(c) Behavioral		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	3.06E-06	1.23E-04	3.32E-04	3.58E-05	2.75E-04	1.14E-03	5.19E-07	1.57E-05	7.57E-05
T=80	0.0523	0.0646	0.0767	0.0583	0.0751	0.1119	0.0510	0.0553	0.0621
T=150	0.0531	0.0714	0.0898	0.0616	0.0866	0.1446	0.0513	0.0574	0.0672
T=200	0.0536	0.0755	0.0980	0.0636	0.0939	0.1661	0.0515	0.0586	0.0703
T=1000	0.0585	0.1228	0.1998	0.0842	0.1819	0.4446	0.0534	0.0710	0.1041

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$ with θ_c given in the columns of Table S7. The empirical distance measure equals $p_{ff}(\theta_0, \theta_c, 0.05, T)$, where T is specified in the last four rows of the Table.

Table S9. Parameter values minimizing the KL criterion

Indeterminacy, LS (2004) model							
		(a) All parameters can vary			(b) $M_{g\epsilon}$ fixed		
	θ_0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
τ	0.69	0.69	0.69	0.70	0.69	0.70	0.72
β	0.997	0.998	0.999	0.999	0.953	0.999	0.999
κ	0.77	0.77	0.77	0.77	0.78	0.85	0.93
ψ_1	0.77	0.77	0.77	0.78	0.76	0.74	0.71
ψ_2	0.17	0.17	0.16	0.13	0.18	0.31	0.47
ρ_r	0.60	0.60	0.60	0.60	0.60	0.61	0.61
ρ_g	0.68	0.68	0.68	0.68	0.68	0.66	0.64
ρ_z	0.82	0.82	0.82	0.82	0.82	0.83	0.83
σ_r	0.23	0.23	0.23	0.23	0.23	0.24	0.24
σ_g	0.27	0.27	0.29	0.31	0.27	0.29	0.32
σ_z	1.13	1.13	1.13	1.14	1.16	1.07	1.02
ρ_{gz}	0.14	0.15	0.11	0.07	0.13	0.17	0.20
$M_{r\epsilon}$	-0.68	-0.69	-0.64	-0.59	-0.66	**-1.18**	**-1.68**
$M_{g\epsilon}$	1.74	**1.84**	**1.24**	**0.74**	1.74	1.74	1.74
$M_{z\epsilon}$	-0.69	-0.70	-0.65	-0.61	-0.66	-0.79	-0.89
σ_ϵ	0.20	0.13	0.39	0.50	**0.10**	0.16	0.01

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$. All values are rounded to the second decimal place except for β. The bold value signifies the binding constraint.

Table S10. KL and empirical distances between θ_c and θ_0

Indeterminacy, LS (2004) model						
	(a) All parameters can vary			(b) $M_{g\epsilon}$ fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	4.06E-07	1.22E-05	5.41E-05	5.45E-06	1.10E-03	3.90E-03
T=80	0.0507	0.0550	0.0604	0.0520	0.1084	0.2021
T=150	0.0510	0.0569	0.0646	0.0531	0.1392	0.2926
T=200	0.0512	0.0580	0.0671	0.0538	0.1593	0.3518
T=1000	0.0529	0.0686	0.0941	0.0604	0.4219	0.8739

Note. KL is defined as $KL_{ff}(\theta_0, \theta_c)$ with θ_c given in the columns of Table S9. The empirical distance measure equals $p_{ff}(\theta_0, \theta_c, 0.05, T)$, where T is specified in the last four rows of the Table.

Table S11. Parameter values minimizing the KL criterion

Determinacy, LS (2004) model							
		(a) All parameters can vary			(b) ψ_2 fixed		
	θ_0^D	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
τ	0.54	0.54	0.55	0.55	0.55	0.56	0.57
β	0.992	0.900	0.900	0.900	0.900	0.900	0.900
κ	0.58	0.62	0.62	0.62	0.63	0.65	0.67
ψ_1	2.19	**2.09**	1.71	1.23	**2.29**	**2.69**	**3.19**
ψ_2	0.30	**0.40**	**0.80**	**1.30**	0.30	0.30	0.30
ρ_r	0.84	0.84	0.84	0.84	0.84	0.86	0.88
ρ_g	0.83	0.83	0.83	0.83	0.83	0.84	0.84
ρ_z	0.85	0.85	0.85	0.85	0.85	0.85	0.85
σ_r	0.18	0.18	0.18	0.18	0.18	0.18	0.19
σ_g	0.18	0.18	0.18	0.18	0.18	0.19	0.19
σ_z	0.64	0.64	0.64	0.64	0.64	0.64	0.64
ρ_{gz}	0.36	0.36	0.36	0.36	0.35	0.32	0.29

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$. All values are rounded to the second decimal place except for β. The bold value signifies the binding constraint.

Table S12. KL and empirical distances between θ_c^D and θ_0^D,

Determinacy, LS (2004) model						
	(a) All parameters can vary			(b) ψ_2 fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	1.06E-07	1.28E-05	6.26E-05	1.05E-04	2.10E-03	6.20E-03
T=80	0.0504	0.0544	0.0602	0.0650	0.1479	0.2729
T=150	0.0506	0.0563	0.0648	0.0713	0.2017	0.4011
T=200	0.0507	0.0573	0.0675	0.0751	0.2373	0.4810
T=1000	0.0515	0.0683	0.0970	0.1178	0.6541	0.9659

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$ with θ_c^D given in the columns of Table S11. The empirical distance measure equals $p_{ff}(\theta_0^D, \theta_c^D, 0.05, T)$, where T is specified in the last four rows of the Table.

Table S13. Empirical distances by altering the Taylor rule parameters, LS(2004) model

	ψ_1	ψ_2	ρ_r	σ_r	$\lVert\theta_j^D - \theta_0^D\rVert$	KL	$T = 80$	$T = 1000$
θ_0^D	2.1900	0.3000	0.8400	0.1800	–	–	–	–
				(a) Direction 1				
θ_1^D	2.1194	0.3720	0.8399	0.1800	0.1009	1.23E-06	0.0516	0.0555
θ_2^D	2.0486	0.4442	0.8398	0.1800	0.2019	5.17E-06	0.0533	0.0617
θ_3^D	1.9779	0.5163	0.8397	0.1801	0.3029	1.22E-05	0.0552	0.0688
θ_4^D	1.9072	0.5884	0.8396	0.1801	0.4039	2.28E-05	0.0572	0.0769
θ_5^D	1.8365	0.6605	0.8395	0.1801	0.5049	3.75E-05	0.0593	0.0863
θ_6^D	1.7658	0.7326	0.8394	0.1802	0.6059	5.70E-05	0.0617	0.0969
θ_7^D	1.6951	0.8048	0.8392	0.1802	0.7069	8.19E-05	0.0642	0.1092
θ_8^D	1.6244	0.8769	0.8391	0.1803	0.8079	1.13E-04	0.0670	0.1233
θ_9^D	1.5537	0.9490	0.8390	0.1803	0.9089	1.52E-04	0.0700	0.1395
θ_{10}^D	1.4830	1.0212	0.8389	0.1804	1.0099	1.98E-04	0.0733	0.1583
				(b) Direction 2				
θ_1^D	2.2193	0.2701	0.8400	0.1800	0.0419	1.99E-07	0.0505	0.0520
θ_2^D	2.2487	0.2401	0.8401	0.1800	0.0839	7.83E-07	0.0511	0.0541
θ_3^D	2.2781	0.2101	0.8401	0.1800	0.1259	1.73E-06	0.0516	0.0562
θ_4^D	2.3075	0.1801	0.8402	0.1800	0.1679	3.02E-06	0.0521	0.0583
θ_5^D	2.3369	0.1501	0.8402	0.1800	0.2099	4.64E-06	0.0526	0.0604
θ_6^D	2.3664	0.1201	0.8403	0.1800	0.2519	6.56E-06	0.0531	0.0626
θ_7^D	2.3958	0.0901	0.8403	0.1800	0.2939	8.77E-06	0.0536	0.0648
θ_8^D	2.4252	0.0601	0.8404	0.1800	0.3359	1.12E-05	0.0541	0.0670
θ_9^D	2.4546	0.0302	0.8404	0.1799	0.3779	1.40E-05	0.0546	0.0692
θ_{10}^D	2.4840	0.0002	0.8405	0.1799	0.4199	1.70E-05	0.0551	0.0714

Note. θ_j^D represent equally spaced points taken from the curve determined by the smallest eigenvalue from changing the four parameters in the monetary policy rule. The curve is extended from θ_0^D along two directions. Along Direction 1, the curve is truncated when $\|\theta_j^D - \theta_0^D\|$ exceeds 1. Along Direction 2, the curve is truncated at the closest point to zero where ψ_2 is still positive. KL is defined as $KL_{ff}(\theta_0^D, \theta_j^D)$. The last two columns are empirical distance measures defined as $p_{ff}(\theta_0^D, \theta_j^D, 0.05, T)$.

Table S14. Parameter values miminizing the KL criterion, SW(2007) model, L_1 norm

		(a) All parameters can vary			(b) φ fixed		
	θ_0^D	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
ρ_{ga}	0.52	0.52	0.52	0.52	0.52	0.52	0.53
μ_w	0.84	0.84	0.84	0.84	0.84	0.84	0.83
μ_p	0.69	0.69	0.69	0.69	0.69	0.68	0.70
α	0.19	0.19	0.19	0.20	0.19	0.18	0.20
ψ	0.54	0.54	0.53	0.52	0.54	0.54	0.54
φ	5.74	**5.78**	**5.97**	**6.21**	5.74	5.74	5.74
σ_c	1.38	1.38	1.38	1.37	1.38	1.37	1.43
λ	0.71	0.71	0.71	0.72	0.71	0.72	0.70
ϕ_p	1.60	1.60	1.60	1.61	1.60	1.59	1.61
ι_w	0.58	0.58	0.57	0.57	0.58	0.59	0.56
ξ_w	0.70	0.70	0.70	0.71	0.70	0.71	0.69
ι_p	0.24	0.24	0.24	0.25	0.24	0.23	0.25
ξ_p	0.66	0.66	0.66	0.66	0.66	0.66	0.66
σ_l	1.83	1.84	1.89	1.95	**1.86**	1.95	1.59
r_π	2.04	2.04	2.02	2.00	2.04	2.05	2.08
$r_{\Delta y}$	0.22	0.22	0.22	0.21	0.22	0.22	0.22
r_y	0.08	0.08	0.08	0.08	0.08	0.08	0.08
ρ	0.81	0.81	0.81	0.81	0.81	0.81	0.81
ρ_a	0.95	0.95	0.95	0.95	0.95	0.95	0.95
ρ_b	0.22	0.22	0.22	0.22	0.22	0.22	0.23
ρ_g	0.97	0.97	0.97	0.97	0.97	0.97	0.97
ρ_i	0.71	0.71	0.71	0.70	0.71	0.71	0.71
ρ_r	0.15	0.15	0.15	0.15	0.15	0.15	0.15
ρ_p	0.89	0.89	0.89	0.89	0.89	0.89	0.89
ρ_w	0.96	0.96	0.96	0.96	0.96	0.96	0.96
σ_a	0.45	0.45	0.45	0.45	0.45	0.45	0.45
σ_b	0.23	0.23	0.23	0.23	0.23	0.23	0.23
σ_g	0.53	0.53	0.53	0.53	0.53	0.53	0.53
σ_i	0.45	0.45	0.45	0.45	0.45	0.45	0.45
σ_r	0.24	0.24	0.24	0.24	0.24	0.24	0.24
σ_p	0.14	0.14	0.14	0.14	0.14	0.14	0.14
σ_w	0.24	0.24	0.24	0.24	0.24	0.24	0.24
$\overline{\gamma}$	0.43	0.42	0.38	0.34	0.45	0.56	0.20
$100(\beta^{-1}-1)$	0.16	0.18	0.24	0.31	**0.13**	**0.01**	**0.45**

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$. The bold values signify parameters that move the most.

Table S15. KL and empirical distances between θ_c^D and θ_0^D, SW(2007) model, L_1 norm

	(a) All parameters can vary			(b) φ fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	2.86E-06	7.05E-05	2.77E-04	4.89E-06	1.18E-04	5.30E-04
T=80	0.0525	0.0632	0.0785	0.0530	0.0662	0.0892
T=150	0.0533	0.0682	0.0905	0.0541	0.0730	0.1077
T=200	0.0538	0.0712	0.0980	0.0548	0.0771	0.1195
T=1000	0.0585	0.1039	0.1884	0.0611	0.1238	0.2707

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$ with θ_c^D given in the columns of Table S14. The empirical distance equals $p_{ff}(\theta_0^D, \theta_c^D, 0.05, T)$, where T is specified in the last four rows of the Table.

Table S16. Parameter values miminizing the KL criterion, SW(2007) model, L_2 norm

		(a) All parameters can vary			(b) φ fixed		
	θ_0^D	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
ρ_{ga}	0.52	0.52	0.52	0.52	0.52	0.52	0.52
μ_w	0.84	0.84	0.84	0.84	0.84	0.85	0.85
μ_p	0.69	0.69	0.69	0.69	0.69	0.68	0.68
α	0.19	0.19	0.19	0.19	0.19	0.18	0.18
ψ	0.54	0.54	0.53	0.53	0.54	0.53	0.52
φ	5.74	**5.84**	**6.24**	**6.74**	5.74	5.74	5.74
σ_c	1.38	1.38	1.38	1.38	1.37	1.35	1.32
λ	0.71	0.71	0.71	0.72	0.71	0.72	0.73
ϕ_p	1.60	1.60	1.61	1.61	1.60	1.59	1.59
ι_w	0.58	0.58	0.57	0.57	0.58	0.58	0.58
ξ_w	0.70	0.70	0.70	0.71	0.70	0.72	0.74
ι_p	0.24	0.24	0.24	0.24	0.24	0.24	0.24
ξ_p	0.66	0.66	0.66	0.66	0.66	0.66	0.66
σ_l	1.83	1.84	1.85	1.87	1.87	**2.27**	**2.79**
r_π	2.04	2.04	2.02	2.01	2.04	2.03	2.02
$r_{\Delta y}$	0.22	0.22	0.22	0.22	0.22	0.21	0.20
r_y	0.08	0.08	0.08	0.08	0.08	0.08	0.08
ρ	0.81	0.81	0.81	0.81	0.81	0.81	0.81
ρ_a	0.95	0.95	0.95	0.95	0.95	0.95	0.95
ρ_b	0.22	0.22	0.22	0.22	0.22	0.21	0.21
ρ_g	0.97	0.97	0.97	0.97	0.97	0.97	0.97
ρ_i	0.71	0.71	0.70	0.70	0.71	0.71	0.71
ρ_r	0.15	0.15	0.15	0.15	0.15	0.14	0.14
ρ_p	0.89	0.89	0.89	0.89	0.89	0.89	0.89
ρ_w	0.96	0.96	0.96	0.96	0.96	0.96	0.96
σ_a	0.45	0.45	0.45	0.45	0.45	0.45	0.45
σ_b	0.23	0.23	0.23	0.23	0.23	0.23	0.23
σ_g	0.53	0.53	0.53	0.53	0.53	0.53	0.53
σ_i	0.45	0.45	0.45	0.45	0.45	0.45	0.45
σ_r	0.24	0.24	0.24	0.24	0.24	0.24	0.24
σ_p	0.14	0.14	0.14	0.14	0.14	0.14	0.14
σ_w	0.24	0.24	0.24	0.24	0.24	0.24	0.24
$\overline{\gamma}$	0.43	0.43	0.42	0.42	0.49	0.61	0.64
$100(\beta^{-1}-1)$	0.16	0.17	0.20	0.23	**0.09**	0.01	0.01

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$. The bold value signifies the parameter that moves the most.

Table S17. KL and empirical distances between θ_c^D and θ_0^D, SW(2007) model, L_2 norm

	(a) All parameters can vary			(b) φ fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	8.06E-06	1.84E-04	6.62E-04	2.07E-05	5.23E-04	1.80E-03
T=80	0.0538	0.0705	0.0939	0.0562	0.0894	0.1365
T=150	0.0553	0.0795	0.1157	0.0586	0.1078	0.1843
T=200	0.0562	0.0850	0.1297	0.0601	0.1196	0.2159
T=1000	0.0646	0.1499	0.3112	0.0747	0.2697	0.6010

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$ with θ_c^D given in the columns of Table S16. The empirical distance equals $p_{ff}(\theta_0^D, \theta_c^D, 0.05, T)$, where T is specified in the last four rows of the Table.

Table S18. Parameter values miminizing KL, SW(2007) model, weighted constraints

	(a) β and $\overline{\gamma}$ fixed				(b) β, $\overline{\gamma}$ and r_π fixed			(c) β, $\overline{\gamma}$, r_π, σ_l fixed		
	c	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5	1.0
ρ_{ga}	0.52	0.52	0.53	0.53	0.52	0.52	0.52	0.52	0.51	0.51
μ_w	0.84	0.84	0.83	0.83	0.84	0.85	0.86	0.84	0.84	0.84
μ_p	0.69	0.69	0.69	0.69	0.69	0.68	0.68	0.69	0.70	0.70
α	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
ψ	0.54	0.54	0.55	0.56	0.54	0.52	0.50	0.54	0.54	0.53
φ	5.74	5.69	5.54	5.41	**6.09**	5.89	6.06	**6.08**	**7.47**	**9.19**
σ_c	1.38	1.39	1.43	1.48	1.38	1.34	1.32	1.38	1.38	1.39
λ	0.71	0.71	0.70	0.69	0.71	0.73	0.74	0.71	0.72	0.72
ϕ_p	1.60	1.60	1.59	1.59	1.60	1.59	1.59	1.60	1.61	1.62
ι_w	0.58	0.58	0.59	0.59	0.58	0.57	0.57	0.58	0.57	0.56
ξ_w	0.70	0.70	0.69	0.69	0.70	0.74	0.77	0.70	0.71	0.71
ι_p	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.25
ξ_p	0.66	0.66	0.65	0.65	0.66	0.66	0.67	0.66	0.67	0.67
σ_l	1.83	1.81	1.75	1.67	1.84	**2.77**	**3.70**	1.83	1.83	1.83
r_π	2.04	**2.10**	**2.33**	**2.63**	2.04	2.04	2.04	2.04	2.04	2.04
$r_{\Delta y}$	0.22	0.22	0.23	0.24	0.22	0.20	0.19	0.22	0.22	0.22
r_y	0.08	0.08	0.10	0.13	0.08	0.08	0.08	0.08	0.08	0.08
ρ	0.81	0.81	0.83	0.85	0.81	0.82	0.82	0.81	0.81	0.82
ρ_a	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
ρ_b	0.22	0.22	0.23	0.24	0.22	0.21	0.21	0.22	0.22	0.22
ρ_g	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
ρ_i	0.71	0.71	0.72	0.72	0.70	0.70	0.70	0.70	0.69	0.67
ρ_r	0.15	0.15	0.14	0.14	0.15	0.13	0.12	0.15	0.15	0.14
ρ_p	0.89	0.89	0.89	0.89	0.89	0.89	0.88	0.89	0.89	0.89
ρ_w	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
σ_a	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
σ_b	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
σ_g	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53
σ_i	0.45	0.45	0.46	0.46	0.45	0.45	0.45	0.45	0.45	0.45
σ_r	0.24	0.24	0.24	0.25	0.24	0.24	0.23	0.24	0.24	0.24
σ_p	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
σ_w	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
$\overline{\gamma}$	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
$100(\frac{1}{\beta}-1)$	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16

Note. The constraint is given by $\{\theta_c : |(\theta_c - \theta_0)./w(\theta_0)|_\infty \geq c\}$, where $w(\theta_0)$ contains the lengths of the 90% credible sets from SW(2007). The bold value signifies the binding constraint.

Table S19. KL and empirical distances, SW(2007) model, weighted constraints

	(a) β and $\overline{\gamma}$ fixed			(b) β, $\overline{\gamma}$ and r_π fixed			(c) β, $\overline{\gamma}$, r_π, σ_l fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	6.81E-05	1.36E-03	4.22E-03	9.64E-05	1.74E-03	4.83E-03	9.67E-05	1.77E-03	5.12E-03
T=80	0.0605	0.1133	0.1950	0.0643	0.1346	0.2299	0.0643	0.1341	0.2333
T=150	0.0653	0.1500	0.2891	0.0703	0.1812	0.3377	0.0703	0.1812	0.3462
T=200	0.0682	0.1744	0.3513	0.0739	0.2120	0.4072	0.0739	0.2124	0.4188
T=1000	0.0994	0.4914	0.8896	0.1142	0.5898	0.9282	0.1142	0.5952	0.9399

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$ with θ_c^D given in the columns of Table S18. The empirical distance equals $p_{ff}(\theta_0^D, \theta_c^D, 0.05, T)$, where T is specified in the last four rows of the Table.

Table S20. Parameter values miminizing the KL criterion, SW(2007) model, subset constraints

		(a) Monetary policy			(b) Exogenous shocks			(c) Behavioral parameters		
	c	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5	1.0
ρ_{ga}	0.52	0.52	0.53	0.47	*0.52*	*0.52*	0.56	0.52	0.52	0.52
μ_w	0.84	0.84	0.83	0.80	*0.85*	*0.86*	*0.81*	0.84	0.84	0.84
μ_p	0.69	0.69	0.69	0.75	**0.59**	**0.19**	*0.68*	0.69	0.69	0.69
α	0.19	0.19	0.19	0.22	0.19	0.19	0.19	*0.19*	*0.19*	*0.19*
ψ	0.54	0.54	0.56	0.52	0.54	0.54	1.00	*0.54*	*0.54*	*0.53*
φ	5.74	5.66	5.43	2.79	5.70	5.73	2.00	**5.84**	**6.24**	**6.74**
σ_c	1.38	1.40	1.46	1.36	1.38	1.38	2.28	*1.38*	*1.38*	*1.38*
λ	0.71	0.71	0.69	0.60	0.71	0.71	0.52	*0.71*	*0.71*	*0.72*
ϕ_p	1.60	1.60	1.59	1.78	1.60	1.59	1.69	*1.60*	*1.61*	*1.61*
ι_w	0.58	0.58	0.59	0.99	0.59	0.60	0.63	*0.58*	*0.58*	*0.57*
ξ_w	0.70	0.70	0.69	0.50	0.70	0.71	0.63	*0.70*	*0.70*	*0.71*
ι_p	0.24	0.24	0.24	0.24	0.18	0.02	0.26	*0.24*	*0.24*	*0.24*
ξ_p	0.66	0.66	0.65	0.61	0.66	0.68	0.58	*0.66*	*0.66*	*0.66*
σ_l	1.83	1.81	1.71	0.52	1.87	1.90	1.03	*1.83*	*1.85*	*1.86*
r_π	2.04	**2.14**	**2.54**	*3.00*	2.05	2.05	3.00	2.04	2.02	2.01
$r_{\Delta y}$	0.22	*0.22*	*0.24*	*0.50*	0.22	0.22	0.29	0.22	0.22	0.22
r_y	0.08	*0.09*	*0.12*	*0.28*	0.08	0.08	0.17	0.08	0.08	0.08
ρ	0.81	*0.82*	*0.85*	*0.50*	0.81	0.81	0.85	0.81	0.81	0.81
ρ_a	0.95	0.95	0.95	0.94	*0.95*	*0.95*	*0.96*	0.95	0.95	0.95
ρ_b	0.22	0.22	0.23	0.40	*0.22*	*0.22*	*0.37*	0.22	0.22	0.22
ρ_g	0.97	0.97	0.97	0.96	*0.97*	*0.97*	*0.97*	0.97	0.97	0.97
ρ_i	0.71	0.71	0.72	0.75	*0.71*	*0.71*	*0.99*	0.71	0.70	0.70
ρ_r	0.15	*0.15*	*0.14*	*0.85*	*0.15*	*0.15*	*0.16*	0.15	0.15	0.15
ρ_p	0.89	0.89	0.89	0.99	*0.87*	*0.81*	*0.91*	0.89	0.89	0.89
ρ_w	0.96	0.96	0.96	0.98	*0.96*	*0.96*	*0.95*	0.96	0.96	0.96
σ_a	0.45	0.45	0.45	0.43	*0.45*	*0.45*	*0.43*	0.45	0.45	0.45
σ_b	0.23	0.23	0.23	0.16	*0.23*	*0.23*	*0.17*	0.23	0.23	0.23
σ_g	0.53	0.53	0.53	0.54	*0.53*	*0.53*	*0.56*	0.53	0.53	0.53
σ_i	0.45	0.45	0.46	0.52	*0.45*	*0.45*	**1.45**	0.45	0.45	0.45
σ_r	0.24	*0.24*	*0.24*	**1.24**	*0.24*	*0.24*	*0.27*	0.24	0.24	0.24
σ_p	0.14	0.14	0.14	0.08	*0.13*	*0.10*	*0.14*	0.14	0.14	0.14
σ_w	0.24	0.24	0.24	0.31	*0.24*	*0.25*	*0.25*	0.24	0.24	0.24
$\overline{\gamma}$	0.43	0.45	0.48	0.65	0.43	0.40	0.10	*0.43*	*0.43*	*0.42*
$100(\beta^{-1}-1)$	0.16	0.14	0.09	0.20	0.13	0.12	0.01	*0.17*	*0.19*	*0.22*

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$. The bold value signifies the binding constraint. The italicized values signify that the parameter belongs to the constrained subset.

Table S21. KL and empirical distances between θ_c^D and θ_0^D, SW(2007) model, subset constraints

	(a) Monetary policy			(b) Exogenous shocks			(c) Behavioral parameters		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	1.86E-04	3.25E-03	6.21E-01	1.00E-03	1.05E-02	1.26E-01	8.15E-06	1.86E-04	6.66E-04
T=80	0.0683	0.1677	0.9983	0.1109	0.3937	0.7609	0.0539	0.0706	0.0941
T=150	0.0771	0.2432	1.0000	0.1413	0.5760	0.8392	0.0553	0.0796	0.1159
T=200	0.0825	0.2936	1.0000	0.1610	0.6759	0.8756	0.0562	0.0852	0.1299
T=1000	0.1469	0.8079	1.0000	0.4161	0.9981	0.9959	0.0646	0.1505	0.3123

Note. KL is defined as $KL_{ff}(\theta_0^D, \theta_c^D)$ with θ_c^D given in the columns of Table S20. The empirical distance equals $p_{ff}(\theta_0^D, \theta_c^D, 0.05, T)$, where T is specified in the last four rows of the Table.