CONFRONTING MODEL MISSPECIFICATION IN MACROECONOMICS

DANIEL F. WAGGONER AND TAO ZHA

ABSTRACT. We estimate a Markov-switching mixture of two familiar macroeconomic models: a richly parameterized DSGE model and a corresponding BVAR model. We show that the Markov-switching mixture model dominates both individual models and improves the fit considerably. Our estimation indicates that the DSGE model plays an important role only in the late 1970s and the early 1980s. We show how to use the mixture model as a data filter for estimation of the DSGE model when the BVAR model is not identified. Moreover, we show how to compute the impulse responses to the same type of shock shared by the DSGE and BVAR models when the shock is identified in the BVAR model. Our exercises demonstrate the importance of integrating model uncertainty and parameter uncertainty to address potential model misspecification in macroeconomics.

I. INTRODUCTION

In this paper we study and estimate a Markov-switching mixture of two familiar macroeconomic models: a medium-scale linearized dynamic stochastic general equilibrium (DSGE) model and a Bayesian vector autoregression (BVAR) model. This

exercise is motivated by policy analysis dealing with situations where there are multiple models and each model may be misspecified. Practical policy discussions often proceed with combining the implications from different models in an informal way. For our exercise to be econometrically coherent and practically implementable, we build on the recent work of Geweke and Amisano (2011) by estimating a time-varying mixture of these two medium-scale macroeconomic models in which the time variation in the weights assigned to the two models follows a Markov-switching process. Our objective is to explore empirical implications of such a Markov-switching mixture model using a standard set of U.S. time series data.

Geweke and Amisano (2011) propose an optimal pool of alternative models, where the pool combines predictive densities of alternative models. Predictive densities of each model in the pool, as well as the parameters for any parametric model, are taken as given. The optimal pool concerns the estimation of model weights only.

We extend Geweke and Amisano (2011)'s approach in two dimensions. First, we study a formal mixture of heterogenous models by estimating the parameters and the combining weights simultaneously. Second, we allow model weights to switch between two regimes to explore the possibility that the importance of a particular model may change over time. To make our exercise relevant and at the same time feasible to implement, we focus on only two macroeconomic models that are often used in the literature.

The DSGE and BVAR models studied in this paper are sufficiently heterogenous to yield different economic implications. The DSGE model is tightly parameterized around carefully chosen economic structures, whereas the BVAR model is loosely parameterized with relatively few theoretical preconceptions. The mixture of these two models enables us to address both model uncertainty and parameter uncertainty jointly. The Markov-switching feature allows us to study two possible regimes and to determine when model weights change.

Our application yields the following key findings. First, the Markov-switching mixture model dominates both the DSGE and BVAR models according to the marginal

data density (MDD) measure.[1] The estimation of the MDD is computationally demanding, but it is necessary for gauging the fit of the Markov-switching mixture model. For the completeness of our analysis, we also compute the MDD for the mixture model with constant weights. We find that much of the improvement in model fit comes from the constant-weight mixture model. However, allowing regime-dependent model weights not only improves the fit further but also enables us to identify the periods in which the DSGE model or the BVAR model plays an important role.

Second, the estimated posterior probabilities of two regimes reveal that in one regime the DSGE model plays an important role only in the late 1970s and the early 1980s, with the estimated model weight being 0.43. The BVAR model weight is 0.57 in the same regime. In the other regime the BVAR model dominates the DSGE model. This regime covers all other periods than the late 1970s and the early 1980s, including the latest three recessions. Thus, only in certain periods can the DSGE model become an important factor for the improvement in model fit.

Third, the estimated Markov-switching mixture model is used to filter the data for estimation of the DSGE model. Since the regime in which the BVAR model dominates covers 76% of the sample, the mixture model effectively discounts the observations in this regime when the DSGE parameters are estimated. Moreover, in the periods when the DSGE model weight is significant, the data is partially discounted due to the continuing influence of the BVAR model. As a result, the estimates of the DSGE parameters differ considerably from those where the DSGE model is estimated alone with the full sample. We use the impulse responses to a capital depreciation shock in the DSGE model as an example to show that both the magnitude and the uncertainty about these impulse responses are substantially larger than those where the DSGE model is estimated alone over the full sample.

Fourth, we show how to compute the impulse responses to the same type of shock when both individual models in the mixture are structural. In our case, we set up the BVAR model as a structural model to identify a shock to monetary policy. Since the

---

[1]The term "marginal data density" used in the macroeconomic literature is the same concept as the "marginal likelihood" used in the statistics literature. That is, the MDD is an integral of the prior density times the likelihood function, with both the prior and the likelihood being proper probability density functions. When the two models are compared, the Bayes factor defined as the ratio of the two MDDs is often used to determine which model fits to the data better.

responses *conditional on each model* are tightly estimated, economic implications about the effect of a monetary policy shock differ from the DSGE model to the BVAR model. Once we recognize that each model may be misspecified and take model uncertainty into account, we show that the estimate of the effect of a monetary policy shock is smaller, but the probability bands around the estimate are larger than what is implied from the BVAR model. We discuss how the effect of a monetary policy shock changes from one regime to the other. Thus, our application illustrates an effective way of using the Markov-switching mixture model for structural analysis.

The rest of the paper is organized as follows. Section II provides a brief literature review. In Section III we lay out a general framework. In Section IV we apply the general framework to our specific case study and estimate a Markov-switching mixture model of the DSGE and BVAR models. Section V reports different measures of fit for the DSGE model, the BVAR model, the simple mixture model, and the Markov-switching mixture model. In Section VI we show how to use the estimated Markov-switching mixture model as a data filter for estimating the impulse responses to a capital depreciation shock in the DSGE model. In Section VII we show how to perform a full structural analysis when a monetary policy shock is identified in the BVAR model. Concluding remarks are made in Section VIII.

II. LITERATURE REVIEW

The key contribution of this paper is an application of a general Markov-switching mixture framework to two medium-scale macroeconomic models. The general framework consists of two components. As there exists a large strand of literature on each component, we focus on only a small handful of references that are most relevant to this paper. The first component is a simple mixture of alternative models. The predictive density in the mixture is a linear combination of predictive densities of individual models. The idea of combining point forecasts can be traced back to Bates and Granger (1969) and Diebold (1991). In a recent work, Geweke and Amisano (2011) propose a method of pooling individual models by combining predictive densities instead of point forecasts.[2] Fisher and Waggoner (2010) extend a pool of models emphasized by Geweke

---

[2]See Geweke and Amisano (2011) for a long list of other references on forecast combinations as well as on predictive distributions.

and Amisano (2011) to a mixture of models and argue for the Bayesian approach to estimation of the mixture model.

The second component is a Markov-switching process applied to model weights. An earlier paper by Harrison and Stevens (1976) introduces a Markov process in a mixture model with an emphasis on changes in the noise and disturbance matrices. West and Harrison (1997) allow model weights to change over time in dynamic forecasting exercises. The modern analysis of Markov-switching dynamic models can be found in, for example, Hamilton (1989), Chib (1996), Kim and Nelson (1999), and Frühwirth-Schnatter (2006).

The literature on model misspecification is large, and different approaches have been taken to confront this issue. Del Negro and Schorfheide (2004) address potential DSGE model misspecification by introducing the prior implied by a DSGE model into a BVAR model. Further discussions about using the DSGE-VAR approach and about how it is related to Ingram and Whiteman (1994) can be found in Del Negro and Schorfheide (2009) and Del Negro and Schorfheide (Forthcoming). Hansen and Sargent (2001) and Brock, Durlauf, and West (2003) provide the robustness-control framework to address model misspecification. Using this idea, Cogley and Sargent (2005) study an economy in which agents, facing model uncertainty, compute the posterior odds ratios over three models and make decisions by Bayesian model averaging. Our analysis centers on a Markov-switching mixture of the DSGE and BVAR models and is mostly related to Geweke and Amisano (2011).

III. A general econometric framework

To integrate model uncertainty and parameter uncertainty, we use the Bayesian approach. In our general setup we allow the weights in a linear combination of predictive densities of *individual* heterogenous models to vary across regimes. We assume that there are a total of n models (M_i for $i = 1, \ldots, n$) in the study and that the observed data at time t, y_t^o, is generated from the following predictive density:

$$p\left(y_t \mid Y_{t-1}^o, \Theta, Q, w, s_t\right) = \sum_{i=1}^{n} w_{i,s_t} p\left(y_t \mid Y_{t-1}^o, \Theta_i, M_i\right),$$

where $p\left(y_t \mid Y_{t-1}^o, \Theta_i, M_i\right)$ is the predictive density of y_t conditional on model i, its parameters, and the observed data up to time $t-1$, $Y_{t-1}^o = \left\{y_1^o, \cdots, y_{t-1}^o\right\}$. The super-script "$o$" denotes the observed data. The notation Θ_i represents a vector of parameters for model i. The notation "M_i" in the predictive density function $p\left(y_t \mid Y_{t-1}^o, \Theta_i, M_i\right)$ is not redundant because in the end we compare the marginal data density of model i, denoted by $p\left(Y_T^o \mid M_i\right)$, with the marginal data density of the mixture model, denoted by $p\left(Y_T^o\right)$.

The regime-dependent weight, $w_{i,s_t} \geq 0$, is assigned to model i with $\sum_{i=1}^n w_{i,s_t} = 1$.[3] The regime variable s_t is an unobservable state and follows a Markov process with the transition matrix $Q = [q_{k,j}]$, where $q_{k,j} = \text{Prob}\left[s_t = k \mid s_{t-1} = j\right]$ for $k, j = 1, \ldots, h$. The total number of regimes is h.

Grouping all the parameters together, we have

$$\Theta = \{\Theta_1, \cdots, \Theta_n\}, \ w = \{w_{i,k}\} \text{ for } k = 1, \ldots, h, i = 1, \ldots, n.$$

Since s_t is unobservable, we integrate out s_t to obtain the conditional likelihood as

$$p\left(y_t \mid Y_{t-1}^o, \Theta, Q, w\right) = \sum_{s_t=1}^h \left[p\left(y_t \mid Y_{t-1}^o, \Theta, Q, w, s_t\right) p\left(s_t \mid Y_{t-1}^o, \Theta, Q, w\right)\right].$$

The log likelihood function is thus given by

$$\log p\left(Y_T^o | \Theta, Q, w\right) = \sum_{t=1}^T \log p\left(y_t^o \mid Y_{t-1}^o, \Theta, Q, w\right), \tag{1}$$

where the parameters Θ, Q, and w are to be *estimated jointly*.

A special case of our Markov-switching framework is a mixture model with constant weights. In model comparison we include this simple mixture model to gauge how important a convex combination of predictive densities is in improving model fit.

Given the prior $p\left(\Theta, Q, w\right)$ and the likelihood function (1), we form the posterior density function proportional to the product of the likelihood function and the prior density function:

$$p\left(\Theta, Q, w \mid Y_T^o\right) \propto p\left(Y_T^o \mid \Theta, Q, w\right) p\left(\Theta, Q, w\right). \tag{2}$$

---

[3]In an earlier draft, we impose the restriction that one of the weights w_{i,s_t} is equal to 1 and others are set to 0. The restriction, similar to the approach taken by McCulloch and Tsay (1994), reflects the idea that only one model is operative at a time. Our current setup is more general and encompasses the special case in which one of the weights w_{i,s_t} is restricted to be 1.

IV. Application

We apply the general framework presented in Section III to two types of widely used models, a medium-scale DSGE model and a BVAR model. Since $n = 2$ in our case, we adopt the notation $i \in \{DSGE, BVAR\}$. We focus on only two regimes, so that $h = 2$.

The DSGE model, based on Liu, Waggoner, and Zha (2011), is built on a combination of Chari, Kehoe, and McGrattan (2000), Altig, Christiano, Eichenbaum, and Linde (2004), and Smets and Wouters (2007).[4] The DSGE model is fit to eight quarterly variables: quarterly growth of real per capita GDP ($\Delta \log Y_t^{\text{Data}}$), quarterly growth of real per capita consumption ($\Delta \log C_t^{\text{Data}}$), quarterly growth of real per capita investment in capital goods unit ($\Delta \log I_t^{\text{Data}}$), quarterly growth of the real wage ($\Delta \log w_t^{\text{Data}}$), the quarterly GDP-deflator inflation rate (π_t^{Data}), quarterly growth of per capita hours ($\Delta \log L_t^{\text{Data}}$), the federal funds rate ($\text{FFR}_t^{\text{Data}}$), and quarterly growth of investment-specific technology ($\Delta \log Q_t^{\text{Data}}$) as measured by the inverse of the relative price of investment. A detailed description of the data is given in Appendix A. The data in the initial four quarters from 1960:I to 1960:IV are used to obtain the initial condition at 1961:I for the Kalman filter. Thus, the effective sample used for model evaluation is from 1961:I to 2010:II.

The BVAR model has the same eight variables as the DSGE model; and it has four lags from 1960:I to 1960:IV, so that the effective sample is also from 1961:I to 2010:II. We use the standard BVAR model with the Sims and Zha (1998) prior.[5]

In estimation of the Markov-switching mixture of the DSGE and BVAR models, we maintain the assumption that agents in the DSGE model form their expectations without accounting for model uncertainty. To focus on the discussion of regime-dependent weights and their posterior probabilities, we leave to Appendix D the presentation and discussion of the prior distribution and the posterior estimates of DSGE parameters in

---

[4]A detailed description of the model is given in Appendix D.

[5]Using the notation of Sims and Zha (1998), $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 1$, where μ_1 controls overall tightness of the random walk prior, μ_2 controls relative tightness of the random walk prior on the lagged coefficients, μ_3 controls relative tightness of the prior on the constant term, and μ_4 controls tightness of the prior that dampens the erratic sampling effects on lag coefficients (lag decay).

our Markov-switching mixture model. The transition matrix for the Markov process is

$$Q = \begin{bmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{bmatrix},$$

where $\sum_{s=1}^{2} q_{s,k} = 1$ for $k = 1, 2$. The prior density for Q is a Dirichlet probability density:

$$p(Q) \propto \sum_{s,k=1}^{2} (q_{s,k})^{(\alpha_{s,k}-1)},$$

where $\alpha_{s,k} > 0$. Following Sims, Waggoner, and Zha (2008), we express a prior belief that the average duration for each regime is between six and seven quarters. The belief implies that the expected value of the probability of staying in the same regime is $Eq_{s,s} = 8.5$ and the corresponding hyperparameter value is $\alpha_{s,s} = 5.6667$. The hyperparameter $\alpha_{s,k}$ for $s \neq k$ is set to 1.0 to allow for the possibility that the regime may be absorbent (i.e., $q_{s,s} = 1$).[6] The prior for model weights in each regime is also of Dirichlet form. Table 1 summarizes the prior distributions of both weights and transition parameters.

Given the prior and the data, the estimation and inference strategy is as follows:

- Obtain the estimates by maximizing the posterior density (2). We use the blockwise optimization algorithm proposed by Sims, Waggoner, and Zha (2008).
- Break the model parameters into several blocks and use the Gibbs sampler across blocks to simulate the Monte Carlo Markov Chain (MCMC) draws for statistical inference.
- Within each block of parameters during the Gibbs sampling steps, use the Metropolis algorithm.
- Use the MCMC draws from the posterior distribution to simulate impulse responses.
- Use the MCMC draws to compute marginal data densities.

Appendix C provides details of implementation at each step of our estimation and simulation approach. Note that since the likelihood is of very high dimension, the MCMC simulations will not deliver a point that is even close to the posterior mode. Moreover, the posterior mode serves as an important benchmark for selecting a good starting point for our MCMC simulator (see Appendix C for detailed discussions).

---

[6]Note that the variance of the prior is a function of $\alpha_{s,k}$.

TABLE 1. Prior distributions of weights and transition parameters

Parameters	Description	Distributions	Hyperparameters	
			$\alpha_{1,i}$	$\alpha_{2,i}$
$w_{1,s}, w_{2,s}$	Weights in the sth regime	Dirichlet	2.0	2.0
			$\alpha_{1,1}$	$\alpha_{2,1}$
$q_{1,1}, q_{2,1}$	Transition from the first regime	Dirichlet	5.6667	1.0
			$\alpha_{1,2}$	$\alpha_{2,2}$
$q_{1,2}, q_{2,2}$	Transition from the second regime	Dirichlet	1.0	5.6667

Table 2 reports the posterior estimates and the 90% probability intervals for weights and transition parameters. The probability intervals are computed from our MCMC posterior draws. For each posterior draw, we label the regimes so that the weight of the DSGE model in the first regime is less than that in the second regime. This label normalization is a computationally efficient way to approximate the Wald normalization discussed in Hamilton, Waggoner, and Zha (2007); and it is similar to the normalization proposed by Sims and Zha (2006) in which the smoothed probabilities of a regime for each posterior draw of the model parameters match closest to the smoothed probabilities of that regime based on the posterior estimates of the parameters.

In the first regime, the BVAR model dominates the DSGE model; the DSGE model receives almost no weight at the posterior mode (even the upper bound of the 90% probability interval gives the weight of only 8%). In the second regime, however, over 40% of the weight is assigned to the DSGE model at the mode estimate. Figure 1 displays marginal posterior probability distributions of the weights for the DSGE and BVAR models. The marginal posterior distribution of the DSGE model's weight is skewed to the right or, symmetrically, the marginal posterior distribution of the BVAR model's weight is skewed to the left. As a result, the 90% probability intervals for these weights differ only by less than 5%. It is evident that the DSGE model plays an important role in the second regime.

The posterior estimates of $q_{1,1}$ and $q_{2,2}$, reported in Table 2, indicate that the probability of staying in the same regime is high. Although both regimes are persistent, the first regime is more persistent than the second regime so that the ergodic probability

TABLE 2. Posterior estimates of regime-dependent weights and transition parameters

	Weights	
	w_{1,s_t} (DSGE)	w_{2,s_t} (BVAR)
$s_t = 1$	0.016 (0.010, 0.077)	0.984 (0.923, , 0.990)
$s_t = 2$	0.426 (0.300, 0.651)	0.574 (0.349, 0.700)
	Transition parameters	
	$q_{1,1}$	$q_{2,2}$
	0.985 (0.948, 0.996)	0.950 (0.835, 0.990)

Note: the parentheses indicate the bounds of the 90% posterior probability interval.
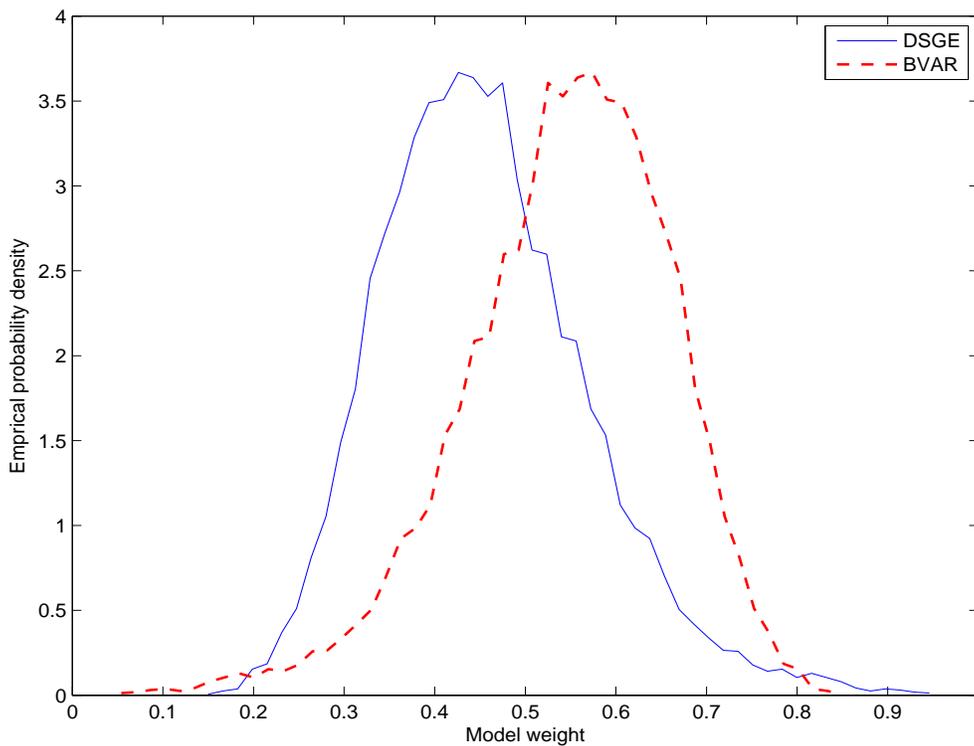
FIGURE 1. The posterior probability densities of model weights in the second regime. The plot of the densities is based on the posterior MCMC draws.

for the first regime is 0.77, implying that on average the second regime occurs only 23% of the time.[7]

---

[7]For only two regimes, as in our case, the values of $q_{1,1}$ and $q_{2,2}$ are all we need to know to compute the ergodic probability. In general, however, the ergodic probability depends on the entire transition matrix Q.
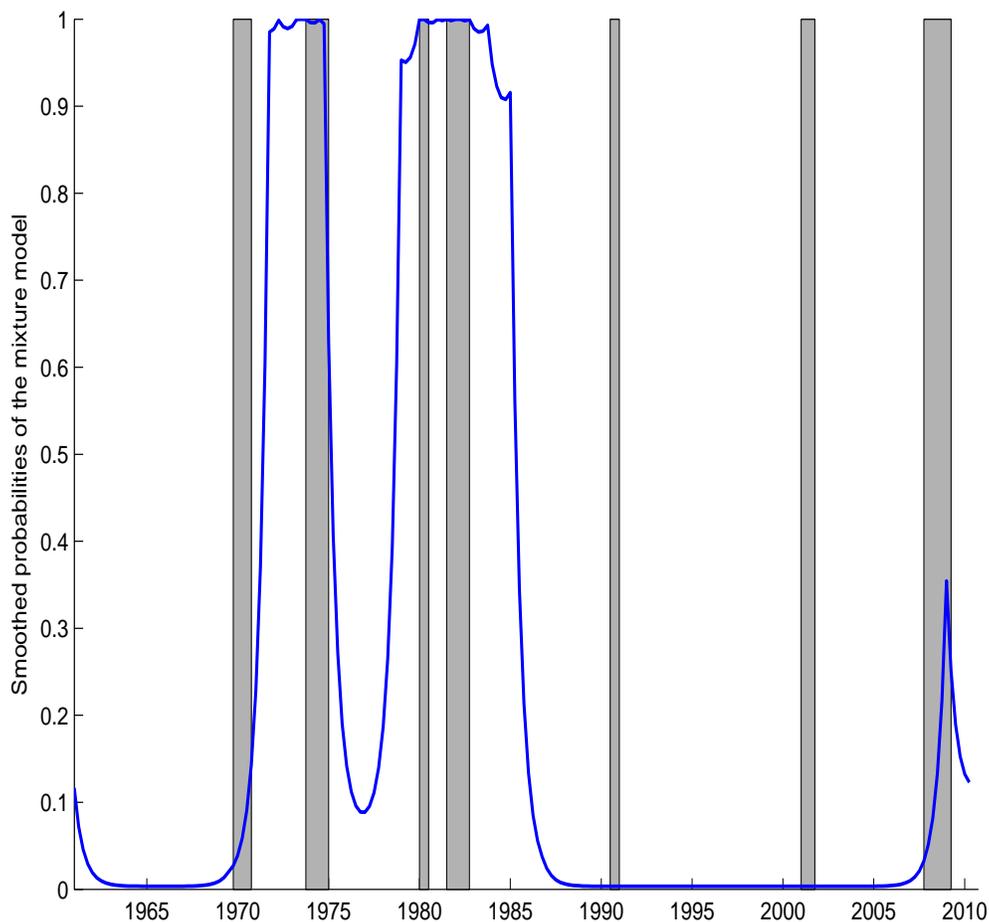
FIGURE 2. The posterior (smoothed) probabilities of the second regime in which both the DSGE model and the BVAR model play an important role. The shaded bars mark the NBER recession dates.

When we restrict model weights to be constant throughout the whole sample and estimate this constant-weight mixture model, the estimate of the DSGE model's weight is only 0.091. The magnitude of this estimate, however, is consistent with the estimates in our Markov-switching case. Since the ergodic probability of the second regime is estimated to be 0.23 and the DSGE model's weight is estimated to be 0.43, the average weight of the DSGE model is $0.23 \times 0.43 = 0.10$. In this respect the constant-weight mixture model and the Markov-switching mixture model reveal the similar information about the average role of the DSGE model throughout the history.

Del Negro and Schorfheide (2004) connect a DSGE model to a BVAR model using a parameter indicating the importance of each model, which they call λ. In their approach, λ is not a weight on the predictive density but has some similar implications.

When λ is small, the data favors their DSGE model more than their BVAR model. When λ approaches infinity, the data prefers the BVAR model. They find that λ is relatively small. One main objective of their paper is to find a good prior for their BVAR model. As discussed in Del Negro and Schorfheide (Forthcoming), the DSGE prior derived in their approach has properties similar to the Sims and Zha (1998) prior. Since our BVAR model has incorporated the Sims and Zha (1998) prior, it is not surprising that the DSGE model's weight in our framework is small on average throughout the sample.

In contrast to the constant-weight case, what is new in the Markov-switching case is information about particular times of the history when the DSGE model is important. Figure 2 displays the posterior (smoothed) probabilities of the second regime, conditional on the posterior estimates of model parameters. In this regime both the DSGE model and the BVAR model play an important role. The DSGE model matters for the late 1970s and the early 1980s, the periods that cover three adjacent recessions. For these periods the predictive densities of the DSGE model are in general much higher than those of the BVAR model. The DSGE model, however, does not receive all the weight because there is a non-trivial probability of switching from this regime to the first regime. For many periods in the first regime, including the recession in the early 1970s and the latest three recessions, the DSGE model plays little role and the BVAR model dominates. While both mixture models can be used to assess the *average* role of the DSGE model throughout the history, it is the Markov-switching result that enables us to determine when the DSGE model is usable and when it is not, as shown in Figure 2. In Sections VI and VII we explore further implications of this regime-switching feature.

V. Model fit

To assess how well our proposed Markov-switching mixture model fits to the data in comparison to other models, we compute the MDD for four models: the DSGE model, the BVAR model, the mixture model with constant weights, and the Markov-switching mixture model. For robustness of analysis we present other measures, such as the log predictive score (LPS), and estimate two versions of a pool of the DSGE and BVAR models, following Geweke and Amisano (2011). One version is based on the predictive

pool with constant weights:

$$p\left(y_t \mid Y_{t-1}^o, w, Pool\right) = w_1 p\left(y_t \mid Y_{t-1}^o, DSGE\right) + w_2 p\left(y_t \mid Y_{t-1}^o, BVAR\right),$$

where $w = (w_1, w_2) \geq 0$ with $w_1 + w_2 = 1$,

$$p\left(y_t \mid Y_{t-1}^o, DSGE\right) = p\left(y_t \mid Y_{t-1}^o, \widehat{\Theta}_{DSGE}\right),$$

$$p\left(y_t \mid Y_{t-1}^o, BVAR\right) = p\left(y_t \mid Y_{t-1}^o, \widehat{\Theta}_{BVAR}\right).$$

The values $\widehat{\Theta}_{DSGE}$ are the posterior estimates of the DSGE parameters when the DSGE model is estimated over the full sample. Similarly, the values $\widehat{\Theta}_{BVAR}$ are the posterior estimates of the BVAR parameters when the BVAR model is estimated over the full sample. Since this is a pool of the two models, we take the estimated parameters for both models as given before we pool the two models. Thus, the pool involves choosing the optimal value of w (not parameters) such that the log predictive score, defined below, is maximized:

$$\text{LPS}\left(Y_T^o | w, Pool\right) \equiv \sum_{t=1}^{T} \log p\left(y_t^o \mid Y_{t-1}^o, w, Pool\right).$$

The second version of a pool of the DSGE and BVAR models is to allow the model weights to vary over time according to the Markov-switching process defined in Section III. Specifically, the predictive pool with Markov-switching weights is

$$p\left(y_t \mid Y_{t-1}^o, Q, w, Pool\right) =$$

$$\sum_{s_t=1}^{2} \left[w_{1,s_t} p\left(y_t \mid Y_{t-1}^o, DSGE\right) + w_{2,s_t} p\left(y_t \mid Y_{t-1}^o, BVAR\right)\right] p\left(s_t \mid Y_{t-1}^o, Q, w\right),$$

where $w = (w_{1,1}, w_{1,2}, w_{2,1}, w_{2,2}) \geq 0$ with $w_{1,s_t} + w_{2,s_t} = 1$, and $s_t = \{1, 2\}$ follows the two-regime Markov-switching process. The parameters Q and w are chosen to maximize the log predictive score

$$\text{LPS}\left(Y_T^o | Q, w, Pool\right) \equiv \sum_{t=1}^{T} \log p\left(y_t^o \mid Y_{t-1}^o, Q, w, Pool\right),$$

where the estimated parameters for both the DSGE and the BVAR models are taken as given before Q and w are optimally chosen.

Table 3 reports the computed MDDs and LPSs for all the models we have discussed. We begin with an analysis of the LPS. The LPS for the BVAR model is overwhelmingly

TABLE 3. Log marginal data densities, log predictive scores, and log concentrated predictive likelihood

Model type	log MDD	LPS	LCPL
DSGE	5735.72	5882.61	4458.65
	(5735.18, 5736.43)		(4458.42, 4459.96)
BVAR	5894.60	6441.68	4619.47
Constant pool	N/A	6441.71	N/A
Markov pool	N/A	6441.71	N/A
Constant mixture	5943.14	6537.73	4680.47
	(5942.38, 5943.18)		(4679.72, 4681.63)
Markov mixture	5952.67	6550.28	4692.78
	(5951.47, 5952.74)		(4691.47, 4692.94)

Note: "N/A" stands for "not applicable." For the "log MDD" column, the value above parentheses is the log MDD estimated with 100 million MCMC draws. The parentheses indicate the minimum and maximum values of the log MDDs estimated from 10 chains of one million MCMC draws with 10 starting points independently drawn from the prior distribution.

higher than that for the DSGE model, by over 500 in log value.[8] In contrast, the simple pool with optimal constant weights improves the LPS of the BVAR model by only 0.03 in log value. To understand why the improvement is so insignificant, we observe that the BVAR model fits to the data so much better than the DSGE model that the log predictive density $p\left(y_t^o \mid Y_{t-1}, BVAR\right)$ is considerably greater than $p\left(y_t^o \mid Y_{t-1}, DSGE\right)$ for 180 data points out of a total of 198 quarters in the sample. In the other 18 periods when $p\left(y_t^o \mid Y_{t-1}, BVAR\right)$ is less than $p\left(y_t^o \mid Y_{t-1}, DSGE\right)$, the differences between the log predictive densities is so small (relative to the differences between $p\left(y_t^o \mid Y_{t-1}, BVAR\right)$ and $p\left(y_t^o \mid Y_{t-1}, DSGE\right)$ in those 180 periods) that the optimal weight for the DSGE model is virtually zero (on the order of $1.0E-10$). The Markov-switching version of an optimal pool yields the same result, with one regime

---

[8]The superiority of the BVAR model with the Sims and Zha (1998) prior over the DSGE model is well documented in the literature. See Del Negro and Schorfheide (Forthcoming).

being an absorbing state and the weight for the BVAR model in this regime being virtually one.

The antecedent finding suggests that it would be worthwhile to explore a mixture of the two models in which the parameters and the weights are jointly estimated. Geweke and Amisano (2011) show that asymptotically a mixture model must be superior to a pool and, when all individual models are false, the pool is superior to individual models. Although this asymptotic result has, in general, no implication of the same ranking in small samples, does it hold for our application? Table 3 indicates that the mixture model with constant weights improves the log LPS by 96. The improvement produced from the Markov-switching mixture increases the log LPS by additional 12. Thus, about 90% of the improvement in model fit is due to a convex combination of predictive densities of individual models. Changes in model weights deliver another 10% of the improvement. Moreover, allowing changes in model weights brings in important economic substance regarding the role of the DSGE model at particular times in history, as stressed in Section IV. We continue to discuss other economic implications of the Markov-switching mixture model in Sections VI and VII.

We now turn to an analysis of the MDD. We use the truncated modified harmonic mean (MHM) method proposed by Sims, Waggoner, and Zha (2008) to calculate the MDD. The details of this method are given in Appendix C. When comparing MDDs, one should bear in mind that the ratio of two MDDs is the Bayes factor. If the difference in log values of MDD between two models is greater than 5, for example, the model with the higher value of MDD is favored decisively by the data. For the BVAR model, there is an analytical solution for calculating the MDD so that the reported log value of MDD has negligible numerical errors. For the DSGE model and the two mixture models, however, numerical errors are small but noticeable. Table 3 reports the minimum and maximum values of the MDDs estimated from 10 chains of one million MCMC draws with 10 starting points independently drawn from the prior distribution.

Note that we do not report MDDs for the two pool models (indicated by N/A). One can compute the MDD by explicitly specifying a prior on model weights, which would not differ much from the LPS value. Since each constituent model in the pool is estimated beforehand, it is equivalent to say that the prior of model parameters is degenerated by centering at the estimates. Reporting such an MDD value would

not be useful. Moreover, since each constituent model in the pool is taken as given before the optimal weights are chosen, there is in general no internal mechanism to prevent in-sample over-fit for an individual model. The mixture approach penalizes model complexity by estimating the parameters in both models jointly.

Nonetheless, the conclusion reached from the MDD results about the ranking of models is the same as that from the LPS results. The two mixture models expand the parameter space by jointly estimating the weights and the parameters of both individual models. The log MDD for the BVAR is higher than the log MDD for the DSGE model by over 150. But the mixture model with constant weights dominates the BVAR model by about 50 in log MDD. For the Markov-switching mixture model, we gain additional 9 value in the log MDD. Again, this result strengthens the previous finding from the LPS analysis that 85% of the improvement in model fit is attributable to a convex combination of predictive densities of individual models.

All this analysis, computationally expensive as it is, indicates that a formal mixture of the two heterogeneous models is important in improving the model's fit to the data.[9] As additional verification, by reducing the influence of the prior distribution, we use the 10-year data at the beginning of the sample to compute the log predictive likelihood, concentrating on the latter part of the sample from 1972:I to 2010:II. To see how to compute this log concentrated predictive likelihood (LCPL), we decompose the log MDD as[10]

$$\log p\left(Y_T^o \mid M\right) = \log\left[\prod_{t=1}^{t^*-1} p\left(y_t^o \mid Y_{t-1}^o, M\right)\right] + \log\left[\prod_{t=t^*}^{T} p\left(y_t^o \mid Y_{t-1}^o, M\right)\right],$$

where M stands for a particular model we study and t^* corresponds to 1972:I in our case. The first term on the right-hand side is the log MDD using the data up to $t^* - 1$. We apply our MCMC simulator on this earlier sample. The LCPL, the second term on the right-hand side, is the difference in log MDDs between the full sample and the earlier sample. The column with the label "LCPL" on the top in Table 3 reports the LCPL values for the four models studied in this paper. It is evident that the LCPLs for both mixture models are significantly higher than those for the DSGE and BVAR models, reinforcing the previous conclusion reached by both LPS and MDD analyses.

---

[9]See Appendix C.4 for a detailed description of computing time.

[10]We thank a referee for this suggestion.

VI. Data filter

One application of our Markov-switching mixture model is to filter the data for estimation of the DSGE model by discounting observations from the periods in which the BVAR dominates. As shown in Table 2 and Figure 2, the DSGE model plays almost no role during the periods under the first regime. This regime covers about 76% of the sample, consistent with the estimated ergodic probability of the first regime presented in Section IV. The data in the first regime is almost completely discounted for estimation of the DSGE parameters.

The DSGE model plays an important role in the second regime. This regime covers a much shorter period of the sample, concentrating on the late 1970s and the early 1980s. The data in the periods under the second regime is partially discounted for estimation of the DSGE parameters because the weight assigned to the DSGE model is less than one. We present and discuss the estimation results for the DSGE parameters in the Markov-switching mixture model in Appendix D.

In this section we focus on the impulse responses of an economic shock in the DSGE model. Since many structural shocks in the DSGE model cannot be identified by the BVAR model, it is important to assess the differences between the responses implied by the Markov-switching mixture model and those from the DSGE model when estimated in isolation. To this end, we use a capital depreciation shock as an example. The capital depreciation shock is a shock to the depreciation rate in the capital accumulation equation in the DSGE model. It is an important shock, as it can be interpreted as a proxy for a shock to efficiency in using the capital or a financial shock.

Let ε_{dt} be an i.i.d. shock to capital depreciation at the time t with $E(\varepsilon_{dt}) = 0$ and $Var(\varepsilon_{dt}) = 1$. A vector of the kth step impulse responses is defined as

$$IR_{DSGE,T+k}^{(j)} = E\left[y_{T+k} \mid \Theta_{DSGE}^{(j)}, \varepsilon_{dT+1} = 1, Y_T, DSGE\right]$$
$$- E\left[y_{T+k} \mid \Theta_{DSGE}^{(j)}, Y_T, DSGE\right], \quad (3)$$

where $\Theta_{DSGE}^{(j)}$ is a jth posterior draw from the Markov-switching mixture model. In our calculation, the initial condition represented by Y_T has no effect on impulse responses because the DSGE model itself is linear. Thus one can choose an arbitrary value of Y_T.

Even if the DSGE parameters are dawn from the posterior distribution of the Markov-switching mixture model, the impulse responses are computed based only on the estimated DSGE model in the mixture because a depreciation shock is not identified in the BVAR model. Since the DSGE model is important only in the second regime, the impulse responses represented by $IR^{(j)}_{DSGE,T+k}$ are effectively those in the second regime. In Section VII we discuss how to obtain impulse responses in the first regime when a common structural shock is identified in both individual models.

Figure 3 contrasts the impulse responses generated by (3), the responses from the DSGE-only model (i.e., when it is estimated in isolation over the full sample), and those from the prior distribution of the DSGE model. The figure displays the impulse responses of output, consumption, real wage, and inflation to a one-standard-deviation shock to capital depreciation. To be compatible with the literature, we follow Sims and Zha (1999) and report the 68% probability bands. The left-hand column displays the responses generated from the estimated Markov-switching mixture model, the middle column displays the responses from the DSGE-only model, and the right column displays the responses generated from the prior distribution of the DSGE model. It is clear that the estimated responses from either the mixture model or the DSGE-only model differ substantially from the responses generated by the prior distribution of the DSGE model.[11] The data are therefore informative.

Comparing the left and middle columns side by side, one can see the notable differences between the mixture model and the DSGE-only model. We begin our analysis with the middle column, when the impulse responses are based on the DSGE-only model. The increase in the depreciation rate reduces the value of capital accumulation, raises the marginal cost of capital, and lowers investment. Since the expected stock of capital wealth declines, the negative wealth effect leads to a fall in consumption. Consequently, aggregate output falls. The decline in output leads to a decline in hours. The equilibrium real wage falls as well, because the declines in hours and in

---

[11]Note that the probability bands from the estimated DSGE model alone tend to be narrower than those from the prior distribution of the DSGE model. Comparing the 95% probability intervals in Table 5 and those in Table 7 in Appendix D, one can see that the posterior standard deviations for some shocks, such as the monetary policy shock and the technology shock, have much tighter ranges than do the prior standard deviations. In those cases, the probability bands from the estimated DSGE model alone are much narrower than those from the prior distribution of the DSGE model.

consumption lower the marginal rate of substitution between labor and consumption, so that the household's desired wage falls. Inflation responses are positive, but it is difficult to detect visually because the magnitude is too small (on the order of 1.0E-06). Inflation increases in response to a depreciation shock because the rise in the marginal cost of capital slightly dominates the fall in the real wage. According to the probability bands, all the responses are sharply estimates.

The corresponding impulse responses generated from the mixture model are considerably larger, both in magnitude and in the width of probability bands (the left-hand column of Figure 3). For the estimates of the DSGE parameters from the mixture model, the fall in the real wage in response to a positive depreciation shock slightly overweighs the rise in the marginal cost of capital so that inflation responses are predominantly negative on impact but are statistically insignificant according to the probability bands. For the most part, the responses of the three real variables (output, consumption, and the real wage) are statistically significant. In this sense, the data discounted or filtered through the mixture model do not make estimation of the DSGE parameters lose its economic meaning.

More interesting is the width of probability bands in the left-hand column of Figure 3. Our estimation indicates that estimation of the DSGE model utilizes about one-tenth of the data points in the sample (taking into account the weight for the DSGE model in the second regime being 0.43). Thus it is unsurprising that the probability bands are wider. What is new in our finding, however, is that the width of probability bands for the mixture model, for most impulse responses, is far more than three (a square root of ten) times the width for the DSGE-only model when it is estimated with all the data points. Insight about this result is revealed in Figure 2. Since the second regime concentrates on three adjacent recessions and excludes many periods of economic expansions, the data in this regime have more similarity than the data in the first regime when both recessions and expansions are covered. Such similarity results in considerably more uncertainty surrounding the estimates than what a simple counting of data points would suggest.

VII. Full structural analysis

Although the analysis of impulse responses to a capital depreciation shock, presented in Section VI, is based on the DSGE model, the analysis is partially structural in the
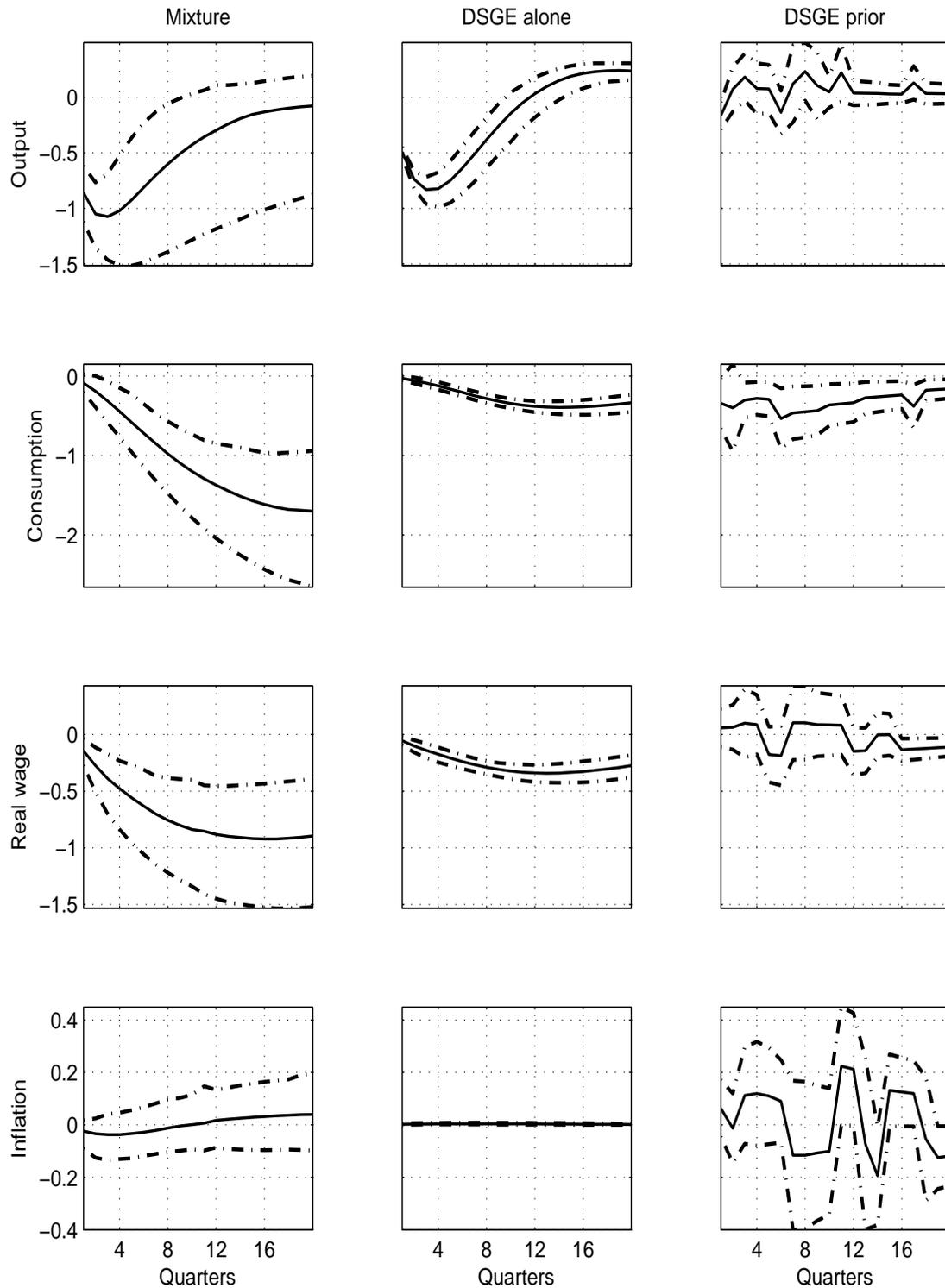
FIGURE 3. Impulse responses (expressed as percentages) to a capital depreciation shock for the Markov-switching mixture model (left column), for the DSGE model when estimated in isolation (middle column), and for the DSGE model with the prior distribution only (right column). The dashed lines represent 68% posterior probability bands and the solid line represents the posterior median estimate.

sense that the BVAR model is not used to interpret the results. The BVAR model plays a role only in the estimation stage. Once the estimation is finished and the posterior draws of the DSGE parameters from the Markov-switching mixture are stored, the BVAR model has no use in the analysis of impulse responses. This exercise is useful only when the BVAR model is *not* treated as a structural model.

There is a large strand of literature on using a BVAR model to identify certain economic shocks, if not all the shocks (Bernanke (1986), Blanchard and Watson (1986), Sims (1986), Leeper, Sims, and Zha (1996), Christiano, Eichenbaum, and Evans (1999)). One prominent example is a monetary policy shock. If we apply the Choleski decomposition to our BVAR model and let the interest rate respond to all other variables contemporaneously, a shock to the interest rate equation is identified as a monetary policy shock by Christiano, Eichenbaum, and Evans (1999). Our original BVAR model has the same ordering of the variables as this structural version, which is now used to identify a monetary policy shock. Building on our econometric framework, we now merge this structural BVAR model (SBVAR model henceforth) with the DSGE model that has the *same* type of shock: a shock to monetary policy.

Figures 4 and 5 display the impulse responses to a one-standard-deviation monetary policy shock under four scenarios: the DSGE-only model (when estimated alone over the full sample), the SBVAR-only model (when estimated in isolation over the full sample), the first regime for the Markov-switching mixture of these two heterogenous models, and the second regime. To avoid wordiness, we use the DSGE model to mean the "DSGE-only" model and the SBVAR model to mean the "SBVAR-only" model, whenever it is clear in the context.

We begin with our analysis on the DSGE-only and SBVAR-only models. Impulse responses to a monetary policy shock from these two heterogenous models are considerably different, both qualitatively and quantitatively. The top two rows of graphs in Figures 4 show that the magnitude of the interest rate response for the DSGE-only model is slightly higher than that for the SBVAR-only model at the beginning of the forecast horizon, but the response for the SBVAR-only model is more persistent. The response of the price level for the DSGE model is small and negative (on the order of 1.0E-06 and thus difficult to detect by eye). The response of the price level for the SBVAR model is relatively large and positive. Although the positive response of the

price level is known in the SBVAR literature as a "price puzzle," the SBVAR model continues to be used to analyze the effect of a monetary policy shock on real variables such as investment and output (Sims (1992), Leeper, Sims, and Zha (1996), Christiano, Eichenbaum, and Evans (1999)).

The top two rows of graphs in Figure 5 show drastically different patterns of the responses of investment and output for the two individual models. The DSGE model implies that the responses of both investment and output are negative at the beginning, but the negative effect disappears after two years (eight quarters). In contrast, the SBVAR model implies that the effect on investment and output of a monetary policy shock is persistently negative throughout the forecast horizon.

To resolve these differences between the DSGE and SBVAR models, the traditional approach in the literature is to estimate the DSGE model subject to the constraint that the impulse responses to a monetary policy matches those of the SBVAR model (Christiano, Eichenbaum, and Evans, 2005). The key argument for this approach is that the SBVAR model dominates the DSGE model in the fit to the data. Indeed, the evidence is supported by the results presented in Table 3, where all three measures indicate that the SBVAR model is decisively favored by the data. This approach suggests that misspecification of the SBVAR model be not serious in practice.

Our approach recognizes that both the DSGE and SBVAR models may be misspecified. Table 3 shows that our Markov-switching mixture model improves the fit considerably when compared to the SBVAR model and that model weights differ substantially across the two regimes. We now describe how to compute impulse responses to a monetary policy shock from the Markov-switching mixture model, and then explore how impulse responses change from one regime to the other and how they differ from those generated from the two individual models.

Let $\Theta^{(j)}$, $Q^{(j)}$, and $w^{(j)}$ denote the jth posterior draw of all the model parameters. A vector of impulse responses at horizon k to a one-standard-deviation monetary policy shock $\varepsilon_{pT+1} = 1$ in regime $\ell \in \{1, 2\}$ is given by

$$IR_{\ell,T+k}^{(j)} = E\left[y_{T+k}|\Theta^{(j)}, w^{(j)}, \varepsilon_{pT+1} = 1, s_{T+1} = \cdots = s_{T+k} = \ell, Y_T\right] -$$
$$E\left[y_{T+k}|\Theta^{(j)}, w^{(j)}, s_{T+1} = \cdots = s_{T+k} = \ell, Y_T\right]. \quad (4)$$
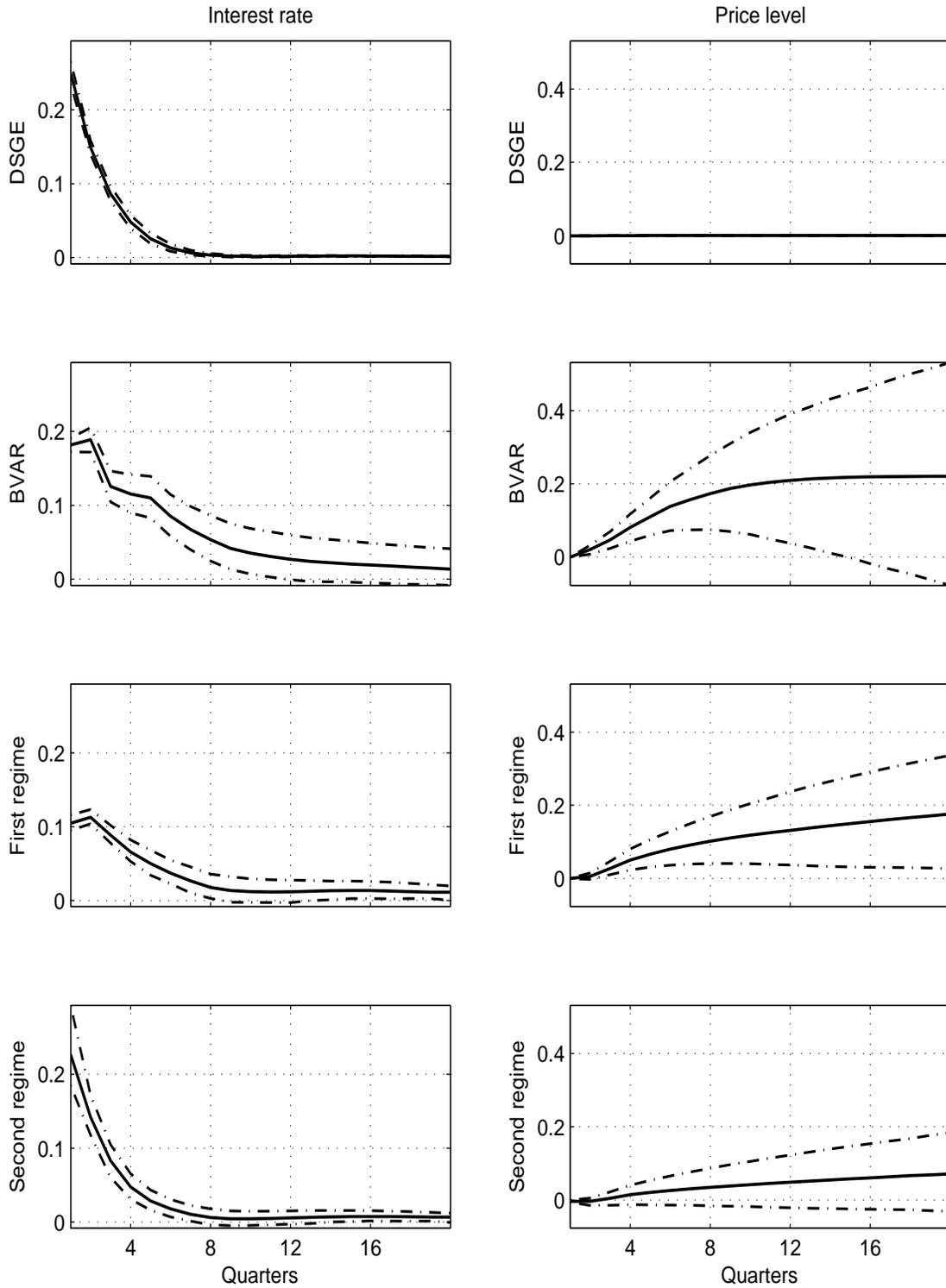
FIGURE 4. Impulse responses (expressed as percentages) to a monetary policy shock from the DSGE-only model (first row), the SBVAR-only model (second row), and Markov-switching mixture model (third and fourth rows). The dashed lines represent 68% posterior probability bands and the solid line represents the posterior median estimate.
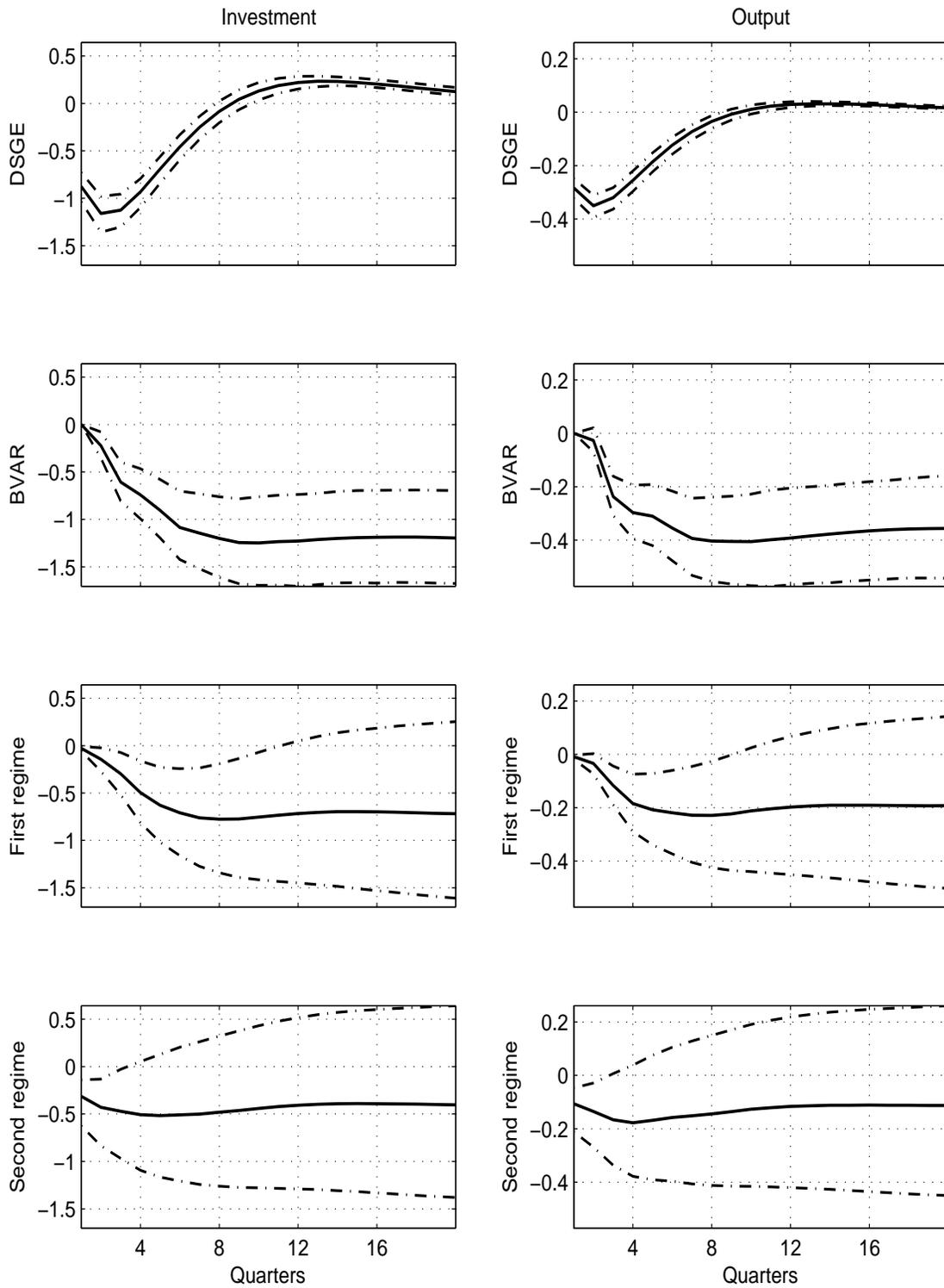
FIGURE 5. Impulse responses (expressed as percentages) to a monetary policy shock from the DSGE-only model (first row), the SBVAR-only model (second row), and Markov-switching mixture model (third and fourth rows). The dashed lines represent 68% posterior probability bands and the solid line represents the posterior median estimate.

Notice the notational difference between $IR^{(j)}_{DSGE,T+k}$ in (3) and $IR^{(j)}_{\ell,T+k}$ in (4). In (3), since a depreciation shock is identified only in the DSGE model, the impulse responses are computed through the DSGE model for every posterior MCMC draw of model parameters. In (4) both the DSGE and SBVAR models identify the same type of structural shock, which is a monetary policy shock in our case. Thus, the subscript in $IR^{(j)}_{\ell,T+k}$ has no reference to any particular model, only to the regime.

To compute the impulse responses defined in (4), we let the state space representation of the ith model be

$$y_t = a_i + H_i f_{i,t}, \tag{5}$$

$$f_{i,t} = b_i + F_i f_{i,t-1} + \Phi_i \varepsilon_{i,t}. \tag{6}$$

The first system (5) represents measurement equations, the second system (6) represents state equations, and f_t is an unobserved state vector. For a monetary policy shock, we have $\varepsilon_{i,pt} = \varepsilon_{pt}$ for $i \in \{DSGE, SBVAR\}$. That is, the DSGE model and the SBVAR model identify same shock. Combining (4)-(6) leads to

$$IR^{(j)}_{\ell,T+k} = \sum_i w^j_{i,\ell} H_i \Big\{ E\left[f_{i,T+k} | \Theta^{(j)}, \varepsilon_{pT+1} = 1, s_{T+1} = \cdots = s_{T+k} = \ell, Y_T \right] -$$

$$E\left[f_{i,T+k} | \Theta^{(j)}, s_{T+1} = \cdots = s_{T+k} = \ell, Y_T \right] \Big\}, \tag{7}$$

where $i \in \{DSGE, SBVAR\}$. Because both the DSGE and BVAR models are linear, the impulse responses at regime ℓ, represented by $IR^{(j)}_{\ell,T+k}$, turn out to be independent of the initial condition Y_T.

We follow Sims and Zha (2006) and report impulse responses to a monetary policy shock under different regimes, represented by $IR^{(j)}_{\ell,T+k}$.[12] The bottom two rows of graphs in Figures 4 and 5 display the resulting responses. We first analyze the first regime in which the SBVAR model dominates. As shown in the third row of Figure 4, the dynamic responses of the price level are smaller both in magnitude and in probability bands than those generated by the SBVAR-only model; nonetheless, the price puzzle continues to be significant. As for real variables, the effects on investment and output of a monetary policy shock, estimated from the Markov-switching mixture model, are

---

[12]We compute the impulse responses using alternative methods discussed in Appendix B and with various values of p_T, including the unconditional probabilities of regimes. We find that the values tend to lie between $IR^{(j)}_{1,T+k}$ (first regime) and $IR^{(j)}_{2,T+k}$ (second regime).

smaller with wider probability bands than those generated from the SBVAR-only model (comparing the second and third rows of Figure 5).

The influence of model uncertainty is evident in this case. The SBVAR model, when estimated alone, indicates that investment and output continue to stay negative after two years (the second row of Figure 5). The Markov-switching mixture model (the third row of Figure 5), however, reveals that there is a nontrivial probability that investment and output become positive after two years. This time horizon is consistent with the time after which the responses of investment and output from the DSGE-only model become positive (the first row of Figure 5). Even though the DSGE model plays little role in the first regime, its importance in the second regime influences the joint estimation of the parameters in both individual models. The influence is large enough to alter the estimates and distributions of impulse responses of both nominal and real variables.

We now analyze the second regime, in which both the DSGE model and the SBVAR model play an important role, and compare the results to those in the first regime. The fourth rows of graphs in Figures 4 and 5 display impulse responses under the second regime in the Markov-switching mixture model. In response to a monetary policy shock, the interest rate rises twice as much as does the interest rate in the first regime. The price puzzle is much weakened. In comparison to the result in the first regime, the magnitude of the responses is smaller and there is a nontrivial probability of no price puzzle (the fourth row of Figure 4). As for real variables, the uncertainty about the responses of investment and output is larger than that in the first regime. According to the probability bands, the negative responses of investment and output in the second regime are short lived and, in general, have very wide probability bands, even at the beginning of the forecast horizon.

To summarize, we show how to compute the impulse responses to the same type of shock when both individual models in the mixture are structural. More important is what we have learned from this exercise. Since the responses based on on each individual model are sharply estimated, economic implications about the effects of a monetary policy shock differ from one model to the other, as shown in the top two rows of graphs in Figures 4 and 5. Comparing the impulse responses in the third and fourth rows to those in the first and second rows of Figures 4 and 5, one can see clearly

that the mixture of the two individual models, the DSGE model and the SBVAR model, has fundamentally different implications about the magnitude and uncertainty of the effects of a monetary policy on both nominal and real variables. The finding of smaller magnitude and larger uncertainty about the impulse responses from the Markov-switching mixture model, in comparison to those from the SBVAR-only model, is consistent with the view that a large effect of monetary policy is predominantly due to its systematic component, not due to its unpredictable (random) shocks (Bernanke, Gertler, and Watson (1997), Sims and Zha (2006)).

VIII. Conclusion

We show in this paper how to apply the Markov-switching mixture methodology to macroeconomic models. We study two types of widely used macroeconomic models: a DSGE model and a BVAR model. Although it is computationally demanding, we show that estimating a Markov-switching mixture of these two heterogenous models is feasible. The estimated mixture model with two regimes improves the fit to the data considerably, implying that both models may be misspecified. Taking into account model uncertainty can alter the estimated results of the parameters in each individual model. Using a capital depreciation shock as an example, we illustrate how impulse responses in the DSGE model are changed when the mixture model is used as a data filter. When the DSGE model and the BVAR model identify a common economic shock, which is a monetary policy shock in our application, we show how to use the Markov-switching mixture model to combine the two individual models and how to compute the impulse responses from the mixture model. The resulting impulse responses differ across the two regimes and have different economic implications about the magnitude and uncertainty when compared to the impulse responses generated by each individual model.

The Markov-switching mixture model studied in this paper should be viewed as a step towards a deep and sophisticated macroeconomic model that we have neither technology nor intellectual capacity to cope with at the present time. To this end, one natural extension is to allow agents in our economic model to take into account both model uncertainty and parameter uncertainty. Works by Hansen and Sargent (2001), Brock, Durlauf, and West (2003), and Hansen and Sargent (2010) provide guidance on how to pursue this line of research in the future. Meanwhile, it is our hope that the

empirical exercise conducted in this paper illustrates how a Markov-switching mixture of heterogenous structural models can be used to integrate model uncertainty and parameter uncertainty in macroeconomics.

Appendix A. Detailed data description

All data are constructed from the original data in the Haver Analytics Database. The constructed data, the original data identifiers, and the data sources are described below.

- $Y_t^{\text{Data}} = \frac{\text{GDPH}}{\text{LN16N@USECON}}$.
- $C_t^{\text{Data}} = \frac{(\text{CN@USECON} + \text{CS@USECON - CSRU@USECON})*100/\text{JCXFE@USNA}}{\text{LN16N@USECON}}$.
- $I_t^{\text{Data}} = \frac{(\text{CD@USECON} + \text{FNE@USECON})*100/\text{JCXFE@USNA}}{\text{LN16N@USECON}}$.
- $w_t^{\text{Data}} = \frac{\text{LXNFC@USECON}/100}{\text{JCXFE@USNA}}$.
- $\pi_t^{\text{Data}} = \frac{\text{JCXFE@USNA}_t}{\text{JCXFE@USNA}_{t-1}}$.
- $L_t^{\text{Data}} = \frac{\text{LXNFH@USECON}}{\text{LN16N@USECON}}$.
- $\text{FFR}_t^{\text{Data}} = \frac{\text{FFED@USECON}}{400}$.
- $Q_t^{\text{Data}} = \frac{\text{JCXFE@USNA}}{\text{GordonPriceCDplusES}}$.

LN16N@USECON: Civilian noninstitutional population: 16 years and over. Breaks in population are eliminated from 10-year censuses and post 2000 American Community Surveys using "error of closure" method. This fairly simple method was used by the Census Bureau to get a smooth population monthly population series. This smooth series reduces the unusual influence of drastic demographic changes. Source: BLS.

GDPH: Real gross domestic product (2005 dollars). Source: BEA.

CN@USECON: Nominal personal consumption expenditures: nondurable goods. Source: BEA.

CS@USECON: Nominal consumption expenditures: services. Source: BEA.

CSRU@USECON: Nominal personal consumption expenditures: housing and utilities. Source: BEA.

CD@USECON: Nominal personal consumption expenditures: durable goods. Source: BEA.

FNE@USECON: Nominal private nonresidential investment: equipment and software. Source: BEA.

JCXFE@USNA: PCE excluding Food and Energy: Chain Price Index (2005=100).
Source: BEA.

LXNFC@USECON: Nonfarm business sector: compensation per hour (1992=100).
Source: BLS.

LXNFH@USECON: Nonfarm business sector: hours of all persons (1992=100).
Source: BLS.

FFED@USECON: Annualized federal funds effective rate. Source: FRB.

GordonPriceCDplusES: Investment deflator. The Tornquist procedure is used
to construct this deflator as a weighted aggregate index from the four quality-
adjusted price indexes: private nonresidential structures investment, private
residential investment, private nonresidential equipment and software invest-
ment, and personal consumption expenditures on durable goods. Each price
index is a weighted one from a number of individual price series within these
categories. For each individual price series from 1947 to 1983, we use Gordon
(1990)'s quality-adjusted price index. Following Cummins and Violante (2002),
we estimate an econometric model of Gordon's price series as a function of a
time trend and a number of NIPA indicators (including the current and lagged
values of the corresponding NIPA price series). The estimated coefficients are
then used to extrapolate the quality-adjusted price index for each individual
price series for the sample from 1984 to 2007. These constructed price series
are annual. Denton (1971)'s method is used to interpolate these annual series
on a quarterly frequency. The Tornquist procedure is then used to construct
each quality-adjusted price index from the appropriate interpolated quarterly
price series.

APPENDIX B. IMPULSE RESPONSES AND DECOMPOSITION OF VARIANCE

In Section VII we discuss how to compute impulse responses conditional on a partic-
ular regime when both models are structural. Because regime-switching model weights
introduce nonlinearity into the mixture model, there are many other ways to compute
impulse responses. One approach is to let impulse responses depend on the probability
of s_{T+k} instead of a particular regime. Let $p_{T+k}^{(j)}$ be the vector whose ℓ^{th} component,
$p_{\ell,T+k}^{(j)}$, is the probability that $s_{T+k} = \ell$. We have $p_{T+k}^{(j)} = \left(Q^{(j)}\right)^k p_T^{(j)}$. Let $w_{i,T+k}^{(j)}$ de-
note the weight associated with model $i \in \{DSGE, BVAR\}$ at the time $T + k$. Thus,

$w_{i,T+k}^{(j)} = \sum_{\ell=1}^{2} w_{i,\ell}^{(j)} p_{\ell,T+k}^{(j)}$. Clearly, $w_{i,T+k}^{(j)}$ does not depend on a particular realization of s_{T+k} but rather on the probability of s_{T+k}. Impulse responses are computed as

$$IR_{T+k}^{(j)} = \sum_{i} w_{i,T+k}^{(j)} H_i \Big\{ E\left[f_{i,T+k}|\Theta^{(j)}, \varepsilon_{pT+1}=1, Y_T\right] - E\left[f_{i,T+k}|\Theta^{(j)}, Y_T\right] \Big\},$$

where $i \in \{DSGE, SBVAR\}$. Note that the impulse responses $IR_{T+k}^{(j)}$ do not depend on any particular model in the mixture and that they depend on Y_T if and only if the initial probability $p_T^{(j)}$ is a function of Y_T. Note that the term in the braces is simply the impulse response of the ith state vector to a structural shock and can easily be computed using the ith state equation.

Impulse responses are nonlinear functions of model parameters Θ. For any function of Θ that has a finite posterior variance, denoted by $f(\Theta)$, we propose the following method to decompose the overall variance of $f(\Theta)$ into the sum of two components, one attributable to parameter uncertainty within the model and the other to uncertainty across models:[13]

$$\underbrace{Var\left(f(\Theta) \mid Y_T\right)}_{\text{Overall uncertainty}} = \underbrace{\int Var\left(f(\Theta) \mid w, Y_T\right) p(w \mid Y_T)\, dw}_{\text{Parameter uncertainty}}$$
$$+ \underbrace{\int \left[E\left(f(\Theta) \mid w, Y_T\right) - E\left(f(\Theta) \mid Y_T\right)\right]^2 p(w \mid Y_T) dw}_{\text{Model uncertainty}}.$$

As shown in Table 3, the MDD for the BVAR model is at least 150 over the MDD for the DSGE model. Thus, the BVAR model overwhelmingly dominates the DSGE model and there is no model uncertainty according to Bayesian model averaging. In contrast, our proposed approach to decomposition overcomes this difficulty by measuring model uncertainty through variations in model weights. Indeed, as shown in Table 2, there exists a considerable variation in model weights. Implementing our decomposition method, however, incurs an additional cost of sampling Θ for each posterior draw of w.

---

[13]This decomposition is for the purpose of structural analysis. For the purpose of pure prediction or forecasting, Geweke and Amisano (Forthcoming) show how to decompose the variance of predictive distributions into extrinsic variance, arising from posterior uncertainty about parameters, and intrinsic variance, arising from predictive errors given the parameters.

Appendix C. The MCMC posterior sampler

In this section we describe, in detail, our algorithm of finding the posterior mode, our posterior simulator for MCMC draws, and our method of computing the MDD. Our approach follows Sims, Waggoner, and Zha (2008). In all succeeding subsections we omit the notation M_i for notational simplicity with the understanding that all the objects analyzed in this appendix are conditioned on a particular model, being the DSGE model, the BVAR model, or the mixture model.

C.1. **Posterior mode.** Estimation of a mixture of our DSGE and BVAR models, the two substantially heterogeneous models, is a challenging task, as the shape of the posterior density tends to be very non-Gaussian, full of local modes and winding ridges. Because of such a non-Gaussian shape of the density function, the posterior mean receives an extremely low probability and thus is a poor approximation to the posterior mode. For the same reason, searching the posterior mode is difficult, as standard optimization routines often converge to different local peaks from different starting points.

We use the block-wise optimization method recommended by Sims, Waggoner, and Zha (2008). We first group the model parameters into four blocks: Θ_{DSGE} (all the parameters for the DSGE model), Θ_{BVAR} (all the parameters for the BVAR model), Q, and w. This separation proves critical in practice because the conditional posterior density $p(\theta_{DSGE} \mid Y_T^o, \theta_{BVAR}, Q, w)$ differs substantially from $p(\theta_{BVAR} \mid Y_T^o, \theta_{DSGE}, Q, w)$. While the density $p(\theta_{DSGE} \mid Y_T^o, \theta_{BVAR}, Q, w)$ is non-Gaussian, the conditional posterior density $p(\theta_{BVAR} \mid Y_T^o, \theta_{DSGE}, Q, w)$ is closer to being Gaussian.

Given an initial guess of the values of the parameters, we use a standard hill-climbing quasi-Newton optimization routine to find the value of each block of parameters that maximizes the posterior density while holding other blocks of parameters fixed at the previous values. We iterate this algorithm through blocks until it converges. For each iteration we employ a *constrained* optimization routine to check whether there are boundary or corner solutions associated with Q, w, or other model parameters. While this block-wise approach at first increases likelihood more efficiently than a quasi-Newton method applied directly to the complete parameter vector, it can become inefficient after initial iterations. For this reason, when the block-wise iterations have converged or nearly converged, we apply the quasi-Newton algorithm to the full

parameter vector, with BFGS (Broyden-Fletcher-Goldfarb-Shanno) updates of the full Hessian matrix. In our experience, these alternative approaches substantially improve the likelihood value. We use a computer cluster to search in parallel for the highest posterior density from as many as ten thousand randomly chosen starting points.

C.2. **MCMC simulations.** Our MCMC simulations are based on the Gibbs-Metropolis algorithm. The general convergence property of the Gibbs-Metropolis algorithm is discussed in Geweke (Geweke, 2005). We use the idea of Gibbs sampling to obtain the empirical joint posterior density $p(\theta_{DSGE}, \theta_{BVAR}, Q, w \mid Y_T^o)$ by sampling alternately from the following conditional posterior distributions represented by

$$p(\Theta_{DSGE} \mid Y_T^o, \Theta_{BVAR}, w),$$
$$p(\Theta_{BVAR} \mid Y_T^o, \Theta_{DSGE}, w),$$
$$p(w \mid Y_T^o, \Theta, Q),$$
$$p(Q \mid Y_T^o, \Theta, w),$$

where $\Theta = (\Theta_{DSGE}, \Theta_{BVAR})$. For each of the first three conditional posterior densities, we use the straight Metropolis algorithm with a Gaussian density as a proposal density.[14] To simulate from the last distribution, we first make a draw of the Markov chain S_T from $p(S_T \mid Y_T^o, \Theta, Q, w)$, and then draw Q from $p(Q \mid Y_T^o, \Theta, w, S_T)$. This approach has the advantage that both of these distributions are sampled directly. To draw S_T, the distributions $p(s_t \mid Y_t^o, \Theta, Q, w)$ and $p(s_t \mid Y_{t-1}^o, \Theta, Q, w)$ are obtained using the forward recursion algorithm documented in Hamilton (1989), Chib (1996), and Kim and Nelson (1999). Then s_T is drawn from $p(s_T \mid Y_T^o, \Theta, Q, w)$ and s_{T-1}, s_{T-2}, ..., s_0 are drawn recursively using

$$p\left(s_t \mid Y_T^o, \Theta, Q, w, s_T, \cdots, s_{t+1}\right) = p\left(s_t \mid Y_t^o, \Theta, Q, w, s_{t+1}\right) = \frac{q_{s_{t+1}, s_t} p\left(s_t \mid Y_t^o, \Theta, Q, w\right)}{p\left(s_{t+1} \mid Y_t^o, \Theta, Q, w\right)}.$$

The distribution $p(Q \mid Y_T^o, \Theta, w, S_T)$ is of Dirichlet if the prior is of Dirichlet, which is true in our case. It is straightforward to sample directly from the Dirichlet using the univariate gamma distribution Devroye (1986, pp. 593-594).

---

[14]We have also experimented with a Beta or Dirichlet density. There is no notable improvement in efficiency mainly because the shape of $p(w \mid Y_T^o, \Theta, Q)$ is campanulate, as shown in Figure 1.

C.3. **Computing the MDD.** Denote $\theta^* = (\Theta, Q, w)$. The marginal data density is defined as

$$p(Y_T^o) = \int p(Y_T^o \mid \theta^*)p(\theta^*) \, d\theta^*. \tag{A1}$$

We use the popular modified harmonic means method discussed in Gelfand and Dey (1994) to approximate (A1) numerically. The method is based on the following equation:

$$p(Y_T^o)^{-1} = \int_{\Theta^*} \frac{h(\theta^*)}{p(Y_T^o \mid \theta^*)p(\theta^*)} p(\theta^* \mid Y_T^o) d\theta^*, \tag{A2}$$

where Θ^* is the support of the posterior probability density and the weighting density (not just kernel) function $h(\theta^*)$ must have support that is contained in Θ^*.

A numerical evaluation of the integral on the right-hand side of (A2) is accomplished through the Monte Carlo (MC) integration

$$\hat{p}(Y_T^o)^{-1} = \frac{1}{N} \sum_{i=1}^{N} m(\theta^{*, (j)}), \tag{A3}$$

where $\theta^{*, (j)}$ is a jth draw of θ^* from the posterior distribution $p(\theta^* \mid Y_T)$ and

$$m(\theta^*) = \frac{h(\theta^*)}{p(Y_T^o \mid \theta^*)p(\theta^*)}.$$

If $m(\theta^*)$ is bounded above, the rate of convergence from this MC approximation is likely to be practical. Geweke (1999) proposes an implementation with $h(\cdot)$ being a truncated multivariate Gaussian density constructed from the posterior simulator. The tail of this Gaussian distribution is truncated to ensure that the support of $h(\cdot)$ is contained in the support of the posterior density function. When the posterior distribution is very non-Gaussian, as in our case, Sims, Waggoner, and Zha (2008) point out three sources of difficulty with this implementation. One prominent source of difficulty is that the likelihood can get to almost zero in the interior points of the parameter space Θ^*. In this situation, truncating the tail of the weight distribution does not guarantee that $m(\theta^*)$ is bounded above.

To overcome this numerical hurdle, we follow Sims, Waggoner, and Zha (2008) and choose the weighting density such that $m(\theta^*)$ is bounded above by construction. Specifically, let U be a positive number and Θ_U^* be the region defined by

$$\Theta_U^* = \left\{ \theta : p(Y_T^o \mid \theta^*)p(\theta^*) > U \right\}.$$

If $g(\theta^*)$ is any known and tractable density that is bounded above, and $h(\theta^*)$ is the density obtained by restricting $g(\theta^*)$ to Θ_U^*, then the function $m(\theta^*)$ is bounded above as well. To compute $h(\theta^*)$ from $g(\theta^*)$, we must know the probability that a draw from $g(\theta^*)$ lies in Θ_U^*. The probability is calculated using draws from the distribution given by $g(\theta^*)$.

We choose $g(\theta^*)$ from the family of elliptical distributions.[15] For instance, Gaussian distributions are elliptical. For our problem, an elliptical density function gives us flexibility to approximate the posterior density function better than a Gaussian density function. Elliptical distributions are characterized by a symmetric and positive definite matrix S, which defines the elliptical contours, a vector c, which defines the center, and non-negative one dimensional function, which defines the density across the contours. We use the estimated posterior mode to define the center, the estimated second moment of the posterior distribution to define the contours, and a step function to define the density across the contours. The step function is chosen so that the probability of lying inside an ellipse is approximately the same for the posterior and proposal distributions.

C.4. **Convergence.** To compute the MDD accurately, we take two steps. First, we must be able to compute the probability that a proposal draw lies in the region Θ_U^*, which can be interpreted as the probability of success in a Bernoulli trial. Because we make as many independent draws from the proposal as desired, this probability is accurately computed in our application.

With this probability in hand, we can use posterior draws to compute the MDD using equation (A3). To check the accuracy of this computation, we use two techniques. First, we use an extremely long sequence, one hundred million, of MCMC draws.[16] We divided this sequence into a hundred subsequences of one million draws and then computed the MDD from the entire sequence and from each of the subsequences. The variation among the subsequences is very small.

While the above technique employs many MCMC chains, the posterior mode is a starting point for each. As an alternative, we use draws from the prior as starting points for multiple MCMC chains, each of which has a length of one million draws. Selecting

---

[15]See Sims, Waggoner, and Zha (2008) for details.

[16]On a standard desktop computer with one core, the computation would have taken more than a month. We have used a cluster of computers to help speed up our computation.

an *appropriate* starting point is crucial for reliable MCMC draws. If the initial value is in an extremely low probability region, then an unreasonably long burn-in period would be required to obtain convergence of the MCMC chain. Most parameter values drawn from the prior have extremely low likelihood values, a majority below -10 in log value. Recall that the likelihood at the posterior density mode is over 6000 in log value. Thus, we draw from the prior until it reaches a reasonable likelihood value (e.g., above 3000 in log value). We use 10 such randomly selected starting points and record the minimum and maximum values of the MDDs calculated from these chains. The MDD value reported in Table 3 uses the long MCMC chain starting from the posterior mode, and the associated interval marks the minimum and maximum values of the MDDs computed from the 10 shorter chains using draws from the prior distribution as starting points.

Appendix D. The DSGE model

In this appendix we describe the complete log-linearized system for the DSGE model studied by Liu, Waggoner, and Zha (2011), along with the prior specification and the posterior estimates. The model is similar to Altig, Christiano, Eichenbaum, and Linde (2004) and Smets and Wouters (2007), with the notable exceptions that (1) some real rigidity is introduced, as in Chari, Kehoe, and McGrattan (2000), by assuming the existence of firm-specific factors (such as land) such that the sum of cost shares of capital and labor inputs is less or equal to one; and (2) a shock to the depreciation in physical capital is introduced as a stand-in for a shock to capital destruction or a financial shock.

D.1. **Linearized system.** We introduce the notation $\Delta x_t = x_t - x_{t-1}$. We use the hat variable, $\hat{x}_t$, to denote the log deviation of the stationary variable X_t from its steady state value (i.e., $\hat{x}_t = \log(X_t/X)$). The log-linearized equilibrium conditions for our DSGE mode, below, summarize the equilibrium dynamics.

$$\hat{\pi}_t - \gamma_p \hat{\pi}_{t-1} = \frac{\kappa_p}{1+\bar{\alpha}\theta_p}(\hat{\mu}_{pt} + \hat{mc}_t) + \beta E_t[\hat{\pi}_{t+1} - \gamma_p \hat{\pi}_t], \text{ (price-Phillips curve)} \qquad \text{(A4)}$$

$$\hat{w}_t - \hat{w}_{t-1} + \hat{\pi}_t - \gamma_w \hat{\pi}_{t-1} = \frac{\kappa_w}{1+\eta\theta_w}(\hat{\mu}_{wt} + \hat{mrs}_t - \hat{w}_t) +$$
$$\beta E_t[\hat{w}_{t+1} - \hat{w}_t + \hat{\pi}_{t+1} - \gamma_w \hat{\pi}_t], \text{ (wage-Phillips curve)} \qquad \text{(A5)}$$

$$\hat{q}_{kt} = S''\lambda_I^2 \left\{ \Delta \hat{i}_t + \frac{1}{1-\alpha_1}(\Delta \hat{q}_t + \alpha_2 \Delta \hat{z}_t) \right.$$

$$-\beta \mathrm{E}_t\left[\Delta\hat{i}_{t+1} + \frac{1}{1-\alpha_1}(\Delta\hat{q}_{t+1} + \alpha_2\Delta\hat{z}_{t+1})\right]\Big\}, \quad \text{(investment decision)} \qquad \text{(A6)}$$

$$\hat{q}_{kt} = \mathrm{E}_t\Big\{\Delta\hat{a}_{t+1} + \Delta\hat{U}_{c,t+1} - \frac{1}{1-\alpha_1}\left[\alpha_2\Delta\hat{z}_{t+1} + \Delta\hat{q}_{t+1}\right]$$

$$+ \frac{\beta}{\lambda_I}\left[(1-\delta)\hat{q}_{k,t+1} - \delta\hat{\delta}_{t+1} + \tilde{r}_k\hat{r}_{k,t+1}\right]\Big\}, \quad \text{(capital decision)} \qquad \text{(A7)}$$

$$\hat{r}_{kt} = \sigma_u\hat{u}_t, \quad \text{(capacity utilization)} \qquad \text{(A8)}$$

$$0 = \mathrm{E}_t\Big[\Delta\hat{a}_{t+1} + \Delta\hat{U}_{c,t+1}$$

$$- \frac{1}{1-\alpha_1}\left[\alpha_2\Delta\hat{z}_{t+1} + \alpha_1\Delta\hat{q}_{t+1}\right] + \hat{R}_t - \hat{\pi}_{t+1}\Big], \quad \text{(bond decision)} \qquad \text{(A9)}$$

$$\hat{k}_t = \frac{1-\delta}{\lambda_I}\left[\hat{k}_{t-1} - \frac{1}{1-\alpha_1}(\alpha_2\Delta\hat{z}_t + \Delta\hat{q}_t)\right]$$

$$- \frac{\delta}{\lambda_I}\hat{\delta}_t + \left(1 - \frac{1-\delta}{\lambda_I}\right)\hat{i}_t, \quad \text{(capital law of motion)} \qquad \text{(A10)}$$

$$\hat{y}_t = c_y\hat{c}_t + i_y\hat{i}_t + u_y\hat{u}_t + g_y\hat{g}_t, \quad \text{(resource constraint)} \qquad \text{(A11)}$$

$$\hat{y}_t = \alpha_1\left[\hat{k}_{t-1} + \hat{u}_t - \frac{1}{1-\alpha_1}(\alpha_2\Delta\hat{z}_t + \Delta\hat{q}_t)\right] + \alpha_2\hat{l}_t, \quad \text{(production function)} \qquad \text{(A12)}$$

$$\hat{w}_t = \hat{r}_{kt} + \hat{k}_{t-1} + \hat{u}_t - \frac{1}{1-\alpha_1}(\alpha_2\Delta\hat{z}_t + \Delta\hat{q}_t) - \hat{l}_t, \quad \text{(labor \& capital demand)} \qquad \text{(A13)}$$

$$\hat{R}_t = \rho_r\hat{R}_{t-1} + (1-\rho_r)\left[\phi_\pi\hat{\pi}_t + \phi_y\hat{y}_t\right] + \sigma_r\varepsilon_{rt}, \quad \text{(interest rate rule)} \qquad \text{(A14)}$$

where

$$\hat{mc}_t = \frac{1}{\alpha_1+\alpha_2}[\alpha_1\hat{r}_{kt} + \alpha_2\hat{w}_t] + \bar{\alpha}\hat{y}_t, \qquad \text{(A15)}$$

$$\hat{mrs}_t = \eta\hat{l}_t - \hat{U}_{ct}, \qquad \text{(A16)}$$

$$\hat{U}_{ct} = \frac{\beta b(1-\rho_a)}{\lambda^* - \beta b}\hat{a}_t - \frac{\lambda^*}{(\lambda^*-b)(\lambda^*-\beta b)}[\lambda^*\hat{c}_t - b(\hat{c}_{t-1} - \Delta\hat{\lambda}_t^*)]$$

$$+ \frac{\beta b}{(\lambda^*-b)(\lambda^*-\beta b)}[\lambda^*\mathrm{E}_t(\hat{c}_{t+1} + \Delta\hat{\lambda}_{t+1}^*) - b\hat{c}_t], \qquad \text{(A17)}$$

Note that $\hat{\pi}_t$ is inflation, $\hat{w}_t$ is real wage, $\hat{q}_{kt}$ is the shadow price of existing capital (Tobin's q), $\hat{i}_t$ is investment, $\hat{q}_t$ is the biased technology shock process, $\hat{z}_t$ is the neutral technology shock process, $\hat{a}_t$ is the risk premium (preference) shock process, $\hat{u}_t$ is the utilization rate of capital, $\hat{r}_{kt}$ is the real rental price of capital, $\hat{\delta}_t$ is the capital depreciation shock process, $\hat{R}_t$ is the nominal rate of interest, $\hat{k}_t$ is the capital stock, $\hat{y}_t$ is output, $\hat{c}_t$ is consumption, $\hat{g}_t$ is government spending, and $\hat{l}_t$ is hours worked.

The steady-state variables are given by

$$\tilde{r}_k = \frac{\lambda_I}{\beta} - (1-\delta), \qquad \text{(A18)}$$

$$u_y \equiv \frac{\tilde{r}_k\tilde{K}}{\tilde{Y}\lambda_I} = \frac{\alpha_1}{\mu_p}, \qquad \text{(A19)}$$

$$i_y = [\lambda_I - (1 - \delta)] \frac{\alpha_1}{\mu_p \tilde{r}_k}, \qquad (A20)$$

$$c_y = 1 - i_y - g_y. \qquad (A21)$$

The new parameters introduced in the above equilibrium conditions are

$$\lambda_I = (\lambda_q \lambda_z^{\alpha_2})^{\frac{1}{1-\alpha_1}},$$

$$\lambda^* = (\lambda_z^{\alpha_2} \lambda_q^{\alpha_1})^{\frac{1}{1-\alpha_1}},$$

$$\Delta \hat{\lambda}_t^* = \frac{1}{1 - \alpha_1}(\alpha_1 \Delta \hat{q}_t + \alpha_2 \Delta \hat{z}_t),$$

$$\theta_p = \frac{\mu_p}{\mu_p - 1},$$

$$\kappa_p = \frac{(1 - \beta \xi_p)(1 - \xi_p)}{\xi_p},$$

$$\bar{\alpha} = \frac{1 - \alpha_1 - \alpha_2}{\alpha_1 + \alpha_2},$$

$$\theta_w \equiv \frac{\mu_w}{\mu_w - 1},$$

$$\kappa_w = \frac{(1 - \beta \xi_w)(1 - \xi_w)}{\xi_w}.$$

Note that g_y is the average ratio of government spending to output, c_y is the average ratio of consumption to output, i_y is the average ratio of investment to output, μ_{pt} is the average price markup, μ_{wt} is the average wage markup, λ_q is the growth rate of investment-specific technology, λ_z is the growth rate of neutral technology, α_1 is the cost share of capital input, α_2 is the cost share of labor input, δ is the average capital depreciation rate, b is internal habit, S'' represents the investment adjustment costs, σ_u represents the curvature of the cost function of variable capital utilization, ξ_p is the probability that a firm cannot adjust its price, γ_p measures the degree of price indexation, ξ_w is a fraction of households who cannot reoptimize their wage decisions, and γ_w measures the degree of wage indexation.

In addition to all the equilibrium conditions, we have 7 shock processes:

$$\log \mu_{wt} = (1 - \rho_w) \log \mu_w + \rho_w \log \mu_{w,t-1} + \sigma_w \varepsilon_{wt} - \phi_w \sigma_w \varepsilon_{w,t-1}, \text{ (price markup)}$$

$$\log \mu_{pt} = (1 - \rho_p) \log \mu_p + \rho_p \log \mu_{p,t-1} + \sigma_p \varepsilon_{pt} - \phi_p \sigma_p \varepsilon_{p,t-1}, \text{ (wage markup)}$$

$$\log z_t = (1 - \rho_z) \log z + \rho_z \log z_{t-1} + \sigma_z \varepsilon_{zt}, \text{ (neutral technology)}$$

$$\log q_t = (1 - \rho_q) \log q + \rho_q \log q_{t-1} + \sigma_q \varepsilon_{qt}, \text{ (embodied technology)}$$

$$\log A_t = (1 - \rho_a) \log A + \rho_a \log A_{t-1} + \sigma_a \varepsilon_{at}, \text{ (risk premium)}$$

$$\log \delta_t = (1 - \rho_d) \log \delta + \rho_d \log \delta_{t-1} + \sigma_d \varepsilon_{dt}, \text{ (capital depreciation)}$$

$$\log \tilde{G}_t = (1 - \rho_g) \log \tilde{G} + \rho_g \log \tilde{G}_{t-1} + \sigma_g \varepsilon_{gt} + \rho_{gz} \sigma_z \varepsilon_{zt}, \text{ (spending)}$$

where ε represents an i.i.d. normal shock and σ represents the corresponding standard deviation.

To compute the equilibrium, we eliminate both $\hat{u}_t$ and $\hat{r}_{kt}$ by using (A8) and (A11), leaving 9 equations and 9 variables $\hat{\pi}_t$, $\hat{w}_t$, $\hat{i}_t$, $\hat{q}_{kt}$, $\hat{c}_t$, $\hat{k}_t$, $\hat{y}_t$, $\hat{l}_t$, and $\hat{R}_t$. Out of these 9 variables, we have 7 corresponding observable variables (except $\hat{q}_{kt}$ and $\hat{k}_t$) for our estimation. Finally, we have one additional observable variable used in our estimation: the biased technology shock $\hat{q}_t$.

In addition to the 9 equilibrium conditions, we have 7 equations describing the AR processes for the 7 structural shocks, 4 equations describing the 2 MA processes, and 7 equations concerning the 7 expectational terms in the system. Thus, there are 27 DSGE equations in total.

A standard solution technique, such as the method proposed by Sims (2002), can be directly applied to these 27 equations. The solution leads to the following VAR(1) form of state equations:

$$f_t = F f_{t-1} + \Phi \varepsilon_t, \tag{A22}$$

where $\varepsilon_t = [\varepsilon_{rt}, \varepsilon_{pt}, \varepsilon_{wt}, \varepsilon_{gt}, \varepsilon_{zt}, \varepsilon_{at}, \varepsilon_{dt}, \varepsilon_{qt}]'$, f_t is a 27×1 vector of variables in the log-linearized system, and F and Φ are matrix functions of model parameters.

Let y_t be a 8×1 vector of *observable* represented as

$$y_t = \left[\Delta \ln Y_t^{\text{Data}}, \Delta \ln C_t^{\text{Data}}, \Delta \ln I_t^{\text{Data}}, \Delta \ln w_t^{\text{Data}}, \ln \pi_t^{\text{Data}}, \Delta \ln Q_t^{\text{Data}}, \ln L_t^{\text{Data}}, \text{FFR}_t^{\text{Data}}\right]'.$$

The observable vector is connected to the model (state) variables through the measurement equations

$$y_t = a + H f_t,$$

where

$$a = \left[\log \lambda^*, \log \lambda^*, \log \lambda^*, \log \lambda^*, \log \pi, \log \lambda_q, \log L, \log R\right]'. \tag{A23}$$

The estimation applies to this state space form.

D.2. **The prior.** The prior for the DSGE model is reported in Tables 4 and 5. Instead of specifying the mean and the standard deviation, we use the 90% probability interval to back out the hyperparameter values of the prior distribution. The intervals are chosen to be wide enough to allow for the possibility that the posterior mode is close to or on the boundary of the parameter space. The wide intervals also allow for the possibility of multiple local posterior peaks (Del Negro and Schorfheide, 2008). This

approach to choosing the prior is useful to deal with skewed distributions. It allows for reasonable hyperparameter values in certain distributions, such as the Inverse-Gamma, where the first two moments may not exist.

For many parameters with the Beta prior distribution, such as the habit parameter and the persistence parameters in shock processes, we insist on a positive probability density at the value zero to allow for the possibility of no habit and no persistence at all. On the other hand, we insist on zero probability density at the value 1 to maintain the assumption that the economy is on the balanced growth path. Consequently, the two hyperparameter values for the Beta prior are set at 1.0 and 2.0.

The prior for the labor share and capital share is the Beta distribution with the restriction $\alpha_1 + \alpha_2 \leq 1$ such that the production technology requires firm-specific factors (Chari, Kehoe, and McGrattan, 2000). If we treated α_1 and α_2 independently, the 90% probability bounds for the α_1 values would be 0.3 and 0.4 and those for α_2 would be 0.5 and 0.7. With the restriction $\alpha_1 + \alpha_2 \leq 1$ imposed in this paper, however, the joint 90% probability region implies that the 90% probability bounds will be different.

The prior for the inverse Frisch elasticity η follows the Gamma distribution. We choose the two hyperparameters of the Gamma distribution such that the lower bound (0.2) and the upper bound (10.0) of η constitute the 90% probability interval. This prior range for η implies that the Frisch elasticity lies between 0.1 and 5.

The lower and upper bounds of prior distributions for the parameters λ_q, λ^*, β, σ_u, S'', δ, ξ_p, γ_p, ξ_w, γ_w, ϕ_π, ϕ_y, and π^* are specified in Table 4. Using these wide bounds, we back out the two hyperparameter values for the corresponding prior distributions.

The Gamma prior for the average net price markup $\mu_p - 1$ is the same as the Gamma prior for the average net wage markup $\mu_w - 1$. By setting the first hyperparameter of this prior to be 1.0, we allow for a positive probability that the net markups may be zero. This generality (a less stringent prior) turns out to be critical, as our posterior estimates of $\mu_p - 1$ and $\mu_w - 1$ are nearly zero. We set the second hyperparameter of the Gamma prior at 5.5 such that the implied 90% probability bounds are wide enough (from 0.0094 to 0.5446).

The prior for the parameter ρ_{gz}, capturing the impact of technological improvement on government spending, is the Gamma distribution with the 90% probability bounds given by $[0.2, 3.0]$.

The standard deviation of each of the 8 shocks has the Inverse Gamma prior distribution with the 90% probability bounds given by $[0.0005, 1.0]$. These wide bounds are necessary to take account of the possibility that some shocks may have very small variances while others may have very large variances. The two hyperparameters implied by these bounds, as reported in Table 5, indicate that there exist no moments for this Inverse Gamma prior.

D.3. **Posterior estimates.** The prior specified for the DSGE model is looser and more agnostic in this paper than most priors used in the DSGE literature. The agnostic prior comes also with a price: since the likelihood function for the Markov-switching mixture model is complicated and full of multiple local peaks, the resulting posterior density function is complicated as well. The non-Gaussian nature of the posterior density implies that the posterior mean may have a very low probability and thus cannot represent the most likely outcome for the model. The posterior mode is, by definition, the most probable point in the parameter space, regardless of how non-Gaussian and complicated the shape of the posterior probability density is. Moreover, using a point in the neighborhood of the posterior mode as a starting point for the MCMC algorithm avoids the situation where a long sequence of posterior draws gets stuck in the low probability region due to a poor starting point.

Tables 6 and 7 report the posterior-mode estimates of the DSGE model parameters along with the 90% marginal probability intervals. In these tables we contrast the estimated results for the Markov-switching mixture model to those for the DSGE model when estimated alone (we call it "DSGE-only"). Despite the fact that the mixture model discounts a great many observations used for estimation of the DSGE parameters, a number of the estimated DSGE parameters from the mixture model are similar to those from the DSGE-only model. For instance, the estimate of the average price markup is close to zero, similar to the estimate in the DSGE-only model. This result implies that the demand curve for differentiated goods is very flat. Thus, a small increase in the relative price can lead to large declines in relative output demand. Even if firms can re-optimize their pricing decisions frequently, they choose not to adjust

their relative prices too much. In other words, the small average markup and thus the large demand elasticity become a source of strategic complementarity in firms' pricing decisions.

The strength of strategic complementarity is measured by the price Phillips-curve slope parameter:

$$ps = \frac{\kappa_p}{1 + \bar{\alpha}\theta_p}.$$

The smaller the value of ps is, the stronger strategic complementarity is. According to the posterior estimates for the DSGE model alone (Table 6), we have $\mu_p = 1.00019$, $\alpha_1 = 0.177$, $\alpha_2 = 0.804$, $\beta = 0.9977$, and $\xi_p = 0.372$. Thus, the Phillips-cure slope parameter is $pc = 0.0103$. If there were no real rigidity (i.e., $\alpha_1 + \alpha_2$ were equal to one exactly), we would have $pc = \kappa_p = 1.0616$. This weak strategic complementarity would imply a fairly large response of inflation or the price level to a structural shock. But our estimated Phillips-curve slope parameter is much smaller. To attain such a small value ($pc = 0.0103$) without any real rigidity, the price sticky parameter would have to be $\xi_p = 0.90$, implying the average duration of two and a half years before prices change.

The general pattern, as indicated by the 90% probability intervals, is that the Markov-switching mixture model exposes more uncertainty about the estimated DSGE parameters than what is implied when the DSGE model is treated as the truth and estimated alone. In many cases, such as the inverse Frisch elasticity of labor supply (η), the curvature of the capital utilization cost function evaluated at the steady state (σ_u), and the curvature of the adjustment cost function at the steady state (S''), the probability distributions have changed to be heavily skewed to higher values. For instance, the posterior distribution of η is so skewed that the mode is outside the 90% probability interval.[17]

In addition to changes of probability intervals, many of the estimated DSGE parameters from the mixture model are different from the DSGE-only model. For instance,

---

[17]Remember that the number of parameters combined from the two models in the mixture is very large and the shape of the posterior probability density over this high-dimensional parameter space is extremely non-Gaussian full of skewness and fat-tails. When we compute the marginal 90% probability interval of one parameter by *integrating out all the rest of the parameters*, it is not uncommon that some posterior mode estimates fall outside the 90% probability intervals as indicated in Tables 6 and 7.

the estimate of β is much smaller for the mixture model than for the DSGE-only model. Both the biased technology growth rate (λ_q) and the output growth rate (λ^*) are estimated to be much smaller from the mixture model than from the DSGE-only model. These results are intuitive because the DSGE model in the mixture plays an important role only in the late 1970s and early 1980s (see Figure 2). These are the times when the U.S. economy experiences three large recessions in a very short period of time and the growth rates are slower than the rest of the sample.

Perhaps the most notable are changes pertaining to every persistence parameter. As shown in Table 7, the 90% probability intervals for persistence parameters are much wider in the mixture model than in the DSGE-only model. Specifically, the posterior distributions for persistence parameters tend to have a long fat tail toward zero, indicating much more uncertainty about the persistence of a shock than the inference from the DSGE-only model. The estimates of persistence parameters themselves from the mixture model are considerably smaller than those estimates from the DSGE-only model.

Another notable example pertains to the estimated results for the capital depreciation shock process. The estimate of the shock standard deviation (σ_d) from the mixture model is considerably larger than that from the DSGE-only model. Moreover, the 90% probability interval indicates that the marginal distribution of the shock standard deviation is skewed heavily to a very high value. The estimated persistence parameter (ρ_d), on the other hand, is smaller than that from the DSGE-only model. The probability interval implies a large amount of probability of lower values of the persistence parameter than the value at the posterior mode. In the main text, we compare the impulse responses to a depreciation shock from the mixture model to those from the DSGE-only model.

TABLE 4. Prior distributions of DSGE structural parameters

Parameters	Description	Prior				
		Distributions	α_{prior}	β_{prior}	5%	95%
General parameters						
b	Habit	Beta	1.0	2.0	0.025	0.776
α_1	Capital share	Beta	85.5869	159.4377	0.3	0.4
α_2	Labor share	Beta	38.4721	25.4535	0.5	0.7
η	1/(Frisch elasticity)	Gamma	1.0576	0.3106	0.2	10
$100(\lambda_q - 1)$	Biased tech growth	Gamma	1.8611	3.0112	0.1	1.5
$100(\lambda^* - 1)$	Output growth	Gamma	1.8611	3.0112	0.1	1.5
$100\left(\beta^{-1} - 1\right)$	Discount factor	Gamma	1.5832	1.0126	0.2	4.0
Firm parameters						
σ_u	Utilization cost	Gamma	3.7790	2.4791	0.5	3.0
S''	Adjustment cost	Gamma	1.0576	0.6213	0.5	5.0
$\mu_p - 1$	Price markup	Gamma	1.0	5.5	0.0094	0.5446
$\mu_w - 1$	Wage markup	Gamma	1.0	5.5	0.0094	0.5446
4δ	Depreciation	Beta	5.4257	41.4890	0.05	0.2
ξ_p	Calvo pricing	Beta	2.0384	3.0426	0.1	0.75
γ_p	Price indexation	Beta	1.0	1.0	0.05	0.95
ξ_w	Calvo wage	Beta	2.0384	3.0426	0.1	0.75
γ_w	Wage indexation	Beta	1.0	1.0	0.05	0.95
Policy parameters						
ρ_r	Interest persistence	Beta	1.0	2.0	0.025	0.776
ϕ_π	Inflation coef	Gamma	2.4373	1.0876	0.5	5.0
ϕ_y	Output coef	Gamma	1.0	1.0	0.05	3.0
$400 \log \pi^*$	Inflation target	Gamma	2.9043	0.7690	1.0	8.0

Note: "5%" and "95%" demarcate the low and high bounds of the 90% probability interval.

TABLE 5. Prior distributions of DSGE shock parameters

Parameters	Description	Prior				
		Distributions	α_{prior}	β_{prior}	5%	95%
Persistence parameters						
ρ_p	Price markup AR	Beta	1.0	2.0	0.025	0.776
ϕ_p	Price markup MA	Beta	1.0	2.0	0.025	0.776
ρ_w	Wage markup AR	Beta	1.0	2.0	0.025	0.776
ϕ_w	Wage markup MA	Beta	1.0	2.0	0.025	0.776
ρ_{gz}	Spending on tech	Gamma	1.8611	1.5056	0.2	3.0
ρ_a	Preference	Beta	1.0	2.0	0.025	0.776
ρ_q	Biased tech	Beta	1.0	1.0	0.05	0.95
ρ_z	Neutral tech	Beta	1.0	1.0	0.05	0.95
ρ_d	Depreciation	Beta	1.0	2.0	0.025	0.776
Standard deviations						
σ_r	Monetary policy	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_p	Price markup	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_w	Wage markup	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_g	Gov spending	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_z	Neutral tech	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_a	Preference	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_q	Biased tech	Inverse Gamma	0.4436	0.0009	0.0005	1.0
σ_d	Depreciation	Inverse Gamma	0.4436	0.0009	0.0005	1.0

Note: "5%" and "95%" demarcate the low and high bounds of the 90% probability interval.

TABLE 6. Posterior distributions of DSGE structural parameters

		DSGE model alone			Markov mixture model		
Parameters	Description	Mode	5%	95%	Mode	5%	95%
General parameters							
b	Habit	0.544	0.493	0.624	0.596	0.544	0.881
α_1	Capital share	0.177	0.151	0.203	0.321	0.251	0.342
α_2	Labor share	0.804	0.747	0.818	0.675	0.566	0.707
η	1/(Frisch elasticity)	0.005	0.003	0.167	0.009	0.122	7.397
$100(\lambda_q - 1)$	Biased tech growth	1.507	1.215	1.911	0.763	0.198	0.948
$100(\lambda^* - 1)$	Output growth	0.483	0.400	0.569	0.253	0.052	0.430
$100\,(\beta^{-1} - 1)$	Discount factor	0.228	0.081	0.909	0.822	0.402	1.441
Firm parameters							
σ_u	Utilization cost	2.018	1.404	3.787	0.620	0.226	1.797
S''	Adjustment cost	0.800	0.608	1.278	0.746	0.288	4.032
$\mu_p - 1$	Price markup	0.000	0.000	0.001	0.000	0.000	0.386
$\mu_w - 1$	Wage markup	0.003	0.015	0.176	0.003	0.010	0.603
4δ	Depreciation	0.145	0.064	0.204	0.076	0.008	0.153
ξ_p	Calvo pricing	0.372	0.308	0.760	0.406	0.349	0.926
γ_p	Price indexation	0.121	0.028	0.408	0.775	0.147	0.968
ξ_w	Calvo wage	0.303	0.269	0.606	0.312	0.231	0.808
γ_w	Wage indexation	0.790	0.088	0.954	0.537	0.075	0.961
Policy parameters							
ρ_r	Interest persistence	0.618	0.572	0.687	0.457	0.327	0.742
ϕ_π	Inflation coef	1.480	1.392	1.693	1.388	1.190	2.472
ϕ_y	Output coef	0.066	0.052	0.101	0.166	0.056	0.558
$400 \log \pi^*$	Inflation target	5.576	3.863	10.109	4.216	1.022	9.399

Note: "5%" and "95%" demarcate the low and high bounds of the 90% probability interval.

TABLE 7. Posterior distributions of DSGE shock parameters

Parameters	Description	DSGE model alone			Markov mixture model		
		Mode	5%	95%	Mode	5%	95%
Persistence parameters							
ρ_p	Price markup AR	0.786	0.587	0.878	0.526	0.087	0.903
ϕ_p	Price markup MA	0.627	0.276	0.820	0.377	0.019	0.756
ρ_w	Wage markup AR	0.992	0.987	0.997	0.730	0.087	0.878
ϕ_w	Wage markup MA	0.530	0.305	0.827	0.048	0.022	0.713
ρ_{gz}	Spending on tech	0.947	0.490	1.348	1.961	0.224	2.118
ρ_a	Preference	0.988	0.973	0.995	0.400	0.112	0.777
ρ_q	Biased tech	0.994	0.988	0.997	0.992	0.962	0.998
ρ_z	Neutral tech	0.942	0.927	0.961	0.923	0.898	0.996
ρ_d	Depreciation	0.915	0.854	0.975	0.813	0.674	0.945
Standard deviations							
σ_r	Monetary policy	0.003	0.002	0.003	0.004	0.003	0.006
σ_p	Price markup	1.012	0.593	2.109	0.707	0.031	1.798
σ_w	Wage markup	0.023	0.017	0.065	0.025	0.053	4.405
σ_g	Gov spending	0.029	0.026	0.031	0.034	0.024	0.050
σ_z	Neutral tech	0.008	0.007	0.009	0.009	0.009	0.017
σ_a	Preference	0.061	0.035	0.137	0.010	0.013	0.053
σ_q	Biased tech	0.006	0.006	0.007	0.006	0.005	0.008
σ_d	Depreciation	0.096	0.065	0.261	0.273	0.210	5.035

Note: "5%" and "95%" demarcate the low and high bounds of the 90% probability interval.

References

ALTIG, D., L. J. CHRISTIANO, M. EICHENBAUM, AND J. LINDE (2004): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," Federal Reserve Bank of Cleveland Working Paper 04-16.

BATES, J. M., AND C. W. J. GRANGER (1969): "The Combination of Forecasts," *Operational Research Quarterly*, 20(4), 451–468.

BERNANKE, B. S. (1986): "Alternative Exploration of the Money-Income Correction," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49–99.

BERNANKE, B. S., M. GERTLER, AND M. W. WATSON (1997): "Systematic Monetary Policy and the Effects of Oil Price Shocks," *Brookings Papers on Economic Activity*, 1, 91–142.

BLANCHARD, O. J., AND M. W. WATSON (1986): "Are Business Cycles All Alike?," in *The American Business Cycle: Continuity and Change*, ed. by R. Gordon, pp. 123–156. University of Chicago Press, Chicago, Illinois.

BROCK, W. A., S. N. DURLAUF, AND K. D. WEST (2003): "Policy Evaluation in Uncertain Economic Environment," *Brookings Papers on Economic Activity*, 1, 235–301.

CHARI, V., P. J. KEHOE, AND E. R. MCGRATTAN (2000): "Sticky Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem?," *Econometrica*, 68(5), 1151–1179.

CHIB, S. (1996): "Calculating Posterior Distributions and Model Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 79–97.

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (1999): "Monetary Policy Shocks: What Have We Learned and To What End?," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1A, pp. 65–148. North-Holland, Amsterdam, Holland.

———— (2005): "Nominal Rigidities and the Dynamics Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

COGLEY, T., AND T. J. SARGENT (2005): "The Conquest of U.S. Inflation: Learning and Robustness to Model Uncertainty," *Review of Economic Dynamics*, 8, 528–563.

CUMMINS, J. G., AND G. L. VIOLANTE (2002): "Investment-Specific Technical Change in the United States (1947-2000): Measurement and Macroeconomic Consequences," *Review of Economic Dynamics*, 5, 243–284.

DEL NEGRO, M., AND F. SCHORFHEIDE (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643–673.

——— (2008): "Forming Priors for DSGE Models (and How It Affects the Assessment of Nominal Rigidities)," Manuscript, Federal Reserve Bank of New York and University of Pennsylvania.

——— (2009): "Monetary Policy Analysis with Potentially Misspecified Models," *American Economic Review*, 99(4), 1415–1450.

——— (Forthcoming): "Bayesian Macroeconometrics," in *Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. van Dijk. Oxford University Press.

DENTON, F. T. (1971): "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization," *Journal of the American Statistical Association*, 66, 99–102.

DEVROYE, L. (1986): *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

DIEBOLD, F. X. (1991): "A Note on Bayesian Forecast Combination Procedures," in *Economic Structural Change: Analysis and Forecasting*, ed. by A. H. Westlund, and P. Hackl, pp. 225–232. Springer-Verlag, New York, NY.

FISHER, M., AND D. F. WAGGONER (2010): "Mixture Models and Bayesian Model Selection," Unpublished manuscript.

FRÜHWIRTH-SCHNATTER, S. (2006): *Finite Mixture and Markov Switching Models*. Springer, New York.

GELFAND, A. E., AND D. K. DEY (1994): "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society (Series B)*, 56, 501–514.

GEWEKE, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18(1), 1–73.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc., Hoboken, New Jersey.

GEWEKE, J., AND G. AMISANO (2011): "Optimal Prediction Pools," *Journal of Econometrics*, 164, 130–141.

———— (Forthcoming): "Analysis of Variance for Bayesian Inference," *Econometric Reviews.*

GORDON, R. J. (1990): *The Measurement of Durable Goods Prices.* University of Chicago Press, Chicago,Illinois.

HAMILTON, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

HAMILTON, J. D., D. F. WAGGONER, AND T. ZHA (2007): "Normalization in Econometrics," *Econometric Reviews*, 26(2-4), 221–252.

HANSEN, L. P., AND T. J. SARGENT (2001): "Acknowledging Misspecification in Macroeconomic Theory," *Review of Economic Dynamics*, 4, 519–535.

———— (2010): "Wanting Robustness in Macroeconomics," Unpublished Manuscript.

HARRISON, P. J., AND C. F. STEVENS (1976): "Bayesian Forecasting," *Journal of the Royal Statistical Society (Series B)*, 38(3), 205–247.

INGRAM, B. F., AND C. H. WHITEMAN (1994): "Supplanting the "Minnesota" Prior: Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors," *Journal of Monetary Economics*, 34(3), 497–510.

KIM, C.-J., AND C. R. NELSON (1999): *State-Space Models with Regime Switching.* MIT Press, London, England and Cambridge, Massachusetts.

LEEPER, E. M., C. A. SIMS, AND T. ZHA (1996): "What Does Monetary Policy Do?," *Brookings Papers on Economic Activity*, 2, 1–78.

LIU, Z., D. F. WAGGONER, AND T. ZHA (2011): "Sources of Macroeconomic Fluctuations: A Regime-Switching DSGE Approach," *Quantitative Economics*, 2, 251–301.

MCCULLOCH, R. E., AND R. S. TSAY (1994): "Bayesian Inference of Trend- and Difference-Stationarity," *Econometric Th*, 10, 596–608.

SIMS, C. A. (1986): "Are Forecasting Models Usable for Policy Analysis?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 2–16.

———— (1992): "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review*, 36, 975–1011.

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

SIMS, C. A., D. F. WAGGONER, AND T. ZHA (2008): "Methods for Inference in Large Multiple-Equation Markov-Switching Models," *Journal of Econometrics*,

146(2), 255–274.

SIMS, C. A., AND T. ZHA (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949–968.

――― (1999): "Error Bands for Impulse Responses," *Econometrica*, 67(5), 1113–1155.

――― (2006): "Were There Regime Switches in US Monetary Policy?," *American Economic Review*, 96, 54–81.

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

WEST, M., AND J. HARRISON (1997): *Bayesian Forecasting and Dynamic Models*. Springer, 2nd edn.

FEDERAL RESERVE BANK OF ATLANTA, FEDERAL RESERVE BANK OF ATLANTA, EMORY UNIVERSITY, SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS, AND NBER