JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **15**: 645–670 (2000)

LOSS FUNCTION-BASED EVALUATION OF DSGE MODELS

FRANK SCHORFHEIDE*
Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA

SUMMARY

In this paper we propose a Bayesian econometric procedure for the evaluation and comparison of DSGE models. Unlike in many previous econometric approaches we explicitly take into account the possibility that the DSGE models are misspecified and introduce a reference model to complete the model space. Three loss functions are proposed to assess the discrepancy between DSGE model predictions and an overall posterior distribution of population characteristics that the researcher is trying to match. The evaluation procedure is applied to the comparison of a standard cash-in-advance (CIA) and a portfolio adjustment cost (PAC) model. We find that the CIA model has higher posterior probability than the PAC model and achieves a better in-sample time series fit. Both models overpredict the magnitude of the negative correlation between output growth and inflation. However, unlike the PAC model, the CIA model is not able to generate a positive real effect of money growth shocks on aggregate output. Overall, the impulse response dynamics of the PAC model resemble the posterior mean impulse response functions more closely than the responses of the CIA model. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

The quantitative evaluation of dynamic stochastic general equilibrium (DSGE) models in macroeconomics is often conducted without formal statistical methods. While DSGE models provide a complete multivariate stochastic process representation for the data, they impose very strong restrictions on actual time series and are in many cases rejected against less restrictive specifications such as vector autoregressions (VAR). Nevertheless, DSGE models are interesting from a theoretical perspective and it is important to assess their empirical adequacy. This paper provides an econometric procedure that enables a coherent model evaluation. Our approach is applied to the evaluation of two types of standard monetary models: a cash-in-advance (CIA) model and its extension with portfolio adjustment costs (PAC).

The potential misspecification of the candidate models poses a conceptual difficulty for the design of econometric procedures. The inference problem is not to determine whether a particular model, for example the CIA or PAC model in our application, is 'true'. Instead, the goal is to determine which model summarizes the regular features of the data more accurately. Our evaluation procedure differs from existing econometric approaches with respect to the treatment of candidate model misspecification and the explicit incorporation of loss functions.

Macroeconomists generally judge DSGE models based on their ability to replicate realistic patterns of co-movements among key macroeconomic variables and impulse responses to structural shocks, such as unexpected changes in the growth rate of money supply or total factor

*Correspondence to: Frank Schorfheide, Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA. e-mail: schorf@ssc.upenn.edu

Received 23 September 1999
Revised 24 August 2000

productivity. However, neither population moments nor impulse response functions are directly observable in the data. Therefore we will construct a probabilistic representation of the data that serves as a benchmark for the DSGE model comparison. *A priori* this benchmark is a mixture of the DSGE models and a reference model. The reference model is introduced to cope with the potential misspecification of the DSGE models. It should impose weaker restrictions on the data than the DSGE models and achieve an acceptable fit. In our application, we use a vector autoregression (VAR). Based on this mixture of DSGE models and a reference model, it is possible to compute posterior distributions for correlations, impulse response functions, and other population characteristics. The mixture weights for DSGE models and reference model are given by the posterior model probabilities and reflect the statistical fit of the various models.

For each DSGE model predicted, correlations and impulse response functions are computed. A loss function that penalizes deviations from actual and predicted population characteristics is chosen. Then the expected loss of the DSGE model predictions is calculated. The expectation is taken with respect to the overall posterior distribution that takes the reference model into account. The DSGE model that attains the smallest posterior prediction loss wins the comparison. This idea is generalized to take parameter uncertainty into account, and to derive loss function-based parameter estimates. Due to the explicit consideration of loss functions defined over time series characteristics that are of interest to macroeconomists, our procedure goes beyond model determination based on posterior probabilities or Bayes factors, e.g. Kass and Raftery (1995) and Phillips (1996). The evaluation is robust, if the results are insensitive to the choice of loss function.

The evaluation approach remains sensible even in a situation where all DSGE models have posterior probabilities that are close to or equal to zero. We will argue that under special conditions our evaluation procedure resembles a calibration approach (Kydland and Prescott, 1996), in which point predictions of DSGE models are simply compared to sample moments without systematically taking standard errors of moment estimates into account. On the other hand, if a parsimonious DSGE model has very high posterior probability in the presence of the more densly parameterized reference model, this DSGE model is likely to win the model comparisons regardless of the employed loss functions.

Many evaluation techniques that were previously proposed in the literature are based on *p*-values for various sample characteristics of the data. The *p*-values measure how far transformations of the data fall in the tails of their respective sampling distributions, that are derived from the DSGE models. A non-exhaustive list of examples are Canova *et al.* (1994), Christiano and Eichenbaum (1992a), Nason and Cogley (1994), Smith (1993), and Söderlind (1994). Canova (1994) computes Bayesian versions of these *p*-values following Box (1980). However, *p*-values are not designed for model comparisons. Moreover, in cases where it is believed that the structural models are severely misspecified it is implausible to use sampling distributions as a benchmark, that were derived from misspecified models.

As an alternative, Diebold *et al.* (1998) propose a loss function-based evaluation of DSGE model predictions under the sampling distribution obtained from a non-parametric spectral representation of the data. In the language of this paper, the conceptual drawback of their approach is that their non-parametric representation (reference model) has always posterior probability one, regardless of the statistical fit of the DSGE models. Our approach takes the potential impact of the DSGE model predictions on the shape of the overall posterior distribution of population characteristics into account. Despite low posterior probability of

DSGE models, this impact can be significant because DSGE models generally deliver more concentrated predictive distributions than densly parameterized reference models.

Dejong *et al.* (1996) propose a Bayesian approach to calibration which also assigns zero probability to the structural model. In particular, they focus on the evaluation of DSGE models that generate singular probability distributions for the data. A measure of overlap between the posterior distribution of unconditional moments obtained from the reference model and the prior predictive distribution from a structural model is proposed. Geweke (1999a) shows that a refinement of the overlap criterion can be interpreted as posterior odds ratio for DSGE models conditional on a reference model. He assumes that DSGE models only claim to describe population characteristics, such as autocorrelations, but not the data itself. We, however, adopt the traditional interpretation of macroeconometric models and regard DSGE models as probabilistic representations of the data. Our posterior model probabilities document to what extent the DSGE models are able to track the observed time series, and the loss function based evaluation documents to what extent DSGE model predictions of population characteristics are consistent with empirical evidence.

The paper is structured as follows. The CIA and the PAC model are presented in Section 2. We will examine three questions. Are the two models able to track US output growth and inflation data? Which model generates a more realistic correlation pattern between output growth and present and lagged inflation? Which model should be used to calculate the predicted change of output growth due to an unanticipated monetary expansion? The model evaluation approach that is used to answer these questions is described in Section 3. Section 4 contains the empirical analysis and Section 5 concludes. Computational details are provided in the Appendix.

2. TWO MONETARY DSGE MODELS

A standard CIA model and a cash-in-advance model with portfolio adjustment costs (PAC) are considered. Detailed discussions of the models can be found, for instance, in Christiano (1991), Christiano and Eichenbaum (1992b), and Nason and Cogley (1994). To make this paper self-contained we briefly review the model specifications. The model economies consist of a representative household, a firm, and a financial intermediary. Output is produced according to a Cobb–Douglas production function

$$GDP_t = K_t^{\alpha}(A_t N_t)^{1-\alpha} \tag{1}$$

where K_t denotes the capital stock (predetermined at the beginning of period t), N_t is the labour input and A_t is a labour-augmenting technology.

The model economies are pertubated by two exogenous processes. Technology evolves according to

$$\ln A_t = \gamma + \ln A_{t-1} + \epsilon_{A,t} \qquad \epsilon_{A,t} \sim \mathcal{N}(0, \sigma_A^2) \tag{2}$$

The central bank lets the money stock M_t grow at rate $m_t = M_{t+1}/M_t$:

$$\ln m_t = (1-\rho)\ln m^* + \rho \ln m_{t-1} + \epsilon_{M,t} \qquad \epsilon_{M,t} \sim \mathcal{N}(0, \sigma_M^2) \tag{3}$$

Equation (3) can be interpreted as a simple monetary policy rule without feedbacks. The innovations $\epsilon_{M,t}$ capture unexpected changes of the money growth rate due to 'normal' policy making in the terminology of Sims (1982). Changes in m^* or ρ correspond to rare regime shifts.

At the beginning of period t, the representative household inherits the entire money stock of the economy, M_t. The aggregate price level is denoted by P_t. In the standard CIA model, all decisions are made after, and therefore completely reflect, the current period surprise change in money growth and technology. The household determines how much money D_t to deposit at the bank. These deposits earn interest at the rate $R_{H,t}-1$. The bank receives household deposits and a monetary injection X_t from the central bank, which it lends to the firm at rate $R_{F,t}-1$.

The firm starts production and hires labour services from the household. After the firm produces its output, it uses the money borrowed from the financial intermediary to pay wages W_tH_t, where W_t is the nominal hourly wage, and H_t is hours worked. The household's cash balance increases to $M_t - D_t + W_tH_t$. The cash-in-advance constraint implies that all consumption purchases must be paid for with the accumulated cash balance. The firm's net cash inflow is paid as dividend F_t to the household. Moreover, the household receives back its bank deposits inclusive of interest and the net cash inflow of the bank as dividend B_t.

In period t, the household chooses consumption C_t, hours worked H_t, and non-negative deposits D_t to maximize the sum of discounted expected future utility. It solves the problem

$$\begin{aligned} \max_{\{C_t,H_t,M_{t+1},D_t\}} \quad & \mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^t[(1-\phi)\ln C_t + \phi\ln(1-H_t)]\right] \\ \text{s.t.} \quad & P_tC_t \le M_t - D_t + W_tH_t \\ & 0 \le D_t \\ & M_{t+1} = (M_t - D_t + W_tH_t - P_tC_t) + R_{H,t}D_t + F_t + B_t \end{aligned} \tag{4}$$

The firm chooses next period's capital stock K_{t+1}, labour demand N_t, dividends F_t and loans L_t. Since households value a unit of nominal dividends in terms of the consumption it enables during period $t+1$, and firms and the financial intermediary are owned by households, date t nominal dividends are discounted by date $t+1$ marginal utility of consumption. Thus, the firm solves the problem

$$\begin{aligned} \max_{\{F_t,K_{t+1},N_t,L_t\}} \quad & \mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^{t+1}\frac{F_t}{C_{t+1}P_{t+1}}\right] \\ \text{s.t.} \quad & F_t \le L_t + P_t\Big[K_t^{\alpha}(A_tN_t)^{1-\alpha} - K_{t+1} + (1-\delta)K_t\Big] - W_tN_t - L_tR_{F,t} \\ & W_tN_t \le L_t \end{aligned} \tag{5}$$

The financial intermediary solves the trivial problem

$$\begin{aligned} \max_{\{B_t,L_t,D_t\}} \quad & \mathbb{E}_0\left[\sum_{t=1}^{\infty}\beta^{t+1}\frac{B_t}{C_{t+1}P_{t+1}}\right] \\ \text{s.t.} \quad & B_t = D_t + R_{F,t}L_t - R_{H,t}D_t - L_t + X_t \\ & L_t \le X_t + D_t \end{aligned} \tag{6}$$

where $X_t = M_{t+1} - M_t$ is the monetary injection. The market-clearing conditions for labour market, money market, and goods market are $H_t = N_t$, $P_tC_t = M_t + X_t$, and $C_t + (K_{t+1} - (1-\delta)K_t) = K_t^{\alpha}(A_tH_t)^{1-\alpha}$. In equilibrium $R_{F,t} = R_{H,t}$.

The portfolio adjustment cost model differs in two respects from the standard CIA model. It adopts the information structure proposed by Fuerst (1992). We assume that the household makes its deposit decision D_t before observing the monetary growth shock and the technology shock. Thus in period t, after observing $\epsilon_{A,t}$ and $\epsilon_{M,t}$, the household chooses consumption C_t, the labour supply H_t, and next periods deposits D_{t+1}. Since the household cannot revise its deposit decision after a surprise change in the money growth rates, the additional cash has to be absorbed by the firm, which forces the nominal interest rate to fall. Let the household's cash holdings be denoted by $Q_t = M_t - D_t$. To make this liquidity effect persistent, Christiano and Eichenbaum (1992b) introduced a portfolio management cost, which is given by

$$\tilde{p}_t = \alpha_1 \left[\exp\left\{ \alpha_2 \left[\frac{Q_t}{Q_{t-1}} - m^* \right] \right\} + \exp\left\{ -\alpha_2 \left[\frac{Q_t}{Q_{t-1}} - m^* \right] \right\} - 2 \right] \quad (7)$$

where $Q_t = M_t - D_t$. The portfolio management cost reduces the time available for leisure. The household's problem then becomes

$$\begin{aligned} \max_{\{C_t, H_t, M_{t+1}, Q_{t+1}\}} \quad & \mathbb{E}_0 \left[\sum_{t=1}^{\infty} \beta^t [(1-\phi) \ln C_t + \phi \ln(1 - H_t - \tilde{p}_t)] \right] \\ \text{s.t.} \quad & P_t C_t \leq Q_t + W_t H_t \\ & Q_t \leq M_t \\ & M_{t+1} = (Q_t + W_t H_t - P_t C_t) + R_{H,t}(M_t - Q_t) + F_t + B_t \end{aligned} \quad (8)$$

Apart from the timing of the deposit decision and the portfolio adjustment cost, the model has the same structure as the standard cash-in-advance model.

To solve the models, optimality conditions are derived for the maximization problems. The real variables are then detrended by the productivity A_t, the price level by M_t/A_t, and X_t, Q_t, and D_t are detrended by M_t. It can be shown that the system in the detrended variables has a deterministic steady state and can be log-linearized around it. A solution to this linear rational expectation system can be obtained by elimination of unstable roots according to the algorithm in Sims (1995, unpublished manuscript).

Let y_t be an $n \times 1$ vector of observable variables. The log-linearized DSGE models yield state-space representations for y_t

$$y_t = \Xi_0 + \Xi_1 s_t + \Xi_* \epsilon_t \quad (9)$$

$$s_t = \Psi_1 s_{t-1} + \Psi_* \epsilon_t \qquad \epsilon_t \sim iid\mathcal{N}(0, \Sigma_\epsilon) \quad (10)$$

where $\epsilon_t = [\epsilon_{A,t}, \epsilon_{M,t}]'$, Σ_ϵ is the diagonal matrix with elements σ_A^2 and σ_M^2, and s_t is a vector of percentage deviations of detrended model variables from their steady state. The system matrices Ξ_0, Ξ_1, Ξ_*, Ψ_1, and Ψ_* are functions of the structural DSGE model parameters

$$\theta = [\alpha, \beta, \gamma, m^*, \rho, \phi, \delta, \sigma_A, \sigma_M, \alpha_1, \alpha_2]'$$

The DSGE models generate a joint probability distribution for the data $Y_T = [y_1, \ldots, y_T]'$.

Traditionally, macroeconometric models were evaluated according to their ability to track and forecast economic time series, e.g. Fair (1994). However, many DSGE models have difficulties tracking time series and macroeconomists started assessing business cycle models based on their ability to reproduce co-movement and impulse response patterns observed in the data. Our loss function-based evaluation procedure incorporates the practice of moment comparisons but also retains an important feature of the traditional evaluation approach.

3. LOSS FUNCTION-BASED DSGE MODEL EVALUATION

The CIA model and the PAC model are denoted by $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. The parameter vector of model $\mathcal{M}_i$ is $\theta_{(i)} \in \Theta_{(i)}$, y_t is an $n \times 1$ vector of observables and $Y_T = [y_1, \ldots, y_T]'$. We adopt a Bayesian approach and place probability on models and their parameters. Let $p(Y_T|\theta_{(i)}, \mathcal{M}_i)$ be the likelihood function for model i, and $p(\theta_{(i)}|\mathcal{M}_i)$ the prior for its parameters. In our application we will examine the implications of the DSGE models with respect to impulse response dynamics and correlations between output growth and inflation. These population characteristics are collected in generic $m \times 1$ vectors φ.

3.1 Accounting for DSGE Model Misspecification

An important aspect of the model evaluation is the potential misspecification of the DSGE models. Since the vector of model parameters θ is generally of low dimension, strong restrictions are imposed on the short- and long-run dynamics of y_t. These restrictions can cause poor statistical fit and forecasting performance.

If the number of structural shocks is less than the dimension n of the vector y_t, then the probability distribution of the observables y_t is degenerate. In this case the DSGE model implies that some linear combinations of y_t are perfectly correlated. In practice these singularities are rejected with a short span of time series data. While this paper focuses on the non-degenerate case, we will argue below that the proposed evaluation approach can also be applied in the degenerate case.

To cope with the potential misspecification of the two DSGE models, we consider a reference model $\mathcal{M}_0$ in addition to $\mathcal{M}_1$ and $\mathcal{M}_2$. The parameter vector of $\mathcal{M}_0$ is denoted by $\theta_{(0)}$, the likelihood function is $p(Y_T|\theta_{(0)}, \mathcal{M}_0)$, and the prior density is $p(\theta_{(0)}|\mathcal{M}_0)$. The reference model has to satisfy two requirements. First, it has to be more densely parameterized than the DSGE models to avoid dynamic misspecifications. Second, the population characteristics φ have to be identified within the context of the reference model, that is, conditional on $\theta_{(0)}$ it must be possible to compute a unique value of φ.

In the application we will use a vector autoregression as a reference model. The VAR is much less restrictive in its dynamic specification than the DSGE models, provided enough lags are included. If φ consists only of unconditional moments of y_t, such as autocorrelations and autocovariances, it can be identified from the VAR estimates without further assumptions. However, the effects of unanticipated changes in the money growth rate measured by the response functions to a money supply shock are not identified in a reduced-form VAR. Therefore, we have to impose restrictions on the VAR that guarantee the identification of impulse response functions. The identification scheme for the reference model will be discussed in Section 4.

The prior probabilities of the three models are $\pi_{0,0}$, $\pi_{1,0}$, and $\pi_{2,0}$, respectively. It is assumed that the mixture of the three models provides a for practical purposes acceptable probabilistic representation of the data upon which an evaluation of the structural models can be based. Two additional assumptions will be made. First, if the structural models are nested within the reference model, the prior $p(\theta_{(0)}|\mathcal{M}_0)$ assigns zero probability to the subset of the parameter space that corresponds to the structural models. Second, to simplify exposition and implementation of the framework it is assumed that the prior distributions of $\theta_{(0)}$, $\theta_{(1)}$, and $\theta_{(2)}$ are independent of each other.

3.2 The Evaluation Procedure

The model evaluation proceeds in four steps:

Step 1
Compute posterior distributions $p(\theta_{(i)}|Y_T,\mathcal{M}_i)$ for model parameters $\theta_{(i)}$ and calculate posterior model probabilities

$$\pi_{i,T} = \frac{\pi_{i,0}p(Y_T|\mathcal{M}_i)}{\sum_{i=0}^{2}\pi_{i,0}p(Y_T|\mathcal{M}_i)} \tag{11}$$

where $p(Y_T|\mathcal{M}_i)$ is the marginal data density

$$p(Y_T|\mathcal{M}_i) = \int p(Y_T|\theta_{(i)},\mathcal{M}_i)p(\theta_{(i)}|\mathcal{M}_i)\mathrm{d}\theta_{(i)} \tag{12}$$

under model $\mathcal{M}_i$. If the distribution of the data Y_T is degenerate under model $\mathcal{M}_i$, as is the case if the dimension of y_t exceeds the number of structural shocks, then the posterior probability of $\mathcal{M}_i$ is equal to zero.

Step 2
The population characteristics φ are a function $f_i(\theta_{(i)})$ of the model parameters $\theta_{(i)}$. Based on the posterior distribution of $\theta_{(i)}$, one can obtain a posterior for φ conditional on model $\mathcal{M}_i$. This posterior will be denoted by $p(\varphi|Y_T,\mathcal{M}_i)$. Since there are three models, including the reference model, the overall posterior of φ is given by the mixture

$$p(\varphi|Y_T) = \sum_{i=0}^{2}\pi_{i,T}p(\varphi|Y_T,\mathcal{M}_i) \tag{13}$$

The posterior probabilities $\pi_{i,T}$ determine the weights of the densities $p(\varphi|Y_T,\mathcal{M}_i)$. If both DSGE models are degenerate, then $\pi_{1,T}=0$, $\pi_{2,T}=0$ and $p(\varphi|Y_T)=p(\varphi|Y_T,\mathcal{M}_0)$.

Step 3
Loss functions are introduced to assess the ability of the DSGE models to replicate patterns of co-movements among key macroeconomic variables and impulse responses to structural shocks. The loss functions $L(\varphi,\hat{\varphi})$ penalize deviations of DSGE model predictions $\hat{\varphi}$ from population characteristics φ.

The prediction from DSGE model $\mathcal{M}_i$ is obtained as follows. Suppose a decision maker bases decisions exclusively on DSGE model $\mathcal{M}_i$. The optimal predictor[1] is

$$\hat{\varphi}_i = \text{argmin}_{\tilde{\varphi} \in R^m} \int L(\varphi, \tilde{\varphi}) p(\varphi | Y_T, \mathcal{M}_i) \mathrm{d}\varphi \tag{14}$$

The two DSGE models are judged according to the expected loss (risk) of $\hat{\varphi}_i$ under the overall posterior distribution $p(\varphi | Y_T)$.

$$R(\hat{\varphi}_i | Y_T) = \int L(\varphi, \hat{\varphi}_i) p(\varphi | Y_T) \mathrm{d}\varphi \tag{15}$$

The posterior risk $R(\hat{\varphi}_i | Y_T)$ provides an absolute measure of how well model $\mathcal{M}_i$ predicts the population characteristics φ. Risk differences across DSGE models yield a relative measure of adequacy that allows model comparisons. For instance, one can select the DSGE model $\mathcal{M}_i$ that minimizes $R(\hat{\varphi}_i | Y_T)$.

Step 4

Since φ is a function $f_i(\cdot)$ of the parameters $\theta_{(i)}$, one can obtain loss function parameter estimates $\hat{\theta}_{l,i}$ by solving the minimization problem

$$\hat{\theta}_{l,i} = \text{argmin}_{\theta_{(i)} \in \Theta_{(i)}} R\big(f_i(\theta_{(i)}) | Y_T\big) \tag{16}$$

These estimates provide a lower bound for the posterior risk attainable through a particular DSGE model. In other words, one can find the structural parameter estimates that achieve the best fit of Model $\mathcal{M}_i$ in a particular dimension.

3.3 Loss Functions

While the choice of loss function $L(\varphi, \hat{\varphi})$ can be tailored to the problem at hand, we propose three specific loss functions that will be used in our application. The first loss function is quadratic in φ:

$$L_q(\varphi, \hat{\varphi}) = (\varphi - \hat{\varphi})' W (\varphi - \hat{\varphi}) \tag{17}$$

where W is a positive definite $m \times m$ weight matrix. Let $I\!E[\varphi | Y_T]$ denote the expectation of the population characteristics with respect to the overall posterior $p(\varphi | Y_T)$. Since

$$\begin{aligned} I\!E\big[(\varphi - \hat{\varphi})' W (\varphi - \hat{\varphi}) | Y_T\big] = {} & I\!E\big[(\varphi - I\!E[\varphi | Y_T])' W (\varphi - I\!E[\varphi_T | Y_T]) | Y_T\big] \\ & + (\hat{\varphi} - I\!E[\varphi | Y_T])' W (\hat{\varphi} - I\!E[\varphi | Y_T]) \end{aligned} \tag{18}$$

the ranking of DSGE model predictions $\hat{\varphi}$ depends only on the weighted distance

$$\tilde{R}_q(\hat{\varphi} | Y_T) = (\hat{\varphi} - I\!E[\varphi | Y_T])' W (\hat{\varphi} - I\!E[\varphi | Y_T]) \tag{19}$$

between $\hat{\varphi}$ and $I\!E[\varphi | Y_T]$ but not on higher-order moments of the posterior distribution. The drawback of the quadratic loss function is that it requires the (subjective) specification of a weight matrix W.

[1] If the data density is degenerate under model $\mathcal{M}_i$, then the predictor $\hat{\varphi}_i$ can be calculated based on the distribution of $\varphi = f_i(\theta_{(i)})$ induced by the prior $p(\theta_{(i)} | \mathcal{M}_i)$.

The second loss function, $L_p(\hat{\varphi},\varphi)$, penalizes point predictions that lie in regions of low posterior density. Let $\mathcal{I}\{x > x_0\}$ denote the indicator function that is equal to one if $x > x_0$ is true and else equal to zero:

$$L_p(\varphi, \hat{\varphi}) = \mathcal{I}\{p(\varphi|Y_T) > p(\hat{\varphi}|Y_T)\} \tag{20}$$

If the posterior density $p(\varphi| Y_T)$ is unimodal, the expected L_p-loss provides a measure of how far the model prediction $\hat{\varphi}$ lies in the tail of the posterior distribution.

At last, we propose the L_{χ^2}-loss. Let V_φ denote the posterior covariance of φ under $p(\varphi| Y_T)$. Provided that V_φ is finite and non-singular, define the weighted discrepancy between φ and the posterior mean $I\!E[\varphi| Y_T]$:

$$C_{\chi^2}(\varphi|Y_T) = (\varphi - I\!E[\varphi|Y_T])' V_\varphi^{-1}(\varphi - I\!E[\varphi|Y_T]) \tag{21}$$

The third loss function is of the form

$$L_{\chi^2}(\varphi, \hat{\varphi}) = \mathcal{I}\{C_{\chi^2}(\varphi|Y_T) < C_{\chi^2}(\hat{\varphi}|Y_T)\} \tag{22}$$

If the posterior distribution of φ is Gaussian, L_{χ^2} and L_p-loss are identical. In general, under the L_p-loss function the DSGE models are compared based on the height of the posterior density at $\hat{\varphi}_i$. Under the L_{χ^2}-loss function the comparison is based on the weighted distance between $\hat{\varphi}_i$ and the posterior mean $I\!E[\varphi| Y_T]$.

The computation of the predictor $\hat{\varphi}_i$ requires solving the minimization problem (14). Under the quadratic loss function it is straightforward to see that $\hat{\varphi}_i$ is the posterior mean $\hat{\varphi}_{q,i} = I\!E[\varphi| Y_T,\mathcal{M}_i]$ of φ under model $\mathcal{M}_i$. The L_{χ^2} and L_p-loss functions depend on the contours of the posterior. Since we require that the calculation of the predictor is solely based on the information contained in $p(\varphi| Y_T,\mathcal{M}_i)$ (see Step 3), we let the loss function used in the minimization (14) be based on the contours of $p(\varphi| Y_T,\mathcal{M}_i)$ rather than $p(\varphi| Y_T)$. Hence, the expected L_p-loss is minimized by the posterior mode of $p(\varphi| Y_T,\mathcal{M}_i)$, $\hat{\varphi}_{p,i}$, and the L_{χ^2}-loss by the posterior mean $\hat{\varphi}_{q,i}$.

3.4 Discussion

In Bayesian analyses, the statistic that is most frequently employed for the comparison of two models $\mathcal{M}_1$ and $\mathcal{M}_2$ is the posterior odds ratio

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \frac{\pi_{1,0}p(Y_T, \mathcal{M}_1)/p(Y_T)}{\pi_{2,0}p(Y_T, \mathcal{M}_2)/p(Y_T)} \tag{23}$$

If $\mathcal{M}_1$ and $\mathcal{M}_2$ are the only models under consideration then $p(Y_T) = \sum_{i=1}^{2} \pi_{i,0}\, p(Y_T|\mathcal{M}_i)$. If the loss is one for choosing the incorrect model and zero for choosing the correct model $(0-1$ loss), then the posterior expected loss is minimized by selecting model $\mathcal{M}_1$ if $\pi_{1,T}/\pi_{2,T} > 1$.

Suppose a third model $\mathcal{M}_0$ is added to the model space, but the goal of the analysis remains to choose between models $\mathcal{M}_1$ and $\mathcal{M}_2$. If the prior model probabilities are adjusted to accomodate the presence of $\mathcal{M}_0$, the posterior odds ratio of $\mathcal{M}_1$ versus $\mathcal{M}_2$ stays the same, provided that $\mathcal{M}_0$ does not change the prior odds ratio. The consideration of model $\mathcal{M}_0$ only changes the marginal data density to $p(Y_T) = \sum_{i=0}^{2} \pi_{i,0}\, p(Y_T\, |\mathcal{M}_i)$. However, the $p(Y_T)$ terms in equation (23) cancel.

Under the $0-1$ loss it remains optimal to choose $\mathcal{M}_1$ instead of $\mathcal{M}_2$ whenever the posterior odds ratio favour model $\mathcal{M}_1$. However, if the DSGE models fit poorly and $\mathcal{M}_0$ has a high posterior probability, the $0-1$ loss function is not appealing. The function does not capture the economic implications of selecting $\mathcal{M}_1$ or $\mathcal{M}_2$ if the reference model is 'correct'. Since

$$\ln p(Y_T|\mathcal{M}_i) = \sum_{t=1}^{T} \ln p(y_t|Y_{t-1}, \mathcal{M}_i) \tag{24}$$

$\ln p(Y_T|\mathcal{M}_i)$ can be interpreted as predictive score (Good, 1951). The calculation of posterior odds only captures the relative one-step-ahead forecasting performance of the DSGE models, which is not the primary interest of many DSGE model evaluations. If both DSGE models create a degenerate probability distribution for y_t, then $p(Y_T|\theta_{(i)},\mathcal{M}_i) = 0$, $i = 1,2$, and the posterior odds of $\mathcal{M}_1$ versus $\mathcal{M}_2$ are not even defined.

Geweke (1999a) interprets DSGE models as structural specifications for population characteristics φ, rather than as models for the data Y_T or functions of Y_T. Under the assumption that $\mathcal{M}_1$ and $\mathcal{M}_2$ claim only to describe φ, he shows that

$$\frac{\pi_{1,0} \int p(\varphi|\mathcal{M}_1)p(\varphi|Y_T, \mathcal{M}_0)\mathrm{d}\varphi}{\pi_{2,0} \int p(\varphi|\mathcal{M}_2)p(\varphi|Y_T, \mathcal{M}_0)\mathrm{d}\varphi} \tag{25}$$

can be interpreted as odds ratio of $\mathcal{M}_1$ versus $\mathcal{M}_2$ conditional on the reference model $\mathcal{M}_0$. The numerator of expression (25) is large, if there is a strong overlap between the predictive densities for the population moments under DSGE model $\mathcal{M}_1$ and reference model $\mathcal{M}_0$.

In this paper we adopt the traditional interpretation of macroeconometric models and regard DSGE models as probabilistic representations of the data Y_T, not just of the population moments φ. The principle of our evaluation approach is to replace the above-mentioned $0-1$ loss function by a loss function with an economic interpretation and to illustrate the sensitivity of the model evaluation with respect to the choice of loss function. For instance, should the CIA or the PAC model be selected to predict the effect of an unanticipated monetary policy change that raises the price level by 1%?

To highlight important features of the proposed evaluation procedures we will discuss two special cases: (1) the posterior probability of the DSGE model $\mathcal{M}_1$ converges to one, and (2) the posterior probability of the reference model $\mathcal{M}_0$ converges to one as the sample size increases. In both cases we will assume that the evaluation is conducted under a quadratic loss function with some positive definite weight matrix W. We will use the fact that the ranking of DSGE models under L_q-loss only depends on $\tilde{R}_q(\hat{\varphi}_i|Y_T)$ defined in equation (19).

Case (1)

Suppose a sequence of observations $\{y_t\}_{t=1}^{T}$ is generated from model $\mathcal{M}_1$ conditional on a parameter vector $\theta_{(1),0}$ and the posterior model probability of $\mathcal{M}_1$ converges to one: $\pi_{1,T} \xrightarrow{p} 1$. Here, '$\xrightarrow{p}$' denotes convergence in probability as T tends to infinity under the distribution of $\{y_t\}_{t=1}^{T}$ induced by the data-generating model. We make the following additional assumptions. For each model $\mathcal{M}_i$, the posterior mean of the population characteristics converges to a finite limit, that is, $\mathbb{E}[\varphi|Y_T,\mathcal{M}_i] \xrightarrow{p} \underline{\varphi}_{(i)}$. $\mathcal{M}_1$ and $\mathcal{M}_2$ are asymptotically distinguishable based on the population characteristics φ in the sense that $\underline{\varphi}_{(1)} - \underline{\varphi}_{(2)} = \delta_\varphi$ and $|\delta_\varphi| > 0$.

From these assumptions it can be deduced that $\tilde{R}_q(\hat{\varphi}_{q,1}) \xrightarrow{p} 0$ and $\tilde{R}_q(\hat{\varphi}_{q,2}) \xrightarrow{p} \delta'_\varphi W\delta_\varphi$. For any positive definite weight matrix W, the probability that $\mathcal{M}_1$ is ranked above $\mathcal{M}_2$ converges to one as the sample size T tends to infinity. Thus, under regularity conditions that are sufficient for the above assumptions to hold, the proposed evaluation procedure is consistent. Regardless of W, in a large sample our procedure favours the model that would be selected based on a comparison of one-step-ahead forecast performances $\sum_{t=1}^{T} \ln p(y_t | Y_{t-1}, \mathcal{M}_i)$. Hence, the procedure retains an important feature of the traditional approach to the evaluation of macroeconometric models.

Case (2)
Suppose a sequence of observations $\{y_t\}_{t=1}^{T}$ is generated from the reference model $\mathcal{M}_0$ conditional on a parameter vector $\theta_{(0),0}$ and the posterior model probability of the reference model $\mathcal{M}_0$ converges to one: $\pi_{0,T} \xrightarrow{p} 1$. Moreover, assume $\mathbb{E}[\varphi | Y_T, \mathcal{M}_i] \xrightarrow{p} \underline{\varphi}_{(i)}$. Under these assumptions $\mathbb{E}[\varphi | Y_T] - \mathbb{E}[\varphi | Y_T, \mathrm{M}_0] \xrightarrow{p} 0$ and the ranking of DSGE models depends in large samples on the discrepancy between $\mathbb{E}[\varphi | Y_T, \mathcal{M}_0]$ and the predictors $\hat{\varphi}_{q,i}$, $i = 1,2$.

In the real business cycle literature moment predictions of DSGE models are often informally compared to sample estimates obtained from time series data. Econometricians have criticized this practice because no attention is paid to the fact that these sample moments are only noisy measures of population moments, and standard errors associated with the sample estimates are not systematically taken into account.

Consider a Gaussian VAR(1) with an $n_z \times 1$ deterministic trend component z_t as reference model:

$$y_t = \Phi_z z_t + \Phi_1 y_{t-1} + u_t \qquad u_t \sim \mathcal{N}(0, \Sigma_u) \tag{26}$$

Define the $T \times n$ matrices Y and Y_{-1} with rows y_t' and y_{t-1}', $t = 1, \ldots, T$. Let Z be the $T \times n_z$ matrix with rows z_t'. Suppose that y_t is trend stationary and φ consists only of autocovariances. Based on the maximum likelihood estimates $\hat{\Phi}_{1,mle}$ and $\hat{\Sigma}_{u,mle}$ of the VAR parameters, one can calculate the implied autocovariances for the detrended y_t, denoted by $\Gamma_{yy,h}$:

$$vec(\hat{\Gamma}_{yy,0}) = \left[I_{n^2 \times n^2} - \hat{\Phi}_{1,mle} \otimes \hat{\Phi}_{1,mle}\right]^{-1} vec(\hat{\Sigma}_{u,mle}) \tag{27}$$

$$\hat{\Gamma}_{yy,h} = \hat{\Phi}_{1,mle}^{h} \hat{\Gamma}_{yy,0} \tag{28}$$

where $vec(A)$ stacks the columns of matrix A, $I_{n^2 \times n^2}$ denotes the $n^2 \times n^2$ identity matrix, and $\otimes$ is the Kronecker product. For the autocovariances of order $h = 0,1$ we obtain

$$\hat{\Gamma}_{yy,0} = \frac{1}{T} Y' M_z Y + \text{small} \qquad \hat{\Gamma}_{yy,1} = \frac{1}{T} Y' M_z Y_{-1} + \text{small} \tag{29}$$

where $M_z = I_{T \times T} - Z(Z'Z)^{-1} Z'$. Thus, for $h = 0,1$ the VAR-based autocovariance estimates are approximately equivalent to the sample autocovariances. The *small*-term arises because $Y' M_z Y$ is not exactly equal to $Y_{-1}' M_z Y_{-1}$. To the extent that the posterior mean $\mathbb{E}[\varphi | Y_T, \mathcal{M}_0]$ of the autocovariances can be well approximated by the maximum likelihood plug-in values given in equation (29),[2] the evaluation under the quadratic loss function resembles the widely applied comparison of DSGE model predictions and sample moments.

[2] Consider the posterior expected value of a function $f(\theta)$:

$$\int f(\theta) p(\theta | Y_T) \mathrm{d}\theta = f(\hat{\theta}_{mle}) + \int \left[f(\theta) - f(\hat{\theta}_{mle})\right] p(\theta | Y_T) \mathrm{d}\theta \tag{30}$$

In large samples the posterior distribution generally concentrates around $\hat{\theta}_{mle}$ and the second term vanishes as $T \to \infty$. For a formal argument and regularity conditions see, for instance, Crowder (1988).

It is a particular feature of the quadratic loss function that for any given W the posterior variance of φ does not affect the ranking of the DSGE models. Informal Kydland and Prescott (1996)-style moment comparisons usually incorporate judgments about the relative importance of individual moments, i.e. the elements of the vector φ. In our approach these judgements are reflected in the weight matrix W.[3]

Model evaluation under the L_p and L_{χ^2}-loss function, on the other hand, is sensitive to the posterior uncertainty with respect to φ. We will discuss the relationship between traditional p-values and the posterior expected L_{χ^2} and L_p-loss in the context of the empirical application.

4. EMPIRICAL ANALYSIS

While the cash-in-advance and the portfolio adjustment cost models presented in Section 2 generate stochastic representations for several observable time series, our empirical analysis focuses on two key macroeconomic variables, stacked in the vector y_t. We will examine to what extent the DSGE models fit quarterly US output growth $\Delta \ln GDP_t$ and inflation data $\Delta \ln P_t$ from 1950:I to 1997:IV. The data were extracted from the DRI database. Aggregate output is real gross domestic product (*GDPQ*), converted into per capita terms by the NIPA population series (*GPOP*). The GDP deflator series (*GD*) is used as aggregate price level. Logarithms and first differences are taken to obtain quarterly output growth and inflation.

In this application the dimension of y_t is 2×1 and matches the number of structural shocks in the DSGE model. Therefore, the likelihood functions $p(Y_T|\theta_{(i)},\mathcal{M}_i)$ are non-degenerate. We will compute posterior distributions for the parameters $\theta_{(i)}$ as well as posterior model probabilities. As pointed out in the previous section, the proposed evaluation procedure can also be applied if the dimension of y_t exceeds the number of structural shocks. However, if the intention is to fit a model to several macroeconomic variables simultaneously, we find it desirable to include enough shocks such that $p(Y_T|\theta_{(i)})$ is non-degenerate. Examples of this modeling approach are Leeper and Sims (1994), Kim (2000), and Dejong *et al.* (2000).[4]

4.1 The Reference Model

A VAR serves as reference model $\mathcal{M}_0$:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_* \epsilon_t \qquad \epsilon_t \sim iid\mathcal{N}(0, \Sigma_\epsilon) \tag{31}$$

where ϵ_t is interpreted as vector of structural shocks $[\epsilon_{A,t}, \epsilon_{M,t}]'$. As in the two DSGE models, the fluctuations of output growth and inflation are due to technology shocks and unanticipated changes in the growth rate of money supply. Without an additional restriction, the four elements of the matrix Φ_* are not identifiable from the data. Both the cash-in-advance and the portfolio

[3] In practice, model DSGE evaluations are often based on population moments of filtered series. If y_t corresponds the filtered data the argument remains valid. However, if y_t corresponds to the raw series and the filtering is incorporated into the definition of the population characteristic φ, then the discrepancy between posterior mean $\mathbb{E}[\varphi | Y_T,\mathcal{M}_0]$ and the sample autocovariances of the filtered series does generally not vanish.

[4] Various approaches have been employed to remove the counterfactual singularities from the DSGE models prior to the empirical analysis. For instance, Altug (1989) introduced measurement errors in the empirical specification and Ireland (1998) constructed a hybrid model. If our evaluation approach is applied to one of these modified DSGE models, it will generally lead to an assessment of the modified DSGE model rather than the original model.

adjustment cost model imply long-run neutrality of money growth shocks. This restriction is imposed on the VAR and the elements of Φ_* are identified by the Blanchard and Quah (1989) scheme.

Vector autoregressions are generally less restrictive in their dynamics than linearized DSGE models. The identification scheme, however, is application-specific and open to criticism. It should be consistent with the DSGE models that are being compared. For our analysis, we do find it plausible to identify money supply shocks through a long-run neutrality restriction and more attractive than identifying them as innovations to a money stock series. The same long-run identification scheme has been used by Nason and Cogley (1994).

The reduced-form disturbances are defined as $u_t = \Phi_* \epsilon_t$ with covariance matrix $\Sigma_u = \Phi_* \Sigma_\epsilon \Phi'_*$. Let Y_T now be the $(T-p) \times n$ matrix with rows y'_t, $t = p+1, \dots, T$ (the first p observations are used to initialize lags). Let $k = 1 + np$, X_T be the $(T-p) \times k$ matrix with rows $[1, y'_{t-1}, \dots, y'_{t-p}]$, U_T be the matrix with rows u'_t, and $B = [\Phi_0, \Phi_1, \dots, \Phi_p]'$. The reference model can be expressed as

$$Y_T = X_T B + U_T \tag{32}$$

For the calculation of posterior model probabilities it is important that the prior distributions of the parameters are proper. To obtain a proper prior for B and Σ a portion of the sample $t = 1, \dots, T_*$ is considered as training sample. Let Y_* and X_* be matrices with rows y'_t and x'_t, $t = p+1, \dots, T_*$. Define

$$\hat{B}_* = (X'_* X_*)^{-1} X'_* Y_*, \quad \hat{\Sigma}_* = (T_* - p)^{-1} (Y_* - X_* \hat{B}_*)'(Y_* - X_* \hat{B}_*) \tag{33}$$

Conditional on the training sample, we use conjugate proper priors for B and Σ:

$$\Sigma | Y_* \sim IW\big((T_* - p)\hat{\Sigma}_*, T_* - k - p\big) \tag{34}$$

$$vec(B) | \Sigma, Y_* \sim \mathcal{N}\big(vec(\hat{B}_*, \Sigma \otimes (X'_* X_*)^{-1}\big) \tag{35}$$

where IW denotes the Inverted Wishart distribution. The appropriate lag length p is determined from the data. The subsequent analysis is conducted conditional on a training sample that ranges from 1950:I to 1954:IV ($T_* = 20$). We use Y_T^* to denote the $(T - T_*) \times n$ matrix with rows y'_t, $t = T_* + 1, \dots, T$.

4.2 The Model Evaluation

The prior distribution for the DSGE model parameters is summarized in Table I. The shapes of the densities are chosen to match the domain of the structural parameters. As in Canova (1994) and Dejong *et al.* (1996, 2000), it is assumed that the structural parameters are *a priori* independent of each other. Thus, the joint prior density is simply the product of the marginal densities. Since all the marginal densities integrate to unity, it follows that the joint prior distribution is proper.

The prior was specified conditional on information from the years 1950:I to 1954:IV. Based on DRI series on Wages and Salaries, Other Labor Income, and National Income, we calculated an average capital share of 0.356. The standard error (SE) for α was chosen to be 0.02 which is about four times the SE of the sample mean of the calculated capital share. To specify a prior for γ we computed the average GDP growth rate during the training sample period. The priors for the steady-state money growth rate ln m_* and the autoregressive parameter ρ are obtained by

Table I. Prior distribution for the parameters of the DSGE models

Name	Range	Density	Mean	SE
α	[0,1]	Beta	0.3560	(0.020)
β	[0,1]	Beta	0.9930	(0.002)
γ	$\mathbb{R}$	Gaussian	0.0085	(0.003)
$\ln m_*$	$\mathbb{R}$	Gaussian	0.0002	(0.007)
ρ	[0,1]	Beta	0.1290	(0.223)
ϕ	[0,1]	Beta	0.6500	(0.050)
δ	[0,1]	Beta	0.0100	(0.005)
κ	$\mathbb{R}^+$	Gamma	50.000	(20.00)
σ_A	$\mathbb{R}^+$	InvGamma	0.020*	2.000*
σ_M	$\mathbb{R}^+$	InvGamma	0.005*	2.000*

Notes: SE denotes standard error. The parameter κ appears only in the portfolio adjustment cost model $\mathcal{M}_2$. For the Inverse Gamma (ν,s) priors we report the parameters ν and s. For $\nu = 2$ the standard error is infinite.

fitting an AR(1) model to the DRI money base series. The priors for β, ϕ, and δ are chosen to obtain plausible steady-state values for the capital–output ratio, the consumption–output ratio, and the leisure–labour ratio. With 95% prior probability, these ratios lie between 8 and 24, 0.67 and 0.82, and 1.3 and 3.2, respectively.

Since model $\mathcal{M}_2$ depends on the adjustment cost parameters α_1 and α_2 only through $\kappa = \alpha_1\alpha_2^2$, we re-parameterize $\mathcal{M}_2$ accordingly. Christiano and Eichenbaum (1992b) choose $\alpha_1 = 0.00005$ and $\alpha_2 = 1000$, which leads to $\kappa = 50$. In the steady state the adjustment costs are equal to zero. The parameter κ affects only the marginal costs (in terms of hours) of changing bank deposits. Christiano and Eichenbaum (1992b) calculated that after an unanticipated one-standard-deviation money growth shock their choice of parameters translates into a loss of 3 minutes in the third quarter after the shock. We centre our prior at $\kappa = 50$ and use a standard error of 20.

Step 1

Draws from the posterior distribution $p(\theta_{(i)} | Y_T, \mathcal{M}_i)$ of the DSGE models cannot be generated directly because the posterior does not belong to a well-known class of distributions. We can only numerically evaluate the product $p(Y_T^* | Y_*, \theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)} | \mathcal{M}_i)$ of likelihood and prior. Therefore, a random walk Metropolis algorithm, described in the Appendix, is used to generate parameter draws from the posterior distributions.

Posterior means and standard errors are calculated from the output of the Metropolis algorithm and summarized in Table II. For both models the posterior mean of the capital share parameter is about 0.42, somewhat larger than the value that is often used in calibration exercises. Conditional on model $\mathcal{M}_1$ the posterior mean of β is 0.99, which implies an annualized steady-state real interest rate of 4%. The posterior-mean estimate of β under $\mathcal{M}_2$ corresponds to a 6% real interest rate. The estimated technology growth rate γ is 0.37 and 0.42% per quarter, respectively. According to our estimates, the implied capital output ratio lies between 16 and 41, the consumption output ratio between 0.74 and 0.97, the leisure–labour ratio between 2 and 5, and the nominal interest rate between 7% and 11%. The magnitude of the estimated simulation standard errors for the posterior means and standard errors is about 1%.

Table II. Posterior means and standard errors for the parameters of the structural models

	Model $\mathcal{M}_1$		Model $\mathcal{M}_2$	
	Mean	SE	Mean	SE
α	0.4168	(0.0218)	0.4193	(0.0233)
β	0.9901	(0.0021)	0.9856	(0.0029)
γ	0.0038	(0.0010)	0.0041	(0.0013)
$\ln m_*$	0.0141	(0.0017)	0.0136	(0.0023)
ρ	0.8623	(0.0343)	0.8610	(0.0433)
ϕ	0.6837	(0.0479)	0.6558	(0.0534)
δ	0.0020	(0.0011)	0.0029	(0.0016)
κ			63.837	(23.211)
σ_A	0.0133	(0.0009)	0.0162	(0.0011)
σ_M	0.0029	(0.0002)	0.0033	(0.0002)

Notes: $\mathcal{M}_1$ is the standard cash-in-advance model and $\mathcal{M}_2$ is the portfolio adjustment cost model. The moments are calculated from the output of the Metropolis Algorithm. The adjustment cost parameter κ does not enter model $\mathcal{M}_1$. The estimated simulation standard errors (see Appendix) for the posterior moments are less than 1%.

The estimates are consistent with the findings of Christiano (1991). To match the data, the models need a high value of α to offset the depressive effect of inflation on the capital stock and on labour supply. The estimated autocorrelation of money growth is large at 0.86. Based on the monetary base series from 1954:I to 1991:IV, Nason and Cogley (1994) estimated the autocorrelation of money growth to be 0.728. Since prices in the model adjust quickly, a large ρ is needed to capture the persistence in inflation. The estimated adjustment cost parameter for $\mathcal{M}_2$ is 63.84, somewhat larger than the value suggested by Christiano and Eichenbaum (1992b).

Prior and posterior model probabilities are summarized in Table III. We assigned prior probabilities of one-third to the two DSGE models and the reference model. Since there is uncertainty with respect to the appropriate lag length p of the VAR, we use a mixture of four vector autoregressions (lags 1 to 4) as reference model.

Table III. Prior and posterior model probabilities

	$\mathcal{M}_1$	$\mathcal{M}_2$	VAR(1)	VAR(2)	VAR(3)	VAR(4)
Prior Prob π_{i,T_*}	1/3	1/3	1/12	1/12	1/12	1/12
$\ln p(Y_T^* \mid Y_*, \mathcal{M}_i)$	N/A	N/A	1250.73	1253.01	1259.12	1249.95
Laplace approx.	1195.01	1182.35	1250.59	1252.77	1258.77	1249.46
Harmonic mean	1195.15	1182.47	1250.73	1253.01	1259.10	1249.94
Post. odds $\pi_{i,T}/\pi_{1,T}$	1.00000	3.1E-6	3.4E23	3.4E24	1.5E27	1.6E23
Post. prob $\pi_{i,T}$	6.6E-28	2.1E-33	0.0002	0.0022	0.9975	0.0001
$RMSE(\Delta \ln GDP_t)$	0.0099	0.0098	0.0093	0.0093	0.0092	0.0091
$RMSE(\Delta \ln P_t)$	0.0039	0.0042	0.0032	0.0030	0.0029	0.0029

Notes: Laplace approx. is the Laplace approximation of $\ln p(Y_T|\mathcal{M}_i)$ given in equation (37), Harmonic mean refers to the simulation-based approximation of $\ln p(Y_T|\mathcal{M}_i)$. $\mathcal{M}_1$ is the standard cash-in-advance model and $\mathcal{M}_2$ is the portfolio adjustment cost model. For the VARs, the harmonic mean approximations are based on 8000 draws from the posterior parameter distribution. For the DSGE models it is based on 80,000 draws. In the latter case the estimated approximation error is about 0.5 (in log-units).

The posterior model probabilities are functions of the marginal data densities

$$p(Y_T^*|Y_*, \mathcal{M}_i) = \int p(Y_T^*|Y_*, \theta_{(i)}, \mathcal{M}_i)p(\theta_{(i)}|Y_*, \mathcal{M}_i)\mathrm{d}\theta_{(i)} \tag{36}$$

Since conjugate priors are used for the VARs, the marginal data density of the reference model can be calculated analytically. For the two DSGE models expression (36) does not have a closed-form solution. Two approximations are reported in Table III. The Laplace approximation is of the form

$$\tilde{p}(Y_T^*|Y_*, \mathcal{M}_i) = (2\pi)^{d_i/2}|\tilde{\Sigma}_{(i)}|^{1/2}p(Y_T^*|Y_*, \tilde{\theta}_{(i)}, \mathcal{M}_i)p(\tilde{\theta}_{(i)}|Y_*, \mathcal{M}_i) \tag{37}$$

where $\tilde{\theta}_{(i)}$ is the posterior-estimator of $\theta_{(i)}$, d_i the dimension of $\theta_{(i)}$, and $\tilde{\Sigma}_{(i)}$ is the inverse Hessian evaluated at the posterior mode:

$$\tilde{\Sigma}_i = \left[-\frac{\partial^2}{\partial\theta_i\partial\theta_i'} \ln p(Y_T^*|Y_*, \theta_i, \mathcal{M}_i)p(\theta_i|Y_*, \mathcal{M}_i) \right]^{-1}_{\theta_{(i)}=\tilde{\theta}_{(i)}} \tag{38}$$

The approximation error is small if ln $[p(Y|\theta_{(i)}, \mathcal{M}_i)p(\theta_{(i)}|\mathcal{M}_i]$ is well approximated by quadratic function of $\theta_{(i)}$. The term $|\tilde{\Sigma}_{(i)}|^{1/2}$ can be interpreted as penalty for the dimensionality of model $\mathcal{M}_i$. Second, we are reporting the posterior simulation-based modified harmonic mean estimator of the marginal data density proposed in Geweke (1999b) and described in the Appendix.

For all models, the two different approximations of the marginal data densities lead to the same substantive conclusions. The posterior odds of $\mathcal{M}_1$ versus $\mathcal{M}_2$ are roughly 1 to 300,000 and favour the standard CIA model. The in-sample root-mean-square-errors (RMSE) indicates that the CIA model tracks the inflation series better than the PAC model. However, compared to the vector autoregressions, the time series fit of the two DSGE models is not good.

The VAR posterior is Gaussian in terms of the autoregressive parameters, but only approximately quadratic with respect to the three covariance parameters. Therefore, the Laplace approximations are quite precise. The approximation errors of the harmonic mean estimator of the VAR marginal data densities are negligible. For the DSGE models the estimated numerical standard errors of the harmonic mean estimates are about 0.5 for the log-densities.

We conducted a sensitivity analysis of the posterior estimates to the choice of prior distributions. As the prior variance of α and β is increased, a better time series fit of the models was achieved. However, the posterior mean estimates of the implied steady-state real interest rates and the capital share became very large. The former was more than 10% annually and the latter above 70%. It is well documented, e.g. by Cogley and Nason (1995) and Rotemberg and Woodford (1996), that standard business cycle models have difficulties in replicating the persistence in output growth. The random walk technology shock leads to small autocorrelated variations in the capital stock. To translate these variations into persistent movements of output growth, a very large capital share α is needed. We found that for priors that lead to very large estimates of α and $1/\beta$ the posterior odds favour model $\mathcal{M}_2$ rather than $\mathcal{M}_1$. Under these priors, the marginal data density of $\mathcal{M}_2$ can reach the magnitude of the VAR(4) data density.

Throughout the remainder of this section we only report numerical results for the prior summarized in Table I. The prior variance of α and β is consistent with training sample evidence and capital share values and steady state interest rates that economists regard as plausible.

Step 2
Since the combined posterior probability of the DSGE models is small, the $0-1$ loss function that assigns the same loss to choosing $\mathcal{M}_1$ or $\mathcal{M}_2$ if the reference model is 'correct' is not appealing. Therefore we use loss functions that are based on the prediction of population characteristics. The following characteristics are considered: (1) the correlations between output growth and inflation $corr(\Delta \ln GDP_t, \Delta \ln P_{t+h})$, $h = -2, \ldots, 2$; (2) the responses $d \ln GDP/d\epsilon_A$ and $d\,\Delta \ln P/d\epsilon_A$ to a permanent (technology) shock that increases output by 1% in the long run and the responses $d \ln GDP/d\epsilon_M$ and $d\Delta \ln P/d\epsilon_M$ to a transitory (money growth) shock that raises the price level by 1% in the long run. Since the probabilistic specification of the DSGE models is deficient, as indicated by their low posterior probability, we normalize the magnitude of the structural shocks by their long-run effect rather than by estimated standard deviations to compare impulse response dynamics. We describe in the Appendix how to generate draws from the density $p(\varphi \mid Y_T) = \sum_{i=0}^{2} \pi_{i,T}\, p(\varphi \mid Y_T, \mathcal{M}_i)$ and how to compute posterior expected losses. Since $\pi_{1,T}$ and $\pi_{2,T}$ are essentially zero and L_p and L_{χ^2}-loss functions are bounded by one, the contributions of the DSGE model to the posterior $p(\varphi \mid Y_T)$ is neglected in the calculation of expected values.

Step 3
This subsection summarizes posterior prediction losses for autocorrelations and impulse response functions. The results for $corr\,(\Delta \ln GDP_t, \Delta \ln P_{t+h})$ are presented in Table IV. The first two rows of Table IV contain 90% highest posterior density (HPD) confidence intervals. The posterior indicates that the correlation among output growth and inflation is between -0.37 and -0.05. According to the posterior mode predictions $\hat{\varphi}_{p,i}$ of the DSGE models (rows 3 and 4 of Table IV), the contemporaneous correlation between $\Delta \ln GDP_t$ and $\Delta \ln P_t$ is less than -0.50. While $\mathcal{M}_1$ predicts a slightly negative correlation between current output growth and future inflation, $\mathcal{M}_2$ implies essentially no correlation. Under $\mathcal{M}_1$ current output growth is nearly uncorrelated with past inflation. $\mathcal{M}_2$ suggests a small positive correlation between output growth and inflation lagged by one period.

A variance decomposition indicates that under both $\mathcal{M}_1$ and $\mathcal{M}_2$ about 99% of the fluctuations in output growth and about 33% of the variation in inflation are due to technology shocks. An increase in productivity causes output to rise and the price level has to fall. Since the

Table IV. Model evaluation statistics for correlations between output growth and inflation (lags $h = -2$ to $h = 2$ individually): 90% highest posterior density intervals, posterior mode predictions $\hat{\varphi}_{p,i}$, and posterior expected L_p-loss (risk) of $\hat{\varphi}_{p,i}$

	Correlations $corr\,(\Delta \ln GDP_t, \Delta \ln P_{t+h})$					
	$\hat{\varphi}_{p,i}$	$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
90% Interval (U)		0.0419	0.0543	−0.0502	0.0222	0.1015
90% Interval (L)		−0.3003	−0.2866	−0.3722	−0.3105	−0.2421
Mode prediction	$i = 1$	0.0005	0.0014	−0.5774	−0.0236	−0.0228
	$i = 2$	0.0042	0.0145	−0.5103	0.0098	0.0048
L_p-risk	$i = 1$	0.7977	0.7375	0.9950	0.7699	0.2942
	$i = 2$	0.8089	0.7786	0.9887	0.8645	0.4711

Note: $i = 1$ is the standard cash-in-advance model and $i = 2$ is the portfolio adjustment cost model.

Table V. Model evaluation statistics for correlations between output growth and inflation (lags $h = -2$ to $h = 2$ jointly)

	$\ln[\pi_{j,T}\tilde{p}(\hat{\varphi}_{q,i}\mid\mathcal{M}_j)]$			
	$j=0$	$j=1$	$j=2$	$C_{\chi^2}(\hat{\varphi}_{q,i}\mid Y_T)$
$i=1$	−78.061	−28.391	−78.201	185.86
$i=2$	−59.537	−41723.	−48.297	148.81

Note: $C_{\chi^2}(\hat{\varphi}_{q,i}\mid Y_T)$ is defined in equation (21). $i=1$ is the standard cash-in-advance model and $i=2$ is the portfolio adjustment cost model.

price level adjusts immediately, the DSGE models predict a negative contemporaneous correlation that is too large.

The informal assessment is confirmed by a formal analysis of L_p and L_{χ^2} risks. The last two columns of Table IV contain posterior expected L_p losses for the prediction of the five correlations individually. For $h=0$ less than 1% of the posterior mass is concentrated at values of φ that have lower density than $\hat{\varphi}_{p,1}$ and $\hat{\varphi}_{p,2}$, respectively.[5] At lags $h=1$ and $h=2$ the CIA model dominates the PAC model.

Our procedure also allows a joint evaluation of the $m=5$ moment predictions. Based on the results of the univariate calculations we expect the DSGE model predictions to lie far in the tails of the posterior distribution, where Kernel density estimates are likely to be inaccurate. We therefore approximate the posterior densities $p(\varphi\mid Y_T,\mathcal{M}_j)$ by Gaussian densities $\tilde{p}(\varphi\mid Y_T,\mathcal{M}_j)$ with mean $\mathbb{E}[\varphi\mid Y_T,\mathcal{M}_j]$ and variance $\text{Var}[\varphi\mid Y_T,\mathcal{M}_j]$. Moreover, the posterior mode predictions $\hat{\varphi}_{p,i}$ of the DSGE models are approximated by the posterior mean predictions $\hat{\varphi}_{q,i}$. The means and covariance matrices are computed from the output of the posterior simulators. Therefore, the height of the posterior density evaluated at the DSGE model prediction is

$$p(\hat{\varphi}_{p,i}\mid Y_T) \approx \sum_{j=0}^{2} \pi_{j,T}\tilde{p}(\tilde{\varphi}_{q,i}\mid Y_T,\mathcal{M}_j) \qquad i=1,2 \tag{39}$$

The natural log of $\pi_{j,T}\tilde{p}(\hat{\varphi}_{q,i}\mid Y_T,\mathcal{M}_j)$ is tabulated in the first three columns of Table V. All three components of $\tilde{p}(\hat{\varphi}_{q,i}\mid Y_T)$ are very small. The results show that for predictor i the height of the density $\tilde{p}(\hat{\varphi}_{q,i}\mid Y_T)$ is determined by the component $\pi_{i,T}\tilde{p}(\hat{\varphi}_{q,i}\mid Y_T,\mathcal{M}_i)$. The other two components are several magnitudes smaller. Thus, $\ln\tilde{p}(\hat{\varphi}_{q,1}\mid Y_T) = -28.39$ and $\ln\tilde{p}(\hat{\varphi}_{q,2}\mid Y_T) = -48.30$. These approximations suggest that the posterior expected L_p-prediction loss is smaller for model $\mathcal{M}_1$ than for model $\mathcal{M}_2$.

[5] Univariate kernel density estimates were used to estimate the posterior densities $p(\varphi\mid Y_T,\mathcal{M}_i)$ based on draws $\varphi^{(s)}$ from the posteriors of $\varphi\mid Y_T,\mathcal{M}_i$, $i=1,2$. To obtain the DSGE model predictions $\hat{\varphi}_{p,i}$ under L_p-loss, we calculate the mode of $\hat{p}(\varphi\mid Y_T,\mathcal{M}_i)$. A first sequence of draws φ is generated from the reference model to construct a density estimate of $p(\varphi\mid Y_T,\mathcal{M}_0)$. Then, a second sequence of draws $\varphi^{(s)}$ from the reference model is used to approximate the L_p-risk:

$$\frac{1}{n_{sim}}\sum_{s=1}^{n_{sim}}\left\{\hat{p}(\varphi^{(s)}|Y_T) < \hat{p}(\hat{\varphi}_{p,i}|Y_T\right\}$$

We used a Gaussian Kernel with bandwidth $\lambda = (4/3)^{(1/5)}\, n_{sim}^{-1/5}\,\sigma_\varphi$, where n_{sim} is the number of posterior draws, and σ_φ is the standard error of the correlation φ, estimated from the posterior simulator output.

The structural models introduce modes into the tails of the overall posterior $p(\varphi \mid Y_T)$. The height of these modes is increasing in the posterior probabilities $\pi_{i,T}$ and decreasing in the posterior variance of φ under model $\mathcal{M}_i$. Since both $\hat{\varphi}_{q,1}$ and $\hat{\varphi}_{q,2}$ are far in the tails of $p(\varphi \mid Y_T, \mathcal{M}_0)$ and $\pi_{1,T}$ is much greater than $\pi_{2,T}$ the CIA model ranks above the PAC model according to the L_p-risk. However, for both predictors the height of the posterior density is so small that the value of the L_p-risk is essentially equal to one.[6]

Now consider the L_{χ^2}-loss. The scaled distances $C_{\chi^2}(\hat{\varphi}_{q,i} \mid Y_T)$, defined in equation (21), are reported in the last column of Table V. Since $C_{\chi^2}(\hat{\varphi}_{q,1} \mid Y_T) > C_{\chi^2}(\hat{\varphi}_{q,2} \mid Y_T)$ the previous ranking of the two models is reversed. The expected L_{χ^2} loss for the portfolio adjustment cost model is lower than for the cash-in-advance model. Under the L_{χ^2}-loss function the comparison is based only on the V_φ^{-1} weighted distance between the predictor $\hat{\varphi}_{q,i}$ and the posterior mean $\mathbb{E}[\varphi \mid Y_T]$. Since $\pi_{1,T}$ and $\pi_{2,T}$ are very small, $\mathbb{E}[\varphi \mid Y_T]$ is practically determined by the reference model. The fact that $\pi_{1,T} > \pi_{2,T}$ does not shift the prediction losses in favour of the cash-in-advance model. Under the assumption that the posterior distribution of φ is approximately normal, the magnitude of the L_{χ^2}-risk can be approximated by $F_{\chi^2_m}[C_{\chi^2}(\hat{\varphi}_{q,i} \mid Y_T)]$, where $F_{\chi^2_m}$ is the cumulative density function (cdf) of a χ^2 random variable with m degrees of freedom. For both predictors the risk is essentially equal to one.

In our application the L_p and the L_{χ^2}-loss functions lead to different rankings of the models. However, since the posterior losses are practically equal to one we conclude that the joint predictions of the correlations are inconsistent with the data.

Figure 1 depicts impulse responses to the structural shocks $\epsilon_{M,t}$ and $\epsilon_{A,t}$. The posterior distribution indicates a positive hump-shaped response of output to a temporary unanticipated increase of the money growth rate. The portfolio adjustment cost model $\mathcal{M}_2$ is able to generate a positive response of output, which is initially smaller and slightly more persistent in the long run than the posterior mean response.

The money growth shock increases the amount of loans that the firm has to absorb since the household's bank deposits are predetermined. To increase the demand for loans the nominal interest rate has to fall which stimulates economic activity. The household can adjust its labour supply in response to the money growth shock. The labour–leisure trade-off determines aggregate real output in the initial period since the capital stock is also predetermined. For the posterior mean values of the structural parameters the incentive for the household to increase its labour supply is not strong enough to generate such a large increase in output as suggested by the overall posterior distribution. The hump-shaped response of output in the PAC model is in part due to the high estimated autocorrelation of the money growth rates. For low values of ρ output rises after impact of the unanticipated money shock and then decays monotonically.

In the standard cash-in-advance model $\mathcal{M}_1$ households can adjust their deposits contemporaneously in response to the money growth shock. The nominal interest rate is approximately equal to the real interest rate plus expected inflation. The money growth shock increases the expected inflation rate. Thus, the nominal interest rate rises and output slightly decreases. The model does not generate a liquidity effect. Both DSGE models predict a sudden increase in the price level which results in too much inflation in the first period.

[6] If the posterior distribution $p(\varphi \mid Y_T, \mathcal{M}_i)$, $i = 1,2$, concentrates in a lower dimensional subspace of $\mathbb{R}^m$, model $\mathcal{M}_i$ introduces a point mass in the overall posterior distribution of φ. To conduct a model evaluation under the L_p-loss, the overall posterior density $p(\varphi \mid Y_T)$ has to be carefully redefined with respect to a dominating measure that assigns non-zero mass to this sub-space.

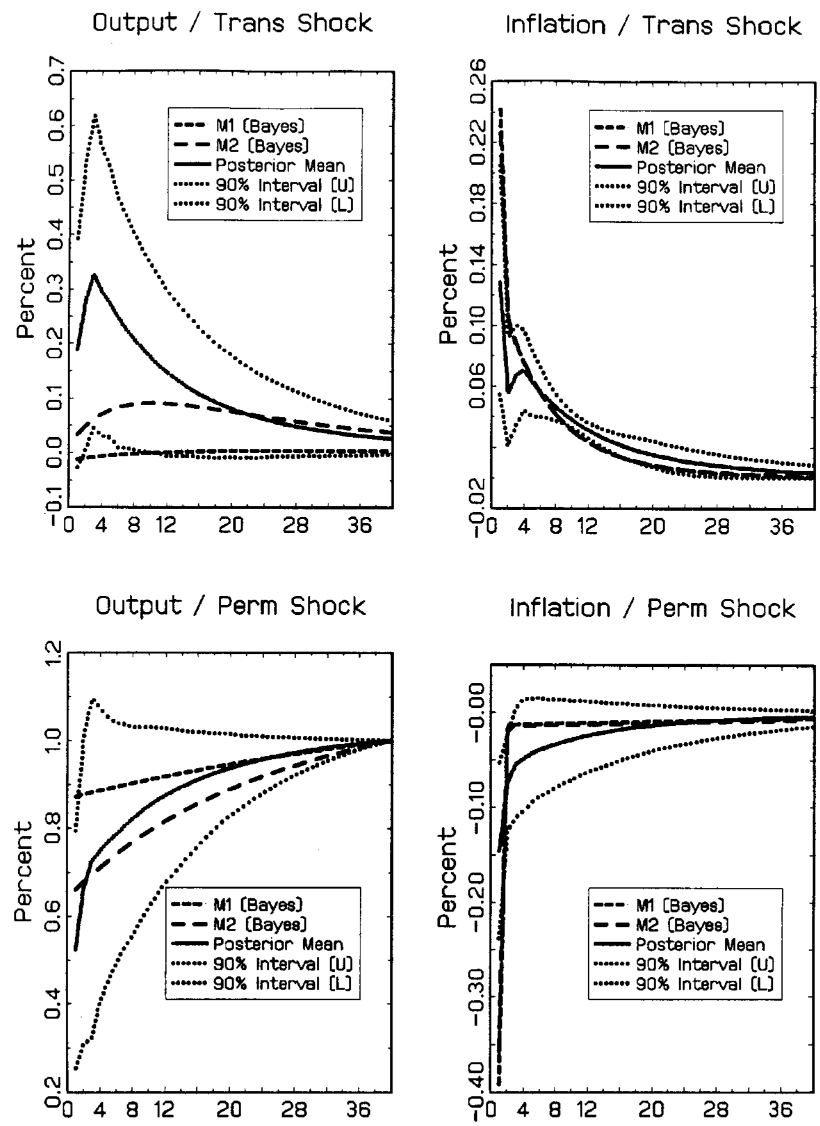

Figure 1. Posterior distribution of impulse response functions and Bayes predictors (lags 1 to 40) for cash-in-advance ($\mathcal{M}1$) and portfolio-adjustment-cost ($\mathcal{M}2$) model

In response to a permanent technology shock output rises steadily. Our posterior estimates do not exhibit an initial overshooting of output. The technology shock is normalized such that the long-run increase of output is 1%. The technology shock causes the price level to drop, leading to a negative inflation rate. According to the DSGE models, the price level adjusts quickly, whereas the overall posterior indicates a delayed response.

Evaluation summary statistics are reported in Table VI. We used the 12×12 identity matrix

Table VI. Model evaluation statistics for impulse response functions (lags 1 to 12 jointly)

	$\hat{\varphi}_{q,i}$	Impulse response function d ln $GDP/\mathrm{d}\epsilon_M$	d Δ ln $P/\mathrm{d}\epsilon_M$	d ln $GDP/\mathrm{d}\epsilon_A$	d Δ ln $P/\mathrm{d}\epsilon_M$
L_q-risk	$i=1$	0.0565	0.0013	0.0221	0.0059
	$i=2$	0.0264	0.0012	0.0036	0.0045
$C_{\chi^2}(\hat{\varphi}_{q,i}\mid Y_T)$	$i=1$	4.7725	15.260	38.733	130.00
	$i=2$	4.4662	14.151	17.020	104.62
L_{χ^2}-risk	$i=1$	0.4142	0.8137	0.9560	0.9939
	$i=2$	0.3856	0.7952	0.8453	0.9914

Notes: As L_q-risk we only report $\tilde{R}_q(\hat{\varphi}_{q,i}|Y_T)$ (see equation (19)). $C_{\chi^2}(\hat{\varphi}_{q,i}|Y_T)$ is defined in equation (21). $i=1$ is the standard cash-in-advance model and $i=2$ is the portfolio adjustment cost model.

W, scaled by the factor 1/12, to compute the L_q-loss. The visual impression from the plots is confirmed by the L_q-statistics: IRFs of the portfolio adjustment cost model resemble the posterior mean response more closely than the cash-in-advance model IRFs do. We also report the weighted distances $C_{\chi^2}(\hat{\varphi}_{q,i}\mid Y_T)$ and the posterior expected L_{χ^2}-loss. Since for all four response functions $C_{\chi^2}(\hat{\varphi}_{q,1}\mid Y_T) > C_{\chi^2}(\hat{\varphi}_{q,2}\mid Y_T)$ the ranking under the L_{χ^2}-loss function is the same as the quadratic loss function. The values of the L_{χ^2}-risk imply that the posterior probability of an IRF that is further apart from the posterior mean response (in L_{χ^2} metric) than the response obtained from the PAC model $\mathcal{M}_2$ is 39.56% for d ln $GDP/\mathrm{d}\epsilon_M$ and 79.52% for dΔ ln $P/\mathrm{d}\epsilon_M$. The L_{χ^2}-risk for the GDP response to a technology shock is larger than the visual inspection of Figure 1 suggests. This is due to the weighting induced by the inverse posterior covariance matrix V_{φ}^{-1}. We conclude that the PAC model is more suitable to predict the effects of an unanticipated increase in the money growth rate and the effect of a change in total factor productivity.

Step 4

Can one obtain better matching impulse response functions to an unanticipated money growth shock by changing the parameter κ of the portfolio adjustment cost model $\mathcal{M}_2$? A loss function estimate of κ is computed to answer this question. Let φ be the 80×1 vector composed of d ln $GDP/\mathrm{d}\epsilon_M$ and dΔ ln $P/\mathrm{d}\epsilon_M$. We use the quadratic loss function

$$L(\varphi, \hat{\varphi}) = (\varphi - \hat{\varphi})'(\varphi - \hat{\varphi}) \tag{40}$$

with the identity matrix as weight matrix. To examine the improvements that can be achieved through a modification of the adjustment cost parameter κ we set all other parameters equal to their posterior mean value (see Table III). The loss function estimate is $\hat{\kappa}_{IRF} = 114.67$. Thus, the marginal costs of changing bank deposits are larger than under the posterior mean estimate. The predicted responses are depicted in Figure 2. While the loss function estimate increases the real impact of the monetary shock, a discrepancy between the posterior mean response and the DSGE model response remains. A further increase of κ amplifies the real effect in the initial period, but also increases its persistence. To match the posterior mean response more closely, other parameters have to change: in particular, a counterfactually large capital share, a high steady-state interest rate and a higher short-run labour supply elasticity $(1-h)/h$ are needed.

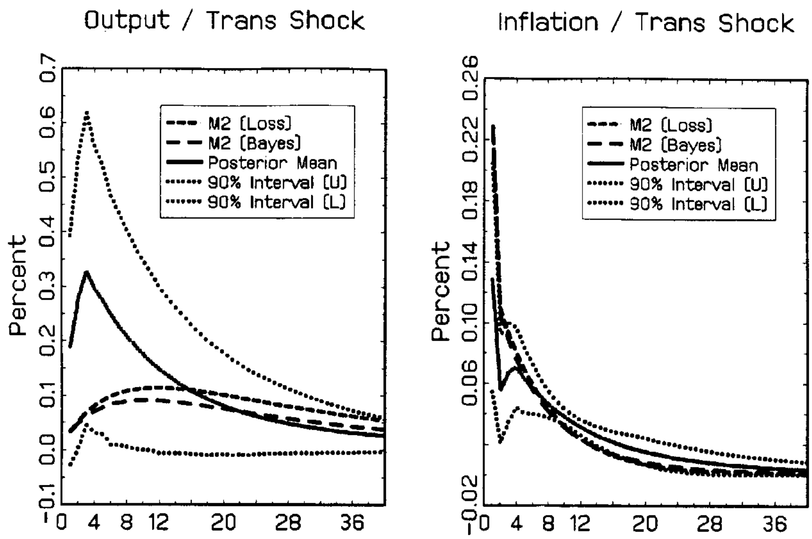

Figure 2. Posterior distribution of impulse response functions, loss function predictor and Bayes predictor (lags 1 to 40) for portfolio-adjustment-cost model

4.3 Discussion

We found in our empirical analysis of the log-linearized DSGE models[7] that the cash-in-advance model has higher posterior probability than the portfolio adjustment cost model and achieves a better in-sample time series fit. Both models overpredict the negative correlation between output growth and inflation. Unlike the PAC model, the CIA model is not able to generate a positive real effect of money growth shocks on aggregate output. Overall, the impulse response dynamics of the PAC model resemble the posterior mean impulse response functions more closely than the responses of the CIA model.

Based on the impulse response function evaluation, our assessment of the DSGE models is more favourable than the results of Nason and Cogley (1994). Both analyses are based on a comparison of DSGE model response functions to response functions estimated through an identified VAR. Nason and Cogley (1994) conducted classical tests of the null hypotheses that the IRFs were generated from the structural models. Conditional on calibrated DSGE parameter values, they derived the sampling distribution of VAR based impulse response function estimates. The variance of the sampling distribution provided a metric to assess the discrepancy between DSGE model responses and VAR responses. For each model $\mathcal{M}_i$ a different covariance matrix is used. Since the parameter values are regarded as known and the DSGE models are well approximated by a fourth-order VAR, Nason and Cogley (1994) obtained small sampling variances which let the discrepancy between DSGE model prediction

[7] To the extent that the log-linearization provides an accurate approximation of the equilibrium in the non-linear model economies, the substantive conclusions drawn from the analysis also hold for the non-linear models. Log-linear approximations are generally accurate if the decision rules of the agents are smooth in the neighbourhood of the steady state and the deviations from the steady state are modest. A numerical assessment of the accuracy of various solution methods for DSGE models can be found in Taylor and Uhlig (1990).

and VAR estimate appear large. In our calculation of the L_{χ^2}-risk, the discrepancy between DSGE model prediction and the posterior mean response function is scaled by the posterior variance of the impulse responses, which is larger than Nason and Cogley's (1994) sampling variances. This lets model predictions and posterior mean appear closer.

In our approach the metric between DSGE model prediction and overall posterior distribution of population characteristics is determined by the researchers' preferences. The L_q, L_p, and L_{χ^2}-loss functions lead to model evaluations that are similar in spirit to many existing evaluation approaches. If a model dominates other structural models under several different loss functions, then it is a clear strike against the its competitors and the inference is robust.

Our approach is particularly interesting for policy analyses because the prediction of policy effects can be convincingly stated within a loss function-based framework. Throughout the paper, we considered a case in which the reference model was just identified to yield predictions with respect to the population characteristics φ. The attraction of tightly parameterized structural models, however, rests in their ability to generate predictions about aspects of economic activity that cannot be analysed in the context of very general and flexible specifications. An example is the prediction of the effects of a rare change in a policy regime, such as switching from an exogenous money supply rule to a feedback rule. To choose a DSGE model for predictions that are beyond the scope of the reference model, it is important to carefully document in which respect the DSGE model fits the data, as emphasized by Sims (1996). Our loss function-based model evaluation approach is a useful and flexible tool for examining the empirical adequacy of DSGE models.

5. CONCLUSION

In this paper we proposed an econometric procedure for the evaluation and comparison of DSGE models. Unlike in many previous econometric approaches, we explicitly take the possibility into account that the DSGE models are misspecified and introduce a reference model to complete the model space. Three loss functions were proposed to assess the discrepancy between DSGE model prediction and the overall posterior distribution of population characteristics that the researcher is trying to match. The evaluation procedure was applied to the comparison of a standard cash-in-advance model and a portfolio adjustment cost model. We examined the empirical adequacy of these two monetary models based on posterior model probabilities and their ability to capture the correlation between output growth and inflation and impulse response dynamics. The evaluation procedure can be easily extended to examine DSGE model predictions of the spectrum of macroeconomic aggregates. We focused on the evaluation of log-linearized DSGE models. The principle of our approach can also be applied to non-linear structural models. However, the implementation of the computations is likely to be more cumbersome.

APPENDIX: COMPUTATIONAL DETAILS

A.1 Reference Model

For each lag-length p, the posterior density of the VAR parameters is of the form

$$p(B, \Sigma | Y_T, \mathcal{M}_0) = p(\Sigma | Y_T, \mathcal{M}_0) p(B | \Sigma, Y_T, \mathcal{M}_0) \quad \text{(A1)}$$

Define $\hat{B}_T = (X_T' X_T)^{-1} X_T' Y_T$ and

$$\hat{\Sigma}_T = \frac{1}{T-p}(Y_T - X_T\hat{B}_T)'(Y_T - X_T\hat{B}_T)$$

The density $p(\Sigma | Y_T, \mathcal{M}_0)$ has the shape of an inverted Wishart distribution with $T-p-k$ degrees of freedom and parameter matrix $H_T = (T-p)\hat{\Sigma}_T$. Conditional on Σ, the posterior distribution of the coefficients $vec(B)$ is multivariate normal with mean $vec(\hat{B}_T)$ and covariance matrix $\Sigma \otimes (X_T' X_T)^{-1}$ (see Zellner, 1971). To obtain independent draws $(B,\Sigma)^{(s)}$, $s = 1, \ldots, n_{sim}$, we successively draw from $\Sigma | Y_T, \mathcal{M}_0$ and $B | \Sigma, Y_T, \mathcal{M}_0$. The results reported in the text are based on $n_{sim} = 50{,}000$. For each draw $(B,\Sigma)^{(s)}$ we calculate the implied correlations and impulse response functions $\varphi^{(s)}$. With some small posterior probability $\epsilon > 0$, the VAR is non-stationary and the unconditional second moments of output growth and inflation do not exist. In the rare event that a draw of B implies non-stationarity it is discarded. This corresponds to a truncated prior $p(B,\Sigma)$ that is zero in the non-stationary regions of the parameter space. Posterior moments $\int g(\varphi)p(\varphi | Y_T, \mathcal{M}_0)d\varphi$ are approximated by simulation averages

$$\frac{1}{n_{sim}}\sum_{s=1}^{n_{sim}} g(\varphi^{(s)})$$

A central limit theorem for sequences of independently distributed random variables can be used to assess the numerical accuracy of the Monte Carlo approximation (see Geweke, 1999b). The marginal data density $p(Y_T^* | Y_*, \mathcal{M}_0) = p(Y_T | \mathcal{M}_0)/p(Y_* | \mathcal{M}_0)$:

$$p(Y_T|\mathcal{M}_0) = \frac{\Gamma[(T-p-k)/2]\Gamma[(T-p-k-1)/2]}{\pi^{(T-k-p)-1/2}} \left|X_T' X_T\right|^{-1} \left|(T-k)\hat{\Sigma}_T\right|^{-(T-k-p)/2}$$

where $\Gamma[x]$ denotes the Gamma-function. $p(Y_* | \mathcal{M}_0)$ is obtained by replacing T, $\hat{\Sigma}_T$, and X_T by T_*, $\hat{\Sigma}_*$, and X_*, defined in Section 4.1.

DSGE Models

Conditional on the data Y_T and a set of parameter values $\theta_{(i)}$ the Kalman Filter is used to evaluate the log-posterior density up to a constant

$$p(\theta_{(i)} | Y_T, \mathcal{M}_i) \propto p(Y_T^* | Y_*, \theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)} | Y_*, \mathcal{M}_i)$$

A numerical optimization routine is used to compute the mode $\tilde{\theta}_{(i)}$ of the posterior density. Then the inverse Hessian $\tilde{\Sigma}_{(i)}$ defined in equation (38) is computed. A random walk Metropolis algorithm is used to generate $n_{sim} = 90{,}000$ draws $\theta_{(i)}^{(s)}$ from the posterior $p(\theta_{(i)} | Y_T, \mathcal{M}_i)$. At each iteration s, a candidate parameter vector $\vartheta_{(i)}$ is drawn from a jumping distribution $J_s(\vartheta_i | \theta_{(i)}^{(s-1)})$ and the density ratio

$$r = \frac{p(Y_T^* | Y_*, \vartheta_{(i)}, \mathcal{M}_i) p(\vartheta_{(i)} | Y_*, \mathcal{M}_i)}{p(Y_T^* | Y_*, \theta_{(i)}^{(s-1)}, \mathcal{M}_i) p(\theta_{(i)}^{(s-1)} | Y_*, \mathcal{M}_i)}$$

is calculated. The jump from $\theta_{(i)}^{(s-1)}$ is accepted ($\theta_{(i)}^{(s)} = \vartheta_{(i)}$) with probability $\min(r,1)$ and rejected ($\theta_{(i)}^{(s)} = \theta_{(i)}^{(s-1)}$) otherwise. The sequence $\{\theta_{(i)}^{(s)}\}_{s=1}^{n_{sim}}$ forms a Markov chain that converges to the target posterior distribution as $n_{sim} \to \infty$. The first $s_0 = 10{,}000$ draws are discarded. We used a Gaussian jumping distribution of the form $J_s \sim \mathcal{N}(\theta_{(i)}^{(s-1)}, c^2\tilde{\Sigma}_{(i)})$. We chose $c = 0.2$ for both models. Based on the draws $\theta_{(i)}^{(s)}$ we calculated the implied population characteristics $\varphi^{(s)}$. To assess the convergence of the Markov chain, we simulated independent sequences, with starting points drawn from an overdispersed distribution. We computed potential scale reduction factors described in Gelman *et al*. (1995, Chapter 11.4). The potential scale reduction factors for model parameters, correlations, and impulse responses were less than 1.005. Hence, we regard the number of draws as large enough to conduct the evaluation. Posterior moments $\int g(\varphi)p(\varphi | Y_T, \mathcal{M}_i)d\varphi$ are approximated by simulation averages

$$\frac{1}{n_{sim}} \sum_{s=s_0+1}^{n_{sim}} g(\varphi^{(s)})$$

Since the sequences $\{g(\varphi^{(s)})\}_{s=s_0+1}^{n_{sim}}$ are serially correlated we used Newey–West standard errors to assess the numerical accuracy of the approximation. The estimated simulation errors of the posterior moments reported in the text are less than 1%. While this procedure is not formally justified (see Geweke, 1999b), it yielded simulation error estimates that were consistent with the variation that we found across different runs of the Markov chain. The modified harmonic mean estimator of $p(Y_T^* | Y_*, \mathcal{M}_i)$ is defined as follows (Geweke, 1999b):

$$\hat{p}(Y_T^* | Y_*, \mathcal{M}_i) = \left[\frac{1}{n_{sim} - s_0} \sum_{s=s_0+1}^{n_{sim}} \frac{f(\theta_{(i)}^{(s)})}{p(\theta_{(i)} | Y_*, \mathcal{M}_i) p(Y_T^* | \theta_{(i)}, Y_*, \mathcal{M}_i)} \right]^{-1}$$

Let $\bar{\theta}_{(i)}$ and $V_{\theta,i}$ be the sample mean and covariance matrix computed from the posterior draws $\theta_{(i)}^{(s)}$. Then for $\tau \in (0,1)$

$$f(\theta) = \tau^{-1}(2\pi)^{-d_i/2} |V_{\theta,i}|^{-1/2} \exp\left[-0.5(\theta - \bar{\theta}_{(i)})' V_{\theta,i}^{-1} (\theta - \bar{\theta}_{(i)}) \right] \times \left\{ (\theta - \bar{\theta}_{(i)})' V_{\theta,i}^{-1} (\theta - \bar{\theta}_{(i)}) \leq F_{\chi^2_{d_i}}^{-1}(\tau) \right\}$$

where d_i is the dimension of $\theta_{(i)}$ and $F_{\chi^2_{k_i}}^{-1}(\tau)$ is the inverse cdf of a χ^2 random variable with d_i degrees of freedom. The discrepancy of ln $\hat{p}(Y_T^* | Y_*, \mathcal{M}_i)$ across truncation levels τ and simulation runs was less than 0.6.

ACKNOWLEDGEMENTS

This paper is based on Chapter 4 of my PhD dissertation. I wish to thank my advisors Peter Phillips and Chris Sims for valuable comments, suggestions, and support. Thanks also to Yongsung Chang, John Geweke (Co-Editor), Kevin Reffett, and two anonymous referees. Financial assistance from the German Academic Exchange Service through Doktoranden-stipendium aus Mitteln des zweiten Hochschulsonderprogramms and the Alfred P. Sloan Foundation are gratefully acknowledged.

REFERENCES

Altug, S. (1989), 'Time-to-build and aggregate fluctuations: some new evidence', *International Economic Review*, **30(4)**, 889–920.

Blanchard, O. J. and D. Quah (1989), 'The dynamic effects of aggregate demand and supply disturbances', *American Economic Review*, **79**, 655–673.

Box, G. E. P. (1980), 'Sampling and Bayes' inference in scientific modelling and robustness', *Journal of the Royal Statistical Society A*, **143**, 383–430.

Canova, F. (1994), 'Statistical inference in calibrated models', *Journal of Applied Econometrics*, **9**, S123–144.

Canova, F., M. Finn and A. R. Pagan (1994), 'Evaluating a real business cycle model', in C. P. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press, Oxford.

Christiano, L. J. (1991), 'Modeling the liquidity effect of a money shock', *Federal Reserve Bank of Minneapolis — Quarterly Review*, Winter, 3–34.

Christiano, L. J. and M. Eichenbaum (1992a), 'Current real-business cycle theory and aggregate labor-market fluctuations', *American Economic Review*, **82**, 430–450.

Christiano, L. and M. Eichenbaum (1992b), 'Liquidity effects and the monetary transmission mechanism', *American Economic Review*, **82**, 346–353.

Cogley, T. and J. M. Nason (1995), 'Output dynamics in real business cycle models', *American Economic Review*, **85**, 492–511.

Crowder, M. (1988), 'Asymptotic expansions of posterior expectations, distributions, and densities for stochastic processes', *Annals of the Institute for Statistical Mathematics*, **40**, 297–309.

DeJong, D. N., B. F. Ingram and C. H. Whiteman (1996), 'A Bayesian approach to calibration', *Journal of Business Economics and Statistics*, **14**, 1–9.

DeJong, D. N., B. F. Ingram and C. H. Whiteman (2000), 'A Bayesian approach to dynamic macroeconomics', *Journal of Econometrics*, (forthcoming).

Diebold, F. X., L. E. Ohanian and J. Berkowitz (1998), 'Dynamic equilibrium economies: a framework for comparing models and data', *Review of Economic Studies*, **65**, 433–452.

Fair, R. (1994), *Testing Macroeconometric Models*, Harvard University Press, Cambridge, MA.

Fuerst, T. (1992), 'Liquidity, loanable funds, and real activity', *Journal of Monetary Economics*, **29**, 3–24.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin (1995), *Bayesian Data Analysis*, Chapman & Hall, New York.

Geweke, J. (1999a), 'Computational experiments and reality', Mimeo, University of Minnesota.

Geweke, J. (1999b), 'Using simulation methods for Bayesian econometric models: inference, development and communication', *Econometric Reviews*, **18**, 1–126.

Good, I. J. (1952), 'Rational decisions', *Journal of the Royal Statistical Society B*, **14**, 107–114.

Ireland, P. N. (1998), 'A method for taking models to the data', Mimeo, Rutgers University.

Kass, R. E. and A. E. Raftery (1995), 'Bayes factors', *Journal of the American Statistical Association*, **90**, 773–795.

Kim, J. (2000), 'Constructing and estimating a realistic optimizing model of monetary policy', *Journal of Monetary Economics*, **45**, 329–360.

Kydland, F. and E. Prescott (1996), 'The computational experiment: an econometric tool', *Journal of Economic Perspectives*, **10**, 69–85.

Leeper, E. M. and C. A. Sims (1994), 'Toward a modern macroeconomic model usable for policy analysis', in S. Fischer and J. J. Rotemberg (Eds), *NBER Macroeconomics Annual 1994*, MIT Press, Cambridge, MA.

Nason, J. M. and T. Cogley (1994), 'Testing the implications of long-run neutrality for monetary business cycle models', *Journal of Applied Econometrics*, **9**, S37–70.

Phillips, C. B. (1996), 'Econometric model determination', *Econometrica*, **64**, 763–812.

Rotemberg, J. and M. Woodford (1996), 'Real business cycle models and the forecastable movements in output, hours, and consumption', *American Economic Review*, **86**, 71–89.

Sims, C. A. (1982), 'Policy analysis with econometric models', *Brookings Paper of Economic Activity*, 107–164.

Sims, C. A. (1996), 'Macroeconomics and methodology', *Journal of Economic Perspectives*, **10**, 105–120.

Smith, A. A. Jr (1993), 'Estimating nonlinear time-series models using simulated vector autoregressions', *Journal of Applied Econometrics*, **8**, S63–S84.

Söderlind, P. (1994), 'Cyclical properties of a real business cycle model', *Journal of Applied Econometrics*, **9**, S113–122.

Taylor, J. B. and H. Uhlig (1990), 'Solving non-linear stochastic growth models: a comparison of alternative solution methods', *Journal of Business Economics and Statistics*, **8**, 1–34.

Zellner, A. (1971), *Introduction to Bayesian Inference in Econometrics*, John Wiley, New York.

