

A Composite Likelihood Framework for Analyzing Singular DSGE Models*

Zhongjun Qu

Boston University

September 15, 2017

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 nonsingular submodels. Second, it provides an MCMC based procedure for parameter estimation. Third, it delivers confidence sets for structural parameters and 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

*First version: October, 2013. I thank the co-editor Bryan Graham, three anonymous referees, Tim Christensen, Atsushi Inoue, Adam McCloskey, Pierre Perron, Frank Schorfheide, Mototsugu Shintani, Denis Tkachenko, seminar participants at BU, Maryland, Vanderbilt, the 2015 NBER EFSF midyear meeting, and the 11th Econometric Society World Congress for comments and suggestions, and Chuqing Jin, Junwen Lu and Andres Sagner for research assistance. Contact: Department of Economics, Boston University, Boston, MA, 02215 (qu@bu.edu).

as a special case when the model is nonsingular. 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.

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, see Sargent (1989), 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. Studies have documented that different choices of observables can have large impacts on 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 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. 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 determine the number of observables used for estimation.

This paper develops a likelihood based framework for analyzing DSGE models, which does not require 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 component likelihoods, each of which corresponds to a marginal or conditional event. It has found applications in diverse areas, particularly in spatial statistics, where complex dependence between variables makes implementing the full likelihood impractical. Here, the issue of complex dependence 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 components. 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 framework consists of 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 data to assess their implications. In addition, for any intermediate model, different composite likelihoods can be constructed and estimated using different sets of submodels. This can potentially reveal shortcomings of the model, therefore being informative about what additional shocks are desirable for model improvement. These features are illustrated through 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 New Keynesian model, studied in Clarida, Gali and Gertler (2000) and Lubik and Schorfheide (2004). The second is the model of Smets and Wouters (2007). 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 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 effect 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 effect 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 to some shocks but very different responses to others. This reflects an identification issue, suggesting 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 and the empirical analysis, the following perspective is fundamental. That is, a DSGE model is an approximation to the true data generating process with stochastic singularity being among the potential misspecifications. This perspective suggests that, as with other misspecifications, 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. The value of the framework is not in providing a unique estimation criterion function that achieves the highest efficiency, but rather in allowing researchers to experiment with different combinations of component likelihoods and to confront all such choices with 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).

This paper is 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, important progresses 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). The above methods are not likelihood based, i.e., they use criteria other than *model implied* densities to link the model with the data.

The paper proceeds as follows. Section 2 characterizes stochastic singularity. Section 3 introduces the composite likelihood. Sections 4 to 7 study identification, inference, impulse responses and forecasting, respectively. Section 8 provides empirical illustrations. Section 9 discusses how to choose between composite likelihoods. The online appendix contains proofs, notes on implementation, and some additional empirical results.

2 Stochastically singular DSGE models

This paper considers DSGE models that are representable as

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

The n -by-1 vector Y_t includes the measured variables, X_t is a vector 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 $\mu(\theta)$, $A(\theta)$, $B(\theta)$, $C(\theta)$, and $D(\theta)$ are functions of θ . Throughout the paper, θ is assumed to take values in a parameter space Θ that is 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 equations can be appended to the system to describe the time evolution of v_t . The representation also encompasses indeterminacy once ε_t and θ are augmented to include the sunspot shocks and the corresponding parameters. Such an extension follows from Proposition 1 in Lubik and Schorfheide (2004).

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

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

where $H(L; \theta) = [C(\theta)(I - A(\theta)L)^{-1}B(\theta), D(\theta)]$ and $\epsilon_t = [\varepsilon_t', v_t']'$. This representation is useful for formulating the theoretical results on identification and inference. The spectral density matrix of Y_t at frequency $\omega \in [-\pi, \pi]$ is $f(\omega; \theta) = \frac{1}{2\pi}H(\exp(-i\omega); \theta)\Sigma(\theta)H(\exp(-i\omega); \theta)^*$, where $\Sigma(\theta) = \text{Var}(\epsilon_t)$ and the superscript “*” stands for the conjugate transpose. The next definition specifies the type of stochastic singularity considered in this paper.

Definition 1 *The 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 variable 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 $\text{Var}(Y_t)$ can still be of full rank. For example, consider $Y_{1,t} = \epsilon_{t-1}$ and $Y_{2,t} = \epsilon_t$. Then, $Y_{1,t}$ is known from observing $Y_{2,t-1}$, although the covariance matrix of Y_t still has full rank. Consequently, the appropriate object to study singularity in the time domain is the entire correlogram. In the frequency domain, (1) is stochastically singular if and only if the spectral density matrix of Y_t is of reduced rank at all frequencies.

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 (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 its lagged values is not well defined because the conditional covariance matrix is singular. Algorithmically, when implementing the Kalman filter, the prediction step produces a singular covariance matrix, causing the updating step to break down. In the frequency domain, the spectral density matrix of Y_t is singular. Because its inverse enters the likelihood, the latter also fails to be well defined.

3 The composite likelihood

The composite likelihood was developed by Lindsay (1988). Its precedents are the pseudo-likelihood of Besag (1974, 1975) and the partial likelihood of Cox (1975). Below, I review it 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 ($i = 1, \dots, N$). Suppose the distribution of y_i conditional on the neighboring observations is given by $(y_i|y_{[i]}) \sim N(\theta w_i' y, \tau^2)$, where $y = (y_1, \dots, y_N)'$ and w_i is an N -by-1 vector whose j -th element equals 1 if i and j are neighbors and 0 otherwise. The Hammersley-Clifford theorem implies that the joint distribution of y is unique and given by $y \sim N(0, \sigma^2 (I - \theta W)^{-1})$, where $W = [w_1, \dots, w_N]$ and σ^2 is a function of τ^2 and θ . Assume σ^2 is known and equals 1. Then, the log likelihood (up to a constant) equals $\kappa(\theta) + \theta y' W y / 2$ with $\kappa(\theta) = (1/2) \log \det (I - \theta W)$. Maximizing this likelihood involves computing $\kappa(\theta)$ and 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 to consider the sum of the conditional log likelihoods:

$$\ell(\theta) = \sum_{i=1}^N \log f(y_i|y_{[i]}; \theta). \quad (3)$$

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

$$\ell(\theta) = \sum_{i=1}^{N-1} \sum_{r=i+1}^N \log f(y_i, y_r|d_{ir}; \theta), \quad (4)$$

where d_{ir} is some measure of the relationship between the two sites. Both (3) and (4) are members of the composite likelihood family. The general principle was laid out in Lindsay (1988): one starts with a set of conditional (e.g., (3)) or marginal (e.g., (4)) 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; see the review 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 (4), involving pairs of assets in the portfolio.

I now relate the composite likelihood to 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

$$Y_{s,t} = P_s \mu(\theta) + P_s C(\theta) X_t + P_s D(\theta) v_t \quad \text{with} \quad X_t = A(\theta) X_{t-1} + B(\theta) \varepsilon_t.$$

Its vector moving average representation is

$$Y_{s,t} = P_s \mu(\theta) + P_s H(L; \theta) \varepsilon_t. \tag{5}$$

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

$$f_s(\omega; \theta) = \frac{1}{2\pi} P_s H(\exp(-i\omega); \theta) \Sigma(\theta) H(\exp(-i\omega); \theta)^* P_s^* = P_s f(\omega; \theta) P_s^*. \tag{6}$$

The relationship (5) can be called a submodel because it carries all the restrictions of the full model, but imposes them only on a subset of the observables.

Definition 2 *The submodel (5) is a nonsingular submodel if it is stochastically nonsingular for all $\theta \in \Theta$. Further, it is a maximal nonsingular submodel if augmenting $Y_{s,t}$ with any variable from Y_t will always make it 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 Kalman filter. In the frequency domain, the inverses of the spectral densities of the nonsingular submodels can be obtained directly from (6). The computational details are in the online 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 (5) 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 can be desirable to match only the nonsingular relationships 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.

Let $Y_{1,t}, \dots, Y_{S,t}$ be some subvectors of Y_t that are stochastically nonsingular, each satisfying (5) for some P_s with S being some integer. Denote their corresponding log likelihood functions by $\ell_s(\theta)$ ($s = 1, \dots, S$). I propose to construct the composite log likelihood as:

$$\ell(\theta) = \sum_{s=1}^S \ell_s(\theta). \quad (7)$$

The above construction has two features. First, it allows for an arbitrary relationship between the number of observables and number of shocks. Second, if the original full model is already nonsingular, then by choosing $S = 1$ and $Y_{S,t} = Y_t$ we obtain the conventional log likelihood. Therefore, the framework encompasses the conventional likelihood analysis as a special case.

Remark 1 *As is clear from Lindsay (1988)'s general principle, the composite likelihood is not unique. It depends on the submodels that the researcher considers. There are two potential approaches to 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 for estimating θ . (Here, high asymptotic efficiency means that the estimator's asymptotic distribution has a low dispersion around the (pseudo) true parameter value. A common measure for it is the asymptotic mean squared error.) However, this can be problematic here because misspecification is clearly present. Under misspecification, the pseudo true parameter value can change when a different set of submodels is considered. Comparing the dispersions alone no longer has a sound theoretical justification. The second approach 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 is the approach that this paper will*

adopt. 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. In practice, the choice of which moments to use is usually guided by what the model intends to capture. It rarely involves only the consideration of estimation efficiency.

Remark 2 *In the empirical illustrations, for each singular model I always start the analysis with the following specification. I choose the first subset, $Y_{1,t}$, to correspond to a maximal nonsingular submodel. This implies that I subject the model to capturing the joint dynamic properties of this vector. Then, I set $Y_{2,t}, \dots, Y_{S,t}$ to be singleton subsets such that their union includes all the remaining variables in Y_t . This implies that I 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 singularity. 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, I 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 ones as special cases.*

The composite likelihood framework allows the researcher to introduce weights on each of the submodels. One can even think about bringing together two different (singular) models and estimating them jointly. Such explorations are left for future work.

3.1 Illustrations

Illustrative example 1. This example leads to analytical results. It also shows that identification failure can occur when excluding variables from the estimation.

Let x_t and c_t be a household's income and consumption. The researcher postulates the following model:

$$x_t = \rho x_{t-1} + e_t, \quad c_t = \gamma x_t. \quad (8)$$

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 consumption. Suppose $|\rho| < 1$ and $e_t \sim i.i.d.N(0, \sigma^2)$. The goal is to estimate ρ, γ and σ .

The model-implied covariance matrix of (x_t, c_t) is singular. A common practice is to use only one variable for estimation. 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 .

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

$$\text{For } x_t : \ell_1(\rho, \gamma, \sigma) = -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (x_t - \rho x_{t-1})^2,$$

$$\text{For } c_t : \ell_2(\rho, \gamma, \sigma) = -\frac{T}{2} \log \sigma^2 - \frac{T}{2} \log \gamma^2 - \frac{1}{2\gamma^2\sigma^2} \sum_{t=1}^T (c_t - \rho c_{t-1})^2.$$

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 parameters are identified from considering the composite likelihood.

The maximizer of the composite likelihood satisfies the following relationship:

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

where $\hat{\rho}$ mirrors an OLS estimator after reweighting the observations from the two equations by their residual variances. As the specification error becomes smaller (i.e., the variance of v_t approaches 0), $\hat{\gamma}^2$ approaches its true value γ in any sample size, and $\hat{\rho}$ and $\hat{\sigma}^2$ approach the conventional MLE under a known γ . Therefore, 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 DSGE model studied in Clarida, Gali and Gertler (2000) and Lubik and Schorfheide (2004). The original model is

$$\begin{aligned}
y_t &= E_t y_{t+1} - \tau(r_t - E_t \pi_{t+1}) + g_t, \\
\pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - z_t), \\
r_t &= \rho_r r_{t-1} + (1 - \rho_r)\psi_1 \pi_t + (1 - \rho_r)\psi_2(y_t - z_t) + \varepsilon_{rt}, \\
g_t &= \rho_g g_{t-1} + \varepsilon_{gt}, \quad z_t = \rho_z z_{t-1} + \varepsilon_{zt},
\end{aligned} \tag{9}$$

where y_t, π_t , and r_t denote log deviations of output, inflation and nominal interest rate. The shocks satisfy $\varepsilon_{rt} \sim \text{i.i.d.}N(0, \sigma_r^2)$, $\varepsilon_{gt} \sim \text{i.i.d.}N(0, \sigma_g^2)$, $\varepsilon_{zt} \sim \text{i.i.d.}N(0, \sigma_z^2)$; ε_{gt} and ε_{zt} are correlated with correlation coefficient ρ_{gz} . The observables are log levels of output, inflation and interest rate: $Y_t = (0, \pi^*, \pi^* + r^*)' + (y_t, 4\pi_t, 4r_t)'$, where the output is pre-filtered and π^* and r^* are annualized steady-state inflation and real interest rates with $\beta = (1 + r^*/100)^{-1/4}$. Let $\theta = (\tau, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r, \sigma_g, \sigma_z, \rho_{gz}, \pi^*, r^*)'$. Henceforth, this model will be referred to as the three shocks model. Here I consider two singular versions of the three shocks model. The first is affected by ε_{zt} only. The second is affected by ε_{gt} and ε_{zt} .

The solutions to the three models are related in a simple way. Consider first the three shocks model. System (9) can be written as (Sims, 2002): $\Gamma_0(\theta)X_t = \Gamma_1(\theta)X_{t-1} + \Psi(\theta)\varepsilon_t + \Pi(\theta)\eta_t$, 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

$$Y_t = \mu(\theta) + C(\theta)X_t \quad \text{with} \quad X_t = A(\theta)X_{t-1} + B(\theta)\varepsilon_t, \tag{10}$$

where $A(\theta)$ and $B(\theta)$ are known functions of structural parameters and $C(\theta)$ is a selection matrix that selects the first three elements of X_t . The solutions to the two singular models can still be represented as (10) 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 $(0, \varepsilon_{gt}, \varepsilon_{zt})'$. 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, the solutions to the three models can be written as $Y_t = \mu(\theta) + C(\theta)X_t$ with $X_t = A(\theta)X_{t-1} + B(\theta)M_i\varepsilon_t$. The VMA representation is $Y_t = \mu(\theta) + C(\theta)[I - A(\theta)L]^{-1}B(\theta)M_i\varepsilon_t$.

Now consider estimation. The component likelihood corresponding to $\{y_t\}$ can be obtained by defining $P_s = [1 \ 0 \ 0]$ and consider $P_s Y_t = P_s \mu(\theta) + P_s C(\theta) X_t$ with $X_t = A(\theta)X_{t-1} + B(\theta)M_i\varepsilon_t$. In the time domain, the likelihood function can be computed using Kalman filtering and, in the frequency domain, by computing its spectral density and periodograms (see the online appendix for details). Other component likelihoods can be computed in the same way by changing P_s accordingly.

The composite likelihood does not use all the information implied by the model when estimating the parameters. Therefore, it can be considered as a limited-information method. It is informative to compare it with minimum distance estimation. The main distinction is that the former uses component likelihood functions to connect the model with the data, while the latter uses moments. This leads to three differences in practice. First, the main estimation steps are different. For the composite likelihood, the most important step is to specify the submodels, after which the estimation becomes fairly similar to the conventional likelihood situation (see Section 5). For minimum distance estimation, the key step is to select a sufficient number of informative moments, and then compute them using both the data and the model. Simulation is often needed for computing the model implied moments. Secondly, the handling of the time series information is different. Because the component likelihoods are proper likelihood functions, they automatically incorporate the time series information implied by the submodels. For minimum distance estimation, such information will need to be handled with suitably chosen moments and a proper weighting matrix. For medium scale models with more than 30 parameters, choosing such moments and weighting

matrix can be a challenging or even daunting task. Thirdly, the implications for forecasting are different. The composite likelihood can be used to obtain forecasts for all the observables irrespective the number of shocks (see Section 7). The minimum distance framework does not encompass forecasting. Meanwhile, the two methods also share some common features. They both require making some choices in either submodels or moments. The computational cost of both methods can be substantial. The computation details related to the composite likelihood will be documented in two ways. First, the Matlab codes are made available on line. Second, the computational time is summarized in Section 9. A more thorough and informative discussion of the minimum distance method for DSGE models can be found in Fernández-Villaverde, Rubio-Ramírez and Schorfheide (2016).

4 Identification

For DSGE models, understanding identification is important for both calibration and formal statistical analysis. Substantial progress has been made recently. Canova and Sala (2009) documented the types of identification issues that can arise in these models. Iskrev (2010) gave sufficient conditions for local identification, while Komunjer and Ng (2011) and Qu and Tkachenko (2012) gave necessary and sufficient conditions. Guerron-Quintana, Inoue and Kilian (2013) are among the first to study weak identification in DSGE models. This section studies the local identification of θ based on the information provided by the submodels. Suppose that Y_t is generated by (1) with $\theta = \theta_0$. I 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 θ is locally identifiable at θ_0 from the first and second order properties of $Y_{s,t}$ ($s = 1, \dots, S$) 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 $s = 1, \dots, S$ 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 autocov-

variance function $\Gamma_s(j; \theta)$ ($j = 0, \pm 1, \dots$) and that $f_s(\omega; \theta)$ is continuous in ω . There is a one-to-one mapping between $\Gamma_s(j; \theta)$ and $f_s(\omega; \theta)$, 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 $f_s(\omega; \theta)$ ($s = 1, \dots, S$) if and only if it is locally identifiable from $\mu_s(\theta)$ and the complete set of autocovariances $\{\Gamma_s(j; \theta)\}_{j=-\infty}^{\infty}$.

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

Assumption 2. 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 $h_j(\theta)$ are defined in $H(L; \theta) = \sum_{j=0}^{\infty} h_j(\theta) L^j$; (ii) The elements of $f(\omega; \theta)$ belong to $\text{Lip}(\beta)$ of $\beta > \frac{1}{2}$ with respect to ω ; (iii) The elements of $f(\omega; \theta)$ are continuously differentiable in θ with $\|\partial \text{vec } f(\omega; \theta) / \partial \theta'\| \leq C < \infty$.

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

$$G_S(\theta) = \sum_{s=1}^S \left\{ \int_{-\pi}^{\pi} \left(\frac{\partial \text{vec } f_s(\omega; \theta)}{\partial \theta'} \right)^* \left(\frac{\partial \text{vec } f_s(\omega; \theta)}{\partial \theta'} \right) d\omega + \frac{\partial \mu_s(\theta)'}{\partial \theta} \frac{\partial \mu_s(\theta)}{\partial \theta'} \right\} \quad (11)$$

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.

Theorem 1 follows from Theorem 2 in Qu and Tkachenko (2012). The main difference is that here the identifying information comes from the nonsingular submodels.

The dimension of $G_S(\theta)$ always equals q . The s -th component in the summation, i.e.,

$$\int_{-\pi}^{\pi} \left(\frac{\partial \text{vec } f_s(\omega; \theta)}{\partial \theta'} \right)^* \left(\frac{\partial \text{vec } f_s(\omega; \theta)}{\partial \theta'} \right) d\omega + \frac{\partial \mu_s(\theta)'}{\partial \theta} \frac{\partial \mu_s(\theta)}{\partial \theta'} \quad (12)$$

measures the contribution from the s -th submodel to identification. It is positive semidefinite by construction. In practice, it is instructive to compare the rank of (12) for $s = 1, \dots, S$, as this can be informative about the source of identification.

I now discuss three extensions. First, to check local identification based on the second order properties only, we only need to delete the term in (11) containing $\partial \mu_s(\theta)' / \partial \theta$. Second, to consider identification based on a subset of frequencies, we replace $G_S(\theta)$ by

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

where $W(\omega)$ is an indicator function that is symmetric around zero. Third, to check the local identification of a subset of parameters, say $\theta^{(1)}$, while fixing the rest at θ_0 , we only need to replace $\partial\theta'$ by $\partial\theta^{(1)'}$. Theorem 1 continues to hold with these changes.

5 Estimation and inference

The estimation can proceed as if $\ell(\theta)$ was the conventional log likelihood. Let $\pi(\theta)$ be a prior density. Then, as in Chernozhukov and Hong (2003), a quasi-posterior density is given by:

$$p(\theta) = \frac{\pi(\theta) \exp(\ell(\theta))}{\int_{\Theta} \pi(\theta) \exp(\ell(\theta)) d\theta}. \quad (13)$$

The estimator 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 $(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(B)})$ corresponding to the density kernel $\pi(\theta) \exp(\ell(\theta))$ and computing $\hat{\theta} = B^{-1} \sum_{j=1}^B \theta^{(j)}$. A useful reference is An and Schorfheide (2007). I refer to the intervals obtained by sorting the MCMC draws as “**MCMC Intervals**”.

The density $p(\theta)$ corresponds to a quasi-posterior because $\ell(\theta)$ is not a log likelihood function. The MCMC Intervals are in general not valid credible sets. They can also differ from frequentist confidence intervals asymptotically. The intuition is as follows. In a regular problem with a large sample size, the length of an MCMC interval is determined by the curvature of the criterion function. For a conventional likelihood, this curvature measures information in the data because of the information matrix identity. For a general criterion function, this curvature no longer represents information except under special circumstances (c.f. discussion on p.223-224 of Lindsay, 1988). Therefore, the MCMC interval alone may not correctly represent uncertainty. Chernozhukov and Hong (2003) have clearly documented this feature in a general context. They also provided intervals with desired frequentist properties.

Misspecification is another complication. When misspecification is present and unaccounted for, standard Bayesian credible sets can differ substantially from valid frequentist

intervals even asymptotically. Müller (2013) studied the risk of Bayesian inference under misspecified models. His results imply that standard credible sets can have higher asymptotic frequentist risk than achievable. To relate such general results to the current study, the online appendix gives two simple examples. There, overconfidence can arise if one applies Bayesian credible sets mechanically without thinking about misspecification. All existing DSGE models exhibit some misspecifications. Therefore accounting for misspecification is of general importance. This issue does not vanish when a DSGE model is analyzed using the composite likelihood instead of the conventional likelihood.

Motivated by the above considerations, below I address the inference issue in two steps. First, I construct confidence intervals with the following two features. (1) They acknowledge model misspecification. (2) They have correct frequentist coverage rates for the pseudo true value asymptotically when the parameters are well identified. I refer to the resulting intervals as “**Asymptotic Intervals**”. Then, I further compare the Asymptotic and the MCMC intervals. This will lead to a suggestion for practice.

Assumption MI. The 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 and serially uncorrelated with finite variance Σ_0 and zero 3rd and 4th order cumulants. 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 model while Assumption MI is about the data. The data can be stochastically nonsingular. The requirements on the 3rd and 4th order cumulants can be relaxed. Doing so will not affect the consistency result (i.e., Lemma 2) but will alter the asymptotic distribution. The procedure for constructing the Asymptotic Intervals (Procedure A) is valid without these cumulant 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 $\ell_{s,\infty}(\theta)$ equals

$$-\frac{1}{4\pi} \left\{ \int_{-\pi}^{\pi} [\log \det(f_s(\omega; \theta)) + \text{tr}(f_s^{-1}(\omega; \theta) f_{s,0}(\omega))] d\omega + (\mu_{s,0} - \mu_s(\theta))' f_s^{-1}(0; \theta) (\mu_{s,0} - \mu_s(\theta)) \right\}.$$

The one-sided Hausdorff distance from a set A to B is given by $h(A, B) = \sup_{x \in A} \inf_{y \in B} \|x - y\|$.

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{C} denote the set of maximizers of $\ell(\theta)$ and C_0 the set of maximizers of $\ell_\infty(\theta)$, then with probability one: $\limsup_{T \rightarrow \infty} \check{C} \subseteq C_0$, i.e., $h(\check{C}, C_0) \rightarrow 0$.
3. Further, if $\ell_\infty(\theta)$ has a unique maximum at θ_0 , then $\check{\theta} \xrightarrow{a.s.} \theta_0$.

The proof of the lemma follows Hansen and Sargent (1993, p.49-53). 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 C_0 depends on model specification. If the model is correctly specified, then C_0 consists of all the parameter values satisfying $f_s(\cdot; \theta) = f_{s,0}(\cdot)$ and $\mu_{s,0} = \mu_s(\theta)$ for $s = 1, \dots, S$. If it is misspecified, then such values do not exist, and C_0 should be interpreted as containing pseudo true values. 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.

Theorem 2 *Suppose θ_0 is the unique maximizer of $\ell_\infty(\theta)$. Let $\hat{\theta}$ denote the mean or mode computed from (13). Then, under Assumptions 1,2 and MI, $\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow^d N(0, M^{-1}VM^{-1})$, 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*

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

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 .

As shown in the online appendix, the result applies to both the time and frequency domain composite likelihoods. Because $M^{-1}VM^{-1}$ is in general different from M^{-1} , the MCMC draws will need to be adjusted in order for their quantiles to provide asymptotically valid confidence intervals. Such an adjustment is given in the following procedure.

Procedure A (for computing Asymptotic Intervals): First, compute $\sqrt{T}(\theta^{(j)} - \hat{\theta})$ with $\hat{\theta}$ being the quasi posterior mean or mode of (13). Use their sample covariance as an estimator for M^{-1} and denote it by \hat{M}^{-1} . Next, compute $\sum_{s=1}^S \partial \ell_{s,t}(\hat{\theta}) / \partial \theta$ ($t = 1, \dots, T$). 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 $\tilde{\theta}^{(j)} = \hat{\theta} + \hat{M}^{-1} \hat{V}^{1/2} \hat{M}^{1/2} (\theta^{(j)} - \hat{\theta})$. Finally, use the quantiles of $\tilde{\theta}^{(j)}$ to form confidence intervals.

Some misspecification 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 may be needed for V . 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 dependence is present. The standard sample covariance matrix can be used if the dependence is considered small.

This concludes the first step. Now I compare the Asymptotic and MCMC Intervals. The Asymptotic Interval is interpretable under misspecification, but requires the parameters to be well identified. The MCMC interval in general does not have a sharp interpretation. But because it does not rely on any asymptotic approximation, it can be more indicative of weak identification than the Asymptotic Interval. I recommend to report both intervals in practice. This recommendation can be related to that in Moon and Schorfheide (2012). Their paper analyzed models with partially identified parameters. They suggested to report the estimates of the identified set and the conditional prior along with the Bayesian credible sets. In both cases, the motivation for the recommendation can be stated as providing a full disclosure of the results when different methods can lead to different conclusions.

6 Impulse response functions

I discuss how to compute impulse responses and to measure the associated uncertainty. Suppose there are no measurement errors. Then, $Y_t = \mu(\theta) + C(\theta)(I - A(\theta)L)^{-1}B(\theta)\varepsilon_t$. Let \mathbf{e}_j be the j -th column of an identity matrix. Then, the impulse response of the j -th variable to the l -th orthogonal shock equals $\mathbf{e}_j'IR(k; \theta)\mathbf{e}_l$, where $IR(k; \theta) = C(\theta)A(\theta)^k B(\theta)\Sigma^{1/2}(\theta)$.

The inference 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. Asymptotic Intervals can be constructed simply by replacing $\theta^{(i)}$ with $\tilde{\theta}^{(i)}$ in Steps 2 and 3.

7 Forecasting

For nonsingular models, one- and multi-step ahead forecasts can be obtained through

$$p(Y_{T+1}|Y_{1:T}) = \int p(Y_{T+1}|Y_{1:T}; \theta)p(\theta|Y_{1:T})d\theta, \quad (14)$$

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. 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, this algorithm is no longer applicable under singularity. In practice, the forecasting often proceeds by introducing measurement errors or by treating some observables as unobserved. The latter approach ignores the information from some observed time series and 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. Specifically, let

$$p_S(Y_{T+1}|Y_{1:T}) = \int p_S(Y_{T+1}|Y_{1:T}; \theta)p_S(\theta|Y_{1:T})d\theta, \quad (15)$$

where $p_S(\theta|Y_{1:T})$ equals $p(\theta)$ in (13) and

$$p_S(Y_{T+1}|Y_{1:T}; \theta) \propto \prod_{s=1}^S p_s(Y_{s,T+1}|Y_{s,1:T}; \theta). \quad (16)$$

Note that $p_s(\cdot|Y_{s,1:T}; \theta)$ is the conditional density of $Y_{s,T+1}$ implied by the s -th submodel, which can be evaluated using the Kalman filter. Using (15) in place of (14) leads to the following forecasting procedure (let $\theta^{(i)}$ denote the MCMC draws from (13)):

Step 1. Sample from $p_S(\cdot|Y_{1:T}; \theta^{(i)})$ in (16) for $i = 1, \dots, B$. Denote the values by $Y_{T+1}^{(i)}$.

Step 2. Compute $\hat{Y}_{T+1} = B^{-1} \sum_{i=1}^B Y_{T+1}^{(i)}$ and use \hat{Y}_{T+1} as the point forecast for Y_{T+1} .

Step 3. To produce multi-step forecasts, 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.

I now apply the above procedure to the first example in Subsection 3.1. In (16), $S = 2$, $Y_{1,t} = x_t$ and $Y_{2,t} = c_t$. Conditional on θ , the joint forecast for Y_{T+1} is a bivariate normal distribution with mean $(\rho x_T, \rho c_T)$ and a diagonal covariance matrix. Two features emerge. First, the forecast for $Y_{1,T+1}$ conditional on $Y_{2,T+1}$ is the same as the unconditional forecast. Therefore, by construction, the information contained in $Y_{2,T+1}$ cannot be used to improve the forecast for $Y_{1,T+1}$. This holds in general situations, as long as the variables are allocated to disjoint submodels. Secondly, the two point forecasts, ρx_T and ρc_T , can be correlated because $\text{corr}(\rho x_T, \rho c_T) = \text{corr}(x_t, c_t)$. They are perfectly correlated if and only if x_t and c_t are perfectly correlated in the data. This suggests that the point forecasts can reflect the cross dependence present in the data, even when the variables are in disjoint submodels. These two features are important for interpreting the forecasts under (20).

8 Empirical illustrations

I apply the composite likelihood framework to analyze both small and medium scale DSGE models. I take the perspective of a modeler who starts with few shocks and then gradually incorporates more shocks to enrich the model. For each singular model, I first study the parameter estimates and then the impulse responses. The results for the medium scale models

(i.e., singular versions of the Smets and Wouters 2007 model) are reported below, while those for the small scale models are in Section S.5 of the online appendix.

In Smets and Wouters (2007) (henceforth SW), the model consists of seven observables: output (y_t), consumption (c_t), investment (i_t), wage (w_t), hours (l_t), inflation (π_t) and interest rate (r_t). It has seven shocks: 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). For reference, I include the equations of the model in Section S.2 and an annotated list of the parameters and their estimates in Table S1 in the online appendix.

Three singular models are considered. The first model has four shocks: $\eta_t^a, \eta_t^g, \eta_t^r$ and η_t^i . The second model includes η_t^p as an additional shock. The third model includes also η_t^w . The same prior distributions and parameter bounds as in SW are used. The only exception is for ν_p , whose lower bound is reduced from 0.5 to 0.1, so that it is not binding when computing the posterior modes. The sample period is 1965:I to 2004:IV.

8.1 The four shocks model

I use $\{y_t, \pi_t, r_t, i_t\}, \{c_t\}, \{w_t\}$ and $\{l_t\}$ to form the composite likelihood. This is based on the following considerations. First, capturing the joint behavior of y_t, π_t and r_t 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 η_t^a, η_t^g and η_t^r . Second, the incorporation of η_t^i into the model permits including i_t in the leading subset. Third, 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.

Parameter estimates. The results are reported in Table 1. Out of the 28 parameters, 21 of them have their confidence intervals (i.e., the unions of the MCMC Interval and the Asymptotic Interval) overlapping with those in SW (reported in Table S1 in the online appendix). Among the remaining 7 parameters $(\xi_p, \rho_a, \lambda, \alpha, r_\pi, \sigma_r, \rho_g), (\xi_p, \rho_a, \lambda)$ take on

quite different values from those in SW. Below I discuss the findings in more detail. All the values are the posterior means unless stated otherwise.

The model's steady state values are similar to those in SW. This reflects a benefit of keeping all the variables in the estimation: if some observables (such as l_t) are not used in the analysis, some steady state parameters (such as \bar{l}) can become unidentified.

Among the shock processes, as in SW, the productivity and exogenous spending processes are found to be persistent while the investment and the monetary policy shock processes are not. The productivity process has an AR(1) coefficient (ρ_a) of 0.99 and a standard deviation parameter (σ_a) of 0.55, higher than the original estimates of 0.95 and 0.45. Under $\rho_a = 0.99$, the half life of a shock equals 68 quarters, much higher than 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 (σ_g) of 0.54, compared with 0.97 and 0.53 in SW. Finally, the AR(1) coefficients for the investment and monetary policy shock processes (ρ_i and ρ_r) are very close to those in SW, while the standard deviation estimates (σ_i and σ_r) are both mildly higher. In summary, the most pronounced difference here pertains to the productivity process.

Now consider the behavioral parameters. The habit parameter (λ) equals 0.37, substantially smaller than 0.71 in SW. The price indexation and rigidity parameters (ι_p and ξ_p) both take on small values. The new ξ_p implies an average price contract of 1.3 quarters, compared with 2.9 quarters in SW. The wage indexation and rigidity parameters (ι_w and ξ_w) are both high. The new ξ_w 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 broadly similar to those in SW. In summary, among the behavioral parameters, those governing habits and price and wage frictions are consistently different from those in the nonsingular model. As shown below, this translates into markedly different responses to productivity shocks.

Next, consider the monetary policy parameters. The inflation weight parameter (r_π) equals 1.41, lower than 2.04 in SW. The output gap weight parameter (r_y) equals 0.17, higher

than 0.08 in SW. 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.

Impulse responses. The responses to a productivity shock are reported in Figures 1-7. The remaining cases are in the appendix (Figures S1 to S21). Each figure consists of 4 subfigures: the first three are for models with 4 to 6 shocks, while the fourth is for the original SW model. The solid line is computed using the posterior mean. The two dashed lines and the shaded area correspond to the 90% Asymptotic and MCMC Intervals respectively.

Figures 1-7 confirm that the productivity shock is prominent in driving business cycle fluctuations. Under a positive shock, the inflation falls sharply (due to the small price inertia), causing real wage to rise sharply (due to the high wage indexation) and real interest to fall (due to the monetary policy reaction). On the real side, the labor supply increases strongly as a result of the higher wage. This leads to a sharp rise in output, accompanied by a strong increase in consumption (due to the low habit persistence and lower real interest rate). Because the productivity shock process is very persistent, its effects on these variables are long lasting. The above responses are substantially more pronounced than in SW.

The responses to the monetary policy shock are close to those in SW (Figures S1-S7). Specifically, the response of inflation is only slightly stronger than in SW. The responses of interest rate and investment are almost identical to those in SW. There is a slight increase in the real wage initially, as opposed to the small decrease seen in SW. This is due to the drop in the price level and the high wage indexation. The initial responses of the output, hours worked and consumption are slightly stronger than in SW. Then, they revert to levels similar to the latter. The initial differences are due to the low habit persistence.

Consider the investment shock (Figures S8-S14). The inflation and interest rate responses are mildly stronger than in SW due to the low price rigidity. The response in real wage shows

an initial slight dip before reverting to levels comparable to SW. This dip follows from the decrease in inflation and the strong wage indexation. The responses in output, investment, hours and consumption are all comparable to those in SW, except that the last two are mildly stronger due to the stronger response in interest rate and the low habit persistence.

Finally, consider the exogenous spending shock (Figures S15-S21). The responses of inflation and interest rate are mildly stronger due to the low price rigidity. Real wage shows a slight decline at short horizons, due to the strong wage indexation. The responses of output, investment, hours worked and consumption are close to SW, except that they exhibit a faster decay to zero due to the smaller AR(1) coefficient of this shock process.

Therefore, while the responses to the monetary shock are similar to those in SW, the responses to the productivity shock are substantially different. The price rigidity parameter is estimated to be small 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.

8.2 The five shocks model

I incorporate η_t^p into the model. This shock breaks the rigid link between inflation and price rigidity, permitting large and frequent changes in the former to be compatible with a high level of the latter. As a result, the model is more flexible in modeling the responses of inflation, and consequently hours worked and output, to productivity shocks. The following 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.

Parameter estimates. The results are reported in Table 1 (Columns 6 to 9). First, the price rigidity parameter (ξ_p) is now higher than in the four shocks model and closer to that in SW. Second, the persistence of the 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 and output

weight parameters (r_π and r_y) are both close to that in SW. In fact, out of the 31 parameters, all except λ and α now have confidence intervals overlapping with those in SW. These two parameter values are little changed relative to the four shocks model.

Impulse responses. See plot (b) of Figures 1-7 and S1-S21. Under a positive productivity shock, the decrease in inflation is much smaller than the four shocks model. The magnitude is close to SW. The increase in real wage is also smaller than the four shocks model. It is still mildly stronger than SW because the productivity process remains more persistent. Since the inflation response is much reduced, the high wage indexation is no longer quantitatively important for determining real wage. Consequently, the response in hours worked is much less pronounced than in the four shocks model.

Responses to the other three shocks are in line with SW. The inflation response is almost identical to SW due to the increased price rigidity. The initial small increase in the real wage seen in the four shocks model is no longer present. The responses of hours worked, output and consumption are slightly stronger than those in SW due to the low habit persistence.

In summary, incorporating the price markup shock leads to a significant increase in the price rigidity and decrease in the persistence of the productivity process. These two changes both lead to milder responses to productivity shocks. Overall, the responses to the five shocks are no longer substantially different from those in SW.

8.3 The six shocks model

I incorporate the wage markup shock (η_t^w) into the model. 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 reported in the last four columns of Table 1. Out of the 34 parameters, all except 2 (λ and α) have confidence intervals overlapping with those in SW; the value of ι_w is now close to that in SW.

Impulse responses. The responses are overall similar to the five shocks model. They are also close to those in SW, except that the initial responses in hours worked, output and consumption are slightly different due to the low habit persistence.

This section has considered models with the number of shocks ranging between 1 and 7. A summary of the main findings can be found in the introduction section of this paper.

9 Experimenting with alternative specifications

This section illustrates how one may decide on which composite likelihood to use. The main idea is to consider a range of alternative specifications, examine their implications in terms of parameter estimates, cross covariances, identification properties, and forecasting performance, and then make the choice by aggregating the results. The discussions use singular versions of SW as illustrating examples. For each of them, three composite likelihood specifications are considered. The first is the specification that leads to Tables 1 and S1 (call it the default specification from now on). The other two are as follows. In the first alternative specification, the variables excluded from the maximal nonsingular submodel under the default specification are used to form the second submodel. The two submodels are then used to form the composite likelihood. In the second alternative specification, the composite likelihood is formed by taking the marginal likelihoods of the seven variables, one at a time. In terms of accounting for dependencies in the data, the default specification can be viewed as an intermediate case and the two alternative specifications as two polar cases. Other specifications in between can also be considered without further conceptual difficulties. For the current application, comparing these three specifications is sufficiently informative. The analysis below examines the four implications outlined above sequentially and also makes an attempt to develop a statistical criterion for selection. The goal of the analysis is not to dictate a particular choice, but to inform how a choice can be made. Therefore, in the end of this section, the results are also summarized to provide a practical procedure for choosing a composite likelihood specification.

9.1 Parameter estimates and cross covariances

This subsection considers models with four, five and six shocks sequentially. For each model, it first studies parameter estimates and then cross covariances. The cross covariances are computed using the full (singular) model. They are used to identify the dimensions along which the fit to the data differs, and also to measure the extent of the differences. If a composite likelihood specification produces a fit that is substantially inferior to the default specification, then this specification will be considered undesirable.

The four shocks model. The estimation results under the first alternative specification are reported in Panel (a) in Table S2. The main findings are as follows. Out of the 28 cases, 25 confidence intervals overlap with those under the default specification. The three exceptions are ρ_g, ρ_{ga} and ι_w . In addition, the value of r_y is smaller, although the confidence intervals still overlap. Overall, the parameter estimates under these two specifications are broadly similar. This suggests that the resulting fit to the data may also be similar.

To study this further, I compare three sets of cross covariances with each other. The first two are model-implied covariances at the posterior means. The third is the sample cross covariance computed directly from the data. The 28 pairs of variables are organized into three figures as follows. They are reported in the online appendix. Figure S22 displays the pairs involving Δy_t . These cases are arguably of leading importance. Figure S23 displays results among $\Delta c_t, \Delta w_t$ and Δl_t . Because these three variables are treated differently under the two specifications, separating them out can help reveal patterns in the results. The remaining pairs are included in Figure S24. The maximum lead and lag orders are set to 8 throughout. The 90% confidence intervals derived from sample cross covariances are also included to gauge the uncertainty.

The figures show the following. First, the two model-implied covariances are close to each other in the majority of the cases. Secondly, there is a trade-off. Moving from the default to the first alternative specification improves the fit along some dimensions, but worsens it along

some others. (See (e)-(f) in Figure S23 and (m)-(o) in Figure S24 for the first situation. These pairs all involve Δc_t . See (e)-(g) in Figure S22 for the second situation. They all belong to the maximal nonsingular submodel.) The differences, however, tend to be fairly mild. Thirdly, the model-implied covariances often fall within the confidence intervals. Therefore, with only four shocks, the model can already fit the data along multiple dimensions. Finally, when substantial deviations from the intervals occur, they typically involve variables across submodels; see (d) in Figure S22, (e) in Figure S23, and (i), (m), (n), (o) in Figure S24.

The parameter estimates under the second alternative specification are reported in (a) of Table S3. The cross covariances are shown in Figures S22-S24. First, the parameter values and cross covariances both exhibit greater differences from the default specification than the above. Second, the fit is disappointing. This shows clearly that important information in the data is missed when applying this specification.

Therefore, while the first two specifications can both be viable candidates for analyzing the four shocks model, the third is undesirable.

As a note on inference, suppose a researcher wants to report a $100(1 - \alpha)\%$ confidence interval for a parameter. Here, three composite likelihood specifications are considered, which deliver three candidate intervals. If the researcher obtains the three intervals at the level $100(1 - \alpha/3)\%$, then, irrespective of which interval he/she chooses to report, it will cover the pseudo true value at the level of $100(1 - \alpha)\%$ or higher. This procedure is conservative.

The five shocks model. The results for the first alternative specification are reported in Table S2 and Figures S25-S27. The posterior means are close to those under the default specification. The confidence intervals all overlap. The two model-implied cross covariances are always in agreement with each other. Substantial deviations from the confidence intervals are limited to variables across submodels, except (b) and (e) in Figure S26. Both cases involve Δw_t . The results for the second alternative specification are in Table S3 and Figures S25-S27. They show the same two patterns as under the four shocks model. Therefore, the conclusion

on the three composite likelihood specifications is the same as in the four shocks case.

The six shocks model. The default and the first alternative specifications coincide. The cross covariances are either inside the intervals, or only show mild deviations from them. Noticeable deviations are again limited to variables across submodels, except (d) in S28 and S30. Both pairs involve Δw_t . The results for the second alternative specification are reported in Panel (c) of Table S3 and Figures S28-S30. The fit continues to be poor.

In summary, for the models considered, the results consistently show that the default and first alternative specifications are similar and that the second alternative specification is undesirable. This finding is not surprising. What is important is the analysis that leads to it. The same analysis can be applied to other models and composite likelihood specifications.

9.2 Identification

Canova, Ferroni and Matthes (2014) suggested that identification should be an important consideration when choosing variables to estimate singular DSGE models. This subsection applies Theorem 1 to compare the identifying power of the three composite likelihood specifications. The matrix $G_S(\theta)$ is computed as follows unless stated otherwise. The derivatives are computed using the symmetric difference quotient with the step size set to 1E-7. The integral is approximated using the Gauss-Legendre quadrature with 1000 evaluation points. The values of θ are set to the posterior means. Because $\bar{\gamma}$, r , $\bar{\pi}$ and \bar{l} are identifiable from the means of Δy_t , r_t , π_t and l_t , it is sufficient to compute only the first term in (11).

For the four shocks model under the default specification, the smallest eigenvalue equals 7.52E-4, above the Matlab default tolerance level of 3.49E-10. This suggests that θ is locally identified at this value. The eigenvalue remains at 7.52E-4 when each of the following changes is made: (1) the five-point method is used to compute the derivative, (2) the step size is set to 1E-6 or 1E-8, or (3) the Gaussian quadrature is replaced by the Riemann sum. Finally, when the number of evaluation points is increased substantially to 10000, the eigenvalue

changes only lightly to 7.53E-4. Therefore, the result is robust to the above changes.

Next, for the four shocks model under the two alternative specifications, the smallest eigenvalues equal 8.73E-4 and 3.87E-6 respectively. The tolerance levels are 3.49E-10 and 8.73E-11. The robustness checks support that θ is locally identified in the two cases. Finally, for the models with five to seven shocks, the results also show that θ is locally identified under the three composite likelihood specifications. The details are omitted.

In summary, the identification analysis alone has not led to information that can distinguish between the three composite likelihood specifications.

9.3 Forecasting

This subsection compares the forecasting performance of the three composite likelihood specifications. The ones with high MSFEs are considered as undesirable. The analysis can be related to Guerron-Quintana (2010). When considering singular models augmented with measurement errors, the paper suggests that forecasting accuracy can be an appealing criterion for making choices between the observables.

I consider the same sample periods as in SW, i.e., 1966:I to 1989:IV for initial estimation and 1990:I to 2004:IV for forecast evaluation. The parameters are reestimated every quarter. The MSFEs for horizons of $k=1, 4, 8$ quarters ahead are reported in Table 2. A relative measure is also reported in parentheses. It is computed as 100 times the difference in the MSFEs between the current and the default specification, divided by \sqrt{k} times the sample standard deviation of the first differenced series. The values exceeding 20% are highlighted in bold. The bold values are all positive, implying that the corresponding MSFEs are all higher than those under the default specification. When cross dependencies are misspecified, imposing them can increase the MSFE. This is important for interpreting the results.

When there are four shocks (see the rows (4,0), (4,1) and (4,2) in Table 2), the default specification performs better than the first alternative specification, particularly at $k = 8$, where the relative measure exceeds 20% for 4 out of the 7 series. The default and the second

alternative specification perform similarly; the relative measure is always below 20%.

When there are five shocks, the results show a different pattern. The default and the first alternative specification are now similar. Meanwhile, the second alternative specification performs substantially worse when forecasting consumption, with the relative measure equal to 7.25%, 31.88% and 59.95% at the three horizons. This follows because the MSFEs under the default and the first alternative specifications have both decreased noticeably, particularly at $k = 8$, while those under the second alternative specification have changed little.

This tendency continues when another shock is added. The default specification now performs notably better than the second alternative specification at $k = 8$. Finally, in the seven shocks model, the forecast for consumption further improves under the default specification. The second alternative specification again shows no improvement.

The comparison shows that the default specification performs the best overall. Combining these results with those in Sections 9.1 and 9.2, the default specification emerges as a preferred choice for analyzing all the models.

9.4 Marginal composite likelihood

The current problem has two special features. First, the submodels considered in Section 9 are non-overlapping. As a result, the composite likelihoods all lead to proper densities. Second, the models are fully parametric and the priors are proper. As a result, the marginal composite likelihoods (henceforth MCL) can be readily computed. These two features suggest that the MCL can be a sensible and feasible criterion for comparing models and composite likelihood specifications.

Let $Y = (Y_1', \dots, Y_T')'$. Let $p(Y|\theta, \mathcal{M}_j, \mathcal{S}_k)$ be the composite likelihood under the specification \mathcal{S}_k ($k = 1, \dots, K$) for the singular model \mathcal{M}_j ($j = 1, \dots, J$). Integrating over the prior distribution $\pi(\theta)$, the MCL is given by

$$p(Y|\mathcal{M}_j, \mathcal{S}_k) = \int p(Y|\theta, \mathcal{M}_j, \mathcal{S}_k)\pi(\theta)d\theta. \quad (17)$$

Let $p(\cdot|\mathcal{M}_0)$ be the true density of Y , where \mathcal{M}_0 stands for the unknown true model. Let θ_0 be the pseudo true value under \mathcal{M}_j and \mathcal{S}_k given in Lemma 2 of the paper. Then, the Kullback-Leibler (KL) divergence from the density $p(\cdot|\theta_0, \mathcal{M}_j, \mathcal{S}_k)$ to $p(\cdot|\mathcal{M}_0)$ is given by

$$d_{KL}((\mathcal{M}_j, \mathcal{S}_k), \mathcal{M}_0) = \int p(Y|\mathcal{M}_0) \log \frac{p(Y|\mathcal{M}_0)}{p(Y|\theta_0, \mathcal{M}_j, \mathcal{S}_k)} dY, \quad (18)$$

which equals zero if and only if $p(\cdot|\theta_0, \mathcal{M}_j, \mathcal{S}_k)$ coincides with $p(\cdot|\mathcal{M}_0)$. As $T \rightarrow \infty$:

$$T^{-1} \log p(Y|\mathcal{M}_j, \mathcal{S}_k) - T^{-1} \log p(Y|\mathcal{M}_0) \xrightarrow{p} -d_{KL}((\mathcal{M}_j, \mathcal{S}_k), \mathcal{M}_0). \quad (19)$$

The result (19) shows that the ranking according to the MCL is consistent with the KL divergence ordering asymptotically. I compute (17) for the composite likelihood specifications considered in Sections 8 and 9. The results are reported in Table 3. First, for the four shocks model, the second alternative specification is ranked the highest, followed by the default and then the first alternative specification. This shows that the four shocks model has substantial difficulty in capturing the dependence between the series. Second, for the five shocks model, the default specification has the highest value, followed by the first and then the second alternative specification. This suggests that model's ability in capturing cross dependence has improved, although it still has difficulty with the dependence between consumption and wage growth. Third, for models with six and seven shocks, the default specification and the first alternative specification coincide. They have higher values than the second alternative specification in both cases. This again shows that the models succeed in capturing some joint dynamics in the data. These findings are in line with those reported in Subsections 9.1 to 9.3 in terms of what the models capture. For example, in the four shock model, because cross dependence is poorly captured, the second specification is favored by the MCL criterion.

9.5 Suggestion for practice

The materials in Sections 8 and 9 can be summarized into the following procedure for choosing a suitable composite likelihood specification:

Step 1. Begin the analysis with a default specification by taking a maximal nonsingular submodel and then the remaining variables one at a time. The leading submodel should be chosen based on what the model is designed to capture.

Step 2. Consider a range of alternative specifications. Although Section 9 has focused on two polar cases, other specifications that maintain the same leading submodel can also be considered. The specifications can be compared with the default specification as in Subsections 9.1 to 9.3. The outcomes can be evaluated as follows: (1) If a specification delivers a substantially inferior fit than the default specification, then this specification can be considered as undesirable. If the fit is in the form of a trade-off, then the choice should be based on what the model is intended to capture. (2) Identification failure should be viewed as a red flag. If it is about some important aspects of the model, then the associated specification can be considered as undesirable. If it is about some relatively minor aspects, then at a minimum, this should be taken into account when interpreting the estimation results. (3) The forecast comparison can help in two ways. First, if a specification leads to a substantially less accurate forecast than the default specification, then the specification can be considered undesirable. Secondly, by observing how the forecast performance changes as more dependencies are imposed, one may learn about whether such dependencies are modeled poorly or adequately. The composite likelihood can then be adjusted to de-emphasize the poorly specified aspects. Depending on the purpose of the DSGE model, the above three criteria can receive different weights when aggregating the results.

Step 3. One may be interested in considering a different leading submodel. If so, then Steps 1 and 2 can be repeated and the outcomes evaluated in the same way as in Step 2.

The computational cost in Steps 1 to 3 is substantial but manageable. The associated computational time is as follows (on a desktop with an 8-core Intel 2.4Ghz processor). It takes 5-7 hours to estimate a model using a given composite likelihood specification (i.e., to obtain a panel of Tables 1 or S2-S4). The computation of cross covariances is immediate. For the

identification analysis, it takes less than 1 hour to compute the $G_S(\theta)$ matrix. The forecasting exercise is more time consuming. It takes 1-2 days to produce the MSFEs in a row of Table 2. However, because different subsamples can be handled independently, the computation time can be substantially reduced if more computer cores are available. Experimentations suggest that the computation time declines almost linearly with the number of cores used.

Finally, the suggested model selection criterion can complement but not substitute for the analysis in Subsections 9.1 to 9.3. If we take a model builder's perspective, then for an intermediate model, we want to learn about its merits and shortcomings in order to improve it. Statistical model selection, such as the one described in Subsection 9.4, offers a ranking of models but does not provide answers to many detailed questions.

10 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 allows for analyses related to parameter identification, estimation, inference and forecasting. Applications to both small and medium scale models show that it can be informative about the similarities and differences between alternative models and also about the sources that generate them. The framework can be further developed. First, extensions to nonlinear models can be possible. Second, multiple singular models can be jointly estimated. Finally, the issue of forecasting with singular models merits further study. The results can shed more light 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.
- 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, 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.

Fernández-Villaverde, J., J. Rubio-Ramírez, and F. Schorfheide (2016): “Solution and Estimation Methods for DSGE models,” *Handbook of Macroeconomics, Volume 2 (forthcoming)*.

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.

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

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

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

- 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.

Table 1: Estimation results for medium scale models

	(a) The four shocks model				(b) The five shocks model				(c) 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]
$\bar{\gamma}$	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]
$\bar{\pi}$	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 Smets and Wouters' (2007) dynare code. MCMC: 90% intervals obtained using the quantiles of the MCMC draws. Asymptotic: 90% intervals obtained using Procedure A. The estimates are based on 200,000 draws.

Table 2. Mean squared forecast errors (MSFEs)

Cases	Output	Inflation	Interest	Hours	Wage	Consumption	Investment
(a) 1 quarter ahead							
(4,0)	0.53	0.24	0.13	0.48	0.62	0.57	1.45
(4,1)	0.51 (-2.33)	0.23 (-3.23)	0.16 (11.54)	0.54 (9.23)	0.60 (-3.57)	0.64 (10.14)	1.39 (-2.67)
(4,2)	0.55 (2.33)	0.21 (-9.68)	0.10 (-11.54)	0.44 (-6.15)	0.61 (-1.79)	0.55 (-2.90)	1.44 (-0.44)
(5,0)	0.53	0.21	0.13	0.48	0.63	0.51	1.43
(5,1)	0.54 (1.16)	0.20 (-3.23)	0.13 (0.00)	0.49 (1.54)	0.63 (0.00)	0.54 (4.35)	1.41 (-0.89)
(5,2)	0.56 (3.49)	0.20 (-3.23)	0.10 (-11.54)	0.44 (-6.15)	0.63 (0.00)	0.56 (7.25)	1.45 (0.89)
(6,0)	0.55	0.26	0.14	0.46	0.59	0.54	1.45
(6,1)	—	—	—	—	—	—	—
(6,2)	0.56 (1.16)	0.20 (-19.35)	0.11 (-11.54)	0.45 (-1.54)	0.62 (5.36)	0.57 (4.35)	1.45 (0.00)
(7,0)	0.57	0.25	0.11	0.46	0.61	0.47	1.46
(7,1)	—	—	—	—	—	—	—
(7,2)	0.56 (-1.16)	0.20 (-16.13)	0.09 (-7.69)	0.44 (-3.08)	0.61 (0.00)	0.53 (8.70)	1.42 (-1.78)
(b) 4 quarters ahead							
(4,0)	1.18	0.32	0.35	1.46	1.58	1.84	4.38
(4,1)	1.38 (11.63)	0.29 (-4.84)	0.47 (23.08)	1.84 (29.23)	1.54 (-3.57)	2.35 (36.96)	4.42 (0.89)
(4,2)	1.39 (12.21)	0.20 (-19.35)	0.35 (0.00)	1.33 (-10.00)	1.57 (-0.89)	1.81 (-2.17)	4.51 (2.89)
(5,0)	1.20	0.19	0.41	1.30	1.62	1.43	4.59
(5,1)	1.23 (1.74)	0.19 (0.00)	0.38 (-5.77)	1.30 (0.00)	1.67 (4.46)	1.64 (15.22)	4.38 (-4.67)
(5,2)	1.44 (13.95)	0.18 (-1.61)	0.35 (-11.54)	1.31 (0.77)	1.63 (0.89)	1.87 (31.88)	4.54 (-1.11)
(6,0)	1.16	0.26	0.44	1.24	1.50	1.67	4.69
(6,1)	—	—	—	—	—	—	—
(6,2)	1.41 (14.53)	0.20 (-9.68)	0.37 (-13.46)	1.41 (13.08)	1.57 (6.25)	1.91 (17.39)	4.59 (-2.22)
(7,0)	1.43	0.19	0.37	1.28	1.60	1.17	4.56
(7,1)	—	—	—	—	—	—	—
(7,2)	1.42 (-0.58)	0.18 (-1.61)	0.33 (-7.69)	1.33 (3.85)	1.56 (-3.57)	1.69 (37.68)	4.50 (-1.33)
(c) 8 quarters ahead							
(4,0)	1.90	0.33	0.45	2.39	2.64	3.60	7.72
(4,1)	2.48 (23.84)	0.31 (-2.28)	0.66 (28.56)	3.14 (40.79)	2.55 (-5.68)	4.38 (39.97)	7.72 (0.00)
(4,2)	2.14 (9.87)	0.24 (-10.26)	0.54 (12.24)	2.27 (-6.53)	2.61 (-1.89)	3.64 (2.05)	7.98 (4.09)
(5,0)	1.90	0.23	0.59	2.00	2.53	2.50	7.23
(5,1)	1.88 (-0.82)	0.24 (1.14)	0.55 (-5.44)	1.96 (2.18)	2.57 (2.53)	2.86 (18.45)	6.96 (-4.24)
(5,2)	2.35 (18.50)	0.21 (-2.28)	0.55 (-5.44)	2.19 (10.33)	2.63 (6.31)	3.67 (59.95)	7.85 (9.74)
(6,0)	1.78	0.24	0.61	1.82	2.33	3.22	6.85
(6,1)	—	—	—	—	—	—	—
(6,2)	2.29 (20.97)	0.24 (0.00)	0.57 (-5.44)	2.43 (33.18)	2.52 (12.00)	3.69 (24.08)	8.19 (21.06)
(7,0)	2.15	0.20	0.52	1.96	2.71	1.90	7.23
(7,1)	—	—	—	—	—	—	—
(7,2)	2.26 (4.52)	0.20 (0.00)	0.51 (-1.36)	2.20 (13.05)	2.53 (-11.36)	3.32 (72.76)	8.15 (14.46)

Note. The table reports the mean squared forecast errors over the period 1990:1 to 2004:4. In (i, j) , i denotes the number of shocks, while $j = 0, 1$ and 2 corresponds to the default, first and second alternative specification, respectively. The initial estimation period is 1966:1–1989:4. The parameters are reestimated every quarter. The value in parentheses is computed as the difference in the MSFEs between the current and the default specification, divided by \sqrt{k} times the sample standard deviation of the first differenced series. The values exceeding 20% are highlighted in bold. The sample standard deviations for the seven series are: 0.86, 0.31, 0.26, 0.65, 0.56, 0.69, 2.25.

Table 3. Log marginal composite likelihood

Four shocks model		Five shocks model		Six shocks model		Seven shocks model	
Case	value	Case	value	Case	value	Case	value
(4,0)	-1099.9	(5,0)	-1019.9	(6,0)	-999.7	(7,0)	-940.4
(4,1)	-1115.2	(5,1)	-1026.0	(6,1)	–	(7,1)	–
(4,2)	-1075.4	(5,2)	-1055.2	(6,2)	-1052.6	(7,2)	-1036.9

Note. The sample period is the same as in Table 1. No training sample is used. In (i, j) , i denotes the number of shocks, while $j = 0, 1$ and 2 corresponds to the default, the first and the second alternative specification. $(6,1)$ and $(7,1)$ are the same as $(6,0)$ and $(7,0)$, therefore not repeated.

Figure 1. Response of output to a productivity shock

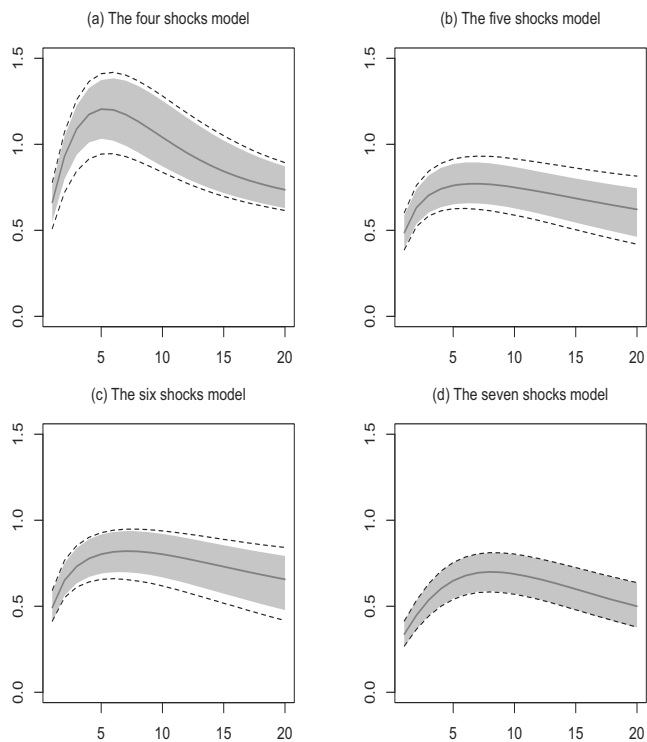


Figure 2. Response of inflation to a productivity shock

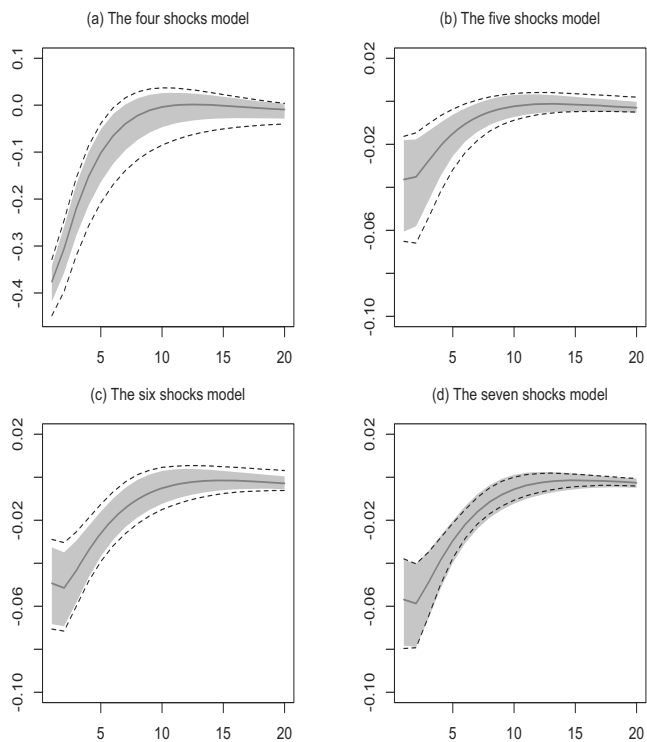


Figure 3. Response of interest to a productivity shock

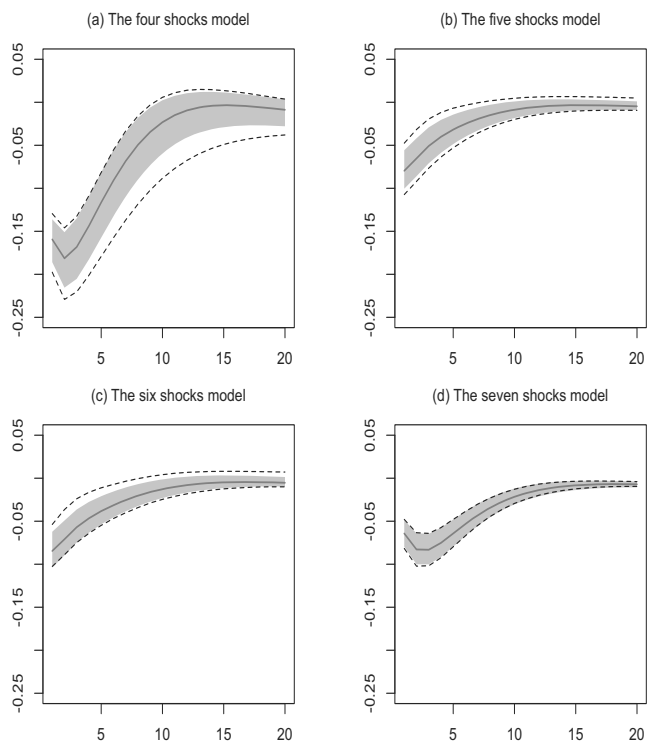


Figure 4. Response of investment to a productivity shock

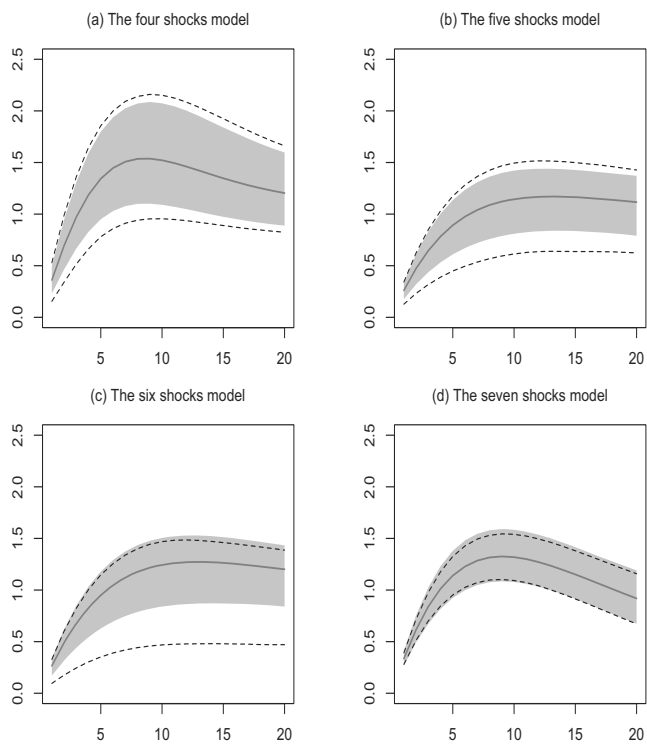


Figure 5. Response of hours worked to a productivity shock

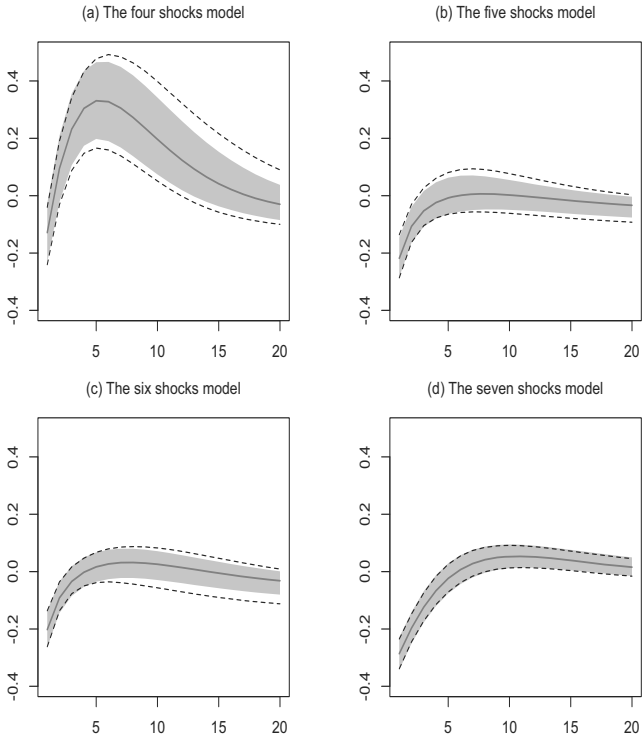


Figure 6. Response of wage to a productivity shock

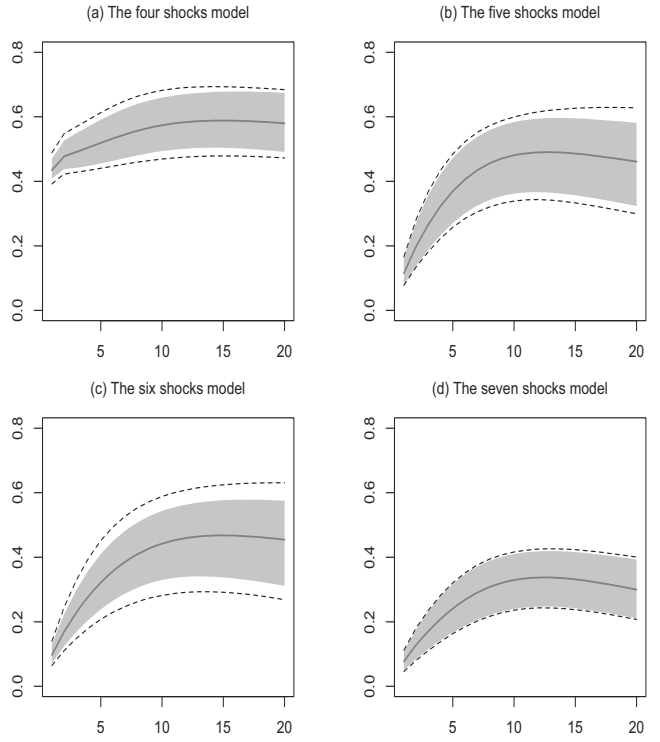
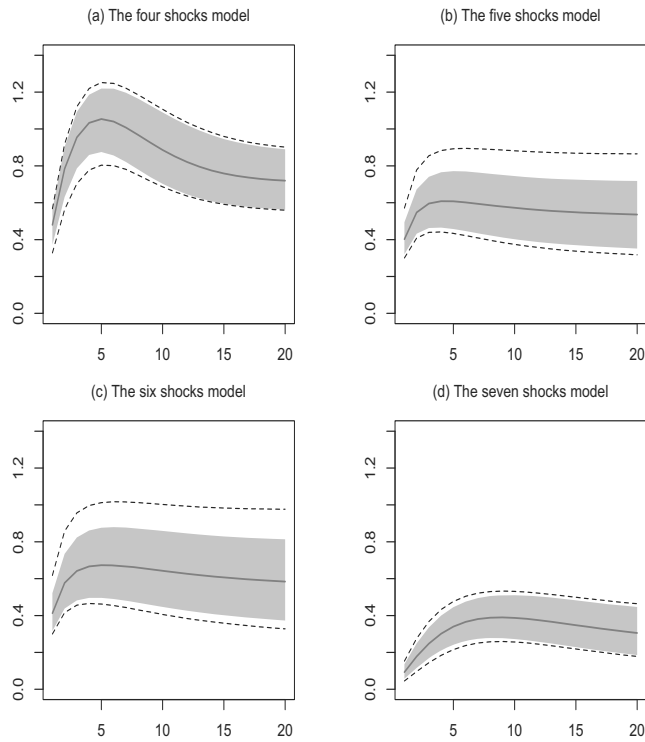


Figure 7. Response of consumption to a productivity shock



Online Appendix

This appendix is structured as follows. Section S.1 contains the proofs. Section S.2 includes the original Smets and Wouters (2007) model. Section S.3 shows how to compute the Gaussian likelihood for a nonsingular submodel in both the time and the frequency domain. Section S.4 gives two examples to illustrate the MCMC and Asymptotic Intervals under misspecification. Section S.5 applies the composite likelihood framework to analyze small scale singular models. Some tables and figures appear at the very end.

S.1 Proofs for the results in the paper

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 I still include the details for the matter of completeness.

Define the following correspondence:

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

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

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

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 adopting 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 :

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

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

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

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

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

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, assume $\{d_k\}$ itself is the convergent subsequence. Then, as $\theta_k \rightarrow \theta_0$, d_k approaches d and

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

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 $\bar{R}_s(\omega; \theta_0)$, which implies $[\partial \bar{R}_s(\omega; \theta_0) / \partial \theta'] d = 0$, further implying $d' [\partial \bar{R}_s(\omega; \theta_0) / \partial \theta']' [\partial \bar{R}_s(\omega; \theta_0) / \partial \theta'] d = 0$. Because the above result holds for arbitrary ω and s , it also holds when integrating over $\omega \in [-\pi, \pi]$ and summing over $s \in \{1, \dots, S\}$:

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

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

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

it follows that

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

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 $s \in \{1, \dots, S\}$. Consequently,

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

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

$$\frac{\partial \bar{R}_s(\omega; \theta(v))}{\partial v} = \frac{\partial \bar{R}_s(\omega; \theta(v))}{\partial \theta(v)'} \frac{\partial \theta(v)}{\partial v} = \frac{\partial \bar{R}_s(\omega; \theta(v))}{\partial \theta(v)'} c(\theta) = 0 \tag{S.3}$$

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

Proof of Lemma 2. First, consider the frequency domain formulation as in (S.15). The arguments in Lemma 1 in Hannan (1973) implies:

$$\begin{aligned} T^{-1} \sum_{j=1}^{T-1} \text{tr} \{ f_s^{-1}(\omega_j; \theta) I_{s,T}(\omega_j) \} &\xrightarrow{a.s} \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \{ f_s^{-1}(\omega; \theta) f_{s,0}(\omega) \} d\omega, \\ T^{-1} \text{tr} \{ f_s^{-1}(0; \theta) I_{s,T}(0; \theta) \} &\xrightarrow{a.s} \frac{1}{2\pi} (\mu_{s,0} - \mu_s(\theta))' f_s^{-1}(0; \theta) (\mu_{s,0} - \mu_s(\theta)). \end{aligned}$$

Note that the key step in Hannan's proof is the uniform approximation of $f_s^{-1}(\omega; \theta)$ by a Cesaro sum of its Fourier series, which also holds 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 $T^{-1} \sum_{j=1}^{T-1} \log \det(f_s(\omega_j; \theta)) \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \det(f_s(\omega; \theta)) d\omega$. For the time domain likelihood, Lemma 2.1 follows from Hansen and Sargent (1993, p.49-52).

Consider the second result of the Lemma. Theorem 7.3.3 in Rockafellar and Wets (2009) and Lemma 2.1 imply that, with probability 1:

$$\limsup_{T \rightarrow \infty} \left(\arg \min_{\theta \in \Theta} -T^{-1} \ell(\theta) \right) \subseteq \arg \min_{\theta \in \Theta} -\ell_{\infty}(\theta),$$

i.e.,

$$\limsup_{T \rightarrow \infty} \check{C} \subseteq C_0. \quad (\text{S.4})$$

Applying Theorem 4.10(b) and Example 4.13 in Rockafellar and Wets (2009), (S.4) is equivalent to

$$\inf \{ \eta \geq 0 \mid \check{C} \subseteq C_0 + \eta \mathbb{B} \} \rightarrow 0, \quad (\text{S.5})$$

where $C_0 + \eta \mathbb{B}$ consists of all the points that lie in a closed ball of radius η around some point of C_0 . For any two compact sets A and B , $\inf \{ \eta \geq 0 \mid A \subseteq B + \eta \mathbb{B} \} < \varepsilon$ if and only if $h(A, B) < \varepsilon$. As a result, (S.5) holds if and only if

$$h(\check{C}, C_0) \rightarrow 0.$$

This proves the second result of the Lemma.

The third result follows immediately. This completes the proof.

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

$$T^{-1/2} \frac{\partial \ell(\hat{\theta}_T)}{\partial \theta} = 0, \quad (\text{S.6})$$

while the pseudo-true value θ_0 satisfies

$$T^{1/2} \frac{\partial \ell_{\infty}(\theta_0)}{\partial \theta} = 0. \quad (\text{S.7})$$

Consider a Taylor expansion of (S.6) around θ_0 :

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

where $\bar{\theta}$ lies between $\hat{\theta}_T$ and θ_0 . Rearrange terms and apply (S.7):

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

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

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

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

Term (I) satisfies a CLT with $\lim_{T \rightarrow \infty} \mathbb{E}(\text{vec} \{I_{s,T}(\omega_j) - f_{s,0}(\omega_j)\} \text{vec} \{I_{h,T}(\omega_j) - f_{h,0}(\omega_j)\}^*) = f_{s,h,0}(\omega_j)' \otimes f_{s,h,0}(\omega_j)$. This leads to $V_{1,s,h}$ defined in the theorem. Term (II) also satisfies a CLT with $\lim_{T \rightarrow \infty} \mathbb{E}(Y_{s,t} - \mu_{s,0})(Y_{h,t} - \mu_{s,0})' = 2\pi f_{s,h,0}(0)$. 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, I 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 $\{\theta : T^{1/2} \|\theta - \theta_0\| \leq M\}$, 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 addition, 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)

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

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 (S.14) in the handling of the initial condition, whose effects vanish asymptotically.

Below, I 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

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

I 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 $\|q_{s,m}(\omega; \theta) - f_s^{-1}(\omega; \theta)\| + \|\partial q_{s,m}(\omega; \theta)/\partial \theta_k - \partial f_s^{-1}(\omega; \theta)/\partial \theta_k\| = O(T^{-3/5})$ 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 $\tilde{H}_s^{-1}(\theta)$ 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):

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

Then, applying the relationship between $q_{s,m}(\omega; \theta)$ and $\tilde{H}_s^{-1}(\theta)$, it follows that

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

The results in the above three displays hold uniformly over the compact set defined above.

They jointly imply

$$(B) = -\frac{1}{2T^{1/2}} \sum_{j=0}^{T-1} I_{s,T}(\omega_j) \frac{\partial f_s^{-1}(\omega; \theta)}{\partial \theta_k} + o_p(1). \quad (\text{S.8})$$

For Term (A), note that

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

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

$$(A) = \frac{1}{2T^{1/2}} \sum_{j=0}^{T-1} \mathbb{E} \text{tr} \left\{ I_{s,T}(\omega_j) \frac{\partial f_s^{-1}(\omega_j; \theta)}{\partial \theta_k} \right\} + o(1) = \frac{1}{2T^{1/2}} \sum_{j=0}^{T-1} \text{tr} \left\{ f_s(\omega_j; \theta) \frac{\partial f_s^{-1}(\omega_j; \theta)}{\partial \theta_k} \right\} + o(1). \quad (\text{S.9})$$

Finally, applying the same argument used for (B) to (C):

$$(C) = \frac{1}{\pi T^{1/2}} \frac{\partial \boldsymbol{\mu}_s(\theta)'}{\partial \theta_k} f_s^{-1}(0; \theta) (Y_s - \boldsymbol{\mu}_s(\theta)) + o(1). \quad (\text{S.10})$$

The sum of the leading terms of (S.9), (S.8) and (S.10) equal the first order derivative of the frequency domain likelihood (S.15) with respect to θ_k . This completes the proof.

S.2 Smets and Wouters (2007)

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: $k_y = \phi_p (L_*/k_*)^{\alpha-1} = \phi_p [((1 - \alpha)/\alpha) (R_*^k/w_*)]^{\alpha-1}$ with $w_* = (\alpha^\alpha(1 - \alpha)^{(1-\alpha)})/[\phi_p (R_*^k)^\alpha]^{1/(1-\alpha)}$, and $R_*^k = \beta^{-1}\gamma^{\sigma_c} - (1 - \delta)$.

Households: The consumption Euler equation is

$$c_t = c_1 c_{t-1} + (1 - c_1) E_t c_{t+1} + c_2 (l_t - E_t l_{t+1}) - c_3 (r_t - E_t \pi_{t+1}) - \varepsilon_t^b. \quad (\text{S.11})$$

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 (S.11) to the habit persistence (λ), steady state labor market mark-up (ϕ_w), and other structural parameters highlighted above is:

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

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

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

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 $i_2 = 1/[(1 + \beta\gamma^{(1-\sigma_c)}) \gamma^2 \varphi]$, 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

$$q_t = q_1 E_t q_{t+1} + (1 - q_1) E_t r_{t+1}^k - (r_t - E_t \pi_{t+1}) - \frac{1}{c_3} \varepsilon_t^b, \quad (\text{S.12})$$

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

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

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 $k_2 = (1 - k_1) (1 + \beta\gamma^{(1-\sigma_c)}) \gamma^2 \varphi$.

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

$$\pi_t = \pi_1 \pi_{t-1} + \pi_2 E_t \pi_{t+1} - \pi_3 \mu_t^p + \varepsilon_t^p, \quad (\text{S.13})$$

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:

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

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

Labor market: The wage mark-up is

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

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):

$$w_1 = \frac{1}{1 + \beta\gamma^{(1-\sigma_c)}}, w_2 = \frac{1 + \beta\gamma^{(1-\sigma_c)}\iota_w}{1 + \beta\gamma^{(1-\sigma_c)}}, w_3 = \frac{\iota_w}{1 + \beta\gamma^{(1-\sigma_c)}},$$

$$w_4 = \frac{1}{1 + \beta\gamma^{(1-\sigma_c)}} \frac{(1 - \beta\gamma^{(1-\sigma_c)}\xi_w)(1 - \xi_w)}{\xi_w((\phi_w - 1)\epsilon_w + 1)}.$$

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.

S.3 Likelihood for a nonsingular submodel

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

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

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 $t - 1$:

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

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

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

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

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

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

$$l_s(\theta) = -\frac{nT}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \log \det(F_t) - \frac{1}{2} \sum_{t=1}^T \eta_t' F_t^{-1} \eta_t. \quad (\text{S.14})$$

The frequency domain Gaussian likelihood for a nonsingular submodel. Let ω_j denote the Fourier frequencies, i.e., $\omega_j = 2\pi j/T$ ($j = 1, 2, \dots, T-1$). The discrete Fourier transforms and periodograms of $Y_{s,t}$ 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

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

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

$$\begin{aligned} & -\frac{1}{2} \sum_{j=1}^{T-1} [\log \det (f_s(\omega_j; \theta)) + \text{tr} \{f_s^{-1}(\omega_j; \theta) I_{s,T}(\omega_j)\}] \\ & -\frac{1}{2} [\log \det (f_s(0; \theta)) + \text{tr} \{f_s^{-1}(0; \theta) I_{s,T}(0; \theta)\}]. \end{aligned} \tag{S.15}$$

S.4 Examples related to misspecification and inference

This appendix uses two examples to illustrate the MCMC and Asymptotic Intervals under misspecification. The examples are kept simple to deliver analytical results. In each example, I obtain the MCMC interval, examine its frequentist coverage, and then repeat the same analysis for the Asymptotic Interval. It will emerge that ignoring misspecification may lead to confidence unwarranted by the data.

Example S1. (Misspecification: omission) Suppose a researcher is interested in learning about the mean of GDP growth series y_t . Suppose the true data generating process is: $y_t = \mu + u_t$ with $u_t = \rho u_t + e_t$, where $|\rho| < 1$ and $e_t \sim i.i.d.N(0, 1)$. However, the researcher wasn't aware that serial correlation could be present and specified u_t as $i.i.d.N(0, \sigma^2)$.

Let $\tau = 1/\sigma^2$ and use the conventional prior $\pi(\mu, \tau) \propto 1/\tau$ for $-\infty < \mu < \infty$ and $0 < \tau < \infty$. Then, the posterior is

$$p(\mu, \tau|y) \propto \tau^{T/2-1} \exp\left(-\frac{T\tau}{2}(\mu - \hat{\mu})^2\right) \exp\left(-\frac{\tau}{2}\sum_{t=1}^T (y_t - \hat{\mu})^2\right),$$

where $\hat{\mu}$ denotes the sample average. The marginal posterior for μ satisfies:

$$\frac{T^{1/2}(\mu - \hat{\mu})}{s} \sim t_{T-1}, \quad (\text{S.16})$$

where $s^2 = (T - 1)^{-1} \sum_{t=1}^T (y_t - \hat{\mu})^2$ and t_{T-1} denotes the standard t density with $T - 1$ degrees of freedom. A level $(1 - \alpha)$ credible set for μ using (S.16) and ignoring the misspecification is given by

$$\left[\hat{\mu} - \frac{s}{\sqrt{T}} q_{T-1, \alpha}, \hat{\mu} + \frac{s}{\sqrt{T}} q_{T-1, \alpha} \right],$$

where $q_{T-1, \alpha}$ is $(1 - \frac{\alpha}{2})$ -th percentile of a t distribution with $T - 1$ degrees of freedom. Because the t distribution approaches the normal distribution as the degrees of freedom increases, in large samples the above interval can be well approximated by

$$\left[\hat{\mu} - \frac{s}{\sqrt{T}} z_{\alpha}, \hat{\mu} + \frac{s}{\sqrt{T}} z_{\alpha} \right],$$

or, further by

$$\left[\hat{\mu} - \frac{1}{\sqrt{T(1 - \rho^2)}} z_{\alpha}, \hat{\mu} + \frac{1}{\sqrt{T(1 - \rho^2)}} z_{\alpha} \right] \quad (\text{S.17})$$

because $s^2 \rightarrow^p 1/(1 - \rho^2)$.

The frequentist coverage of (S.17) is sensitive to ρ . The coverage equals $(1 - \alpha)$ only when $\rho = 0$. In particular, if α is set to 0.1, then the asymptotic frequentist coverage will be 77% and 66% when $\rho = 0.3$ and 0.5. Therefore, using this interval without thinking about misspecification can lead to overly optimistic inference, at least from a frequentist perspective. This does not improve when the sample size is increased.

The Asymptotic Interval for μ involves computing $M^{-1}VM^{-1}$. Because M is diagonal, it suffices to compute the entries in M and V corresponding to μ only. They equal τ and

$\tau^2/(1 - \rho)^2$ respectively. Therefore, replacing $(1 - \rho^2)$ in (S.17) by $M^{-1}VM^{-1}$, we obtain

$$\left[\hat{\mu} - \frac{1}{\sqrt{T(1 - \rho)^2}} z_\alpha, \hat{\mu} + \frac{1}{\sqrt{T(1 - \rho)^2}} z_\alpha \right].$$

This interval is wider than (S.17) if $\rho > 0$ and coincides with it only if $\rho = 0$. Because the central limit theorem implies

$$\sqrt{T}(\hat{\mu} - \mu) \rightarrow^d N(0, 1/(1 - \rho)^2),$$

this interval has correct asymptotic coverage.

This example reflects the reality that it is often hard to be fully aware in advance which features should be put into a DSGE model. Computing the Asymptotic Interval can help avoid overly optimistic inference under such a reality.

Example S2. (Misspecification: wrong structure) Suppose the researcher misspecifies an MA(1) as an AR(1) process. That is, the actual data are generated by

$$y_t = e_t + \theta e_{t-1} \quad \text{with } e_t \sim i.i.d.N(0, 1),$$

while the researcher estimates the following model

$$y_t = \rho y_{t-1} + v_t$$

with $v_t \sim i.i.d.N(0, 1/\tau)$ and $y_0 = 0$.

Under the prior $\pi(\rho, \tau) \propto 1/\tau$ for $-\infty < \rho < \infty$ and $0 < \tau < \infty$, the posterior density is given by

$$p(\rho, \tau | y) \propto \tau^{T/2-1} \exp\left(-(\tau/2)(\rho - \hat{\rho})^2 \sum_{t=1}^T y_{t-1}^2\right) \exp\left(-(\tau/2) \sum_{t=1}^T (y_t - \hat{\rho} y_{t-1})^2\right),$$

where

$$\hat{\rho} = \left(\sum_{t=1}^T y_{t-1}^2 \right)^{-1} \sum_{t=1}^T y_{t-1} y_t.$$

The marginal posterior for ρ satisfies,

$$\frac{T^{1/2}(\rho - \hat{\rho})}{s} \sim t_{T-1}, \tag{S.18}$$

where $s^2 = (T - 1)^{-1} \sum_{t=1}^T (y_t - \hat{\rho}y_{t-1})^2$ and t_{T-1} denotes the standard t density with $T - 1$ degrees of freedom. A level $(1 - \alpha)$ credible set for ρ using (S.18) and ignoring the misspecification is given by

$$\left[\hat{\rho} - \frac{s}{\sqrt{\sum_{t=1}^T y_{t-1}^2}} q_{T-1, \alpha}, \hat{\rho} + \frac{s}{\sqrt{\sum_{t=1}^T y_{t-1}^2}} q_{T-1, \alpha} \right].$$

Because the t distribution approaches the normal distribution as T increases, in large samples the above interval can be approximated by

$$\left[\hat{\rho} - \frac{s}{\sqrt{\sum_{t=1}^T y_{t-1}^2}} z_\alpha, \hat{\rho} + \frac{s}{\sqrt{\sum_{t=1}^T y_{t-1}^2}} z_\alpha \right].$$

Because

$$\frac{s^2}{T^{-1} \sum_{t=1}^T y_{t-1}^2} \rightarrow^p 1 - \left(\frac{\theta}{1 + \theta^2} \right)^2 = \frac{1 + \theta^4 - \theta^2}{(1 + \theta^2)^2},$$

the interval can be further approximated by

$$\left[\hat{\rho} - z_\alpha \frac{1}{1 + \theta^2} \sqrt{\frac{1 + \theta^4 - \theta^2}{T}}, \hat{\rho} + z_\alpha \frac{1}{1 + \theta^2} \sqrt{\frac{1 + \theta^4 - \theta^2}{T}} \right]. \quad (\text{S.19})$$

Note that this interval is centered at $\hat{\rho}$. As the sample size increases,

$$\hat{\rho} \rightarrow^p \frac{\theta}{1 + \theta^2}.$$

This limiting value corresponds to the best AR(1) approximation to the MA(1) process in the sense of minimizing the Kullback Leibler divergence. The resulting AR(1) has the same first order autocorrelation as the MA(1) process.

The Asymptotic Interval is also centered at $\hat{\rho}$. Its length is determined by $M^{-1}VM^{-1}$. Because M is diagonal, it suffices to compute the entries in M and V corresponding to ρ only. They are given by $\tau \text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T y_{t-1}^2$ and $\lim_{T \rightarrow \infty} \text{Var} \left(\tau T^{-1/2} (\hat{\rho} - \rho) \sum_{t=1}^T y_{t-1}^2 \right)$. The corresponding element of $M^{-1}VM^{-1}$ is therefore

$$\lim_{T \rightarrow \infty} \text{Var} \left(T^{1/2} (\hat{\rho} - \rho) \right) = \frac{\theta^8 + \theta^6 + 4\theta^4 + \theta^2 + 1}{(1 + \theta^2)^4}.$$

The resulting Asymptotic Interval is

$$\left[\hat{\rho} - z_\alpha \frac{1}{(1 + \theta^2)^2} \sqrt{\frac{\theta^8 + \theta^6 + 4\theta^4 + \theta^2 + 1}{T}}, \quad \hat{\rho} + z_\alpha \frac{1}{(1 + \theta^2)^2} \sqrt{\frac{\theta^8 + \theta^6 + 4\theta^4 + \theta^2 + 1}{T}} \right].$$

The length is wider than (S.19) whenever $\theta \neq 0$. This interval has correct asymptotic coverage because, by the CLT,

$$T^{1/2} \left(\hat{\rho} - \frac{\theta}{1 + \theta^2} \right) \rightarrow^d N \left(0, \frac{\theta^8 + \theta^6 + 4\theta^4 + \theta^2 + 1}{(1 + \theta^2)^4} \right).$$

This example reflects the reality that it is often hard to know in advance what data generating structure should be used as part of a bigger model. Computing the Asymptotic Interval can again help avoid overly optimistic inference under such a reality.

S.5 Application to small scale singular models

The original model is (9). The first singular model is only affected by ε_{zt} . The second is also affected by ε_{gt} . The sample period is 1982:IV-1997:IV. The same priors are used throughout. They are reported in the first four columns of Table S4. The estimation results for the original three shocks model are also included. This provides a useful point of reference when interpreting the results.

S.5.1 The one shock model

The following subsets are used to form the composite likelihood: $\{y_t\}$, $\{\pi_t\}$, and $\{r_t\}$.

Parameter estimates. See Panel (a) in Table S4. There are two notable differences relative to the nonsingular model: the inflation weight parameter (ψ_1) is smaller, while the standard deviation of the productivity shock (σ_z) is higher. The remaining parameters are similar.

The parameters ψ_1 and σ_z are important for the economy's responses to a one standard deviation shock in productivity. A lower ψ_1 alone implies a deeper decline in the aggregate price level, which can potentially dampen the increase in the output. A higher σ_z implies

more pronounced responses in all three variables. I now use impulse response functions to further quantify their effects.

Impulse responses. Panel (A) in Figure S31 displays the responses to a one standard deviation shock in productivity. The solid line is computed using the posterior mean. The two dashed lines and the shaded area correspond to 90% Asymptotic and MCMC Intervals respectively. The responses in the nonsingular model are included in Panel (C) for comparison. The three responses are all stronger than in the nonsingular model. Therefore, the effect of a higher σ_z dominates that of a lower ψ_1 . The Asymptotic Intervals are consistently wider than the MCMC Intervals due to more frequent occurrence of a low ψ_1 with high κ and σ_z .

Therefore, leaving out the exogenous spending and monetary policy shocks has significantly affected the assessment of the effect of the productivity shock on the three variables.

S.5.2 The two shocks model

The following subsets are used to form the composite likelihood: $\{y_t, r_t\}$ and $\{\pi_t\}$.

Parameter estimates. See Panel (b) in Table S4. The value of σ_z and the associated intervals are now very close to those in the nonsingular model. The value of ψ_1 is also closer to the latter. The point estimate for ρ_{gz} is noticeably lower. However, this difference can be interpreted as moderate because of the high uncertainty associated with this parameter. The remaining parameter values are similar to those in the nonsingular model.

Impulse responses. The results are reported in Panel (B) in Figure S31. The point estimates and the associated intervals are similar to the nonsingular model.

In summary, I find from the two small scale singular models that the parameter estimates and impulse responses to the productivity shock can be either similar or substantially different, depending on what other shocks are in the model.

References

Andrews, D.W.K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–858.

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

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

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

Rockafellar, R.T. and R.J.B. Wets (2009): *Variational Analysis (Vol. 317)*, Springer.

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

Table S1: Parameters 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 intertemporal 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]
$\bar{\gamma}$	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]
$\bar{\pi}$	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 S2: Estimation results under the first alternative specification

	(a) The four shocks model				(b) The five shocks model			
	Mode	Mean	MCMC	Asymptotic	Mode	Mean	MCMC	Asymptotic
α	0.20	0.20	[0.17,0.24]	[0.16,0.24]	0.29	0.29	[0.26,0.32]	[0.25,0.32]
ψ	0.24	0.29	[0.14,0.50]	[0.13,0.50]	0.44	0.50	[0.34,0.67]	[0.33,0.67]
φ	5.66	5.98	[4.27,7.87]	[4.89,7.87]	5.00	5.13	[3.48,6.94]	[4.03,6.94]
σ_c	1.80	1.86	[1.53,2.23]	[1.49,2.23]	1.61	1.61	[1.30,1.97]	[1.31,1.97]
λ	0.40	0.40	[0.33,0.47]	[0.33,0.47]	0.35	0.36	[0.29,0.43]	[0.29,0.43]
ϕ_p	1.35	1.33	[1.22,1.45]	[1.24,1.45]	1.45	1.48	[1.35,1.63]	[1.38,1.63]
ι_w	0.56	0.57	[0.43,0.68]	[0.41,0.68]	0.85	0.83	[0.70,0.93]	[0.75,0.93]
ξ_w	0.91	0.91	[0.89,0.93]	[0.90,0.93]	0.63	0.63	[0.52,0.73]	[0.48,0.73]
ι_p	0.11	0.13	[0.05,0.23]	[0.08,0.23]	0.20	0.21	[0.09,0.35]	[0.13,0.35]
ξ_p	0.13	0.15	[0.11,0.21]	[0.12,0.21]	0.75	0.74	[0.66,0.81]	[0.68,0.81]
σ_l	3.77	3.81	[2.92,4.74]	[3.37,4.74]	3.12	2.13	[2.49,3.85]	[2.80,3.85]
r_π	1.13	1.22	[1.04,1.47]	[1.00,1.47]	2.09	2.08	[1.80,2.36]	[1.90,2.36]
$r_{\Delta y}$	0.30	0.31	[0.25,0.37]	[0.26,0.37]	0.24	0.24	[0.20,0.29]	[0.19,0.29]
r_y	0.04	0.05	[0.01,0.10]	[0.00,0.10]	0.12	0.13	[0.09,0.18]	[0.08,0.18]
ρ	0.75	0.76	[0.69,0.83]	[0.63,0.83]	0.79	0.79	[0.74,0.83]	[0.73,0.83]
ρ_a	0.98	0.97	[0.96,0.99]	[0.96,0.99]	0.99	0.99	[0.98,0.99]	[0.97,0.99]
ρ_b	-	-	-	-	-	-	-	-
ρ_g	0.98	0.98	[0.98,0.99]	[0.98,0.99]	0.94	0.94	[0.91,0.97]	[0.91,0.97]
ρ_i	0.64	0.64	[0.54,0.73]	[0.54,0.73]	0.65	0.64	[0.54,0.72]	[0.54,0.72]
ρ_r	0.22	0.23	[0.13,0.34]	[0.14,0.34]	0.08	0.11	[0.04,0.19]	[0.04,0.19]
ρ_p	-	-	-	-	0.98	0.97	[0.95,0.99]	[0.95,0.99]
μ_p	-	-	-	-	0.84	0.81	[0.68,0.90]	[0.64,0.90]
ρ_w	-	-	-	-	-	-	-	-
μ_w	-	-	-	-	-	-	-	-
ρ_{ga}	0.85	0.84	[0.71,0.98]	[0.71,0.98]	0.32	0.33	[0.15,0.52]	[0.14,0.52]
σ_a	0.51	0.54	[0.47,0.61]	[0.45,0.61]	0.51	0.51	[0.45,0.57]	[0.44,0.57]
σ_b	-	-	-	-	-	-	-	-
σ_g	0.62	0.63	[0.58,0.69]	[0.57,0.69]	0.60	0.66	[0.55,0.81]	[0.48,0.81]
σ_i	0.49	0.51	[0.43,0.60]	[0.42,0.60]	0.57	0.58	[0.50,0.68]	[0.44,0.68]
σ_r	0.37	0.38	[0.34,0.43]	[0.33,0.43]	0.28	0.29	[0.26,0.32]	[0.24,0.32]
σ_p	-	-	-	-	0.15	0.15	[0.12,0.18]	[0.11,0.18]
σ_w	-	-	-	-	-	-	-	-
$\bar{\gamma}$	0.40	0.39	[0.36,0.43]	[0.34,0.43]	0.40	0.39	[0.32,0.45]	[0.33,0.45]
r	0.11	0.13	[0.06,0.23]	[0.10,0.23]	0.12	0.14	[0.07,0.24]	[0.11,0.24]
$\bar{\pi}$	0.65	0.67	[0.51,0.85]	[0.62,0.85]	0.63	0.66	[0.50,0.83]	[0.62,0.83]
\bar{l}	-0.84	-0.94	[-2.65,0.85]	[-2.35,0.85]	0.41	0.37	[-0.99,1.80]	[-0.73,1.80]

Note. See Table S1.

Table S3: Estimation results under the second alternative specification

	(a) The four shocks model				(b) The five shocks model			
	Mode	Mean	MCMC	Asymptotic	Mode	Mean	MCMC	Asymptotic
α	0.31	0.31	[0.27,0.35]	[0.26,0.35]	0.33	0.32	[0.28,0.37]	[0.24,0.37]
ψ	0.52	0.55	[0.34,0.76]	[0.42,0.76]	0.37	0.41	[0.22,0.61]	[0.13,0.61]
φ	3.70	4.13	[2.33,6.41]	[2.00,6.41]	4.59	4.49	[2.43,6.80]	[2.00,6.80]
σ_c	1.22	1.23	[1.02,1.47]	[0.92,1.47]	1.13	1.15	[0.77,1.51]	[0.25,1.51]
λ	0.42	0.43	[0.35,0.51]	[0.35,0.51]	0.33	0.34	[0.25,0.44]	[0.23,0.44]
ϕ_p	1.00	1.05	[1.00,1.17]	[1.00,1.17]	1.38	1.38	[1.23,1.53]	[1.24,1.53]
ι_w	0.58	0.62	[0.42,0.80]	[0.44,0.80]	0.73	0.72	[0.55,0.87]	[0.53,0.87]
ξ_w	0.30	0.33	[0.30,0.39]	[0.31,0.39]	0.34	0.39	[0.31,0.52]	[0.30,0.52]
ι_p	0.22	0.21	[0.08,0.40]	[0.01,0.40]	0.22	0.22	[0.11,0.36]	[0.12,0.36]
ξ_p	0.27	0.27	[0.16,0.39]	[0.15,0.39]	0.56	0.55	[0.41,0.68]	[0.23,0.68]
σ_l	0.75	0.82	[0.40,1.32]	[0.25,1.32]	1.09	1.20	[0.76,1.68]	[0.32,1.68]
r_π	1.81	1.86	[1.62,2.14]	[1.23,2.14]	1.98	2.05	[1.75,2.34]	[1.30,2.34]
$r_{\Delta y}$	0.11	0.12	[0.04,0.20]	[0.09,0.20]	0.07	0.07	[0.01,0.14]	[0.03,0.14]
r_y	0.12	0.12	[0.05,0.21]	[0.09,0.21]	0.11	0.12	[0.06,0.18]	[0.05,0.18]
ρ	0.53	0.55	[0.51,0.62]	[0.52,0.62]	0.50	0.55	[0.50,0.61]	[0.50,0.61]
ρ_a	0.98	0.98	[0.95,0.99]	[0.94,0.99]	0.51	0.53	[0.19,0.84]	[0.33,0.84]
ρ_b	-	-	-	-	-	-	-	-
ρ_g	0.42	0.46	[0.18,0.73]	[0.04,0.73]	0.30	0.38	[0.11,0.72]	[0.15,0.72]
ρ_i	0.89	0.89	[0.83,0.94]	[0.74,0.93]	0.81	0.82	[0.70,0.93]	[0.33,0.93]
ρ_r	0.44	0.42	[0.13,0.79]	[0.20,0.79]	0.48	0.44	[0.15,0.76]	[0.18,0.76]
ρ_p	-	-	-	-	0.98	0.98	[0.96,0.99]	[0.95,0.99]
μ_p	-	-	-	-	0.76	0.72	[0.55,0.86]	[0.54,0.86]
ρ_w	-	-	-	-	-	-	-	-
μ_w	-	-	-	-	-	-	-	-
ρ_{ga}	0.40	0.40	[0.25,0.61]	[0.01,0.61]	0.51	0.56	[0.17,0.96]	[0.36,0.96]
σ_a	0.55	0.58	[0.49,0.70]	[0.37,0.70]	0.05	0.09	[0.03,0.23]	[0.01,0.23]
σ_b	-	-	-	-	-	-	-	-
σ_g	0.21	0.20	[0.11,0.29]	[0.01,0.29]	0.28	0.28	[0.17,0.40]	[0.01,0.40]
σ_i	0.48	0.49	[0.38,0.63]	[0.33,0.63]	0.48	0.49	[0.39,0.65]	[0.01,0.65]
σ_r	0.04	0.07	[0.03,0.14]	[0.02,0.14]	0.05	0.09	[0.03,0.19]	[0.01,0.19]
σ_p	-	-	-	-	0.19	0.20	[0.15,0.25]	[0.01,0.25]
σ_w	-	-	-	-	-	-	-	-
$\bar{\gamma}$	0.45	0.44	[0.41,0.48]	[0.40,0.48]	0.45	0.45	[0.41,0.49]	[0.39,0.49]
r	0.22	0.25	[0.12,0.41]	[0.18,0.41]	0.22	0.25	[0.12,0.42]	[0.18,0.42]
$\bar{\pi}$	0.77	0.76	[0.62,0.91]	[0.66,0.91]	0.64	0.66	[0.51,0.82]	[0.58,0.82]
\bar{l}	-0.77	-0.73	[-1.75,0.25]	[-2.06,0.25]	-0.52	-0.55	[-1.96,0.90]	[-1.96,0.90]

Note. See Table S1.

Table S3: Estimation results under the second alternative specification (cont'd)

	(c) The six shocks model				(d) The seven shocks model			
	Mode	Mean	MCMC	Asymptotic	Mode	Mean	MCMC	Asymptotic
α	0.36	0.35	[0.30,0.40]	[0.29,0.40]	0.32	0.32	[0.27,0.36]	[0.22,0.36]
ψ	0.44	0.43	[0.24,0.65]	[0.27,0.65]	0.45	0.47	[0.31,0.63]	[0.33,0.63]
φ	3.98	3.77	[2.18,5.91]	[2.00,5.91]	4.01	4.08	[2.58,5.96]	[2.00,5.96]
σ_c	1.40	1.46	[1.14,1.84]	[0.70,1.84]	1.18	1.30	[1.04,1.66]	[0.38,1.66]
λ	0.33	0.35	[0.25,0.46]	[0.21,0.46]	0.80	0.77	[0.66,0.86]	[0.57,0.86]
ϕ_p	1.40	1.40	[1.25,1.56]	[1.24,1.56]	1.44	1.46	[1.32,1.61]	[1.34,1.61]
ι_w	0.51	0.52	[0.27,0.77]	[0.42,0.77]	0.56	0.55	[0.30,0.78]	[0.38,0.78]
ξ_w	0.54	0.52	[0.35,0.68]	[0.34,0.68]	0.40	0.44	[0.34,0.58]	[0.30,0.58]
ι_p	0.39	0.38	[0.18,0.61]	[0.23,0.61]	0.37	0.37	[0.18,0.58]	[0.28,0.58]
ξ_p	0.76	0.75	[0.67,0.81]	[0.62,0.81]	0.73	0.73	[0.65,0.80]	[0.62,0.80]
σ_l	1.68	1.68	[1.03,2.44]	[0.82,2.44]	1.64	1.47	[0.60,2.51]	[0.25,2.25]
r_π	1.71	1.75	[1.35,2.10]	[1.32,2.10]	1.67	1.65	[1.20,2.01]	[1.00,2.01]
$r_{\Delta y}$	0.16	0.17	[0.08,0.25]	[0.08,0.25]	0.15	0.17	[0.08,0.25]	[0.13,0.25]
r_y	0.16	0.16	[0.08,0.24]	[0.10,0.24]	0.12	0.13	[0.05,0.21]	[0.06,0.21]
ρ	0.59	0.61	[0.51,0.77]	[0.50,0.77]	0.57	0.63	[0.52,0.74]	[0.50,0.74]
ρ_a	0.55	0.57	[0.34,0.81]	[0.36,0.81]	0.51	0.52	[0.19,0.84]	[0.41,0.84]
ρ_b	–	–	–	–	0.11	0.15	[0.05,0.26]	[0.06,0.26]
ρ_g	0.34	0.36	[0.13,0.65]	[0.09,0.65]	0.50	0.50	[0.17,0.82]	[0.29,0.82]
ρ_i	0.87	0.90	[0.83,0.95]	[0.71,0.95]	0.92	0.93	[0.89,0.96]	[0.85,0.96]
ρ_r	0.96	0.90	[0.77,0.97]	[0.34,0.97]	0.48	0.46	[0.15,0.78]	[0.11,0.78]
ρ_p	0.35	0.36	[0.15,0.60]	[0.18,0.60]	0.37	0.36	[0.13,0.65]	[0.10,0.65]
μ_p	0.60	0.59	[0.35,0.78]	[0.41,0.78]	0.60	0.59	[0.35,0.79]	[0.48,0.79]
ρ_w	0.52	0.50	[0.20,0.80]	[0.07,0.80]	0.49	0.48	[0.16,0.83]	[0.30,0.83]
μ_w	0.46	0.53	[0.22,0.85]	[0.01,0.85]	0.51	0.52	[0.20,0.84]	[0.40,0.84]
ρ_{ga}	0.72	0.72	[0.51,0.94]	[0.38,0.94]	0.51	0.55	[0.16,0.95]	[0.40,0.95]
σ_a	0.65	0.61	[0.40,0.78]	[0.09,0.78]	0.05	0.07	[0.03,0.12]	[0.06,0.12]
σ_b	–	–	–	–	0.28	0.26	[0.23,0.31]	[0.16,0.31]
σ_g	0.05	0.08	[0.03,0.20]	[0.01,0.20]	0.05	0.08	[0.03,0.17]	[0.02,0.17]
σ_i	0.55	0.59	[0.44,0.83]	[0.01,0.83]	0.49	0.50	[0.37,0.66]	[0.29,0.66]
σ_r	0.05	0.09	[0.04,0.20]	[0.01,0.20]	0.05	0.09	[0.03,0.19]	[0.01,0.19]
σ_p	0.21	0.21	[0.17,0.24]	[0.17,0.24]	0.21	0.21	[0.17,0.24]	[0.15,0.24]
σ_w	0.26	0.26	[0.20,0.32]	[0.18,0.32]	0.05	0.13	[0.04,0.25]	[0.01,0.25]
$\bar{\gamma}$	0.43	0.43	[0.40,0.47]	[0.39,0.47]	0.44	0.44	[0.41,0.47]	[0.39,0.47]
r	0.22	0.25	[0.12,0.41]	[0.19,0.41]	0.22	0.25	[0.11,0.41]	[0.18,0.41]
$\bar{\pi}$	0.69	0.73	[0.57,0.90]	[0.57,0.90]	0.78	0.75	[0.59,0.90]	[0.59,0.90]
\bar{l}	-0.75	-0.66	[-1.76,0.52]	[-1.98,0.52]	-0.75	-0.77	[-1.76,0.27]	[-2.05,0.27]

Note. See Table S1.

Table S4. 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): $\{1.0, 1E-5, 1E-5, 1E-5, 1E-5, 1E-5, 1E-5, 1E-5, -0.999, 1E-5, 1E-5, 1E-5\}$. The upper bounds are: $\{10, 0.999, 0.999, 10, 100, 10, 100, 0.999, 0.999, 0.999, 10, 10, 10\}$.

Figure S1. Response of output to a monetary policy shock

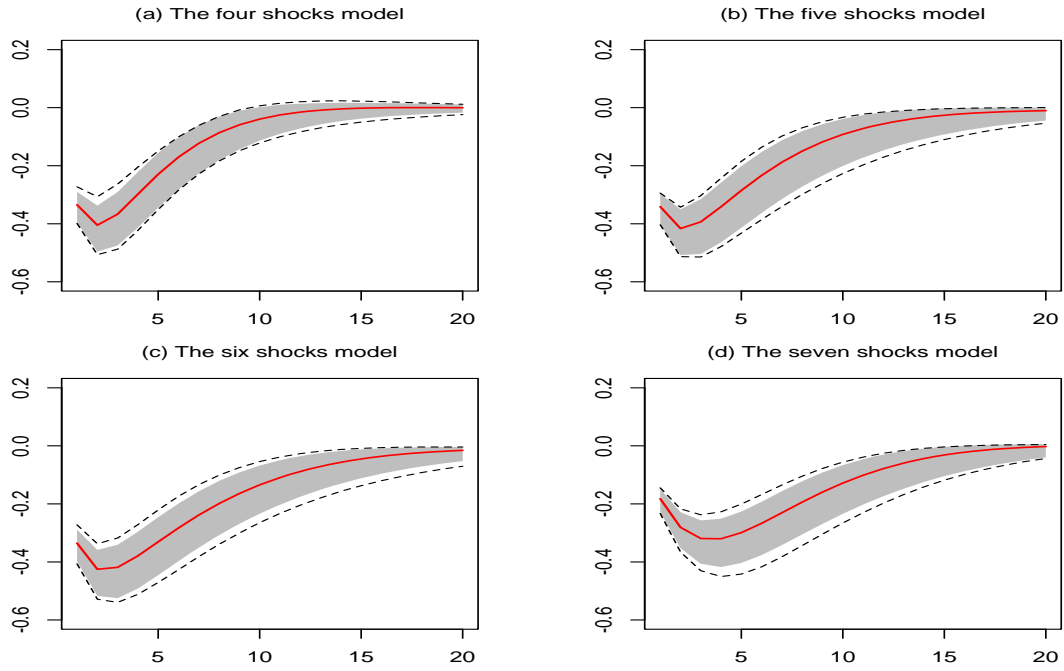
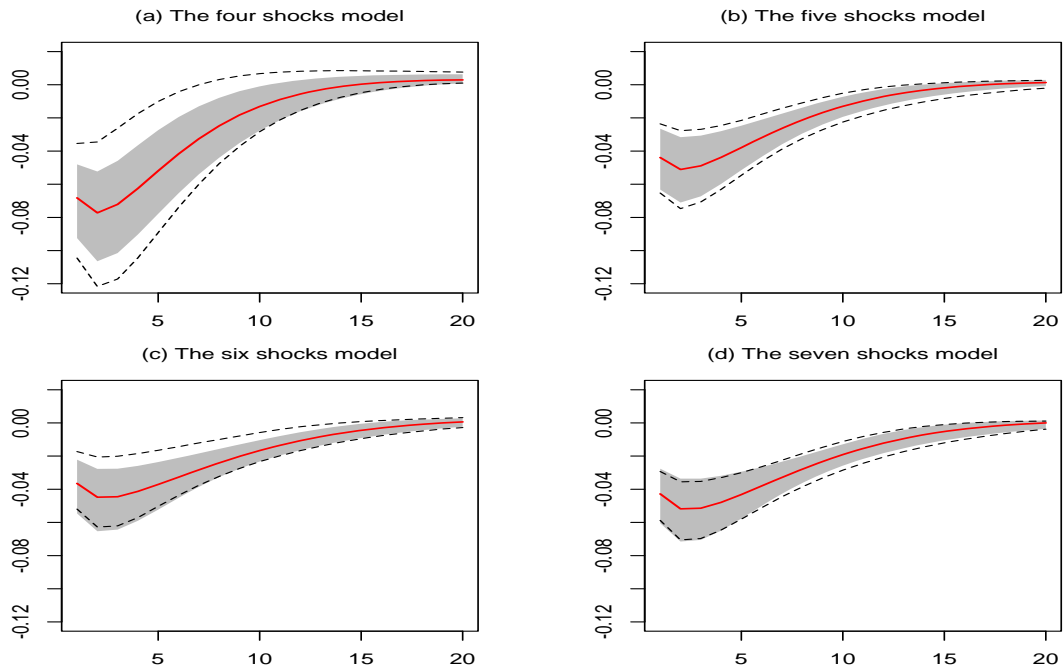


Figure S2. Response of inflation to a monetary policy shock



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

Figure S3. Response of interest rate to a monetary policy shock

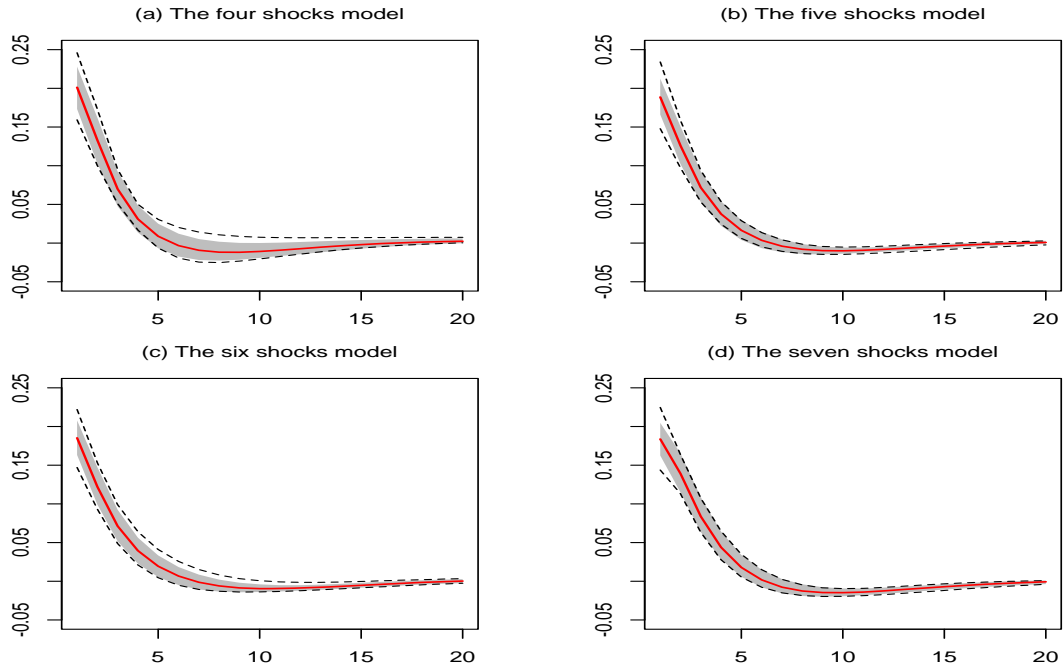
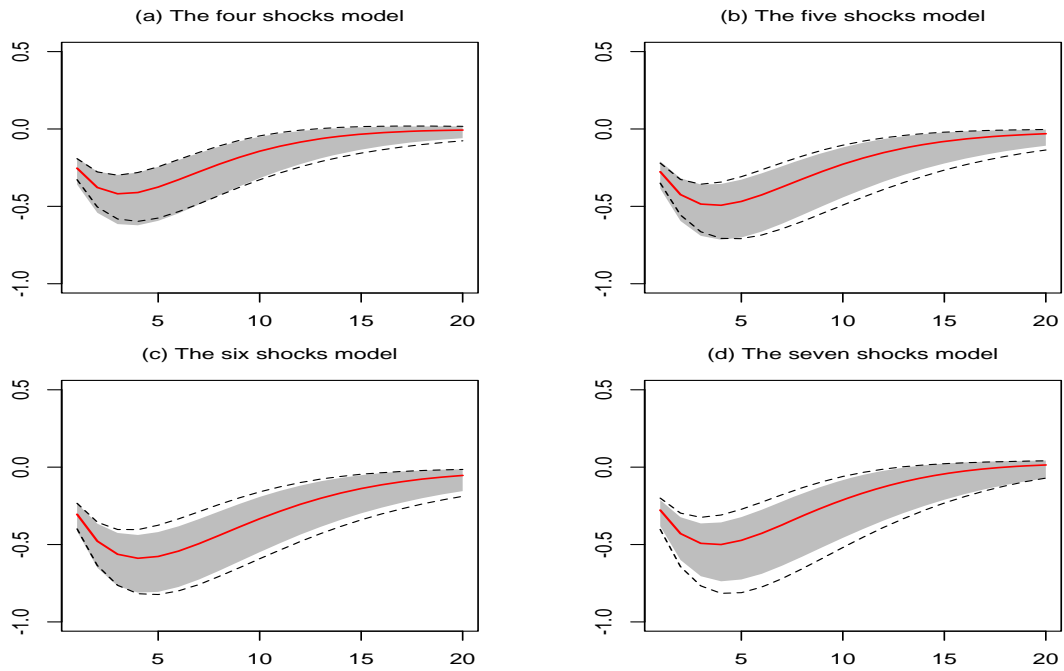


Figure S4. Response of investment to a monetary policy shock



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

Figure S5. Response of hours worked to a monetary policy shock

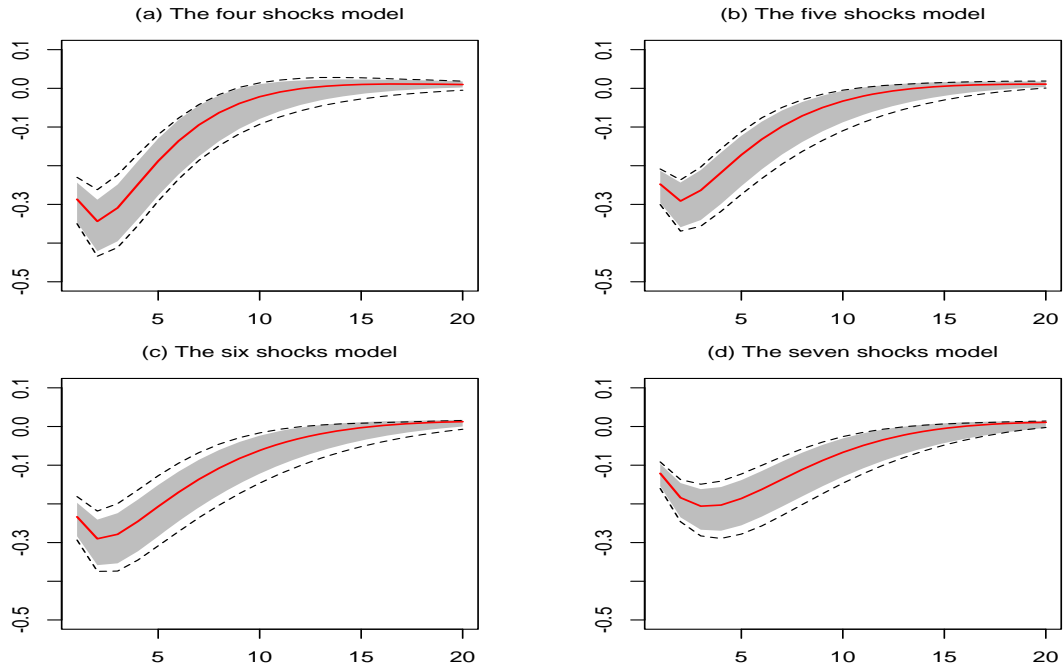
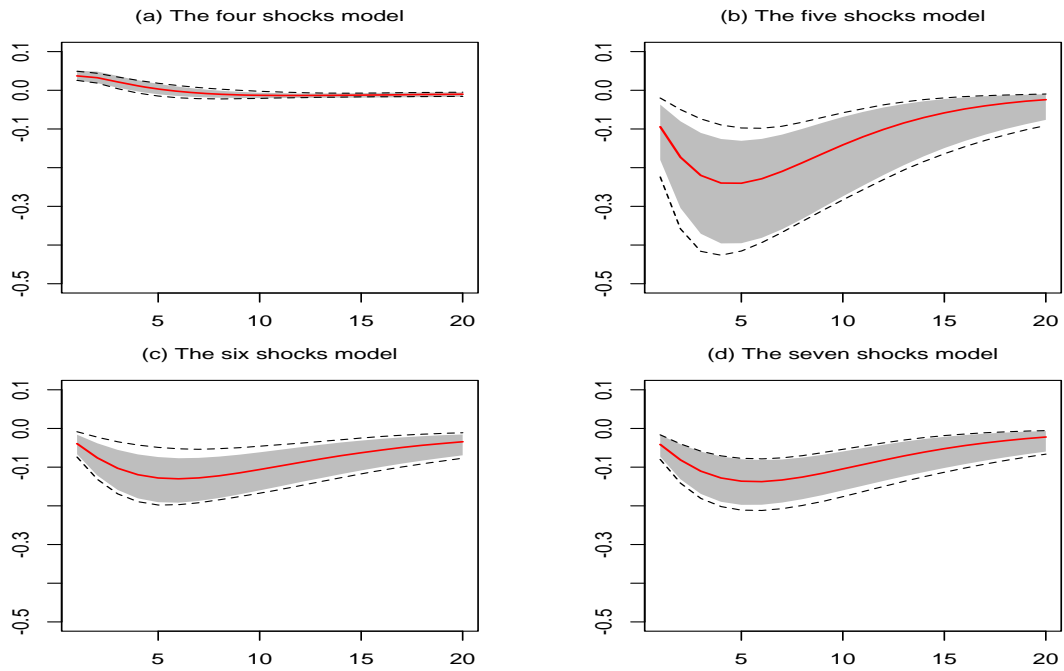
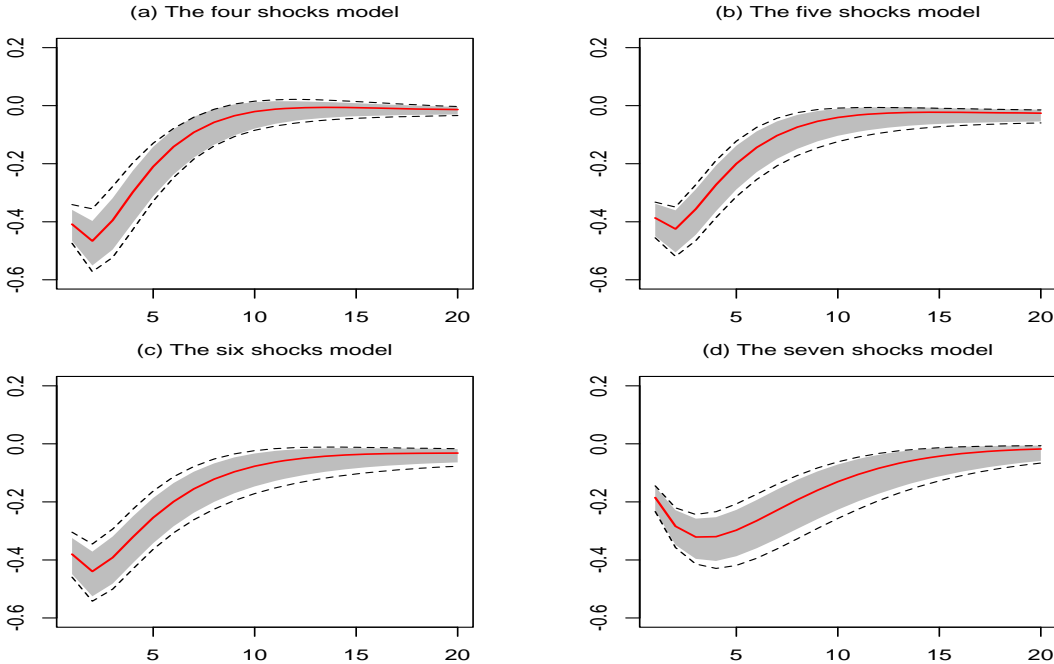


Figure S6. Response of wage to a monetary policy shock



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

Figure S7. Response of consumption to a monetary policy shock



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

Figure S8. Response of output to an investment shock

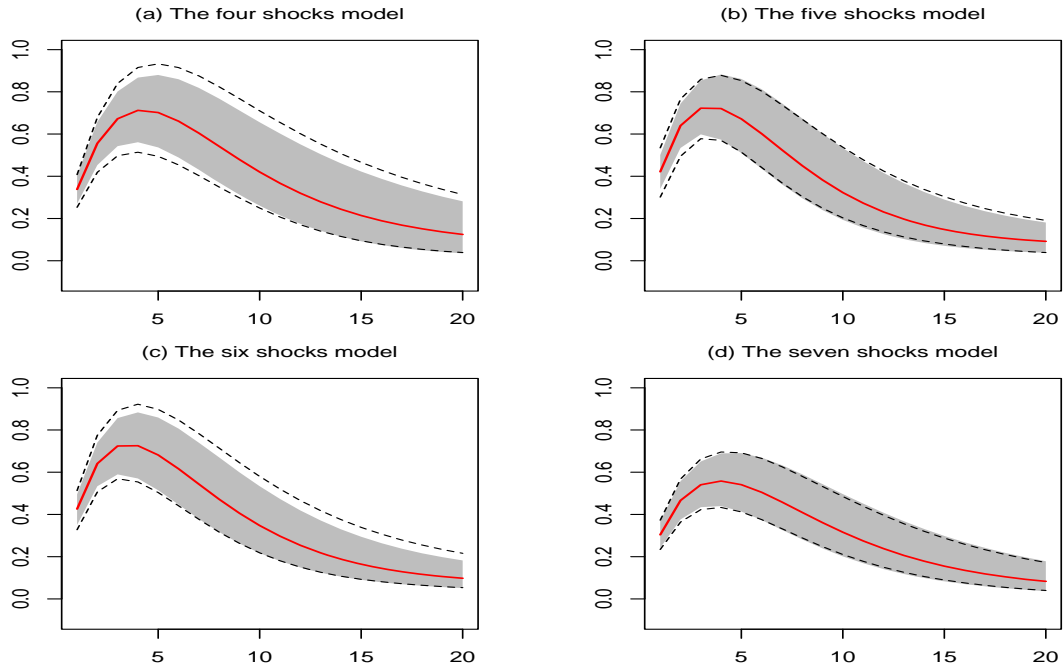
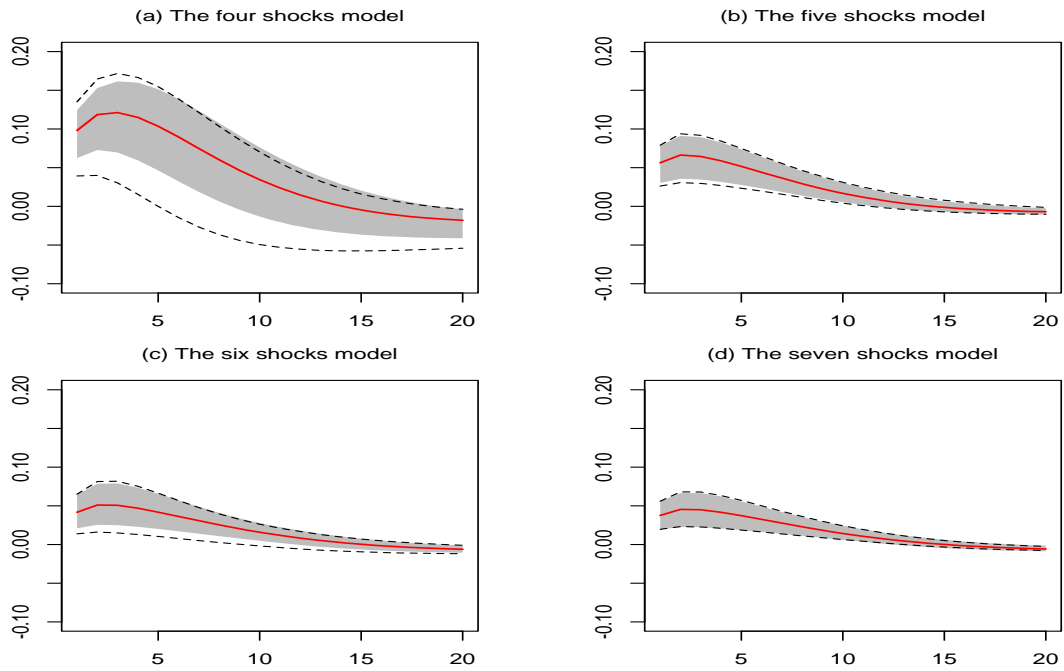


Figure S9. Response of inflation to an investment shock



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

Figure S10. Response of interest rate to an investment shock

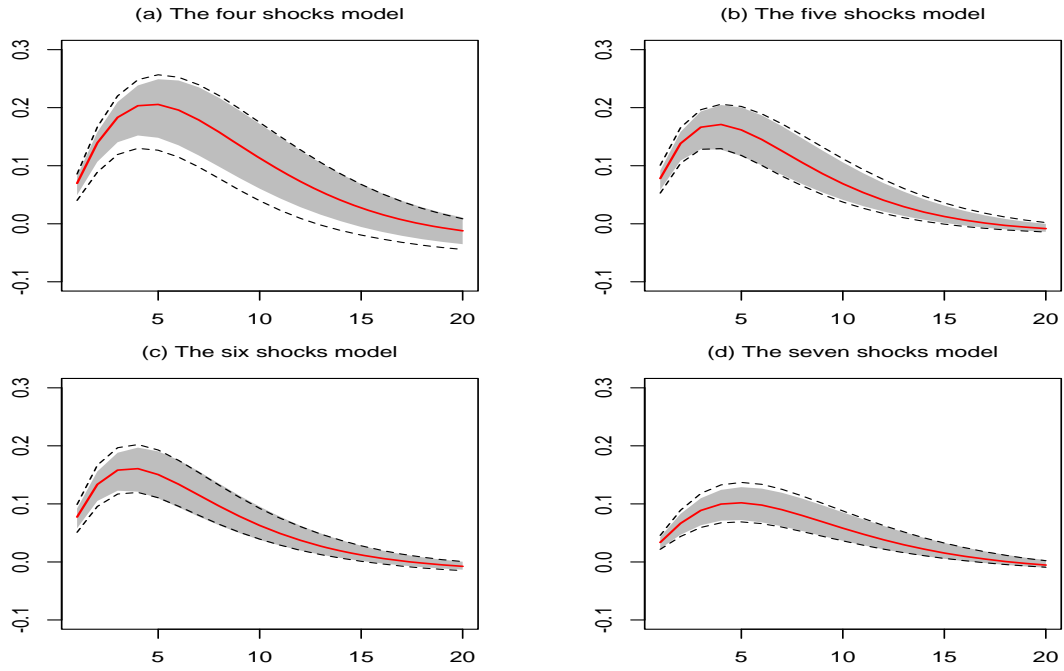
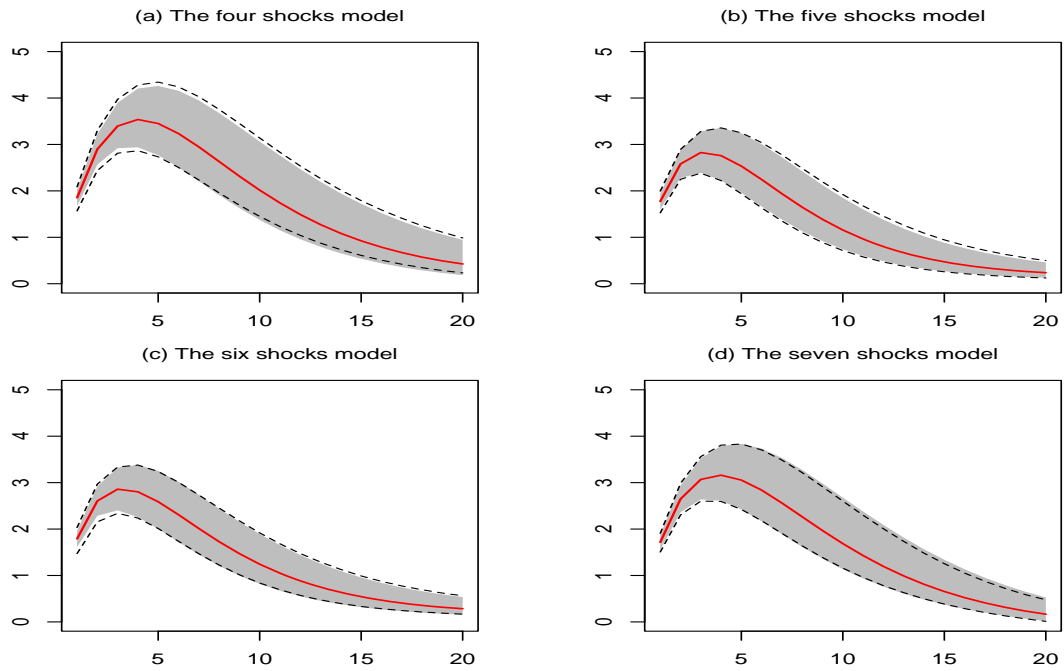


Figure S11. Response of investment to an investment shock



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

Figure S12. Response of hours worked to an investment shock

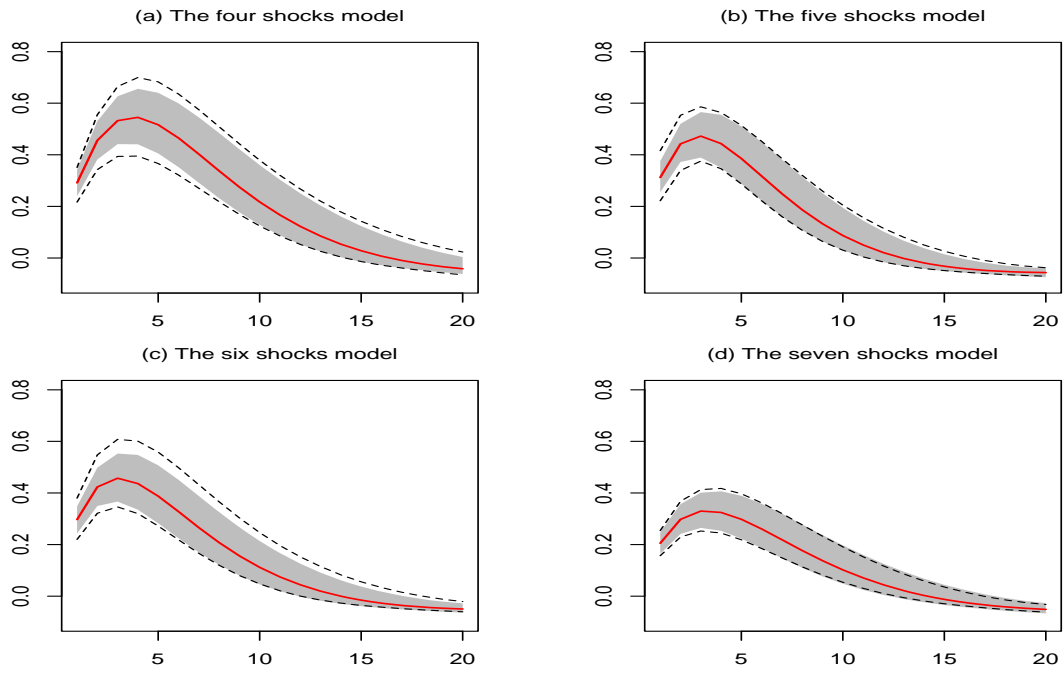
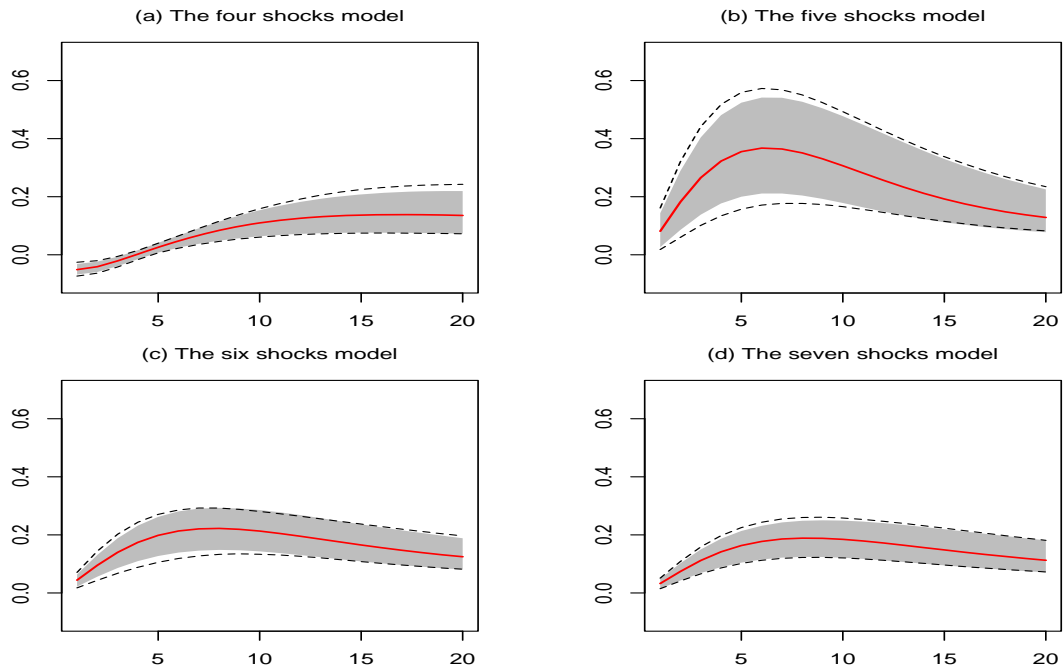
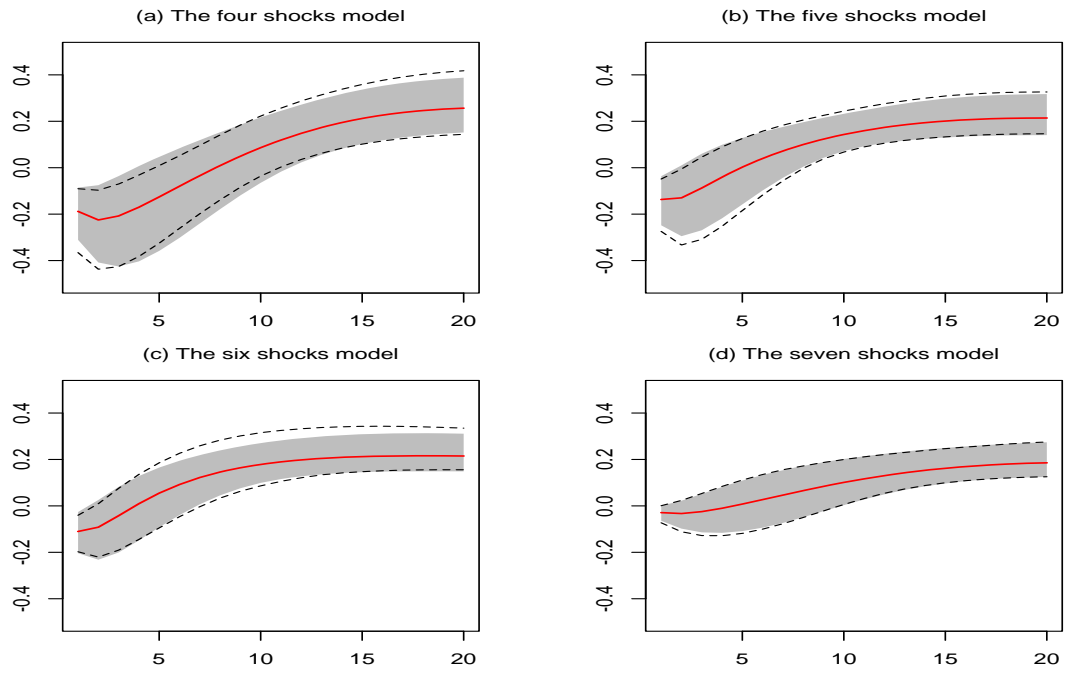


Figure S13. Response of wage to an investment shock



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

Figure S14. Response of consumption to an investment shock



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

Figure S15. Response of output to an exogenous spending shock

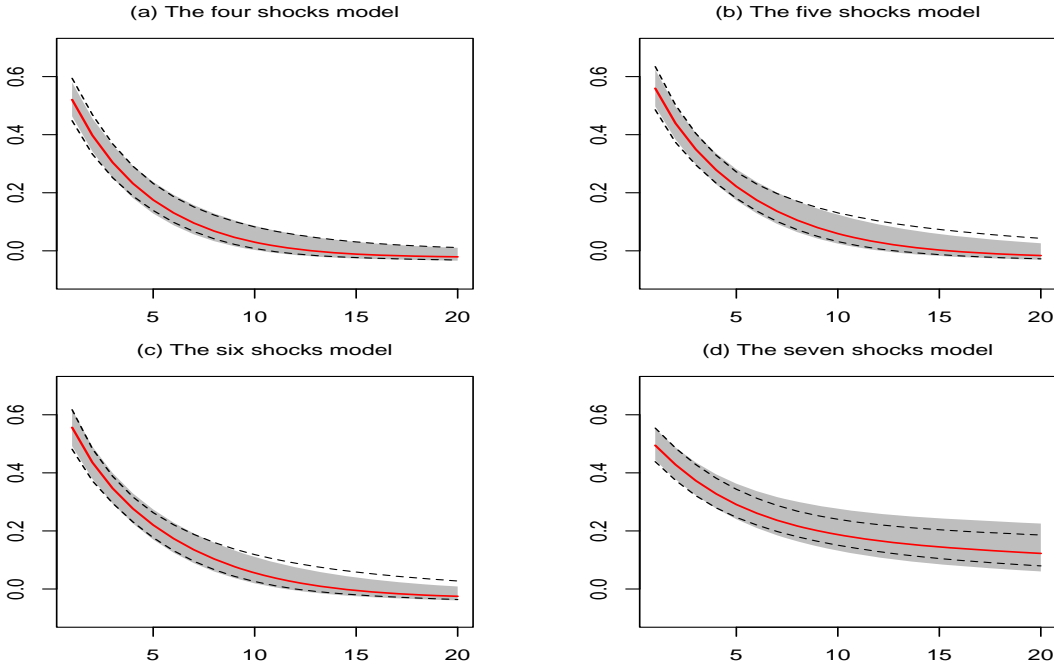
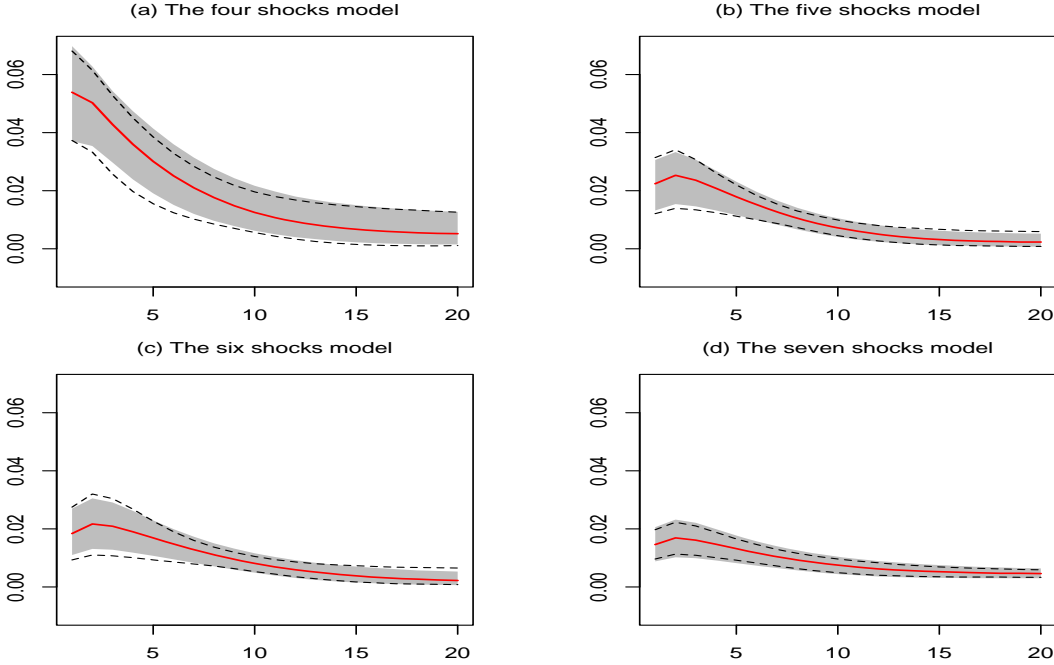


Figure S16. Response of inflation to an exogenous spending shock



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

Figure S17. Response of interest rate to an exogenous spending shock

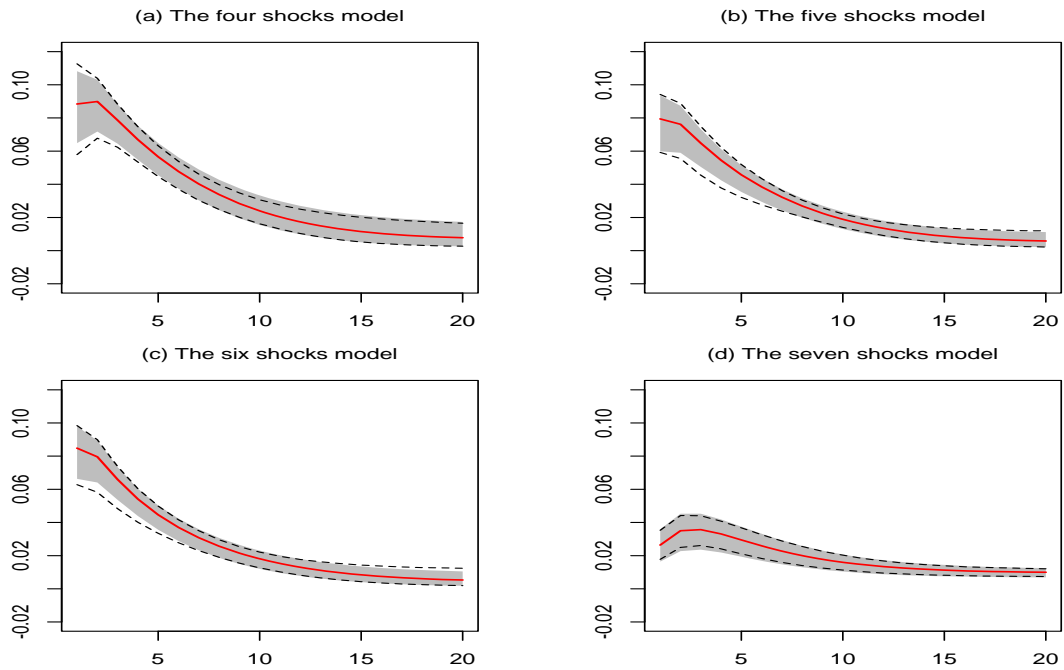
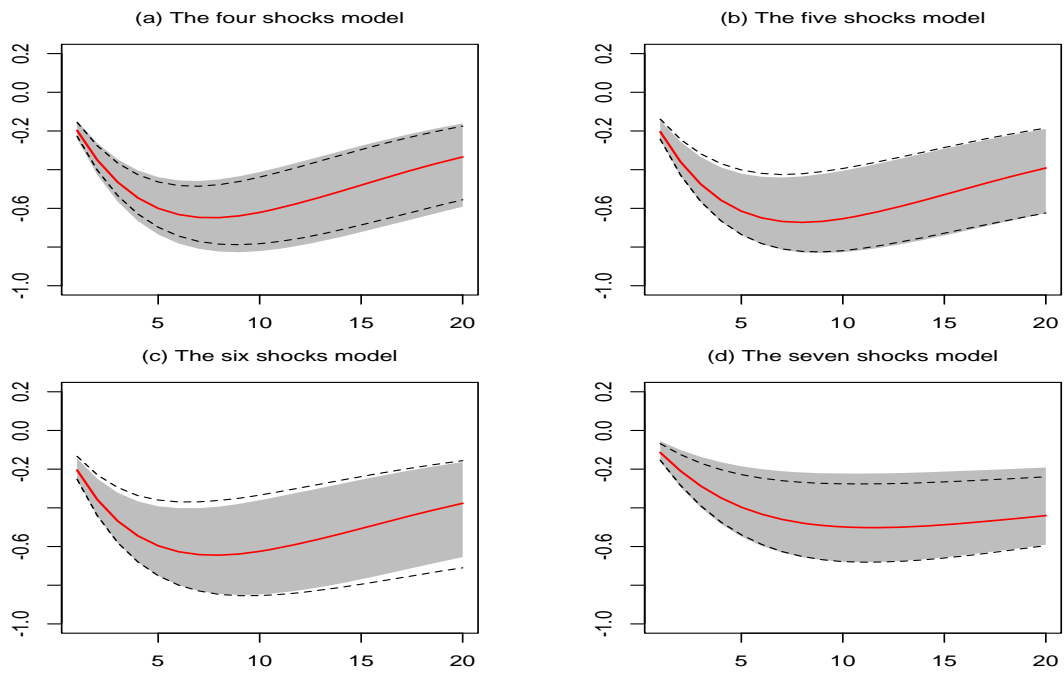


Figure S18. Response of investment to an exogenous spending shock



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

Figure S19. Response of hours worked to an exogenous spending shock

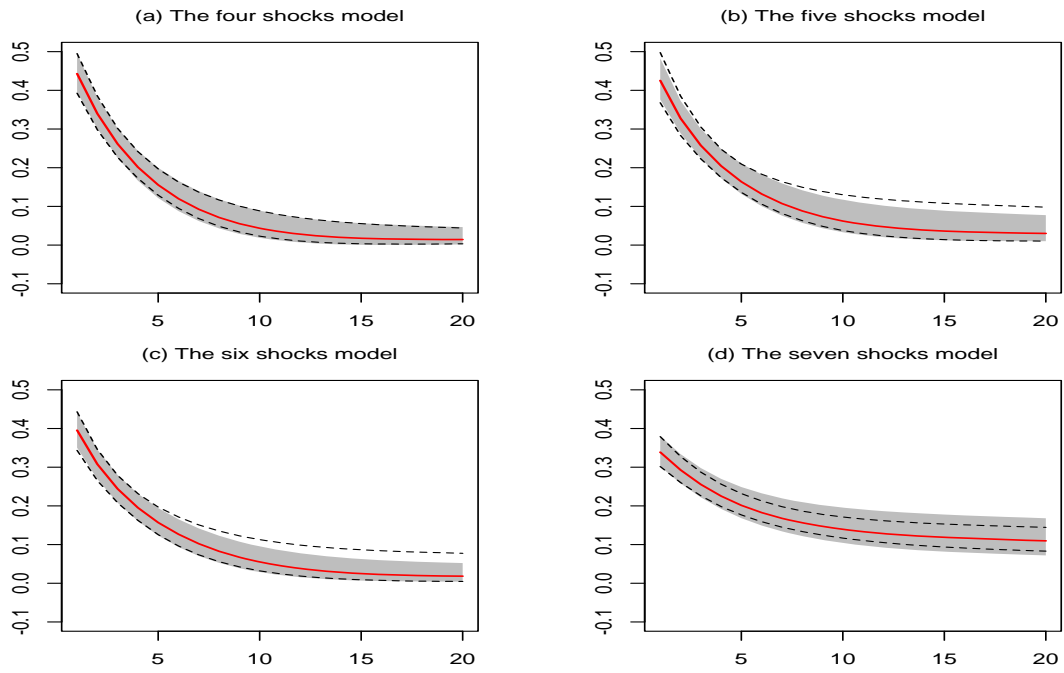
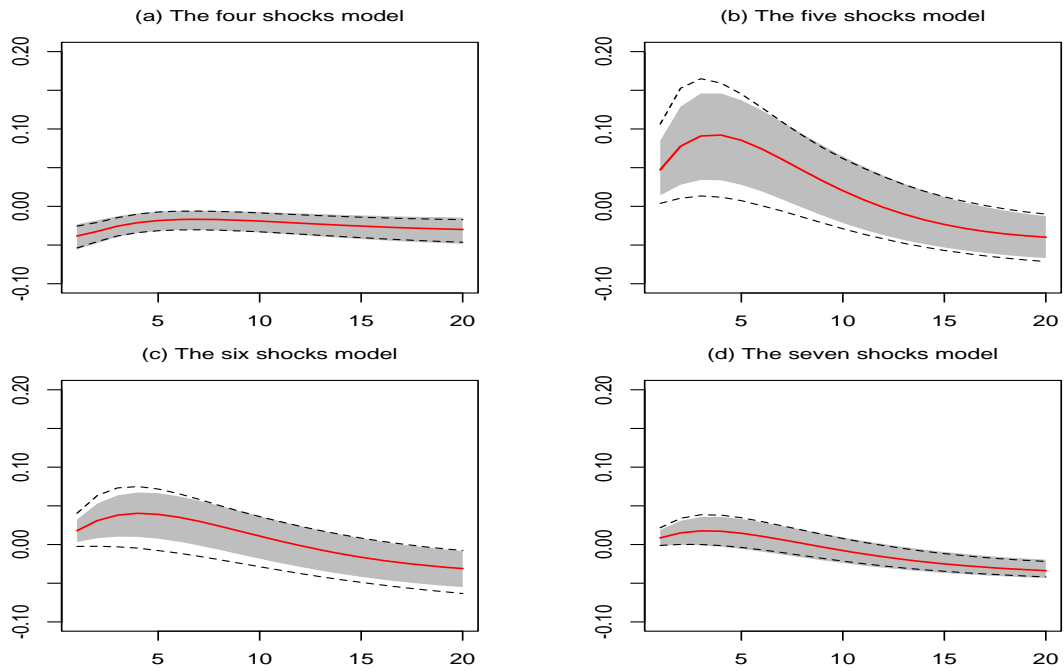
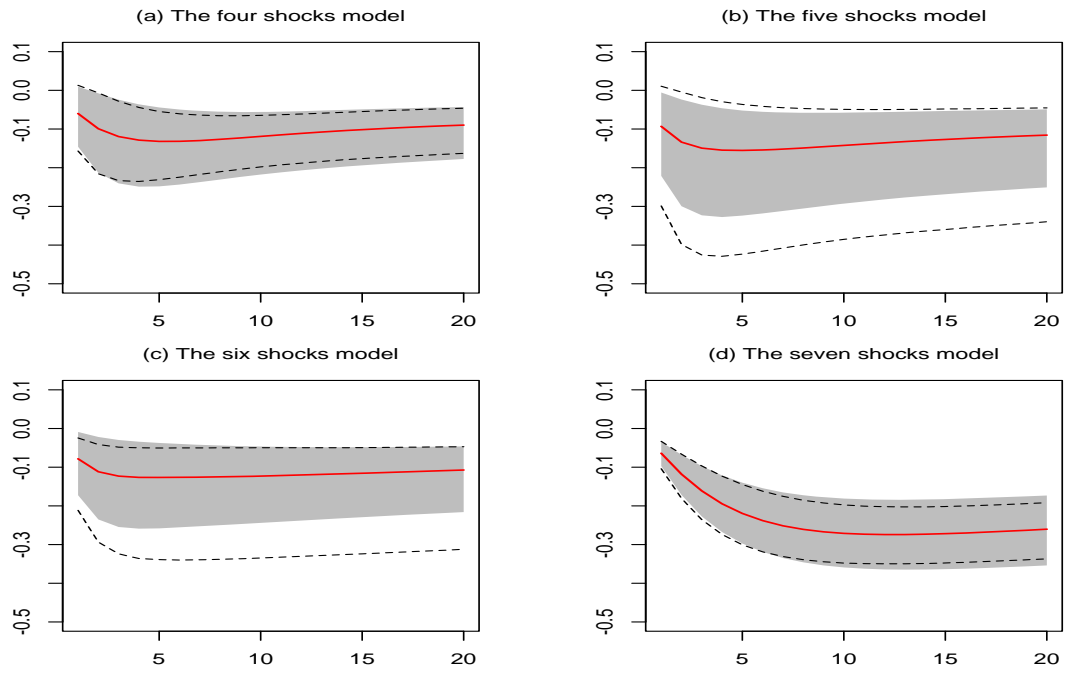


Figure S20. Response of wage to an exogenous spending shock



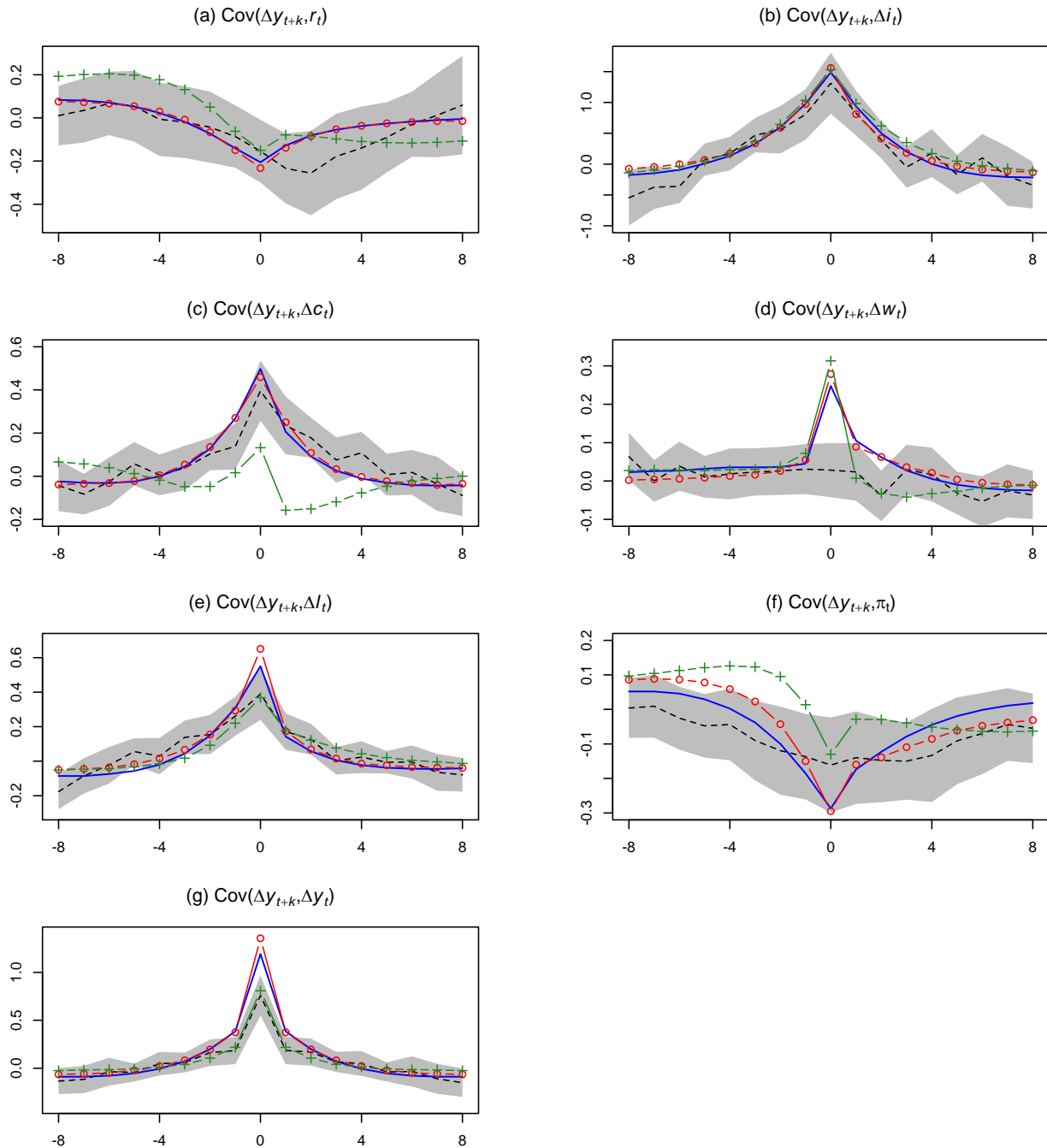
Note. See Figure 1. The variables are not annualized.

Figure S21. Response of consumption to an exogenous spending shock



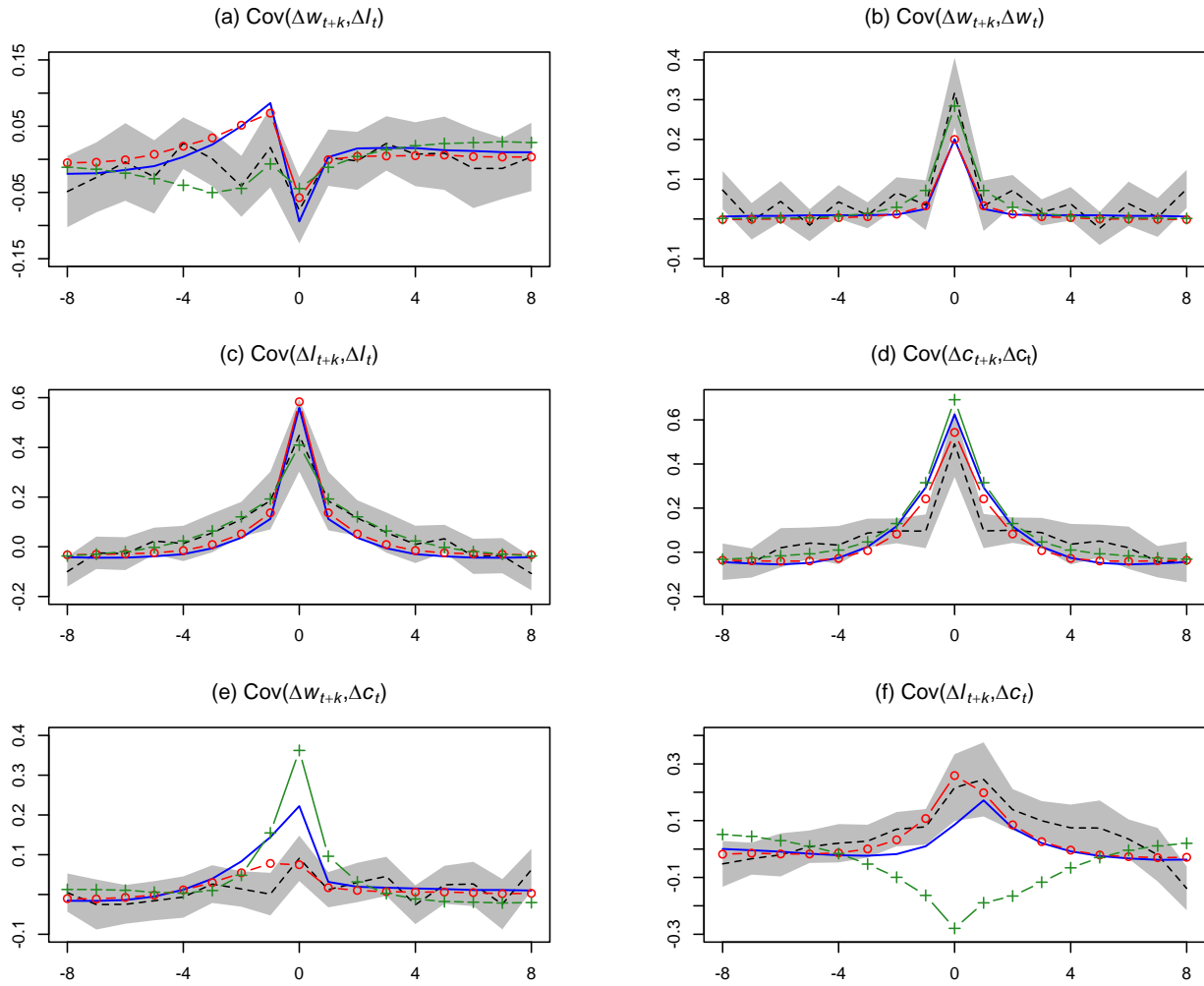
Note. See Figure 1. The variables are not annualized.

Figure S22. Cross covariances involving the output variable (four shocks)



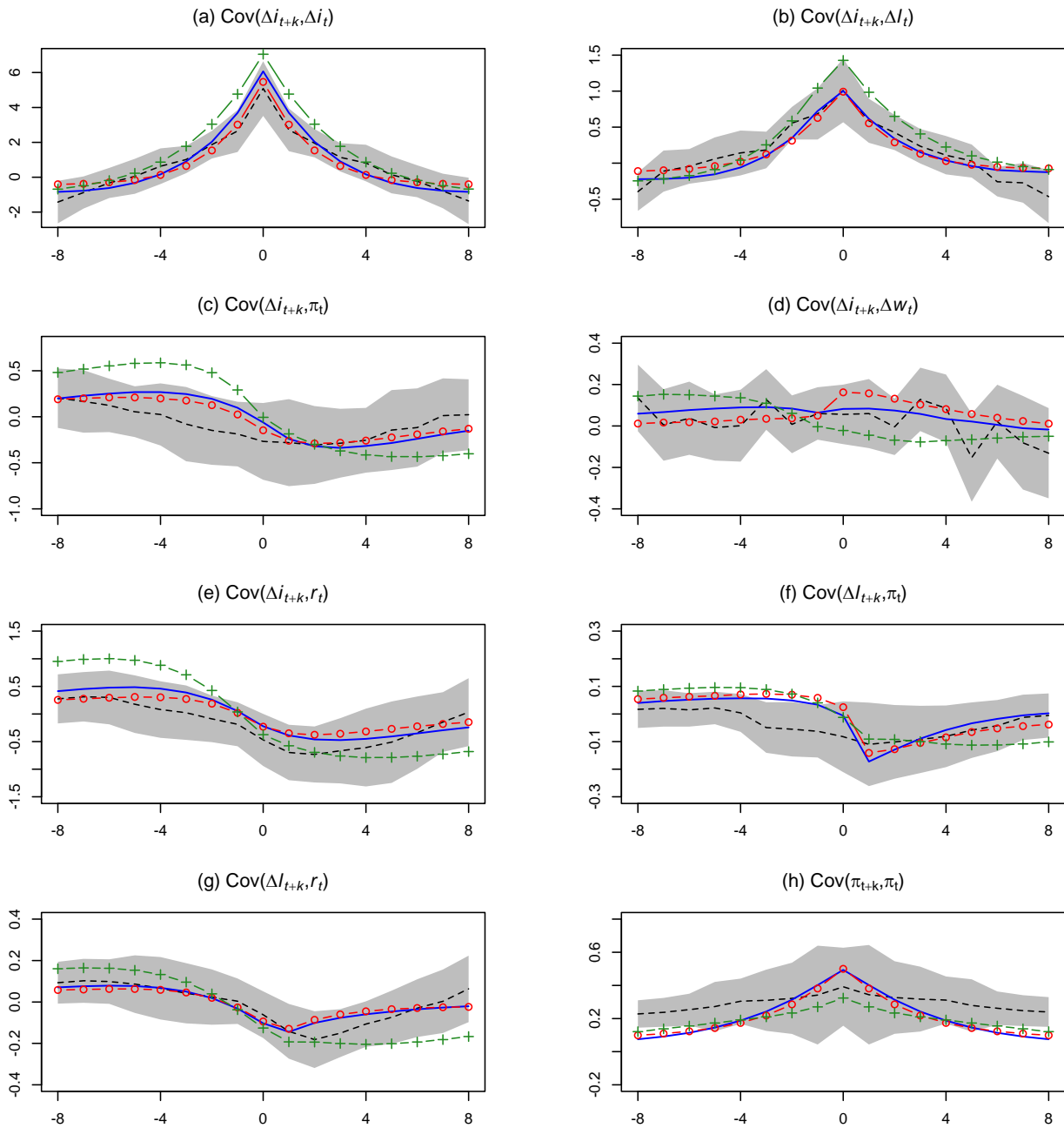
Note. The solid lines, lines with open circles and plus signs are computed using the posterior means obtained under the default, first and second alternative specifications, respectively. The dashed line is the sample cross covariance. The shaded area corresponds to 90% confidence intervals, computed using the Matlab command hac with a quadratic spectral kernel and AR(1) pre-whitening where the bandwidth is determined using Andrews' (1991) method.

Figure S23. Cross covariances among consumption, real wage and hours worked (four shocks)



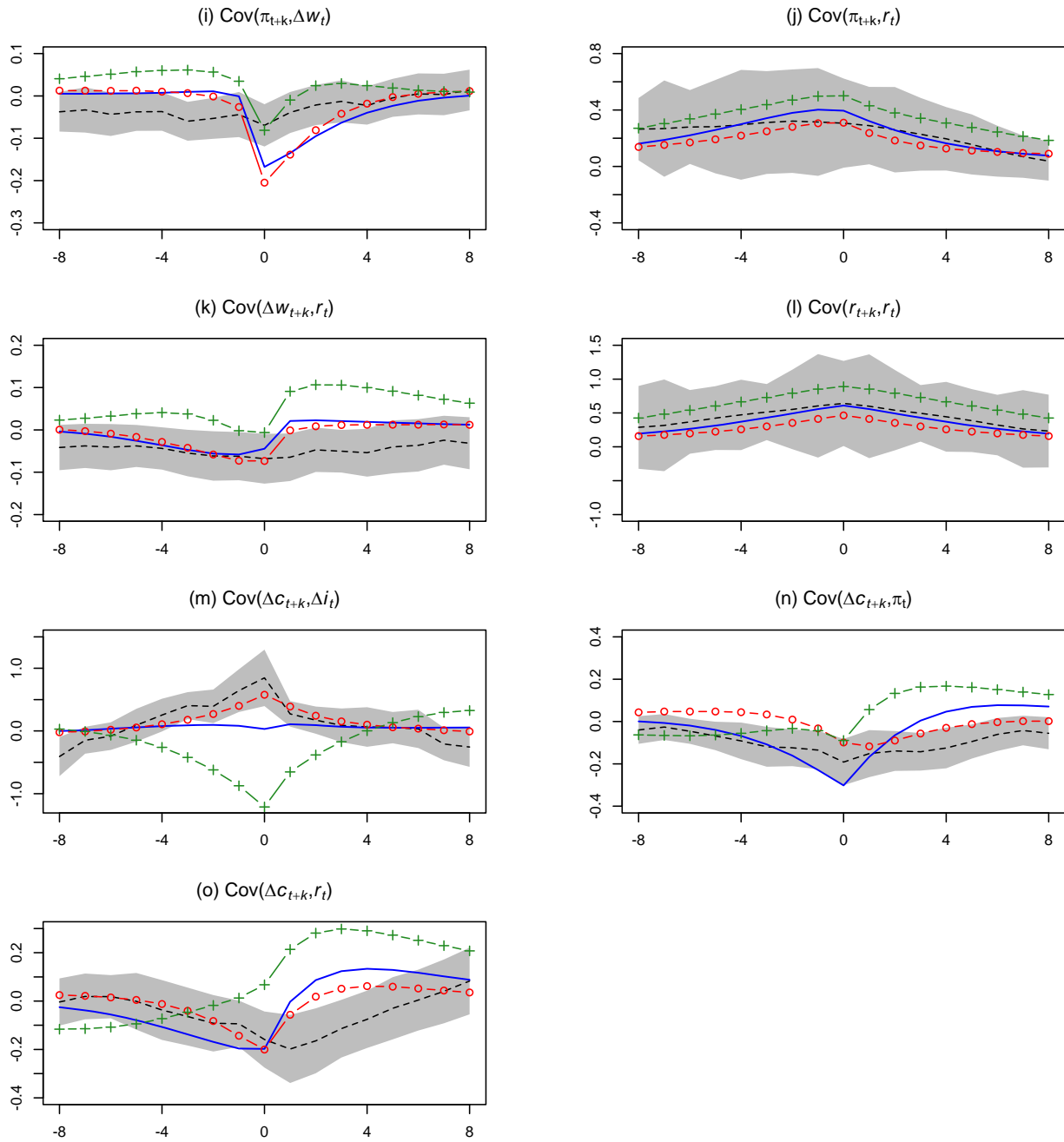
Note. See Figure S22.

Figure S24. Other cross covariances (four shocks)



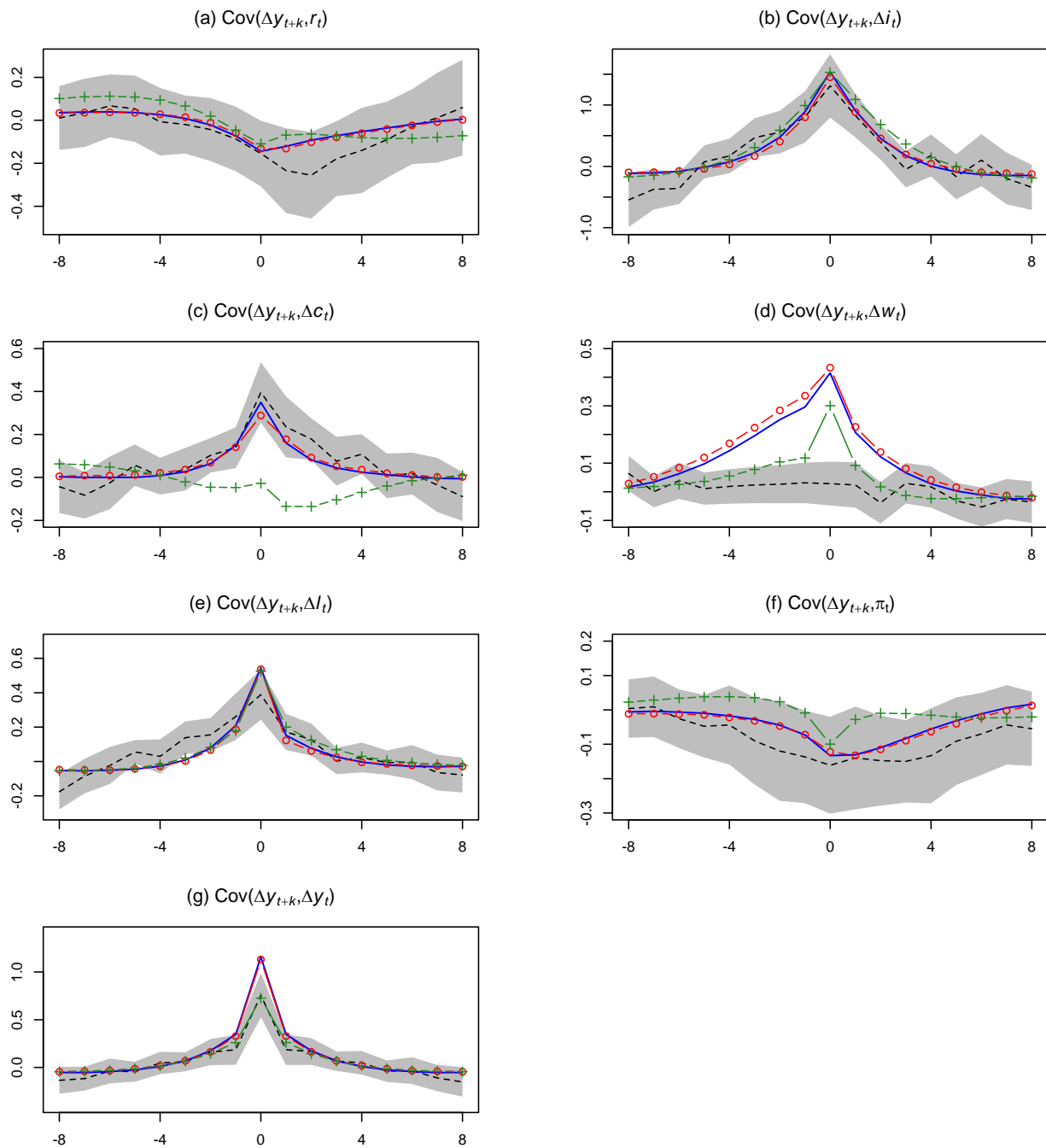
Note. See Figure S22.

Figure S24 (cont'd). Other cross covariances (four shocks)



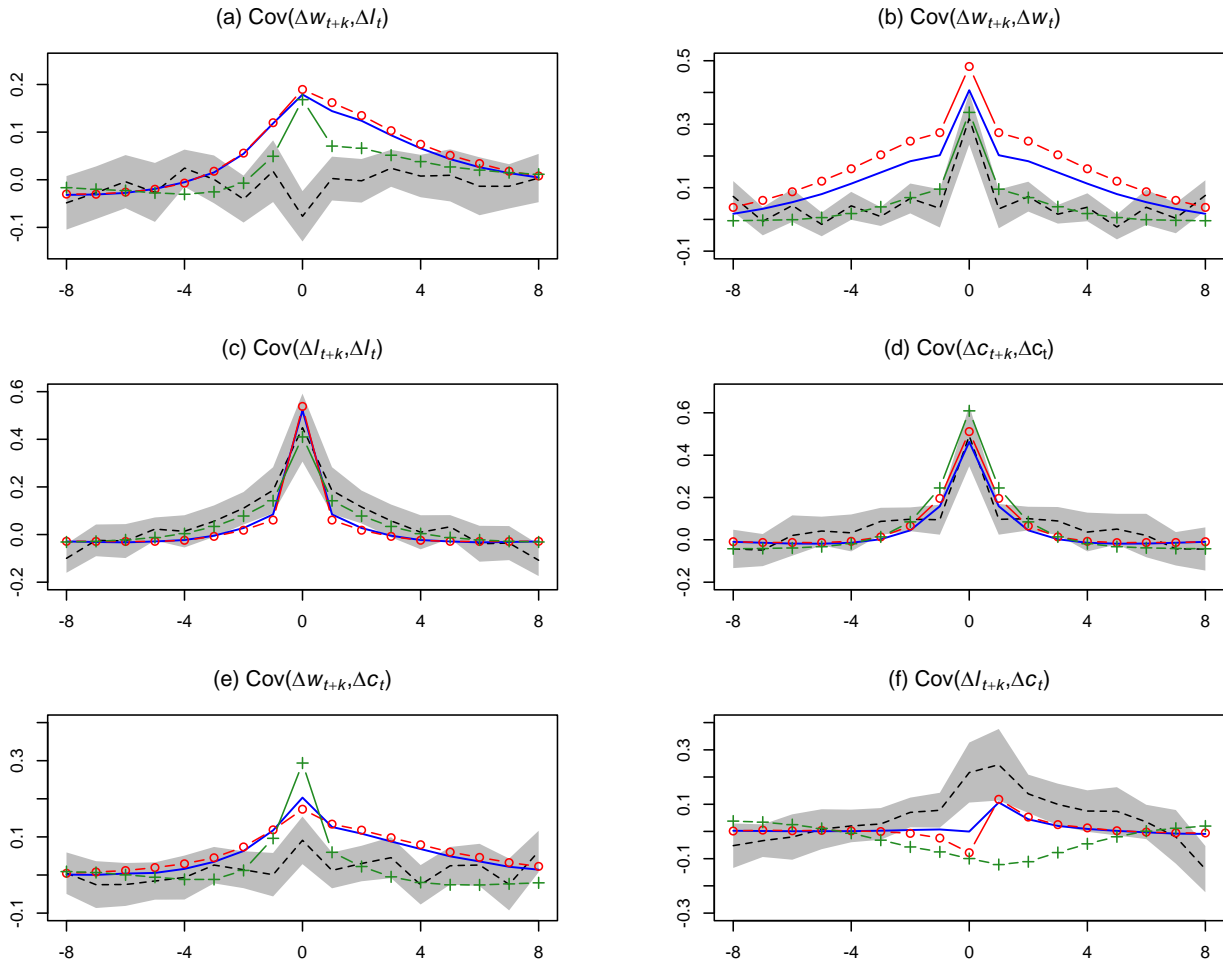
Note. See Figure S22.

Figure S25. Cross covariances involving the output variable (five shocks)



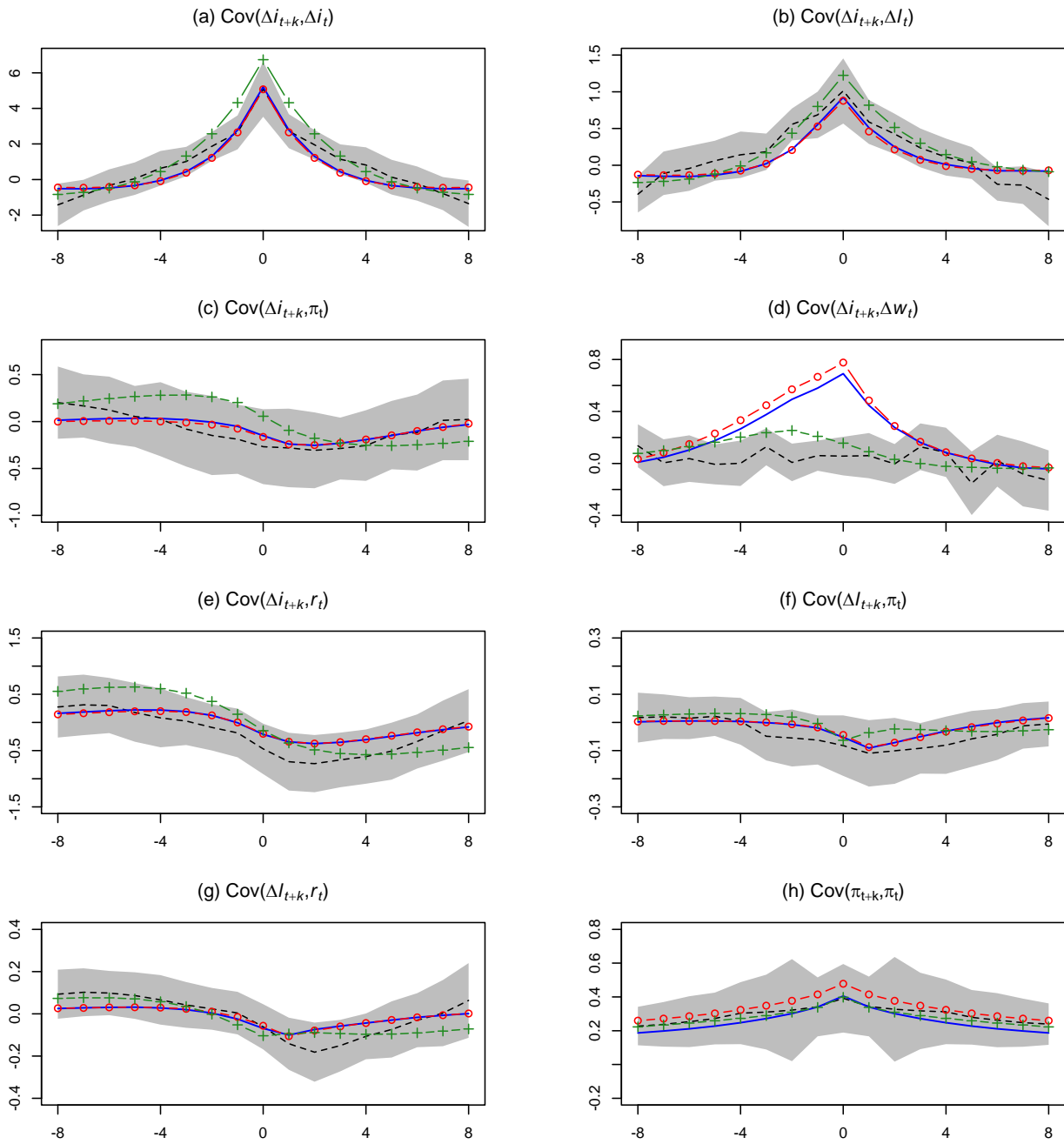
Note. See Figure S22.

Figure S26. Cross covariances among consumption, real wage and hours worked (five shocks)



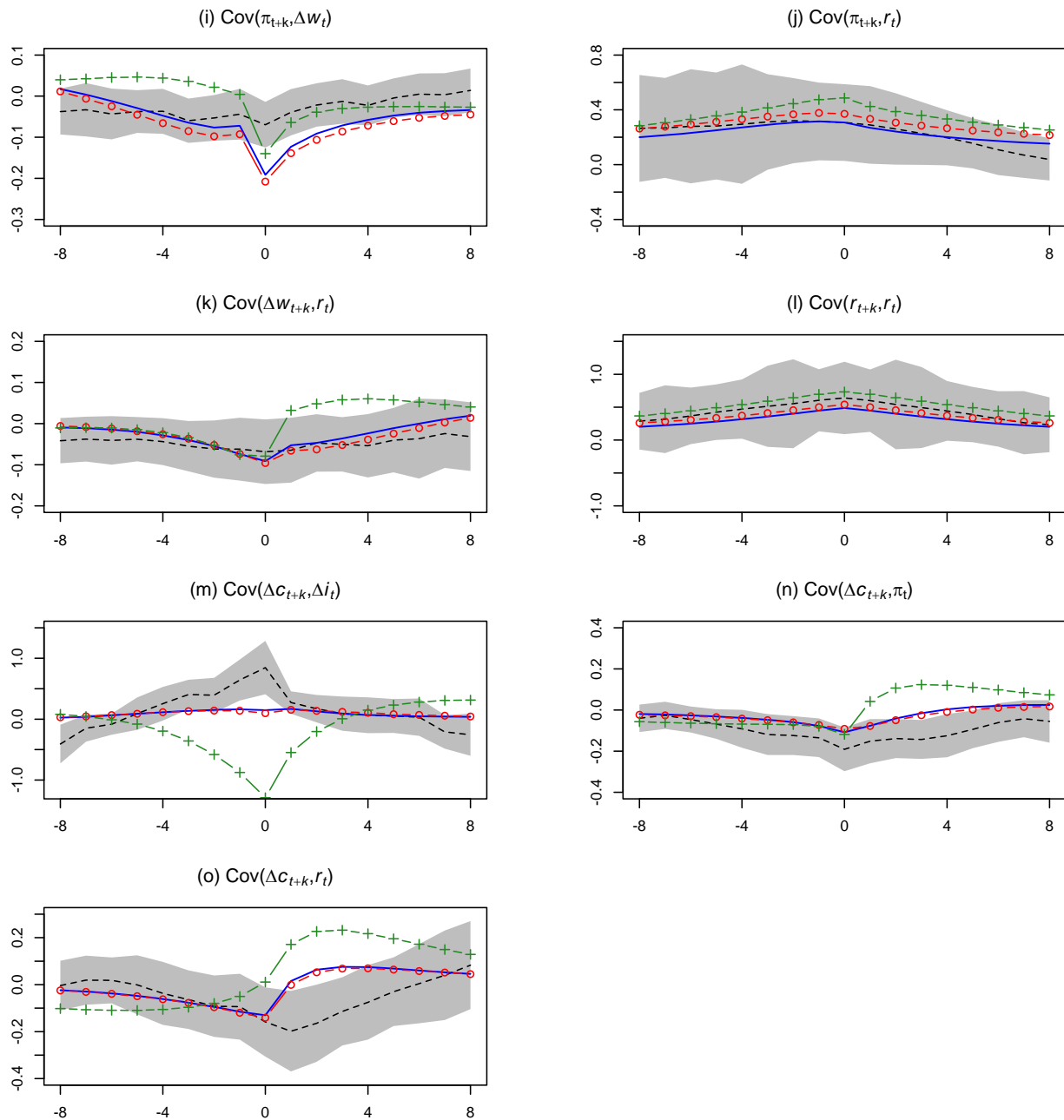
Note. See Figure S22.

Figure S27. Other cross covariances (five shocks)



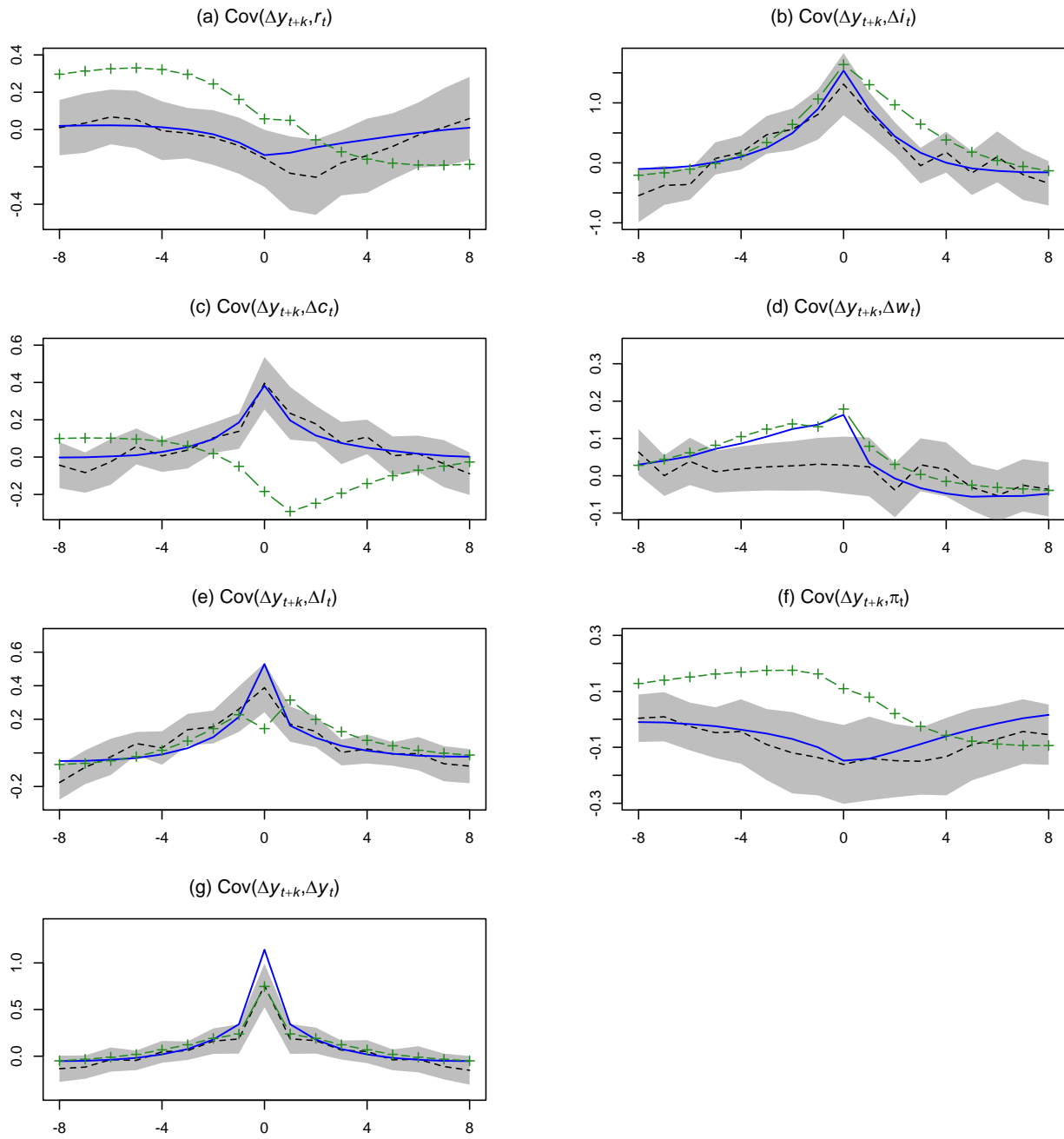
Note. See Figure S22.

Figure S27 (cont'd). Other cross covariances (five shocks)



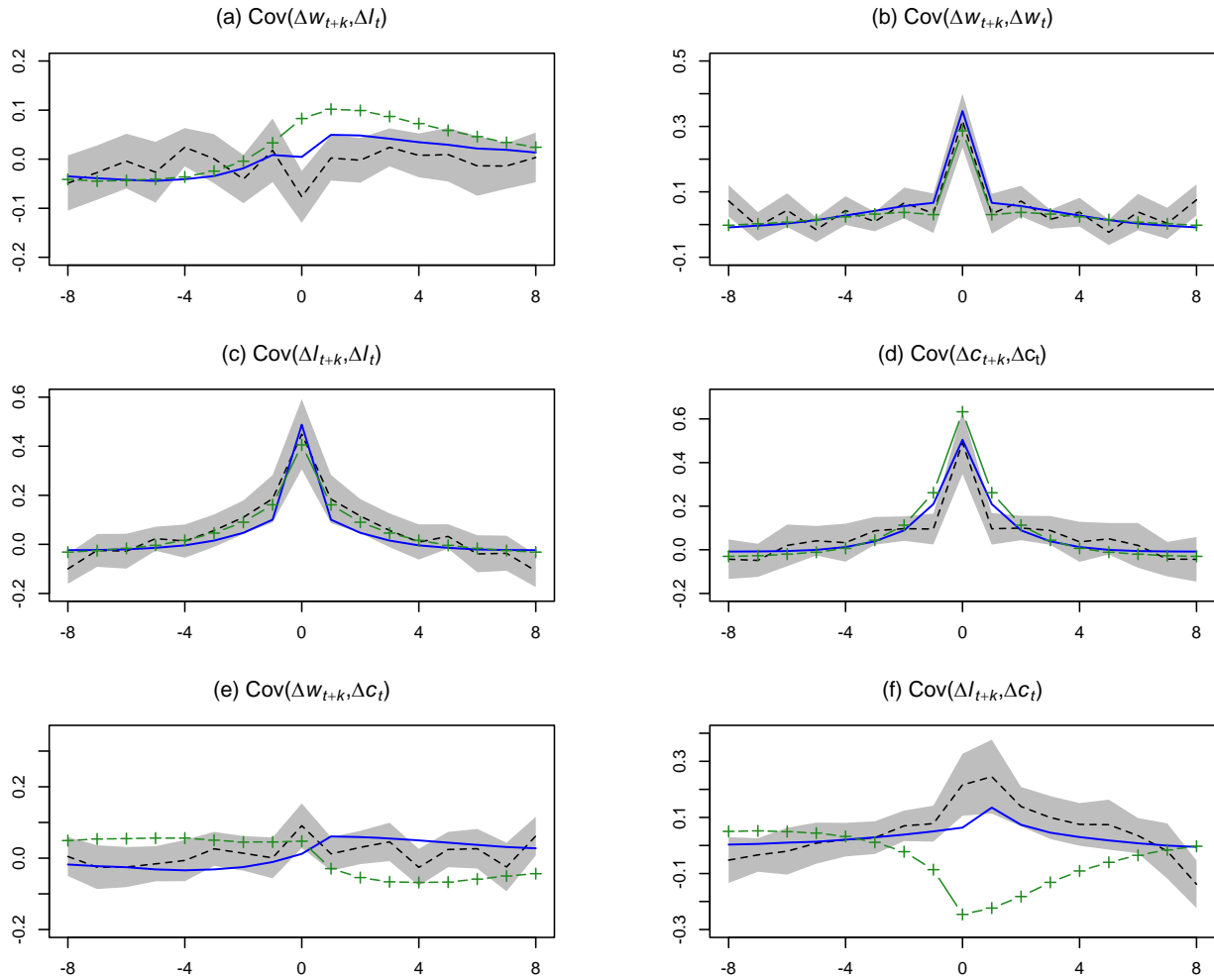
Note. See Figure S22.

Figure S28. Cross covariances involving the output variable (six shocks)



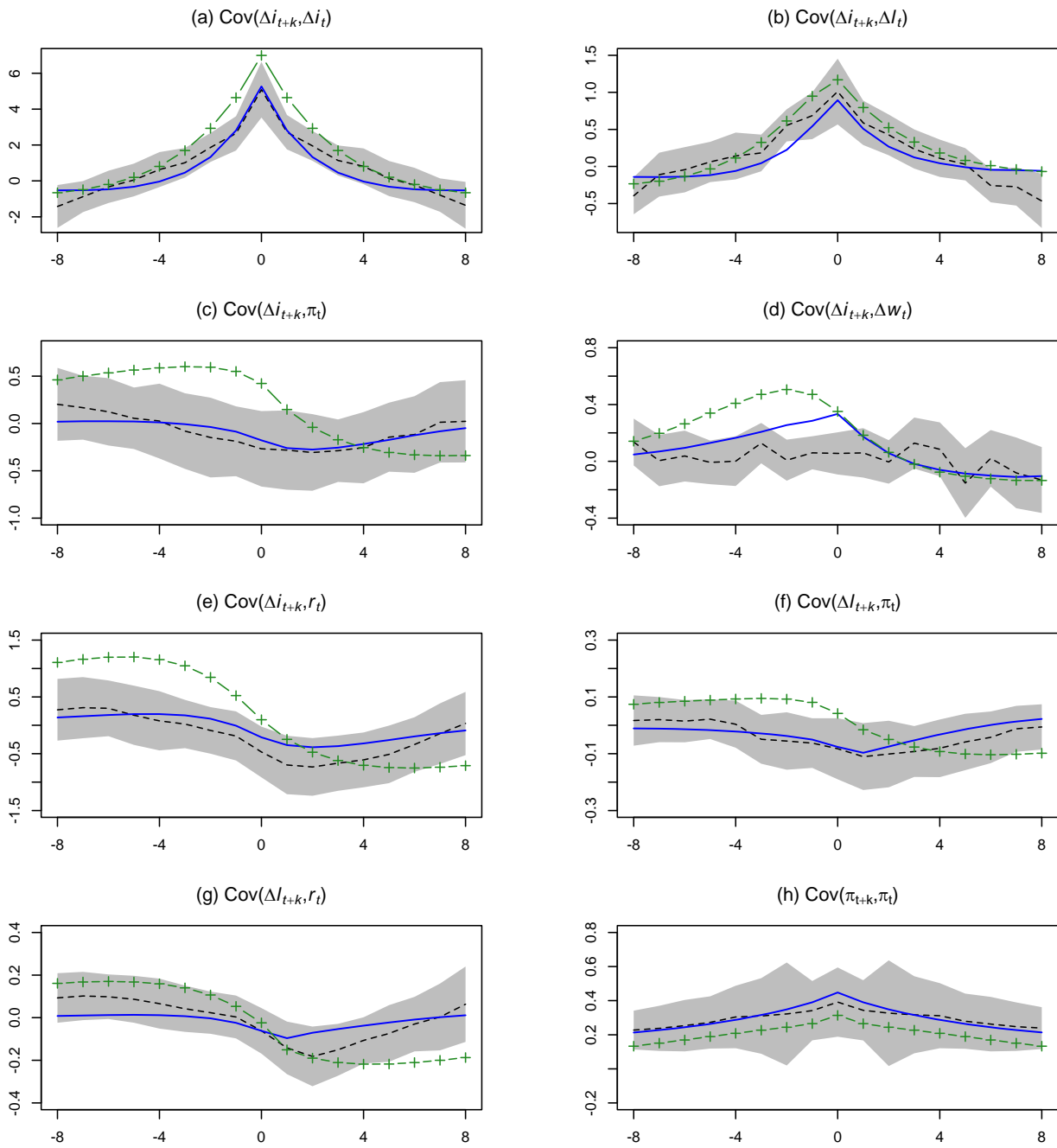
Note. See Figure S22.

Figure S29. Cross covariances among consumption, real wage and hours worked (six shocks)



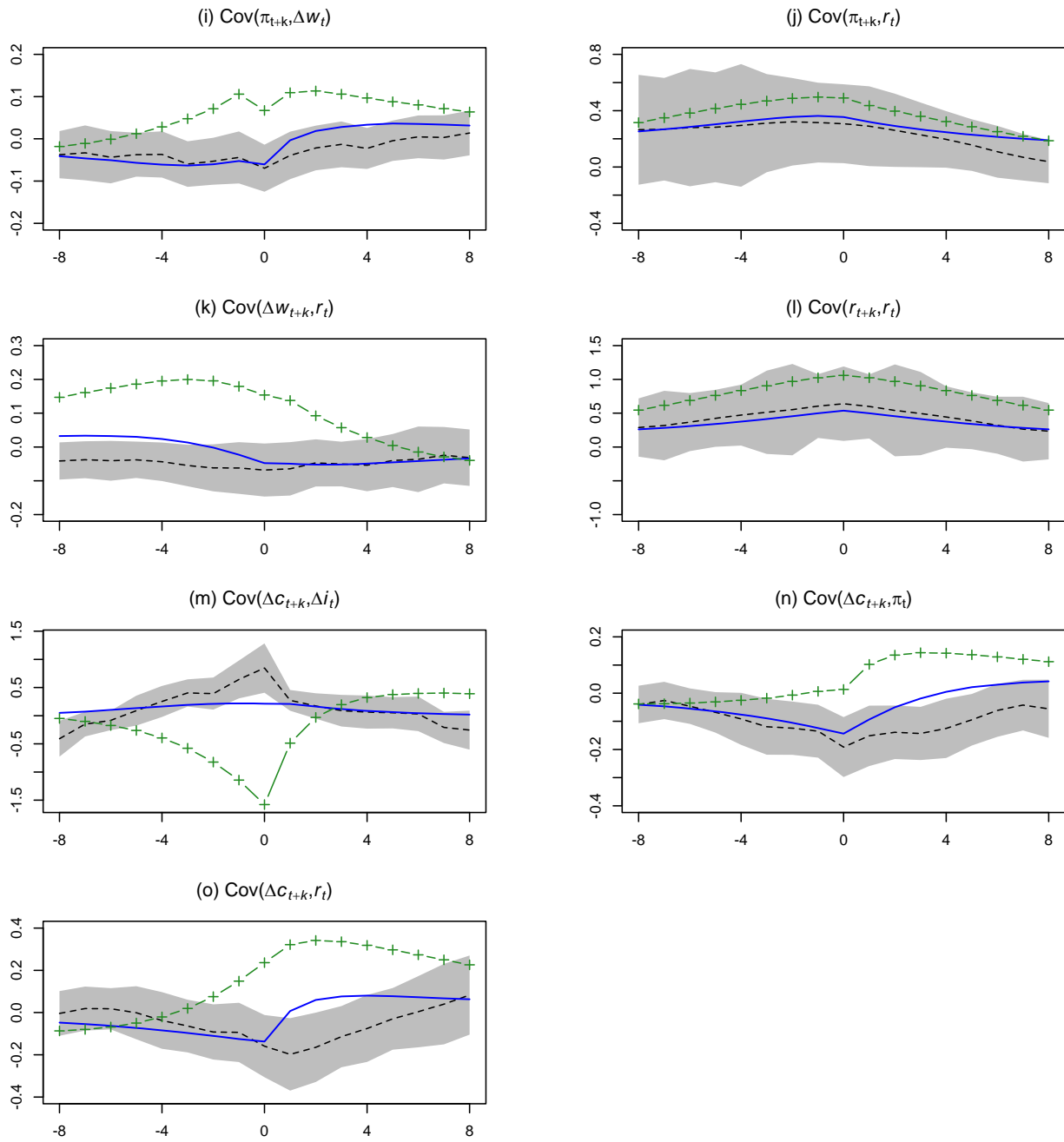
Note. See Figure S22.

Figure S30. Other cross covariances (six shocks)



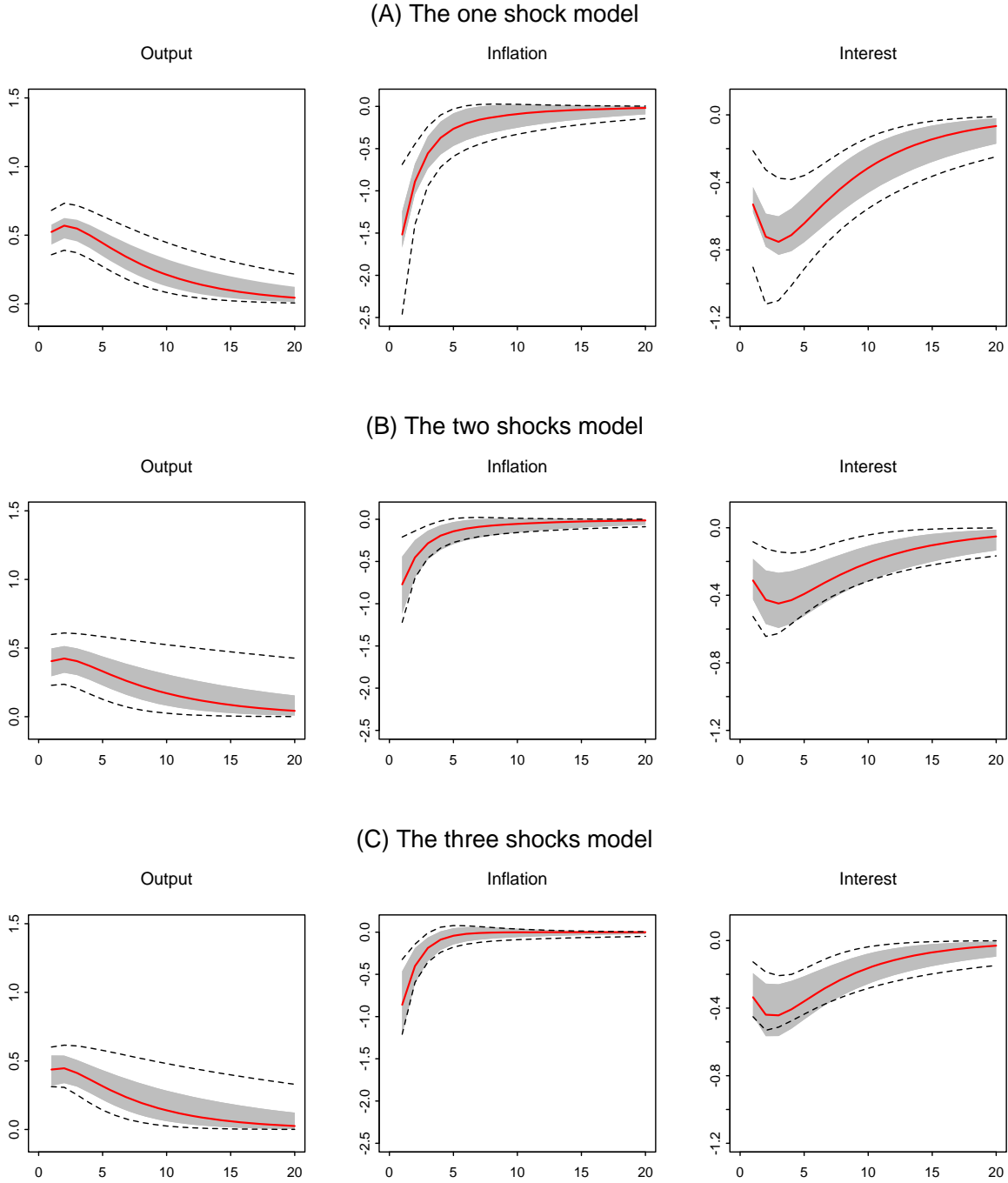
Note. See Figure S22.

Figure S30 (cont'd). Other cross covariances (six shocks)



Note. See Figure S22.

Figure S31. 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.