A Composite Likelihood Framework for Analyzing Singular DSGE Models*

Zhongjun Qu

Boston University

June 24, 2015

Abstract

This paper builds upon the composite likelihood concept of Lindsay (1988) to develop a framework for parameter identification, estimation, inference and forecasting in DSGE models allowing for stochastic singularity. The framework consists of the following four components. First, it provides a necessary and sufficient condition for parameter identification, where the identifying information is provided by the first and second order properties of the nonsingular submodels. Second, it provides an MCMC based procedure for parameter estimation. Third, it delivers confidence sets for the structural parameters and the impulse responses that allow for model misspecification. Fourth, it generates forecasts for all the observed endogenous variables, irrespective of the number of shocks in the model. The framework encompasses the conventional likelihood analysis as a special case when the model is nonsingular. Importantly, it enables the researcher to start with a basic model and then gradually incorporate more shocks and other features, meanwhile confronting all the models with the data to assess their implications. The methodology is illustrated using both small and medium scale DSGE models. These models have numbers of shocks ranging between one and seven.

Keywords: Business cycle, dynamic stochastic general equilibrium models, identification, impulse response, MCMC, stochastic singularity.

JEL Codes: C13, C32, C51, E1.

*First version: October, 2013. We thank Adam McCloskey, Pierre Perron, Mototsugu Shintani, seminar participants at BU, Maryland and Vanderbilt for comments and suggestions, and Junwen Lu and Andres Sagner for excellent research assistance. Contact: Department of Economics, Boston University, Boston, MA, 02215 (qu@bu.edu).

1 Introduction

Economic theory allows the number of structural shocks in DSGE models to be different from the number of observed endogenous variables. When the former is smaller than the latter, the model becomes stochastically singular. This poses a challenge for estimation, inference and forecasting. Several approaches have been undertaken to bridge the gap between likelihood based methods and stochastic singularity. The first approach allows for measurement errors in the observables by building upon the idea of Sargent (1989); see Altug (1989), McGrattan (1994), Hall (1996), McGrattan, Rogerson and Wright (1997) and Ireland (2004). Although this approach is widely applicable, the actual content of these errors can be ambiguous. The second approach adds structural shocks to the model to make it nonsingular. This alters the economic model, which may or may not reflect the intention of the researcher. As theory progresses, DSGE models are expected to take on the challenge of incorporating additional endogenous variables (e.g., those from the financial or the fiscal sector). Therefore, allowing for a flexible link between the number of structural shocks and endogenous variables can become even more desirable.

The third approach involves treating some of the observables as unobserved when constructing the likelihood. This approach maintains the model's structure and avoids the introduction of measurement errors. Studies have documented that different choices of observables can have large impacts on parameter identification, estimation and forecasting, see Fernández-Villaverde and Rubio-Ramírez (2007), Guerron-Quintana (2010) and Del Negro and Schorfheide (2013). Recently, Canova, Ferroni and Matthes (2014) drew further attention to this issue. They proposed two methods for choosing exactly k observables for a model with k structural shocks by building on the convolution idea of Bierens (2007) and the identification condition in Komunjer and Ng (2011). However, under stochastic singularity, the decision to exclude observables often is *not* motivated by economic considerations, but rather because otherwise limited econometric methods are available. Thus, it is desirable to break this rigid link, embracing that there is often no compelling economic reason for why the number of structural shocks should necessarily determine the number of observables used for estimation.

This paper develops a likelihood based framework for analyzing DSGE models, without requiring adding measurement errors, introducing new structural shocks or excluding observables from the estimation. It builds on the composite likelihood concept of Lindsay (1988). The composite likelihood is a likelihood based object formed by multiplying individual component likelihoods, each of which corresponds to a marginal or conditional event. It has found applications in diverse areas, particularly in spatial statistics, where the complex dependence between variables makes implementing the full likelihood impractical. Here, the issue of complex dependencies is irrelevant. Rather, the idea of considering component likelihoods provides a solution for handling singularity. Specifically, in a model with n observables and k ($k < n$) shocks, the subsets that include no more than k observables are typically nonsingular. For any such subset, one can write down the likelihood in either the time or the frequency domain. A composite likelihood can then be formed by multiplying some or all of these individual components. Consequently, all the observables can enter the estimation through the component likelihoods, irrespective of the number of shocks in the model. The researcher can still flexibly add structural shocks or measurement errors, but only when doing so is considered desirable. The composite likelihood reduces to the conventional likelihood if the model is stochastically nonsingular.

The framework developed here consists of the following four components. First, it provides a necessary and sufficient condition for local identification, where the identifying information is provided by the first and second order properties of the nonsingular submodels. This condition extends the results in Qu and Tkachenko (2012). Second, it provides an MCMC based procedure for parameter estimation. The procedure builds on the work of Chernozhukov and Hong (2003) and An and Schorfheide (2007). Third, it proposes methods for obtaining confidence sets for the structural parameters and the impulse responses using the MCMC draws and the properties of the model. Finally, it suggests a procedure that can generate forecasts for all the observed endogenous variables, even if the number of structural shocks is as small as one.

In practice, arriving at a satisfactory model can be a gradual process. The composite likelihood framework enables the researcher to start with a basic model and then gradually incorporate more shocks and other features, meanwhile confronting all the models with the data to assess their implications. In addition, conditional on any intermediate model, different composite likelihoods can be constructed and estimated using different sets of submodels. This can potentially reveal the shortcomings of the model, therefore being informative about what additional shocks are desirable to further improve it. Later in the paper, these features will be illustrated through considering both small and medium scale DSGE models.

Specifically, the models considered are singular versions of two influential models in the literature. The first is a prototypical three-equation model, studied in Clarida, Gali and Gertler (2000) and Lubik and Schorfheide (2004). The second is the model of Smets and Wouters (2007). The resulting models have between one and seven shocks. The findings can be summarized as follows. (1) Among the structural parameters, the estimates related to the steady state tend to remain stable across specifications, while those related to the productivity process and the frictions can vary substantially. (2) The estimated effect of a particular shock (e.g., the productivity shock) can crucially depend on what other shocks are allowed in the model. (3) For the small scale models considered, whether or not to include the monetary policy shock has little effects on the estimated responses to the productivity shock, while for the medium scale models, whether or not to include the wage markup and risk premium shocks has little effects on the estimated responses to the productivity, investment, monetary policy and exogenous spending shocks. (4) There can exist different parameter values that yield similar impulse responses for some shocks while very different responses for others. This reflects an identification issues which implies that relying on matching impulse responses to a particular shock can be insufficient for determining all the parameters. (5) Overall, the composite likelihood framework is informative not only for detecting the above similarities and differences, but also for pinpointing the sources (i.e., which parameters and their values) that generate them.

In this paper, for both the theoretical analysis and the empirical illustration, the following perspective has been fundamental. That is, a DSGE model is an imperfect approximation to the true data generating process with the stochastic singularity being among the potential misspecifications. This perspective suggests, as with other misspecifications, that one should carefully assess the effect of the singularity on the model rather than assuming it away (i.e., treating some observables as unobserved) or ruling out singular models altogether. The composite likelihood framework provides a platform for analyzing such models with the results explicitly acknowledging misspecification. Importantly, the value of the framework is not in providing a unique estimation criterion function that achieves highest efficiency, but rather in allowing the researcher to experiment with different combinations of component likelihoods and confronting all such choices with the data. In this regard, it is related to the literature that studies dynamic general equilibrium models while explicitly acknowledging their misspecifications. This includes, among others, Watson (1993), Hansen and Sargent (1993), Diebold, Ohanian and Berkowitz (1998), Schorfheide (2000), Bierens (2007) and Del Negro and Schorfheide (2009).

The work here is also related to the following contributions that embrace stochastic singularity: the generalized method of moments (Hansen, 1982), the simulated method of moments (Lee and Ingram, 1991 and Duffie and Singleton, 1993), the indirect inference (Smith, 1993, Gouriéroux, Monfort and Renault, 1993, and Gallant and Tauchen, 1996). Recently, valuable efforts have been made in adapting these methods to the current generation of DSGE models; see Ruge-Murcia (2007) and Andreasen, Fernández-Villaverde and Rubio-Ramírez (2013). In contrast to this paper, the above methods are not likelihood based. That is, they all use criteria other than *model implied* densities to link the model with the data.

The paper is structured as follows. Section 2 specifies the model and characterizes stochastic singularity. Section 3 introduces the composite likelihood idea to the current problem. Section 3 studies parameter identification, while Section 4 discusses parameter estimation and obtains confidence sets. Section 5 studies the impulse response function. Section 6 considers forecasting. Section 7 includes some empirical illustrations. The three appendices, A, B and C, contain the proofs, a medium scale model and some details on implementation, respectively.

2 Stochastically singular DSGE models

This paper considers DSGE models that are representable as

$$\begin{aligned} Y_t &= \mu(\theta) + C(\theta)X_t + D(\theta)v_t, \\ X_t &= A(\theta)X_{t-1} + B(\theta)\varepsilon_t, \end{aligned} \tag{1}$$

where the n-by-1 vector Y_t includes the measured variables, X_t is a vector of state variables that includes the endogenous variables, conditional expectation terms and exogenous shocks processes if they are serially correlated, ε_t includes serially uncorrelated structural disturbances and v_t contains measurement errors if there are any. The vector θ consists of the structural parameters. The coefficients matrices $\mu(\theta), A(\theta), B(\theta), C(\theta), D(\theta)$ are functions of θ. Throughout the paper, we assume θ takes values in a parameter space Θ that is a subset of a Euclidean space of dimension q.

The above representation encompasses the current generation of DSGE models, for example Smets and Wouters (2007). For simplicity, the measurement errors are assumed to be serially uncorrelated. Otherwise, as in Ireland (2004), a subset of equation can be appended to the system to describe the time evolution of v_t. The representation also encompasses indeterminacy after augmenting ε_t and θ to include the sunspot shocks and the corresponding parameters. Such an extension follows from the Proposition 1 in Lubik and Schorfheide (2004).

The system (1) has a vector moving average representation:

$$Y_t = \mu(\theta) + H(L;\theta)\epsilon_t, \tag{2}$$

where

$$H(L;\theta) = \begin{bmatrix} C(\theta)(I - A(\theta)L)^{-1}B(\theta), & D(\theta) \end{bmatrix}, \tag{3}$$

and

$$\epsilon_t = \begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix}.$$

This representation is not needed for estimation, but is useful for formulating the theoretical results on identification and inference. The spectral density matrix of Y_t at the frequency $\omega \in [-\pi, \pi]$ equals

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

where $\Sigma(\theta) = \mathrm{Var}(\epsilon_t)$ and the superscript "$*$" stands for the conjugate transpose. The next definition specifies the type of stochastic singularity considered in this paper.

Definition 1 *The DSGE model (1) is stochastically singular at $\theta = \theta_0$ if there exists a partition of the observables $Y_t = [Y_{1,t}, Y_{2,t}']'$ with $Y_{1,t} \in \mathbb{R}^1$ such that, for all t,*

$$Y_{1,t} = \sum_{j=0}^{\infty} g_j(\theta_0) Y_{t-j},$$

where $\{g_j(\theta_0)\}_{j=0}^{\infty}$ are coefficients matrices with the (1,1)-th element of $g_0(\theta_0)$ being zero.

The model is stochastically singular when some variables can be perfectly predicted from its own past values and the current and lagged values of the other variables. Under stochastic singularity, the covariance matrix, $\mathrm{Var}(Y_t)$, can still be of full rank. For example, suppose $Y_{1,t} = \epsilon_{t-1}$ and $Y_{2,t} = \epsilon_t$. Then, $Y_{1,t}$ is known from observing $Y_{2,t-1}$ (therefore Y_t is stochastic singular), although the covariance matrix of Y_t still has full rank. Consequently, characterizing stochastic singularity requires more than just $\mathrm{Var}(Y_t)$. In contrast, from a frequency domain perspective, (1) is stochastic singular if and only if the spectral density matrix of Y_t is of reduced rank at all frequencies. This makes the spectral density a valid characterizing condition for stochastic singularity.

The next lemma relates the above definition to the most common cause of stochastic singularity in DSGE models. Let $\dim(\cdot)$ denote the dimension of a vector.

Lemma 1 *If $\dim(v_t) + \dim(\varepsilon_t) < \dim(Y_t)$, then the model (1) is stochastically singular at all $\theta \in \Theta$.*

It is well known that the conventional time and frequency domain Gaussian likelihoods are not well defined when the model is stochastically singular. Specifically, in the time domain, the density of Y_t given Y_{t-j} $(j = 1, 2, ...)$ is not well defined because the resulting conditional covariance matrix is singular (c.f., Definition 1). Algorithmically, when implementing the Kalman filter, the prediction step (i.e., predicting Y_t given Y_{t-j} $(j = 1, 2, ...)$) produces a singular covariance matrix, causing the updating step (i..e., estimating the distribution of the state vector $C(\theta)X_t$ after observing Y_t) to break down. In the frequency domain, the spectral density matrix of Y_t is singular. Because its inverse enters the likelihood, the later also fails to be well defined.

3 The composite likelihood

The composite likelihood method was proposed by Lindsay (1988). Its precedents are the pseudo-likelihood of Besag (1974, 1975) and the partial likelihood of Cox (1975). A major motivation and application for this method has been geostatistics. Below, we review this method in its original setting (using Example 3A in Lindsay, 1988) to contrast with its application in the current context.

Suppose we observe y_i on a lattice of sites indexed by i ($i = 1, ..., N$). The observation at site i depends on the values at its neighbors. That is, the conditional distribution of y_i given the remaining observations (denoted by $y_{[i]}$) is given by $(y_i|y_{[i]}) \sim N(\theta w_i' y, \tau^2)$, where $y = (y_1, ..., y_N)$, θ and τ^2 are parameters, and w_i is a N-by-1 vector whose j-th element equals to 1 if i and j are neighbors and zero otherwise. The Hammersley-Clifford theorem implies that the joint distribution of y is unique and given by $y \sim N(0, \sigma^2 (1 - \theta W)^{-1})$, where $W = [w_1, ..., w_N]$ and σ^2 is a function of τ^2 and θ. Assuming σ^2 is known and equals 1, the log likelihood equals (up to a constant) $\kappa(\theta) - \theta y' W y/2$ with $\kappa(\theta) = (1/2) \log \det (1 - \theta W)$. Maximizing the likelihood involves evaluating $\kappa(\theta)$ and computing its derivative with respect to θ, both of which can lead to computational difficulties because N is typically large. To bypass this difficulty, Besag (1974) suggested considering the sum of the conditional log likelihoods:

$$\ell(\theta) = \sum_{i=1}^{N} \log f(y_i|y_{[i]}; \theta). \tag{4}$$

Taking the first order derivative delivers $y' W y - \theta y' W^2 y/2 = 0$, which is now straightforward to evaluate. Hjort and Omre (1994) suggested considering the pairwise log likelihood:

$$\ell(\theta) = \sum_{i=1}^{N-1} \sum_{r=i+1}^{N} \log f\left(y_i, y_r | d_{ir}; \theta\right), \tag{5}$$

where d_{sr} is some measure for the relationship between the two sites. Both (4) and (5) are members of the composite likelihood family. The general principle for the latter was laid out in Lindsay (1988): one starts with a set of conditional (e.g., (4)) or marginal (e.g., (5)) events for which one can write the log likelihood; then one constructs the composite log likelihood as the sum of the component log likelihoods. The composite likelihood has found applications in diverse areas featuring complex dependencies between variables. This includes: spatial data, genetics/genomics data, image data and longitudinal data. Reviews with more applications can be found in Varin, Reid and Firth (2011). Recently, Engle, Shephard and Sheppard (2007) introduced the method to estimate the time varying covariances of a portfolio with a vast number of assets. There, the component likelihoods are as in (5), involving pairs of assets in the portfolio.

We now bring the idea to analyze singular DSGE models. Let $Y_{s,t}$ be a subvector of Y_t in (1), i.e., $Y_{s,t} = P_s Y_t$ with P_s being a selection matrix. Then, $Y_{s,t}$ satisfies

$$\begin{aligned} Y_{s,t} &= P_s\mu(\theta) + P_s C(\theta) X_t + P_s D(\theta) v_t, \\ X_t &= A(\theta) X_{t-1} + B(\theta)\varepsilon_t. \end{aligned}$$

Its vector moving average representation is

$$Y_{s,t} = P_s\mu(\theta) + P_s H(L;\theta)\epsilon_t. \tag{6}$$

Its spectral density at $\omega \in [-\pi, \pi]$ equals

$$f_s(\omega;\theta) = \frac{1}{2\pi} P_s H(\exp(-i\omega);\theta)\Sigma(\theta) H(\exp(-i\omega);\theta)^* P_s^* = P_s f(\omega;\theta) P_s^*. \tag{7}$$

The relationship (6) can be called a submodel because it is consistent with the full model (1) but involves only a subset of its restrictions. The vector of observables for this submodel is $Y_{s,t}$, a subset of Y_t. The next definition defines a (maximal) nonsingular submodel.

Definition 2 *The submodel (6) is called a nonsingular submodel if it is stochastically nonsingular for all $\theta \in \Theta$. It is a maximal nonsingular submodel if, in addition, augmenting $Y_{s,t}$ with any variable from Y_t will always make the resulting submodel stochastically singular for some $\theta \in \Theta$.*

The likelihood functions for the nonsingular submodels are simple to obtain. In the time domain, the Gaussian likelihoods can be obtained using the standard Kalman filtering algorithm. In the frequency domain, the inverses of the spectral densities of the nonsingular submodels can be obtained directly from (7). The computational details are included in the appendix.

The motivation for applying the composite likelihood concept to DSGE models is different from the geostatistics setting and can be stated as follows. First, the nonsingular submodels (6) are consistent with the full model, all possessing well defined likelihood functions. The Hammersley-Clifford theorem is easily applicable. This special feature provides the opportunity for constructing component likelihoods and subsequently the composite likelihood. Second, DSGE models are imperfect approximations to the data generating process. Stochastic singularity is typically a misspecification. It is therefore desirable to match only the nonsingular relationships (i.e., those implied by

(6)) with the data. This makes the composite likelihood not a shortcut to circumvent singularity, but a desirable method to relate misspecified models to the data.

Specifically, let $Y_{1,t}, ..., Y_{S,t}$ be some subvectors of Y_t that are stochastically nonsingular, each satisfying (6) for some P_s with S being some positive integer. Denote their corresponding log likelihood functions by

$$\ell_s(\theta) \quad (s = 1, ..., S).$$

We propose to construct the composite log likelihood as:

$$\ell(\theta) = \sum_{s=1}^{S} \ell_s(\theta). \tag{8}$$

The above construction has two features. First, it allows for arbitrary relationships between the numbers of observables and structural shocks. Therefore, it is feasible to keep all the observed endogenous variables for estimation even if the model has only one shock. Second, if the original full model is already nonsingular, then $\ell_s(\theta) = \ell(\theta)$ and we obtain the conventional log likelihood. This implies the framework encompasses the conventional likelihood analysis as a special case.

Remark 1 *As is clear from Lindsay's (1988) general principle, we will arrive at different composite likelihood functions depending on the events (here the submodels) that we consider. There are two potential perspectives for approaching this issue. The first is to treat it as an efficiency issue. That is, we assume that the model is correctly specified and choose submodels to maximize the asymptotic efficiency. However, this is irrelevant here because misspecifications are clearly present. The second option is to treat it as a specification issue. That is, we decide on what the model is intended to capture and then choose the submodels accordingly. This option is what we will implement. Consequently, the value of the composite likelihood framework developed here is not in delivering a unique criterion function that achieves the highest efficiency, but in providing a platform that allows for flexible choices of criterion functions, and in letting all such choices speak to the data. In that regard, it can be related to the generalized method of moments. There, a wide range of unconditional moment restrictions can arise within a very simple model. The practice of choosing which moments to use is usually guided by what the model is intended to capture. It rarely involves only the consideration of estimation efficiency.*

Remark 2 *In the empirical illustrations, for each singular model we will always start the analysis with the following specification. We choose the first subset, $Y_{1,t}$, to correspond to a maximal non-singular submodel. This implies that we subject the model to capturing the joint dynamic properties* *of this vector. Then, we set $Y_{2,t},...,Y_{S,t}$ to be singleton subsets such that their union includes all the remaining variables in Y_t. This implies that we also subject the model to capturing the marginal behaviors of these variables. These two considerations are natural from a modeling perspective, and are also feasible under stochastic singularity. On the technical side, under this specification the composite likelihood function equals one when integrated over the values of the observables. This makes the interpretation of the prior's effects across different models more straightforward. Starting with this specification, we will also experiment with alternative specifications and examine the differences. Note that the results on identification, inference and forecasting apply to general specifications that include the above as a special case.*

3.1 Illustrations

We consider two simple examples. More complex models are considered in Section 8.

Illustrative example 1. This example allows us to provide analytical results in a simple setting. It also shows that loss of identification can occur when excluding variables from the estimation.

Let x_t and c_t be a household's income and consumption respectively. The researcher postulates the following model, in which the income follows an AR(1) process and the consumption is a fixed proportion of the income:

$$x_t = \rho x_{t-1} + e_t, \;\; c_t = \gamma x_t. \tag{9}$$

Suppose this model is misspecified in the sense that the actual relationship is given by the first equation in the preceding display and $c_t = \gamma x_t + v_t$, where v_t is a transitory fluctuation in the consumption. Suppose $|\rho| < 1$ and $e_t \sim i.i.d.N(0, \sigma^2)$. The goal is to estimate ρ, γ and σ.

For this model increasing the number of shocks is a natural option. We emphasize that in more general situations this can be debatable. King and Watson (2012, p. 124) commented on an important feature of the Smets and Wouters' (2007) model in a different context: "As a medium-scale DSGE model, the Smets and Wouters framework model contained predictions for a substantial number of macroeconomic variables (about 40)." There, quite a few variables are related to each other through equalities. Although Smets and Wouters chose seven variables of major interest as observables, investigating the model's implications for other variables can also be valuable. It is hardly convincing to state that one should always increase the number of shocks when additional variables are measured and brought to estimate the model.

Because the model-implied covariance matrix of (x_t, c_t) is singular, the density of (x_t, c_t) does not exist and the conventional full information maximum likelihood is not applicable. A common practice in the DSGE literature is to use only one variable for estimation. Specifically, dropping the variable c_t leaves: $x_t = \rho x_{t-1} + e_t$. This identifies (ρ, σ) but not γ. Dropping the variable x_t leaves: $c_t = \rho c_{t-1} + \gamma e_t$. This identifies $(\rho, \gamma\sigma)$ but does not separately identify γ and σ. Therefore, in both cases, we fail to identify some of the structural parameters.

The model has two nonsingular submodels that correspond to $\{x_t\}$ and $\{c_t\}$. Further,

$$\begin{aligned} \text{For } x_t \quad &: \quad \ell_1(\rho,\gamma,\sigma) = \; -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(x_t - \rho x_{t-1})^2 , \\ \text{For } c_t \quad &: \quad \ell_2(\rho,\gamma,\sigma) = -\frac{T}{2}\log\sigma^2 - \frac{T}{2}\log\gamma^2 - \frac{1}{2\gamma^2\sigma^2}\sum_{t=1}^{T}(c_t - \rho c_{t-1})^2 . \end{aligned}$$

The parameters ρ, σ are identified from $\ell_1(\rho,\gamma,\sigma)$ while γ is further identified from $\ell_2(\rho,\gamma,\sigma)$. Therefore, all the structural parameters are identified from considering the composite likelihood.

The maximizer of the composite likelihood satisfies the following relationship:

$$\begin{aligned} \hat{\rho} &= \left(\sum_{t=1}^{T} x_{t-1}^2 + \frac{1}{\hat{\gamma}^2}\sum_{t=1}^{T} c_{t-1}^2\right)^{-1}\left(\sum_{t=1}^{T} x_t x_{t-1} + \frac{1}{\hat{\gamma}^2}\sum_{t=1}^{T} c_t c_{t-1}\right), \\ \hat{\sigma}^2 &= \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{\rho}x_{t-1})^2, \qquad \hat{\gamma}^2 = \frac{\sum_{t=1}^{T}(c_t - \hat{\rho}c_{t-1})^2}{\sum_{t=1}^{T}(x_t - \hat{\rho}x_{t-1})^2}. \end{aligned}$$

The formula for $\hat{\rho}$ mirrors the OLS estimator obtained by weighting the information from the two equations with their error variances; $\hat{\sigma}^2$ equals the sample residual variance of the first equation; $\hat{\gamma}^2$ equals the ratio of two residual variances.

Interestingly, as the specification error becomes smaller (i.e., the variance of v_t approaches 0), $\hat{\gamma}^2$ will approach its true value γ for any sample size, and $\hat{\rho}$ and $\hat{\sigma}^2$ will reduce to the conventional MLE under a known γ. (Note that $x_t = \rho x_{t-1} + e_t$ and $c_t = \rho c_{t-1} + \gamma e_t$ coincide when the variance of v_t equals 0, therefore using one is equivalent to using both.). Thus, in this simple model, the composite likelihood delivers an intuitive estimator that coincides with the ideal estimator under correct model specification.

Illustrative example 2. This example illustrates how to algorithmically compute the composite likelihood, by considering singular versions of the prototypical DSGE model considered in Clarida, Gali and Gertler (2000) and Lubik and Schorfheide (2004). The same procedure can be applied to medium scale DSGE models such as that of Smets and Wouters (2007).

The original model in Lubik and Schorfheide (2004) is

$$\begin{aligned} y_t &= E_t y_{t+1} - \tau(r_t - E_t \pi_{t+1}) + g_t, \\ \pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - z_t), \\ r_t &= \rho_r r_{t-1} + (1-\rho_r)\psi_1 \pi_t + (1-\rho_r)\psi_2(y_t - z_t) + \varepsilon_{rt}, \\ g_t &= \rho_g g_{t-1} + \varepsilon_{gt}, \\ z_t &= \rho_z z_{t-1} + \varepsilon_{zt}, \end{aligned} \tag{10}$$

where y_t, π_t and r_t denote the log deviations of output, inflation and nominal interest rate from their steady states, g_t is the exogenous spending and z_t captures shifts of the marginal costs of production. The structural shocks are serially uncorrelated and satisfy $\varepsilon_{rt} \sim N(0, \sigma_r^2), \varepsilon_{gt} \sim N(0, \sigma_g^2)$, $\varepsilon_{zt} \sim N(0, \sigma_z^2)$. Among the three shocks, ε_{gt} and ε_{zt} are correlated with correlation coefficient ρ_{gz}. The data contains observations on the levels of output, inflation and interest rate, which are related to the log deviations via

$$Y_t = \begin{pmatrix} 0 \\ \pi^* \\ \pi^* + r^* \end{pmatrix} + \begin{pmatrix} y_t \\ 4\pi_t \\ 4r_t \end{pmatrix}, \tag{11}$$

where the output variable is pre-filtered and therefore has mean zero, π^* and r^* are annualized steady-state inflation and real interest rates in percentages and $\beta = (1 + r^*/100)^{-1/4}$. The vector of structural parameters is

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

We label this model as the three shocks model.

Here, and also in empirical applications, we consider two singular versions of the three shocks model. The first is a one shock model, affected by ε_{zt} only. The second is a two shocks model, affected by ε_{gt} and ε_{zt}. We purposely nest these two models under the nonsingular model to show how the model solution and estimation can be implemented in a unified manner.

The solutions to the three models are related in a simple way. Consider first the three shocks model. The system (10) can be written as (Sims, 2002):

$$\Gamma_0(\theta) X_t = \Gamma_1(\theta) X_{t-1} + \Psi(\theta)\varepsilon_t + \Pi(\theta)\eta_t, \tag{12}$$

where $X_t = (r_t, y_t, \pi_t, g_t, z_t, E_t(\pi_{t+1}), E_t(y_{t+1}))'$, $\eta_t = (\pi_t - E_{t-1}(\pi_t), y_t - E_{t-1}(y_t))'$ and $\varepsilon_t = (\varepsilon_{rt}, \varepsilon_{gt}, \varepsilon_{zt},)'$. The coefficients matrices $\Gamma_0(\theta), \Gamma_1(\theta), \Psi(\theta)$ and $\Pi(\theta)$ are known functions of θ.

Under determinacy, the model's solution can be represented as

$$\begin{aligned} Y_t &= \mu(\theta) + C(\theta) X_t, \\ X_t &= A(\theta) X_{t-1} + B(\theta)\varepsilon_t, \end{aligned} \tag{13}$$

where $A(\theta)$ and $B(\theta)$ are again known functions of structural parameters, $C(\theta)$ is a selection matrix that selects the first three elements of X_t, and $\mu(\theta)$ is given in (11).

The solutions to the two singular models can still be represented as (13) after modifying the shocks accordingly. In the one shock model, ε_t needs to be replaced by $(0, 0, \varepsilon_{zt})'$; in the two shocks model, by $(0, \varepsilon_{gt}, \varepsilon_{zt})'$. Let $i = 1, 2, 3$ be the index for the model with i shocks and define

$$M_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, M_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, M_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then, in a unified manner, the solutions to the three models can be written as: $Y_t = \mu(\theta) + C(\theta)X_t$, $X_t = A(\theta)X_{t-1} + B(\theta)M_i\varepsilon_t$. The vector moving average representation equals $Y_t = \mu(\theta) + C(\theta)\left[I - A(\theta)L\right]^{-1} B(\theta)M_i\varepsilon_t$.

Now consider estimation. Suppose we wish to construct the component likelihood corresponding to the first element of Y_t. Then, define $P_s = [1 \;\; 0 \;\; 0]$ and consider

$$\begin{aligned} P_s Y_t &= P_s\mu(\theta) + P_s C(\theta) X_t, \\ X_t &= A(\theta) X_{t-1} + B(\theta) M_i \varepsilon_t. \end{aligned}$$

Its likelihood can be easily obtained in the time domain by Kalman filtering and in the frequency domain by computing its spectral density and periodograms (c.f. Appendix C). Other component likelihoods can be computed in the same way by simply changing P_s accordingly.

In the empirical illustrations, we will use the three singleton subsets $\{y_t\}, \{\pi_t\}, \{r_t\}$ to form the composite likelihood for the one shock model. For the two shocks model, we will start with the following specification: $\{y_t, r_t\}$ and $\{\pi_t\}$, but will also consider the following alternative: $\{y_t, \pi_t\}$ and $\{r_t\}$. We find the subset $\{r_t, \pi_t\}$ to be nearly singular to the extent that the Kalman filtering algorithm frequently reports degenerate covariance matrices. This pair is therefore not considered when implementing the composite likelihood. It will emerge that the parameter estimates and impulse responses from the latter two alternative specifications are both similar.

4 Identification

Suppose the distribution of Y_t is given by the full system (1) with $\theta = \theta_0$. This section considers the issue of distinguishing θ_0 from alternative parameter values using the information provided by the nonsingular submodels. Note that here identification is considered as a property of the model. No data is involved. Therefore, the issue of misspecification does not play a role. The analysis is from a local identification perspective, which builds on the results in Qu and Tkachekno (2012). We continue to denote the spectral density of $Y_{s,t}$ at θ by $f_s(\omega;\theta)$ and its mean by $\mu_s(\theta)$.

Definition 3 *The parameter vector θ is locally identifiable at θ_0 from the first and second order properties of $Y_{s,t}$ $(s = 1, ..., S)$ if there exists an open neighborhood of θ_0 in which $\mu_s(\theta_1) = \mu_s(\theta_0)$ and $f_s(\omega;\theta_1) = f_s(\omega;\theta_0)$ for all $s = 1, ..., S$ and all $\omega \in [-\pi, \pi]$ implies $\theta_0 = \theta_1$.*

The above definition is formulated in the frequency domain. There is an equivalent formulation in the time domain in terms of autocovariance functions. Suppose $Y_{s,t}$ has autocovariance function $\Gamma_s(j;\theta)$ $(j = 0, \pm 1, ...)$ satisfying $\Gamma_s(j;\theta) = \Gamma_s(-j;\theta)$ and that $f_s(\omega;\theta)$ is continuous in ω. Then, there is a one-to-one mapping between $\Gamma_s(j;\theta)$ $(j = 0, \pm 1, ...)$ and $f_s(\omega;\theta)$ $(\omega \in [-\pi, \pi])$, given by $\Gamma_s(k;\theta) = \int_{-\pi}^{\pi} \exp(ij\omega) f_s(\omega;\theta) d\omega$. Therefore, θ_0 is locally identifiable from $\mu_s(\theta)$ and $f_s(\omega;\theta)$ $(s = 1, ..., S)$ if and only if it is locally identifiable from $\mu_s(\theta)$ and the complete set of autocovariances $\{\Gamma_s(j;\theta)\}_{j=-\infty}^{\infty}$. The next two assumptions impose some regularity conditions on the parameter space and the elements of the spectral density matrix.

Assumption 1. $\theta_0 \in \Theta \subset \mathbb{R}^q$ with θ_0 being an interior point. Assume Θ is compact.

Assumption 2. Assume the following conditions hold for all $\theta \in \Theta$ and $\omega \in [-\pi, \pi]$: (i) $\sum_{j=0}^{\infty} \|h_j(\theta)\| \leq C < \infty$ and $\|\Sigma(\theta)\| \leq C < \infty$, where $h_j(\theta)$ $(j = 0, ..., \infty)$ are defined in $H(L;\theta) = \sum_{j=0}^{\infty} h_j(\theta) L^j$; (ii) The elements of $f(\omega;\theta)$ belong to Lip(β) of $\beta > \frac{1}{2}$ with respect to ω[1]; (ii) The elements of $f_\theta(\omega)$ are continuously differentiable in θ with $\left\| \partial \operatorname{vec}(f_\theta(\omega))/\partial\theta' \right\| \leq C < \infty$.

Theorem 1 *Let Assumptions 1-2 hold. Define*

$$G_S(\theta) = \sum_{s=1}^{S} \left\{ \int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} f_s(\omega;\theta)}{\partial \theta'} \right)^* \left(\frac{\partial \operatorname{vec} f_s(\omega;\theta)}{\partial \theta'} \right) d\omega + \frac{\partial \mu_s(\theta)'}{\partial \theta} \frac{\partial \mu_s(\theta)}{\partial \theta'} \right\} \tag{14}$$

Assume $G_S(\theta)$ has a constant rank in an open neighborhood of θ_0. Then, θ is locally identifiable at θ_0 if and only if $G_S(\theta_0)$ is nonsingular.

[1] Let $g(\omega)$ be a scalar valued function defined on an interval B. We say g belongs to Lip(β) if there exists a finite constant M such that $|g(\omega_1) - g(\omega_2)| \leq M|\omega_1 - \omega_2|^\beta$ for all ω_1, $\omega_2 \in B$.

Theorem 1 follows from Theorem 2 in Qu and Tkachenko (2012). The latter provides a necessary and sufficient condition for local identification based on the mean and spectrum of Y_t, i.e., based on the full model. Here, the identifying information comes from the nonsingular submodels.

The dimension of $G_S(\theta)$ always equals the number of structural parameters. In particular, it is invariant to the number of equations, observables and shocks in the model. This feature is advantageous for analyzing identification in models with a high number of equations or variables. The s-th component in the summation, i.e.,

$$\int_{-\pi}^{\pi}\left(\frac{\partial \operatorname{vec} f_s(\omega;\theta)}{\partial\theta'}\right)^*\left(\frac{\partial \operatorname{vec} f_s(\omega;\theta)}{\partial\theta'}\right)d\omega+\frac{\partial\mu_s(\theta)'}{\partial\theta}\frac{\partial\mu_s(\theta)}{\partial\theta'} \tag{15}$$

measures the contribution from the s-th submodel to identification. It is semidefinite by construction. Therefore, if local identification is achieved by considering a particular $Y_{s,t}$, then it is also achieved by considering all the submodels $Y_{s,t}$ $(s=1,...,S)$. In practice, it is informative to compare the rank of (15) for $s=1,...,S$. This can be informative about the source of the identification.

As in Qu and Tkachenko (2012), Theorem 1 can be extended in several directions. We summarize three such extensions. The proofs are essentially the same as in Qu and Tkachenko (2012), therefore omitted. First, to check local identification based on the second order properties only (i.e., ignoring the steady state restrictions), we simply delete the term in (14) related to $\partial\mu_s(\theta)'/\partial\theta$. Second, to consider identification based on a subset of frequencies, say those corresponding to business cycle fluctuations, we replace $G_S(\theta)$ by

$$\sum_{s=1}^{S}\left\{\int_{-\pi}^{\pi}W(\omega)\left(\frac{\partial \operatorname{vec} f_s(\omega;\theta)}{\partial\theta'}\right)^*\left(\frac{\partial \operatorname{vec} f_s(\omega;\theta)}{\partial\theta'}\right)d\omega+W(0)\frac{\partial\mu_s(\theta)'}{\partial\theta}\frac{\partial\mu_s(\theta)}{\partial\theta'}\right\},$$

where $W(\omega)$ denotes an indicator function defined on $[-\pi,\pi]$ that is symmetric around zero and equals to one over a finite number of closed intervals corresponding to the desired frequencies. Third, to check local identification of a subset of parameters say $\theta^{(1)}$ while fixing the others at θ_0, we divide θ as $\theta'=[\theta^{(1)\prime},\theta^{(2)\prime}]$ and replace $G_S(\theta)$ by

$$\sum_{s=1}^{S}\left\{\int_{-\pi}^{\pi}W(\omega)\left(\frac{\partial \operatorname{vec} f_s(\omega;\theta)}{\partial\theta^{(1)\prime}}\right)^*\left(\frac{\partial \operatorname{vec} f_s(\omega;\theta)}{\partial\theta^{(1)\prime}}\right)d\omega+W(0)\frac{\partial\mu_s(\theta)'}{\partial\theta^{(1)}}\frac{\partial\mu_s(\theta)}{\partial\theta^{(1)\prime}}\right\}.$$

For all the three cases, the statement in Theorem 1 continues to hold after changing $G_S(\theta)$.

5 Estimation and inference

The estimation can be carried out by treating $\ell(\theta)$ as if it was the conventional log likelihood. Specifically, let $\pi(\theta)$ be a prior density. Then, as in Chernozhukov and Hong (2003), we can consider a quasi-posterior distribution with the following density function:

$$p(\theta) = \frac{\pi(\theta)\exp\left(\ell(\theta)\right)}{\int_\Theta \pi(\theta)\exp\left(\ell(\theta)\right)d\theta}. \tag{16}$$

The estimate for θ_0 can be taken to be the quasi-posterior mean: $\hat{\theta} = \int_\Theta \theta p(\theta) d\theta$. Computationally, $\hat{\theta}$ can be obtained using Markov Chain Monte Carlo (MCMC) methods, such as the Metropolis–Hastings algorithm, by drawing a sequence of values $(\theta^{(1)}, \theta^{(2)}, ..., \theta^{(B)})$ corresponding to the density kernel $\pi(\theta)\exp\left(\ell(\theta)\right)$ and computing $\hat{\theta} = B^{-1}\sum_{j=1}^{B}\theta^{(j)}$. We omit the details on the construction of the Markov Chains, since they follow standard procedures. A useful reference is An and Schorfheide (2007). We refer to the intervals obtained by sorting the MCMC draws as "**MCMC Intervals**".

The reason for why $p(\theta)$ is a quasi-posterior is twofold. First, $\ell(\theta)$ is a more general criterion function than the conventional log likelihood. Second, DSGE models are only approximations to the true data generating process. The latter becomes apparent when contrasting the model's singularity with the data's non-singularity. In fact, even nonsingular models can still be considerably misspecified along other dimensions. For example, their low frequency behavior can be quite different from the actual time series data (Schorfheide, 2013). As long as important misspecifications are present, the true likelihood will be typically infeasible to obtain and a prudent interpretation of $p(\theta)$ should be as a quasi-posterior. This statement applies not only to singular models analyzed using the composite likelihood, but also to singular models analyzed using a particular component likelihood say $\ell_s(\theta)$, although the latter is often overlooked in practice.

The complication caused by a quasi posterior is that the MCMC Intervals need not correspond to valid credible intervals. That is, their lengths can differ significantly from the intervals obtained under a correctly specified model and likelihood, even asymptotically. Because the latter intervals have correct asymptotic frequentist coverage, this implies that the MCMC Intervals are invalid confidence intervals when viewed from a frequentist perspective. Chernozhukov and Hong (2003) clearly documented this feature in the context of $\ell(\theta)$ being a general criterion function and provided intervals that have desired asymptotic frequentist properties. Müller (2013) further studied the risk of Bayesian inference under misspecified models. His results imply that the latter intervals can have lower asymptotic frequentist risk than the MCMC intervals.

Building upon the above literature, below we address the inference issue in two steps. In the first step, we construct confidence intervals that have the following two features. (1) It acknowledges model misspecification. (2) It achieves correct frequentist coverage rates for the pseudo true value asymptotically. The analysis builds on Chernozhukov and Hong (2003). We refer to the resulting intervals as "**Asymptotic Intervals**". In the second step, we further contrast the "Asymptotic

Intervals" with the "MCMC intervals", by taking into account both DSGE models' identification properties and data limitations that we face. This will allow us to offer suggestions for practice.

We now state an assumption that specifies the types of misspecifications we allow for. Then, we will provide two results for the probability limit of the estimator and its asymptotic distribution.

Assumption MI. The observed data $\{Y_t\}_{t=1}^T$ are generated by a covariance stationary vector process: $Y_t = \mu_0 + \sum_{j=0}^{\infty} h_{0j}\zeta_{t-j}$, where $\{\zeta_t\}$ are mean zero, serially uncorrelated processes with $\mathrm{Var}(\zeta_t) = \Sigma_0$ and the 3rd and 4th order cumulants being zero. Assume Y_t has spectral density $f_0(\omega)$, satisfying Assumption 2 with $f(\omega;\theta)$, $h_j(\theta)$ and ϵ_t replaced by $f_0(\omega), h_{0j}$ and ζ_t, respectively.

Assumption 2 is about the properties of the model while Assumption MI is about the data. The mean μ_0 and spectral density $f_0(\omega)$ can be different from $\mu(\theta)$ and $f(\omega;\theta)$ for all θ. In particular, the data can be stochastically nonsingular or exhibit low frequency variations different from what are implied by the model. The requirements on the cumulants can be relaxed. Doing so will not affect the consistency result (i.e., Lemma 2) but will alter the asymptotic distribution (i.e., Theorem 2). As will be seen, the procedure for constructing the Asymptotic Intervals (i.e., Procedure A) is valid even without these zero cumulants requirements.

Let $\mu_{s,0}$ and $f_{s,0}(\omega)$ contain the elements of μ_0 and $f_0(\omega)$ that correspond to the s-th nonsingular submodel. Define $\ell_\infty(\theta) = \sum_{s=1}^{S} \ell_{s,\infty}(\theta)$, where

$$\begin{aligned}\ell_{s,\infty}(\theta) \quad = \quad & -\frac{1}{4\pi}\left\{\int_{-\pi}^{\pi}\left[\log\det(f_s(\omega;\theta)) + \mathrm{tr}(f_s^{-1}(\omega;\theta)f_{s,0}(\omega))\right]d\omega\right. \\ & \left. + \left(\mu_{s,0} - \mu_s(\theta)\right)' f_s^{-1}(0;\theta)\left(\mu_{s,0} - \mu_s(\theta)\right)\right\}.\end{aligned}$$

The next result establishes the limits of the composite likelihood and its maximizers.

Lemma 2 *Let Assumptions 1,2 and MI hold. Then:*

1. $T^{-1}\ell(\theta)$ *converges uniformly almost surely to* $\ell_\infty(\theta)$ *over* $\theta \in \Theta$.
2. *Let* $\check{\theta}$ *denote the set of maximizers of* $\ell(\theta)$ *and* θ_0 *the set of maximizers of* $\ell_\infty(\theta)$*, then with probability one we have:* $\limsup_{T\to\infty} \check{\theta} \subseteq \theta_0$.
3. *Further, if* $\ell_\infty(\theta)$ *has a unique maximizer* θ_0*, then* $\check{\theta} \to^{a.s} \theta_0$.

The proof of the lemma follows closely the arguments in Hansen and Sargent (1993, p.49-53). In the Lemma, $\ell(\theta)$ can be either the Gaussian likelihood in the time domain, or its approximation in the frequency domain. The first two results do not assume that the parameters are identified, while all the results hold irrespective of whether the data have Gaussian distributions.

The interpretation of θ_0 depends on the model specification. If it is correctly specified, then the set θ_0 will consist of all the parameter values that satisfy $f_{s,\theta}(\cdot) = f_{s,0}(\cdot)$ and $\mu_{s,0} = \mu_s(\theta)$ for all $s = 1, ..., S$. If it is misspecified, then such values do not exist, and θ_0 should be interpreted as pseudo-true values. Further, if $\ell_\infty(\theta)$ has a unique maximizer, then under correct specifications, because all the components of $\ell_\infty(\theta)$ are maximized at θ_0, using a particular $\ell_s(\theta)$ rather than $\ell(\theta)$ to implement (16) will also yield consistency provided that $\ell_{s,\infty}(\theta)$ also has a unique maximizer. In contrast, under misspecification, the maximizers of $\ell_{s,\infty}(\theta)$ can be different. Consequently, replacing $\ell(\theta)$ in (16) by a particular $\ell_s(\theta)$ can lead to substantially different estimates even asymptotically. The lemma therefore reinforces the discussion in Remark 1, that if misspecification is present, then constructing the composite likelihood $\ell(\theta)$ should be viewed as a specification rather than an efficiency issue. Finally, the feature that different component likelihoods can be associated with different pseudo-true parameter values may in itself serve as the basis for a test for misspecification.

The lemma has abstracted away from the effect of the prior $\pi(\theta)$. The second and third results continue to hold when $\check{\theta}$ is replaced by the mode of (16), provided that the prior is independent of T and that its support includes θ_0 defined in these two results.

The next result provides the asymptotic distribution. It allows misspecification but requires the parameters to be well identified. The practical implication of the latter is further discussed later.

Theorem 2 *Suppose θ_0 is the unique minimizer of $\ell_\infty(\theta)$. Let $\hat{\theta}$ denote the mean or mode computed from (16). Then, under Assumptions 1,2 and MI, we have*

$$\sqrt{T}(\hat{\theta} - \theta_0) \to^d N(0, M^{-1} V M^{-1}),$$

where $M = \frac{1}{4\pi} \sum_{s=1}^{S} (M_{1,s} + M_{2,s})$ *and* $V = \frac{1}{4\pi} \sum_{s=1}^{S} \sum_{h=1}^{S} (V_{1,s,h} + V_{2,s,h})$ *with*

$$\begin{aligned}
M_{1,s} &= \int_{-\pi}^{\pi} \frac{\partial^2}{\partial\theta\partial\theta'} \log\det(f_s(\omega;\theta_0)) + \frac{\partial^2}{\partial\theta\partial\theta'} \operatorname{tr}\left\{ f_s^{-1}(\omega;\theta_0) f_{s,0}(\omega) \right\} d\omega, \\
M_{2,s} &= 2 \frac{\partial \mu_s(\theta_0)'}{\partial\theta} f_s^{-1}(0;\theta_0) \frac{\partial \mu_s(\theta_0)}{\partial\theta'}, \\
V_{1,s,h} &= \int_{-\pi}^{\pi} \left(\frac{\partial \operatorname{vec} f_s(\omega;\theta_0)}{\partial\theta'} \right)^* \left(\overline{f_s^{-1}(\omega;\theta_0)} \otimes f_s^{-1}(\omega;\theta_0) \right) \left(\overline{f_{s,h,0}(\omega)} \otimes f_{s,h,0}(\omega) \right) \\
&\qquad \left(\overline{f_h^{-1}(\omega;\theta_0)} \otimes f_h^{-1}(\omega;\theta_0) \right) \left(\frac{\partial \operatorname{vec} f_h(\omega;\theta_0)}{\partial\theta'} \right) d\omega \ , \\
V_{2,s,h} &= 2 \frac{\partial \mu_s(\theta_0)'}{\partial\theta} f_s^{-1}(0;\theta_0) f_{s,h,0}(\omega) f_h^{-1}(0;\theta_0) \frac{\partial \mu_h(\theta_0)}{\partial\theta'}.
\end{aligned}$$

In the above, $\overline{f_s^{-1}(\omega;\theta_0)}$ *is the conjugate of* $f_s^{-1}(\omega;\theta_0)$ *and* $f_{s,h,0}(\omega)$ *is the cross spectrum of the data vectors* $Y_{s,t}$ *and* $Y_{h,t}$ *at the frequency* ω.

Although the result is formulated in the spectral domain, as shown in the appendix, it applies to both the time and frequency domain composite likelihoods. The matrix M stems from the second order derivative of the log composite likelihood, while V corresponds to the covariance of the score function. The dependence between the submodels shows up through $V_{1,s,h}$ and $V_{2,s,h}$ with $s \neq h$. Without misspecification, M equals V and the asymptotic variance then reduces to M^{-1}. With misspecification, $M^{-1}VM^{-1}$ is usually different from M^{-1}. Therefore, the MCMC draws will need to be adjusted to have variance $M^{-1}VM^{-1}$ in order for their quantiles to deliver asymptotically valid confidence intervals. Such an adjustment is given in the following procedure.

Procedure A for computing the Asymptotic Intervals: First, compute $\sqrt{T}(\theta^{(j)} - \hat{\theta})$ with $\hat{\theta}$ being the quasi posterior mean or mode of (16). Use their sample covariance as an estimator for M^{-1} and denote it by $\hat{M}^{-1}$. Next, compute $\sum_{s=1}^{S} \partial \ell_{s,t}(\hat{\theta})/\partial\theta$ $(t = 1, ..., T)$. Use their (long-run) sample covariance as an estimator for V and denote it by $\hat{V}$. Then, obtain $\hat{M}^{1/2}$ and $\hat{V}^{1/2}$ using the singular value decomposition and compute

$$\widetilde{\theta}^{(j)} = \hat{\theta} + \hat{M}^{-1}\hat{V}^{1/2}\hat{M}^{1/2}(\theta^{(j)} - \hat{\theta}).$$

Finally, use the quantiles of $\widetilde{\theta}^{(j)}$ to form confidence intervals.

Some misspecifications can cause $\sum_{s=1}^{S} \partial \ell_{s,t}(\theta_0)/\partial\theta$ to be serially correlated. This is why a long run covariance estimator is potentially needed for V. In practice, it is informative to regress $\sum_{s=1}^{S} \partial \ell_{s,t}(\hat{\theta})/\partial\theta$ on the lagged values to examine whether such dependencies are present. The standard sample covariance matrix can be used if the latter are considered to be small.

This concludes the first step. Now, we further contrast the Asymptotic Intervals with the MCMC Intervals (see page 15 for how the latter intervals are defined). As discussed above, the Asymptotic Intervals are interpretable under misspecification, but are based on approximating $\sqrt{T}(\hat{\theta} - \theta_0)$ with a multivariate normal distribution. The justification for the latter often involves the model's parameters being well identified and the sample size being not too small. In contrast, the MCMC Intervals are interpretable with any sample sizes without requiring statements about identification, but require the assumption that the model is well specified. While in other contexts one concern, say misspecification, can clearly dominate the other, in the current context the two concerns are both substantively important. This suggests that the two types of intervals can be viewed as complementary. Based on the above consideration, we recommend to report both intervals, and interpret them taking into account their respective strengths and limitations. This will be implemented in the empirical illustrations. We note that the above recommendation can be related to that in Moon and Schorfheide (2012), who analyzed the differences between Bayesian and frequentist confidence sets in models with partially identified parameters. They recommended reporting the estimates of the identified set and the conditional prior along with the Bayesian credible sets. Although the issues studied are different, the motivations for the recommendations can both be stated as providing a full disclosure of the results when the Bayesian and Frequentist perspectives can potentially arrive at substantially different conclusions.

6 Impulse response functions

Impulse responses play a central role for assessing the implications of a DSGE model. Below we first discuss how to compute them and measure the associated uncertainty, and then comment on their interpretation in the presence of stochastic singularity.

The computation follows from the vector moving average representation of the full model. For simplicity, suppose there are no measurement errors. Then we have:

$$Y_t = \mu(\theta) + C(\theta)(I - A(\theta)L)^{-1}B(\theta)\varepsilon_t. \tag{17}$$

The impulse responses at the horizon k are then given by $IR\,(k;\theta) = C(\theta)A(\theta)^k B(\theta)\Sigma^{1/2}\,(\theta)$. Let $\mathbf{e}_j$ be the j-th column of an identity matrix. Then, the response of the j-th variable to the l-th orthogonal shock equals

$$\mathbf{e}_j' IR\,(k;\theta)\,\mathbf{e}_l \tag{18}$$

The inference on (18) can be carried out in three steps using the MCMC draws $\theta^{(i)}$:

- Step 1. Compute $\mathbf{e}_j' IR(k;\hat{\theta})\mathbf{e}_l$, where $\hat{\theta}$ denotes the mean (or the median) of $\theta^{(i)}$.
- Step 2. Compute $\mathbf{e}_j' IR(k;\theta^{(i)})\mathbf{e}_l$.
- Step 3. Sort the resulting values. Use their relevant percentiles to form an interval.

This procedure leads to pointwise MCMC Intervals for the impulse responses. We can construct Asymptotic Intervals simply by replacing $\theta^{(i)}$ with $\widetilde{\theta}^{(i)}$ in Steps 2 and 3. We will report both intervals in the empirical applications.

Consistent with the view elsewhere in this paper, here the impulse responses are interpreted as a summarizing measure for a potentially misspecified model. In other words, the impulse response function is treated as a deterministic function of the model's structural parameters. Such a view is compatible with stochastic singularity and also provides internal consistency for comparing impulse responses across models. In contrast, if we equalized Y_t in (17) with the actual observed time series, then the analysis would make no sense, because under singularity there would exist no ε_t to make (17) hold. The inconsistency associated with the latter view has also been discussed in Ingram, Kocherlakota and Savin (1994), who then suggested that one should consider only nonsingular models. Here, we hold the view that the impulse response analysis can still be meaningfully applied to singular models, but only after explicitly acknowledging the models as approximations.

7 Forecasting

For nonsingular models, one- and muti-step ahead forecasts can be obtained by utilizing the following relationship (see, e.g., An and Schorfheide, 2007)

$$p(Y_{T+1}|Y_{1:T}) = \int p(Y_{T+1}|Y_{1:T};\theta)p(\theta|Y_{1:T})d\theta, \tag{19}$$

where $Y_{1:T}$ denotes the observed sample, $p(\theta|Y_{1:T})$ denotes the posterior distribution of θ given $Y_{1:T}$, and $p(\cdot|Y_{1:T};\theta)$ is the conditional density of Y_{T+1} given $Y_{1:T}$ and θ which can be evaluated using the Kalman filter. Consequently, the left hand side distribution can be generated by first sampling from the posterior distribution of θ and then drawing from the multivariate normal distribution implied by $p(Y_{T+1}|Y_{1:T};\theta)$. However, the same algorithm is no longer applicable in the presence of singularity. In practice, the forecasting typically proceeds either by introducing measurement errors or by treating some observables as unobserved. The latter approach ignores the information from some observed time series and also only yields forecasts for a subset of the observables.

The composite likelihood framework offers an opportunity for obtaining forecasts for all the observed endogenous variables. This can be done by exploiting a relationship that is analogous to (19). Specifically, we consider

$$p_S(Y_{T+1}|Y_{1:T}) = \int p_S(Y_{T+1}|Y_{1:T};\theta)p_S(\theta|Y_{1:T})d\theta, \tag{20}$$

where $p_S(\theta|Y_{1:T})$ equals $p(\theta)$ in (16) and

$$p_S(Y_{T+1}|Y_{1:T};\theta) \propto \prod_{s=1}^{S} p_s(Y_{s,T+1}|Y_{s,1:T};\theta). \tag{21}$$

On the right hand side of (21), $p_s(\cdot|Y_{s,1:T};\theta)$ denotes the conditional density of $Y_{s,T+1}$ corresponding to the s-th submodel, which can be evaluated using the standard Kalman filtering algorithm. Using (20) in place of (19) leads to the following forecasting procedure (Let $\theta^{(i)}$ $(i = 1, ..., B)$ denote the MCMC draws from (16)):

- Step 1. Sample from $p_S(\cdot|Y_{1:T};\theta^{(i)})$ defined in (21) for $i = 1, ..., B$. Denote the values by $Y^{(i)}_{T+1}$.
- Step 2. Compute $\hat{Y}_{T+1} = B^{-1}\sum_{i=1}^{B} Y^{(i)}_{T+1}$ and use $\hat{Y}_{T+1}$ as the point forecast for Y_{T+1}.
- Step 3. If muti-step forecasts are needed, then let $Y_{1:(T+1)} = [Y'_T, \hat{Y}'_{T+1}]'$ and repeat Steps 1 and 2 with T replaced by $T + 1$. Continue this step until the desired horizon is reached.

In the procedure, Step 1 needs to be tailored depending on whether the subsets $Y_{s,t}$ are disjoint (i.e., whether $Y_{s_1,t} \cap Y_{s_2,t}$ are empty for all $s_1 \neq s_2$). If they are all disjoint, then sampling from $p_S(\cdot|Y_{1:T};\theta^{(i)})$ is equivalent to sampling separately from $p_s(Y_{s,T+1}|Y_{s,1:T};\theta^{(i)})$ for $s = 1, ..., S$. Consequently, Step 1 consists only of sampling from multivariate normal distributions. If some subsets are overlapping (e.g., $Y_{s_1,t} \cap Y_{s_2,t}$ is not empty for some $s_1 \neq s_2$), then the sampling needs to account for such dependencies and a Metropolis step is in general needed. This can be achieved by first obtaining a draw, say Y^*_{T+1}, from a proposal distribution and then deciding whether or not to keep it by evaluating $p_S(Y^*_{T+1}|Y_{s,1:T};\theta)$. Here, a natural proposal distribution is the multivariate normal distribution, with the conditional mean and variance set to those implied by an estimated finite order vector autoregression for Y_t $(t = 1, ..., T)$. Importantly, irrespective of whether the subsets are disjoint, $p_S(\theta|Y_{1:T})$ is always obtained using all the S submodels in the composite likelihood.

A key feature of the above procedure is that it allows us to obtain forecasts for all the observed endogenous variables irrespective of the number of the structural shocks. This provides an opportunity for comparing forecasting performance between different singular models, as well as between singular and nonsingular models. Such information can then be used to shed light on what aspects of the model need to be further improved.

8 Empirical illustrations

This section applies the composite likelihood framework to analyze both small and medium scale DSGE models. The models considered are singular versions of two influential models in the literature. The first is a prototypical three-equation model, studied in Clarida, Gali and Gertler (2000) and Lubik and Schorfheide (2004). It is designed to model the behavior of the output, inflation and nominal interest rate and is often considered as a starting point for building more elaborate models. The second is the model of Smets and Wouters (2007). This model features a rich array of shocks and frictions and has become the workhorse model in both academia and central banks.

We take the perspective of a modeler, who starts with simple models with few shocks and then gradually incorporates more shocks to examine the similarities and differences. For each singular model, we first study the parameter estimates and then the impulse responses. Because the identification analysis can be conducted in the same way as in Qu and Tkachenko (2012, Section 3), such studies are omitted here. The analysis, besides illustrating the composite likelihood framework, also sheds lights on the following two issues. (1) What behavioral and policy parameters are sensitive to singularity? (2) How does the effect of a particular shock depend on the inclusion of other shocks? As will be seen, after conditioning on the structure of the model, some structural parameters and shock processes can have substantively different estimates depending on what other shocks are allowed in the system. The experimentation will also reveal important model features that remain essentially invariant across the different specifications.

8.1 Small scale singular models

We relate the analysis to that of Lubik and Schorfheide (2004). The two singular models are specified as follows. The first model is affected only by the productivity shock ε_{zt}, while the second is affected also by the exogenous spending shock ε_{gt}. The sample consists of quarterly observations over 1982:IV-1997:IV. This corresponds to the determinant monetary policy regime considered in Lubik and Schorfheide (2004). We do not study the indeterminate regime, although the framework permits such an analysis. The same priors are used for both models. They are given in Table 1 in Lubik and Schorfheide (2004), and are also reported here in the first four columns in Table 1. The table (the last four columns) also reports the results for the original three shocks model. This provides a useful point of reference for interpreting the estimates of the singular models.

8.1.1 The one shock model

The following subsets are used to form the composite likelihood: $\{y_t\}, \{\pi_t,\}$ and $\{r_t\}$. Consequently, we subject the model to the marginal behavior of these three processes.

Parameter estimates. The results (i.e., the mean, mode, 90% MCMC and Asymptotic Intervals) are reported in Panel (a) in Table 1. The estimates show two notable differences when compared with the nonsingular model. First, the inflation weight parameter (ψ_1) is small. The mean and mode are at 1.36 and 1.23 respectively, compared with 2.20 and 2.15 in the nonsingular model. Because the remaining policy parameters are similar, the former estimate implies a more dovish attitude towards inflation than the latter. Second, the standard deviation of the productivity shock (σ_z) is high. The mean and mode are at 0.90 and 0.89, relative to 0.64 and 0.61 in the nonsingular model. The other parameter values are relatively close across the two models.

The two parameters ψ_1 and σ_z are important for the economy's response to a one standard deviation shock in productivity. A lower ψ_1 leads to a deeper decline in the aggregate price level, which can potentially dampen the increase in the output. At the same time, a higher σ_z leads to more pronounced responses in all the three variables. We now report impulse response functions to further quantify the above effects.

Impulse responses. Panel (A) in Figure 1 shows the responses to a one standard deviation shock in productivity for horizons up to 20 quarters. In each subfigure, the solid line corresponds to the responses computed using the posterior mean. The two dashed lines correspond to 90% Asymptotic Intervals, while the shaded area corresponds to 90% MCMC Intervals. We also include in Panel (C) the responses in the nonsingular model for the purpose of comparison.

The three responses are all stronger than in the nonsingular model. This shows that the effect of a high σ_z dominates that of a low ψ_1. Between the two intervals, the Asymptotic Intervals are consistently wider. There, the large values in the responses are typically associated with the simultaneous occurrence of a low ψ_1 and a high κ and σ_z. Such instances are more frequent in the output of Procedure A than among the MCMC draws.

In summary, the results show that leaving out the exogenous spending and monetary policy shocks can affect significantly the assessment of the productivity shock on the three variables. It also pinpoints that the standard deviation of the productivity shock is the main source of such a difference. Next, we incorporate the exogenous spending shock into the analysis.

8.1.2 The two shocks model

The following subsets are used to form the composite likelihood: $\{y_t, r_t\}$ and $\{\pi_t\}$. Consequently, we subject the model to the joint dynamics of the output and the interest rate process and the marginal behavior of the inflation rate process.

Parameter estimates. The values are reported in Panel (b) in Table 1. Interestingly, the value of σ_z and the associated intervals are now very close to the nonsingular model. The value of ψ_1 is also higher and close to the latter. The point estimates for ρ_{gz} are noticeably lower , however the difference can be interpreted as moderate because of the high uncertainty associated with this parameter. The remaining parameter values are all similar to those in the nonsingular model.

Impulse responses. The results are reported in Panel (B) in Figure 1. There, the point estimates and the associated intervals are all similar to the nonsingular model. Therefore, leaving out the monetary policy shock alone has had little effects on the assessment of the technology shock on the three variables. Between the two intervals, the Asymptotic Intervals are consistently wider for the output response case. There, the persistent strong responses are typically associated with high values of ρ_z, which are more frequent within Procedure A than among the MCMC draws.

Alternative specifications. We consider two alternative specifications of the composite likelihood. First, we construct it using the following subsets: $\{y_t, \pi_t\}$ and $\{r_t\}$. In this specification, the mode is located at (in the same ordering as in Table 1): $\{1.71, 0.21, 0.83, 3.48, 2.77, 0.34, 1.71, 0.80, 0.89, -0.16, 0.18, 0.66\}$. The MCMC Intervals are: $[1.28, 2.19], [0.24, 0.29], [0.76, 0.89]$, $[3.02, 3.98]$, $[1.87, 3.72]$, [0.21,0.57],[1.49,2.20], $[0.77, 0.85]$, $[0.70, 0.99]$, $[-0.40, 0.44]$, $[0.15, 0.23]$, $[0.58, 0.86]$. Next, we construct it using the following subsets: $\{y_t, \pi_t\}$ and $\{y_t, r_t\}$. Now the mode is at $\{1.85, 0.21, 0.82, 3.43, 2.72, 0.46, 1.69, 0.84, 0.91, 0.02$,0.16,0.57$\}$. The MCMC Intervals are: $[1.26, 2.51]$, $[0.12, 0.50]$, $[0.75, 0.87]$, $[2.83, 4.09]$, $[1.52, 3.70]$, $[0.29, 0.72]$, $[1.13, 2.48]$, $[0.79, 0.89]$, $[0.83, 0.95]$, $[-0.24, 0.42]$, $[0.13, 0.20]$, $[0.52, 0.70]$. The above values are close to that in Panel (b) in Table 1, except that in the first alternative specification the interval for ψ_2 is much narrower. The posterior means and the Asymptotic Intervals are also similar to the respective values reported in Table 1. In addition, the impulse responses to the productivity shock, as well as and their Asymptotic and MCMC intervals, are all close to those in Panel (B) in Figure 1. The details are omitted here.

Therefore, depending on what other shocks are included in the model, the parameter estimates and impulse responses to the productivity shock can be similar or substantially different. We now further study such issues in the context of medium scale models.

8.2 Medium scale singular models

The Smets and Wouters (2007, henceforth SW) model consists of seven observed endogenous variables: output (y_t), consumption (c_t), investment (i_t), wage (w_t), hours (l_t), inflation (π_t) and nominal interest rate (r_t). It features seven shocks: total factor productivity (η_t^a), exogenous spending (η_t^g), monetary policy (η_t^r), investment specific technology (η_t^i), price markup (η_t^p), wage markup (η_t^w) and risk premium (η_t^b). The model is therefore exactly nonsingular. To facilitate the discussion, we include the log linearized equations of the original model in Appendix B and an annotated list of the parameters in Table 2. We also include in the table the posterior means, modes and 90% intervals of the model's parameters as reported in Tables 1A and 1B in SW.

Below, we consider three singular versions of this model. The first model consists of four shocks: $\eta_t^a, \eta_t^g, \eta_t^r$ and η_t^i. This is a natural starting point, because the first three shocks are common building blocks for even small scale models and the fourth is important for linking the dynamics of the final goods sector with that of the labor market. The second model includes η_t^p as an additional shock. The third model incorporates also η_t^w. These shocks, when included, follow the same specifications as in the original model. The same prior distributions and parameter bounds are used throughout the analysis. The only exception is the price indexation parameter (ι_p), whose lower bound is further reduced from 0.5 to 0.1, such that the latter is not binding when computing the posterior modes. As in the original study, the following parameters are kept fixed $\delta, \phi_w, g_y, \epsilon_p, \epsilon_w$. The sample consists of quarterly observations from 1965.I to 2004.IV, the same as in the original analysis.

8.2.1 The four shocks model

The following nonsingular subsets are used to form the composite likelihood: $\{y_t, \pi_t, r_t, i_t\}, \{c_t\}, \{w_t\}$ and $\{l_t\}$. Using $\{y_t, \pi_t, r_t, i_t\}$ as the maximum nonsingular subset follows from two considerations. First, capturing the joint behavior of output, inflation and nominal interest rate is a key requirement for even small scale models. The medium scale model considered here has a more flexible structure, therefore is naturally positioned for such a task when endowed with three basic shocks η_t^a, η_t^g and η_t^r. Second, the allowance for the investment shock (η_t^i) permits incorporating the investment series into the subset. Meanwhile, the three subsets, $\{c_t\}, \{w_t\}$ and $\{l_t\}$, ensure that the parameter estimates will also be disciplined by the marginal behaviors of these three processes.

The results are reported in Table 3 and Figures 2-29. Below, we first discuss the parameter estimates and then the impulse responses. These results will be contrasted with those reported in SW. This will help to disentangle model features that have different sensitivities to the singularity.

Parameter estimates. As a preview, out of the 28 parameters, 21 of them have their confidence intervals (i.e., the unions of the MCMC Interval and the Asymptotic Interval) overlap with those in SW reported in Table 2. Among the remaining seven parameters ($\xi_p, \rho_a, \lambda, \alpha, r_\pi, \sigma_r, \rho_g$), some parameters (i.e., ξ_p, ρ_a, λ) take on quite different values relative to SW. To discuss the above findings in more detail, we start with the steady state parameters, and then turn to the exogenous shocks parameters and the behavior parameters, and finally the parameters in the monetary policy reaction function. All the values we refer to are the posterior means unless stated otherwise.

The model's steady state values are similar to those in SW. The trend growth rate ($\overline{\gamma}$) equals 0.38, smaller than but close to the original estimate of 0.43. The steady-state inflation rate ($4\bar{\pi}$) equals 2.9 percent on an annual basis, close to the original estimate of 3.1 percent. The annual discount rate ($400(\beta^{-1} - 1)$) is about 0.48 percent, close to the original estimate of 0.65 percent.

The implied mean steady-state nominal and real interest rates are, respectively, about 5.9% and 2.9% on an annual basis, close to 6.0% and 3.0% in the original model. Finally, the steady state of the hours worked ($\bar{l}$) equals -0.65. This value is not far from the original estimate of 0.53 given the series' high variability. The outcome of the above comparison reflects one benefit of keeping all variables in the estimation: if some observables (such as l_t) are not used in the analysis, then some steady state parameters (such as $\bar{l}$) can become unidentified.

Among the exogenous shock processes, as in SW, the productivity and exogenous spending processes are estimated to be persistent while the investment and the monetary policy shock processes are not. Further, for the productivity shock process, the AR(1) and the standard deviation parameter (ρ_a and σ_a) are estimated to be 0.99 and 0.55, higher than the original estimates of 0.95 and 0.45. Importantly, when $\rho_a = 0.99$, the half life of a shock equals 68 quarters, much higher than the 14 quarters implied by $\rho_a = 0.95$. The exogenous spending process has an AR(1) coefficient (ρ_g) of 0.90 and a standard deviation parameter (σ_g) of 0.54, compared with 0.97 and 0.53 in SW. Finally, the AR(1) coefficients for the investment (ρ_i) and monetary policy (ρ_r) shock processes are very close to those in SW (i.e., 0.75 and 0.15 compared with 0.71 and 0.15), while the standard deviation parameters (σ_i and σ_r) are both estimated to be mildly higher (i.e., 0.50 and 0.32 versus 0.45 and 0.24). In summary, here the most pronounced difference pertains to the productivity process. The implication is further studied below through impulse responses.

Now consider the behavioral parameters. The habit persistence parameter (λ) is estimated to be substantially smaller, being 0.37 compared with 0.71 in the original model. The price indexation (ι_p) and rigidity (ξ_p) parameters both take on small values. While the difference in ι_p from SW is mild, the difference in ξ_p is substantial. The latter equals 0.22 (i.e., an average price contract of 1.3 quarters) compared with 0.66 (i.e., an average price contract of 2.9 quarters) in SW. Meanwhile, the wage indexation (ι_w) and rigidity (ξ_w) parameters are both estimated to be high. The former equals 0.85, compared with 0.58 in SW. The latter implies an average wage contract of 6.3 quarters, compared with 3.3 quarters in SW. The remaining parameter values ($\alpha, \psi, \varphi, \sigma_c, \phi_p, \sigma_l$) are close to their respective prior means, and are also broadly similar to the estimates in SW. In summary, among the behavioral parameters, these governing habits and price and wage frictions are consistently different that in the nonsingular model. As further demonstrated below, this translates into markedly different responses to productivity shocks.

Next, consider the parameters from the monetary reaction function. The inflation weight parameter (r_π) equals 1.41, lower than 2.04 in SW. Meanwhile, the output weight parameter (r_y) equals 0.17, higher than 0.08 in SW. Because the differences are in opposite directions, the overall effect is unclear and will be further studied through impulse responses. The other parameters are broadly similar: the policy reacts fairly strongly to changes in output gap while there is a considerable degree of interest rate smoothing.

In summary, the estimated four shocks model features a highly persistent productivity shock process, low price rigidity, high wage rigidity and indexation, and low habit persistence. We now turn to impulse responses to further quantity the effects of such differences.

Impulse responses. Figures 2 to 29 depict the responses of the seven observables to the four shocks for horizons up to 20 quarters. Each figure contains 4 subfigures. The first three corresponds to the singular models with 4 to 6 shocks, while the fourth corresponds to the nonsingular model. In each subfigure, the solid line corresponds to the impulse response function computed using the posterior mean. The two dashed lines provide the 90% Asymptotic Intervals, while the shaded area corresponds to the MCMC Intervals. To facilitate the comparison across different shocks, the y-axis are specified such that the same outcome variable has the same axis limits across the subfigures.

The first subfigures in Figures 2 to 29 confirm that the productivity shock plays a prominent role in driving business cycle fluctuations. More specifically, under a positive productivity shock, the inflation falls sharply (due to the small price inertia; see Figure 6(a)), causing the real wage to rise sharply (due to the high wage indexation; see Figure 22(a)) and the real interest to fall (because of the monetary policy reaction; see Figure 10(a)). On the real side of the economy, the labor supply increases strongly piqued by the higher wage (Figure 18(a)). This leads to a sharp rise in the output (Figure 2(a)), accompanied by a strong increase in the consumption (due to the low habit persistence and lower real interest rate; see Figure 26(a)). Because the productivity shock process is very persistent, its effects on these variables are long lasting. The above responses are substantially more pronounced than those in SW.

In contrast to the productivity shock, the responses to the monetary policy shock are close to that in SW. Specifically, the response of inflation is slightly stronger than in SW (Figure 7(a)). The responses of interest rate and investment are almost identical to the latter (Figures 11(a) and 15(a)). There is initially a slight increase in the real wage growth (Figure 23(a)), as opposed to the small decrease seen in SW. This is due to the drop in the price level and the high wage indexation. This increase vanishes after 4 quarters. The initial responses in the output, hours worked and consumption are all slightly stronger than in SW (Figures 3(a), 19(a) and 27(a)). Then, they revert to levels similar to the latter. These initial differences are due to the low habit persistence.

Now consider the investment shock. The responses in inflation and interest rate are mildly stronger than in SW due to low price rigidity (Figures 8(a) and 12(a)). The response in real wage growth shows a initial slight dip before reverting to positive levels comparable to those in SW (Figure 24(a)). This initial decrease follows from the decrease in inflation and the strong wage indexation. The responses in output, investment, hours worked and consumption are all comparable to that in SW (Figures 4(a), 16(a), 20(a) and 28(a)), except that the responses in the latter two are mildly stronger due to the stronger response in interest rate and the low habit persistence.

Finally, consider the exogenous spending shock. The responses of inflation and interest rate are mildly stronger due to the low price rigidity (Figure 9(a) and 13(a)). The wage growth shows a slight decline at short horizons due to the strong wage indexation (Figure 25(a)). The responses of output, investment, hours worked and consumption are close to that in SW (Figures 5(a), 17(a), 21(a) and 29(a)), except that they exhibit a faster reversion to zero due to the smaller AR(1) coefficient of this shock process.

Therefore, while the responses to the monetary shock are similar to that in SW, the responses to the productivity shock are substantially different. At the core of this difference is the price rigidity parameter. It is estimated to a small value in order to account for the highly volatile inflation seen in the data. This results in unusually strong responses of inflation to productivity shocks, which further lead to very strong responses of labor hours and consequently the output.

For some further comparison, we have also estimated the model using only the variables y_t, π_t, r_t and i_t. The mode is located at (the parameters are in the same order as in Table 3): $\{0.24, 0.40, 5.75, 1.28, 0.83, 1.26, 0.48, 0.83, 0.10, 0.43, 2.27, 1.00, 0.19,$ $0.20,$ $0.80,$ $0.99,$ $0.92,$ $0.64,$ $0.19,$ $0.43, 0.52, 0.50, 0.56, 0.26, 0.37, 0.17, 0.62, 0.00\}$. We note the following. First, the inflation weight parameter (r_π) is substantively smaller than in the second column of Table 3. It is estimated at 1.00, which equals the lower bound for this parameter. In fact, when this bound is further relaxed, its value further decreases to hit the boundary between determinacy and indeterminacy. In addition, an examination of the MCMC draws shows that the resulting values are distributed tightly near 1.00. These values, along with the values of the other policy parameters, imply a monetary policy that is more tolerant of inflation and an economy that can potentially exhibit multiple equilibria. Second, the price rigidity parameter implies an average price contract of 1.8 quarters. Although it is higher than the 1.3 quarters implied by the second column of Table 3, it is still substantially below the 2.9 quarters reported in SW. Third, the parameter estimates related to the productivity process are similar to that in the second column of Table 3. Fourth, when evaluated at the posterior mode, the model implies that the variables c_t, w_t and l_t have the following variances: 0.09, 0.56 and 0.09, while the sample variances computed directly from the data equal 0.49, 0.38

and 0.31, respectively. The substantial differences in c_t and l_t reflect the effects of excluding these observables from the estimation. In contrast, when evaluated at the values in the second column of Table 3, the respective values now equal 0.61, 0.51 and 0.20, respectively. In addition, the variances of y_t, π_t, r_t, i_t in the three cases are: 5.06, 0.75, 3.84, 0.67; 5.08, 0.76, 8.45, 0.68; 5.66, 1.15, 4.15, 0.57. In summary, this exercise shows that treating some observables as unobserved can lead not only to a qualitatively very different model, but also a notably different fit to the data.

8.2.2 The five shocks model

We incorporate the price markup (η_t^p) shock as additional exogenous shock process. This shock plays two roles. First, it constitutes an additional source for business cycle fluctuations. Second, it breaks the rigid link between inflation and price rigidity, permitting large and frequent changes in the former to be compatible with a high level in the latter (c.f. the Phillips curve (B.3)). Consequently, the model allows for a more flexible scope for modeling the responses of inflation, and consequently hours worked and output, to productivity shocks.

The following nonsingular subsets are used to form the composite likelihood: $\{y_t, \pi_t, r_t, i_t, l_t\}$, $\{c_t\}$ and $\{w_t\}$. The incorporation of l_t into the nonsingular subset exploits the above increased flexibility, while the sets $\{c_t\}$ and $\{w_t\}$ continue to use their marginal behaviors to discipline the estimates.

Parameter estimates. The results are summarized in Table 3 (Columns 6 to 9). The values can be contrasted with both the four shocks and the nonsingular model. First, the price rigidity (ξ_p) is now higher and close to that in SW. Second, the persistence of productivity shock process is lower and closer to that in SW, with the half life of a shock reduced to 34 quarters. Third, the inflation weight (r_π) and the output weight parameter (r_y) are both close to SW. In fact, out of the 31 parameters, all except 2 of them (λ and α) now have confidence intervals overlap with those in SW. These three parameters remain close to their respective values in the four shocks model.

Impulse responses. The results are reported in the second subfigures in Figures 2-29. Under a positive shock in productivity, the decrease in inflation is now much smaller than the four shocks model (Figure 6(b)). The magnitude is now close to SW. The increase in real wage is also smaller than the four shocks model (Figure 22(b)). It is still mildly more pronounced that in SW because the productivity process remains more persistent. Note that because the inflation response is much reduced, the high level of wage indexation is no longer quantitatively important for determining the real wage. Consequently, the response in hours worked is much less pronounced than in the four shocks model and now similar to that in SW (Figure 18(b)). Overall, the responses of the seven variables to productivity shocks are no longer substantially different from those in SW.

Responses to the monetary policy shock continue to be close to SW. Specifically, the inflation response is now almost identical to that in SW due to the increased price rigidity (Figure 7(b)). As a result, the initial small increase in the real wage seen in the four shocks model is no longer present (Figure 23(b)). The responses of hours worked, output and consumption are close to, but slightly stronger than, those in SW due to the low habit persistence (Figures 19(b), 3(b), 27(b)).

Responses to the investment and exogenous spending shocks are now close to those in SW. As in the monetary policy shock case, the inflation responses are now almost identical to that in SW due to the increased price rigidity (Figures 8(b), 9(b)). The initial small increases in the real wage are no longer present (Figures 24(b), 25(b)). The initial responses in hours worked, output and consumption are only slightly stronger that those in SW (Figures 20(b), 21(b), 4(b), 5(b), 28(b), 29(b)). The difference follows from the low habit persistence.

In summary, the addition of the price markup shock leads to a significant increase in the price rigidity and decrease in the persistence of the productivity process. These two factors both lead to milder responses to productivity shocks, while also bring the responses to the other shocks closer to those in SW. These responses are no longer substantially different from those in the latter.

8.2.3 The six shocks model

We incorporate the wage markup shock (η_t^w) as an additional shock process. The following subsets are used to form the composite likelihood: $\{y_t, \pi_t, r_t, i_t, l_t, w_t\}$ and $\{c_t\}$.

Parameter estimates. The estimation results are summarized in the last four columns in Table 3. The values move further toward those in SW. The wage indexation parameter (ι_w) is now at a value close to that in SW. Out of the 34 parameters, all except 2 (λ and α) have confidence intervals overlap with those in SW.

Impulse responses. The results are reported in the third subfigures in Figures 2-29. The responses are overall similar to those in the five shocks model. Therefore, adding the new shock has not affected the assessment of the four basic shocks. These responses are also close to that in SW, except that the initial responses in hours worked, output and consumption are slightly different due to the low habit persistence.

In this section, we have considered both small and medium scale models with the number of shocks ranging between 1 and 7. We have confronted all these models with the data. Some findings are as follows. (1) Among the structural parameters, those governing the steady state tend to remain stable across specifications, while those related to the productivity process and the frictions can vary substantially. (2) The estimated effects of a particular shock can crucially depend on what other shocks are permitted in the model. (3) Meanwhile, for the small scale models considered, whether or not to include the monetary policy shock has little effects on the estimated responses to the productivity shock, while for the medium scale models, whether or not to include the wage markup and risk premium shocks has little effects on the estimated responses to the productivity, monetary policy, investment and exogenous spending shocks. (4) There can exist different parameter values that yield similar impulse responses for some shocks while very different responses for others. This reflects an identification issues, implying that relying on matching impulse responses to a particular shock can be insufficient for determining all the parameters. (5) Overall, the composite likelihood framework is informative not only for detecting the above similarities and differences, but also for pinpointing the sources (i.e., the parameters and their values) that generate them.

9 Conclusion

This paper has developed a unified econometric framework for analyzing both singular and nonsingular DSGE models. The value of this framework is not in providing a unique criterion function that achieves the highest efficiency, but in providing a platform that allows for flexible choices of criterion functions, and in letting all such choices speak to the data. The framework naturally allows for carrying out analyses related to parameter identification, estimation, inference and forecasting. Applications to both small and medium scales model show that it can be informative about revealing the similarities and differences between different models and also for pinpointing the sources that generate them.

The framework can be further developed along several dimensions. First, although the paper has focused on linearized models, extensions to nonlinear models can be possible. This is because for the latter one can still define nonsingular submodels, and use them to form the composite likelihood. Second, in some situations, one may have different measurements for the same variable (say GDP or hours worked) and wish to use all of them to discipline the estimates. The current framework can be applicable. That is, one can construct component likelihoods using separately these measurements and then form an overall criterion function. Finally, on the application side, the paper has not considered forecasting comparisons due to the space constraint. This merits a further study as the results can shed further lights on the usefulness of singular models.

References

Altug, S., (1989): "Time-to-build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, 30, 889–920.

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

Andreasen, M.M., J. Fernández-Villaverde, and J. Rubio-Ramírez (2013): "The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications," *NBER Working Paper 18983.*

Besag, J. (1974): "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society (Series B)*, 36, 192-236.

—— (1975): "Statistical Analysis of Non-Lattice Data," *The Statistician,* 24, 179-195.

Bierens, H. J. (2007): "Econometric analysis of linearized Singular Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, 136, 595–627.

Brockwell, P., and R. Davis (1991): *Time Series: Theory and Methods.* New York: Springer-Verlag, 2nd Edition.

Canova, F., F. Ferroni., and C. Matthes (2014): "Choosing the Variables to Estimate Singular DSGE Models," *Journal of Applied Econometrics,* 29, 1099-1117.

Chernozhukov, V., and H. Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.

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

Cox, D.R. (1975): "Partial Likelihood, " *Biometrika,* 62, 269-276.

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

—— (2013): "DSGE Model-Based Forecasting," *Handbook of Economic Forecasting*, 2, Part A, 57 - 140.

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

Duffie, D., and K.J. Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica,* 61, 929-952.

Engle, R.F., N. Shephard, and K. Sheppard (2007): "Fitting Vast Dimensional Time-varying Covariance Models, " *Working paper,* Oxford-Man Institute, University of Oxford.

Fernández-Villaverde J, and J. Rubio-Ramírez (2007): "How Structural are Structural Parameters Values," *NBER Macroeconomics Annual,* MIT Press: Cambridge, MA.

Gallant, A.R., and G. Tauchen (1996): "Which Moments to Match?", *Econometric Theory*, 12, 657–681.

Gouriéroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics* 8, S85-S118.

Guerron-Quintana, P.A. (2010): "What You Match does Matter: the Effects of Data on DSGE Estimation," *Journal of Applied Econometrics*, 25, 774-804.

Hall, G.J. (1996): "Overtime, Effort, and the Propagation of Business Cycle Shocks," *Journal of Monetary Economics*, 38, 139–160.

Hannan, E.J. (1973): "The Asymptotic Theory of Linear Time-Series Models, " *Journal of Applied Probability, 10, 130-145.*

Hansen, L.P., (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica,* 50, 1929-1954.

Hansen, L.P., and T.J. Sargent (1993): "Seasonality and Approximation Errors in Rational Expectations Models," *Journal of Econometrics*, 55, 21–55.

Hjort, N., and H. Omre (1994): "Topics in Spacial Statistics (with Discussion, Comments and Rejoinder)," *Scandinavian Journal of Statistics*, 21, 289-357.

Ireland, P.N. (2004): "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control*, 28, 1205-1226.

Ingram, B.F., N. Kocherlakota, and N.E. Savin (1994): "Explaining Business Cycles: A Multiple-Shock Approach," *Journal of Monetary Economics* 34, 415-428.

King, R.G. and M.W. Watson (2012): "Inflation and Unit Labor Cost," *Journal of Money Credit and Banking* 44, 111–149.

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

Lee, B.S., and B.F. Ingram (1991): "Simulation Estimation of Time-Series Models," *Journal of Econometrics* 47, 195-205.

Lindsay, B.G. (1988): "Composite Likelihood Methods," *Contemporary Mathematics,* 80, 221-239.

Lubik, T., and F. Schorfheide (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94, 190-217.

McGrattan, E.R. (1994): "The Macroeconomic Effects of Distortionary Taxation," *Journal of Monetary Economics*, 33, 573–601.

McGrattan, E.R., R. Rogerson, and R. Wright (1997): "An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy," *International Economic Review,* 38, 267–290.

Moon, H.R. and F. Schorfheide (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80, 755-782.

Müller, U. (2013): "Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix," *Econometrica*, 81, 1805-1849.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591.

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

Ruge-Murcia, F.J. (2007): "Methods to Estimate Dynamic Stochastic General Equilibrium Models," *Journal of Economic Dynamics and Control,* 31, 2599-2636.

Sargent, T.J. (1989): "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy,* 97, 251–287.

Schorfheide, F. (2000): "Loss Function-based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15, 645–670.

—— (2013): "Estimation and Evaluation of DSGE Models: Progress and Challenges," *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, 184–230, Cambridge University Press.

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

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

Smith, A.A. (1993): "Estimating Nonlinear Time-Series Models using Simulated Vector Autoregressions," *Journal of Applied Econometrics* 8, 63-84.

Varin, C., N. Reid., and D. Firth (2011): "An Overview of Composite Likelihood Methods," *Statistica Sinica,* 21, 5-42.

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

Appendix A. Proofs

Proof of Lemma 1. Consider the matrix polynomial $H(z;\theta)$ and treat it as a function of a scalar z with $|z|>1$. Because its row dimension exceeds the column dimension, there exists a sequence of elementary row operations (i.e., left multiplying $H(z;\theta)$ by elementary matrices that depend on nonnegative powers of z) to reduce its last row to zeros. Call the last row of the product of these elementary matrices $T(z;\theta)$. Then: $T(z;\theta)Y_t = T(z;\theta)H(z;\theta)\epsilon_t = 0$ for all t. This in term implies $T(L;\theta)Y_t = 0$. The result follows because $T(L;\theta)$ depends only on nonnegative powers of L.

Proof of Theorem 1. The result follows from Theorems 1 and 2 in Qu and Tkachenko (2012). Here we still include the complete details for the matter of completeness.

We first simplify the representation for $G_S(\theta)$. Define the following correspondence:

$$f_s(\omega;\theta) \longleftrightarrow f_s(\omega;\theta)^R \quad \text{with} \;\; f_s(\omega;\theta)^R = \begin{bmatrix} \mathrm{Re}(f_s(\omega;\theta)) & \mathrm{Im}(f_s(\omega;\theta)) \\ -\mathrm{Im}(f_s(\omega;\theta)) & \mathrm{Re}(f_s(\omega;\theta)) \end{bmatrix}, \tag{A.1}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real and the imaginary part of a complex matrix, i.e., if $C = A + Bi$, then $\mathrm{Re}(C) = A$ and $\mathrm{Im}(C) = B$. Let $R_s(\omega;\theta) = \mathrm{vec}(f_s(\omega;\theta)^R)$, then

$$\left(\frac{\partial\, \mathrm{vec}\, f_s(\omega;\theta)}{\partial\theta'}\right)^* \left(\frac{\partial\, \mathrm{vec}\, f_s(\omega;\theta)}{\partial\theta'}\right) = \frac{1}{2}\left(\frac{\partial R_s(\omega;\theta)}{\partial\theta'}\right)' \left(\frac{\partial R_s(\omega;\theta)}{\partial\theta'}\right).$$

In addition, let

$$\bar{R}_s(\omega;\theta) = \begin{bmatrix} R_s(\omega;\theta) \\ \frac{1}{\sqrt{\pi}}\mu_s(\theta) \end{bmatrix}.$$

Then, $G_S(\theta)$ can be equivalently represented as

$$G_S(\theta) = \frac{1}{2}\sum_{s=1}^{S}\left\{\int_{-\pi}^{\pi}\left(\frac{\partial \bar{R}_s(\omega;\theta_0)}{\partial\theta'}\right)'\left(\frac{\partial \bar{R}_s(\omega;\theta_0)}{\partial\theta'}\right)d\omega\right\}.$$

This representation is useful because the elements of the function $\bar{R}_s(\omega;\theta)$ are all real valued. This allows us adopt the arguments in Theorem 1 in Rothenberg (1971) to prove the result.

Suppose θ_0 is *not* locally identified. Then, there exists a sequence of vectors $\{\theta_k\}_{k=1}^{\infty}$ approaching θ_0 such that for every k:

$$\bar{R}_s(\omega;\theta_0) = \bar{R}_s(\omega;\theta_k) \text{ for all } \omega \in [-\pi,\pi] \text{ and all } s = 1, ..., S.$$

For an arbitrary $\omega \in [-\pi,\pi]$, $s \in \{1, ...S\}$ and $j \in \{1, ..., \dim(\theta)\}$, by the mean value theorem and the differentiability of $f_s(\omega;\theta)$ and $\mu_s(\theta)$ in θ, we have

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

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

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

The sequence $\{d_k\}$ is an infinite sequence on the unit sphere. Therefore it has a convergent subsequence with a limit d (note that d does not depend on s, j or ω). Without loss of generality, we assume $\{d_k\}$ itself is the convergent subsequence. As $\theta_k \to \theta_0$, d_k approaches d and we have

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

where the convergence holds because $f(\omega;\theta)$ is continuously differentiable in θ. Because this holds for an arbitrary j, it holds for the full vector $\bar{R}_s(\omega;\theta_0)$, implying $[\partial \bar{R}_s(\omega;\theta_0)/\partial\theta']d = 0$, which further implies $d'[\partial \bar{R}_s(\omega;\theta_0)/\partial\theta']'[\partial \bar{R}_s(\omega;\theta_0)/\partial\theta']d = 0$. Because the above result holds for arbitrary ω and s, it also holds when integrating over $\omega \in [-\pi,\pi]$ and summing over $s \in \{1,...S\}$, leading to

$$d' \sum_{s=1}^{S} \left\{ \int_{-\pi}^{\pi} \left(\frac{\partial \bar{R}_s(\omega;\theta_0)}{\partial \theta'} \right)' \left(\frac{\partial \bar{R}_s(\omega;\theta_0)}{\partial \theta'} \right) d\omega \right\} d = 0.$$

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

To show the converse, suppose that $G_S(\theta)$ has constant rank $\rho < q$ in a neighborhood of θ_0 denoted by $\delta(\theta_0)$. Then, consider the characteristic vector $c(\theta)$ associated with one of the zero roots of $G_S(\theta)$. Because

$$\sum_{s=1}^{S} \left\{ \int_{-\pi}^{\pi} \left(\frac{\partial \bar{R}_s(\omega;\theta)}{\partial \theta'} \right)' \left(\frac{\partial \bar{R}_s(\omega;\theta)}{\partial \theta'} \right) d\omega \right\} \times c(\theta) = 0$$

we have

$$\sum_{s=1}^{S} \int_{-\pi}^{\pi} \left(\frac{\partial \bar{R}_s(\omega;\theta)}{\partial \theta'} c(\theta) \right)' \left(\frac{\partial \bar{R}_s(\omega;\theta)}{\partial \theta'} c(\theta) \right) d\omega = 0.$$

Since the integrand is continuous in ω and always non-negative, we must have

$$\left(\frac{\partial \bar{R}_s(\omega;\theta)}{\partial \theta'} c(\theta) \right)' \left(\frac{\partial \bar{R}_s(\omega;\theta)}{\partial \theta'} c(\theta) \right) = 0$$

for all $\omega \in [-\pi,\pi]$, all $\theta \in \delta(\theta_0)$ and all $s \in \{1,...S\}$. Consequently,

$$\frac{\partial \bar{R}_s(\omega;\theta)}{\partial \theta'} c(\theta) = 0 \tag{A.2}$$

for all $\omega \in [-\pi,\pi]$, all $\theta \in \delta(\theta_0)$ and all $s \in \{1,...S\}$. Because $G_S(\theta)$ is continuous and has constant rank in $\delta(\theta_0)$, the vector $c(\theta)$ is continuous in $\delta(\theta_0)$. Consider the curve χ defined by the function $\theta(v)$ which solves for $0 \leq v \leq \bar{v}$ the differential equation: $\partial\theta(v)/\partial v = c(\theta)$ with $\theta(0) = \theta_0$. Then,

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

for all $\omega \in [-\pi,\pi]$, $0 \leq v \leq \bar{v}$ and all $s \in \{1,...S\}$, where the last equality uses (A.2). Thus, $\bar{R}_s(\omega;\theta)$ is constant on the curve χ for all $s \in \{1,...S\}$. This implies that along the curve we have observational equivalence. This completes the proof.

Proof of Lemma 2. First consider the frequency domain formulation as in (C.2). The arguments in Lemma 1 in Hannan (1973) implies:

$$T^{-1}\sum_{j=1}^{T-1}\text{tr}\left\{f_s^{-1}(\omega_j;\theta)I_{s,T}\left(\omega_j\right)\right\}\overset{a.s}{\to}\frac{1}{2\pi}\int_{-\pi}^{\pi}\text{tr}\left\{f_s^{-1}(\omega;\theta)f_{s,0}(\omega)\right\}d\omega,$$

$$T^{-1}\text{tr}\left\{f_s^{-1}(0;\theta)I_{s,T}\left(0;\theta\right)\right\}\overset{a.s}{\to}\frac{1}{2\pi}\left(\mu_{s,0}-\mu_s(\theta)\right)'f_s^{-1}(0;\theta)\left(\mu_{s,0}-\mu_s(\theta)\right).$$

Note that the key step in Hannan's proof is in uniformly approximating $f_s^{-1}(\omega;\theta)$ with a Cesaro sum of its Fourier series, which holds also for multivariate series provided that the smallest eigenvalues of $f_s^{-1}(\omega;\theta)$ are strictly bounded away from 0 for all $\omega\in[-\pi,\pi]$. Using these two results, Lemma 2.1 then follows because $T^{-1}\sum_{j=1}^{T-1}\log\det(f_s(\omega_j;\theta))\to\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det(f_s(\omega;\theta))d\omega$. For the time domain likelihood, Lemma 2.1 follows from Hansen and Sargent (1993, p.49-52). The remaining two results in the Lemma follow from the arguments in Hansen and Sargent (1993, p.53).

Proof of Theorem 2. We first analyze the frequency domain likelihood and then verify that the time domain estimation yields the same asymptotic distribution. Because the effect of the prior vanishes asymptotically, it can be omitted from the derivations. Let $\hat{\theta}_T$ denote the mode. We have

$$T^{-1/2}\frac{\partial\ell(\hat{\theta}_T)}{\partial\theta}=0, \tag{A.3}$$

while the pseudo-true value θ_0 satisfies

$$T^{1/2}\frac{\partial\ell_\infty\left(\theta_0\right)}{\partial\theta}=0. \tag{A.4}$$

Consider a Taylor expansion of (A.3) around θ_0 :

$$\frac{\partial\ell\left(\theta_0\right)}{\partial\theta}+\frac{\partial^2\ell(\bar{\theta})}{\partial\theta\partial\theta'}(\hat{\theta}_T-\theta_0)=0,$$

where $\bar{\theta}$ lies between $\hat{\theta}_T$ and θ_0. Rearrange terms and apply (A.4):

$$T^{1/2}(\hat{\theta}_T-\theta_0)=\left[-\frac{1}{T}\frac{\partial^2\ell(\bar{\theta})}{\partial\theta\partial\theta'}\right]^{-1}\left(T^{-1/2}\frac{\partial\ell\left(\theta_0\right)}{\partial\theta}-T^{1/2}\frac{\partial\ell_\infty\left(\theta_0\right)}{\partial\theta}\right).$$

On the right hand side, the term inside the brackets converges to M defined in the theorem, while the term in the parentheses equals

$$\frac{1}{2T^{1/2}}\sum_{s=1}^{S}\left\{\sum_{j=0}^{T-1}\left(\frac{\partial\text{vec}f_s(\omega_j;\theta_0)}{\partial\theta'}\right)^*\left[f_s^{-1}(\omega_j;\theta_0)'\otimes f_s^{-1}(\omega_j;\theta_0)\right]\text{vec}\left(I_{s,T}\left(\omega_j\right)-f_{s,0}(\omega_j)\right)\quad\text{(I)}\right.$$

$$\left.+\frac{1}{\pi}\sum_{j=1}^{T}\frac{\partial\mu_s(\theta_0)'}{\partial\theta}f_s^{-1}(0;\theta_0)\left(Y_{s,t}-\mu_{s,0}\right)\right\}+o_p\left(1\right). \tag{II}$$

Term (I) satisfies a CLT with $\lim_{T\to\infty}\mathbb{E}\left(\text{vec}\left\{I_{s,T}\left(\omega_j\right)-f_{s,0}(\omega_j)\right\}\text{vec}\left\{I_{h,T}\left(\omega_j\right)-f_{h,0}(\omega_j)\right\}^*\right)=f_{s,h,0}(\omega_j)'\otimes f_{s,h,0}(\omega_j)$. This leads to $V_{1,s,h}$ defined in the theorem. Term (II) also satisfies a CLT with $\lim_{T\to\infty}\mathbb{E}(Y_{s,t}-\mu_{s,0})(Y_{h,t}-\mu_{s,0})' = 2\pi f_{s,h,0}(0)$. This leads to $V_{2,s,h}$ defined in the theorem. The covariance of Terms (I) and (II) is zero because of Assumption MI.

Next, we consider the time domain formulation. Because $T^{1/2}(\hat{\theta}_T-\theta_0) = O_p(1)$, we can restrict our analysis to the following compact set $\left\{\theta : T^{1/2}\left\|\theta-\theta_0\right\| \leq M\right\}$, where for any $\epsilon > 0$, M can be chosen such that $\hat{\theta}_T$ falls into the set with probability at least $1-\epsilon$ in large samples.

Up to some constant, the time domain Gaussian log likelihood that corresponds to the submodel s can also be represented as (see Hannan, 1973 and Hansen and Sargent, 1993)

$$\ell_s(\theta) = -\frac{1}{2}\log\det G_s^{-1}(\theta) - \frac{1}{2}\left[Y_s-\boldsymbol{\mu}_s(\theta)\right]' G_s^{-1}(\theta)\left[Y_s-\boldsymbol{\mu}_s(\theta)\right],$$

where Y_s is a matrix whose t-th row is given by $Y_{s,t}'$ and $\boldsymbol{\mu}_s(\theta)$ and $G_s(\theta)$ correspond to the mean and covariance matrix of Y_s implied by the model. This representation differs from (C.1) in the handling of the initial condition, whose effects vanish asymptotically.

Below, we show that the first order condition in the time domain is asymptotically equivalent to that in the frequency domain. The derivative of $\ell_s(\theta)$ with respect to the k-th element of θ multiplied by $T^{-1/2}$ equals

$$\begin{aligned}
& \frac{1}{2T^{1/2}}\frac{\partial\log\det G_s^{-1}(\theta)}{\partial\theta_k} - \frac{1}{2T^{1/2}}\left[Y_s-\boldsymbol{\mu}_s(\theta)\right]'\frac{\partial G_s^{-1}(\theta)}{\partial\theta_k}\left[Y_s-\boldsymbol{\mu}_s(\theta)\right] \\
& +T^{-1/2}\frac{\partial\boldsymbol{\mu}_s(\theta)'}{\partial\theta_k}G_s^{-1}(\theta)\left[Y_s-\boldsymbol{\mu}_s(\theta)\right] \\
= \; & (A)+(B)+(C).
\end{aligned}$$

We now analyze (A), (B) and (C) separately.

The analysis of Term (B) uses the results in Brockwell and Davis (1991, p.392-393) but applied to multivariate processes. First, define $q_{s,m}(\omega;\theta)$ to be the m-th order Fourier series approximation to $f_s^{-1}(\omega;\theta)$. With $m = O(T^{1/5})$, the approximation error satisfies $\left\|q_{s,m}(\omega;\theta)-f_s^{-1}(\omega;\theta)\right\| + \left\|\partial q_{s,m}(\omega;\theta)/\partial\theta_k - \partial f_s^{-1}(\omega;\theta)/\partial\theta_k\right\| = O(T^{-3/5})$ uniformly over ω and θ. This implies

$$T^{-1/2}\sum_{j=0}^{T-1} I_{s,T}(\omega_j)\left[\frac{\partial f_s^{-1}(\omega;\theta)}{\partial\theta_k} - \frac{\partial q_{s,m}(\omega;\theta)}{\partial\theta_k}\right] = o_p(1).$$

Next, view $(4\pi^2)^{-1}q_{s,m}(\omega;\theta)$ as a spectral density and let $\widetilde{H}_s^{-1}(\theta)$ be the covariance matrix of the resulting VMA(m) process. Similar to Displays (10.8.17) and (10.8.45) in Brockwell and Davis (1991, p.392-393), we have

$$T^{-1/2}\left[Y_s-\boldsymbol{\mu}_s(\theta)\right]'\left(\frac{\partial G_s^{-1}(\theta)}{\partial\theta_k} - \frac{\partial\widetilde{H}_s^{-1}(\theta)}{\partial\theta_k}\right)\left[Y_s-\boldsymbol{\mu}_s(\theta)\right] = o_p(1).$$

Then, applying the relationship between $q_{s,m}(\omega;\theta)$ and $\widetilde{H}_s^{-1}(\theta)$, we have

$$T^{-1/2}\left[Y_s-\boldsymbol{\mu}_s(\theta)\right]'\frac{\partial\widetilde{H}_s^{-1}(\theta)}{\partial\theta_k}\left[Y_s-\boldsymbol{\mu}_s(\theta)\right] - T^{-1/2}\sum_{j=0}^{T-1} I_{s,T}(\omega_j)\frac{\partial q_{s,m}(\omega;\theta)}{\partial\theta_k} = o_p(1).$$

The results in the above three displays hold uniformly over the compact set defined above. Combining them, we have

$$(B) = -\frac{1}{2T^{1/2}} \sum_{j=0}^{T-1} I_{s,T}\left(\omega_j\right) \frac{\partial f_s^{-1}(\omega;\theta)}{\partial \theta_k} + o_p\left(1\right). \tag{A.5}$$

For Term (A), note that

$$\begin{aligned}(A) &= \frac{1}{2T^{1/2}} \operatorname{tr}\left\{G_s\left(\theta\right) \frac{\partial G_s^{-1}\left(\theta\right)}{\partial \theta_k}\right\} = \frac{1}{2T^{1/2}} \operatorname{tr}\mathbb{E}\left\{\left[Y_s(\theta) - \boldsymbol{\mu}_s(\theta)\right]\left[Y_s(\theta) - \boldsymbol{\mu}_s(\theta)\right]' \frac{\partial G_s^{-1}\left(\theta\right)}{\partial \theta_k}\right\} \\ &= \frac{1}{2T^{1/2}} \operatorname{tr}\mathbb{E}\left\{\left[Y_s(\theta) - \boldsymbol{\mu}_s(\theta)\right]' \frac{\partial G_s^{-1}\left(\theta\right)}{\partial \theta_k}\left[Y_s(\theta) - \boldsymbol{\mu}_s(\theta)\right]\right\},\end{aligned}$$

where $Y_s(\theta)$ denote a random vector with mean $\boldsymbol{\mu}_s(\theta)$ and covariance $G_s\left(\theta\right)$. Applying the argument for proving (B) and then take the expectation, we have

$$(A) = \frac{1}{2T^{1/2}} \sum_{j=0}^{T-1} \mathbb{E}\operatorname{tr}\left\{I_{s,T}\left(\omega_j\right) \frac{\partial f_s^{-1}(\omega_j;\theta)}{\partial \theta_k}\right\} + o\left(1\right) = \frac{1}{2T^{1/2}} \sum_{j=0}^{T-1} \operatorname{tr}\left\{f_s(\omega_j;\theta) \frac{\partial f_s^{-1}(\omega_j;\theta)}{\partial \theta_k}\right\} + o\left(1\right). \tag{A.6}$$

Finally, applying the same argument used for (B) to Term (C), we have

$$(C) = \frac{1}{\pi T^{1/2}} \sum_{j=1}^{T} \frac{\partial \boldsymbol{\mu}_s(\theta)'}{\partial \theta_k} f_s^{-1}(0;\theta)\left(Y_s - \boldsymbol{\mu}_s(\theta)\right) + o\left(1\right). \tag{A.7}$$

Combining the leading terms of (A.6), (A.5) and (A.7), we arrive at the first order derivative of the frequency domain likelihood (C.2) with respect to θ_k. This completes the proof.

Appendix B. The Smets and Wouters (2007) Model

The model has seven observable endogenous variables and seven shocks. Below is an outline of the log linearized system. Its singular versions, as explained in the main text, are obtained by removing a subset or all of the following three shocks: the risk premium shock, the price market shock and the wage mark up shock.

The aggregate resource constraint: It satisfies

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

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

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

The coefficients c_y, i_y and z_y are functions of the steady state spending-output ratio (g_y), steady state output growth ($\gamma = 1 + \bar{\gamma}/100$), capital depreciation (δ), household discount factor (β), intertemporal elasticity of substitution (σ_c), fixed costs in production (ϕ_p), and share of capital in production (α): $i_y = (\gamma - 1 + \delta)k_y$, $c_y = 1 - g_y - i_y$, and $z_y = R_*^k k_y$. Here, k_y is the steady state capital-output ratio, and R_*^k is the steady state rental rate of capital: $k_y = \phi_p \left(L_*/k_*\right)^{\alpha-1} = \phi_p \left[((1-\alpha)/\alpha)\left(R_*^k/w_*\right)\right]^{\alpha-1}$ with $w_* = \left(\alpha^\alpha(1-\alpha)^{(1-\alpha)}/[\phi_p \left(R_*^k\right)^\alpha]\right)^{1/(1-\alpha)}$, and $R_*^k = \beta^{-1}\gamma^{\sigma_c} - (1-\delta)$.

Households: The consumption Euler equation is

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

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

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

The relationship of the coefficients in (B.1) to the habit persistence (λ), steady state labor market mark-up (ϕ_w), and other structural parameters highlighted above is:

$$c_1 = \frac{\lambda/\gamma}{1+\lambda/\gamma}, c_2 = \frac{(\sigma_c - 1)\left(w_*^h L_*/c_*\right)}{\sigma_c\left(1+\lambda/\gamma\right)}, c_3 = \frac{1-\lambda/\gamma}{\left(1+\lambda/\gamma\right)\sigma_c}, \text{ where } w_*^h L_*/c_* = \frac{1}{\phi_w}\frac{1-\alpha}{\alpha}R_*^k k_y \frac{1}{c_y}$$

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

The dynamics of households' investment are given by

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

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

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

The coefficients satisfy $i_1 = 1/(1+\beta\gamma^{(1-\sigma_c)})$ and $i_2 = 1/[\left(1+\beta\gamma^{(1-\sigma_c)}\right)\gamma^2\varphi]$, where φ is the steady state elasticity of the capital adjustment cost function. The corresponding arbitrage equation for the value of capital is given by

$$q_t = q_1 E_t q_{t+1} + (1-q_1)\, E_t r_{t+1}^k - (r_t - E_t \pi_{t+1}) - \frac{1}{c_3}\varepsilon_t^b\ , \tag{B.2}$$

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

Final and intermediate goods market: The aggregate production function is

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

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

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

The current capital service usage (k_t^s) is a function of capital installed in the previous period (k_{t-1}) and the degree of capital utilization (z_t): $k_t^s = k_{t-1} + z_t$. Furthermore, the capital utilization is a positive fraction of the rental rate of capital (r_t^k): $z_t = z_1 r_t^k$, where $z_1 = (1-\psi)/\psi$,and ψ is a positive function of the elasticity of the capital utilization adjustment cost function and normalized to be between zero and one. The accumulation of installed capital (k_t) satisfies

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

where ε_t^i is the investment specific technology process as defined before, and k_1 and k_2 satisfy $k_1 = (1-\delta)/\gamma$ and $k_2 = (1-k_1)\left(1+\beta\gamma^{(1-\sigma_c)}\right)\gamma^2\varphi$.

The price mark-up satisfies $\mu_t^p = \alpha\left(k_t^s - l_t\right) + \varepsilon_t^a - w_t$, where w_t is the real wage. The New Keynesian Phillips curve is

$$\pi_t = \pi_1 \pi_{t-1} + \pi_2 E_t \pi_{t+1} - \pi_3 \mu_t^p + \varepsilon_t^p, \tag{B.3}$$

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

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

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

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

Cost minimization by firms implies that the rental rate of capital satisfies $r_t^k = -\left(k_t^s - l_t\right) + w_t$.

Labor market: The wage mark-up is

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

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

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

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

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

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

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

Monetary policy: The empirical monetary policy reaction function is

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

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

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

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

Appendix C. Some details on implementation

This appendix shows how to compute the Gaussian likelihood for a nonsingular submodel in both the time and the frequency domain. The material is not new; it is included to facilitate the methods' implementation in practice.

The time domain Gaussian likelihood for a nonsingular submodel. The model is

$$Y_{s,t} = P_s\mu(\theta) + P_sC(\theta)X_t + P_sD(\theta)v_t, \quad X_t = A(\theta)X_{t-1} + B(\theta)\varepsilon_t,$$

where $\varepsilon_t \sim i.i.d.N(0,Q)$, $v_t \sim i.i.d.N(0,H)$ and $E(\varepsilon_t v_t') = V$.

Initialization. Suppose the initial condition satisfies $X_{0|0} \sim N(0, P_{0|0})$, where $P_{0|0}$ is a symmetric positive definite matrix.

Prediction. Obtain the optimal forecast of X_t and its mean squared forecast error (MSE) using the information available at $t-1$:

$$\begin{aligned} X_{t|t-1} &= A(\theta)X_{t-1|t-1} \\ P_{t|t-1} &= A(\theta)P_{t-1|t-1}A(\theta)' + B(\theta)QB(\theta)', \end{aligned}$$

The corresponding prediction error for $Y_{s,t}$ and its MSE then equal

$$\begin{aligned} \eta_t &= Y_{s,t} - P_s\mu(\theta) - P_sC(\theta)X_{t|t-1}, \\ F_t &= P_s\left[C(\theta)P_{t|t-1}C(\theta)' + D(\theta)HD(\theta)' + C(\theta)B(\theta)VD(\theta)' + D(\theta)V'B(\theta)'C(\theta)'\right]P_s'. \end{aligned}$$

Updating. Upon observing Y_t, compute the optimal estimator for the state and its MSE as

$$\begin{aligned} X_{t|t} &= X_{t|t-1} + \left(P_{t|t-1}C(\theta)' + B(\theta)VD(\theta)'\right)P_s'F_t^{-1}\eta_t \\ P_{t|t} &= P_{t|t-1} - \left(P_{t|t-1}C(\theta)' + B(\theta)VD(\theta)'\right)P_s'F_t^{-1}P_s\left(C(\theta)P_{t|t-1} + D(\theta)V'B(\theta)'\right). \end{aligned}$$

After implementing the predication and updating steps sequentially for $t = 1, 2, ..., T$, we obtain the log likelihood:

$$l_s(\theta) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log\det(F_t) - \frac{1}{2}\sum_{t=1}^{T}\eta_t'F_t^{-1}\eta_t. \tag{C.1}$$

The frequency domain Gaussian likelihood for a nonsingular submodel. Let ω_j denote the Fourier frequencies, i.e., $\omega_j = 2\pi j/T$ $(j = 1, 2, ..., T-1)$. The discrete Fourier transforms and periodograms of $Y_{s,t}$ at such frequencies are: $w_{s,T}(\omega_j) = (2\pi T)^{-1/2}\sum_{t=1}^{T} Y_{s,t}\exp(-i\omega_j t)$ and $I_{s,T}(\omega_j) = w_{s,T}(\omega_j)\, w_{s,T}(\omega_j)^*$. At the zero frequency, let $w_{s,T}(0;\theta) = (2\pi T)^{-1/2}\sum_{t=1}^{T}(Y_{s,t} - P_s\mu(\theta))$ and $I_{s,T}(0;\theta) = w_{s,T}(0;\theta)\, w_{s,T}(0;\theta)^*$. The spectral density matrix satisfies

$$f_s(\omega;\theta) = (2\pi)^{-1}P_sH(\exp(-i\omega);\theta)\Sigma(\theta)H(\exp(-i\omega);\theta)^*P_s^*.$$

An approximate log-likelihood for θ up a constant is then given by (see Hansen and Sargent, 1993)

$$\begin{aligned} &-\frac{1}{2}\sum_{j=1}^{T-1}\left[\log\det\left(f_s(\omega_j;\theta)\right) + \mathrm{tr}\left\{f_s^{-1}(\omega_j;\theta)I_{s,T}(\omega_j)\right\}\right] \\ &-\frac{1}{2}\left[\log\det\left(f_s(0;\theta)\right) + \mathrm{tr}\left\{f_s^{-1}(0;\theta)I_{s,T}(0;\theta)\right\}\right]. \end{aligned} \tag{C.2}$$

Table 1. Estimation results for small scale models

	Prior			(a) One shock				(b) Two shocks				(c) Three shocks			
	Density	Mean	SD	Mode	Mean	MCMC	Asymptotic	Mode	Mean	MCMC	Asymptotic	Mode	Mean	MCMC	Asymptotic
ψ_1	Gamma	1.10	0.50	1.23	1.36	[1.04,1.85]	[1.00,1.89]	1.56	1.64	[1.11,2.34]	[1.08,2.25]	2.15	2.20	[1.45,3.07]	[1.41,3.05]
ψ_2	Gamma	0.25	0.13	0.20	0.24	[0.09,0.46]	[0.19,0.30]	0.21	0.27	[0.09,0.52]	[0.24,0.30]	0.23	0.29	[0.10,0.57]	[0.24,0.34]
ρ_r	Beta	0.50	0.20	0.77	0.78	[0.71,0.84]	[0.69,0.86]	0.79	0.79	[0.71,0.86]	[0.71,0.85]	0.84	0.84	[0.79,0.89]	[0.79,0.89]
π^*	Gamma	4.00	2.00	3.45	3.45	[2.59,4.34]	[2.66,4.23]	3.38	3.37	[2.64,4.10]	[2.55,4.13]	3.43	3.42	[2.87,3.98]	[2.95,3.93]
r^*	Gamma	2.00	1.00	2.82	2.83	[1.48,4.21]	[1.67,4.03]	2.85	2.84	[1.63,4.08]	[1.91,3.76]	3.04	3.02	[2.24,3.80]	[2.18,3.88]
κ	Gamma	0.50	0.20	0.36	0.42	[0.20,0.72]	[0.13,0.82]	0.38	0.39	[0.21,0.62]	[0.17,0.66]	0.55	0.57	[0.30,0.91]	[0.24,0.90]
τ^{-1}	Gamma	2.00	0.50	1.72	1.97	[1.38,2.63]	[1.75,2,22]	1.72	1.83	[1.17,2.63]	[1.50,2.22]	1.79	1.86	[1.14,2.72]	[1.36,2.38]
ρ_g	Beta	0.70	0.10	–	–	–	–	0.81	0.81	[0.75,0.87]	[0.76,0.85]	0.83	0.83	[0.76,0.89]	[0.77,0.88]
ρ_z	Beta	0.70	0.10	0.85	0.86	[0.79,0.91]	[0.77,0.94]	0.88	0.87	[0.79,0.93]	[0.70,0.99]	0.85	0.84	[0.76,0.92]	[0.70,0.98]
ρ_{gz}	Normal	0.00	0.40	–	–	–	–	0.18	0.17	[-0.17,0.51]	[-0.18,0.55]	0.41	0.35	[0.02,0.64]	[-0.13,0.81]
σ_r	IGamma	0.31	0.16	–	–	–	–	–	–	–	–	0.17	0.18	[0.15,0.22]	[0.14,0.22]
σ_g	IGamma	0.38	0.20	–	–	–	–	0.18	0.19	[0.15,0.25]	[0.16,0.23]	0.17	0.18	[0.14,0.23]	[0.15,0.22]
σ_z	IGamma	1.00	0.52	0.89	0.90	[0.70,1.19]	[0.73,1.15]	0.57	0.59	[0.50,0.77]	[0.50,0.75]	0.61	0.64	[0.53,0.78]	[0.49,0.81]

Note. SD: standard deviation. MCMC: 90% intervals using the quantiles of the MCMC draws. Asymptotic: 90% intervals using Procedure A. The estimates are based on 200,000 draws.The lower bounds for the parameters are (same order as in the table): {1.0,1E-5,1E-5,1E-5,1E-5,1E-5,1E-5,1E-5,1E-5,-0.999,1E-5,1E-5,1E-5}. The upper bounds are: {10, 0.999, 0.999,10, 100, 10, 100, 0.999, 0.999, 0.999,10, 10, 10}.

Table 2: Parameters and estimates of the original Smets and Wouters (2007) model

	Parameter interpretation	Prior			Posterior		
		Distribution	Mean	SD	Mode	Mean	MCMC
α	Share of capital in production	Normal	0.30	0.05	0.19	0.19	[0.16,0.21]
ψ	Elasticity of capital utilization adjustment cost	Beta	0.50	0.15	0.54	0.54	[0.36,0.72]
φ	Investment adjustment cost	Normal	4.00	1.50	5.48	5.74	[3.97,7.42]
σ_c	Elasticity of inertemporal substitution	Normal	1.50	0.38	1.39	1.38	[1.16,1.59]
λ	Habit persistence	Beta	0.70	0.10	0.71	0.71	[0.64,0.78]
ϕ_p	Fixed costs in production	Normal	1.25	0.13	1.61	1.60	[1.48,1.73]
ι_w	Wage indexation	Beta	0.50	0.15	0.59	0.58	[0.38,0.78]
ξ_w	Wage stickiness	Beta	0.50	0.10	0.73	0.70	[0.60,0.81]
ι_p	Price indexation	Beta	0.50	0.15	0.22	0.24	[0.10,0.38]
ξ_p	Price stickiness	Beta	0.50	0.10	0.65	0.66	[0.56,0.74]
σ_l	Labor supply elasticity	Normal	2.00	0.75	1.92	1.83	[0.91,2.78]
r_π	Taylor rule: inflation weight	Normal	1.50	0.25	2.03	2.04	[1.74,2.33]
$r_{\Delta y}$	Taylor rule: output gap change weight	Normal	0.13	0.05	0.22	0.22	[0.18,0.27]
r_y	Taylor rule: output gap weight	Normal	0.13	0.05	0.08	0.08	[0.05,0.12]
ρ	Taylor rule: interest rate smoothing	Beta	0.75	0.10	0.81	0.81	[0.77,0.85]
ρ_a	Productivity shock AR	Beta	0.50	0.20	0.95	0.95	[0.94,0.97]
ρ_b	Risk premium shock AR	Beta	0.50	0.20	0.18	0.22	[0.07,0.36]
ρ_g	Exogenous spending shock AR	Beta	0.50	0.20	0.97	0.97	[0.96,0.99]
ρ_i	Investment shock AR	Beta	0.50	0.20	0.71	0.71	[0.61,0.80]
ρ_r	Monetary policy shock AR	Beta	0.50	0.20	0.12	0.15	[0.04,0.24]
ρ_p	Price mark-up shock AR	Beta	0.50	0.20	0.90	0.89	[0.80,0.96]
μ_p	Price mark-up shock MA	Beta	0.50	0.20	0.74	0.69	[0.54,0.85]
ρ_w	Wage mark-up shock AR	Beta	0.50	0.20	0.97	0.96	[0.94,0.99]
μ_w	Wage mark-up shock MA	Beta	0.50	0.20	0.88	0.84	[0.75,0.93]
ρ_{ga}	Cross-corr.: tech. and exog. spending shocks	Normal	0.50	0.25	0.52	0.52	[0.37,0.66]
σ_a	Productivity shock std. dev.	IGamma	0.10	2.00	0.45	0.45	[0.41,0.50]
σ_b	Risk premium shock std. dev.	IGamma	0.10	2.00	0.24	0.23	[0.19,0.27]
σ_g	Exogenous spending shock std. dev.	IGamma	0.10	2.00	0.52	0.53	[0.48,0.58]
σ_i	Investment shock std. dev.	IGamma	0.10	2.00	0.45	0.45	[0.37,0.53]
σ_r	Monetary policy shock std. dev.	IGamma	0.10	2.00	0.24	0.24	[0.22,0.27]
σ_p	Price mark-up shock std. dev.	IGamma	0.10	2.00	0.14	0.14	[0.11,0.16]
σ_w	Wage mark-up shock std. dev.	IGamma	0.10	2.00	0.24	0.24	[0.20,0.28]
$\overline{\gamma}$	Trend growth: real GDP, Infl., Wages	Normal	0.40	0.10	0.43	0.43	[0.40,0.45]
r	Discount rate	Gamma	0.25	0.10	0.16	0.16	[0.07,0.26]
$\overline{\pi}$	Steady state inflation rate	Gamma	0.62	0.10	0.81	0.78	[0.61,0.96]
$\bar{l}$	Steady state hours worked	Normal	0.00	2.00	-0.1	0.53	[-1.3,2.32]

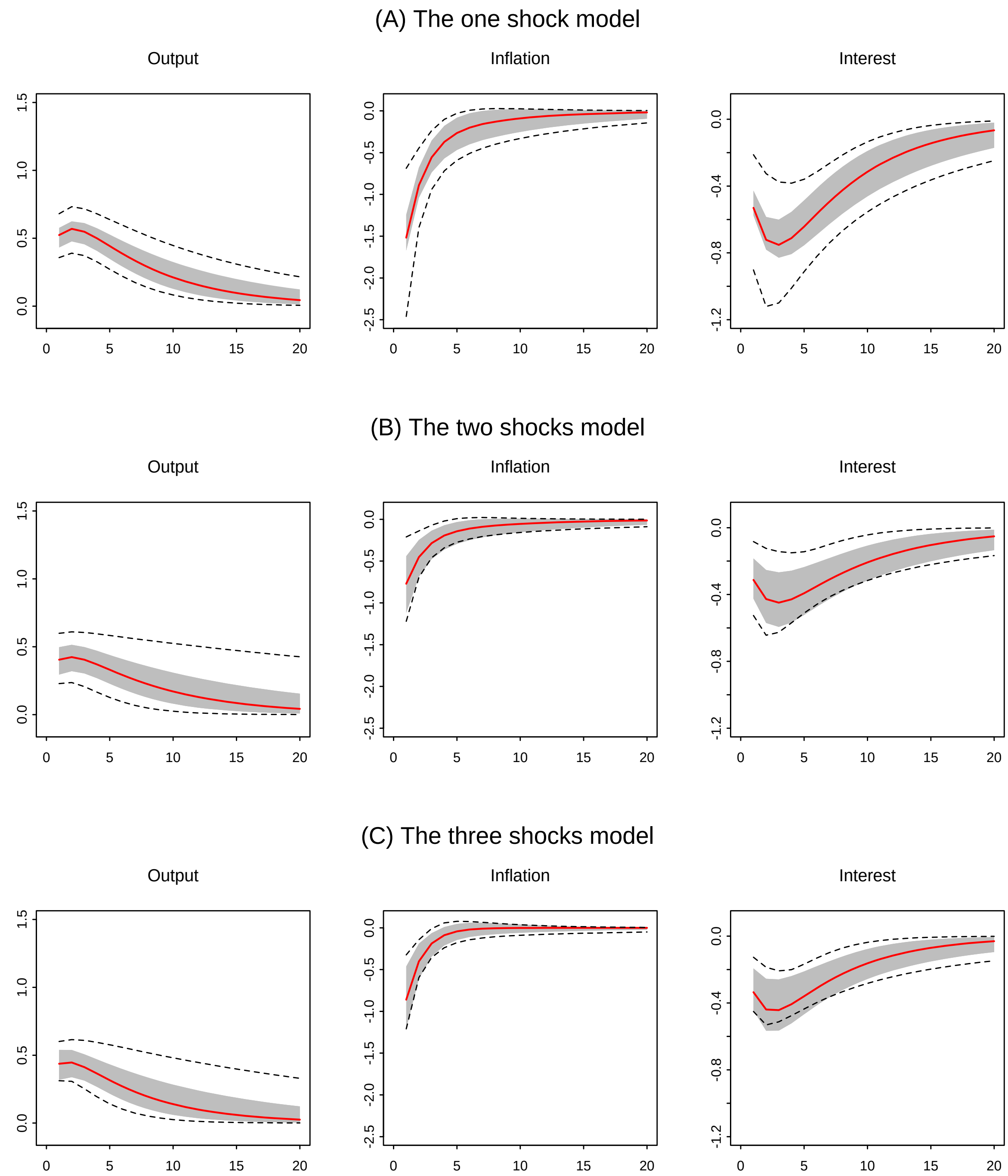
Note. The prior distributions are taken from Smets and Wouters' (2007) Dynare code. MCMC stands for 90% intervals obtained using the quantiles of the MCMC draws. The discount rate $r = 100(\beta^{-1} - 1)$. The following five parameters are kept fixed: capital depreciation rate ($\delta = 0.025$), steady state labor market mark-up ($\phi_w = 1.50$), steady state exogenous spending-output ratio ($g_y = 0.18$), curvatures of Kimball goods and labor market aggregators ($\epsilon_p = \epsilon_w = 10$).

Table 3: Estimation results for medium scale models

	The four shocks model				The five shocks model				The six shocks model			
	Mode	Mean	MCMC	Asymptotic	Mode	Mean	MCMC	Asymptotic	Mode	Mean	MCMC	Asymptotic
α	0.25	0.25	[0.21,0.29]	[0.21,0.29]	0.28	0.28	[0.25,0.32]	[0.25,0.32]	0.28	0.28	[0.25,0.31]	[0.24,0.32]
ψ	0.41	0.45	[0.24,0.67]	[0.30,0.59]	0.43	0.48	[0.31,0.64]	[0.28,0.65]	0.43	0.44	[0.29,0.61]	[0.28,0.62]
φ	5.65	5.74	[3.96,7.70]	[4.91,6.55]	5.10	5.29	[3.60,7.18]	[4.03,6.55]	5.04	5.20	[3.54,7.02]	[3.94,6.57]
σ_c	1.62	1.60	[1.31,1.91]	[1.32,1.91]	1.73	1.70	[1.35,2.11]	[1.27,2.12]	1.79	1.74	[1.42,2.11]	[1.45,2.04]
λ	0.34	0.37	[0.29,0.45]	[0.25,0.49]	0.29	0.31	[0.24,0.39]	[0.24,0.38]	0.32	0.34	[0.26,0.40]	[0.27,0.40]
ϕ_p	1.32	1.34	[1.20,1.49]	[1.21,1.48]	1.47	1.49	[1.36,1.63]	[1.40,1.59]	1.58	1.59	[1.47,1.72]	[1.50,1.69]
ι_w	0.87	0.85	[0.74,0.94]	[0.77,0.92]	0.86	0.83	[0.71,0.93]	[0.76,0.90]	0.58	0.56	[0.35,0.76]	[0.42,0.71]
ξ_w	0.86	0.84	[0.76,0.89]	[0.72,0.95]	0.60	0.60	[0.44,0.73]	[0.38,0.79]	0.78	0.76	[0.67,0.84]	[0.62,0.87]
ι_p	0.12	0.13	[0.05,0.23]	[0.06,0.22]	0.21	0.22	[0.10,0.37]	[0.14,0.31]	0.24	0.27	[0.13,0.43]	[0.13,0.42]
ξ_p	0.23	0.22	[0.14,0.31]	[0.15,0.28]	0.71	0.70	[0.61,0.80]	[0.61,0.80]	0.66	0.65	[0.57,0.73]	[0.55,0.74]
σ_l	2.66	2.38	[1.27,3.49]	[1.03,3.57]	2.15	2.11	[1.27,3.06]	[1.24,3.00]	2.71	2.68	[1.78,3.64]	[2.11,3.27]
r_π	1.39	1.41	[1.20,1.64]	[1.15,1.69]	2.08	2.07	[1.79,2.35]	[1.88,2.26]	2.00	2.00	[1.73,2.28]	[1.83,2.16]
$r_{\Delta y}$	0.23	0.23	[0.16,0.29]	[0.16,0.30]	0.24	0.25	[0.20,0.30]	[0.20,0.30]	0.26	0.26	[0.21,0.31]	[0.21,0.31]
r_y	0.18	0.17	[0.09,0.24]	[0.07,0.26]	0.11	0.12	[0.08,0.17]	[0.06,0.19]	0.12	0.12	[0.07,0.18]	[0.04,0.21]
ρ	0.75	0.73	[0.67,0.79]	[0.66,0.80]	0.77	0.77	[0.71,0.81]	[0.70,0.82]	0.79	0.79	[0.74,0.83]	[0.74,0.83]
ρ_a	0.99	0.99	[0.98,0.99]	[0.98,0.99]	0.98	0.98	[0.96,0.99]	[0.95,0.99]	0.98	0.98	[0.96,0.99]	[0.94,0.99]
ρ_b	–	–	–	–	–	–	–	–	–	–	–	–
ρ_g	0.90	0.90	[0.86,0.94]	[0.86,0.94]	0.92	0.92	[0.87,0.95]	[0.86,0.97]	0.91	0.91	[0.86,0.95]	[0.83,0.99]
ρ_i	0.74	0.75	[0.67,0.82]	[0.66,0.84]	0.64	0.65	[0.55,0.74]	[0.54,0.75]	0.64	0.64	[0.55,0.73]	[0.55,0.73]
ρ_r	0.13	0.15	[0.05,0.27]	[0.03,0.29]	0.09	0.12	[0.04,0.22]	[0.04,0.21]	0.08	0.10	[0.03,0.18]	[0.04,0.17]
ρ_p	–	–	–	–	0.97	0.97	[0.94,0.99]	[0.93,0.99]	0.86	0.85	[0.75,0.93]	[0.75,0.93]
μ_p	–	–	–	–	0.81	0.78	[0.63,0.89]	[0.62,0.91]	0.70	0.66	[0.45,0.82]	[0.44,0.82]
ρ_w	–	–	–	–	–	–	–	–	0.98	0.96	[0.91,0.99]	[0.82,0.99]
μ_w	–	–	–	–	–	–	–	–	0.92	0.87	[0.77,0.94]	[0.64,0.99]
ρ_{ga}	0.48	0.46	[0.30,0.63]	[0.30,0.63]	0.40	0.42	[0.24,0.59]	[0.24,0.60]	0.41	0.44	[0.25,0.63]	[0.24,0.67]
σ_a	0.53	0.55	[0.47,0.64]	[0.45,0.65]	0.50	0.50	[0.45,0.56]	[0.44,0.56]	0.47	0.47	[0.42,0.53]	[0.41,0.55]
σ_b	–	–	–	–	–	–	–	–	–	–	–	–
σ_g	0.53	0.54	[0.48,0.62]	[0.48,0.61]	0.57	0.59	[0.51,0.70]	[0.50,0.70]	0.56	0.59	[0.51,0.67]	[0.51,0.67]
σ_i	0.49	0.50	[0.42,0.57]	[0.39,0.61]	0.56	0.57	[0.49,0.67]	[0.43,0.73]	0.57	0.58	[0.49,0.67]	[0.44,0.73]
σ_r	0.31	0.32	[0.28,0.36]	[0.27,0.37]	0.29	0.30	[0.27,0.34]	[0.25,0.36]	0.29	0.30	[0.27,0.33]	[0.24,0.36]
σ_p	–	–	–	–	0.15	0.15	[0.12,0.18]	[0.11,0.19]	0.14	0.14	[0.11,0.17]	[0.11,0.17]
σ_w	–	–	–	–	–	–	–	–	0.25	0.25	[0.21,0.28]	[0.18,0.30]
$\overline{\gamma}$	0.37	0.38	[0.33,0.42]	[0.33,0.42]	0.40	0.40	[0.36,0.44]	[0.34,0.46]	0.35	0.36	[0.32,0.39]	[0.31,0.40]
r	0.10	0.12	[0.06,0.20]	[0.09,0.16]	0.11	0.14	[0.07,0.22]	[0.11,0.16]	0.12	0.14	[0.07,0.24]	[0.11,0.18]
$\overline{\pi}$	0.70	0.72	[0.55,0.90]	[0.64,0.81]	0.65	0.67	[0.51,0.84]	[0.63,0.71]	0.66	0.69	[0.52,0.86]	[0.60,0.80]
$\bar{l}$	-0.66	-0.65	[-1.8,0.46]	[-1.9,0.57]	0.52	0.38	[0.85,1.77]	[-0.83,1.57]	0.69	0.42	[-0.97,1.91]	[-0.93,1.67]

Note. The prior distributions are the same as in Table 2. MCMC: 90% intervals obtained using the quantiles of the MCMC draws. Asymptotic: 90% intervals obtained using Procedure A. The discount rate $r = 100(\beta^{-1} - 1)$. The estimates are based on 200,000 draws.

Figure 1. Responses to productivity shocks in small scale models



Note. Solid line: impulse response at the posterior mean. Shaded area: intervals formed using the MCMC draws. Dashed lines: intervals formed using Procedure A. Y-axis: percent. X-axis: horizon. The interest and inflation rates are annualized.

Figure 2. Response of output to a productivity shock

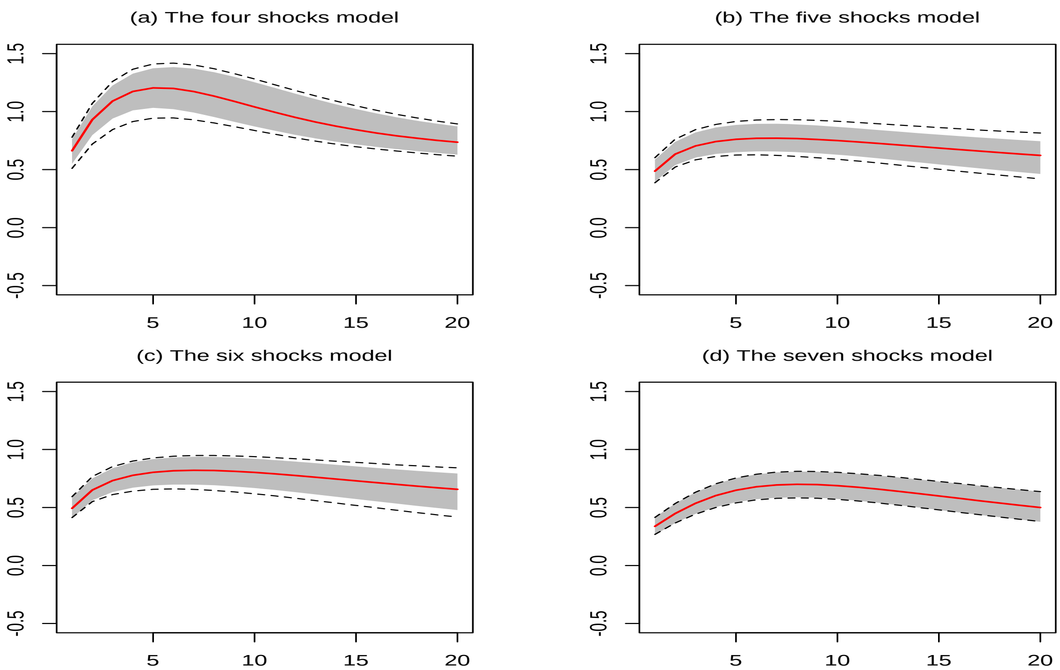
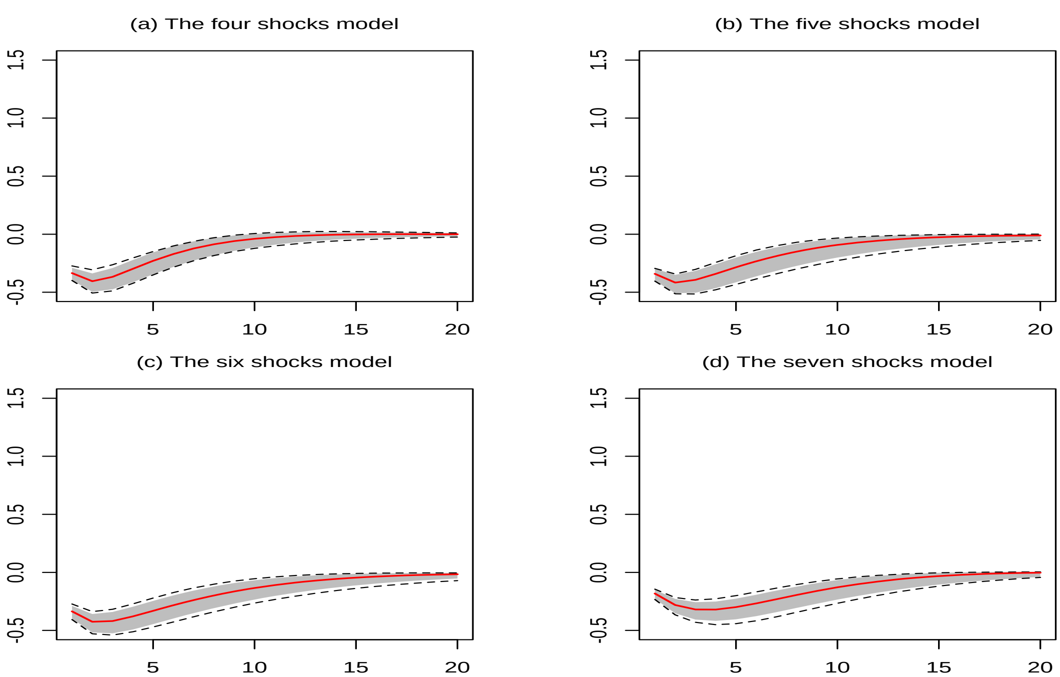


Figure 3. Response of output to a monetary policy shock



Note. See Figure 1. The variables are not annualized.

Figure 4. Response of output to an investment shock

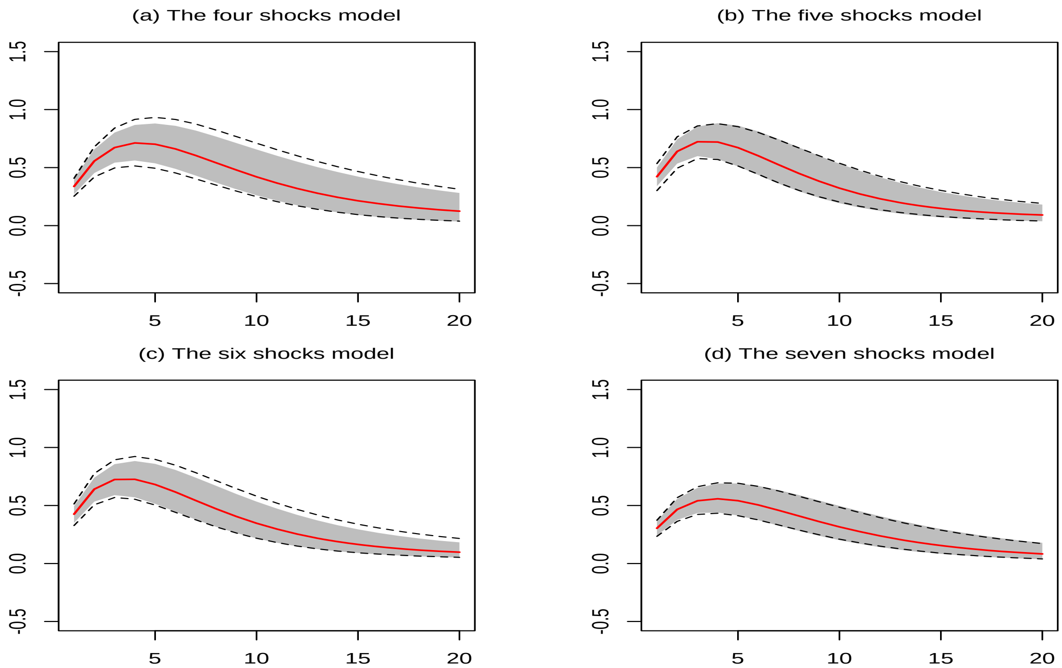
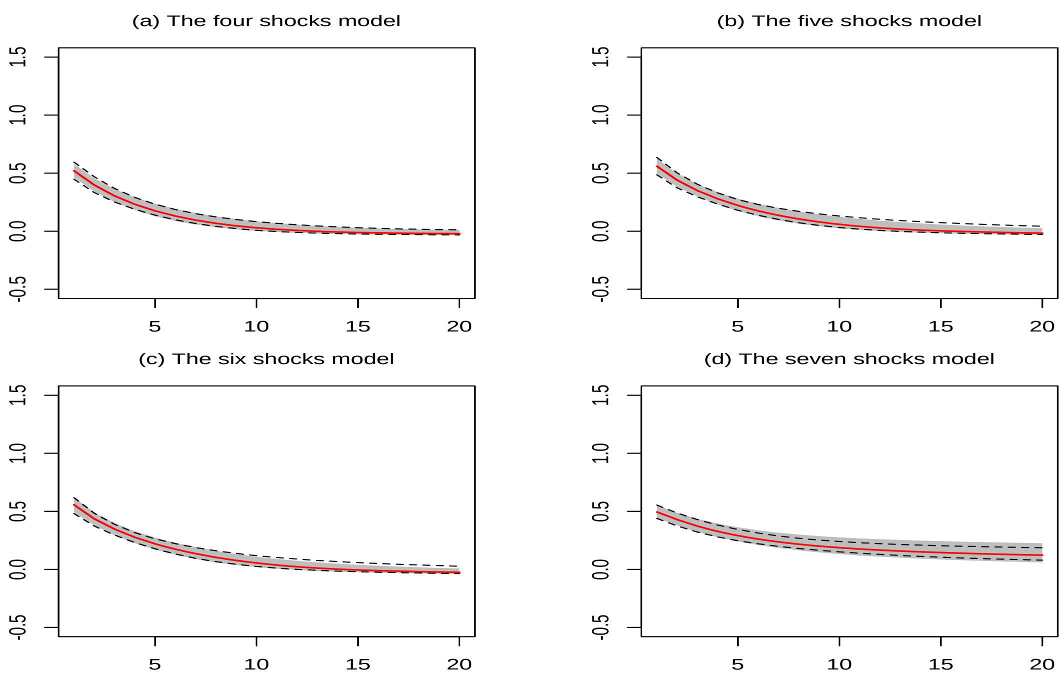


Figure 5. Response of output to an exogenous spending shock



Note. See Figure 1. The variables are not annualized.

Figure 6. Response of inflation to a productivity shock

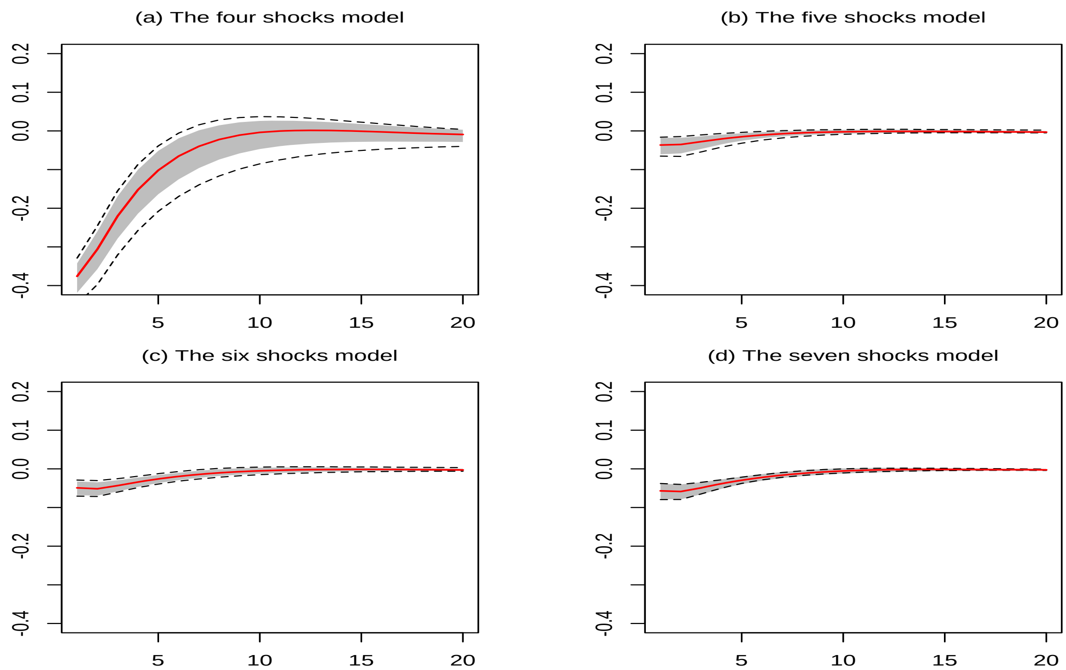
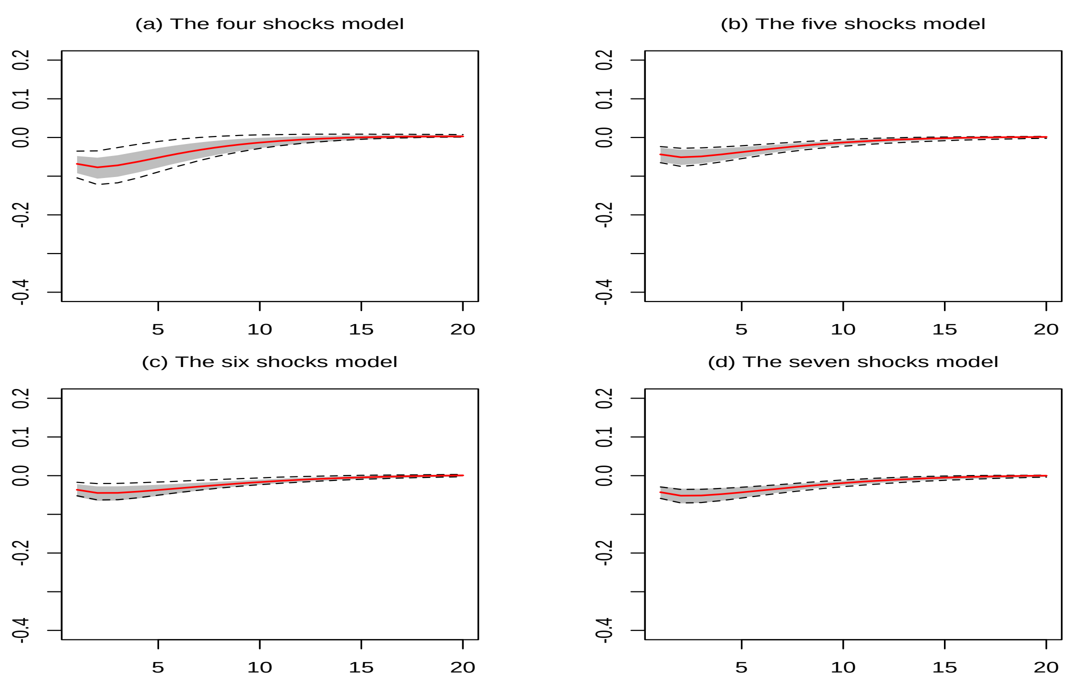


Figure 7. Response of inflation to a monetary policy shock



Note. See Figure 1. The variables are not annualized.

Figure 8. Response of inflation to an investment shock

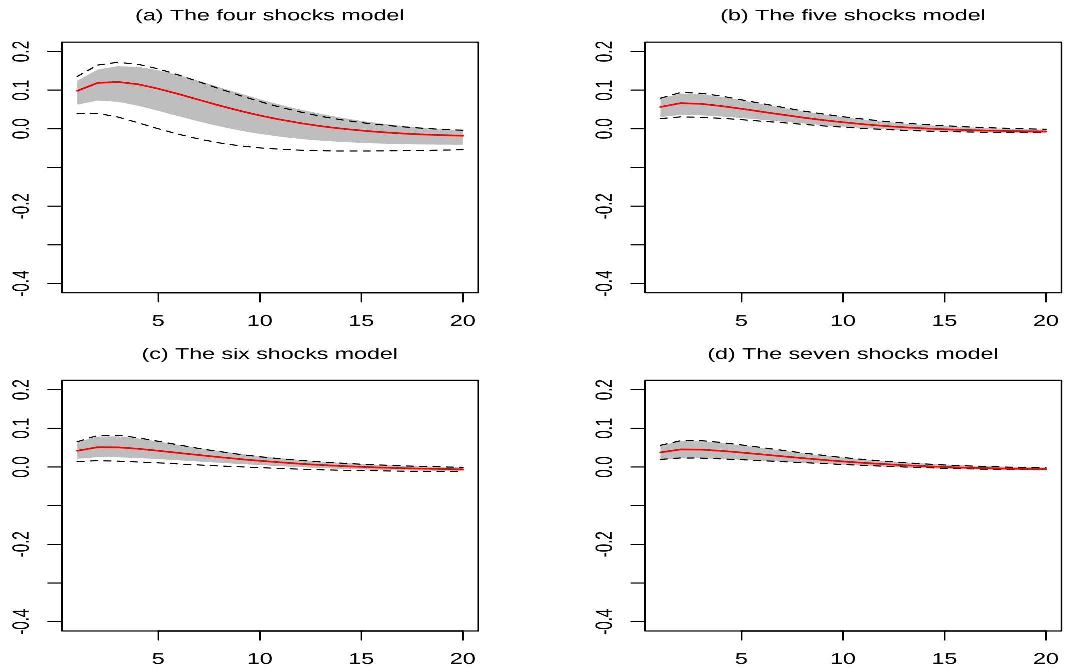
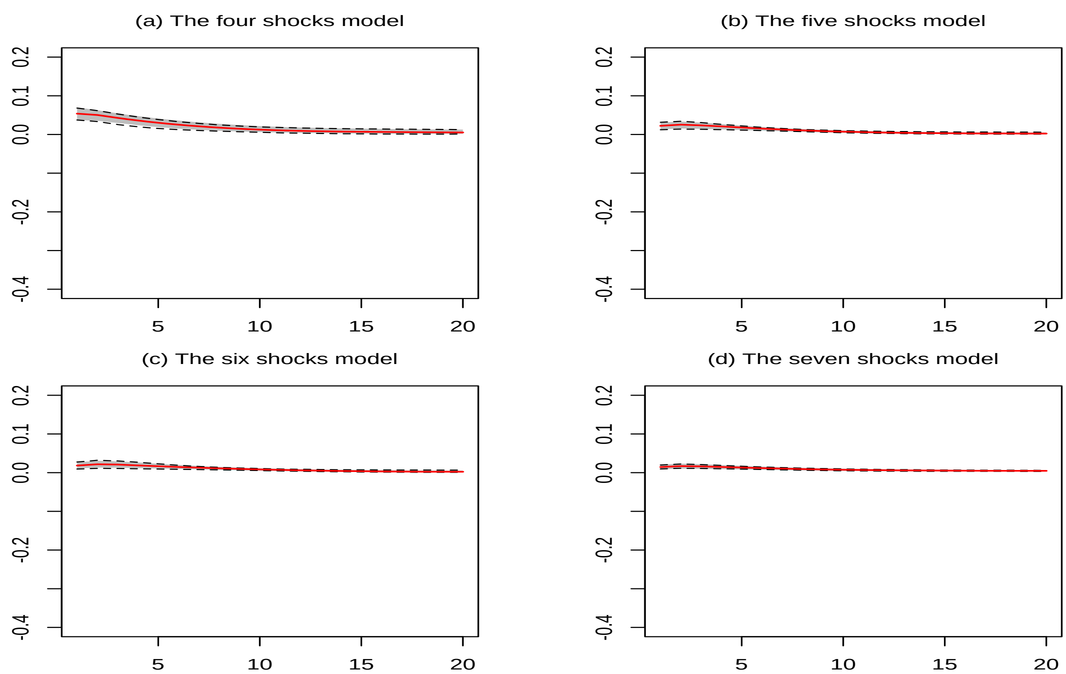


Figure 9. Response of inflation to an exogenous spending shock



Note. See Figure 1. The variables are not annualized.

Figure 10. Response of interest rate to a productivity shock

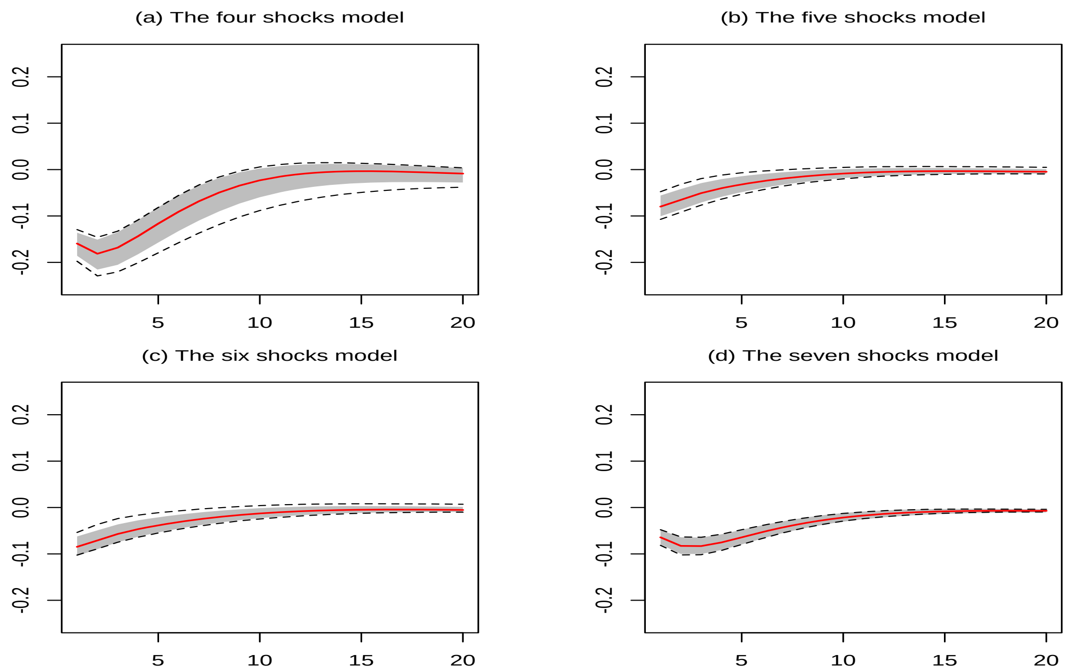
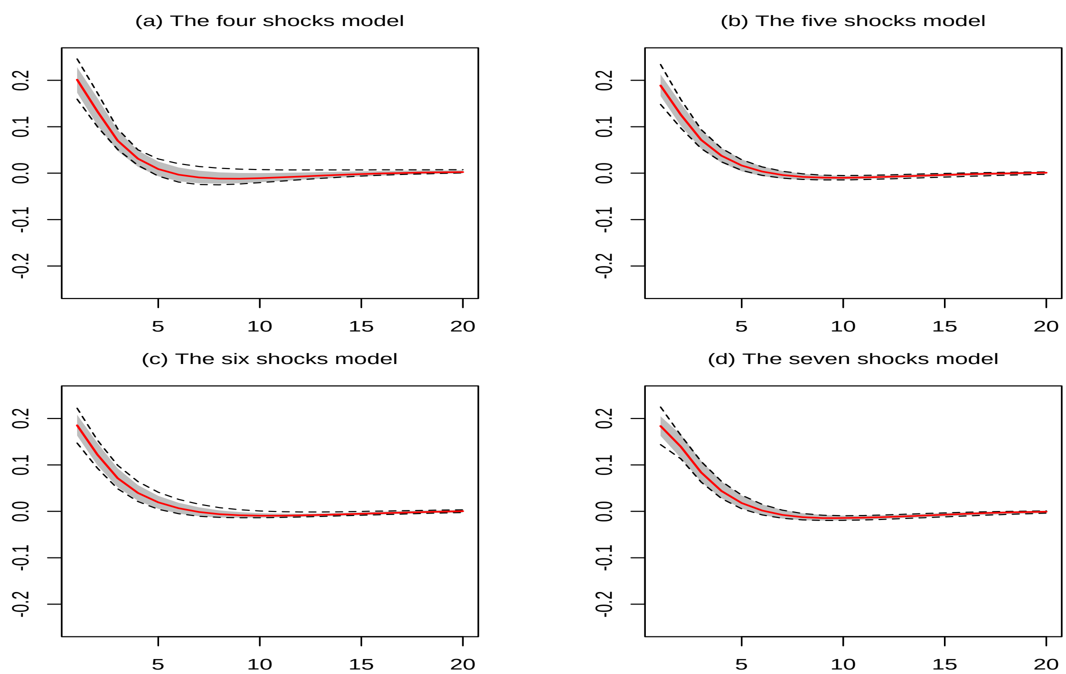


Figure 11. Response of interest rate to a monetary policy shock



Note. See Figure 1. The variables are not annualized.

Figure 12. Response of interest rate to an investment shock

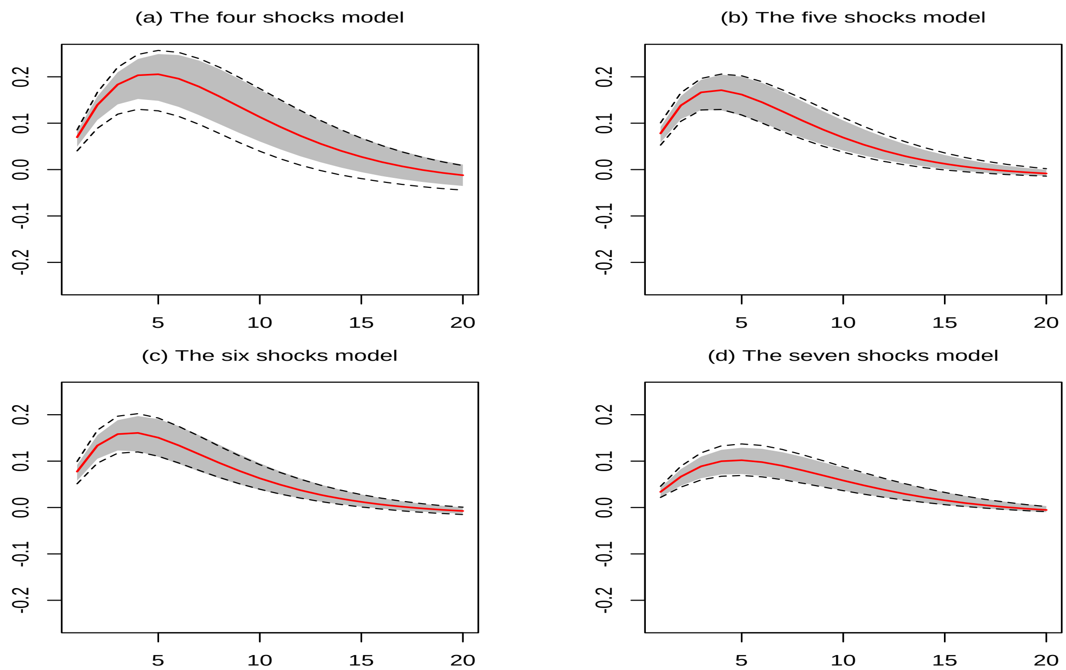
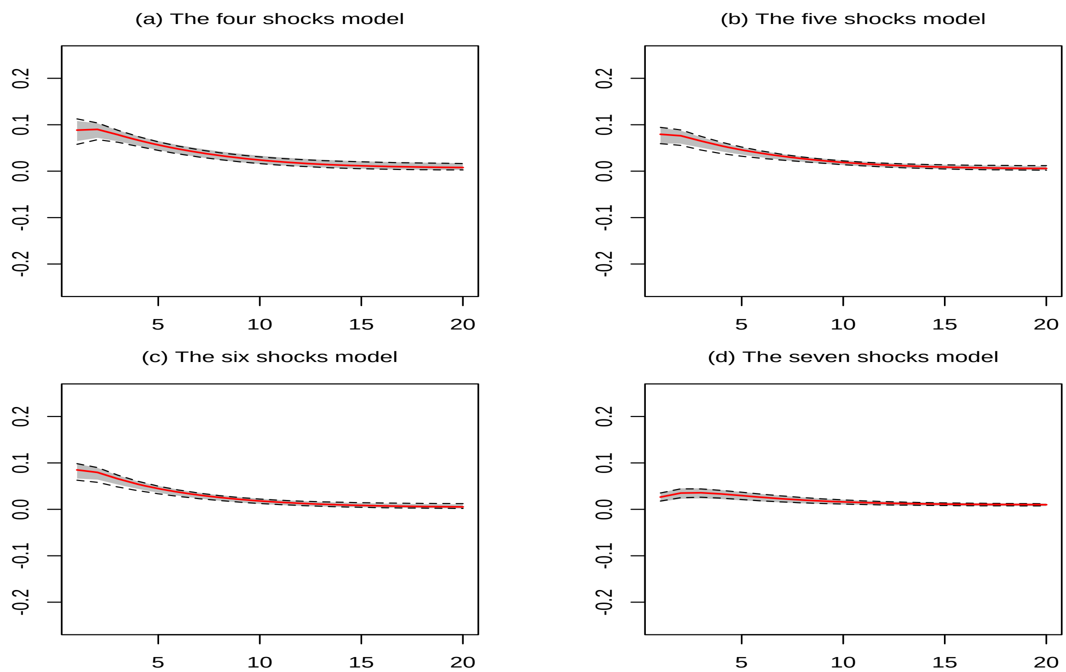


Figure 13. Response of interest rate to an exogenous spending shock



Note. See Figure 1. The variables are not annualized.

Figure 14. Response of investment to a productivity shock

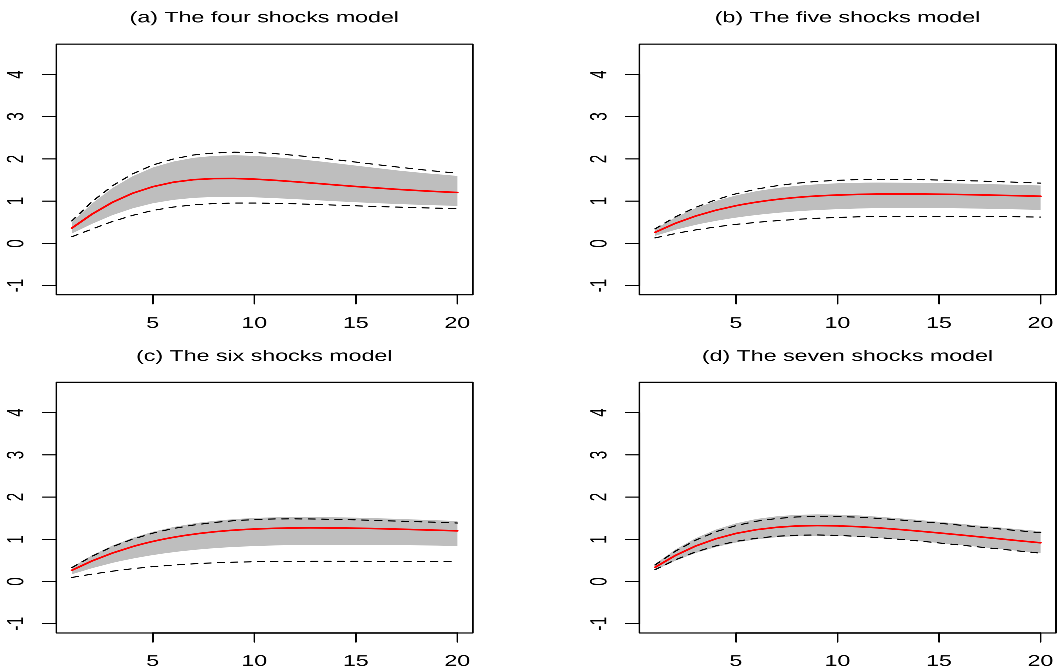
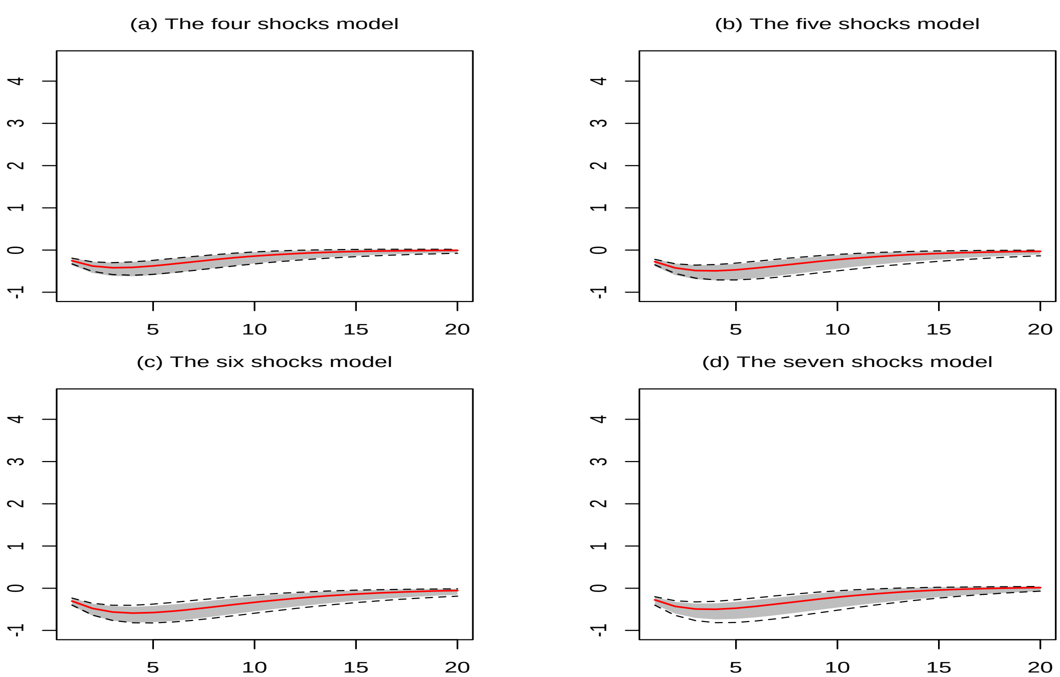


Figure 15. Response of investment to a monetary policy shock



Note. See Figure 1. The variables are not annualized.

Figure 16. Response of investment to an investment shock

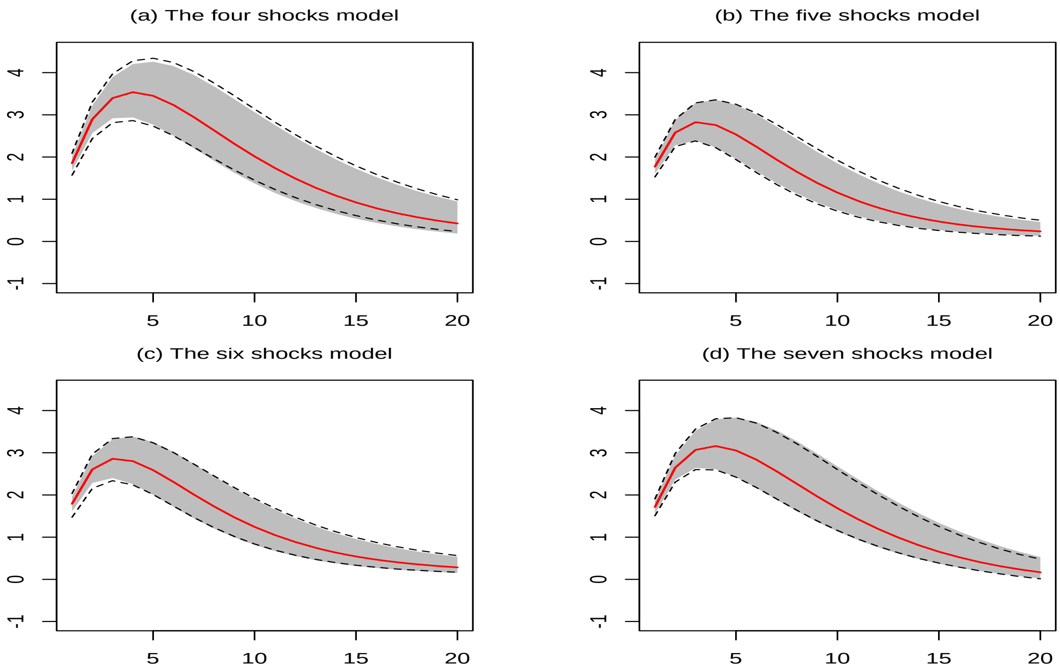
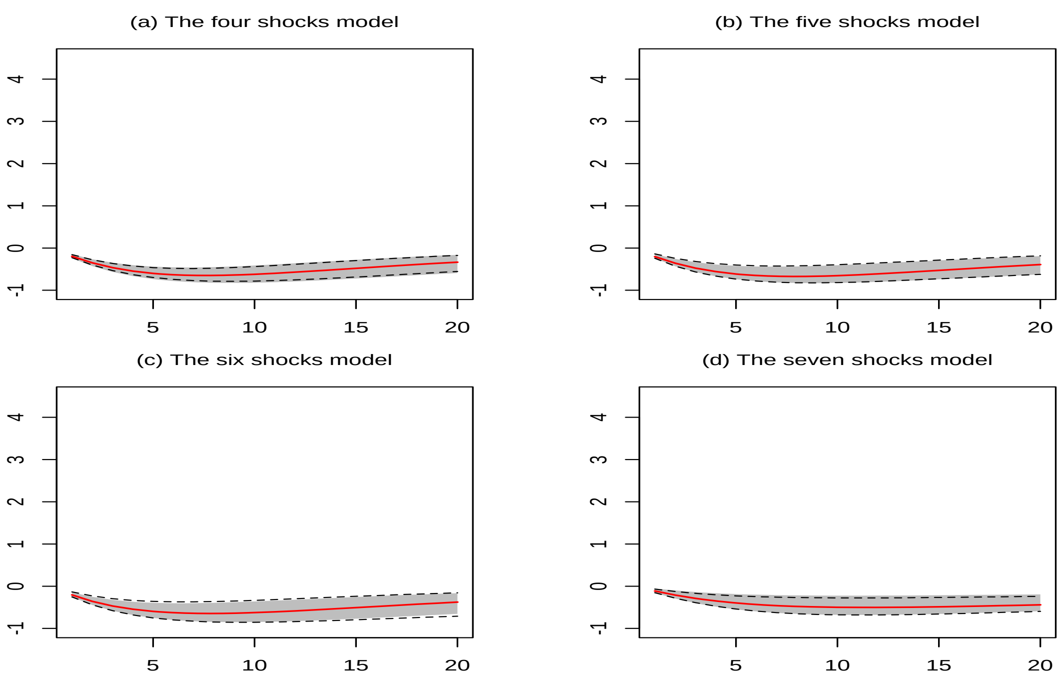


Figure 17. Response of investment to an exogenous spending shock



Note. See Figure 1. The variables are not annualized.

Figure 18. Response of hours worked to a productivity shock

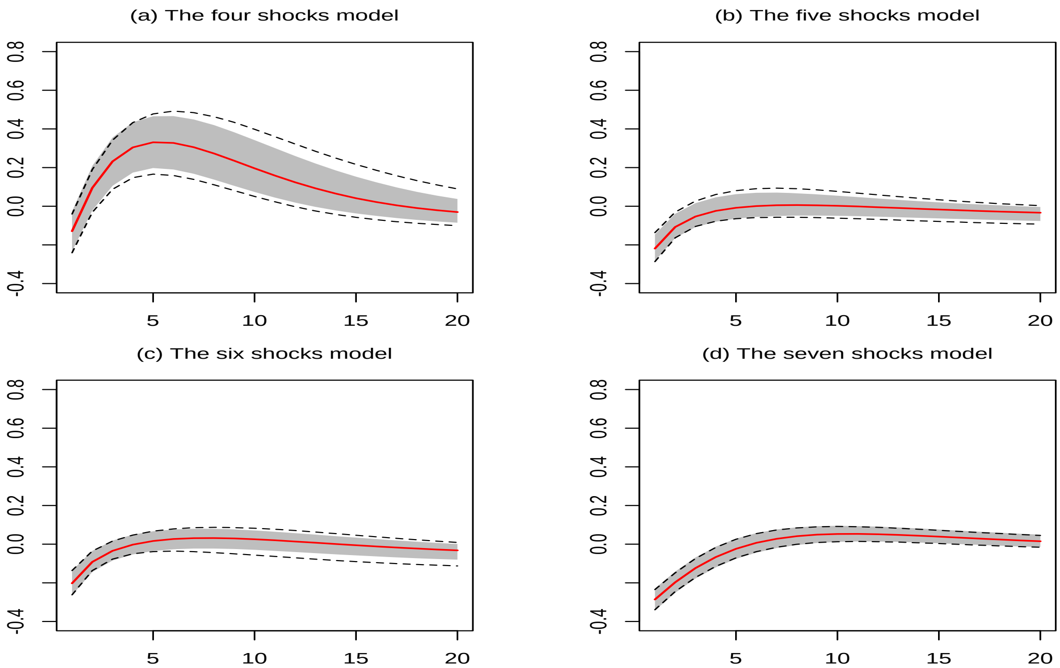
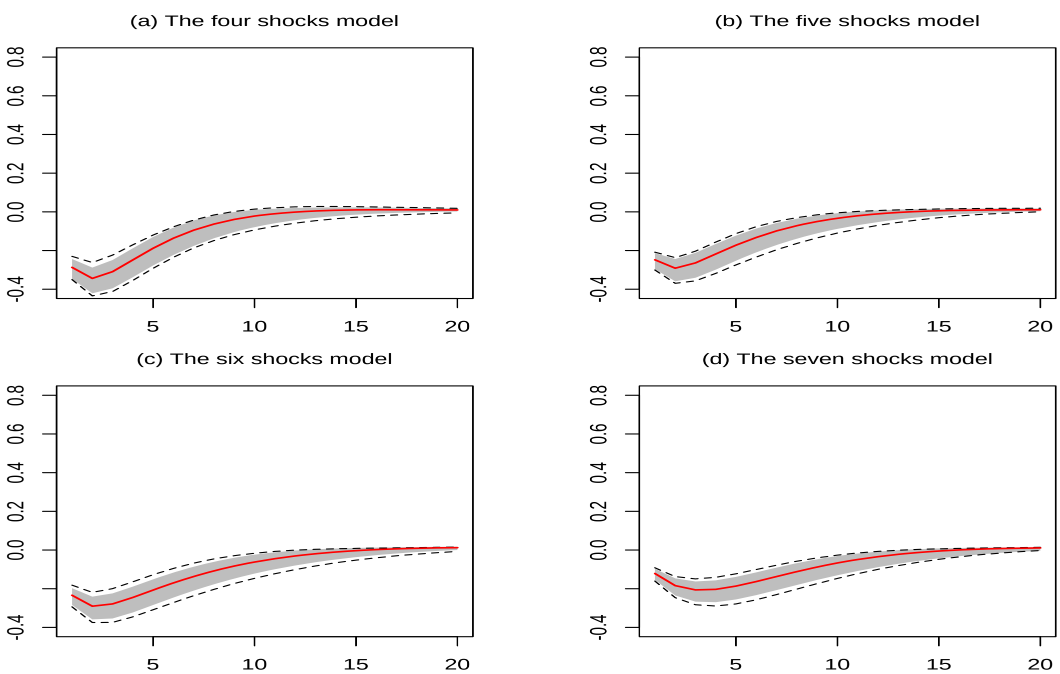


Figure 19. Response of hours worked to a monetary policy shock



Note. See Figure 1. The variables are not annualized.

Figure 20. Response of hours worked to an investment shock

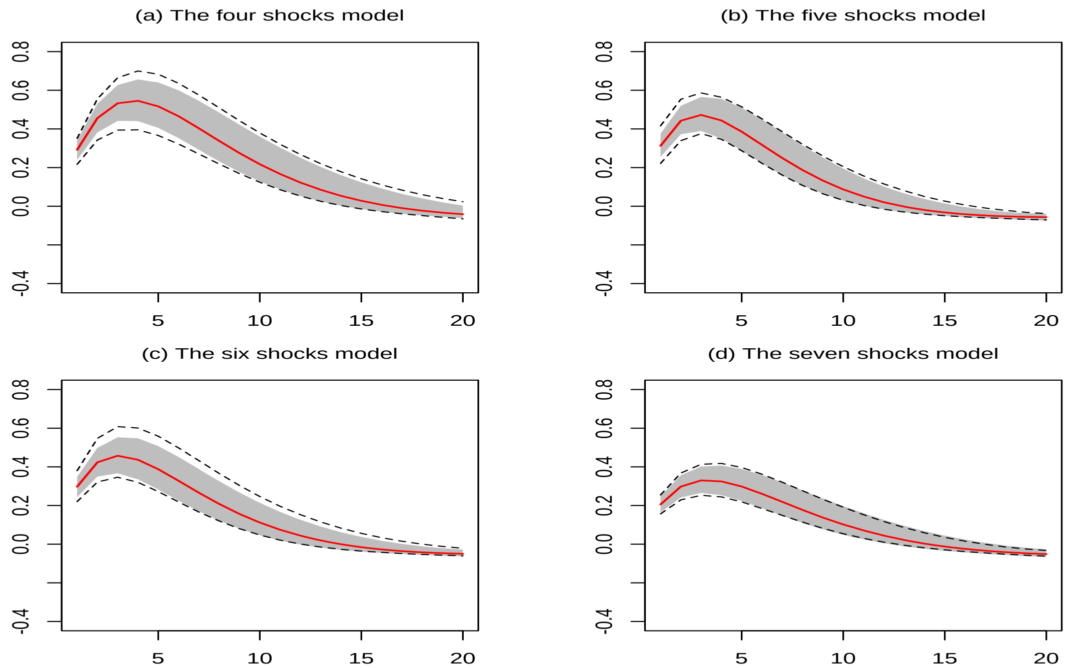
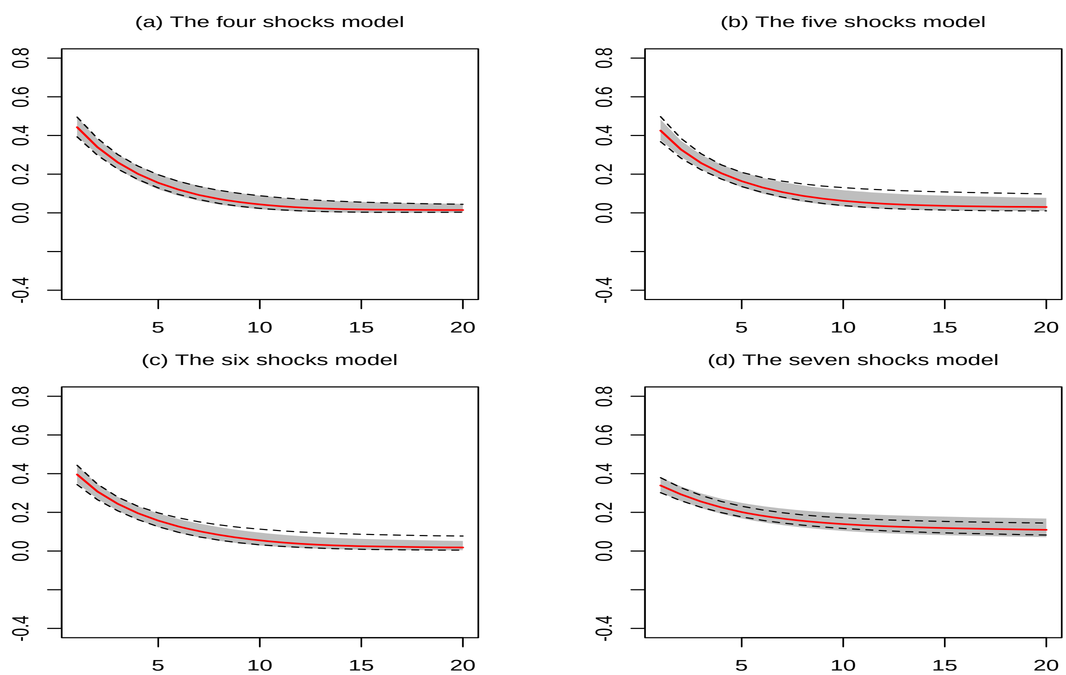


Figure 21. Response of hours worked to an exogenous spending shock



Note. See Figure 1. The variables are not annualized.

Figure 22. Response of wage to a productivity shock

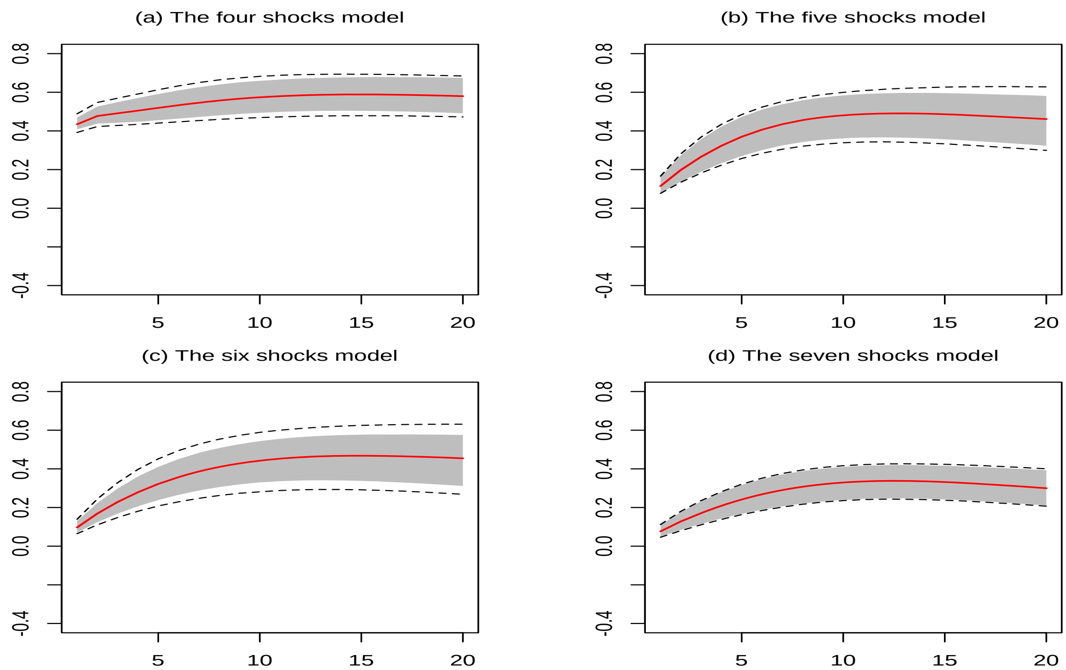
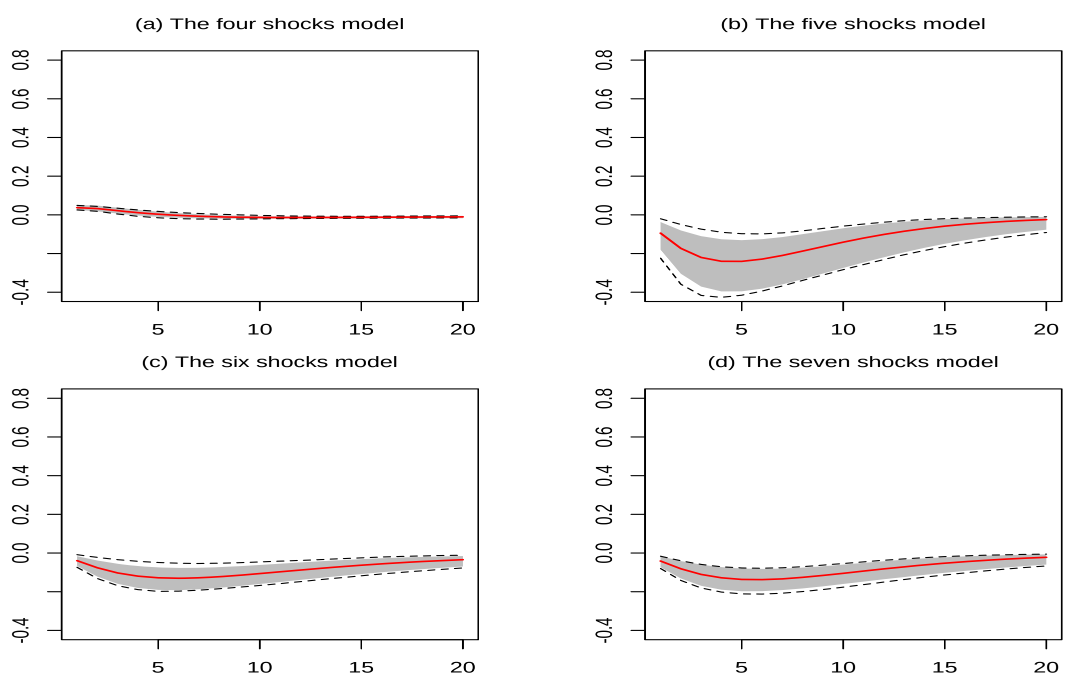


Figure 23. Response of wage to a monetary policy shock



Note. See Figure 1. The variables are not annualized.

Figure 24. Response of wage to an investment shock

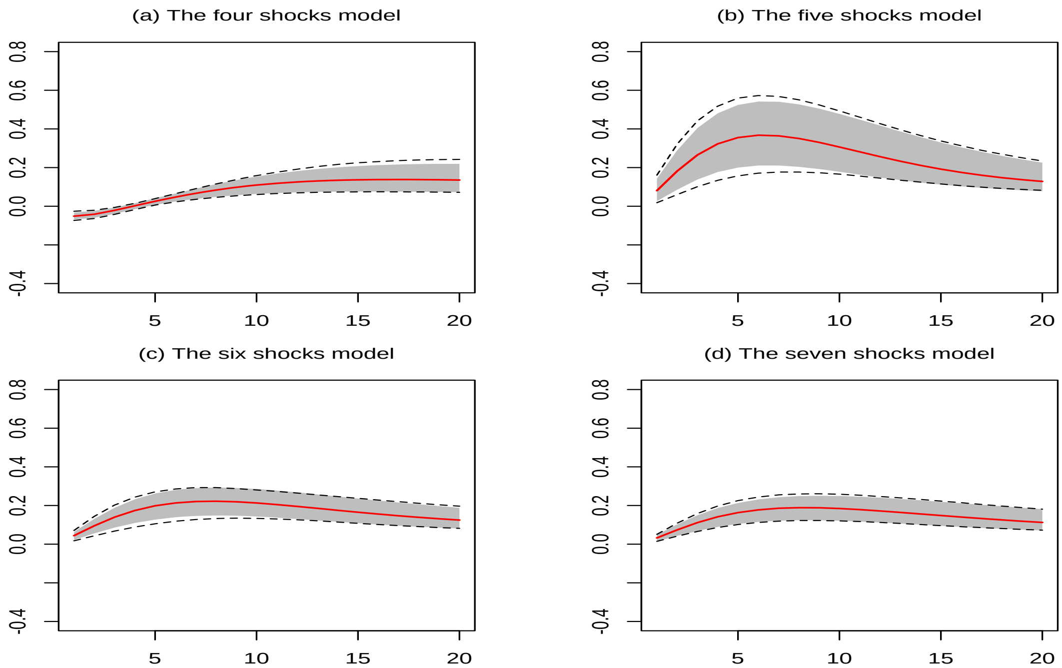
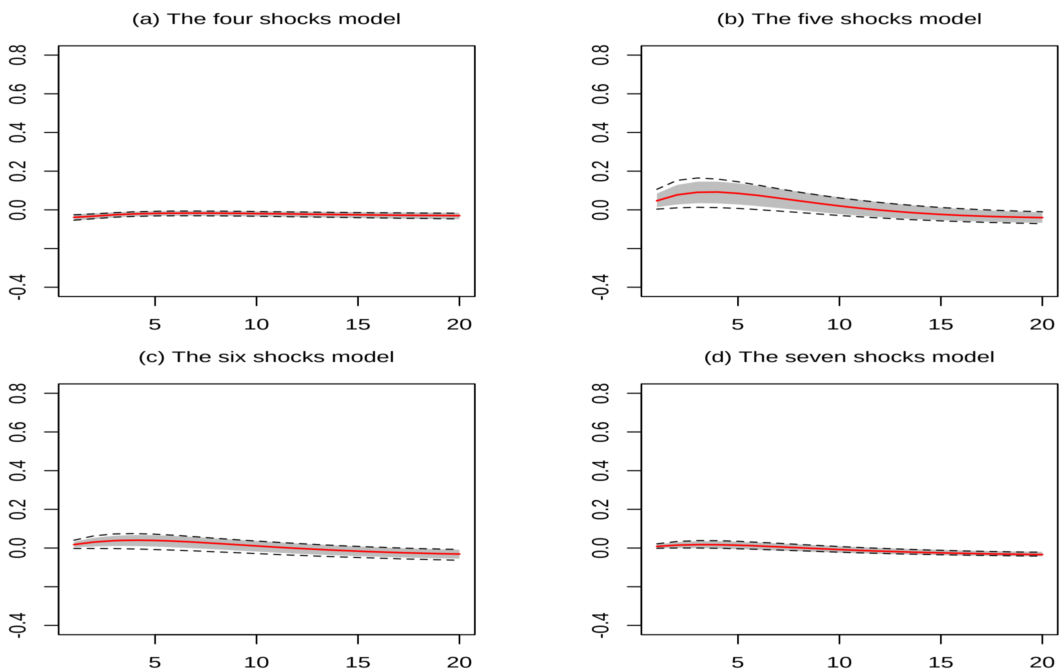


Figure 25. Response of wage to an exogenous spending shock



Note. See Figure 1. The variables are not annualized.

Figure 26. Response of consumption to a productivity shock

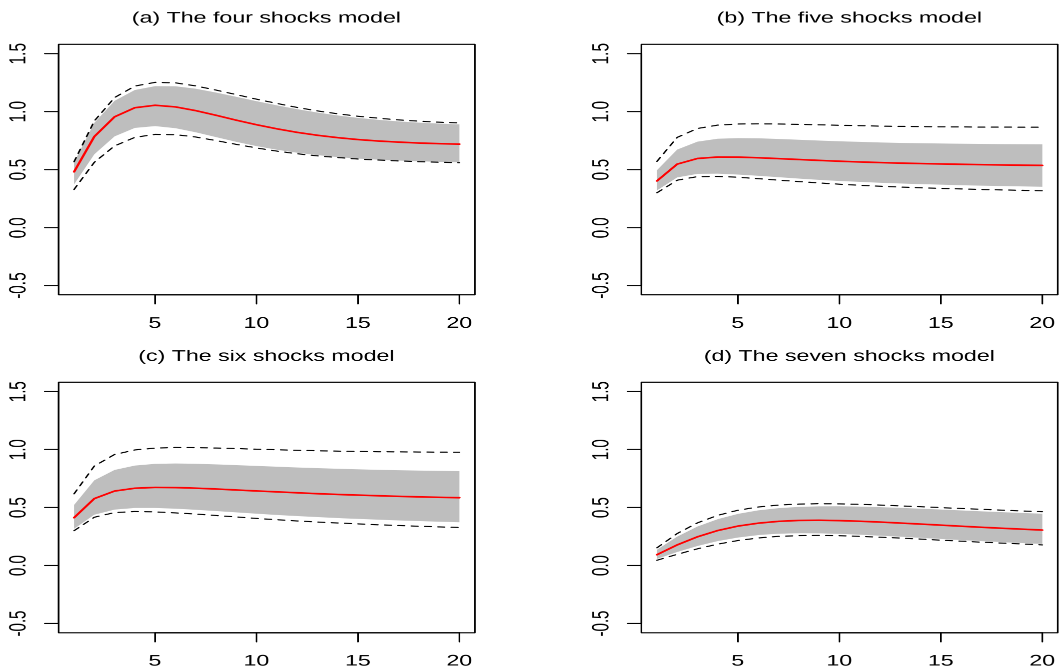
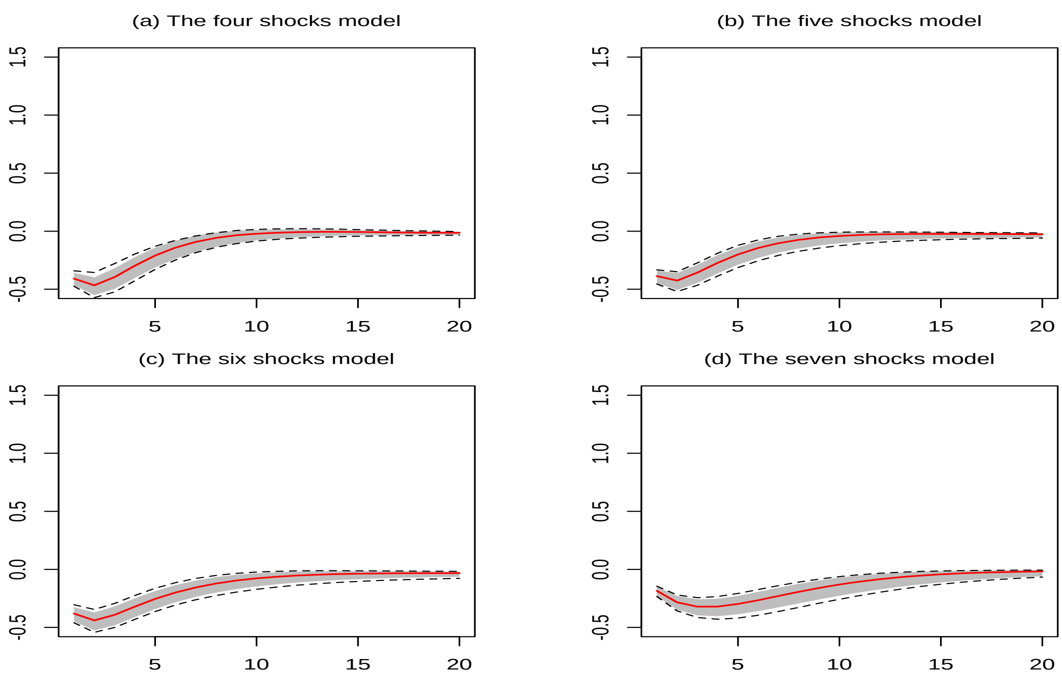


Figure 27. Response of consumption to a monetary policy shock



Note. See Figure 1. The variables are not annualized.

Figure 28. Response of consumption to an investment shock

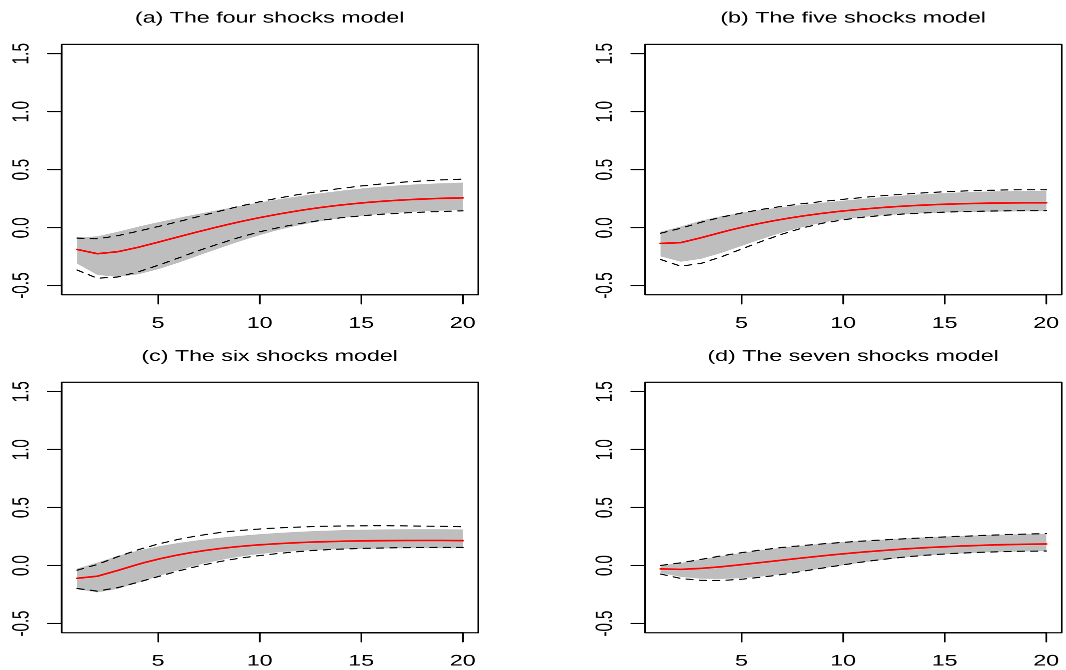
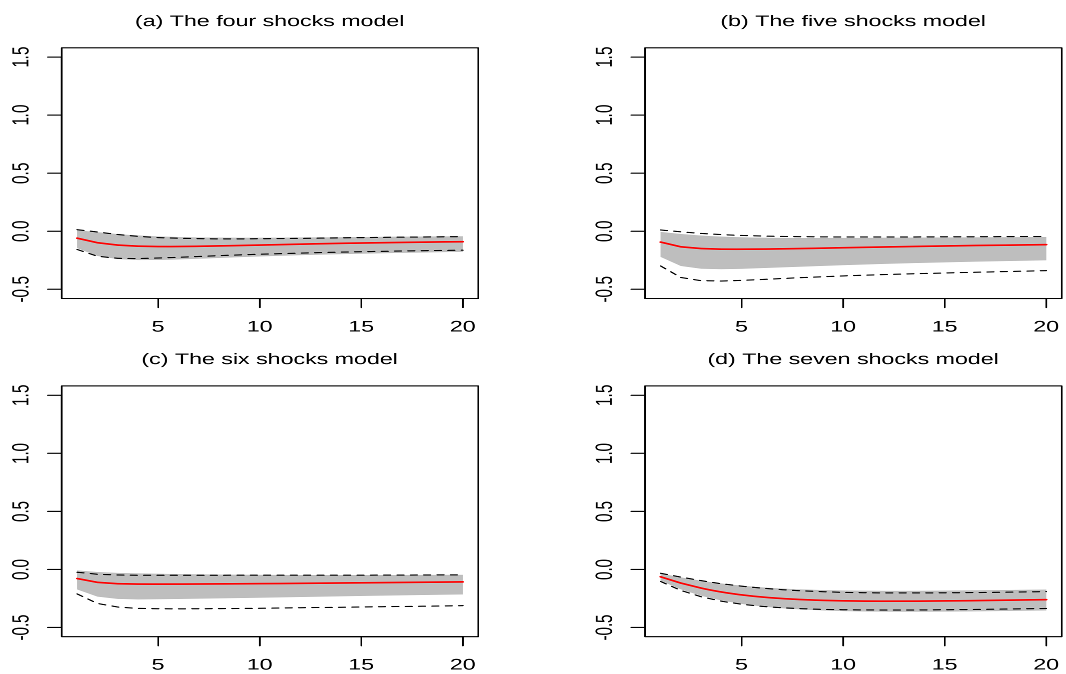


Figure 29. Response of consumption to an exogenous spending shock



Note. See Figure 1. The variables are not annualized.