Using Arbitrary Precision Arithmetic to Sharpen Identification Analysis for DSGE Models*

Zhongjun Qu† Boston University

Denis Tkachenko‡ National University of Singapore

December 19, 2019

Abstract

This paper is at the intersection of macroeconomics and modern computer arithmetic. It seeks to apply arbitrary precision arithmetic to resolve practical difficulties arising in the identification analysis of log linearized DSGE models. The main focus is on methods in Qu and Tkachenko (2012, 2017) since the framework appears to be the most comprehensive to date. Working with this arithmetic, we develop the following three-step procedure for analyzing local and global identification. (1) The DSGE model solution algorithm is modified so that all the relevant objects are computed as multiprecision entities allowing for indeterminacy. (2) The rank condition and the Kullback-Leibler distance are computed using arbitrary precision Gauss-Legendre quadrature. (3) Minimization is carried out by combining double precision global and arbitrary precision local search algorithms, where the criterion for convergence is set based on the chosen precision level, so that it can be effectively examined whether the minimized value equals zero. In an application to a model featuring monetary and fiscal policy interactions (Leeper, 1991 and Tan and Walker, 2015), we find that the arithmetic removes all ambiguity in the analysis. As a result, we reach clear conclusions showing observational equivalence both within the same policy regime and across different policy regimes under generic parameter values. We further illustrate the application of the method to medium scale DSGE models by considering the model of Schmitt-Grohé and Uribe (2012), where the use of extended precision again helps remove ambiguity in cases where near observational equivalence is detected.

Keywords: Arbitrary precision arithmetic, dynamic stochastic general equilibrium models, local identification, global identification.

JEL classification: C1, E1.

*We thank Pavel Holoborodko and seminar and conference participants at the 30th (EC)^2 Conference on Identification in Macroeconomics, NBER-NSF Time Series Conference, Frontiers in Econometrics Conference (Connecticut), Identification in Econometrics Conference (Vanderbilt), Academia Sinica, Columbia, and HKUST for helpful comments and suggestions.

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (qu@bu.edu).

‡Department of Economics, National University of Singapore, 1 Arts Link, 117570, Singapore (ecstd@nus.edu.sg).

1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models have become a widely applied modeling framework both in academia and in policy institutions due to important insights about the consequences of monetary, fiscal and macro-prudential policies as well as economic forecasts obtained through their lens. DSGE models typically feature various frictions, often involving a relatively large number of equations and parameters with complex cross-equation restrictions. Identification and computation are central issues in developing DSGE models. This paper seeks to introduce arbitrary precision arithmetic to this literature in order to resolve important practical difficulties arising in the identification analysis of DSGE models.

Identification analysis in DSGE models can potentially touch on the following four issues. Firstly, within a given DSGE model, are there parameter values close to the default parameter value that can generate observational equivalence (i.e., local identification)? Secondly, within a model, can there be parameter values that are potentially distant from the default value that can lead to observational equivalence (global identification)? Thirdly, comparing two different DSGE model structures (e.g., models with different policy rules), can there exist different parameter values that make them observationally equivalent (identification allowing for different model structures)? Lastly, in the above three situations, can there be parameter values that lead to near observational equivalence, i.e., lead to models that are hard to tell apart based on sample sizes typically encountered in practice (weak identification)?

Substantial theoretical progresses have been made in answering these issues. Canova and Sala (2009) documented the types of identification failure that can arise in DSGE models. Iskrev (2010) gave sufficient conditions, while Komunjer and Ng (2011) and Qu and Tkachenko (2012) gave necessary and sufficient conditions for local identification. Qu and Tkachenko (2017) developed a framework for addressing the second to the fourth issues allowing for indeterminacy. In particular, they considered the frequency domain expression for the Kullback-Leibler distance between two DSGE models and showed that global identification fails if and only if the minimized distance equals zero. Further, they proposed a measure for the empirical closeness between two DSGE models for assessing the strength of the identification. This measure represents the highest possible power when testing a default DSGE model specification against a local alternative under Gaussianity. The measure is related to but different from Cox's (1961) test for non-nested hypotheses; Pesaran (1987) provided a fairly thorough study of the local asymptotic power properties of Cox's test.

Implementing the above theoretical results requires at least three of the following steps: (i) solving the DSGE model; (ii) constructing a criterion function for identification; (iii) evaluating whether

the criterion function is rank deficient, and (iv) minimizing a distance function and determining whether the minimized value is zero. Among them, (i), (iii) and (iv) always require numerical operations for DSGE models that are relevant in practice. These operations introduce numerical uncertainty that can blur the difference between weak identification and lack of identification. From the model builder's perspective, it is important to distinguish the two cases, as exact identification failure found numerically may point to useful analytical results such as those we document later in the paper about equivalence between different policy regimes. This distinction is also important, as a starting point, if one wishes to further quantify the strength of the identification within or between different models.

Progresses have been made toward reducing the numerical uncertainty. Iskrev (2010) provided closed-form expressions for first order derivatives of covariances of observables with respect to structural parameters. Nevertheless, to determine these covariances, a Lyapunov equation still needs to be solved numerically. As a result, the procedure still faces challenges (i) and (iii). Komunjer and Ng (2011) experimented with using different step sizes to compute the numerical derivatives in (ii), and with different tolerance levels when determining the rank in (iii). However, occasionally the result can still be ambiguous. Qu and Tkachenko (2012, 2017) suggested to use non-identification curves to check local identification and also conducted sensitivity checks. However, the same concern as in Komunjer and Ng (2011) still lingers. So far, it has remained unclear whether the numerical uncertainty can be effectively eliminated in practice, which may prevent the identification analysis from being effectively carried out in applied research.

This paper develops a solution by utilizing arbitrary precision arithmetic, also commonly referred to as multiple precision or bignum arithmetic. It involves operating on numbers whose precision digits are in principle limited only by the memory resources of the host system. This contrasts with the fixed-precision arithmetic natively supported by most modern processors, which typically offers up to 64 bits of precision. Working with this arithmetic, we develop the following three-step procedure for analyzing local and global identification. Step 1: The DSGE model solution algorithm is modified so that all the relevant objects are computed as multiprecision entities. Step 2: The rank condition and the Kullback-Leibler distance are computed using arbitrary precision symmetric difference quotient rule and Gauss-Legendre quadrature. Step 3: Minimization is carried out using double precision hybrid global search (such as `Matlab`'s *ga* and *particleswarm* followed by *fmincon*) algorithms followed by arbitrary precision local search algorithm (`Matlab`'s *fminsearch*). The criterion for convergence is set based on the chosen precision level, so that whether the minimized value is zero can be evaluated effectively.

We developed a `Matlab` code with the aid of the Multiprecision Computing Toolbox by Ad-

vanpix. The code can run on a usual desktop computer with `Matlab` version 2016b or later. To apply the code to the four issues outlined above, the researchers mainly need to write down the log linearized model, the default parameter values, the tolerance levels, and the step sizes for numerical differentiation. Beside requiring the multiprecision toolbox, the implementation is similar to what it would be under the standard double precision environment. We provide a thorough illustration of the steps involved and will make the code available. The code can be useful for solving DSGE models with user specified precision when the interest is not in identification analysis.

We apply the procedure to a small scale DSGE model featuring monetary and fiscal policy interactions. The model is a cashless version of Leeper (1991) analyzed in Tan and Walker (2015). It exhibits three stable regimes characterized by: i) Active monetary and passive fiscal policy (AMPF, there is a unique equilibrium, i.e., determinacy); ii) Passive monetary and active fiscal policy (PMAF, determinacy); and iii) Passive monetary and fiscal policies (PMPF, there are a continuum of equilibria, i.e., indeterminacy). The model has analytical solutions under both determinacy and indeterminacy, which we make use of to examine to what extent the arbitrary precision arithmetic sharpens the identification analysis. That is, for each regime, we first obtain results using the standard double precision arithmetic, then apply the arbitrary precision arithmetic, and finally use the analytical solutions to compare the results and convey the intuition behind them. The following results are worth noting. i) The parameters are not locally identified in the two determinate policy regimes, while they are locally identified in the indeterminate regime if the sunspot variance is not zero. ii) When each of the two determinate regimes is taken as the default specification, we always find parameter values in the indeterminate regimes that generate observational equivalence. This implies that, if the data were generated by this model, and if the monetary and fiscal authorities operated according to either the AMPF or the PMAF rule, then we would be unable to rule out the PMPF rule possibility even with an infinite sample size. iii) We also trace out parameter values in the PMPF regime that generate the equivalence. The values demonstrate how the parameters can be altered without causing any change in the dynamic properties of the model. The analysis here further advances the work by Leeper (1991), Leeper and Walker (2013), and Tan and Walker (2015) about potential observational equivalence between different fiscal-monetary regimes. Comparison with their results is included in Subsection 5.5.

The procedure we develop is directly applicable to medium scale DSGE models, which is illustrated with the model of Schmitt-Grohé and Uribe (2012). The salient feature of the model is that its exogenous shock processes contain anticipated, or news, components, which have been argued to explain a large portion of business cycle fluctuations. The model does not have an analytical solution, and hence the researcher must rely on numerical results when conducting identification

analysis. The following results are worth noting. i) We find that the model is locally identified at the posterior median reported in Schmitt-Grohé and Uribe (2012), but there are small eigenvalues that can potentially be labelled zero, and hence arbitrary precision computation is helpful in verifying this. ii) The global identification analysis within the model structure points to weak identification of some of the news shock parameters. iii) We conduct additional analysis and document differences in the relative importance of news shocks by process and by horizon in terms of generating the model dynamics. We find that the 8-horizon shocks tend to be in general less important, and that news in the wage markup plays the most important role. iv) We further examine whether a different model structure featuring unanticipated moving average shocks can be effectively distinguished from the news benchmark. The findings suggest that, although imposing the alternative structure on all shock processes at once results in a model sufficiently distinct to tell apart in finite samples, doing so for individual shocks may produce near observational equivalence. Somewhat surprisingly in light of earlier results, it is found that such near equivalence is obtained for the model with unanticipated moving average wage markup shocks.

Besides identification analysis, the arithmetic may find applications in other contexts where numerical stability is of high importance, such as solving nonlinear DSGE models, inverting large matrices, and solving Lyapunov equations. These issues are all of substantial importance to macroeconomic research. In implementation, the researcher can switch on the arbitrary precision arithmetic when higher resolution is needed, and switch off when the double precision is sufficient. This simplicity permits achieving an acceptable balance between precision and speed of computation.

The paper proceeds as follows. Section 2 describes identification procedures in Qu and Tkachenko (2012, 2017). It also explains where the numerical issues for identification analysis arise. Section 3 gives an easily accessible and self-contained explanation of the idea behind the arithmetic with relevant resources in order to facilitate its adoption in macroeconomics. Section 4 explains how the arbitrary precision arithmetic can be applied to identification analysis to deliver clearer conclusions. Sections 5 and 6 provide the applications. Section 7 concludes. The three online appendices contain some additional details on derivation, model solution, and results from the two applications.

2 Identification analysis allowing for indeterminacy

This section outlines the identification procedures in Qu and Tkachenko (2012, 2017). The results themselves are not new. However, without first explaining them in some detail, the subsequent analysis (i.e., how the arbitrary precision arithmetic operates inside these results and the empirical applications) will be confusing. We first describe the underlying framework, and then discuss the conditions for local and global identification, together with the implementation details.

2.1 The framework

Consider a DSGE model log linearized around its steady state (Sims, 2002):

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

where S_t is a vector that includes the endogenous variables, the conditional expectations, and variables from exogenous shock processes if they are serially correlated. The vector ε_t contains serially uncorrelated structural shocks and η_t contains expectation errors. The elements of Γ_0, Γ_1, Ψ and Π are known functions of structural parameters of the model. Depending on Γ_0 and Γ_1, the system can have none, a unique, or multiple stable solutions (indeterminacy). Under indeterminacy, the structural parameters alone do not uniquely determine the dynamics of the observables. Below we briefly explain the steps Qu and Tkachenko (2017) take to overcome this obstacle, paying attention to the numerical issues that can arise.

Lubik and Schorfheide (2003) show that the full set of solutions can be represented as

$$S_t = \Theta_1 S_{t-1} + \Theta_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t, \tag{2}$$

or equivalently, $S_t = (1-\Theta_1 L)^{-1}[\Theta_\varepsilon \;\; \Theta_\epsilon](\varepsilon_t' \;\; \epsilon_t')'$, where L is the lag operator. For completeness, the appendix outlines the main steps of the solution method. This helps to reveal the main numerical issues involved and to better understand the arbitrary precision code that we make available.

In (2), Θ_1, Θ_ε and Θ_ϵ depend only on Γ_0, Γ_1, Ψ and Π, therefore are functions of the structural parameters only. The vector ϵ_t contains the sunspot shocks. The DSGE model alone imposes few restriction on ϵ_t: it needs to be a martingale difference, i.e., $E_t \epsilon_{t+1} = 0$, however can be arbitrarily contemporaneously correlated with the fundamental shocks ε_t. Intuitively, the properties of ϵ_t depend on how agents form their expectations, which, under indeterminacy, is not fully revealed by the model. To reflect this, Qu and Tkachenko (2017) adopt the following parameterization that expresses ϵ_t as an orthogonal projection onto ε_t and a residual term: $\epsilon_t = M\varepsilon_t + \widetilde{\epsilon}_t$, where M is a matrix of constants and $\widetilde{\epsilon}_t$ is now uncorrelated with ε_t. Let θ^D be a p-by-1 vector consisting of all the structural parameters in (1). Let θ^U be a q-by-1 vector consisting of the sunspot parameters $\theta^U = (\text{vec}\,(\Sigma_\epsilon)', \text{vec}\,(M)')'$. Then, an augmented parameter vector can be denoted as $\theta = (\theta^{D\prime}, \theta^{U\prime})'$.

In practice, the estimation is typically based on a subset of S_t or some linear transformations involving its current and lagged values. To be consistent with this, let $A(L)$ denote a matrix of finite order lag polynomials to specify the observables and write $Y_t(\theta) = A(L)S_t = H(L;\theta)(\varepsilon_t' \;\; \epsilon_t')'$, where $H(L;\theta) = A(L)(1-\Theta_1 L)^{-1}[\Theta_\varepsilon \;\; \Theta_\epsilon]$. Then, the spectral density of $Y_t(\theta)$ is given by

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

where * denotes the conjugate transpose and

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

The elements of $f_\theta(\omega)$ can be considered as mappings from the augmented parameter space to a space of complex valued functions defined over $[-\pi, \pi]$. Local or global identification holds if and only if the overall mapping is locally or globally injective. Henceforth, we let $\{Y_t\}$ denote a stochastic process whose spectral density is given by $f_{\theta_0}(\omega)$ with $\omega \in [-\pi, \pi]$.

Numerical issues. When solving the model to obtain (2) from (1), the important steps are the generalized Schur decomposition (see (A.1)) and the singular value decomposition (see (A.3)). These steps are done numerically, thus contributing to the numerical uncertainty when computing the solution, and ultimately the identification criterion functions. These two numerical issues are not particular to the current identification method. They affect all methods that apply the generalized Schur decomposition and the singular value decompositions to solve the model.

2.2 Local identification

The parameter vector θ is said to be locally identifiable from the second order properties of $\{Y_t\}$ at $\theta = \theta_0$ 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 $\theta_1 = \theta_0$. Intuitively, if this holds, then it is potentially possible to locally determine both the parameters describing technology and preferences (θ^D) and those governing the equilibrium beliefs of agents (θ^U).

Define

$$G(\theta) = \int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} f_\theta(\omega)}{\partial \theta'} \right)^* \left(\frac{\partial \operatorname{vec} f_\theta(\omega)}{\partial \theta'} \right) d\omega \tag{4}$$

and assume θ_0 is a regular point, that is, $G(\theta)$ has a constant rank in an open neighborhood of θ_0. Then, Qu and Tkachenko (2017) show that θ is locally identifiable from the second order properties of $\{Y_t\}$ at $\theta = \theta_0$ if and only if $G(\theta_0)$ has full rank. If the regular point assumption is dropped, then the nonsingularity of $G(\theta_0)$ is sufficient but not necessary for local identification.

Some useful extensions follow immediately from the above result. Firstly, one could be interested in the identification of a subset of θ holding the rest fixed at some values, e.g., of θ^D conditional on $\theta^U = \theta_0^U$. Then, when computing the condition, we only need to take the derivative with respect to the parameters in the subset of interest. Moreover, if $G(\theta_0)$ has already been computed for some θ_0, investigating conditional local identification does not bring about any further computational cost: it suffices to examine the rank of the submatrix of $G(\theta_0)$ formed by the intersection of the rows and

columns of $G(\theta_0)$ corresponding to the parameters of interest. Secondly, one could be interested in studying identification using a subset of frequencies, e.g., those corresponding to business cycle fluctuations. The relevant result obtains by modifying the integral in (4):

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

where $W(\omega)$ is an indicator function symmetric about zero to select the frequencies of interest. Thirdly, one could be interested in studying identification using both the steady state restrictions and the dynamic properties, in which case one should replace $G(\theta)$ by

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

where $\mu(\theta)$ is a vector of steady states.

If local identification failure is detected, one could be interested in which parameters, or parameter subsets, are not separately identifiable. Qu and Tkachenko (2012) show that the number of zero eigenvalues of $G(\theta)$ implies that observational equivalence can be maintained by varying some parameters along the same number of orthogonal directions, and hence corresponds to the minimum number of parameters that need to be fixed in order to achieve local identification. In order to pinpoint the nonidentified parameter subsets, they propose a procedure that involves recursive application of the conditional local identification analysis outlined above and can be summarized in four steps. Step 1: check whether $G(\theta)$ has any zero eigenvalues, and proceed to the next step if identification failure is present. Step 2: Apply the conditional identification analysis to each individual parameter. If a zero eigenvalue is found, then the corresponding parameter is not locally identified. Step 3: Increase the number of parameters in the considered parameter subsets by one, and single out those that satisfy: i) they do not contain any previously detected subset as a proper subset; ii) the corresponding submatrix of $G(\theta)$ has only one zero eigenvalue. Such subsets will have the following property: varying the parameters jointly in a particular way would maintain observational equivalence and thus they are not separately identified. Fixing one parameter in such a subset will identify the rest. Step 4: repeat Step 3 until all possible subsets are considered. Since this procedure does not enforce orthogonality, it is possible to find more nonidentified parameter subsets than there are zero eigenvalues of $G(\theta)$. This analysis could be taken further by tracing out nonidentification curves corresponding to each detected subset that show how the parameters inside the subset change in the neighborhood of θ_0 to preserve observational equivalence, see Qu and Tkachenko (2012, Subsection 3.1) for more details.

Numerical issues. Two numerical issues are present when implementing the local identification conditions. First, the derivatives need to be computed numerically, involving the choice of a step

size and approximation method. Second, the integral needs to be approximated using, say, the Gaussian quadrature method. Note that if an eigenvalue of $G(\theta)$ is zero, then the error from Gaussian quadrature alone will not make it nonzero. In this sense, the quadrature introduces approximation error only when the eigenvalue is not zero.

Some further numerical issues arise when interpreting the result. The matrix $G(\theta)$ is real, symmetric and positive semidefinite by construction, hence its rank can be determined by counting the number of nonzero eigenvalues. In practice, the computed eigenvalues are usually not exactly zero, and hence a certain tolerance level must be chosen, so that the eigenvalues falling below this level are considered zero. There is no generally accepted way to choose tolerance in the literature, but computational software often provides certain rules of thumb. For example, the default tolerance level for determining matrix rank in `Matlab` depends on the properties of the matrix - it equals the product of the larger dimension of the matrix and the positive distance from its largest singular value to the next largest floating point number of the same precision. Using the nonidentification subset and curve analysis discussed above for verifying the robustness of rank determination, Qu and Tkachenko (2012) found that, for a suitably chosen numerical differentiation step, using this default tolerance level generally delivers sensible results. Hence, we will also adopt the same default tolerance level for the baseline analysis under double precision.

However, at times the results could be unclear due to some eigenvalues being close to the tolerance level from either side, or the tolerance level not being appropriate given the structure of the matrix, e.g., as documented by Komunjer and Ng (2011) for the case of sparse matrices. We will show how such situations can be effectively dealt with using arbitrary precision arithmetic. Note that when some of the true eigenvalues of $G(\theta)$ equal zero, it is possible to obtain small negative numbers of the order of magnitude close to the roundoff error despite the theoretical positive semidefiniteness property of $G(\theta)$. This is a well known phenomenon and is due to some degree of numerical error inherent in any eigenvalue computation routine. When we obtain such results, we report the small negative eigenvalues as is.

2.3 Global identification

This section considers the global identification of θ at θ_0 allowing for different model structures. Relevant examples of the two structures can correspond to two DSGE models with different monetary policy rules, models with different determinacy properties, or models of different scale. Global identification within the same model structure can be considered as a special case.

Suppose $Y_t(\theta)$ and $Z_t(\phi)$ are two vector linear processes generated by two DSGE structures (Structures 1 and 2) with spectral densities $f_\theta(\omega)$ and $h_\phi(\omega)$, where $\theta \in \Theta, \phi \in \Phi$, and Θ and Φ

are finite dimensional and compact. Suppose we treat Structure 1 with $\theta = \theta_0$ as the benchmark specification and are interested in whether Structure 2 can generate the same dynamic properties. As a definition, we say that Structure 2 is distinct from Structure 1 at $\theta = \theta_0$ if, for any $\phi \in \Phi$, $h_\phi(\omega) \neq f_{\theta_0}(\omega)$ for some $\omega \in [-\pi, \pi]$. Define

$$KL_{fh}(\theta, \phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \{\text{tr}(h_\phi^{-1}(\omega) f_\theta(\omega)) - \log\det(h_\phi^{-1}(\omega) f_\theta(\omega)) - n_Y\} d\omega.$$

Then, Qu and Tkachenko (2017) prove that Structure 2 is distinct from Structure 1 at $\theta = \theta_0$ if and only if $\inf_{\phi \in \Phi} KL_{fh}(\theta_0, \phi) > 0$. Similarly to local identification analysis, one can consider global identification based on a subset of frequencies. In this case, we just need to replace $KL_{fh}(\theta_0, \phi)$ by

$$KL_{fh}^W(\theta, \phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} W(\omega) \{\text{tr}(h_\phi^{-1}(\omega) f_\theta(\omega)) - \log\det(h_\phi^{-1}(\omega) f_\theta(\omega)) - n_Y\} d\omega. \tag{5}$$

When studying global identification within the same model structure, let

$$KL(\theta_0, \theta_1) = KL_{ff}(\theta_0, \theta_1) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \{\text{tr}(f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)) - \log\det(f_{\theta_1}^{-1}(\omega) f_{\theta_0}(\omega)) - n_Y\} d\omega. \tag{6}$$

Qu and Tkachenko (2017) show that θ is globally identified from the second order properties of $\{Y_t\}$ at θ_0 if and only if $KL(\theta_0, \theta_1) > 0$ for any $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$.

The above results reduce the problem of checking global identification to minimizing a deterministic function. For illustration, suppose the condition in the previous subsection shows that θ_0 is locally identified. Then, to study global identification within the same model structure, we proceed to check whether

$$\inf_{\theta_1 \in \Theta \setminus B(\theta_0)} KL(\theta_0, \theta_1) > 0, \tag{7}$$

where $B(\theta_0)$ is an open neighborhood of θ_0 that serves two purposes. First, it excludes parameter values that are arbitrarily close to θ_0. Second, its shape and size can be varied to examine the sensitivity of identification. For example, one can examine how identification improves when successively larger neighborhoods are excluded or when some parameters are fixed.

Numerical issues. The implementation encounters two numerical issues. First, the integral needs to be approximated. We continue to use Gaussian quadrature. As in the local identification case, if the true distance is zero, then the error from Gaussian quadrature alone will not make it nonzero. Second, the optimization needs to be done numerically with a specified tolerance level, and hence one needs to decide on whether the minimized value of KL is zero or only close to zero. This is similar to determining whether the eigenvalue is zero in the local identification case, but facing more uncertainty due to the usage of a convergence tolerance level.

2.4 Empirical distance between two models

When global identification holds, there may still exist parameter values difficult to distinguish when faced with finite sample sizes. For example, Del Negro and Schorfheide (2008) observed that the data provide similar support for a model with moderate price rigidity and trivial wage rigidity, and the one in which both rigidities are high. More generally, even models with different structures (e.g., different policy rules) can produce data dynamics that are quantitatively similar. This section describes a measure developed by Qu and Tkachenko (2017) for gauging the feasibility of distinguishing a model structure with spectral density $h_{\phi_0}(\omega)$ from a structure with $f_{\theta_0}(\omega)$ constrained by a hypothetical sample size T. The measure reflects the feasibility of distinguishing one model from another using likelihood ratio tests based on a finite number of observations generated by the two models. It represents the highest power under Gaussianity when considering local alternatives.

Define $V_{fh}(\theta_0,\phi_0) = 1/(4\pi)\int_{-\pi}^{\pi}\mathrm{tr}\{[I - f_{\theta_0}(\omega)h_{\phi_0}^{-1}(\omega)][I - f_{\theta_0}(\omega)h_{\phi_0}^{-1}(\omega)]\}d\omega$ and $V_{hf}(\phi_0,\theta_0) = 1/(4\pi)\int_{-\pi}^{\pi}\mathrm{tr}\{[I - h_{\phi_0}(\omega)f_{\theta_0}^{-1}(\omega)][I - h_{\phi_0}(\omega)f_{\theta_0}^{-1}(\omega)]\}d\omega$. Let $Z \sim N(0,1)$, and

$$q_\alpha = -T^{1/2}KL_{fh}(\theta_0,\phi_0) + \sqrt{V_{fh}(\theta_0,\phi_0)}z_{1-\alpha},$$

where $z_{1-\alpha}$ is the $100(1-\alpha)$th percentile of Z. The empirical distance measure of h from f equals

$$p_{fh}(\theta_0,\phi_0,\alpha,T) = \Pr\left(Z > \frac{q_\alpha - T^{1/2}KL_{hf}(\phi_0,\theta_0)}{\sqrt{V_{hf}(\phi_0,\theta_0)}}\right).$$

To obtain it, the main work is in computing $KL_{fh}(\theta_0,\phi_0)$, $KL_{hf}(\phi_0,\theta_0)$, $V_{fh}(\theta_0,\phi_0)$ and $V_{hf}(\phi_0,\theta_0)$. They depend only on the spectral densities $f_{\theta_0}(\omega)$ and $h_{\phi_0}(\omega)$ without any reference to any data. Computing them thus only requires solving the two models once to compute the respective spectral densities. No simulation is required.

As with the KL criterion, the empirical distance measure can be applied to structures with overlapping but different sets of observables. This, in particular, permits measuring the distance between a small and a medium scale DSGE model. It can also be computed using a subset of frequencies. In this case, $KL_{fh}(\theta_0,\phi_0)$ and $KL_{hf}(\phi_0,\theta_0)$ need to be replaced with $KL_{fh}^W(\theta_0,\phi_0)$ and $KL_{hf}^W(\phi_0,\theta_0)$ (see (5) for their definition) and $V_{fh}(\theta_0,\phi_0)$ and $V_{hf}(\phi_0,\theta_0)$ with $V_{fh}^W(\theta_0,\phi_0) = 1/(4\pi)\int_{-\pi}^{\pi}W(\omega)\,\mathrm{tr}\{[I - f_{\theta_0}(\omega)h_{\phi_0}^{-1}(\omega)][I - f_{\theta_0}(\omega)h_{\phi_0}^{-1}(\omega)]\}d\omega$ and $V_{hf}^W(\phi_0,\theta_0) = 1/(4\pi)\int_{-\pi}^{\pi}W(\omega)\,\mathrm{tr}\{[I - h_{\phi_0}(\omega)f_{\theta_0}^{-1}(\omega)][I - h_{\phi_0}(\omega)f_{\theta_0}^{-1}(\omega)]\}d\omega$, respectively. This extension is valuable as it can decompose to what extent the quantitative differences between models are driven by, for example, the business cycle frequencies as opposed to others.

Numerical issues. They include matrix inversion, computing integrals, and computing a tail probability with respect to a standard normal distribution.

3 Introduction to arbitrary precision arithmetic

This section presents an accessible and self-contained explanation of the idea behind this arithmetic, along with relevant resources to facilitate its wider adoption in macroeconomics. This will not only pave the way for applying the arithmetic to DSGE models, but also show why it can be useful for other applications in macroeconomics, such as accurately solving nonlinear DSGE models.

3.1 The basic idea

On modern computers, real numbers are represented using a floating point arithmetic system. This subsection examines the factors determining the precision of such a system, and then explains how they can be engineered to achieve arbitrary precision. The discussion mainly follows Kneusel (2017, Chapter 3) and Muller et al. (2010, Chapter 2), which can be consulted for more details.

A floating point number system F is a subset of the real numbers whose elements have the form

$$(-1)^s d_0.d_1d_2...d_{p-1} \times \beta^e, \tag{8}$$

where $d_0.d_1d_2...d_{p-1}$ is the p-digit *significand* with $0 \leq d_i < \beta$ for all $i = 0, ..., p-1$, representing the value $d_0 + d_1\beta^{-1} + d_2\beta^{-2} + ... + d_{p-1}\beta^{-(p-1)}$. Note that $s, d_0, d_1, ..., d_{p-1}, \beta$ and e are all integers.

This system is thus characterized by four integers: the base $\beta \geq 2$; the precision p, which controls the number of bits allocated to store the significand; the exponent range $e_{\min}$ and $e_{\max}$ such that $e_{\min} \leq e \leq e_{\max}$. The most commonly used base is 2, as it is efficient on binary computers. Henceforth, we will only consider $\beta = 2$. This means that all the numbers are binary and the only allowed digits in the significand are 0 and 1. The first digit d_0 is required to be 1 to make the representation unique. An equivalent (normalized) representation of (8) is therefore given by

$$(-1)^s 1.d_1d_2...d_{p-1} \times \beta^e,$$

which implies that to store a number, it suffices to save the integers s, $d_1d_2...d_{p-1}$ and e.

For general values of $e_{\min}, e_{\max}$ and p, the range of positive floating point numbers that can be represented is given by $[2^{e_{\min}}, 2^{e_{\max}}(1 - 2^{1-p})]$. The lower bound is obtained by letting $d_1 = d_2 = ... = d_{p-1} = 0$ and $e = e_{\min}$. The upper bound is obtained by letting $d_1 = d_2 = ... = d_{p-1} = 1$ and $e = e_{\max}$. It is important to note that floating point numbers do not cover all real numbers. For a positive floating point number with exponent e, the next larger floating point number is different from it by $2^{1-p}2^e$. This shows that the distribution of floating point numbers on the real line is not uniform. It also follows that for any real number in the range of F, it can always be approximated by an element of F with a *relative* rounding error no larger than

$$u = (1/2)2^{1-p}. \tag{9}$$

The value $\epsilon_M = 2^{1-p}$ is often referred to as the *machine epsilon*.

It is now clear what determines the precision of the system: (1) Because $e_{\min}$ and $e_{\max}$ are finite, numbers that are too large or too small in absolute values will have to be approximated. (2) Because p is finite, the spacing between two floating point numbers is always nonzero. Therefore, the precision can be improved by increasing the exponent range $[e_{\min}, e_{\max}]$ and by increasing p. Consider an example of storing 0.001956 on a computer as a floating point number. Suppose $p = 3$, and $e_{\min} = -8$, then the closest floating point number to it is $1.00 \times 2^{-8} \approx 0.003906$. The difference is about 0.001950, roughly of the same magnitude as the original number. If $e_{\min}$ is reduced to -9, then the closest floating point number is $1.00 \times 2^{-9} \approx 0.001953$. The difference is reduced to 0.000003. If p is further increased to 11, then the closest floating point number is $1.0000000001 \times 2^{-9} \approx$0.001955. The difference is further reduced to 0.000001.

The above description can be used to understand the IEEE 754, the universally accepted standard for floating point arithmetic, which was created in 1985 and revised in 2008 (IEEE, 2008). Most software commonly used in economic research, e.g., `Matlab`, `R` and `Stata`, use the IEEE 754 binary double precision (binary64) as default. In this standard, the first of the 64 bits is allocated to the sign $s \in \{0, 1\}$, the next 11 digits are allocated to store the exponent e in excess-1023 format (that is, instead of e, $e + 1023$ is stored using 11 digits taking values between $2^0 = 1$ and $2^{11} - 1 = 2047$; as a result, $e_{\min} = -1022$ and $e_{\max} = 1023$), and the final 52 bits are used to store $d_1 d_2 ... d_{52}$ (therefore, $p = 53$). The corresponding maximum relative roundoff error is $u = 2^{1-53}/2 \approx$ 1.1102E-16, so that the roundoff level is about 16 decimal digits. Here, the group of 64 bits form what is called a "word" - the number of bits that the computer can process in a single operation. Therefore, every number is stored using one word.

Higher precision arithmetic involves specifying p to values higher than 53 used in the IEEE 754 double format and widening the exponent range. For example, IEEE 754 also specifies the quadruple precision as a basic binary format where $p = 113$, $e_{\min} = -16382$, $e_{\max} = 16383$, and the roundoff error is approximately 9.6296E-35 yielding approximately 34 decimal digits. Further increasing the number of significand bits p would result in increased range and smaller roundoff errors. There is no practical limit to increasing the precision except the one implied by the available memory of the machine the program runs on. This gives rise to the so called arbitrary precision arithmetic where the user is able to set p to achieve the desired level of accuracy.

Remark 1 *In practice, the researcher can first decide the desired decimal digits for the computation. Then, the corresponding precision parameter p can be determined by solving*

$$floor(p \times log_{10}(2)) = desired\ decimal\ digits. \tag{10}$$

This operation, though simple, entails an important conceptual change. That is, the computing precision is no longer a constraint that one has to take as given, but an input that the researcher can choose, and experiment with, to better serve the task at hand.

There are four immediate consequences from increasing p. First, a number is now expressed using more than one word. Second, the computer needs to keep track of more than one address in order to access one number. Third, operations such as addition need to operate on an increased number of digits. Finally, some standard double precision operations such as addition and multiplication are implemented at the hardware level through the processor's integrated floating point unit. With a larger p, software instead of hardware implementations for these operations are needed. These four factors increase memory requirements and also potentially slow down the operations. Fast algorithms for implementation are thus essential for the practicality of the arithmetic.

3.2 Software implementation

Software tools are currently actively developed to tackle the above mentioned computational cost challenge. The GNU Multiple Precision Arithmetic Library (GMP) and its extensions (GNU MPFR and others) is by far the most successful free library for arbitrary precision arithmetic. It is available at `https://gmplib.org/`. The basic interface is for `C` but wrappers exist for other languages including `Julia`, `Python`, `R`, and the Wolfram Language. As stated there, GMP operates on signed integers, rational numbers, and floating point numbers. It provides fast algorithms for multiplication, division, raising to power, factorial, etc.; see the "Algorithms" section of the manual.

The GMP libraries are often used as foundations for building more user oriented software applications. These softwares bridge the gap between basic arbitrary precision arithmetic operations (such as addition of two numbers) and the more specific user applications (such as inversion of a matrix or numerical integration). For example, `mpmath` is a free (BSD licensed) `Python` toolbox for arbitrary precision floating point arithmetic, developed by Fredrik Johansson since 2007. Among other things, it supports numerical integration and differentiation, Fourier and Taylor series approximation, and linear algebra functions such as matrix inverse and singular value decomposition. `Rmpfr` is an `R` package available from CRAN since August 2009 that supports the use of arbitrary precision numbers with some `R` functions through the interface with GNU MPFR. The functionality of these software packages is expected to substantially expand with further development.

In our implementation, we utilize the Multiprecision Computing Toolbox (MPC) for Matlab by Advanpix. (A free 1-week trial of the toolbox can be downloaded at www.advanpix.com.) To our knowledge, this is so far the most feature rich software implementing arbitrary precision arithmetic. Although `Matlab` has a built-in Variable Precision Arithmetic (VPA) toolbox, it lacks the support

for some functions crucial for obtaining our results and is substantially slower. In particular, the MPC toolbox supports the fully arbitrary precision versions of the main numerical aspects of our computations: 1) the complex generalized Schur decomposition that is necessary for solving the DSGE model; 2) efficient numerical integration techniques such as Gaussian quadrature; 3) optimization using Nelder-Mead simplex method (overloading `Matlab`'s *fminsearch* function). The MPC toolbox uses quadruple precision (34 decimal digits) by default and most of its functions are heavily optimized at this level to the extent that they sometimes execute faster than the corresponding `Matlab` commands in double precision. We always start with this default level of precision in our applications. Then, we alter the precision level, e.g., to 50 digits by executing the command mp.Digits(50), where 50 refers to decimal digits and the binary precision p is as in (10).

The MPC toolbox defines a new numeric object type *mp* in the `Matlab` environment that stores values in the specified precision format. Once the object is defined as *mp* type, all further operations applied to it and supported by the toolbox, from basic arithmetic to most commonly used `Matlab` functions, will be automatically performed with their overloaded arbitrary precision versions. The *mp* objects can be constructed by passing a string with a mathematical expression to the *mp()* function, e.g., mp('pi'), or converting an existing `Matlab` matrix or expression of any numeric type, e.g., mp(pi). The important caveat to note is that objects previously stored as double and converted to *mp* will not have the same accuracy as floating point numbers calculated directly in arbitrary precision. E.g., in the example above, conversion from double (mp(pi)) yields a less accurate answer than direct computation (mp('pi')):

$$\begin{aligned} \text{mp(pi)} &= 3.141592653589793\ \ 115997963468544185, \\ \text{mp('pi')} &= 3.141592653589793\ \ 238462643383279503, \end{aligned}$$

where the difference can be seen after the 15th decimal digit. Because of the above features, the existing `Matlab` code can be ported to arbitrary precision by redefining inputs as *mp* objects and in some cases precision independent code can be produced by utilizing a wrapper function that unifies processing of different numeric types.

We emphasize that the use of software environments other than `Matlab` should be eventually feasible. This is important to note because the scope of potential uses for arbitrary precision arithmetic in macroeconomics can be much wider than the applications in this paper.

4 Arbitrary precision arithmetic and identification analysis

As highlighted in Section 2, three numerical steps reside within local identification analysis: (i) The model is solved numerically. (ii) The integral is approximated. (iii) The derivative is computed

numerically. For the global analysis, the first two are still present, in addition: (iv) the convergence of the minimization is up to some tolerance level. We now discuss these aspects in more detail and explain how we address them using arbitrary precision arithmetic.

4.1 Model solution

All the available methods for solving empirically relevant DSGE models involve some numerical steps. Here, as in Qu and Tkachenko (2012, 2017), we build the analysis on the GENSYS algorithm of Sims (2002) with the Lubik and Schorfheide's (2003) generalization to allow for indeterminacy. As seen from the Appendix A, starting with the representation (1), the main steps of the solution algorithm are as follows. (i) Apply the generalized Schur decomposition to transform the system to have an upper triangular structure, c.f. Display (A.1). (ii) Apply the singular value decomposition and reduced column echelon form to isolate the effective restrictions imposed by the model on the conditional expectation errors η_t, c.f. (A.3). (ii) Solve a linear system of equations with η_t as unknowns to obtain its entire set of solutions, c.f., (A.6). This expresses η_t as a function of the structural shocks and sunspot shocks if there are any. (iv) Plug the expression for η_t into the triangular system and then reverse Step (i) to express the state vector S_t in terms of its lagged value, structural shocks and sunspot shocks. Note that Steps (ii) and (iv) involve only simple matrix inversions and multiplications. Among the software mentioned above, only the Advanpix toolbox has implemented the generalized Schur decomposition in multiple precision. This appears to be the only constraint precluding the use of other software for local identification analysis.

We will make the `Matlab` code for solving a DSGE model to any precision available. The user can provide a desired precision level. Anderson (2008) compared the numerical precision of different solution methods under double precision. Here, the constraint is lifted and the solution precision becomes an input that can be altered by the user. Fernández-Villaverde, Rubio-Ramírez and Schorfheide (2016) discussed the importance of solution precision. They in fact pointed to the possibility of using GNU multiple precision arithmetic library to achieve any arbitrary level of accuracy, but without providing details on implementation.

4.2 Integration

The local and global analysis both require computing an integral over a compact set $[-\pi, \pi]$. We approximate the integral using Gauss-Legendre quadrature. Generally, the integral $I(f) = \int_{-\pi}^{\pi} f(x)dx$ is approximated as $\pi \sum_{j=1}^{n} w_j f(\pi x_j)$, where $w_j = 2/\{(1-x_j^2)[P_n'(x_j)]^2\}$, and x_j $(j = 1, ..., n)$ are the n roots of the n-th order Legendre polynomial $P_n(x)$.

The integral is exact for polynomials of order not exceeding $2n-1$. In the general case, its

precision depends on n, the abscissas x_j and weights w_j. Note that these values are independent of the function to integrate. We pre-compute these values once, save them as multiple precision entities, and then load the values to compute the summation as needed.

It is useful to consider what happens to the approximation when the parameters are indeed not identified. In this case, $f(x) = 0$ for all x belonging to $[-\pi, \pi]$. Assuming there was no error in computing $f(x)$, even if the abscissas x_j and weights w_j were imprecisely computed, we would still have $\pi \sum_{i=1}^{n} w_i f(\pi x_j) = 0$. Therefore, the increased precision helps only when the parameters are weakly identified. Higher precision can better reveal that the value is indeed distinct from zero. Because of its wide usage, high precision computation of the quadrature is readily available in several language environments. For example, it is also implemented in the *mpmath* `Python` library.

4.3 Differentiation

The local identification requires computing derivatives $\partial \operatorname{vec} f_\theta(\omega)/\partial\theta_j$ for $j = 1, 2, ..., q$. In implementation, we mainly use the symmetric difference quotient method. This involves solving the model at $\theta - h\mathrm{e}_j$ and $\theta + h\mathrm{e}_j$ and compute the derivative as $[\operatorname{vec} f_{\theta+h\mathrm{e}_j}(\omega) - \operatorname{vec} f_{\theta-h\mathrm{e}_j}(\omega)]/(2h)$ for $j = 1, 2, ..., q$. With a step size h, the error is of rate $O(h^2)$. We also experiment with a four point method, for which the error rate is $O(h^5)$. Previous experimentations with double precision in Qu and Tkachenko (2012) suggest to set h to values between 1E-6 and 1E-7. Here, when higher precision is used, we can further decrease h to improve derivative accuracy. When quadruple precision is used, we set h =1E-10. This makes the error from differentiation of orders O(E-20) and O(E-50) respectively. When an n-point method is used, the model will need to be solved $n \times q$ times, where q stands for the number of parameters. Thus, the computational cost increases linearly with n.

4.4 Optimization

As discussed in Qu and Tkachenko (2017), the minimization of $KL(\theta_0, \theta)$ over θ is a first order challenge to the suggested approach. It faces two difficulties: $KL(\theta_0, \theta)$ may have multiple local minima and the dimension of θ can be high. Meanwhile, the problem has two desirable features, i.e., $KL(\theta_0, \theta)$ is a deterministic function of θ and it is typically infinitely differentiable with respect to it. These two features help make global optimization possible.

For implementation, we first carry out the minimization in two steps under double precision. The first step conducts global searches with gradient free methods. This permits exploring wide parameter regions under multimodality. These searches return a range of parameter values that correspond to the regions where the values of $KL(\theta_0, \theta)$ are small. The second step applies multiple local searches, using the values returned from the first step along with additional uniformly ran-

domly generated initial values within the relevant parameter bounds. These local searches exploit the smoothness in $KL(\theta_0, \theta)$ and are efficient in locating the nearby locally optimal solution. In both steps, we employ optimizers that are parallelizable, thus maintaining computational feasibility even when the dimension of θ is high.

After these two steps in double precision are completed, further local minimization is carried out using *fminsearch* in multiple precision. The initial values are set to the outputs from the second step. The resulting optimizers are compared with their initial values. We check whether the KL values and the parameter vectors are nontrivially different. For the non-identified cases, the KL values are expected to be substantially smaller due to the increased precision. For the identified cases, the KL values should remain essentially the same. If the result is still not clear cut, then we further increase the precision level and repeat the optimization using *fminsearch*.

5 Application: a small scale model

The model is a cashless version of the Leeper (1991) model studied by Tan and Walker (2015) and can be solved analytically. The latter feature is useful because it can reveal to what extent the arbitrary precision arithmetic helps in sharpening the numerical analysis.

5.1 Model solutions under alternative policy regimes

The model describes an endowment economy with lump-sum taxes. A representative household chooses $\{c_t, B_t\}$ to solve $\max E_0 \sum_{t=0}^{\infty} \beta^t [\log(c_t)]$, subject to $c_t + B_t/P_t + \tau_t = y + R_{t-1}B_{t-1}/P_t$, where c_t is consumption, B_t is a bond that costs \$1 and pays $R_t = 1 + i_t$ dollars at $t+1$, y is a constant quantity of goods, τ_t is lump-sum taxes (or transfers if negative). The government chooses $\{B_t, \tau_t\}$ to satisfy the budget constraint:

$$\frac{B_t}{P_t} + \tau_t = \frac{R_{t-1}B_{t-1}}{P_t}. \tag{11}$$

Denote real debt by $b_t = B_t/P_t$ and the gross rate of inflation by $\pi_t = P_t/P_{t-1}$. After imposing the market clearing condition $c_t = y$, the household's Euler equation reduces to the Fisher relation:

$$\frac{1}{R_t} = \beta E_t \left(\frac{1}{\pi_{t+1}} \right). \tag{12}$$

The policy rules for setting interest rate R_t and taxes τ_t are specified as follows:

$$R_t = R^*(\pi_t/\pi^*)^{\alpha} \exp(\varepsilon_t^r) \text{ and } \tau_t = \tau^*(b_{t-1}/b^*)^{\gamma} \exp(\varepsilon_t^{\tau}), \tag{13}$$

where

$$\varepsilon_t^r = e_t^r + \phi_r e_{t-1}^r \text{ with } e_t^r \sim i.i.d.N(0, \sigma_r^2), \tag{14}$$

and

$$\varepsilon_t^\tau = e_t^\tau + \phi_\tau e_{t-1}^\tau \text{with } e_t^\tau \sim i.i.d.N(0, \sigma_\tau^2). \tag{15}$$

Linearizing (11), (12) and (13), and solving R_t and τ_t out of the system yields:

$$E_t\widehat{\pi}_{t+1} = \alpha\widehat{\pi}_t + \varepsilon_t^r, \tag{16}$$

$$\widehat{b}_t + \frac{\widehat{\pi}_t}{\beta} = (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} + \frac{\alpha}{\beta}\widehat{\pi}_{t-1} - (1/\beta - 1)\varepsilon_t^\tau + \frac{1}{\beta}\varepsilon_{t-1}^r, \tag{17}$$

where "$\widehat{\ }$" denotes the log deviation of a variable from its steady state value.

Henceforth, we study identification properties based on dynamic properties of $\{\widehat{b}_t, \widehat{\pi}_t\}$ specified by (16) and (17) with the shocks satisfying (14) and (15). As shown by Leeper (1991) and Tan and Walker (2015), the existence and uniqueness of solutions to this linearized system depend on the values of α and γ (they are assumed to be nonnegative throughout the paper):

(i) When $\alpha > 1$ and $\gamma > 1$, this corresponds to a regime with active monetary and passive fiscal policy (AMPF). There is determinacy; the solution exists and is unique.

(ii) When $\alpha < 1$ and $\gamma < 1$, this corresponds to a regime with passive monetary and active fiscal policy (PMAF). The solution exists and is unique.

(iii) When $\alpha < 1$ and $\gamma > 1$, this corresponds to a regime with passive monetary and fiscal policies (PMPF). There is indeterminacy; there is a continuum of solutions.

(iv) When $\alpha > 1$ and $\gamma < 1$, this corresponds to a regime with active monetary and fiscal policies. There exist no nonexplosive solutions.

Tan and Walker (2015) obtained analytical solutions for (i) and (ii), but not for (iii). Because we use a different way to solve the model, we provide analytical solutions for (i) to (iii). The derivation is in the appendix.

Case (i): $\alpha > 1$ and $\gamma > 1$ (AMPF). The unique solution is given by

$$\widehat{\pi}_t = \left(-\frac{1}{\alpha} - \frac{1}{\alpha^2}\phi_r\right) e_t^r - \frac{1}{\alpha}\phi_r e_{t-1}^r, \tag{18}$$

$$\widehat{b}_t = (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} + \frac{1}{\beta}\left[\left(\frac{1}{\alpha} + \frac{1}{\alpha^2}\phi_r\right) e_t^r\right] - (1/\beta - 1)\varepsilon_t^\tau.$$

In general, $\widehat{\pi}_t$ is an MA(1) process and $\widehat{b}_t$ an ARMA(1,1) process. However, as documented in Tan and Walker (2015), if we let $\phi_r = -\alpha$ and $\phi_\tau = -(1/\beta - \gamma(1/\beta - 1))$, then the solution becomes

$$\widehat{\pi}_t = e_{t-1}^r, \widehat{b}_t = -(1/\beta - 1)e_t^\tau. \tag{19}$$

Both $\widehat{\pi}_t$ and $\widehat{b}_t$ are now i.i.d. processes. Below, we will consider both generic parameter values and also this special set of parameter values.

Case (ii): $\alpha < 1$ and $\gamma < 1$ (PMAF). The unique solution is given by

$$\begin{aligned} \widehat{\pi}_t &= \alpha\widehat{\pi}_{t-1} - \beta(1/\beta - 1)\left(\frac{1}{1/\beta - \gamma(1/\beta - 1)}\phi_\tau + 1\right)e_t^\tau + \varepsilon_{t-1}^r, \\ \widehat{b}_t &= \frac{1/\beta - 1}{1/\beta - \gamma(1/\beta - 1)}\phi_\tau e_t^\tau. \end{aligned} \tag{20}$$

In general, $\widehat{b}_t$ is i.i.d. and $\widehat{\pi}_t$ is an ARMA(1,3) process. However, if we set $\phi_\tau = -(1/\beta - \gamma(1/\beta - 1))$ and $\phi_r = -\alpha$, then, again, $\widehat{\pi}_t$ and $\widehat{b}_t$ both follow i.i.d. processes as given by (19).

Case (iii): $\alpha < 1$ and $\gamma > 1$ (PMPF). There is a continuum of solutions in this regime. The entire set of solutions can be written as

$$\begin{aligned} \widehat{\pi}_t &= \alpha\widehat{\pi}_{t-1} + \varepsilon_{t-1}^r + \eta_t, \\ \widehat{b}_t &= (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} - (1/\beta - 1)\varepsilon_t^\tau - \frac{1}{\beta}\eta_t, \end{aligned}$$

where η_t is a sunspot shock that satisfies $E_{t-1}\eta_t = 0$ and can be correlated with e_t^r and e_t^τ. To model this, we consider projections of η_t onto e_t^r and e_t^τ : $\eta_t = M_r e_t^r + M_\tau e_t^\tau + \epsilon_t$, where ϵ_t is now uncorrelated with e_t^r and e_t^τ. Because of the potential correlation between η_t and e_t^r, $\widehat{\pi}_t$ is in general an ARMA(1,4) process. Because of the potential correlation between η_t and e_t^τ, $\widehat{b}_t$ is in general an ARMA(1,2) process. Therefore, under indeterminacy, the potential dynamics can be richer than under determinacy, a feature that is also documented in Lubik and Schorfheide (2004).

If we let $M_r = M_\tau = \sigma_\epsilon = 0$, $\phi_r = -\alpha$ and $\phi_\tau = -(1/\beta - \gamma(1/\beta - 1))$, then, again, the solution reduces to (19). This implies that we can have observational equivalence between the three policy regions under a particular set of parameter values. Meanwhile, whether there can be observational equivalence under generic parameter values is not known. This provides an opportunity to evaluate the algorithms' capability to detect such cases.

The analysis below is organized as follows. We first consider local identification, then global identification, and finally empirical distances between policy regimes. For each case, we first obtain results using the standard double precision arithmetic. Then, we apply the arbitrary precision arithmetic. In these two steps, the models and identification conditions are computed *numerically* using the algorithms described in the previous sections. Finally, we further validate the results and provide intuition for them in light of the analytical solutions. Note that in cases (i) and

(ii), the structural parameter vector is $\theta = [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau]$. In case (iii), it is given by $\theta = [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau, M_r, M_\tau, \sigma_\epsilon]$.

Presentation of results. Since more than a few results will be reported, we adopt the following structure for clarity. For each analysis, we first use a few sentences to highlight the main result (e.g., Result 1). Subsequently, we provide details on how this finding is reached.

5.2 Local Identification

We focus on the full spectrum case in the main paper and report the results based on the business cycle frequencies in the online appendix.

5.2.1 Local identification based on the full spectrum

We start with a generic parameter value, and then consider the special parameter value chosen according to Tan and Walker (2015) that leads to i.i.d. processes in all three regimes.

Case (i): AMPF. We begin with considering the following baseline generic parameter value

$$\theta^1_{AMPF} = [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau] = [1.5, 0.9804, 1.2, 0.5, 0.5, 1, 1],$$

where α, β and γ are set to values in Tan and Walker (2015), and $\phi_r, \phi_\tau, \sigma_r$ and σ_τ are set arbitrarily since the literature has not considered calibrations with MA shocks.

Result 1 *α, ϕ_r and σ_r are not separately identifiable; fixing any of them leads to local identification.*

Under double precision, the eigenvalues of $G(\theta^1_{AMPF})$ in decreasing order are 8.72E+06, 8941.019, 4.689, 0.005, 8.24E-07, 1.81E-09 and -6.46E-11. The two smallest eigenvalues are below the `Matlab` default tolerance level of 1.30E-08. This suggests that the rank equals 5 with 7 parameters in total. However, because the two eigenvalues 8.24E-07 and 1.81E-09 are close to the tolerance level, the result is not clear cut**.** Using quadruple precision, the eigenvalues of $G(\theta^1_{AMPF})$ are: 8.72E+06, 8941.019, 4.689, 0.005, 8.24E-07, 1.81E-09 and -1.70E-29. Note that the smallest eigenvalue drops significantly to -1.70E-29, while other eigenvalues all remain essentially unchanged. Therefore, the results show that the rank of the matrix in fact equals 6. The precision can be increased even further. We increase it to 50 digits and find that the first 6 eigenvalues remains virtually the same, while the smallest eigenvalue further shrinks to 2.32E-45. This reaffirms the conclusion.

The above result shows that there is either one unidentified parameter or one subset of parameters that are not separately identifiable. In order to pinpoint this parameter subset, we apply the

procedure outlined in Subsection 2.2. This analysis ultimately points to $(\alpha, \phi_r, \sigma_r)$. When all the other parameters are fixed, the smallest eigenvalue of the relevant submatrix of $G(\theta^1_{AMPF})$ equals 3.39E-12 under double precision. For the computation in quadruple precision, its magnitude is reduced to -9.20E-30. Therefore, the three parameters α, ϕ_r and σ_r are not separately identifiable, but fixing any one of them can lead to identification.

This identification feature can also be seen from the analytical solution. Looking at (18), we can see that the three parameters enter the solution only through two transformations: $\alpha^{-1}\phi_r\sigma_r$ and $\left(\alpha^{-1} + \alpha^{-2}\phi_r\right)\sigma_r$. Therefore, there is one degree of freedom to change the three parameters jointly without affecting the above two quantities, which explains the obtained result.

We repeat the above analysis using the 4-point rule for numerical derivative and find that the conclusions about identification are the same. This follows because the smallest eigenvalue is of similar magnitude and the tolerance level does not change. Therefore, using progressively higher precision resolves the uncertainty about the rank and sources of identification failure thus sharpening the conclusion about identification. Further, although the rank is computed only at θ^1_{AMPF}, once the nonidentified parameter subsets are pinned down and the mechanism behind identification failure is established, it follows that the lack of identification is a generic feature not confined to this parameter value. This suggests one way in practice how identification features at a single point can be generalized to other parameter values.

Now, consider the special parameter value for the first regime:

$$\theta^2_{AMPF} = [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau] = [1.5, 0.9804, 1.2, -1.5, -0.996, 1, 1].$$

Result 2 *θ^2_{AMPF} can be changed along three orthogonal directions with observational equivalence.*

In this case, the rank of $G(\theta^2_{AMPF})$ equals 4 in both double and quadruple precision, implying that three parameters need fixing to achieve local identification. In the former case, the eigenvalues equal 9.972, 0.598, 2.76E-04, 8.11E-07, 2.39E-15, 5.42E-17, 1.02E-18, of which the last three fall below the default tolerance of 1.24E-14. In the latter case, the largest four eigenvalues remain essentially the same at 9.972, 0.598, 2.76E-04, 8.11E-07, while the three smallest eigenvalues fall to -6.98E-34, -7.22E-35 and -1.66E-37. Increasing precision to 50 digits, we again see no discernible change in the eigenvalues above the tolerance level, while the three smallest eigenvalues are further reduced to 2.03E-40, 8.89E-42, -6.86E-52, thus confirming our conclusion.

This result shows that the parameters can be moved along each of three orthogonal directions with observational equivalence. We again apply the procedure from Subsection 2.2 to locate the relevant parameter subsets. The same tolerance level of 1.24E-14 is used throughout. The analysis pinpoints subsets (α, ϕ_r), (γ, ϕ_τ), $(\beta, \gamma, \sigma_\tau)$ and $(\beta, \phi_\tau, \sigma_\tau)$. In double precision, the smallest

eigenvalues pertaining to the submatrices are 4.44E-16 for (α, ϕ_r), 1.10E-18 for (γ, ϕ_τ), 4.35E-20 for $(\beta, \gamma, \sigma_\tau)$ and 2.48E-20 for $(\beta, \phi_\tau, \sigma_\tau)$. In quadruple precision, the respective values are 3.85E-34, 3.15E-39, -5.43E-41 and -4.56E-38. Note that the number of subsets exceeds 3 because orthogonality is not enforced. These parameter subsets can again be understood using the analytical solution. Note that the model remains unchanged as long as the conditions for the root cancellation $\phi_r = -\alpha$ and $\phi_\tau = -(1/\beta - \gamma(1/\beta - 1))$ are satisfied, and the values of $(1/\beta - 1)\sigma_\tau$ and σ_r stay the same. Moving ϕ_r and α in exactly opposite directions will satisfy these requirements. This leads to the first subset. Increasing γ by one unit and increasing ϕ_τ by $(1/\beta - 1)$ also satisfies this requirement. This leads to the second subset. Further, because the restriction $\phi_\tau = -(1/\beta - \gamma(1/\beta - 1))$ also involves β, moving γ and β together can also keep it satisfied. In the meantime, $(1/\beta - 1)\sigma_\tau$ needs to be kept fixed, so σ_τ has to adjust accordingly. This gives rise to the third set. The fourth set follows from an analogous mechanism to the third.

In summary, the arbitrary precision arithmetic effectively removes the ambiguity regarding the rank of the matrix $G(\theta)$. Subsequently, the procedure outlined in Subsection 2.2 pinpoints the parameter subsets that contribute to the identification failure. Although the parameter values considered here are subject to arbitrariness, the methods themselves are of general applicability. For example, in bigger models with more parameters, the number of parameter combinations that the procedure needs to search over is greater but the implementation remains the same.

Case (ii): PMAF. We structure the analysis in the same way as for the AMPF regime. Some details are omitted to avoid repetition. The baseline generic parameter values we consider are

$$\theta^1_{PMAF} = [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau] = [0.3, 0.9804, 0.1, 0.5, 0.5, 1, 1].$$

Result 3 *θ^1_{PMAF} can be changed along two orthogonal directions with observational equivalence.*

In double precision, the eigenvalues equal 7.926, 0.418, 0.020, 3.12E-04, 2.49E-09, 9.81E-18 and 2.41E-18, with the latter two being below the default tolerance level of 6.22E-15. This suggests that $G(\theta^1_{PMAF})$ has rank 5 with 7 parameters. The computation in quadruple precision shows that the first five eigenvalues do not change noticeably, but the remaining two go down to -8.17E-35 and 2.12E-37 respectively. Further increasing precision to 50 digits shows that these eigenvalues shrink to 2.90E-51 and -2.65E-53, while the rest are virtually unaffected. Therefore, the arbitrary precision arithmetic confirms the rank. Switching the numerical derivative to the 4-step method does not alter the conclusion. Applying the procedure in Subsection 2.2, we find that the parameter subsets responsible for identification failure are (γ, ϕ_τ), $(\beta, \gamma, \sigma_\tau)$ and $(\beta, \phi_\tau, \sigma_\tau)$. The analytical

solution (20) shows that both $\widehat{\pi}_t$ and $\widehat{b}_t$ contain a term multiplying e_t^τ that can be maintained by varying parameter combinations given in the nonidentified subsets.

The special parameter value is given by:

$$\theta^2_{PMAF} = [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau] = [0.3, 0.9804, 0.1, -0.3, -1.018, 1, 1].$$

Result 4 *θ^2_{PMAF} can be changed along three orthogonal directions with observational equivalence.*

In double precision, the tolerance level equals 7.77E-16 with the eigenvalues being 0.700, 0.637, 2.66E–04, 5.37E-08, 1.46E-16, 4.22E-17 and -1.13E-17. This suggests that $G(\theta^2_{PMAF})$ has rank 4. In quadruple precision, the only notable change is that the three smallest eigenvalues shrink to -9.75E-39, 3.91E-36 and -1.49E-34. When using 50 digits precision, the values fall further to 2.33E-54, 4.00E-51 and 2.04E-40. The nonidentified parameter subsets and their interpretation in light of the analytical solution coincide with those in the case of θ^2_{AMPF} due to the identical structure of the solution given parameter restrictions in both special cases.

Case (iii): PMPF. The parameter value considered in the indeterminacy regime is

$$\theta_{PMPF} = [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau, M_r, M_\tau, \sigma_\epsilon] = [0.3, 0.9804, 1.5, -0.3, -1.018, 1, 1, 0.3, 0.3, 1],$$

where M_r and M_τ are parameters determining the correlations between the sunspot shock and the monetary and fiscal shocks, respectively, and σ_ϵ is the standard deviation of the sunspot shock.

Result 5 *θ_{PMPF} is locally identified, a result clear under arbitrary but not double precision.*

Using double precision, the following eigenvalues are obtained: 1.01E+08, 5501.153, 274.015, 3.098, 0.372, 0.021, 0.002, 4.26E-05, 4.25E-07 and 9.07E-09. As the default tolerance level equals 1.49E-07, one would conclude that the rank of $G(\theta_{PMPF})$ equals 9 instead of 10. However, the computation in quadruple precision shows that $G(\theta_{PMPF})$ has full rank. The smallest eigenvalue remains of similar magnitude, being equal to 1.60E-08, and not going toward zero when the 50-digit precision is used. Therefore, the model is locally identified at θ_{PMPF}.

We take 1000 uniform random draws of parameter points to examine to what extent this is a generic feature of the model under indeterminacy. The parameters are random uniform draws within the following bounds: Lower bound: $[0.01, 0.90, 1.01, 0.005, 0.005, -0.99, -0.99, -3, -3, 0.005]$; Upper bound: $[0.99, 0.999, 2, 1, 1, 0.99, 0.99, 3, 3, 1]$. Note that the variance of the sunspot shock is constrained to be nonzero. We find that the computation in double precision produces quite heterogeneous results. The rank of the criterion matrices fluctuates between 6 and full rank of 10

across the 1000 points, with 689 out of 1000 points indicating some degree of local identification failure. The default tolerance levels fluctuate between 1.42E-13 and 3.81E-05, while the smallest eigenvalues range from 3.96E-15 to 1.96E-05. In contrast, using quadruple precision, we find that all of the 1000 considered parameter vectors are locally identified. The smallest eigenvalues remain largely the same in all cases and do not shrink to zero. Therefore, local identification under indeterminacy seems a generic property of the regime.

5.2.2 Computational time

We now document the computational costs, measured as the median runtime over 50 runs of the program, associated with computing the model solution and $G(\theta)$ on an 8-core Intel 2.4 Ghz processor. In double precision, it takes on average 0.001 seconds to solve the model for all cases considered. In quadruple and 50-digit precision, the computational cost rises to 0.021 and 0.028 seconds, respectively. When computing $G(\theta)$ for cases (i) and (ii), it takes about 0.5, 8.6 and 14.3 seconds to compute $G(\theta)$ in double, quadruple and 50-digit precision respectively using the symmetric quotient rule and 500 points for Gauss-Legendre quadrature. It takes longer to compute $G(\theta)$ in case (iii) due to the higher number of parameters in the model. Double precision computation completes in 0.6 seconds on average, while quadruple and 50-digit precision computations take 12.1 and 22.9 seconds respectively. When the derivative is computed using the four point rule, the respective times rise to 1, 16.1 and 27.2 seconds when AMPF and PMAF regimes are considered, and 1.2, 22.8 and 43.7 seconds in case of indeterminacy. We can see that using arbitrary precision arithmetic leads to substantial increases in computation time. The increase is exponential when switching from double to quadruple precision, but appears close to linear when going from quadruple to 50-digit precision. Nonetheless, it is clear that it is feasible to utilize the arithmetic for local identification analysis of the small scale models even using serial computation on a relatively low frequency CPU.

To further reduce computational cost, we parallelize part of the code to utilize all 8 cores. In case of double precision, the parallel overhead prevents any speedup. However, arbitrary precision computations benefit substantially: 1) the quadruple precision computations using 2-point derivative take 5 and 6.9 seconds for determinate and indeterminate regimes, respectively, and increase to 7.5 and 10.6 seconds when the 4-point rule is used; 2) the respective 50-digit computation times fall to 5.2 and 10.2 seconds using the 2-point rule, and to 8.3 and 20.5 seconds using the 4-point rule. Thus, the largest gains from parallelization are obtained for cases where the 4-point rule is used. There is also ample room for further reduction of computational cost by using modern CPU's with higher clock frequency, more cores or both.

5.2.3 Summary

In the above example, using double precision with default tolerance for rank determination delivers correct conclusions about local identification in 8 out of 10 cases considered. The use of arbitrary precision arithmetic leads to sharp results in all the cases. In the two cases, which concern θ^1_{AMPF} and θ_{PMPF}, the common feature is that the determined rank of $G(\theta)$ is spuriously lower by 1, which occurs due to the large disparity between the largest and the smallest nonzero eigenvalues. Interestingly, when identification from only the business cycle frequencies is considered in both cases (c.f. Appendix C), this issue is no longer present and the computed ranks are higher. Since limiting the frequencies considered cannot improve identification, this points to a potential problem with the full spectrum result. Thus, computing the identification condition at business cycle frequencies, apart from the obvious insight it provides, could act as another informal robustness check in double precision in cases where the low frequency properties of the model may result in $G(\theta)$ possessing disproportionate eigenvalues. These findings suggest two conclusions. First, the numerical procedure of Qu and Tkachenko (2012) does tend to deliver relatively robust results in double precision. Second, utilizing arbitrary precision arithmetic is indeed helpful in sharpening the results whenever there are any doubts about the conclusions.

The analysis also suggests that the following steps for carrying out the local identification analysis can be effective. First, apply the standard double precision arithmetic to estimate the rank. Then, use the arbitrary precision arithmetic to remove the ambiguity if there is any. Subsequently, use the procedure in Subsection 2.2 to pinpoint the parameter subsets and obtain their values that contribute to the identification failure. This way, we will know not only whether local identification holds, but also know which parameters are behind the identification failure and what values they may take to yield observational equivalence.

5.3 Global identification

We examine the global identification properties of the model both within and across the three regimes. Finding observational equivalence, or lack thereof, across regimes would be particularly interesting, as this can reveal whether different monetary-fiscal policy combinations can generate the same observed outcomes. The analysis focuses on the full spectrum. Using the business cycle frequencies only leads to the same conclusions. The presentation of the results follows the same structure as in the local identification case.

5.3.1 Potential observational equivalence between the $AMPF$ and the $PMPF$ regime

We examine whether there can be observational equivalence between the AMPF regime at the generic parameter value θ^1_{AMPF} and the PMPF regime at some unknown parameter values. Specifically, we treat the AMPF regime at θ^1_{AMPF} as the default specification and minimize the KL divergence between it and models in the indeterminacy region over all 10 parameters. It has been noted above that in the special case where the solutions of the model are white noise in all three regimes, it is possible to have observational equivalence across all of them.

Result 6 *Starting with the AMPF regime at $\theta = \theta^1_{AMPF}$ as the default, the method shows that there are a continuum of parameter values in the PMPF regime that can lead to observational equivalence. Furthermore, this is a generic feature of the model because the equivalence also holds at parameters different from θ^1_{AMPF}. Consequently, even if the the monetary policy has been active, it is impossible to rule out that it has been passive, even with an infinite sample size.*

Detecting observational equivalence. The minimized KL equals 2.16E-12 in double precision, which suggests observational equivalence. The minimized parameter values are

$$\begin{aligned} & [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau, M_r, M_\tau, \sigma_\epsilon] \qquad (21) \\ = \; & [0.6644, 0.9804, 1.2000, 0.8609, 0.5000, 0.2572, 1.0000, -3.4554, 2.80\text{E-}07, 1.46\text{E-}04]. \end{aligned}$$

It is apparent that three parameters differ substantially from the values under the default specification: α, ϕ_r and σ_r, while the rest stay relatively unchanged from those in θ^1_{AMPF}. Among the sunspot parameters, which do not appear in the determinate regime, all except M_r take on values close to zero. This leads to the tentative conclusion that some values of α, ϕ_r,σ_r and M_r within the PMPF regime can generate observational equivalence to the AMPF regime at θ^1_{AMPF}.

Meanwhile, one can argue that the value 2.16E-12 could be different from zero. We further apply the arbitrary precision arithmetic to see whether this can be the case. In quadruple precision, the minimized KL equals 5.09E-29 with the parameter values being

$$[0.6645, 0.9804, 1.2000, 0.8607, 0.5000, 0.2574, 1.0000, -3.4540, 2.63\text{E-}15, 1.39\text{E-}08].$$

Further increasing precision to 50 digits leads to a minimized KL of 1.95E-35, with parameters at the minimizer remaining the same as above except for M_τ and σ_ϵ, which now equal 1.64E-18 and 3.46E-10 respectively. It should be noted that KL can still be reduced further in 50-digit precision, however, the optimization algorithm converges slowly in this case, making small improvements at every iteration, most likely due to values of M_τ and σ_ϵ slowly approaching zero. We interpret the

obtained substantial reduction in the minimized KL over 200000 function evaluations in this case as further evidence of observational equivalence.

We notice that the parameters M_τ and σ_ϵ get closer to 0 when the precision is increased. Does it imply that equivalence obtains when they are set to exactly zero? To see this, we repeat the analysis in the previous two paragraphs imposing this restriction. In double precision, we find that KL reaches exactly 0 with the parameter values equal to

$$\begin{aligned} & [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau, M_r, M_\tau, \sigma_\epsilon] \\ = \; & [0.6193, 0.9804, 1.2000, 0.9506, 0.5000, 0.2172, 1.0000, -4.0929, 0, 0]. \end{aligned} \tag{22}$$

Minimizing KL in quadruple and 50-digit precision yields 3.57E-37 and 4.74E-51, respectively, with virtually the same parameter values as above (differences are of order E-06 and lower). We can see that making relevant restrictions makes the equivalence much easier to detect even in double precision. Furthermore, validating the result in increased precision is less costly as convergence is much faster. It may seem counter intuitive that an exact zero result is obtained in double precision, while small nonzero numbers are obtained in extended precision. Exact zeros do not arise often in nontrivial floating point computation output, however, they can be a consequence of integer subtraction or accidental cancellation. Also, sufficiently close real numbers may be represented by the same number in binary floating point format so that their subtraction will produce an exact zero. This implies that obtaining an exact zero does not necessarily mean that the result is absolutely precise. When higher precision is used, such close numbers can be approximated more accurately and hence their subtraction will yield a small number comparable to the machine epsilon of the precision level. This is indeed what seems to happen in this example. This reasoning also applies to exact zero results obtained in some other cases below.

Tracing out observationally equivalent parameter values. We also notice the following feature from the optimization outcomes. Aside from the point reported, the procedure returns over 10 points for which the KL divergence is of orders E-12 and E-11, all of which have different values of the above mentioned four parameters (α, ϕ_r,σ_r and M_r), while the other 6 parameter values stay relatively constant and close to their values under AMPF. Furthermore, when the zero restrictions on M_τ and σ_ϵ are imposed, the number of the minima with such features grows to 50. Thus, there can be multiple values in the PMPF region that are observationally equivalent to θ^1_{AMPF}.

To study this, we first check whether the parameters are locally identifiable at (22). Identification failure is detected with the only nonidentified parameter subset being $(\alpha, \phi_r, \sigma_r, M_r)$. We then trace out the nonidentification curve that corresponds to this subset using the method in Qu

and Tkachenko (2012) in order to collect observationally equivalent points. The step size of 1E-04 is chosen for the Euler approximation. Along the curve, the parameters α, σ_r and M_r move in the opposite direction to ϕ_r, while the rest of the parameters are fixed at their values in (22). In the direction where α increases (call it Direction 1), the curve is terminated right before α reaches 1, which is the boundary of the PMPF region. In the other direction, where ϕ_r increases (call it Direction 2), the curve is truncated right before ϕ_r reaches the invertibility bound of 1. Table 1 reports 10 equally spaced points in both directions, together with the KL distances and empirical distance measures to convey the scope of the nonidentified parameter set. From the table, we can see that the curve covers a wide range of values of the four parameters. Along the curve, both α and ϕ_r take values between 0.60 and 0.99, σ_r varies between 0.20 and 0.56, and M_r ranges from -4.44 to -1.60. The corresponding KL and empirical distances suggest exact or near equivalence to (22) along the whole curve. To verify whether equivalence is indeed exact, we perform additional minimization in quadruple precision using the reported points on the curve as starting values. The largest parameter differences at the resulting minimizers are of order E-03 in one case and E-04 in the rest, while the corresponding KL values are of order E-33 and below. The results thus further confirm observational equivalence at all considered points on this curve.

Comparing the parameter values from the nonidentification curve computed above and the multiple minima from the original KL minimization procedure, we can see that almost half of these points appear to be on this curve. The other half features much smaller values for α and opposite signs for ϕ_r and M_r. One such point is given by

$$\begin{aligned} & [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau, M_r, M_\tau, \sigma_\epsilon] \\ = \; & [0.0522, 0.9804, 1.2000, -0.0607, 0.5000, 0.2869, 1.0000, 3.0980, 0, 0]. \end{aligned}$$

The KL divergence at this point equals 5.99E-13 in double precision, and further shrinks to 3.79E-36 in quadruple precision without noticeable changes in parameter values. Local identification analysis again pinpoints the subset $(\alpha, \phi_r, \sigma_r, M_r)$ as the source of identification failure. We extend the nonidentification curve from this point in the same way as in the previous paragraph. Along this curve, the parameters α and M_r move in the opposite direction from σ_r and ϕ_r. The values along the curve are reported in Table 2. In the direction where α increases (Direction 1), the curve is terminated at the point where ϕ_r is closest to the boundary of the invertible region at -1. In the opposite direction (Direction 2), the curve is truncated when α reaches zero. Along the curve, α varies between 0 and 0.27, ϕ_r ranges from being close to 0 to being close to -1, σ_r moves between 0.09 and 0.33, and M_r takes values in the range from 2.70 to 9.78. Similarly to the previous paragraph, we use optimization in quadruple precision to verify exact equivalence for the

values reported in Table 2. This curve incorporates all the remaining minima found.

This suggests that, in practice, observing multiple parameter vectors with KL close to zero can be a signal that there are multiple observationally equivalent points. All such points can be evaluated using arbitrary precision arithmetic to validate exact equivalence. Tracing out nonidentification curves can help further characterize the observationally equivalent parameter values.

Analytical characterization. We now demonstrate this result analytically. The solution for $\widehat{\pi}_t$ under the default AMPF regime takes on an MA(1) form $\widehat{\pi}_t = (-1 - \phi_r^*/\alpha^* - \phi_r^* L)\, e_t^{*r}/\alpha^*$, where an asterisk has been added because these quantities can differ from their values under PMPF in (21). Multiply both sides of the preceding equation by $(1 - \alpha L)$:

$$\begin{aligned}
(1-\alpha L)\widehat{\pi}_t &= (1-\alpha L)\left(-\frac{1}{\alpha^*} - \frac{1}{\alpha^{*2}}\phi_r^* - \frac{\phi_r^*}{\alpha^*}L\right) e_t^{*r} \\
&= \left(-\frac{1}{\alpha^*} - \frac{1}{\alpha^{*2}}\phi_r^* - \frac{\phi_r^*}{\alpha^*}L + \frac{\alpha}{\alpha^*}L + \frac{\alpha}{\alpha^{*2}}\phi_r L + \frac{\alpha\phi_r^*}{\alpha^*}L^2\right) e_t^{*r} \\
&= \left(\left(-\frac{1}{\alpha^*} - \frac{1}{\alpha^{*2}}\phi_r\right) + \left(\frac{\alpha}{\alpha^*} + \frac{\alpha\phi_r^*}{\alpha^{*2}} - \frac{\phi_r^*}{\alpha^*}\right)L + \frac{\alpha\phi_r^*}{\alpha^*}L^2\right) e_t^{*r}.
\end{aligned}$$

The solutions for $\widehat{\pi}_t$ under the PMPF regime satisfy $\widehat{\pi}_t = \alpha\widehat{\pi}_{t-1} + \varepsilon_{t-1}^r + \eta_t, \eta_t = M_r e_t^r + M_\tau e_t^\tau + \epsilon_t$. Setting $M_\tau = \sigma_\epsilon = 0$, we obtain, in lag operator notation

$$(1-\alpha L)\widehat{\pi}_t = (M_r + L + \phi_r L^2) e_t^r.$$

The two representations of $\widehat{\pi}_t$ would be equivalent if the coefficients on the right hand side match:

$$\begin{aligned}
\left(-\frac{1}{\alpha^*} - \frac{1}{\alpha^{*2}}\phi_r^*\right)\sigma_r^* &= (-1)^s M_r \sigma_r, \\
\left(\frac{\alpha}{\alpha^*} + \frac{\alpha\phi_r^*}{\alpha^{*2}} - \frac{\phi_r^*}{\alpha^*}\right)\sigma_r^* &= (-1)^s \sigma_r, \\
\frac{\alpha\phi_r^*}{\alpha^*}\sigma_r^* &= (-1)^s \phi_r \sigma_r,
\end{aligned} \tag{23}$$

where either $s = 0$ or $s = 1$ holds in all three equations. The solution for $\widehat{b}_t$ under PMPF is

$$\widehat{b}_t = (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} - (1/\beta - 1)\varepsilon_t^\tau - \frac{M_r}{\beta} e_t^r,$$

which is the same form as (18) except for the term involving e_t^r. Thus, for equivalence in $\widehat{b}_t$ between regimes, it suffices to have

$$\left(-\frac{1}{\alpha^*} - \frac{1}{\alpha^{*2}}\phi_r^*\right)\sigma_r^* = (-1)^s M_r \sigma_r,$$

which is also one of the conditions for equivalence of $\widehat{\pi}_t$ in (23). Treating $(M_r, \sigma_r, \phi_r, \alpha)$ as four unknowns, each set of values for $(\sigma_r^*, \phi_r^*, \alpha^*)$ leads to a continuum of solutions because there are

only three restrictions in (23). Setting $s = 0$ leads to the first curve, while $s = 1$ leads to the second. Indeed, all the points reported above satisfy the restrictions in (23) numerically.

Noninvertibility. The above analysis has restricted the process for ε_t^r to be invertible, i.e., $|\phi_r| < 1$. Further observationally equivalent points can be obtained by considering noninvertible specifications of the MA(1) shocks, i.e., $|\phi_r| > 1$. One such point is given by

$$\begin{aligned} & [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau, M_r, M_\tau, \sigma_\epsilon] \\ = \; & [0.5755, 0.9804, 1.2000, 1.0765, 0.5000, 0.1782, 1.0000, -4.9886, 0, 0]. \end{aligned}$$

At this value the KL equals 0 under double precision and remains at 0 in quadruple precision.

5.3.2 Potential observational equivalence between the $PMAF$ and the $PMPF$ regime

We treat the PMAF regime at θ^1_{PMAF} as the default and search for observationally equivalent points in the PMPF regime. The analysis is organized in the same way as in the previous case.

Result 7 *Starting with the PMAF regime at θ^1_{PMAF} as the default, the method finds that there are a continuum of parameter values in the PMPF regime that can lead to observational equivalence. Furthermore, this feature is generic because the equivalence also holds for parameter values different from θ^1_{PMAF}. Consequently, even if the fiscal policy has been active, it is impossible to rule out that it has been passive, even with an infinite sample size.*

Detecting observational equivalence. The minimized KL in double precision equals 3.66E-11 with the parameter values at the optimum being

$$\begin{aligned} & [\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau, M_r, M_\tau, \sigma_\epsilon] \\ = \; & [0.3000, 0.9036, 1.1587, 0.5000, 0.4285, 1.0000, 0.2112, 2.46\text{E-}08, -0.1384, 2.28\text{E-}05]. \end{aligned}$$

It can be seen that in this case only β, γ, ϕ_τ and σ_τ differ from their original values at θ^1_{PMAF}. Out of the three sunspot parameters, only M_τ appears to be different from zero. Thus, here we could tentatively conclude that: 1) varying five parameters $(\beta, \gamma, \phi_\tau, \sigma_\tau, M_\tau)$ can produce observational equivalence between the PMAF and the PMPF regimes; 2) it could be impossible to test whether fiscal policy has been active if the monetary policy has been passive using samples of infinite size.

However, the KL divergence of 3.66E-11 could be perceived as a manifestation of near rather than exact equivalence. We perform further optimization in arbitrary precision restricting $M_r =$

$\sigma_\epsilon = 0$ to reach a sharper conclusion. In both quadruple and 50-digit precision, the minimized KL equals 0. The parameter values in both cases equal

$$[0.3000, 0.9625, 1.0208, 0.5000, 0.4775, 1.0000, 0.5270, 0, -0.0555, 0]. \tag{24}$$

Tracing out observationally equivalent parameter values. As in the previous case, in double precision, the optimization returns 10 other points with minimized KL of order E-11 and below. Applying the local identification condition to (24) reveals that two parameters need fixing to guarantee local identification. The corresponding nonidentified subsets are (γ, ϕ_τ), $(\beta, \gamma, \sigma_\tau, M_\tau)$ and $(\beta, \phi_\tau, \sigma_\tau, M_\tau)$. We extend the corresponding nonidentification curves and report the associated parameter values in Tables 3-5. Along the curve corresponding to (γ, ϕ_τ), the two parameters move in the opposite directions until γ reaches the upper bound of 3 set in optimization in Direction 1 or reaches the boundary of the PMPF region at 1 in Direction 2. The values of ϕ_τ along this curve range between 0.4406 and 0.4779. The curve corresponding to $(\beta, \gamma, \sigma_\tau, M_\tau)$ involves all parameters moving in the same direction. It is truncated for the same range of values of γ as in the previous case. The set of observationally equivalent points implied by this curve has β ranging between 0.9619 and 0.9838, σ_τ ranging between 0.5199 and 1.2103, and M_τ ranging between -0.0562 and -0.0241. The final curve involves parameters $(\beta, \phi_\tau, \sigma_\tau, M_\tau)$ moving in the same direction. It is truncated when σ_τ reaches its upper bound of 10 imposed in optimization in Direction 1, and at the point where β is closest to 0.5 in Direction 2. Along this curve, ϕ_τ moves between 0.1977 and 0.5044, σ_τ moves between 0.0486 and 10, and M_τ moves between -0.6010 and -0.0029. As in the case of the AMPF regime above, we verify exact equivalence along the curves using optimization in quadruple precision. These curves cover all of the multiple minima found during the search for equivalent PMPF values while imposing invertibility of ε_t^r and ε_t^τ.

Analytical characterization. The above results can be verified analytically. First, consider the solution for $\widehat{\pi}_t$. In the PMPF regime it takes the form $\widehat{\pi}_t = \alpha\widehat{\pi}_{t-1} + \varepsilon_{t-1}^r + M_\tau e_t^\tau$. Contrasting this with the expression for $\widehat{\pi}_t$ in (20), we can see that, keeping α, ϕ_r and σ_r the same across the two regimes, observation equivalence can be achieved if the following holds:

$$-\beta^*(1/\beta^* - 1)\left(\frac{1}{(1/\beta^* - \gamma^*(1/\beta^* - 1))}\phi_\tau^* + 1\right)\sigma_\tau^* = (-1)^s M_\tau \sigma_\tau, \tag{25}$$

where, again, parameters taking potentially different values under the PMAF regime are marked by asterisks, and $s = 0$ or 1. Moving on to the solution for $\widehat{b}_t$, in the PMPF regime it is given by

$$\begin{aligned}\widehat{b}_t &= (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} - (1/\beta - 1)\varepsilon_t^\tau - \frac{M_\tau}{\beta}e_t^\tau \\ &= (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} - \left((1/\beta - 1) + \frac{M_\tau}{\beta}\right)\left(e_t^\tau + \frac{(1/\beta - 1)\phi_\tau}{(1/\beta - 1) + \frac{M_\tau}{\beta}}e_{t-1}^\tau\right),\end{aligned}$$

which is an ARMA(1,1) process. The solution for $\widehat{b}_t$ under the PMAF regime is a white noise process given in (20). Therefore, two conditions need to be satisfied for observational equivalence: 1) ARMA(1,1) root cancellation under PMPF, implying the restriction

$$(1/\beta - \gamma(1/\beta - 1) = -\frac{(1/\beta - 1)\phi_\tau}{(1/\beta - 1) + M_\tau/\beta}; \tag{26}$$

2) equal variance of the resulting white noise process and the one in (20), which requires that

$$\frac{(1/\beta^* - 1)\phi_\tau^*\sigma_\tau^*}{(1/\beta^* - \gamma^*(1/\beta^* - 1))} = -(-1)^s\left((1/\beta - 1) + M_\tau/\beta\right)\sigma_\tau, \tag{27}$$

where $s = 0$ or 1 as in (25).

The values reported by the minimization procedure all satisfy (25), (26), and (27) with $s = 0$. There are no points with $s = 1$ found because, given θ^1_{PMAF}, the restrictions (25), (26), and (27) cannot be satisfied in this case by theoretically admissible values, as β in the PMPF regime would need to exceed 1. One way to show this is to plug in numerical values for θ^1_{PMAF} parameters on the left hand sides of the two restrictions. The first restriction yields: $-0.0292 = -M_\tau\sigma_\tau$. The second, after rewriting in terms of $M_\tau\sigma_\tau$, becomes $M_\tau\sigma_\tau = \beta \times 0.0098 + \beta\sigma_\tau - \sigma_\tau$. Combining with the first restriction, we get $\beta \times 0.0098 + \sigma_\tau(\beta - 1) = 0.0292$, which cannot be satisfied unless $\beta > 1$.

Noninvertibility. Further observationally equivalent points can be obtained by considering parameter values with noninvertible specification for ε_t^r. For example, one such point is given by

$$\begin{aligned}&[\alpha, \beta, \gamma, \phi_r, \phi_\tau, \sigma_r, \sigma_\tau, M_r, M_\tau, \sigma_\epsilon] \\ = \;&[0.3000, 0.9081, 1.0425, 2.000, 0.4372, 0.5000, 0.2211, 0, -0.1322, 0].\end{aligned}$$

At this point the KL equals 2.07E-12 in double precision, and is further reduced to 0 in quadruple precision with minor changes to parameter values. No observational equivalence is found when ε_t^τ is restricted to be noninvertible.

5.3.3 Potential observational equivalence between the $AMPF$ and $PMAF$ regimes

We consider θ^1_{AMPF} and θ^1_{PMAF} as benchmark specifications and search for the closest point in the opposite regimes.

Result 8 *No observational equivalence is detected. This no equivalence result can be viewed jointly with the equivalence results reported in the previous two subsections. They imply that whether a passive monetary (resp. fiscal) policy always leads to different economic dynamics than an active monetary (fiscal) policy depends on how the fiscal (monetary) policy is conducted. Within this model, the statement that "an active monetary policy is more beneficial to the economy" alone has no content, because a passive monetary policy can lead to observational equivalence, or no equivalence, depending on the fiscal policy. Similarly, the statement that "a passive fiscal policy is more beneficial to the economy" alone also has no content, because an active fiscal policy can lead to observational equivalence, or no equivalence, depending on monetary policy.*

The results are reported in Tables 6 and 7, respectively. In the former case, the minimized KL equals 8.8717 when ε^τ_t is restricted to be invertible, and 5.6198 when it is allowed to be nonivertible. In the latter case, noninvertible specifications of ε^r_t lead to minimized KL of 0.2665, with larger KL of 0.9715 and 0.9807 in cases where the invertibility ε^r_t is imposed. Additional optimization in quadruple precision delivers the same results to four digits after the decimal.

5.4 Empirical distance between different policy regimes

Although no cross-regime observational equivalence between $AMPF$ and $PMAF$ was found when generic parameter values were considered as benchmarks, it is of interest to assess how well one could distinguish between the base model and the closest model in the alternative regime in finite samples. We thus compute the corresponding empirical distance measures in both double and quadruple precision in order to interpret our identification results relative to sample sizes commonly encountered in empirical applications. Specifically, we fix the significance level at 5% and compute the empirical distance measures for sample sizes of 80, 150, 200 and 1000 observations using 100 points to approximate the integrals inside the KL divergence and asymptotic variance terms.

Result 9 *It is quite feasible to distinguish between the base model at θ^1_{AMPF} (resp. θ^1_{PMAF}) and the closest model in the $PMAF$ ($AMPF$) regime in empirically relevant sample sizes.*

We first consider the empirical distances between θ^1_{AMPF} and the closest point in the $PMAF$ regime with invertible shock specifications (Table 6, row 3). The computations in double precision

for all the sample sizes considered yield 1.0000 when rounded to 4 digits. The deviations from 1 are of order E-05, E-09 and E-11 for $T = 80, 150, 200$ respectively, while the distance equals 1 exactly at $T = 1000$. The computation in quadruple precision yields results that differ only slightly from their double precision counterparts: for $T = 80$ and 150, the distances are larger by values of orders E-18 and E-17, respectively and for $T = 200$ the resulting distance is smaller by E-18, while for $T = 1000$ we again obtain exactly 1. Increasing the precision to 50 digits shows that for T=1000 the difference between 1 and the resulting empirical distance is of order E-49, so that the subtraction of such a small quantity from 1 is rounded to exactly 1 in both double and quadruple precision. We next turn to the closest $PMAF$ parameter vector with both ε_t^r and ε_t^τ noninvertible (Table 6, row 6). The results are qualitatively similar: the empirical distance stays very slightly below 1 for $T = 80, 150, 200$ (deviations of order E-06, E-11 and E-14, respectively) and reaches 1 exactly at $T = 1000$ in double precision. Utilizing quadruple precision produces very similar empirical distances. The results for the remaining cases (Table 6, rows 4 and 5) are virtually identical to the second and the first cases considered above, respectively, and hence are omitted.

We further consider the $AMPF$ parameter values found to be closest to θ^1_{PMAF}. When both ε_t^r and ε_t^τ are invertible (Table 7, row 3), the empirical distances in double precision equal 0.9157, 0.9704, 0.9854 and 1.0000 for $T = 80, 150, 200, 1000$ when rounded to 4 digits. The difference between 1 and the latter value is of order E-07. When quadruple precision is used, the values are slightly smaller than those obtained in double precision, differing by values of order E-16 for $T = 80, 150, 200$ and E-17 for $T = 1000$. When both MA shock processes are noninvertible (Table 7, row 6), the empirical distances obtained in double precision equal 0.9987 for $T = 80$ and 1.0000 for the rest of sample sizes considered. For $T = 150$ and 200 the deviations of the empirical distances from 1 are of order E-06 and E-07 respectively, while at $T = 1000$ the distance equals 1 exactly. When the computation is performed in quadruple precision, the empirical distances are slightly above their double precision counterparts by values of order E-18 for $T = 80$ and E-17 for $T = 150, 200$, while at $T = 1000$ the value is smaller than 1 by 2.27E-31. The results for the remaining two cases are qualitatively similar and hence are omitted.

5.5 Discussion

The possibility of observational equivalence between the AMPF and PMAF regimes, as well as that between determinacy and indeterminacy is well recognized in the literature. We therefore recount selected studies in these areas to help put our results in context. To our knowledge, Leeper and Walker (2013) were the first to provide an analytical example of observational equivalence between the AMPF and PMAF regimes in a simple model with monetary-fiscal policy interactions. Their

policy rules are purposely simplified to be deterministic. Tan and Walker (2015), whose model we use in the illustration above, document observational equivalence between the two determinate policy regimes only under a special parameterization of shock processes (see (19)). Regarding the potential equivalence between determinate and indeterminate equilibria, Beyer and Farmer (2004) and Cochrane (2011) provide numerical and analytical examples, respectively, illustrating such observational equivalence in stylized New Keynesian models that do not contain fiscal rules.

Our results in Subsections 5.3.1-5.3.2 demonstrate explicitly, for the first time, the observational equivalence between the determinate and indeterminate regimes in a model with monetary-fiscal policy interactions. Finally, the existing literature does not appear to contain analytical results showing such cross-regime equivalence in medium scale DSGE models, however, some recent empirical studies involving such models, e.g., Leeper et al. (2017), report nearly identical fit of the two policy regimes to data[1]. Therefore, one may conjecture that such identification problems may still persist in more sophisticated models. The global identification methodology showcased here can thus be useful in investigating these issues in more general setups. The feasibility of application to medium scale DSGE models is demonstrated in the next section.

6 Application: a medium scale model

We analyze identification in the model of Schmitt-Grohé and Uribe (2012). It is an interesting application since the model features anticipated shocks, the quantitative importance of which is actively investigated in the literature. Some recent studies highlighting the important role of such shocks include Milani and Treadwell (2012), Christiano et al. (2014) and Forni et al. (2017), among others. The model is founded on a real business cycle model with endogenous capital accumulation and is augmented with real rigidities in investment, capital and wages, nonstandard consumer preferences and anticipated components in all of the exogenous shock processes. We outline the main building blocks of the model here, to ease the understanding of the results, while Schmitt-Grohé and Uribe (2012) can be consulted for full details. The full list of model parameters and their interpretations can be found in Table C1 of the online appendix.

The economy is populated with agents maximizing lifetime utility $E_0 \sum_{t=0}^{\infty} \beta^t \zeta_t U(V_t)$, where ζ_t is an exogenous preference shock, and $U(V_t) = (V_t^{1-\sigma} - 1)/(1-\sigma)$ with $V_t = C_t - bC_{t-1} - \psi h_t^{\theta} S_t$, where $S_t = (C_t - bC_{t-1})^{\gamma} S_{t-1}^{1-\gamma}$ so that consumer preferences are defined over V_t which represents a bundle of consumption (C_t), labor (h_t) and an additional variable S_t. Jaimovich and Rebelo (2009) found that this form of preferences, together with other real rigidities, is key for generating

[1] See Leeper and Leith (2016) for a review of the empirical literature on discerning policy regimes.

aggregate comovement in response to news about fundamental shocks.

Households own physical capital stock K_t, which evolves according to: $K_t = (1-\delta(u_t))K_{t-1} + z_t^I I_t\left[1 - S\left(I_t/I_{t-1}\right)\right]$. Here I_t is gross investment, and u_t measures capacity utilization in period t, so that the effective amount of capital supplied to firms is $u_t K_{t-1}$. The depreciation rate $\delta(u_t)$ satisfies $\delta(u_t) = \delta_0 + \delta_1(u_t - 1) + (\delta_2/2)(u_t - 1)^2$, with $\delta_0, \delta_1, \delta_2 > 0$. The investment adjustment cost function $S(\cdot)$, due to Christiano et al. (2005), is given by $S(x) = (\kappa/2)(x - \mu^i)^2$, where μ^i is the steady state growth rate of investment. Finally, the stationary exogenous shock z_t^I affects the technology transforming investment goods into capital goods.

The production function is of Cobb-Douglas form:

$$Y_t = z_t(u_t K_{t-1})^{\alpha_k}(X_t h_t)^{\alpha_h}(X_t L)^{1-\alpha_k-\alpha_h}, \tag{28}$$

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

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

There are seven exogenous shocks in the model, and all of them are assumed to have anticipated components. They are: 1) the stationary neutral productivity shock z_t, 2) the nonstationary neutral productivity shock X_t, 3) the stationary investment-specific productivity shock z_t^i, 4) the nonstationary investment-specific productivity shock A_t, 5) the government spending shock G_t, 6) the wage markup shock μ_t, 7) the preference shock ζ_t. The shocks X_t and A_t are made stationary using growth rates, with the respective variables being $\mu_t^x = X_t/X_{t-1}$ and $\mu_t^a = A_t/A_{t-1}$. G_t is detrended to form $g_t \equiv G_t/X_t^G$, where $X_t^G = (X_{t-1}^G)^{\rho_{xg}}(X_{t-1}A_{t-1}^{\alpha_K/(\alpha_K-1)})^{1-\rho_{xg}}$ is the trend in government spending . All seven processes x_t ($x = \{z, \mu^x, z^i, \mu^a, g, \mu, \zeta\}$) are assumed to follow: $\ln(x_t/x) = \rho_x \ln(x_{t-1}/x) + \varepsilon_{x,t}$ with $\varepsilon_{x,t} = \varepsilon_{x,t}^0 + \varepsilon_{x,t-4}^4 + \varepsilon_{x,t-8}^8$, where x denote the nonstochastic steady state values of the variables and $\varepsilon_{x,t}^j$ are i.i.d. Normal disturbances with mean zero and standard deviation σ_x^j. Since each exogenous process contains three disturbances realized at different horizons, the total number of fundamental shocks is 21.

After stationarity inducing transformation and log linearizing the equilibrium conditions around a nonstochastic steady state, Schmitt-Grohé and Uribe (2012) estimate the model on seven de-

meaned observables: real GDP growth, real consumption growth, real investment growth, labor hours, real government spending growth, TFP growth and relative price of investment growth. We consider the same set of variables, except that for the first six variables we consider log deviations from the steady state rather than growth rates, because otherwise the spectral density at the frequency zero will be singular. For the seventh variable, we continue to use the growth rate because the variable itself is nonstationary. The measurement variables are therefore

$$\widehat{y}_t \quad \widehat{c}_t \quad \widehat{i}_t \quad \widehat{h}_t \quad \widehat{g}_t \quad \widehat{z}_t \quad 100\widehat{\mu}_t^a \ . \tag{29}$$

For analysis based on the business cycle frequencies, we consider both the levels and the growth rates specifications to examine the result sensitivity.

6.1 Local identification

We first examine the local identification at the posterior median reported in Schmitt-Grohé and Uribe (2012) Table II. The full vector of structural parameters is given by:

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

The parameter values corresponding to the posterior median are denoted ϕ_0 and, for ease of reference, are included in Table C1. Schmitt-Grohé and Uribe (2012) previously applied Iskrev's (2010) condition in double precision and found that the model is locally identified from their observables at the estimated parameter values when the autocovariances of orders 0 and 1 are considered. Here we conduct the local identification analysis for the two sets of observables: 1) variables in (29); 2) the original observables in Schmitt-Grohé and Uribe (2012). In both cases, we focus on the full spectrum case in the main paper and report the results based on the business cycle frequencies in the online appendix. It should be noted that since the observables in (29) do not contain the measurement error in output, $\sigma^{me}_{g^y}$ is not considered in the corresponding identification analysis.

Result 10 *The model is locally identified at the posterior median based on the full spectrum. However, relatively low magnitudes of the smallest eigenvalues indicate that there may potentially be near observational equivalence.*

Starting with the observables in (29), the five smallest eigenvalues of the $G(\phi_0)$ matrix computed in double precision are: 1.83E-10, 3.63E-11, 1.21E-11, 3.48E-12, 7.55E-13. The default tolerance level equals 6.04E-14, so the matrix is declared full rank. Even though the smallest 5 eigenvalues above exceed the default tolerance level, they are small enough to be considered zero by practitioners

using rule of thumb tolerance levels. We repeat the computation in both quadruple and 50-digit precision, and find that the smallest eigenvalues remain virtually the same in both cases. Hence, this confirms that the model is locally identified at ϕ_0 using the observables in (29).

We consider the original set of observables from Schmitt-Grohé and Uribe (2012). The $G(\phi_0)$ matrix has full rank in double precision, as its five smallest eigenvalues equal 1.20E-04, 8.92E-05, 6.37E-05, 1.65E-05 and 3.06E-06, and the default tolerance level equals 1.99E-12. Recomputing the matrices in quadruple and 50-digit precision yields virtually the same eigenvalues, thus confirming local identification.

6.1.1 Computational time

In double precision, the model solution can be obtained in 0.014 seconds on average, while the corresponding times in quadruple and 50-digit precision are 0.50 and 0.78 seconds, respectively. It takes approximately 21 seconds to obtain $G(\phi_0)$ on an 8-core Intel 2.4 Ghz processor using 500 points Gaussian quadrature and symmetric difference quotient rule. In quadruple precision, the mean computation time increases to 29.3 minutes on the same hardware, and further grows to 110.7 minutes on average when 50-digit precision is used. When the parallelized code is used, the three times are reduced to 5.5 seconds for double, 13.9 minutes for quadruple, and 35.6 minutes in 50-digit precision. Although the costs are substantial, the computation remains feasible.

6.2 Global identification

In this subsection we examine the global identification properties of the model, which, to our knowledge, have not been previously formally studied in the literature. We also study whether the model structure can be altered to yield (near) observational equivalence.

6.2.1 Searching for observational equivalence

We search outside of a progressively larger neighborhood $B(\phi_0)$ defined as

$$B(\phi_0) = \{\phi : \|\phi - \phi_0\|_\infty \leq c\},$$

with c increasing from 0.1 to 0.5, and then to 1. For all cases, we report the minimized KL divergence and empirical distance measures for $T = 80, 150, 200, 1000$ in both double and quadruple precision. Full results can be found in Tables 8 and 9**.**

Result 11 *The model is globally identified at ϕ_0. However, seven parameters need to be fixed in order to have nontrivial ability to distinguish the closest alternative model outside the neighborhood*

given by $c = 1.0$ in sample sizes of empirical relevance. Most of the relatively weakly identified parameters in the model pertain to the anticipated shock components. This motivates a closer investigation of the relative importance of these and other news shocks in the model specification, which is pursued in the next subsection.

When searching for the closest model with $c = 0.1$, we find that the constraint is binding for σ_μ^8. The minimized KL equals 1.44E-07 in double precision. The empirical distance measures are close to 0.05 and gradually increase to 0.0518 for $T = 1000$. Therefore, although not observationally equivalent, the closest model outside the neighborhood is virtually impossible to distinguish in practice from that at ϕ_0. Increasing c to 0.5 yields a higher KL of 1.14E-06, with empirical distance reaching 0.0551 for $T = 1000$. The constraint is again binding for σ_μ^8. We can see that the identification hardly improves when a larger neighborhood is considered, and the closest model is still nearly observationally equivalent for empirically relevant sample sizes. The results imply that the standard deviation of the 8-quarter anticipated wage markup shock is difficult to determine, as nearly observationally equivalent models are obtained, with small adjustments to the rest of the parameters, when it is reduced from its base value of 0.51 to 0.41 and further to 0.01. This finding is in line with the result obtained from a Bayesian perspective by Herbst and Schorfheide (2014) using US data, where a bimodal posterior was obtained for σ_μ^8. The analysis here helps to show that the weak identification of σ_μ^8 is an intrinsic property of the model at ϕ_0.

Finally, for $c = 1$, KL equals 1.02E-05 and empirical distance measures equal 0.0527, 0.0539, 0.0546 and 0.0617 for $T = 80, 150, 200, 1000$, respectively. Thus, near observational equivalence still persists outside a fairly large neighborhood around ϕ_0. Differently from the previous two cases, here the constraint is binding for $\sigma_{z^i}^4$.

We repeat the analysis with σ_μ^8 fixed at 0.51 in order to examine to what extent identification improves. The minimized KL equal 1.71E-07, 3.48E-06 and 1.03E-05 for $c = 0.1, 0.5, 1$ respectively. The empirical distance measures remain similarly low, reaching only 0.0665 for $T = 1000$ when $c = 1$. In all cases, the constraint binds for $\sigma_{z^i}^4$, and in the case where $c = 1$ the results are virtually the same as in the previously examined respective case when σ_μ^8 was allowed to vary.

For all the cases considered above, we perform additional optimization in quadruple precision in order to clarify whether the low KL values obtained in some cases imply near or exact observational equivalence. We find that, after 2000 iterations, there are no visible changes both in the value of the objective function and in the parameter values of the minimizer. Therefore, we can conclude that there are at least two parameters that can generate near equivalence and all of them pertain to variances of anticipated shocks.

We continue this process and successively fix parameters for which the constraint is binding at their ϕ_0 values until the empirical distance for the case $c = 1$ at $T = 200$ is above 0.2. The parameters thus fixed are:

$$\sigma_{zi}^0, \sigma_{zi}^4, \sigma_{zi}^8, \sigma_{\mu}^0, \sigma_{\mu}^8, \sigma_{\zeta}^4, \sigma_{\zeta}^8. \tag{30}$$

The resulting KL equals 0.0024, and the empirical distance at $T = 200$ only equals 0.2863.

6.2.2 Identification of anticipated shocks

Motivated by the finding above, we proceed to evaluate the effect of shutting down anticipated shocks on identification. We consider scenarios where such shocks are disabled all at once, as well as one by one, in order to pinpoint the most important ones for generating model dynamics. The empirical distance measures reported below correspond to $T = 200$, unless otherwise indicated.

Result 12 *Anticipated shocks play an important role in generating model dynamics at ϕ_0, with the closest model without such components still being feasible to distinguish even in small samples. However, the anticipated components of the seven exogenous shock processes and their anticipation horizons are not equally relevant. In some cases, the entire anticipated component or its 8-horizon part can be shut down with only a modest effect on the spectrum of the observables, so that the resulting model becomes very difficult to differentiate from the benchmark in finite samples. In line with existing empirical findings, the news shocks in the wage markup are by far the most important in generating the dynamics implied by the model at ϕ_0.*

After shutting down all the anticipated shocks, the KL of the closest such model equals 0.0773, with empirical distances above 0.99. The finding suggests that the news shocks play a distinct role in shaping the dynamics of the observables. When building models featuring news shocks, the researcher typically needs to make specification choices regarding: 1) the anticipation horizon(s) to be considered; 2) which exogenous shock processes should feature the anticipated component. We proceed to examine these aspects of the model in turn.

First, we consider the relative importance of anticipation horizons, by shutting down only the 4-period, and later only the 8-period news shocks. In the former case, the closest KL divergence equals 0.0110, and the empirical distance equals 0.6828. This implies that it can be feasible to distinguish between the model featuring news shocks at both horizons and the one with only the 8-period shocks using typical sample sizes. In the latter case, the minimized KL equals 0.0037 and the corresponding empirical distance equals 0.3129. Therefore, the 8-period news shocks are relatively less important in matching the benchmark model dynamics. As Table 10 shows, the main mechanism in obtaining the closest model in this case stems mainly from inflating the 4-period news

shock standard deviations, the most pronounced changes being the increases in σ^4_{zi} (from 1.93 to 6.29) and σ^4_ζ (from 1.89 to 2.64). The increases in the rest of the 4-period standard deviations are less drastic, and σ^4_μ is the only such parameter that decreases slightly from 4.79 to 4.74. Among the unanticipated components, the notable changes are the increases in σ^0_μ (from 0.50 to 0.82) and σ^0_ζ (from 4.03 to 4.33). Thus, it may be conjectured that at least for some of the exogenous shock processes the 4-period and 8-period news shocks can have a similar effect on dynamics when standard deviations are suitably scaled.

Second, we examine whether the anticipated components of each of the exogenous shock processes can be shut down without affecting dynamics noticeably. The corresponding results are reported in Table 11. When comparing the resulting KL and empirical distances, it becomes clear that the anticipated innovations in the wage markup are by far the most important in this respect. We obtain the KL of 0.0536 and the empirical distance of 0.9972 for the closest model when σ^4_μ and σ^8_μ are set to zero. Interestingly, shutting down this anticipated component is costlier than removing either all of the 4-period or all of the 8-period news shocks, and is of qualitatively similar magnitude to the case of shutting down all of the news shocks in terms of the resulting KL of the closest model obtained under such restrictions. This result is in agreement with the findings of Schmitt-Grohé and Uribe (2012) and Sims (2016), who found that the anticipated shocks to wage markup (especially the 4-horizon component) explain the largest fractions of the forecast error variance of the key observables when either the unconditional variance or only its pure news share is considered, respectively. Here we show that such importance is a structural feature of the model at ϕ_0.

The second most important news component appears to be that of the investment-specific productivity shock $\widehat{z}^i_t$. The minimized KL equals 0.0046, resulting in the empirical distance of 0.3669. This is qualitatively similar to disabling all of the 8-period news shocks based on the results above. This is also broadly in line with the variance decomposition results of Schmitt-Grohé and Uribe (2012) and Sims (2016), who find it the second most important news shock in driving the output growth. The third and fourth largest KL magnitudes of the closest models equal 0.0031 and 0.0019, corresponding to cases where the news shocks in $\widehat{\mu}^a_t$ and $\widehat{g}_t$ are shut down, respectively. The empirical distances equal 0.2912 and 0.2214. The relative importance of these news shocks can be explained by the fact that both $\widehat{g}_t$ and $\widehat{\mu}^a_t$ are among the considered observables, and over 50% of their own variance is due to their anticipated component (see Table VI in Schmitt-Grohé and Uribe (2012)), which is seemingly hard to replicate by changing other parameters.

The remaining news shocks in $\widehat{\mu}^x_t$, $\widehat{\zeta}_t$ and $\widehat{z}_t$ appear much less important. The respective minimized KL values obtained are 0.0009, 0.0006 and 0.0002, while the respective empirical distances equal 0.1394, 0.1212 and 0.0823. All empirical distance measures fall below 0.1 when $T = 80$.

This implies that it may be practically impossible to distinguish between models with and without anticipated components in these shocks. Such results are not surprising for the neutral productivity shocks $\widehat{\mu}_t^x$ and $\widehat{z}_t$, as Schmitt-Grohé and Uribe (2012) documented that nearly all contribution of these shocks to the variance of observables comes from their unanticipated components. In the case of $\widehat{\zeta}_t$, despite previous findings on its news component explaining relatively large fractions of consumption and output variance, our result shows that the model closely matching the original dynamics can be obtained by mainly increasing σ_ζ^0 and decreasing σ_z^8.

Finally, we examine whether shutting down individual 8-horizon news shock could lead to (near) observational equivalence. We find that the 8-horizon news shock in $\widehat{z}_t^i$ is the most costly in terms of KL divergence to shut down: the closest model has KL equal to 0.0030, implying an empirical distance of 0.2708. The closest model is obtained mainly by amplifying σ_{zi}^4 from 1.93 to 6.20. The next two most important 8-horizon news shocks are those in $\widehat{\mu}_t^a$ and $\widehat{\mu}_t^x$, for which the minimized KL values are 0.0004 and 0.0001, and the empirical distances equal 0.1070 and 0.0756, respectively. For these shocks appropriately rescaling the 4-horizon standard deviations (in the closest models, $\sigma_{\mu^x}^4$ and $\sigma_{\mu^a}^4$ increase by roughly 60% and 40%, respectively) coupled with relatively small changes in other parameters can closely replicate the effect of 8-horizon shocks. For the remaining four exogenous processes, we find near observational equivalence. The minimized KL values for the cases of $\widehat{\zeta}_t$, $\widehat{g}_t$, $\widehat{z}_t$ and $\widehat{\mu}_t$ equal 6.34E-05, 5.53E-05, 6.91E-06 and 1.14E-06, respectively. The corresponding empirical distances equal 0.0684, 0.0672, 0.0555 and 0.0522, and all of them remain below 0.1 even for $T = 1000$. Optimization in quadruple precision results in essentially the same objective function values, thus ruling out exact observational equivalence. In all four cases, near equivalence is achieved mainly by increasing the standard deviations of the 4-horizon news shocks.

6.2.3 Distinguishability from models with ARMA shocks

We consider whether an alternative information structure can deliver observationally equivalent dynamics. Specifically, for all the seven shocks we consider the following specification:

$$\begin{aligned} \ln(x_t/x) &= \rho_x \ln(x_{t-1}/x) + \varepsilon_{x,t}, \\ \varepsilon_{x,t} &= b_0 e_{x,t} + b_4 e_{x,t-4} + b_8 e_{x,t-8}, \end{aligned} \tag{31}$$

where $e_{x,t} \sim i.i.d.N(0,1)$, $\varepsilon_{x,t}$ is allowed to be non-invertible, and b_0 is normalized to be nonnegative. This leads to a model with a total of 7 structural shocks in contrast to the 21 in the original model.

Result 13 *The dynamics generated by the original news shock specification are not easily replicable with the ARMA specification (31). However, near observational equivalence can arise when $\varepsilon_{x,t}$ is*

replaced by the moving average in (31) one by one. This is the case for the wage markup shock $\widehat{\mu}_t$, which implies that the importance of the wage markup shock dynamics does not necessarily stem from its news component, but from introducing dynamics at 4 and 8 period horizons in some form.

When all the original shocks are replaced by (31), the minimized KL equals 0.02226 with the corresponding empirical distance of 0.7991 at $T = 200$. Additional optimization in quadruple precision yields essentially the same values. Thus the model with the considered alternative structure can be differentiated from the benchmark in finite samples of empirical relevance.

We now consider alternative model structures where the shocks are replaced by (31) one by one. The closest models thus obtained for $\widehat{\mu}_t^a$, $\widehat{\mu}_t^x$ and $\widehat{z}_t$ are essentially the same as in the previous subsection when the entire anticipated components of these shock processes were shut down, indicating that augmenting their unanticipated components with lags does not help in matching dynamics at ϕ_0. Relatively marginal reductions in empirical distances for $\widehat{z}_t^i$, $\widehat{g}_t$, and $\widehat{\zeta}_t$ are obtained, which implies that the lags of unanticipated shocks help in generating news like dynamics only to a limited extent in these cases. Finally, we find that for $\widehat{\mu}_t$ the specification in (31) yields near equivalence: the closest model yields the KL of 1.25E-05 with the empirical distance of 0.0575 at $T = 200$. Even at $T = 1000$, the empirical distance remains low at 0.0683. This result may be viewed as surprising, given that the anticipated component of the wage markup shock is of key importance among such components in generating model dynamics and as such the costliest to shut down.

The consideration of (31) is motivated by Walker and Leeper (2011), who contrasted different shock specifications analytically within a stylized model, and numerically using an earlier version of the model considered here (Schmitt-Grohé and Uribe (2008)), with a particular focus on the nonstationary technological shock. Another related paper is Leeper et al. (2013), where tax news are studied under different information flow specifications. It is desirable to have a method that can comprehensively study to what extent these alternative information structures can be distinguished from each other. The method of this paper can potentially play such a role.

6.2.4 Computational time

The computation times are reported for a 12-core 2.3 Ghz Intel Xeon processor. The computation times in double precision vary between 6 and 24 hours. For additional optimization in quadruple precision the computational time varies between 5.5 and 11 hours. These time estimates can be treated as conservative, as in all cases we let the algorithm run for 2000 iterations even though no noticeable improvement is eventually obtained. Capping the iteration count at a lower but still sufficiently large number, say, 1000, would reduce the computation time roughly by half. Although the computation here does not benefit to the same extent from the increased number of cores

(some modest speedup can be expected due to the built-in parallelism of some of the multiprecision toolbox operations), it would run much faster on processors with fewer cores but faster clock speeds.

7 Conclusion

This paper has applied arbitrary precision arithmetic to resolve practical difficulties arising in the identification analysis of DSGE models. We develop a three-step procedure for analyzing both local and global identification. The nonidentification curve further traces out observationally equivalent parameter values when identification failure is detected. The empirical distance measure quantifies the closeness between different models. A `Matlab` code is developed to implement all these methods. The applications suggest that it is feasible to apply the methods to small as well as medium scale models to deliver informative results. We conjecture that resolving the numerical uncertainty can substantially facilitate the identification analysis for DSGE models in practice.

References

Advanpix LLC. (2016): "Multiprecision Computing Toolbox for MATLAB," Yokohama, Japan.

Anderson, G. S. (2008): "Solving Linear Rational Expectations Models: A Horse Race," *Computational Economics*, 31, 95–113.

Beyer, A., and R. E. A. Farmer (2004): "On the Indeterminacy of New-Keynesian Economics," European Central Bank Working Paper 323.

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

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

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

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

Cox, D. R. (1961): "Tests of Separate Families of Hypotheses," in *Proceedings of the 4th Berkeley Symposium in Mathematical Statistics and Probability*, vol. 1, pp. 105–123. Berkeley: University of California Press.

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

Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016): "Solution and Estimation Methods for DSGE Models," *Handbook of Macroeconomics*, 2, 527–724.

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

Greenwood, J., Z. Hercowitz, and G. Huffman (1988): "Investment, Capacity Utilization, and the Real Business Cycle," *American Economic Review*, 78, 402–417.

Herbst, E. P., and F. Schorfheide (2014): "Sequential Monte Carlo Sampling For DSGE Models," *Journal of Applied Econometrics*, 29, 1073–1098.

IEEE (2008): "IEEE Standard for Floating-Point Arithmetic," *IEEE Std 754-2008*, pp. 1–70.

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

Jaimovich, N., and S. Rebelo (2009): "Can News About the Future Drive the Business Cycle?," *American Economic Review*, 99, 1097–1118.

King, R., C. Plosser, and S. Rebelo (1988): "Production, Growth and Business Cycles: I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21, 195–232.

Kneusel, R. T. (2017): *Numbers and Computers*. Springer, 2 edn.

Komunjer, I., and S. Ng (2011): "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," *Econometrica*, 79, 1995–2032.

Leeper, E. M. (1991): "Equilibria Under 'Active' and 'Passive' Monetary and Fiscal Policies," *Journal of Monetary Economics*, 27, 129–147.

Leeper, E. M., and C. Leith (2016): "Understanding Inflation as a Joint Monetary-Fiscal Phenomenon," *Handbook of Macroeconomics*, 2, 2305–2416.

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

Leeper, E. M., and T. B. Walker (2013): "Perceptions and Misperceptions of Fiscal Inflation," in *Fiscal Policy after the Financial Crisis*, ed. by A. Alesina, and F. Giavazzi, pp. 255–299. University of Chicago Press.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013): "Fiscal Foresight and Information Flows," *Econometrica*, 81, 1115–1145.

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

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

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

Muller, J.-M., N. Brisebarre, F. de Dinechin, C.-P. Jeannerod, V. Lefèvre, G. Melquiond, N. Revol, D. Stehlé, and S. Torres (2010): *Handbook of Floating-Point Arithmetic*. Birkhäuser Boston.

Pesaran, H. (1987): "Global and Partial Non-nested Hypotheses and Asymptotic Local Power," *Econometric Theory*, 3, 69–97.

Qu, Z., and D. Tkachenko (2012): "Identification and Frequency Domain Quasi-maximum Likelihood Estimation of Linearized Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 3, 95–132.

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

Schmitt-Grohé, S., and M. Uribe (2008): "What's News in Business Cycles," NBER Working Paper 14215.

——— (2012): "What's News in Business Cycles," *Econometrica*, 80, 2733–2764.

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

Sims, E. (2016): "What's News in News? A Cautionary Note on Using a Variance Decomposition to Assess the Quantitative Importance of News Shocks," *Journal of Economic Dynamics & Control*, 73, 41–60.

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

Tan, F., and T. B. Walker (2015): "Solving Generalized Multivariate Linear Rational Expectations Models," *Journal of Economic Dynamics & Control*, 60, 95–111.

Walker, T. B., and E. M. Leeper (2011): "Information Flows and News Driven Business Cycles," *Review of Economic Dynamics*, 14, 55–71.

Table 1. Nonidentification curve 1, PMPF regime equivalent to AMPF Leeper (1991) model

	α	ϕ_r	σ_r	M_r	$\|\theta_j - \theta_0\|$	KL	$T = 100$	$T = 1000$
θ_0	0.6193	0.9506	0.2172	-4.0929	–	–	–	–
				(a) Direction 1				
θ_1	0.6355	0.9147	0.2316	-3.8381	0.2582	1.96E-10	0.0500	0.0501
θ_2	0.6541	0.8789	0.2481	-3.5835	0.5165	9.71E-10	0.0500	0.0501
θ_3	0.6754	0.8432	0.2670	-3.3293	0.7747	2.76E-09	0.0501	0.0502
θ_4	0.7001	0.8075	0.2890	-3.0756	1.0329	6.37E-09	0.0501	0.0504
θ_5	0.7293	0.7719	0.3149	-2.8228	1.2910	1.33E-08	0.0502	0.0505
θ_6	0.7639	0.7366	0.3457	-2.5711	1.5488	2.64E-08	0.0502	0.0507
θ_7	0.8058	0.7015	0.3829	-2.3215	1.8061	5.19E-08	0.0503	0.0510
θ_8	0.8570	0.6668	0.4284	-2.0749	2.0625	1.04E-07	0.0505	0.0515
θ_9	0.9204	0.6328	0.4848	-1.8334	2.3170	2.26E-07	0.0507	0.0522
θ_{10}	0.9999	0.6000	0.5554	-1.6004	2.5680	7.33E-07	0.0512	0.0541
				(b) Direction 2				
θ_1	0.6173	0.9555	0.2153	-4.1279	0.0355	2.95E-12	0.0500	0.0500
θ_2	0.6152	0.9604	0.2135	-4.1631	0.0711	1.15E-11	0.0500	0.0500
θ_3	0.6132	0.9654	0.2117	-4.1982	0.1067	2.53E-11	0.0500	0.0500
θ_4	0.6112	0.9703	0.2100	-4.2334	0.1423	4.39E-11	0.0500	0.0500
θ_5	0.6093	0.9753	0.2082	-4.2685	0.1779	6.69E-11	0.0500	0.0500
θ_6	0.6074	0.9802	0.2065	-4.3037	0.2135	9.40E-11	0.0500	0.0500
θ_7	0.6055	0.9852	0.2049	-4.3389	0.2491	1.25E-10	0.0500	0.0501
θ_8	0.6036	0.9901	0.2032	-4.3740	0.2847	1.59E-10	0.0500	0.0501
θ_9	0.6018	0.9950	0.2016	-4.4092	0.3203	1.97E-10	0.0500	0.0501
θ_{10}	0.6000	0.9999	0.2000	-4.4444	0.3559	2.37E-10	0.0500	0.0501

Note. θ_j represent equally spaced points taken from the curve determined by the smallest eigenvalue from changing the four parameters. The curve is extended from θ_0 along two directions. Along Direction 1, the curve is truncated when α reaches 1. Along Direction 2, the curve is truncated at the point where ϕ_r is closest to 1. KL is defined as $KL_{ff}(\theta_0, \theta_j)$. The last two columns are empirical distance measures defined as $p_{ff}(\theta_0, \theta_j, 0.05, T)$. The KL and empirical distance measures are computed in double precision using 500 quadrature points to approximate the integral.

Table 2. Nonidentification curve 2, PMPF regime equivalent to AMPF Leeper (1991) model

	α	ϕ_r	σ_r	M_r	$\|\theta_j - \theta_0\|$	KL	$T = 100$	$T = 1000$
θ_0	0.0522	-0.0607	0.2869	3.0980	–	–	–	–
				(a) Direction 1				
θ_1	0.1092	-0.1541	0.2363	3.7623	0.6752	3.50E-09	0.0501	0.0503
θ_2	0.1492	-0.2478	0.2007	4.4290	1.3504	8.72E-09	0.0501	0.0504
θ_3	0.1788	-0.3417	0.1744	5.0967	2.0255	1.32E-08	0.0502	0.0505
θ_4	0.2015	-0.4357	0.1542	5.7648	2.7005	1.67E-08	0.0502	0.0506
θ_5	0.2196	-0.5297	0.1382	6.4332	3.3755	1.94E-08	0.0502	0.0506
θ_6	0.2342	-0.6237	0.1252	7.1017	4.0505	2.15E-08	0.0502	0.0507
θ_7	0.2463	-0.7177	0.1144	7.7704	4.7256	2.32E-08	0.0502	0.0507
θ_8	0.2565	-0.8117	0.1053	8.4390	5.4006	2.45E-08	0.0502	0.0507
θ_9	0.2652	-0.9058	0.0976	9.1078	6.0757	2.56E-08	0.0502	0.0507
θ_{10}	0.2727	-0.9998	0.0909	9.7765	6.7507	2.65E-08	0.0502	0.0507
				(b) Direction 2				
θ_1	0.0476	-0.0546	0.2910	3.0548	0.0440	2.78E-11	0.0500	0.0500
θ_2	0.0429	-0.0485	0.2952	3.0115	0.0881	1.17E-10	0.0500	0.0500
θ_3	0.0381	-0.0424	0.2995	2.9684	0.1322	2.75E-10	0.0500	0.0501
θ_4	0.0331	-0.0364	0.3039	2.9252	0.1763	5.11E-10	0.0500	0.0501
θ_5	0.0280	-0.0303	0.3084	2.8820	0.2204	8.36E-10	0.0500	0.0501
θ_6	0.0228	-0.0242	0.3131	2.8389	0.2645	1.26E-09	0.0501	0.0502
θ_7	0.0173	-0.0182	0.3179	2.7959	0.3086	1.80E-09	0.0501	0.0502
θ_8	0.0117	-0.0121	0.3229	2.7528	0.3527	2.47E-09	0.0501	0.0502
θ_9	0.0060	-0.0061	0.3280	2.7098	0.3968	3.28E-09	0.0501	0.0503
θ_{10}	3.00E-05	-2.82E-05	0.3333	2.6669	0.4409	4.26E-09	0.0501	0.0503

Note. θ_j represent equally spaced points taken from the curve determined by the smallest eigenvalue from changing the four parameters. The curve is extended from θ_0 along two directions. Along Direction 1, the curve is truncated when ϕ_r is closest to -1. Along Direction 2, the curve is truncated at the point where α is closest to 0. KL is defined as $KL_{ff}(\theta_0, \theta_j)$. The last two columns are empirical distance measures defined as $p_{ff}(\theta_0, \theta_j, 0.05, T)$. The KL and empirical distance measures are computed in double precision using 500 quadrature points to approximate the integral.

Table 3. Nonidentification curve 1, PMPF regime equivalent to PMAF Leeper (1991) model

	γ	ϕ_τ	$\|\theta_j - \theta_0\|$	KL	$T = 100$	$T = 1000$
θ_0	1.0208	0.4775	–	–	–	–
			(a) Direction 1			
θ_1	1.2186	0.4738	0.1978	4.95E-17	0.0500	0.0500
θ_2	1.4164	0.4701	0.3957	1.14E-16	0.0500	0.0500
θ_3	1.6143	0.4664	0.5936	1.70E-16	0.0500	0.0500
θ_4	1.8122	0.4628	0.7915	1.44E-16	0.0500	0.0500
θ_5	2.0100	0.4591	0.9894	1.56E-16	0.0500	0.0500
θ_6	2.2079	0.4554	1.1873	1.53E-16	0.0500	0.0500
θ_7	2.4058	0.4517	1.3852	1.60E-16	0.0500	0.0500
θ_8	2.6036	0.4480	1.5831	1.53E-16	0.0500	0.0500
θ_9	2.8015	0.4443	1.7810	1.56E-16	0.0500	0.0500
θ_{10}	3.0000	0.4406	1.9795	1.57E-16	0.0500	0.0500
			(b) Direction 2			
θ_1	1.0188	0.4775	0.0020	6.52E-17	0.0500	0.0500
θ_2	1.0167	0.4776	0.0041	7.51E-17	0.0500	0.0500
θ_3	1.0146	0.4776	0.0062	8.54E-17	0.0500	0.0500
θ_4	1.0125	0.4777	0.0083	5.66E-17	0.0500	0.0500
θ_5	1.0104	0.4777	0.0104	5.83E-17	0.0500	0.0500
θ_6	1.0083	0.4777	0.0125	8.33E-17	0.0500	0.0500
θ_7	1.0062	0.4778	0.0146	6.94E-17	0.0500	0.0500
θ_8	1.0041	0.4778	0.0167	7.40E-17	0.0500	0.0500
θ_9	1.0020	0.4779	0.0188	8.36E-17	0.0500	0.0500
θ_{10}	1.0000	0.4779	0.0208	6.75E-17	0.0500	0.0500

Note. θ_j represent equally spaced points taken from the curve determined by the smallest eigenvalue from changing the two parameters. The curve is extended from θ_0 along two directions. Along Direction 1, the curve is truncated when γ reaches the upper bound of 3. Along Direction 2, the curve is truncated at the point where γ is closest to 1. KL is defined as $KL_{ff}(\theta_0, \theta_j)$. The last two columns are empirical distance measures defined as $p_{ff}(\theta_0, \theta_j, 0.05, T)$. The KL and empirical distance measures are computed in double precision using 500 quadrature points to approximate the integral.

Table 4. Nonidentification curve 2, PMPF regime equivalent to PMAF Leeper (1991) model

	β	γ	σ_τ	M_τ	$\|\theta_j - \theta_0\|$	KL	$T = 100$	$T = 1000$
θ_0	0.9625	1.0208	0.5270	-0.0555	–	–	–	–
				(a) Direction 1				
θ_1	0.9669	1.2185	0.5953	-0.0491	0.2093	5.41E-11	0.0500	0.0500
θ_2	0.9703	1.4164	0.6636	-0.0440	0.4187	1.78E-10	0.0500	0.0501
θ_3	0.9731	1.6142	0.7319	-0.0399	0.6281	3.30E-10	0.0500	0.0501
θ_4	0.9754	1.8121	0.8002	-0.0365	0.8375	4.89E-10	0.0500	0.0501
θ_5	0.9774	2.0101	0.8685	-0.0336	1.0469	6.45E-10	0.0500	0.0501
θ_6	0.9791	2.2080	0.9369	-0.0312	1.2563	7.95E-10	0.0500	0.0501
θ_7	0.9805	2.4059	1.0052	-0.0291	1.4656	9.36E-10	0.0500	0.0501
θ_8	0.9818	2.6038	1.0735	-0.0272	1.6750	1.07E-09	0.0500	0.0501
θ_9	0.9829	2.8017	1.1418	-0.0256	1.8844	1.19E-09	0.0500	0.0501
θ_{10}	0.9838	3.0000	1.2103	-0.0241	2.0941	1.31E-09	0.0500	0.0502
				(b) Direction 2				
θ_1	0.9624	1.0188	0.5263	-0.0555	0.0021	6.68E-15	0.0500	0.0500
θ_2	0.9624	1.0167	0.5256	-0.0556	0.0043	2.81E-14	0.0500	0.0500
θ_3	0.9623	1.0147	0.5249	-0.0557	0.0065	6.44E-14	0.0500	0.0500
θ_4	0.9623	1.0126	0.5242	-0.0558	0.0087	1.16E-13	0.0500	0.0500
θ_5	0.9622	1.0105	0.5235	-0.0558	0.0109	1.82E-13	0.0500	0.0500
θ_6	0.9622	1.0084	0.5228	-0.0559	0.0131	2.64E-13	0.0500	0.0500
θ_7	0.9621	1.0064	0.5220	-0.0560	0.0153	3.61E-13	0.0500	0.0500
θ_8	0.9621	1.0043	0.5213	-0.0561	0.0175	4.74E-13	0.0500	0.0500
θ_9	0.9620	1.0022	0.5206	-0.0561	0.0197	6.02E-13	0.0500	0.0500
θ_{10}	0.9619	1.0000	0.5199	-0.0562	0.0220	7.53E-13	0.0500	0.0500

Note. θ_j represent equally spaced points taken from the curve determined by the smallest eigenvalue from changing the four parameters. The curve is extended from θ_0 along two directions. Along Direction 1, the curve is truncated when γ reaches 3. Along Direction 2, the curve is truncated at the point where γ is closest to 1. KL is defined as $KL_{ff}(\theta_0, \theta_j)$. The last two columns are empirical distance measures defined as $p_{ff}(\theta_0, \theta_j, 0.05, T)$. The KL and empirical distance measures are computed in double precision using 500 quadrature points to approximate the integral.

Table 5. Nonidentification curve 3, PMPF regime equivalent to PMAF Leeper (1991) model

	β	ϕ_τ	σ_τ	M_τ	$\|\theta_j - \theta_0\|$	KL	$T = 100$	$T = 1000$
θ_0	0.9625	1.0208	0.5270	-0.0555	–	–	–	–
				(a) Direction 1				
θ_1	0.9867	0.4958	1.4727	-0.0198	0.9469	1.45E-08	0.0501	0.0505
θ_2	0.9919	0.4998	2.4201	-0.0121	1.8939	2.16E-08	0.0502	0.0506
θ_3	0.9942	0.5015	3.3675	-0.0087	2.8411	2.52E-08	0.0502	0.0507
θ_4	0.9955	0.5025	4.3149	-0.0068	3.7884	2.73E-08	0.0502	0.0507
θ_5	0.9963	0.5031	5.2623	-0.0056	4.7357	2.87E-08	0.0502	0.0507
θ_6	0.9969	0.5035	6.2097	-0.0047	5.6830	2.96E-08	0.0502	0.0507
θ_7	0.9973	0.5039	7.1571	-0.0041	6.6304	3.04E-08	0.0502	0.0507
θ_8	0.9976	0.5041	8.1045	-0.0036	7.5777	3.10E-08	0.0502	0.0508
θ_9	0.9979	0.5043	9.0519	-0.0032	8.5251	3.14E-08	0.0502	0.0508
θ_{10}	0.9981	0.5044	10.0000	-0.0029	9.4732	3.18E-08	0.0502	0.0508
				(b) Direction 2				
θ_1	0.9533	0.4707	0.4254	-0.0687	0.1032	1.94E-09	0.0500	0.0502
θ_2	0.9386	0.4600	0.3259	-0.0897	0.2061	1.25E-08	0.0501	0.0505
θ_3	0.9135	0.4419	0.2342	-0.1248	0.3070	4.63E-08	0.0502	0.0509
θ_4	0.8731	0.4138	0.1627	-0.1796	0.4003	1.17E-07	0.0504	0.0515
θ_5	0.8206	0.3789	0.1179	-0.2477	0.4839	2.00E-07	0.0506	0.0520
θ_6	0.7622	0.3420	0.0914	-0.3197	0.5640	2.81E-07	0.0508	0.0525
θ_7	0.7004	0.3050	0.0746	-0.3918	0.6452	3.64E-07	0.0511	0.0530
θ_8	0.6360	0.2685	0.0631	-0.4629	0.7290	4.51E-07	0.0514	0.0536
θ_9	0.5692	0.2327	0.0548	-0.5326	0.8156	5.42E-07	0.0517	0.0542
θ_{10}	0.5000	0.1977	0.0486	-0.6010	0.9048	6.33E-07	0.0520	0.0547

Note. θ_j represent equally spaced points taken from the curve determined by the smallest eigenvalue from changing the four parameters. The curve is extended from θ_0 along two directions. Along Direction 1, the curve is truncated when σ_τ reaches 10. Along Direction 2, the curve is truncated at the point where β is closest to 0.5. KL is defined as $KL_{ff}(\theta_0, \theta_j)$. The last two columns are empirical distance measures defined as $p_{ff}(\theta_0, \theta_j, 0.05, T)$. The KL and empirical distance measures are computed in double precision using 500 quadrature points to approximate the integral.

Table 6. Parameter values minimizing KL between θ^1_{AMPF} and the PMAF region

Leeper (1991) model

	α	β	γ	ϕ_r	ϕ_τ	σ_r	σ_τ		
θ^1_{AMPF}	1.5000	0.9804	1.2000	0.5000	0.5000	1.0000	1.0000		
					All parameters vary				
$(\varepsilon^r_t, \varepsilon^\tau_t)$	α	β	γ	ϕ_r	ϕ_τ	σ_r	σ_τ	KL	KL_{34}
(Inv, Inv)	**0.0100**	**0.9000**	**0.9900**	0.3663	**-0.9990**	0.8901	**10.0000**	8.8717	8.8717
$(Inv, Noninv)$	**0.0100**	**0.9000**	0.2279	0.8299	-3.0997	1.3140	**10.0000**	5.6198	5.6198
$(Noninv, Inv)$	**0.0100**	**0.9000**	**0.9900**	2.7299	**-0.9990**	0.3260	**10.0000**	8.8717	8.8717
$(Noninv, Noninv)$	**0.0100**	**0.9000**	0.2184	1.2050	-3.1027	1.0905	**10.0000**	5.6198	5.6198

Note. The parameter values represent the minimizers of $KL(\theta^1_{AMPF}, \theta)$ with θ restricted to the PMAF region. KL and KL_{34} signify the minimized KL values obtained in double and quadruple precision respectively. The notation $(Inv, Noninv)$ indicates whether each of $(\epsilon^r_t, \epsilon^\tau_t)$ is restricted to an invertible or noninvertible specification. Bolded values denote the binding boundary constraints.

Table 7. Parameter values minimizing KL between θ^1_{PMAF} and the AMPF region

Leeper (1991) model

	α	β	γ	ϕ_r	ϕ_τ	σ_r	σ_τ		
θ^1_{PMAF}	0.3000	0.9804	0.1000	0.5000	0.5000	1.0000	1.0000		
					All parameters vary				
$(\varepsilon^r_t, \varepsilon^\tau_t)$	α	β	γ	ϕ_r	ϕ_τ	σ_r	σ_τ	KL	KL_{34}
(Inv, Inv)	**1.0100**	0.9980	**1.0100**	**-0.9990**	0.0629	1.3285	**10.0000**	0.9715	0.9715
$(Inv, Noninv)$	**1.0100**	**0.9000**	**3.0000**	**-0.9990**	**5.0000**	1.3285	0.0359	0.9807	0.9807
$(Noninv, Inv)$	2.9839	**0.9000**	**3.0000**	-2.9842	-0.7778	1.3053	0.0884	0.2665	0.2665
$(Noninv, Noninv)$	**3.0000**	**0.9000**	**3.0000**	-3.0003	-1.2857	1.3053	0.0687	0.2665	0.2665

Note. The parameter values represent the minimizers of $KL(\theta^1_{PMAF}, \theta)$ with θ restricted to the AMPF region. KL and KL_{34} signify the minimized KL values obtained in double and quadruple precision respectively. The notation $(Inv, Noninv)$ indicates whether each of $(\epsilon^r_t, \epsilon^\tau_t)$ is restricted to an invertible or noninvertible specification. Bolded values denote the binding boundary constraints.

Table 8. Parameter values miminizing the KL criterion, SGU(2012) model

		(a) All parameters can vary			(b) σ_μ^8 fixed		
	ϕ_0	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
θ	4.74	4.74	4.74	4.74	4.74	4.74	4.74
γ	0.0019	0.0019	0.0019	0.0019	0.0019	0.0019	0.0019
κ	9.11	9.11	9.11	9.08	9.11	9.09	9.08
δ_2/δ_1	0.34	0.34	0.34	0.34	0.34	0.34	0.34
b	0.91	0.91	0.91	0.91	0.91	0.91	0.91
ρ_{xg}	0.72	0.72	0.72	0.72	0.72	0.72	0.72
ρ_{μ^a}	0.48	0.48	0.48	0.48	0.48	0.48	0.48
ρ_{μ^x}	0.38	0.38	0.38	0.38	0.38	0.38	0.38
ρ_{z^i}	0.47	0.47	0.47	0.48	0.47	0.47	0.48
ρ_z	0.92	0.92	0.92	0.92	0.92	0.92	0.92
ρ_μ	0.98	0.98	0.98	0.98	0.98	0.98	0.98
ρ_g	0.96	0.96	0.96	0.96	0.96	0.96	0.96
ρ_ζ	0.17	0.17	0.17	0.17	0.17	0.17	0.17
$\sigma^0_{\mu^a}$	0.21	0.21	0.21	0.21	0.21	0.21	0.21
$\sigma^4_{\mu^a}$	0.16	0.16	0.16	0.16	0.16	0.16	0.16
$\sigma^8_{\mu^a}$	0.16	0.16	0.16	0.16	0.16	0.16	0.16
$\sigma^0_{\mu^x}$	0.38	0.38	0.38	0.38	0.38	0.38	0.38
$\sigma^4_{\mu^x}$	0.08	0.08	0.08	0.08	0.08	0.08	0.08
$\sigma^8_{\mu^x}$	0.10	0.10	0.10	0.10	0.10	0.10	0.10
$\sigma^0_{z^i}$	11.72	11.72	11.72	11.64	11.71	11.67	11.64
$\sigma^4_{z^i}$	1.93	1.92	1.91	**0.93**	**1.83**	**1.43**	**0.93**
$\sigma^8_{z^i}$	5.50	5.50	5.50	5.53	5.50	5.52	5.53
σ^0_z	0.65	0.65	0.65	0.65	0.65	0.65	0.65
σ^4_z	0.11	0.11	0.11	0.11	0.11	0.11	0.11
σ^8_z	0.09	0.09	0.09	0.09	0.09	0.09	0.09
σ^0_μ	0.50	0.51	0.52	0.54	0.50	0.52	0.54
σ^4_μ	4.79	4.80	4.82	4.79	4.79	4.79	4.79
σ^8_μ	0.51	**0.41**	**0.01**	0.45	0.51	0.51	0.51
σ^0_g	0.62	0.62	0.62	0.62	0.62	0.62	0.62
σ^4_g	0.57	0.57	0.57	0.57	0.57	0.57	0.57
σ^8_g	0.37	0.37	0.37	0.37	0.37	0.37	0.37
σ^0_ζ	4.03	4.03	4.03	4.02	4.03	4.03	4.02
σ^4_ζ	1.89	1.89	1.89	1.89	1.89	1.89	1.89
σ^8_ζ	2.21	2.21	2.21	2.22	2.21	2.21	2.22

Note. All values are rounded to the second decimal place except for γ. The bold value signifies the binding constraint.

Table 9. KL and empirical distances between ϕ_0 and ϕ_c, SGU(2012) model

	(a) All parameters can vary			(b) σ_μ^8 fixed		
	c=0.1	c=0.5	c=1.0	c=0.1	c=0.5	c=1.0
KL	1.44E-07	1.14E-06	1.02E-05	1.71E-07	3.48E-06	1.03E-05
T=80	0.0505	0.0514	0.0527	0.0505	0.0524	0.0543
T=150	0.0507	0.0519	0.0539	0.0507	0.0534	0.0559
T=200	0.0508	0.0522	0.0546	0.0509	0.0539	0.0569
T=1000	0.0518	0.0551	0.0617	0.0519	0.0592	0.0665

Note. KL is defined as $KL_{ff}(\phi_0, \phi_c)$ with ϕ_c given in the columns of Table 9. The empirical distance equals $p_{ff}(\phi_0, \phi_c, 0.05, T)$, where T is specified in the last four rows of the Table. The KL and empirical distance values computed in quadruple precision are identical to the ones reported above when rounded to the same decimal.

Table 10. The closest models with news shocks shut down, SGU (2012) model

	ϕ_0	All news shut down	4-period news shut down	8-period news shut down
KL	–	0.0773	0.0110	0.0037
T=200	-	0.9993	0.6828	0.3129
θ	4.74	4.42	4.68	4.71
γ	0.0019	0.0011	0.0011	0.0019
κ	9.11	10.80	8.86	9.24
δ_2/δ_1	0.34	0.29	0.34	0.33
b	0.91	0.93	0.91	0.91
ρ_{xg}	0.72	0.72	0.72	0.72
ρ_{μ^a}	0.48	0.47	0.48	0.48
ρ_{μ^x}	0.38	0.38	0.38	0.38
ρ_{z^i}	0.47	0.73	0.50	0.47
ρ_z	0.92	0.93	0.92	0.92
ρ_μ	0.98	0.99	0.98	0.98
ρ_g	0.96	0.96	0.96	0.96
ρ_ζ	0.17	0.24	0.17	0.18
$\sigma^0_{\mu^a}$	0.21	0.31	0.23	0.22
$\sigma^4_{\mu^a}$	0.16	0	0	0.22
$\sigma^8_{\mu^a}$	0.16	0	0.21	0
$\sigma^0_{\mu^x}$	0.38	0.40	0.38	0.38
$\sigma^4_{\mu^x}$	0.08	0	0	0.13
$\sigma^8_{\mu^x}$	0.10	0	0.13	0
$\sigma^0_{\tilde{z}^i}$	11.72	10.28	11.26	11.70
$\sigma^4_{\tilde{z}^i}$	1.93	0	0	6.29
$\sigma^8_{\tilde{z}^i}$	5.50	0	5.12	0
$\sigma^0_{\tilde{z}}$	0.65	0.67	0.65	0.65
$\sigma^4_{\tilde{z}}$	0.11	0	0	0.14
$\sigma^8_{\tilde{z}}$	0.09	0	0.14	0
σ^0_μ	0.50	4.16	1.28	0.82
σ^4_μ	4.79	0	0	4.74
σ^8_μ	0.51	0	4.60	0
σ^0_g	0.62		0.68	0.63
σ^4_g	0.57	0.92	0	0.67
σ^8_g	0.37	0	0.62	0
σ^0_ζ	4.03	0	3.80	4.33
σ^4_ζ	1.89	6.73	0	2.64
σ^8_ζ	2.21	0	3.10	0

Note. All values are rounded to the second decimal place except for γ. The KL and empirical distance values computed in quadruple precision are identical to the ones reported above when rounded to the same decimal.

Table 11. The closest models with individual news shocks shut down, SGU (2012) model

	ϕ_0	μ_t^a	μ_t^x	z_t^i	z_t	μ_t	g_t	ζ_t
KL	–	0.0031	9.07E-04	0.0046	1.75E-04	0.0536	0.0019	5.65E-04
T=200	-	0.2912	0.1394	0.3669	0.0823	0.9972	0.2214	0.1212
θ	4.74	4.80	4.71	4.49	4.67	4.21	4.66	4.69
γ	0.0019	0.0020	0.0021	0.0014	0.0019	7.00E-06	0.0018	0.0022
κ	9.11	9.22	9.28	8.63	9.26	11.36	9.22	9.16
δ_2/δ_1	0.34	0.34	0.34	0.29	0.34	0.27	0.34	0.34
b	0.91	0.91	0.91	0.91	0.91	0.92	0.91	0.91
ρ_{xg}	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72
ρ_{μ^a}	0.48	0.47	0.48	0.48	0.48	0.48	0.48	0.48
ρ_{μ^x}	0.38	0.38	0.38	0.38	0.38	0.39	0.38	0.38
ρ_{z^i}	0.47	0.47	0.47	0.64	0.47	0.61	0.47	0.47
ρ_z	0.92	0.92	0.92	0.92	0.92	0.93	0.92	0.92
ρ_μ	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
ρ_g	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
ρ_ζ	0.17	0.17	0.18	0.18	0.17	0.19	0.17	0.16
$\sigma^0_{\mu^a}$	0.21	0.31	0.21	0.21	0.21	0.21	0.21	0.21
$\sigma^4_{\mu^a}$	0.16	0	0.16	0.16	0.16	0.16	0.16	0.16
$\sigma^8_{\mu^a}$	0.16	0	0.16	0.16	0.16	0.16	0.16	0.16
$\sigma^0_{\mu^x}$	0.38	0.38	0.40	0.38	0.38	0.38	0.38	0.38
$\sigma^4_{\mu^x}$	0.08	0.08	0	0.06	0.08	4.54E-06	0.08	0.10
$\sigma^8_{\mu^x}$	0.10	0.10	0	0.11	0.10	0.11	0.10	0.08
$\sigma^0_{z^i}$	11.72	11.90	11.93	9.59	11.91	12.55	11.83	11.78
$\sigma^4_{z^i}$	1.93	2.08	2.03	0	2.08	6.26E-05	1.98	2.14
$\sigma^8_{z^i}$	5.50	5.58	5.51	0	5.52	4.23	5.48	5.50
σ^0_z	0.65	0.65	0.65	0.65	0.67	0.67	0.65	0.65
σ^4_z	0.11	0.10	0.08	0.09	0	2.81E-06	0.10	0.14
σ^8_z	0.09	0.11	0.11	0.11	0	7.16E-06	0.09	1.87E-07
σ^0_μ	0.50	0.31	0.66	1.43	0.59	3.90	0.56	0.53
σ^4_μ	4.79	4.85	4.76	4.35	4.71	0	4.70	4.75
σ^8_μ	0.51	0.62	0.37	1.63E-04	0.68	0	0.71	0.38
σ^0_g	0.62	0.62	0.62	0.63	0.62	0.63	0.92	0.62
σ^4_g	0.57	0.57	0.56	0.56	0.57	0.54	0	0.58
σ^8_g	0.37	0.37	0.38	0.37	0.37	0.39	0	0.36
σ^0_ζ	4.03	4.06	4.20	4.39	4.09	4.87	4.10	4.97
σ^4_ζ	1.89	1.82	1.70	2.73	1.83	1.75	1.83	0
σ^8_ζ	2.21	2.24	2.23	0.0011	2.27	2.46	2.25	0

Note. All values are rounded to the second decimal place except for γ. The KL and empirical distance values computed in quadruple precision are identical to the ones reported above when rounded to the same decimal.

Online Appendix A. Model Solution under Indeterminacy

This appendix outlines the main steps for solving a DSGE model allowing for indeterminacy. The algorithm is essentially the same as Lubik and Schorfheide's (2003) algorithm that is a generalization of Sims (2002). The same algorithm is used by Qu and Tkachenko (2017). The reason for presenting this algorithm here is to help understand the relevant numerical issues that arise.

Applying the QZ decomposition to (1), we have $Q^*\Lambda Z^* = \Gamma_0$, $Q^*\Omega Z^* = \Gamma_1$, where Q and Z are unitary, Λ and Ω are upper triangular. Let $w_t = Z^*S_t$ and premultiply (1) by Q:

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ 0 & \Omega_{22} \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_{1\cdot} \\ Q_{2\cdot} \end{bmatrix} (\Psi\varepsilon_t + \Pi\eta_t), \tag{A.1}$$

where an ordering has been imposed such that the diagonal elements of Λ_{11} (Λ_{22}) are greater (smaller) than those of Ω_{11} (Ω_{22}) in absolute values. Then, because the generalized eigenvalues corresponding to the pair Λ_{22} and Ω_{22} are unstable and ε_t and η_t are serially uncorrelated, the block of equations corresponding to $w_{2,t}$ has a stable solution if and only if $w_{2,0} = 0$ and

$$Q_{2\cdot}\Pi\eta_t = -Q_{2\cdot}\Psi\varepsilon_t \quad \text{for all } t > 0. \tag{A.2}$$

The condition (A.2) determines $Q_{2\cdot}\Pi\eta_t$ as a function of ε_t. However, it may be insufficient to determine $Q_{1\cdot}\Pi\eta_t$, in which case it will lead to indeterminacy.

Because the rows of $Q_{2\cdot}\Pi$ can be linearly dependent, Sims (2002) and Lubik and Schorfheide (2003) suggested to work with its SVD to isolate the effective restrictions imposed on η_t:

$$Q_{2\cdot}\Pi = \begin{bmatrix} U_{\cdot 1} & U_{\cdot 2} \end{bmatrix} \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{\cdot 1}^* \\ V_{\cdot 2}^* \end{bmatrix} = U_{\cdot 1} D_{11} V_{\cdot 1}^*, \tag{A.3}$$

where $\begin{bmatrix} U_{\cdot 1} & U_{\cdot 2} \end{bmatrix}$ and $\begin{bmatrix} V_{\cdot 1} & V_{\cdot 2} \end{bmatrix}$ are unitary matrices and D_{11} is nonsingular. The submatrices $U_{\cdot 1}$ and $V_{\cdot 1}^*$ are unique up to multiplication by a unit-phase factor $\exp(i\varphi)$ (for the real case, up to sign). The spaces spanned by $U_{\cdot 2}$ and $V_{\cdot 2}$ are also unique, although the matrices themselves are not if their column dimensions exceed one. In the latter case, as a normalization, we use the reduced column echelon form for $V_{\cdot 2}$ when implementing the relevant procedures. Note that matrices with the same column space have the same unique reduced column echelon form.

Applying (A.3), (A.2) can be equivalently represented as

$$U_{\cdot 1} D_{11} V_{\cdot 1}^* \eta_t = -Q_{2\cdot}\Psi\varepsilon_t \quad \text{for all } t > 0. \tag{A.4}$$

Premultiplying (A.4) by the conjugate transpose of $\begin{bmatrix} U_{\cdot 1} & U_{\cdot 2} \end{bmatrix}$ does not alter the restrictions because the latter is nonsingular. Thus, (A.4) is equivalent to (using $U_{\cdot 1}^* U_{\cdot 1} = I$ and $U_{\cdot 2}^* U_{\cdot 1} = 0$)

the following two sets of equations: $D_{11}V_{.1}^{*}\eta_t = -U_{.1}^{*}Q_{2.}\Psi\varepsilon_t$ and $0 = -U_{.2}^{*}Q_{2.}\Psi\varepsilon_t$ for all $t > 0$. The second set of equations places no restrictions on η_t. The first set is equivalent to: $V_{.1}^{*}\eta_t = -D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t$. This can be viewed as a system of linear equations of the form $Ax = b$ with $A = V_{.1}^{*}$ and $b = -D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t$. The full set of solutions for such a system is given by

$$\{p + v : v \text{ is any solution to } Ax = 0 \text{ and } p \text{ is a specific solution to } Ax = b\}. \tag{A.5}$$

Here, a specific solution is given by $p = -V_{.1}D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t$, while η_t solves $V_{.1}^{*}\eta_t = 0$ if and only if $\eta_t = V_{.2}\epsilon_t$ with ϵ_t being an arbitrary vector conformable with $V_{.2}$. Therefore, the full set of solutions to (A.4) can be represented as

$$\left\{\eta_t : \eta_t = -V_{.1}D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t + V_{.2}\epsilon_t \text{ with } E_{t-1}\epsilon_t = 0\right\}. \tag{A.6}$$

The restriction $E_{t-1}\epsilon_t = 0$ follows because η_t is an expectation error and $E_{t-1}\varepsilon_t = 0$. This representation is the same as in Proposition 1 in Lubik and Schorfheide (2003).

We now provide some computational details on how to use (A.6) to solve for S_t in (1) as in Sims (2002). Define Φ as the projection of the rows of $Q_{1.}\Pi$ onto those of $Q_{2.}\Pi$: $\Phi = Q_{1.}\Pi V_{.1}D_{11}^{-1}U_{.1}^{*}$. Note that $Q_{1.}\Pi - \Phi Q_{2.}\Pi = Q_{1.}\Pi - Q_{1.}\Pi V_{.1}V_{.1}^{*} = Q_{1.}\Pi\left(I - V_{.1}V_{.1}^{*}\right)$, which equals zero under determinacy. Multiplying (A.1) by

$$\begin{bmatrix} I & -\Phi \\ 0 & I \end{bmatrix}$$

and imposing the restrictions (A.2):

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} - \Phi\Lambda_{22} \\ 0 & I \end{bmatrix}\begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} - \Phi\Omega_{22} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_{1.} - \Phi Q_{2.} \\ 0 \end{bmatrix}\left(\Psi\varepsilon_t + \Pi\eta_t\right).$$

Further, using the expression (A.6),

$$\begin{aligned} \left(Q_{1.} - \Phi Q_{2.}\right)\Pi\eta_t &= \left(Q_{1.}\Pi - \Phi Q_{2.}\Pi\right)\left(-V_{.1}D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t + V_{.2}\epsilon_t\right) \\ &= -Q_{1.}\Pi\left(I - V_{.1}V_{.1}^{*}\right)V_{.1}D_{11}^{-1}U_{.1}^{*}Q_{2.}\Psi\varepsilon_t + Q_{1.}\Pi\left(I - V_{.1}V_{.1}^{*}\right)V_{.2}\epsilon_t. \end{aligned}$$

The first term on the right hand side equals zero. Therefore

$$\begin{aligned} \begin{bmatrix} \Lambda_{11} & \Lambda_{12} - \Phi\Lambda_{22} \\ 0 & I \end{bmatrix}\begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = &\begin{bmatrix} \Omega_{11} & \Omega_{12} - \Phi\Omega_{22} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_{1.} - \Phi Q_{2.} \\ 0 \end{bmatrix}\Psi\varepsilon_t \\ &+ \begin{bmatrix} Q_{1.}\Pi\left(I - V_{.1}V_{.1}^{*}\right) \\ 0 \end{bmatrix}V_{.2}\epsilon_t. \end{aligned}$$

Call the most left hand side matrix G_0. Multiply the above equation by ZG_0^{-1} and using $w_t = Z^* S_t$, we obtain $S_t = \Theta_1 S_{t-1} + \Theta_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t$, where

$$\begin{aligned}
\Theta_1 &= ZG_0^{-1} \begin{bmatrix} \Omega_{11} & \Omega_{12} - \Phi\Omega_{22} \\ 0 & 0 \end{bmatrix} Z^*, \quad \Theta_\varepsilon = ZG_0^{-1} \begin{bmatrix} Q_{1.} - \Phi Q_{2.} \\ 0 \end{bmatrix} \Psi, \\
\Theta_\epsilon &= ZG_0^{-1} \begin{bmatrix} Q_{1.}\Pi \left(I - V_{.1} V_{.1}^*\right) \\ 0 \end{bmatrix} V_{.2}.
\end{aligned}$$

Further, applying the triangular structure of G_0^{-1}, the above matrices can be represented as $\Theta_1 = Z_{.1}\Lambda_{11}^{-1}[\ \Omega_{11} \quad \Omega_{12} - \Phi\Omega_{22}\ ]Z^*$, $\Theta_\varepsilon = Z_{.1}\Lambda_{11}^{-1}\left(Q_{1.} - \Phi Q_{2.}\right)\Psi$ and $\Theta_\epsilon = Z_{.1}\Lambda_{11}^{-1} Q_{1.}\Pi\left(I - V_{.1}V_{.1}^*\right)V_{.2}$, where $Z_{.1}$ includes the first block of columns of Z conformable with Λ_{11}.

Online Appendix B. Analytical Solutions to the Small Scale Model

Case (i): $\alpha > 1$ and $\gamma > 1$ (AMPF). Iterate Equation (16) forward:

$$\widehat{\pi}_t = \frac{1}{\alpha^k} E_t \widehat{\pi}_{t+k} - E_t \sum_{j=0}^{k-1} \frac{1}{\alpha^{j+1}} \varepsilon^r_{t+j},$$

which holds for any $k > 0$. Let $k \to \infty$, we obtain

$$\widehat{\pi}_t = -E_t \sum_{j=0}^{\infty} \frac{1}{\alpha^{j+1}} \varepsilon^r_{t+j}. \tag{B.1}$$

To obtain the solution for $\widehat{b}_t$, apply (B.1) to (17):

$$\widehat{b}_t - \frac{1}{\beta} \sum_{j=0}^{\infty} \frac{1}{\alpha^{j+1}} E_t \varepsilon^r_{t+j} = (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} - \frac{\alpha}{\beta} \sum_{j=0}^{\infty} \frac{1}{\alpha^{j+1}} E_{t-1} \varepsilon^r_{t+j-1} - (1/\beta - 1)\varepsilon^\tau_t + \frac{1}{\beta}\varepsilon^r_{t-1},$$

which is equivalent to

$$\begin{aligned} \widehat{b}_t &= (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} \\ &\quad + \frac{1}{\beta} \sum_{j=0}^{\infty} \frac{1}{\alpha^{j+1}} E_t \varepsilon^r_{t+j} - \frac{\alpha}{\beta} \sum_{j=0}^{\infty} \frac{1}{\alpha^{j+1}} E_{t-1} \varepsilon^r_{t+j-1} - (1/\beta - 1)\varepsilon^\tau_t + \frac{1}{\beta}\varepsilon^r_{t-1}. \end{aligned} \tag{B.2}$$

The solution formulas (B.1) and (B.2) hold for generic stationary shock processes $\{\varepsilon^r_t\}$ and $\{\varepsilon^\tau_t\}$.

In particular, when $\{\varepsilon^r_t\}$ is specified by (14), the solution for $\widehat{\pi}_t$ simplifies to

$$\widehat{\pi}_t = -\frac{1}{\alpha^2} E_t \varepsilon^r_{t+1} - \frac{1}{\alpha}\varepsilon^r_t = \left(-\frac{1}{\alpha} - \frac{1}{\alpha^2}\phi_r\right) e^r_t - \frac{1}{\alpha}\phi_r e^r_{t-1}.$$

The solution for $\widehat{b}_t$ simplifies to

$$\begin{aligned} \widehat{b}_t &= (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} + \frac{1}{\beta}\left(\frac{1}{\alpha}\varepsilon^r_t + \frac{1}{\alpha^2} E_t \varepsilon^r_{t+1}\right) - \frac{\alpha}{\beta}\left(\frac{1}{\alpha}\varepsilon^r_{t-1} + \frac{1}{\alpha^2} E_{t-1}\varepsilon^r_t\right) - (1/\beta - 1)\varepsilon^\tau_t + \frac{1}{\beta}\varepsilon^r_{t-1} \\ &= (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} + \frac{1}{\beta}\left[\left(\frac{1}{\alpha} + \frac{1}{\alpha^2}\phi_r\right) e^r_t + \frac{1}{\alpha}\phi_r e^r_{t-1}\right] - \frac{1}{\alpha\beta}\phi_r e^r_{t-1} - (1/\beta - 1)\varepsilon^\tau_t. \end{aligned}$$

After cancellation, the right hand side expression reduces to that in the paper.

Case (ii): $\alpha < 1$ and $\gamma < 1$ (PMAF). Take the conditional expectation of (17):

$$E_t \widehat{b}_{t+1} + E_t \frac{\widehat{\pi}_{t+1}}{\beta} = (1/\beta - \gamma(1/\beta - 1))\widehat{b}_t + \frac{\alpha}{\beta}\widehat{\pi}_t - (1/\beta - 1)E_t\varepsilon^\tau_{t+1} + \frac{1}{\beta}\varepsilon^r_t.$$

Because of (16), the above equality implies $E_t\widehat{b}_{t+1} = (1/\beta - \gamma(1/\beta - 1))\widehat{b}_t - (1/\beta - 1)E_t\varepsilon^\tau_{t+1}$. To simplify the notation, let $\xi = (1/\beta - \gamma(1/\beta - 1))$. Then, $\widehat{b}_t = (1/\xi)\, E_t\widehat{b}_{t+1} + (1/\xi)\,(1/\beta - 1)E_t\varepsilon^\tau_{t+1}$.

Because $\gamma < 1$, we have $(1/\beta - \gamma(1/\beta - 1)) > 1$. Therefore, the above equation for $\widehat{b}_t$ can be solved forward, leading to

$$\widehat{b}_t = (1/\beta - 1)E_t \sum_{j=1}^{\infty} \frac{1}{\xi^j} \varepsilon^{\tau}_{t+j}.$$

Plugging this expression into (17), we obtain

$$\begin{aligned} \widehat{\pi}_t &= \alpha \widehat{\pi}_{t-1} - \beta(1/\beta - 1)E_t \sum_{j=1}^{\infty} \frac{1}{\xi^j} \varepsilon^{\tau}_{t+j} \\ &\quad + \beta(1/\beta - \gamma(1/\beta - 1))(1/\beta - 1)E_{t-1} \sum_{j=1}^{\infty} \frac{1}{\xi^j} \varepsilon^{\tau}_{t+j-1} - \beta(1/\beta - 1)\varepsilon^{\tau}_t + \varepsilon^{r}_{t-1}. \end{aligned}$$

The above solution for $\widehat{b}_t$ and $\widehat{\pi}_t$ holds for generic stationary shock processes $\{\varepsilon^r_t\}$ and $\{\varepsilon^\tau_t\}$.

In particular, if $\{\varepsilon^\tau_t\}$ is an MA(1) process as in the paper, then

$$\widehat{b}_t = \frac{1/\beta - 1}{\xi} E_t \varepsilon^{\tau}_{t+1} = \frac{1/\beta - 1}{1/\beta - \gamma(1/\beta - 1)} \phi_\tau e^{\tau}_t.$$

This further implies that $\widehat{\pi}_t$ satisfies

$$\begin{aligned} \widehat{\pi}_t &= \alpha \widehat{\pi}_{t-1} - \frac{\beta(1/\beta - 1)}{1/\beta - \gamma(1/\beta - 1)} \phi_\tau e^{\tau}_t \\ &\quad + \beta(1/\beta - 1)\phi_\tau e^{\tau}_{t-1} - \beta(1/\beta - 1)\varepsilon^{\tau}_t + \varepsilon^{r}_{t-1} \\ &= \alpha \widehat{\pi}_{t-1} - \beta(1/\beta - 1)\left(\frac{1}{1/\beta - \gamma(1/\beta - 1)} \phi_\tau + 1 \right) e^{\tau}_t + \varepsilon^{r}_{t-1}. \end{aligned}$$

Case (iii): $\alpha < 1$ and $\gamma > 1$ (PMPF). Define $\eta_t = \widehat{\pi}_t - E_{t-1}\widehat{\pi}_t$. Then, (16) can be written as $\widehat{\pi}_t = \alpha \widehat{\pi}_{t-1} + \varepsilon^r_{t-1} + \eta_t$. Apply this to (17), we obtain

$$\widehat{b}_t = (1/\beta - \gamma(1/\beta - 1))\widehat{b}_{t-1} - (1/\beta - 1)\varepsilon^{\tau}_t - \frac{1}{\beta}\eta_t.$$

Now, view η_t as a disturbance (sunspot shock) to the system. Because the lagged coefficients in the above two equations, α and $(1/\beta - \gamma(1/\beta - 1))$, are both strictly less than 1, these two equations both correspond to stationary and well defined solutions. In contrast, in case (i), $\alpha > 1$; in case (ii), $(1/\beta - \gamma(1/\beta - 1)) > 1$. In either case, one of the two equations needs to be further solved forward to pin down the corresponding expectation error η_t. This is why the multiplicity of solutions does not arise in those two cases.

Online Appendix C. Additional Results from the Applications

1 Local identification of the small scale model based on business cycle frequencies

Result 14 *The local identification properties of the model at the business cycle frequencies are the same as in the full spectrum case reported in Subsection 5.2.1 under the same parameter values considered there.*

The matrix $G^W(\theta^1_{AMPF})$ has rank 6 in both double and quadruple precision. In the former case, the eigenvalues are 25.334, 0.890, 0.035, 1.70E-04, 2.73E-07, 2.91E-10 and -3.18E-15 with the tolerance level of 2.49E-14. In the latter, only the smallest eigenvalue differs noticeably, being much closer to zero at -3.46E-33. We note that the magnitudes of all the eigenvalues are smaller than in the full spectrum case above, reflecting the fact that identification is weaker since less information is used. The subset $(\alpha, \phi_r, \sigma_r)$ is the cause of identification failure, as found previously using the full spectrum. Furthermore, the conclusion about identification in double precision now coincides with that in quadruple precision, which was not the case when the full spectrum was considered.

In the case of θ^2_{AMPF}, the rank of $G^W(\theta^2_{AMPF})$ remains equal to 4 in both double and quadruple precision. As above, the eigenvalues are generally lower in magnitude compared to the full spectrum case: 0.660, 0.046, 7.49E-05, 1.23E-11, 4.13E-17, 8.66E-19 and -1.51E-16. The last three eigenvalues fall below the tolerance level of 7.77E-16. When quadruple precision is used, the three smallest eigenvalues go down further to 1.44E-38, -3.94E-36 and -2.01E-35. The nonidentified subsets and hence the overall conclusions remain the same as in the full spectrum case.

For the generic value of the PMAF regime, the rank of $G^W(\theta^1_{PMAF})$ equals 5 at both precision levels. In double precision, the tolerance level is 6.22E-15, with the eigenvalues being equal to 4.936, 0.110, 1.26E-04, 2.63E-05, 6.86E-10, 1.03E-18 and -9.79E-18. In quadruple precision, the two smallest eigenvalues decrease to 5.93E-36 and -7.29E-40. Again, the conclusions about identification remain the same as in the full spectrum case.

At θ^2_{PMAF}, the rank of $G^W(\theta^2_{PMAF})$ equals 4 across the precision levels. The tolerance level in double precision equals 3.89E-16 and the eigenvalues, the last three of which fall below this level, are 0.438, 0.015, 7.20E-05, 2.19E-08, 7.34E-17, 1.02E-18 and -8.75E-18. In quadruple precision, the three eigenvalues fall further below the tolerance level and equal 2.04E-35, 6.44E-39 and -6.95E-38. The conclusion about identification thus is the same as in the full spectrum case.

Finally, we consider the parameter θ_{PMPF}, as well as the 1000 randomly drawn parameter

values from the PMPF region. The matrix $G^W(\theta_{PMPF})$ is found to be full rank in both double and quadruple precision. The default tolerance level in double precision is 1.42E-13, and the eigenvalues obtained are all above it: 64.752, 16.886, 2.421, 0.696, 0.114, 0.001, 4.84E-05, 3.72E-06, 1.65E-08 and 5.05E-10. In quadruple precision the smallest eigenvalue does not change noticeably remaining at 5.05E-10. As with θ^1_{AMPF}, the spurious lack of local identification that appeared in the full spectrum case is no longer detected in double precision. Among the 1000 randomly generated points, the ranks of $G^W(\theta)$ vary between 8 and 10, with 248 cases found to have local identification failure. Thus, the detection of spurious identification failure due to numerical issues in double precision still persists when using only the business cycle frequencies, although to a notably lesser extent than in the full spectrum case. In quadruple precision all the criterion matrices have full rank as in the full spectrum case. The parameter vectors in the indeterminacy region thus are also locally identifiable using only the business cycle frequencies.

Therefore, the conclusions regarding local identifiability are always the same as in the full spectrum case. The results show that, using the suggested procedures and algorithms, it is feasible to obtain clear cut conclusions regarding local identification using business cycle frequencies.

2 Local identification of the medium scale model based on business cycle frequencies

Result 15 *The local identification properties of the model at the business cycle frequencies are the same as in the full spectrum case reported in Subsection 6.1 under the same parameter values considered there.*

Starting with the observables in (29), we find that $G^W(\phi_0)$ has full rank in double precision, with five smallest eigenvalues being 6.16E-12, 1.66E-12, 9.80E-13, 6.71E-13, 1.07E-13. The respective tolerance level equals 4.72E-16. Recomputing the result in quadruple and 50-digit precision does not show any noticeable change in eigenvalues thus confirming local identification.

We now consider the original set of observables from Schmitt-Grohé and Uribe (2012). We find that $G^W(\phi_0)$ has full rank in double precision, with five smallest eigenvalues being 3.53E-05, 2.83E-05, 1.24E-05, 4.24E-06, 1.18E-06. The respective tolerance level equals 4.97E-13. Obtaining the result in quadruple and 50-digit precision again confirms local identification.

3 Global identification of the medium scale model based on business cycle frequencies

We repeat the global identification analysis using only the business cycle frequencies. We also examine whether our results obtained for the observables in (29) are robust to switching to the growth rate observables considered in Schmitt-Grohé and Uribe (2012). The empirical distance measures reported below refer to the case of $T = 200$ unless specified otherwise

Result 16 *The global identification properties are qualitatively similar to those found based on the full spectrum. The main difference is that near observational equivalence is more pronounced: the resulting empirical distances are often three times lower than those obtained with the full spectrum. In particular, the empirical distance equals only 0.2617 at* $T = 200$ *when the original shock specification is replaced by (31). These findings are corroborated by the results obtained using observables in growth rates.*

When searching outside the neighborhoods $B(\phi_0)$, σ_μ^8 and σ_{zi}^4 again emerge as weakly identified as c increases from 0.1 to 1.0, and the corresponding minimizing parameter values remain largely the same as the full spectrum case. The KL measures equal 4.24E-08, 3.37E-07 and 3.57E-06 for c =0.1, 0.5 and 1.0, respectively, while the corresponding empirical distance measures equal 0.0509, 0.0507 and 0.0594. Virtually the same results are obtained from additional optimization in quadruple precision. Repeating the searches for the closest models while fixing the weakly identified parameters sequentially yields the same parameters as found previously in (30). When these seven weakly identified parameters are fixed, the minimized KL for $c = 1.0$ equals 2.06E-04 and the empirical distance measure is 0.0961. The parameter values minimizing KL do not noticeably change when compared to the full spectrum case. We can see that it becomes substantially more difficult to distinguish this model from the benchmark at the business cycle frequencies. These results are robust to changing the observables to growth rates: very similar parameter values are obtained, with the resulting KL and empirical distances generally slightly below those shown above.

Next, we repeat the global identification analysis of the anticipated shocks at business cycle frequencies. The results remain qualitatively similar when the news shocks are shut down altogether as well as horizon by horizon. The closest model without the news shocks still remains relatively easily distinguishable from the benchmark in finite samples with the minimized KL of 0.0191 and the empirical distance of 0.8681. The 4-horizon news shocks still remain more important than the 8-horizon ones, however, in both cases the ability to differentiate the closest model without respective shocks decreases substantially. In the former case, when compared to the full spectrum results obtained earlier, the KL decreases from 0.0110 to 0.0041 and the empirical distance shrinks

to 0.3637. In the latter case, the KL equals 0.0010 with the corresponding empirical distance 0.1549. When the relative importance of individual news shocks is considered, the wage markup shock again emerges as the most crucial in forming the dynamics. Using only business cycle frequencies, the closest model without news in the wage markup is still readily distinguishable in finite samples with the KL of 0.0150 and the empirical distance of 0.8074. As in the full spectrum case, news in the investment-specific productivity shock $\widehat{z}_t^i$ remain relatively important, with the difference that at the business cycle frequencies news in $\widehat{\mu}_t^a$ have a similar, even marginally higher cost of shutting down. The minimized KL are very similar for the two cases: 0.0009 for the former and 0.0010 for the latter, which lead to respective empirical distances of 0.1507 and 0.1523. The findings for the rest of the shocks also closely mirror the full spectrum case. Shutting down news in $\widehat{g}_t$ delivers the fourth largest KL of 0.0007 with the corresponding empirical distance of 0.1273. The other three shocks, $\widehat{\mu}_t^x$, $\widehat{\zeta}_t$ and $\widehat{z}_t$, remain the least important: the respective empirical distances equal 0.0905, 0.0840 and 0.0659. When growth rates are used as observables instead, the results obtained are qualitatively very similar, with slightly lower KL and empirical distances in all cases.

Finally, we revisit the analysis of identification from alternative shock structures. When all the shocks are replaced by (31), the minimized KL equals 0.0025 and the empirical distance equals 0.2617. Therefore, when only business cycle frequencies are considered, it is possible to approximate the original model dynamics much closer using models with the ARMA shock processes as in (31). When the alternative shock structure is imposed on one exogenous process at a time, we obtain results that qualitatively mirror those obtained using the full spectrum. In particular, in the case of the wage markup shock, we again find near equivalence. Specifically, the minimized KL equals 2.77E-06 and the empirical distance equals 0.0519. Furthermore, the empirical distance at $T = 1000$ is still only 0.0579. These findings are corroborated by the results obtained using observables in growth rates: there, qualitatively similar parameter values lead to the closest models with marginally lower KL and empirical distances.

Table C1. List of structural parameters in the SGU(2012) model

Parameter	SGU posterior median	Interpretation
θ	4.74	Frisch elasticity of labor supply (when $\gamma = b = 0$)
γ	0.0019	Governs wealth elasticity of labor supply
κ	9.11	Investment adjustment cost parameter
δ_2/δ_1	0.34	Ratio of depreciation parameters (see Section 6)
b	0.91	Habit parameter
ρ_{xg}	0.72	AR coeff. of government spending trend
ρ_{μ^a}	0.48	AR coeff. of nonstationary investment-specific prod. shock
ρ_{μ^x}	0.38	AR coeff. of nonstationary neutral productivity shock
ρ_{z^i}	0.47	AR coeff. of stationary investment shock
ρ_z	0.92	AR coeff. of stationary neutral productivity shock
ρ_μ	0.98	AR coeff. of wage markup shock
ρ_g	0.96	AR coeff. of gov. spending shock
ρ_ζ	0.17	AR coeff. of the preference shock
$\sigma^0_{\mu^a}$	0.21	Std. dev. of unanticipated shock in μ^a_t
$\sigma^4_{\mu^a}$	0.16	Std. dev. of 4-period anticipated shock in μ^a_t
$\sigma^8_{\mu^a}$	0.16	Std. dev. of 8-period anticipated shock in μ^a_t
$\sigma^0_{\mu^x}$	0.38	Std. dev. of unanticipated shock in μ^x_t
$\sigma^4_{\mu^x}$	0.08	Std. dev. of 4-period anticipated shock in μ^x_t
$\sigma^8_{\mu^x}$	0.10	Std. dev. of 8-period anticipated shock in μ^x_t
$\sigma^0_{z^i}$	11.72	Std. dev. of unanticipated shock in z^i_t
$\sigma^4_{z^i}$	1.93	Std. dev. of 4-period anticipated shock in z^i_t
$\sigma^8_{z^i}$	5.50	Std. dev. of 8-period anticipated shock in z^i_t
σ^0_z	0.65	Std. dev. of unanticipated shock in z_t
σ^4_z	0.11	Std. dev. of 4-period anticipated shock in z_t
σ^8_z	0.09	Std. dev. of 8-period anticipated shock in z_t
σ^0_μ	0.50	Std. dev. of unanticipated shock in μ_t
σ^4_μ	4.79	Std. dev. of 4-period anticipated shock in μ_t
σ^8_μ	0.51	Std. dev. of 8-period anticipated shock in μ_t
σ^0_g	0.62	Std. dev. of unanticipated shock in g_t
σ^4_g	0.57	Std. dev. of 4-period anticipated shock in g_t
σ^8_g	0.37	Std. dev. of 8-period anticipated shock in g_t
σ^0_ζ	4.03	Std. dev. of unanticipated shock in ζ_t
σ^4_ζ	1.89	Std. dev. of 4-period anticipated shock in ζ_t
σ^8_ζ	2.21	Std. dev. of 8-period anticipated shock in ζ_t
$\sigma^{me}_{g^y}$	0.30	Std. dev. of measurement error in output growth