



Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19103-6297
pier@ssc.upenn.edu
<http://www.econ.upenn.edu/pier>

PIER Working Paper 99-007

“A Unified Econometric Framework for the Evaluation of DSGE Models”

by

Frank Schorfheide

A Unified Econometric Framework for the Evaluation of DSGE Models

Frank Schorfheide^{*}

This version: August 1999

^{*}Correspondence: c/o University of Pennsylvania, Department of Economics, Philadelphia, PA 19104-6297. E-mail: schorf@ssc.upenn.edu. URL: <http://www.econ.upenn.edu/~schorf>. This paper is based on Chapter 4 of my Ph.D. dissertation. I wish to thank my advisors Peter Phillips and Chris Sims for valuable comments, suggestions, and support. Financial Assistance from the German Academic Exchange Service through "Doktorandenstipendium aus Mitteln des zweiten Hochschulsonderprogramms" and the Alfred P. Sloan Foundation are gratefully acknowledged.

Abstract

A unified Bayesian framework for the econometric evaluation of dynamic stochastic general equilibrium (DSGE) models is presented. The evaluation is coherent under misspecification, that is, low posterior probability of all DSGE models in a candidate set, as well as no misspecification. The framework encompasses many of the existing evaluation schemes as special cases, including Kydland and Prescott's (1996) informal calibration and the traditional macroeconomic approach of judging models according to their ability to track and forecast aggregate time series. A detailed illustrative application of the framework to a standard cash-in-advance model and a liquidity model is provided. The models are evaluated according to their predictions of co-movements between output growth and inflation, and responses to discretionary changes in the growth rate of money supply.

JEL CLASSIFICATION: C11, C32, C52

KEY WORDS: Bayesian, Dynamic Stochastic General Equilibrium Model, Econometric Model Evaluation, Loss Function.

1 Introduction

While dynamic stochastic general equilibrium (DSGE) models provide a complete multivariate stochastic process representation for the data, and are also interesting from a theoretical perspective, they impose very strong restrictions on actual time series and are in many cases rejected against reduced form models such as vector autoregressions (VAR). The debate among macroeconomists and econometricians about the empirical evaluation of these models is controversial, e.g., Kydland and Prescott (1996), Hansen and Heckman (1996), and Sims (1996). The purpose of this paper is to provide a unified econometric framework that enables a coherent model evaluation and determination regardless of the degree of misspecification of the DSGE models. It encompasses many existing approaches as special cases. Although this framework will deliver some absolute measures of fit for structural models, the focus is on model comparisons. Which DSGE model, among a candidate set of two or more models, summarizes the regular features of the data most accurately?

The potential misspecification of all candidate models poses a conceptual difficulty for the design of econometric procedures. The inference problem is not to determine the “true” model among an exhaustive set. Instead, the goal is to find an approximation to an appropriate probabilistic representation of the data from a restricted choice set. Two questions have to be addressed: how should one construct the probabilistic representation of the data that serves as a benchmark, and how should the distance to its second best approximation be measured?

The proposed framework is Bayesian. It is based on a joint probability distribution for models, parameters, and the data. A reference model is considered in addition to the structural candidate models to cope with their potential misspecification. In general, such a reference model should impose weaker restrictions on the data than the structural models and achieve an acceptable fit. Vector autoregressions, widely used in empirical macroeconomics, can serve as a reference for the evaluation of DSGE models. The probability distribution for the data is specified as a mixture of the structural models and the reference model. Based on this mixture, it is possible

to compute posterior distributions for features of the aggregate time series, such as patterns of covariation. If the statistical fit of the structural models is poor, the overall posterior distributions are dominated by the contribution of the reference model.

A notion of distance to the mixture of candidate models and reference model is needed to determine which structural model provides a second best approximation. A frequently used measure of discrepancy between probability distributions is the Kullback-Leibler distance. Each structural model as well as the mixture of structural and reference model define a probability distribution for the data. In principle, the econometric comparison of the candidate models could be based on their respective proximity in the Kullback-Leibler sense to the overall mixture distribution. However, macroeconomists are generally more interested in the evaluation of DSGE models based on their ability to generate realistic patterns of co-movements among macroeconomic aggregates or impulse responses to structural shocks.

The proposed evaluation framework will incorporate this practice as follows: based on a specific research question the investigator determines which time series characteristics are important for the model comparison, e.g., unconditional second moments or impulse response functions. Moreover, a loss function that penalizes deviations of predicted characteristics from actual characteristics is chosen. For each structural model the implications, or predictions, with respect to the characteristics of interest are derived. Then the expected loss of the model predictions is calculated. The expectation is taken with respect to the overall posterior distribution that takes the contribution of the reference model into account. The structural model that attains the smallest posterior prediction loss wins the comparison. This idea is generalized to take parameter uncertainty into account, derive loss function estimators, and construct mixtures of structural models that approximate the overall probabilistic representation of the data.

The evaluation approach remains sensible in situations where all structural models have negligible posterior probability compared to the reference model. If in addition, the loss function is quadratic, then the framework leads to an evaluation

procedure that closely resembles Kydland and Prescott's (1982, 1996) calibration approach. On the other hand, if the parsimonious DSGE models have high posterior probability compared to a more general reference model, then the structural models are essentially judged according to their posterior probabilities. Since the posterior probabilities provide a measure for one-step ahead out-of-sample forecasting performance, the procedure resembles the macroeconometric tradition of evaluating models based on their ability to track and forecast aggregate time series (cf. Fair, 1984 and 1994). The forecasting performance of a prototypical DSGE model has been examined by Dejong *et al.* (1997).

Many evaluation techniques that were previously proposed in the literature are based on p -values for various 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 structural models. A non-exhaustive list of examples are Christiano and Eichenbaum (1992a), Burnside *et al.* (1993), Söderlind (1994), Smith (1993), Canova *et al.* (1994), and Nason and Cogley (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 the misspecified models.

Alternatively, the implications of the structural models can be compared to predictions of a reference model. Diebold *et al.* (1998) proposed a frequentist procedure, that evaluates structural models under a sampling distribution obtained from non-parametric spectral estimates. Their framework is similar to our approach in the sense that it makes explicit use of loss functions. Its conceptual drawback is that the reference model has always posterior probability one and the structural models have posterior probability zero regardless of their fit. We will demonstrate in an empirical illustration that despite low posterior probability the structural models can have a significant impact on the shape of the overall posterior distribution because they deliver more concentrated predictive distributions than the VAR.

Dejong *et al.* (1996) propose a Bayesian approach to calibration which also assigns zero probability to the structural model. In particular, the authors 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. This methodology is further discussed and applied in Geweke (1999). Our framework can be viewed as a generalization in two dimensions. For non-singular DSGE models we do not impose zero posterior probability, prior predictive distributions obtained from the structural models are replaced by posterior predictive distributions, and several additional criteria and loss functions that can be useful for model evaluations are introduced.

Due to the explicit consideration of loss functions defined over time series characteristics that are of particular interest to researchers, the framework goes beyond model determination based on posterior probabilities or Bayes factors, e.g, Kass and Raftery (1995). However, asymptotic approximations (e.g. Phillips, 1996), as well as computational aspects, for instance discussed in Geweke (1995), are very important for the successful implementation of the framework.

Section 2 of this paper provides the details of the loss function based model evaluation approach. Section 3 contains the empirical illustration, adopted from the work of Nason and Cogley (1994). A standard cash-in-advance (CIA) model and a CIA model with liquidity effect (Christiano, 1991, and Christiano and Eichenbaum, 1992b) are evaluated according to their implications about covariation patterns among output growth and inflation, and the effects of a discretionary change in the growth rate of money supply. Section 4 concludes.

2 A Decision Theoretic Model Evaluation Approach

2.1 Notation and Setup

Suppose that the goal is to evaluate two DSGE models, \mathcal{M}_1 and \mathcal{M}_2 , and determine which of the two models provides a better summary of macroeconomic time series features. To be more explicit, the reader may assume that we are particularly interested in a collection of second moment statistics φ , where φ is an $m \times 1$ vector. The time series is denoted by $Y_T = \{y_t\}_{t=1}^T$, where y_t is a $n \times 1$ vector. The parameters of model \mathcal{M}_i are denoted by θ_i . Each model consists of a parametric density function for the data, $p(Y_T|\theta_i, \mathcal{M}_i)$. Since the framework is Bayesian, parameters are treated as random variables and we introduce prior distributions $p(\theta_i)$.

To cope with the potential misspecification of the candidate models, a reference model \mathcal{M}_* is considered in addition to \mathcal{M}_1 and \mathcal{M}_2 . The parameter vector of \mathcal{M}_* is denoted by θ_* , the data density is $p(Y_T|\theta_*, \mathcal{M}_*)$, and the prior density is $p(\theta_*)$. The prior probabilities of the three models are $\pi_{1,0}$, $\pi_{2,0}$, and $\pi_{*,0}$, 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. Three additional assumptions will be made.

- (i) If the structural models are nested within the reference model, the prior $p(\theta_*)$ assigns zero probability to the subset of the parameter space that corresponds to the structural models.
- (ii) To simplify exposition and implementation of the framework it is assumed that the prior distributions of θ_1 , θ_2 , and θ_* , are independent of each other.
- (iii) It is possible to compute predictions for φ based on the reference model.

The third assumption is obviously satisfied if φ is composed of second moments. Based on the parameter values of a VAR it is easily possible to calculate autocovariances for y_t . If φ is composed of response functions to orthogonal structural disturbances then Assumption (iii) is satisfied if an identification scheme for the reference model is available. If φ consists of the effects of a permanent change in

a fiscal or monetary policy rule then it is not possible to derive predictions from a reference model such as a VAR. The discussion of this case is deferred to Section 4.

The first part of the evaluation procedure consists of the computation of posterior densities $p(\theta_i|Y_T, \mathcal{M}_i)$ for the parameter vectors θ_i , $i = 1, 2, *$, posterior model probabilities $\pi_{i,T}$, and the posterior distributions of the characteristics φ conditional on the three models¹ of

$$p(\varphi|Y_T, \mathcal{M}_i) = \int p(\varphi|\theta_i, Y_T, \mathcal{M}_i)p(\theta_i|Y_T, \mathcal{M}_i)d\theta_i \quad (1)$$

for $i = 1, 2, *$. This leads to the overall posterior distribution

$$p(\varphi|Y_T) = \sum_{i=*,1,2} \pi_{i,T}p(\varphi|Y_T, \mathcal{M}_i) \quad (2)$$

where

$$\pi_{i,T} = \frac{\pi_{i,0}p(Y_T|\mathcal{M}_i)}{\sum_{i=1,2,*} \pi_{i,0}p(Y_T|\mathcal{M}_i)}, \quad p(Y_T|\mathcal{M}_i) = \int p(Y_T|\theta_i, \mathcal{M}_i)p(\theta_i|\mathcal{M}_i)d\theta_i$$

We will refer to the latter quantity as the marginal density of the data conditional on model \mathcal{M}_i . The posterior probabilities $\pi_{i,T}$ determine the relative weight of the posterior densities $p(\varphi|Y_T, \mathcal{M}_i)$. If the structural models fit poorly relative to the reference model, then their posterior probabilities are low and the overall posterior density is dominated by $p(\varphi|Y_T, \mathcal{M}_*)$. Vice versa, if the parsimonious structural models fit better than the reference model the posterior mainly determined by $p(\varphi|Y_T, \mathcal{M}_1)$ or $p(\varphi|Y_T, \mathcal{M}_2)$. If the distribution of the data Y_T is singular under the structural models, as it is the case if the dimension of y_t exceeds the number of structural shocks, then the posterior probabilities of \mathcal{M}_1 and \mathcal{M}_2 are zero and $p(\varphi|Y_T) = p(\varphi|Y_T, \mathcal{M}_*)$. This special case is considered in DeJong *et al.* (1996), Geweke (1999), and from a frequentist perspective in Diebold *et al.* (1998).

¹If the vector φ is composed of second moments or impulse response functions then the distribution of φ conditional on θ_i, Y_T and model \mathcal{M}_i is simply a point mass at $\varphi(\theta_i, \mathcal{M}_i)$.

2.2 Loss Function Based Evaluation

At first we describe how to derive predictions from the structural models \mathcal{M}_1 and \mathcal{M}_2 . Consider a decision maker who has to report a point prediction of φ . The decision maker is myopic and bases decisions exclusively on the set of structural models \mathcal{M}_1 and \mathcal{M}_2 . If the set were exhaustive then the optimal predictor $\hat{\varphi}$ would be a solution to the problem

$$\hat{\varphi} = \operatorname{argmin}_{\hat{\varphi} \in \mathbb{R}^m} \int L(\varphi, \hat{\varphi}) \left[\sum_{i=1,2} \tilde{\pi}_{i,T} \int p(\varphi|\theta_i, Y_T, \mathcal{M}_i) p(\theta_i|Y_T, \mathcal{M}_i) d\theta_i \right] d\varphi \quad (3)$$

where $\tilde{\pi}_{i,T}$ denotes the myopic posterior probability $\tilde{\pi}_{i,T} = \pi_{i,T}/(\pi_{1,T} + \pi_{2,T})$, $i = 1, 2$. Since the structural models are potentially misspecified ($\pi_{*,T} > 0$), we will judge the predictions $\hat{\varphi}$ based on their expected loss under the overall posterior distribution of φ , given in Equation (2). The model evaluation and determination involves the examination of the posterior expected loss of $\hat{\varphi}$ and a re-calculation of the posterior weights for the structural models and its parameters, such that the predictions of the myopic decision maker yield a small overall expected prediction loss.

Formally, define the vector of model weights $\lambda = [\lambda_1, \lambda_2] \in S^1$, where S^1 is the simplex $\{\lambda \in \mathbb{R}^2 : \lambda_i \geq 0, \lambda_1 + \lambda_2 = 1\}$. The posterior densities $p(\theta_i|Y_T, \mathcal{M}_i)$ can be replaced by general weight functions $f_i(\theta_i)$ with the property that $\int_{\Theta_i} f_i(\theta_i) d\theta_i = 1$, where Θ_i denotes the domain of θ_i . The predictor $\hat{\varphi}(\lambda, \{f_i\}_{i=1}^J)$ solves the minimization problem

$$\hat{\varphi}(\lambda, f_1, f_2) = \operatorname{argmin}_{\hat{\varphi} \in \mathbb{R}^m} \int L(\varphi, \hat{\varphi}) \left[\sum_{i=1,2} \lambda_i \int p(\varphi|\theta_i, Y_T, \mathcal{M}_i) f_i(\theta_i) d\theta_i \right] d\varphi \quad (4)$$

We denote the domain of densities $f_i(\theta_i)$ by F_i . The expected prediction loss under the overall posterior distribution is

$$\mathcal{R}(\hat{\varphi}(\lambda, f_1, f_2|Y_T)) = \int L(\varphi, \hat{\varphi}(\lambda, f_1, f_2)) \left[\sum_{i=1,2,*} \pi_{i,T} p(\varphi|Y_T, \mathcal{M}_i) \right] d\varphi \quad (5)$$

The model evaluation and selection are based on the problem

$$\min_{\lambda \in S^1, f_i \in F_i} \mathcal{R}(\hat{\varphi}(\lambda, f_1, f_2)|Y_T) \quad (6)$$

and consists of finding weights that solve the minimization under various restrictions on \mathcal{S}^{J-1} and the F_i 's, documenting the expected prediction losses and their sensitivity to changes in λ and f_i .

Equation (6) defines a loss function estimator for model weights and parameters. Loss function estimators are frequently used to estimate parameters of forecasting model. Suppose that a forecaster wants to use a first order autoregressive model to predict the inflation rate four quarters into the future. If the AR(1) model is regarded as potentially misspecified it may be sensible to estimate the autoregressive parameter by the minimization of in-sample four step ahead forecast errors $\sum_{t=1}^T (y_t - \theta y_{t-4})^2$. The justification for Equation (6) is similar. The difference is that the model predictions $\hat{\varphi}(\lambda, f_1, f_2)$ are not compared to actual outcomes φ which are unobservable. Instead they are evaluated with respect to a more general posterior distribution that adjusts for misspecification. This analogy is further explored in Schorfheide (1998).

Loss function based estimation and selection of models is also part of the frequentist framework proposed by Diebold *et al.* (1998). Sampling uncertainty with respect to the parameters of their “true” reference model translates into sampling uncertainty about the parameters of the structural model that best approximates the reference model. In our framework, a posterior distribution for φ is obtained model averaging. Conditional on this posterior distribution, there is no uncertainty about which structural model minimizes the posterior expected loss. Instead, there is sensitivity of the posterior expected loss to the choice of model and parameter weights.

Since it is difficult to solve the minimization (6) for general classes of weight functions we will consider the following restrictions: (i) the weights for the parameters of the structural models are restricted to be equal to the posterior density of these parameters²; (ii) the parameter weights are point masses, represented by the dirac function $\delta_{\{\theta_i = \theta_{i,0}\}}$.³ In the second case, we are searching for a single parameter

²The approach by DeJong *et al.* (1996) and Geweke (1999) restricts the parameter weights to be equal to the prior densities.

³The dirac function has the properties $\delta_{\{x=x_0\}} = 0$ for $x \neq x_0$ and $\int \delta_{\{x=x_0\}} dx = 1$.

value, say $\hat{\theta}_i$, that yields the best prediction for model \mathcal{M}_i . If the model weights are restricted to $[1, 0]$ and $[0, 1]$, the procedure leads to a loss function based model selection. The optimal weights λ and f_1, f_2 define a loss function based pseudo-true mixture of structural models conditional on the observations Y_T that approximates the full posterior distribution. If no restrictions are imposed on λ and the f_i 's then the loss function procedure has the following important characteristic:

Proposition 1 (i) If $\pi_{*,T} = 0$ then $\lambda_i = \pi_{i,T}$, $f_i = p(\theta_i|Y_T)$, $i = 1, 2$, solves the minimization problem (6). This solution does not depend on the choice of loss function. (ii) If $\pi_{*,T} > 0$ then $\lambda_i = \pi_{i,T}$, $f_i = p(\theta_i|Y_T)$, for $i = 1, 2$ is generally not a solution to (6). The solution to the minimization problem becomes loss function dependent.

The consequences of the presence of a reference model on Bayesian model determination have been pointed out previously in the literature, for instance Min and Zellner (1993) in the context of a forecasting problem, and Poirier (1997). If the impact of $\mathcal{R}(\hat{\varphi}(\lambda, \{f_i\}_{i=1}^J)|Y_T, \mathcal{M}_*)$ is expected to be small, then we are willing to ignore the term for practical purposes and might not even consider a reference model *a priori*. The model determination procedure is consistent under general conditions that ensure the consistency of Bayes estimators. If the data are generated from the density $p(Y_T|\theta_{i,0}, \mathcal{M}_i)$, the posterior probability of Model \mathcal{M}_i will converge to one and the posterior distribution of θ_i will concentrate in arbitrary small neighborhoods of $\theta_{i,0}$ for large enough sample sizes.

2.2.1 Loss Functions

While the choice of loss function $L(\varphi, \hat{\varphi})$ may depend on the problem the researcher is trying to solve, we consider three loss functions that are widely applicable in practice. The quadratic loss function is of the form

$$L_q(\varphi, \hat{\varphi}) = (\varphi - \hat{\varphi})'W(\varphi - \hat{\varphi}) \quad (7)$$

where W is a $m \times m$ weighting matrix. Let \mathbb{E}_T denote the expectation with respect to the overall posterior distribution. Since

$$\begin{aligned} \mathbb{E}_T[(\varphi - \hat{\varphi})'W(\varphi - \hat{\varphi})] &= \mathbb{E}_T[(\varphi - \mathbb{E}_T\varphi)'W(\varphi - \mathbb{E}_T\varphi)] \\ &\quad + (\hat{\varphi} - \mathbb{E}_T\varphi)'W(\hat{\varphi} - \mathbb{E}_T\varphi) \end{aligned} \quad (8)$$

the posterior ranking of predictions $\hat{\varphi}$ depends only on the weighted distance between $\hat{\varphi}$ and $\mathbb{E}_T[\varphi]$ 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 . A reasonable choice is $W = V_\varphi^{-1}$, where V_φ denotes the posterior covariance matrix of φ . This weight matrix places less weights on elements of φ that are imprecisely measured.

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

$$L_p(\varphi, \hat{\varphi}) = \mathcal{I}\left\{p(\varphi|Y_T) > p(\hat{\varphi}|Y_T)\right\} \quad (9)$$

The expected L_p loss is similar to a p -value if the posterior density is unimodal. However, its interpretation is different from traditional p -values. Many procedures that have been used to evaluate DSGE models, e.g. Christiano and Eichenbaum (1992a) and Nason and Cogley (1994), are based on classical p -values or their Bayesian counterparts (Canova, 1994). These traditional p -values measure how far the observed sample analog of the population characteristics φ fall into the tail of their sampling distribution derived from the various structural models. Although the p -values can be used to test for each model \mathcal{M}_i the null hypothesis that \mathcal{M}_i is “true”, there is no formal statistical justification for ranking possibly misspecified models according to these p -values. Our approach is to use the data Y_T to derive a posterior distribution for population characteristics and to determine how far the population characteristics implied by the structural models lie in the tails of this posterior distribution.

At last, a L_{χ^2} loss is proposed. The posterior distribution of φ is approximated by a multivariate normal distribution centered at the posterior mean $\bar{\varphi}$. Let V_φ again

denote the posterior variance.

$$L_{\chi^2}(\varphi, \hat{\varphi}) = \mathcal{I} \left\{ (\varphi - \bar{\varphi})' V_{\varphi}^{-1} (\varphi - \bar{\varphi}) < (\hat{\varphi} - \bar{\varphi})' V_{\varphi}^{-1} (\hat{\varphi} - \bar{\varphi}) \right\} \quad (10)$$

If the posterior distribution of φ is Gaussian, L_{χ^2} and L_p loss are identical. However, if the posterior distribution is skewed or multi-modal, evaluations under the two loss functions can lead to different results as we will illustrate in Section 3.

2.2.2 Predictions from Mixtures of Structural Models

This Section briefly discusses the derivation of the predictor $\hat{\varphi}(\lambda, f_1, f_2)$, defined in Equation (4), for the L_q , L_{χ^2} and L_p loss functions. As a special case, the predictive distribution of φ under the mixture of structural models can be discrete. Suppose φ is a vector of unconditional moments of the data. Conditional on a vector of parameters $\theta_i = \theta_{i,0}$ each model i leads to a specific value for φ which we will denote by $\varphi_{i,0}$. If the weight function for θ_i is of the form $\delta_{\{\theta_i=\theta_{i,0}\}}$ then the distribution of φ conditional on Y_T and \mathcal{M}_i degenerates to a point mass on $\varphi_{i,0}$. In our notation,

$$\int p(\varphi|\theta_i, Y_T, \mathcal{M}_i) \delta_{\{\theta_i=\theta_{i,0}\}} d\theta_i = \delta_{\{\varphi=\varphi_{i,0}\}} \quad (11)$$

where $\varphi_{i,0}$, the prediction of model \mathcal{M}_i with $\theta_{i,0}$ plugged in.

Under the quadratic loss function L_q the predictor that solves the minimization (4) is the weighted averaged

$$\hat{\varphi}_q(\lambda, f_1, f_2) = \int \varphi \left(\sum_{i=1,2} \lambda_i \int p(\varphi|\theta_i, Y_T, \mathcal{M}_i) f_i(\theta_i) d\theta_i \right) d\varphi \quad (12)$$

which reduces to $\sum_{i=1,2} \lambda_i \varphi_{i,0}$ in the discrete case.

For the χ^2 -loss function we have to solve

$$\begin{aligned} \min_{\hat{\varphi} \in \mathbf{R}^m} \quad & \int \mathcal{I} \left\{ (\varphi - \bar{\varphi})' V^{-1} (\varphi - \bar{\varphi}) < (\hat{\varphi} - \bar{\varphi})' V^{-1} (\hat{\varphi} - \bar{\varphi}) \right\} \\ & \times \left(\sum_{i=1,2} \lambda_i \int p(\varphi|\theta_i, Y_T, \mathcal{M}_i) f_i(\theta_i) d\theta_i \right) d\varphi \end{aligned} \quad (13)$$

For the computation of $\hat{\varphi}_q$ the mean $\bar{\varphi}$ and variance V parameters for the L_{χ^2} loss function are determined from the mixture obtained by weighting the \mathcal{M}_i models with

λ and their parameters with $f_i(\theta_i)$ instead of the overall posterior distribution. Let $\hat{\varphi}_{\chi^2}(\lambda, f_1, f_2) = \tilde{\varphi}$. The loss function then simplifies to $\mathcal{I}\{(\varphi - \tilde{\varphi})'V^{-1}(\varphi - \tilde{\varphi}) < 0\}$. Since the quadratic function of $(\varphi - \tilde{\varphi})$ is always non-negative, it can be deduced that the expected loss is zero and that $\hat{\varphi}_{\chi^2}$ is indeed a solution to (4). Note that it need not be a unique solution, particularly if the predictive distribution of φ under the mixture is discrete.

The predictor $\hat{\varphi}_p(\lambda, f_1, f_2)$ corresponds to the value of φ that maximizes the function $\sum_{i=1,2} \lambda_i \int p(\varphi|\theta_i, Y_T, \mathcal{M}_i) f_i(\theta_i) d\theta_i$. As under L_{χ^2} , the expected loss is zero. To cover the discrete plug-in case we adopt the convention that $\lambda_i \delta_{\{x=x_0\}} < \lambda_j \delta_{\{x=x_0\}}$ if $\lambda_i < \lambda_j$. Technically, there is the possibility that the predictor $\hat{\varphi}_p$ is not uniquely defined by the argmax. However, small changes in the weights $\lambda_i, f_i(\theta_i)$ will overcome the non-uniqueness so that it poses no practical problem for the evaluation procedure.

2.2.3 A General Measure of Divergence

In situations where the specification of a loss function is difficult but the posterior distribution of φ and the predictive distribution of φ under the pseudo-true mixture of candidate models are absolutely continuous with respect to each other, the candidate models can be evaluated based on the following entropy measure

$$KL(\lambda, f_1, \dots, f_J) = \int \left(\sum_{i=1,2,*} \pi_{i,TP}(\varphi|Y_T, \mathcal{M}_i) \right) \times \ln \left[\frac{\sum_{i=1,2,*} \pi_{i,TP}(\varphi|Y_T, \mathcal{M}_i)}{\sum_{i=1,2} \lambda_i \int p(\varphi|Y_T, \theta_i, \mathcal{M}_i) f(\theta_i) d\theta_i} \right] d\varphi \quad (14)$$

Model and parameter weights are selected to minimize the Kullback Leibler distance between the posterior distribution of φ and the predictive distribution of the candidate models. This information theoretic measure of divergence complements the confidence interval overlap criterion proposed by DeJong *et al.* (1996).

2.2.4 Special Cases

Two frequently used and seemingly very different model evaluation approaches can be interpreted as approximations to evaluation procedures that arise as special cases within the proposed framework.

Informal Moment Matching: In the real business cycle literature DSGE models are often informally judged by their ability to replicate patterns of covariation among macroeconomic variables. Model predictions are simply compared to sample estimates without paying much attention to standard errors of estimates. Suppose the prior probability of the reference model $\pi_{*,0}$ is one, the loss function is quadratic, the prior distribution of the structural model parameters is degenerate and concentrates at a single value, and φ consists of a collection of unconditional second moments of y_t . Selecting the DSGE model that minimizes the posterior expected loss is equivalent to

$$\min_{\lambda \in S^{J-1}} (\hat{\varphi}_* - \hat{\varphi}(\lambda, f_1, f_2))' W (\hat{\varphi}_* - \hat{\varphi}(\lambda, f_1, f_2))$$

where $\hat{\varphi}_*$ is the posterior mean of the population moment φ based on the reference model \mathcal{M}_* . In large enough samples, the posterior mean of φ obtained from a VAR is not very different from the corresponding sample moment. Due to the properties of the quadratic loss function, higher order moments of the posterior distribution of φ have no impact on the model evaluation. Thus, under special assumptions, which may or may not be justified in any particular application, the framework leads to an evaluation procedure in which essentially point predictions from the DSGE models are compared to point estimates from the data.

Traditional Evaluation of Macroeconomic Models: Macroeconometric models based on the Cowles Commission approach, e.g. Fair (1994), were evaluated according to their ability to track and forecast the time series Y_T . Posterior probabilities implicitly provide such a measure of one-step ahead forecast performance. The logarithm of the joint data density $\ln p(Y_T | \mathcal{M}_i)$ can be decomposed as $\ln p(Y_k | \mathcal{M}_i) + \sum_{t=k+1}^T \ln p(y_t | Y_{t-1}, \mathcal{M}_i)$ and $\ln p(y_t | Y_{t-1}, \mathcal{M}_i)$ can be interpreted as predictive score

(Good, 1952). Proposition 1 implies that if the posterior probability of the reference model is approximately zero, the optimal weights for the structural models are given by their respective posterior probabilities. The DSGE model evaluation is essentially based on their forecast performance regardless of the loss function.

3 Empirical Illustration

A step-by-step illustration of the proposed framework is provided to demonstrate the differences and similarities to existing model evaluation and selection procedures. We consider a standard cash-in-advance (CIA) model versus a CIA model with portfolio adjustment costs, cf. Christiano (1991), Christiano and Eichenbaum (1992b), and Nason and Cogley (1994). Both models are driven by two exogenous processes, namely, a random walk production technology

$$\ln A_t = \gamma + \ln A_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2) \quad (15)$$

and an autoregressive money growth process of the form

$$\ln m_t = (1 - \rho) \ln m^* + \rho \ln m_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \quad (16)$$

Equation (16) can be interpreted as a simple monetary policy rule. The innovations η_t reflect discretionary actions of the central bank or “normal” policy making. Changes in m^* or ρ correspond to regime shifts (Sims 1982, 1986). The analysis of regime shifts will be briefly discussed in Section 4. The models are solved in terms of the observables output growth and inflation.

The CIA model implies that a surprise increase in the money supply growth rate leads to a temporary contraction of output due to an anticipated inflation effect. According to the CIA model with portfolio adjustment costs increased money supply forces the interest rate to fall and stimulates economic activity. This liquidity effect potentially dominates the anticipated inflation effect. Although the two structural models might not forecast output growth and inflation data very well, they may be able to reproduce patterns of covariation between the two series and the economy’s response to a discretionary change in the money growth rate.

3.1 Model Specifications

3.1.1 The Two Monetary DSGE Models

The model economy consists of a representative household (HH), a firm (F), and a financial intermediary (FI). Let $m_t \equiv M_{t+1}/M_t$, where M_{t+1} is the stock of money at the end of period t . The money growth rate evolves according to Equation (16). At the beginning of period t , the representative household inherits the entire money stock of the economy, M_t . In the standard CIA model, all decisions are made after, and therefore completely reflect, the current period surprise change in money growth or technology. The household determines how much money d_t to deposit at the bank (FI). These deposits earn interest at the rate $RH_t - 1$. The bank receives household deposits and a monetary injection X_t from the central bank, which it lends to the firm at rate $RF_t - 1$.

The firm starts production and hires labor services h_t from the household. After the firm produced its output, it uses the money borrowed from the FI to pay wages $w_t h_t$. The household's cash balance increases to $M_t - d_t + w_t h_t$. 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 deposits d_t to maximize the sum of discounted expected future utility. It solves the problem

$$\begin{aligned} \max_{\{C_t, h_t, M_{t+1}, d_t\}} \quad & E_0 \left[\sum_{t=0}^{\infty} \beta^t [(1 - \phi) \ln c_t + \phi \ln(1 - h_t)] \right] \\ \text{s.t.} \quad & p_t c_t \leq M_t - d_t + w_t h_t \\ & M_{t+1} = (M_t - d_t + w_t h_t - p_t c_t) + RH_t d_t + f_t + b_t \end{aligned} \quad (17)$$

The firm chooses next periods capital stock k_{t+1} , labor 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

(FI) are owned by households, date t nominal dividends are discounted by date $t + 1$ marginal utility of consumption. Thus, the firm solves the problem

$$\begin{aligned} \max_{\{f_t, k_{t+1}, n_t, l_t\}} \quad & E_0 \left[\sum_{t=0}^{\infty} \beta^{t+1} \frac{f_t}{c_{t+1} p_{t+1}} \right] \\ \text{s.t.} \quad & f_t \leq l_t + p_t [k_t^\alpha (A_t n_t)^{1-\alpha} - k_{t+1} + (1-\delta)k_t] - w_t n_t - l_t R F_t \\ & w_t n_t \leq l_t \end{aligned} \quad (18)$$

The financial intermediary solves the trivial problem

$$\begin{aligned} \max_{\{b_t, l_t, d_t\}} \quad & E_0 \left[\sum_{t=1}^{\infty} \beta^{t+1} \frac{b_t}{c_{t+1} p_{t+1}} \right] \\ \text{s.t.} \quad & b_t = d_t + R F_t l_t - R H_t d_t - l_t + X_t \\ & l_t \leq X_t + d_t \end{aligned} \quad (19)$$

where $X_t = M_{t+1} - M_t$ is the monetary injection. The market clearing conditions for labor market, money market, and goods market are $h_t = n_t$, $p_t c_t = M_t + X_t$, and $c_t + (k_{t+1} - (1-\delta)k_t) = k_t^\alpha (A_t h_t)^{1-\alpha}$. To solve the model, optimality conditions are derived for the optimization problems. The real variables are then detrended by the productivity A_t , and the nominal variables by M_t/A_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). Let $e_t = [\epsilon_t, \eta_t]'$ and z_t be a vector of percentage deviations of detrended model variables from their steady state. The solution to the model can be expressed as

$$z_t = \tilde{T} z_{t-1} + \tilde{R} e_t \quad (20)$$

The liquidity 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 ϵ_t and η_t , the household chooses consumption c_t , the labor 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 the liquidity effect persistent, Christiano and Eichenbaum (1992b) introduced a portfolio management cost, which is given by

$$\tilde{p}_t = \alpha_1 \left[\exp \left\{ \alpha_2 \left[\frac{Q_t}{Q_{t-1}} - m^* \right] \right\} + \exp \left\{ -\alpha_2 \left[\frac{Q_t}{Q_{t-1}} - m^* \right] \right\} - 2 \right] \quad (21)$$

where $Q_t = M_t - d_t$. The household's problem then becomes

$$\begin{aligned} \max_{\{C_t, h_t, M_{t+1}, Q_{t+1}\}} \quad & \mathbb{E}_0 \left[\sum_{t=1}^{\infty} \beta^t [(1 - \phi) \ln c_t + \phi \ln(1 - h_t - \tilde{p}_t)] \right] \\ \text{s.t.} \quad & p_t c_t \leq Q_t + w_t h_t \\ & M_{t+1} = (Q_t + w_t h_t - p_t c_t) + RH_t(M_t - Q_t) + f_t + b_t \end{aligned} \quad (22)$$

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 and can be solved in the same manner.

3.1.2 Potential Misspecification and Reference Model

Both DSGE models imply that the vector of observables y_t , composed of output growth ($\Delta \ln gdp_t$) and inflation ($\Delta \ln p_t$), follow a process of the form

$$y_t = \begin{bmatrix} \Delta \ln gdp_t \\ \Delta \ln p_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \Delta \ln \widehat{gdp}_t \\ \Delta \ln \widehat{p}_t \\ \ln m_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} (\gamma + \epsilon_t) \quad (23)$$

where $\ln m_{t-1}$ is a function of the η_t 's, \widehat{gdp}_t is the stochastically detrended output gdp_t/A_t , and \widehat{p}_t is the detrended price level $P_t/(M_t/A_t)$. The money growth process $\ln m_t$ is exogenous and the laws of motion for $\Delta \ln \widehat{gdp}_t$, $\Delta \ln \widehat{p}_t$ are obtained from the solutions of the DSGE models.

More generally, log-linear approximations to DSGE models can be represented in a generic state space form

$$y_t = b + Zs_t + Ue_t \quad (24)$$

$$s_t = Ts_{t-1} + Re_t \quad (25)$$

where $e_t \sim iid(0, \Sigma_e)$ is a $k \times 1$ vector of structural shocks with covariance matrix Σ_e , y_t is the $n \times 1$ vector of observables, and s_t is a $m \times 1$ vector of state variables. The system matrices b , Z , U , T , and R are functions of the structural parameters θ . Let $y_{0,t} = ZT^t s_0$, $B_h = ZT^h R$ for $h \geq 0$, and $B_0 = ZR + U$. Equations (24) and (25) lead to the following MA representation for y_t :

$$y_t = y_{0,t}(\theta) + b(\theta) + \sum_{h=0}^{t-1} B_h(\theta) e_{t-h} \quad (26)$$

The random variable $y_{0,t}(\theta)$ is determined by the initialization of the state vector s_0 .

There are two dimensions in which DSGE models are potentially misspecified. First, if the number of structural disturbances k is less than the number of observables then the conditional distribution of y_t generated by the model is singular, meaning that some linear combinations of observables should be perfectly correlated. Such singularities are usually rejected with a short span of time series data. Nevertheless, the proposed model evaluation framework remains applicable despite the presence of such singularities. The singularities will lead to the degenerate case in which the structural models have posterior probability zero. Moreover, it is not possible to compute posterior densities for the parameters of the DSGE model. The prior densities can be used instead.

Various approaches have been employed to remove the counterfactual singularities from the DSGE models prior to the empirical analysis. Further structural shocks can be added to the specification of the DSGE model such that $k \geq n$, e.g. Leeper and Sims (1994), one can introduce measurement errors, e.g. Altug (1989), or construct a hybrid model as in Ireland (1998). Although we prefer the first modeling approach and consider two models for which $k = n$, the framework remains applicable and useful in the other cases.

Second, since the vector of structural parameters θ is generally of low dimension, strong restrictions are imposed on the moving average representation, the trend component, and the short-run dynamics of y_t . These restrictions can cause poor statistical fit and forecasting performance.

To cope with the potential misspecification of the structural models a VAR(p) is considered as a reference model:

$$y_t = C_0 + C_1 y_{t-1} + \dots + C_p y_{t-p} + G \nu_t \quad (27)$$

where $\nu_t = [\nu_{1,t}, \nu_{2,t}] \sim iid(0, I_{2 \times 2})$ with diagonal covariance matrix Σ_ν . We will interpret $\nu_{1,t}$ as standardized technology shock $\epsilon_t/\sigma_\epsilon$ and $\nu_{2,t}$ as money growth shock η_t/σ_η . Due to a larger number of parameters the VAR imposes fewer restrictions on the MA representation of y_t .

The VAR specified in Equation (27) is not identified. The lack of identification does not pose a problem for the calculations of moments and correlations of y_t . However, further assumptions are needed to derive posterior distributions for money growth impulse response functions. The two DSGE models considered in this application imply long-run neutrality of money growth shocks, that is, the long-run level of output is determined by the technology shocks ϵ_t only. The neutrality restriction will be imposed on the VAR and enable identification via the Blanchard and Quah (1989) decomposition. Nason and Cogley (1994) provide further details.

3.2 Likelihoods, Priors and Computational Issues

The evaluation procedure requires the computation of posterior distributions for the model parameters θ_i , posterior model probabilities $\pi_{i,T}$, and the corresponding posterior predictive densities for the vector of characteristics φ . In general, it is intractable to find closed form solutions for these quantities. However, Bayesian simulation techniques can be used to obtain draws from the various posterior densities and to approximate posterior moments. If the DSGE models are singular then their posterior probabilities are zero and the computations have to be carried out only for the reference model.

3.2.1 Reference Model

The VAR(p), specified in Equation (27) can be rewritten as

$$y'_t = x_t C' + u'_t, \quad u'_t \sim \mathcal{N}(0, \Sigma) \quad (28)$$

where $x_t = [1, y'_{t-1}, \dots, y'_{t-p}]$, $C = [C_0, \dots, C_p]'$, and $u_t = G\nu_t$. The number of regressors is $k = 1 + np$. Let X be the $T \times k$ matrix with rows x_t . The generic parameter vector θ_* is composed of the non-redundant elements of C and Σ .

An improper prior of the form $p(C, \Sigma) \propto |\Sigma|^{-(n+1)/2}$ is chosen for the parameters of the reference model. As a consequence, the marginal density of the data $p(Y_T | \mathcal{M}_i)$ is also improper. This leaves density ratios and posterior probabilities essentially undetermined. A practical solution to this problem is to condition on a training sample $Y_0 = (y_{-\tau}, \dots, y_0)$ and to compute posterior probabilities based on the densities $p(Y_T | Y_0, \mathcal{M}_i)$.⁴ The conditional data density of the VAR can be expressed as

$$p(Y_T | Y_0, \mathcal{M}_*) = \sum_{t=1}^T \int p(y_t | C, \Sigma, Y_{t-1}, Y_0, \mathcal{M}_*) p(C, \Sigma | Y_{t-1}, Y_0, \mathcal{M}_*) d(C, \Sigma) \quad (29)$$

It is well known that the conditional densities $p(C, \Sigma | Y_{t-1}, Y_0, \mathcal{M}_*)$ are proper if the training sample is long enough, namely, if $\tau \geq p + k + n$. The first p observations are used to initialize lags. Therefore, $p(Y_T | Y_0, \mathcal{M}_*)$ is a proper density that integrates to unity.

Under the multivariate normal model, the predictive density $p(y_{t+1} | Y_t, Y_0, \mathcal{M}_*)$ has the shape of a t -distribution (see Zellner, 1971), that is,

$$p(y_{t+1} | Y_t, Y_0, \mathcal{M}_*) = \frac{\nu^{\nu/2} \Gamma((\nu + n)/2) |H_t|^{-1/2}}{\pi^{n/2} \Gamma(\nu/2)} \times \left[\nu + (y_{t+1} - \hat{y}_{t+1|t})' H_t^{-1} (y_{t+1} - \hat{y}_{t+1|t}) \right]^{-(n+\nu)/2} \quad (30)$$

with degrees of freedom $\nu = t - k - n + 1$, and

$$\begin{aligned} \hat{y}_{t+1|t} &= \hat{C}'_t x_{t+1} \\ \hat{C}_t &= (X'_t X_t)^{-1} X'_t Y_t \\ \hat{\Sigma}_t &= \frac{1}{t} (Y_t - X_t \hat{C}_t)' (Y_t - X_t \hat{C}_t) \\ H_t &= (t/\nu) (1 + x'_{t+1} (X'_t X_t)^{-1} x_{t+1}) \hat{\Sigma}_t \end{aligned}$$

The analysis of $p(Y_T | Y_0, \mathcal{M}_*)$ for the reference model is straightforward. Draws $(C, \Sigma)_{(s)}$, $s = 1, \dots, n_{sim}$ from the posterior distribution of the VAR parameters

⁴A discussion of training samples can be found, e.g., in Kass and Raftery (1995).

can be easily generated because the density conveniently factorizes as

$$p(C, \Sigma | Y_T, Y_0, \mathcal{M}_*) = p(\Sigma | Y_T, Y_0, \mathcal{M}_*) p(C | \Sigma, Y_T, Y_0, \mathcal{M}_*) \quad (31)$$

The density $p(\Sigma | Y_T, Y_0, \mathcal{M}_*)$ has the shape of an inverted Wishart distribution with $T - k$ degrees of freedom and parameter matrix $H_T = T\hat{\Sigma}_T$. Conditional on Σ , the posterior distribution of the coefficients $vec(C)$ is multivariate normal with mean $vec(\hat{C})$ and covariance matrix $\Sigma \otimes (X_T' X_T)^{-1}$. The draws from the posterior distribution of φ are obtained by calculating the implied autocovariance matrices $\Gamma_y(h)$ ⁵ and impulse response functions $\partial y_{t+h} / \partial \eta_t$ for each $(C, \Sigma)_{(s)}$. Posterior moments of the form $\int g(\varphi) p(\varphi | Y_T, Y_0, \mathcal{M}_i) d\varphi$ are approximated by sample means $\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} g(\varphi^{(s)})$, where $\varphi^{(s)}$ is a draw from $p(\varphi | Y_T, Y_0, \mathcal{M}_i)$.

3.2.2 DSGE Models

To cast the structural models \mathcal{M}_1 and \mathcal{M}_2 into state space form (Equations 24, 25) define the $2m \times 1$ state vector $s_t = [z_t, z_{t-1}]$. The vector z_t denotes the percentage deviations of detrended DSGE model variable from their steady state values. The measurement equation is based on Equation (23) and the transition equation is derived from Equation (20). The system matrices are

$$T = \begin{bmatrix} \tilde{T} & 0_{m \times m} \\ I_{m \times m} & 0_{m \times m} \end{bmatrix} \quad R = \begin{bmatrix} \tilde{R} \\ 0_{m \times 2} \end{bmatrix} \quad b = \begin{bmatrix} \gamma \\ \ln m^* - \gamma \end{bmatrix} \quad U = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$$

Moreover, Z is a $2m \times 2$ matrix that links the state vector s_t to the observables $\Delta \ln gdp_t$ and $\Delta \ln p_t$. The vector θ_i of structural model parameters is

$$\theta_i = [\alpha, \beta, \gamma, m^*, \rho, \phi, \delta, \sigma_\epsilon, \sigma_\eta, \alpha_1, \alpha_2]', \quad i = 1, 2$$

The portfolio adjustment cost parameters α_1 and α_2 have no effect on the standard CIA model \mathcal{M}_1 . The system matrices of the state space representation are non-linear functions of θ_i . Under the assumption that the structural shocks $e_t = [\epsilon_t, \eta_t]$

⁵With some posterior probability $\epsilon > 0$, the vector autoregression in non-stationary and the unconditional moments of output growth and inflation do not exist. In the rare event that a draw of C implies non-stationarity, it is discarded. This corresponds to a prior $p(C, \Sigma)$ that is zero in non-stationary regions of the parameter space.

are normally distributed the unconditional likelihood function $p(Y_T|\theta_i, \mathcal{M}_i)$ can be computed with the Kalman Filter algorithm. To make the analysis compatible to the treatment of the reference model we condition on the training sample Y_0 .

$$p(Y_T|Y_0, \theta_i, \mathcal{M}_i) = p(Y_T, Y_0|\theta_i, \mathcal{M}_i)/p(Y_0|\theta_i, \mathcal{M}_i) \quad i = 1, 2 \quad (32)$$

A posteriori inference will be based upon the conditional likelihood $p(Y_T|Y_0, \theta_i, \mathcal{M}_i)$.

We will now specify a prior distribution for θ_i . A common approach in the calibration literature is to evaluate models based on parameter values are regarded as economically plausible. Such values are obtained by matching steady state characteristics of the DSGE model to first moment properties of the time series data, from micro econometric studies with cross sectional data, or by pure introspection. This feature of the calibration approach can be interpreted as a prior distribution that concentrates on a single point of the parameter space. Following previous Bayesian DSGE model evaluation procedures we relax the tightness of the prior and consider non-degenerate distributions.

The marginal prior densities that are used in the empirical application are summarized in Table 1. The shapes of the densities are chosen to match the domain of the structural parameters. The prior means correspond to the values in Nason and Cogley's (1994) study and are consistent with the values that appear throughout the literature on monetary DSGE models. The prior densities for α_1 and α_2 are only used for the liquidity model with portfolio adjustment costs. As in Canova (1994) and Dejong *et al.* (1996, 1997), 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.

To document the impact of the prior distribution we will consider two specifications that differ in the relative weight placed on the prior. The standard errors for the parameters α , β , ρ , ϕ , δ , α_1 , and α_2 are larger for Prior 2 than for Prior 1. Under the more diffuse Prior 2 the posterior will more closely resemble the likelihood function than under Prior 1. The Bayes estimation can be interpreted as follows:

find values for the structural parameters such that the DSGE models fit the data in a likelihood sense, without deviating too far from parameter values that are economically plausible and consistent with additional information that might be available.

Table about here

Table 1: Prior Distribution for the Parameters of the DSGE Models.

The computations to obtain draws from the various posterior distributions involve several steps. Conditional on the actual data Y_T, Y_0 , a set of parameter values θ_i , the Kalman Filter is used to evaluate the posterior density up to a constant

$$p(\theta_i|Y_T, Y_0, \mathcal{M}_i) \propto p(Y_T|Y_0, \theta_i, \mathcal{M}_i)p(\theta_i|\mathcal{M}_i) \quad (33)$$

A numerical optimization routine is used to compute the mode $\tilde{\theta}_i$ of the posterior density. Let $\tilde{\Sigma}_i$ be

$$\tilde{\Sigma}_i = \left[-\frac{\partial^2}{\partial \theta_i \partial \theta_i'} \ln p(Y_T|Y_0, \theta_i, \mathcal{M}_i)p(\theta_i|\mathcal{M}_i) \right]_{\theta_i=\tilde{\theta}_i} \quad (34)$$

the Hessian of the log posterior density evaluated at the mode.

The Metropolis Algorithm is used to generate $n_{sim} = 30,000$ draws $\theta_i^{(s)}$ from the posterior distribution. At each iteration s a candidate parameter vector ϑ_i is drawn from a jumping distribution $J_s(\vartheta_i|\theta_i^{(s-1)})$ and the density ratio

$$r = \frac{p(Y_T|Y_0, \vartheta_i, \mathcal{M}_i)p(\vartheta_i|\mathcal{M}_i)}{p(Y_T|Y_0, \theta_i^{(s-1)}, \mathcal{M}_i)p(\theta_i^{(s-1)}|\mathcal{M}_i)} \quad (35)$$

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. It can be shown that the sequence $\{\theta_i^{(s)}\}$ converges to the target posterior distribution (Gelman *et al.* 1995 and references cited therein). 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 the standard CIA model, and $c = 0.1$ for the model with liquidity effect. The rejection rates were about 37 percent, respectively. We found

that $\tilde{c} = 2.4/\sqrt{d_i}$, where d_i is the number of parameters 9 and 11, respectively, recommended in Gelman *et al.* (1995) leads to too many rejections. Based on the draws $\{\theta_i^{(s)}\}$ one can calculate the implied autocovariances, autocorrelations, and impulse response functions $\{\varphi^{(s)}\}$ for the DSGE models.

Finally, the marginal data densities have to be computed at the observed Y_T, Y_0 .

$$p(Y_T|Y_0, \mathcal{M}_i) = \int p(Y_T|Y_0, \theta_i, \mathcal{M}_i)p(\theta_i|\mathcal{M}_i)d\theta_i \quad (36)$$

For the structural models this integral cannot be evaluated analytically. There are several approximation methods available. An overview of different approximation methods can be found in Kass and Raftery (1995). We will use a Laplace approximation

$$\tilde{p}(Y_T|Y_0, \mathcal{M}_i) = (2\pi)^{d/2}|\tilde{\Sigma}_i|^{-1/2}p(Y_T|Y_0, \tilde{\theta}_i, \mathcal{M}_i)p(\tilde{\theta}_i|Y_0, \mathcal{M}_i) \quad (37)$$

based on a log-quadratic expansion around the posterior mode.

3.3 Empirical Results

The two DSGE models and the VAR are fitted to quarterly U.S. data from 1950:I to 1997:IV. The data were extracted from the DRI/Citibase 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.

3.3.1 Posterior Distributions and Model Probabilities

The first step consists of the analysis of the overall posterior density $p(\varphi|Y_T, Y_0)$ given in Equation (2). Table 2 summarizes posterior means and standard errors for the structural models \mathcal{M}_1 and \mathcal{M}_2 . The posterior moments are calculated from the output of the Metropolis algorithm. For the parameters ϕ , α_1 , and α_2 the likelihood function is essentially uninformative. The moments of the marginal posterior distributions for these parameters are very similar to the prior moments. For other

Range	Density	Mean	Std.Error (1)	Parametrization (1)	Std.Error (2)	Parametrization (2)
α	$Beta(a, b)$	0.345	0.050	30.84	0.060	21.31
β	$Beta(a, b)$	0.933	0.005	275.1	0.010	68.03
γ	$\mathcal{N}(\mu, \sigma^2)$	0.003	0.001	0.003		(same)
m^*	$Gamma(a, b)$	1.011	0.002	2.5E6		(same)
ρ	$Beta(a, b)$	0.728	0.050	56.93	0.100	13.68
ϕ	$Beta(a, b)$	0.773	0.050	53.48	0.100	12.79
δ	$Beta(a, b)$	0.022	0.005	18.91	0.010	4.711
σ_ϵ	$InvGamma(s, \nu)$	0.035	∞	0.020		(same)
σ_η	$InvGamma(s, \nu)$	0.009	∞	0.005		(same)
$\sqrt{\alpha_1}$	$Gamma(a, b)$	0.007	0.001	49.00	0.003	5.444
$\sqrt{\alpha_2}$	$\mathcal{N}(\mu, \sigma^2)$	31.62	2.000	31.62	5.000	31.62
				2.000		5.000

Table 1: Prior Distribution for the Parameters of the DSGE Models. The columns “Parametrization” refer to the values of (a, b) , (μ, σ) , and (s, ν) , respectively.

parameters, such as the capital share α , the discount factor β , the depreciation rate δ , and the persistence of the money supply shock, the posterior means are very different from the prior means. A seemingly small increase in the prior variance from Prior 1 to Prior 2 leads to a substantial change in the posterior mean. To gain a better understanding of this effect, consider a Gaussian linear regression model with conjugate prior. The posterior mean is a weighted average of the prior mean and the maximum likelihood estimate. The relative weight of the prior decreases as the variance of the prior distribution increases. If we use Prior 2 instead of Prior 1 in our analysis the posterior mode moves more closely toward the maximum of the likelihood function.

[Table about here](#)

Table 2: Posterior Moments for the Parameters of the DSGE Models.

The estimates of the capital share parameter α are greater than 0.7, and, thus, larger than estimates reported elsewhere in the literature. The estimates of the discount factor β are about 0.97 for the standard CIA model and 0.95 for the liquidity model. The low estimates of β imply a counterfactual risk free rate of more than 3 percent per quarter. Under Prior 2 the estimate of the depreciation rate δ is 0.29 percent for \mathcal{M}_1 and 0.62 percent for \mathcal{M}_2 which leads to a low investment to capital ratio along the balanced growth path of the model economies.

For the interpretation of the estimates it is important to note that most GMM and maximum-likelihood estimates reported elsewhere in the literature are based on long-run averages and low frequency characteristics of aggregate data, such as output, consumption, investment and capital stock. In our empirical application, the models are fitted to differenced data and the likelihood estimates are dominated by short-run dynamics. It is well documented, e.g. Cogley and Nason (1995) and Rotemberg and Woodford (1996) that standard business cycle models have difficulties to replicate the stochastic behavior of output growth. Thus, it is not surprising that the transformation of the data from levels to differences leads to a different set of estimates. This sensitivity to the choice of criterion function is an indication of

model misspecification. The model has no “true” parameter values. Instead there are only “pseudo-true” parameter values that depend on the criterion function of the estimators and can be thought of as their hypothetical limit as the sample size tends to infinity. The use of prior distributions provides a coherent way of imposing that large deviations of the parameter estimates from values that are consistent with other information is penalized.

Table about here

Table 3: Approximations to the Marginal Data Densities.

Table 3 contains likelihood statistics and approximations of the marginal data densities $p(Y_T|\mathcal{M}_i)$. Not surprisingly, the reference model attains the highest likelihood, followed by the liquidity model \mathcal{M}_2 . However, it is not meaningful to rank models according to maximum likelihood values. The Schwarz approximation (BIC) to the marginal data density suggests to penalize the likelihood values by the model dimensionality d_i according to the term $(d_i/2) \ln T$. The liquidity model is more strongly penalized than the standard CIA model because it has the additional parameters α_1 and α_2 . Row 4 of Table 3 lists the BIC values for the three models. The information criterion can be converted into a posterior odds scale by the transformation $\exp\{BIC_i - BIC_j\}$. Our results imply that the BIC odds of the VAR versus the liquidity model are approximately 90:1, and the odds of the standard CIA model versus the liquidity model are 1:8.5.

Exact posterior probabilities depend on the marginal data density

$$p(Y_T|Y_0, \mathcal{M}_i) = \int p(Y_T|Y_0, \theta_i, \mathcal{M}_i)p(\theta_i|Y_T)d\theta_i \quad (38)$$

As pointed out in the previous section, the marginal density cannot be calculated analytically for the two structural models. Instead we report a Laplace approximation. According to the approximation the liquidity model is preferred to the standard CIA model for Prior 1 and Prior 2, respectively. Surprisingly, the liquidity model with Prior 2 is also slightly preferred to the VAR (5:1). Unfortunately, the

	Model $\mathcal{M}_{1(1)}$		Model $\mathcal{M}_{1(2)}$		Model $\mathcal{M}_{2(1)}$		Model $\mathcal{M}_{2(2)}$	
	mean	(stdd)	mean	(stdd)	mean	(stdd)	mean	(stdd)
α	0.7232	(0.0322)	0.7935	(0.0272)	0.7180	(0.0398)	0.7936	(0.0289)
β	0.9736	(0.0064)	0.9786	(0.0069)	0.9506	(0.0119)	0.9418	(0.0189)
γ	0.0023	(0.0009)	0.0019	(0.0009)	0.0027	(0.0009)	0.0026	(0.0009)
m^*	1.0137	(0.0011)	1.0128	(0.0016)	1.0133	(0.0011)	1.0125	(0.0018)
ρ	0.8621	(0.0218)	0.9310	(0.0139)	0.8616	(0.0232)	0.9233	(0.0187)
ϕ	0.7727	(0.0472)	0.7713	(0.1068)	0.7829	(0.0481)	0.7918	(0.0865)
δ	0.0102	(0.0025)	0.0029	(0.0015)	0.0129	(0.0031)	0.0062	(0.0028)
σ_ϵ	0.0309	(0.0041)	0.0425	(0.0063)	0.0336	(0.0052)	0.0460	(0.0075)
σ_η	0.0023	(0.0002)	0.0020	(0.0001)	0.0026	(0.0002)	0.0024	(0.0003)
$\sqrt{\alpha_1}$	N/A		N/A		0.0070	(0.0011)	0.0072	(0.0028)
$\sqrt{\alpha_2}$	N/A		N/A		31.987	(1.9696)	31.124	(4.8127)

Table 2: Posterior Moments for the Parameters of the DSGE Models. Posterior moments are based on output from Metropolis algorithm. The label $\mathcal{M}_{i(j)}$ refers to Model \mathcal{M}_i with prior distribution j .

	$\mathcal{M}_{1(1)}$	$\mathcal{M}_{1(2)}$	$\mathcal{M}_{2(1)}$	$\mathcal{M}_{2(2)}$	VAR(4)
Posterior density $p(Y_{\mathcal{T}} Y_0, \tilde{\theta}_i, \mathcal{M}_i)p(\tilde{\theta}_i \mathcal{M}_i)$	1289.13	1313.84	1296.68	1320.13	1318.78
Likelihood at mode $p(Y_{\mathcal{T}} Y_0, \tilde{\theta}_i, \mathcal{M}_i)$	1296.49	1324.47	1305.81	1327.25	1352.85
Maximum likelihood $p(Y_{\mathcal{T}} Y_0, \hat{\theta}_{i,mle}, \mathcal{M}_i)$	1338.15		1345.46		1373.24
BIC (variance param. not penalized)	1320.03		1322.17		1326.66
Exact marginal density $p(Y_{\mathcal{T}} Y_0, \mathcal{M}_i)$	N/A		N/A		1275.90
Laplace approx. of marginal density	1246.61	1271.41	1251.25	1277.69	1276.00

Table 3: Approximations to the marginal data densities of structural models and the reference model. The label $\mathcal{M}_{i(j)}$ refers to Model \mathcal{M}_i with prior distribution j .

good statistical fit of model \mathcal{M}_2 comes at the price of unusual structural parameter estimates.

The trade-off between economically plausible parameter values and statistical fit deserves closer inspection. Consider the expression

$$p(Y_T|Y_0, \theta_i, \mathcal{M}_i)p(\theta_i|Y_T) \quad (39)$$

The informative prior distribution pulls the posterior mode $\tilde{\theta}_i$ away from the mode of the likelihood function $\hat{\theta}_{i,mle}$ toward the prior mode. If the prior distribution is very concentrated (Prior 1), the difference between the likelihood function evaluated at its maximum and at the posterior mode can be substantial, in our case roughly 40. The tight prior leads ultimately to a low value of the marginal density. Thus, somebody who strongly believes that the parameter values should not deviate much from the mean specified in Table 1 (Prior 1) will deduce that the structural models have essentially zero posterior probability. A person who is less certain about the parameter values *a priori* (Prior 2) can conclude that the liquidity model fits output growth and inflation data quite well, compared to a VAR(4).

A simple Bayesian model determination stops with the calculation of posterior odds. The posterior probabilities of DSGE models, however, may often be small, and macroeconomists are interested in evaluating such models according to a different metric than the one implied by likelihood based econometric procedures. In our evaluation framework the calculation posterior distributions and model probabilities is an intermediate step, that is necessary to obtain the overall posterior distribution of population moments and impulse response functions φ .

Although according to the Laplace approximation the liquidity model with Prior 2 has a higher posterior probability than the VAR(4), a qualification of the result is necessary. It is possible to find a plausible prior distribution for the VAR parameters, e.g., Minnesota Prior with hyperparameters integrated out as in Schorfheide (1998), such that the posterior probability of the reference model is greater than the posterior probabilities of all structural models. However, many evaluations of DSGE models are conducted based on a very diffuse reference model, such as a

fourth or sixth order VAR (Nason and Cogley 1994, 1995), or nonparametric estimates (Diebold *et al.* 1998). Our application suggests it is not always justified to neglect the predictions of the best fitting structural model. The subsequent calculations will be conducted with a posterior probability of 80 percent for the VAR (4) and 20 percent for the liquidity model with Prior 2, $\mathcal{M}_{2(2)}$.

3.3.2 Evaluation Based on Bayes Estimation of DSGE Models

In this section it is demonstrated how the model evaluation framework can be used to evaluate models based on their ability to generate realistic patterns of co-movements among macroeconomic aggregates or realistic impulse responses to structural shocks. At first parameter weight functions $f_i(\theta)$ that appear in Equation (4) are specified. The standard calibration approach uses degenerate weight functions that concentrate all their mass on parameter values such as the posterior means given in Table 1. Dejong *et al.* (1996) and Geweke (1999) suggest to replace the degenerate parameter weights by prior densities. Since these authors consider DSGE models that generate a singular probability distribution for the data there is no posterior parameter distribution. In our application posterior densities $p(\theta_i|Y_T, Y_0, \mathcal{M}_{i(j)})$ are available and can be used as weight functions $f_{i(j)}$. The subscript $i(j)$ refers to model \mathcal{M}_i with Prior j . Define

$$\hat{\varphi}_{b,ij} = \operatorname{argmin}_{\hat{\varphi} \in \mathbf{R}^m} \int L(\varphi, \hat{\varphi}) p(\varphi|Y_T, \mathcal{M}_{i(j)}) d\varphi \quad (40)$$

Under L_q and L_χ^2 loss the optimal predictor $\hat{\varphi}_{b,i}$ is the posterior mean, for the L_p loss it is the posterior mode. Below we report the overall posterior expected prediction loss of $\hat{\varphi}_{b,ij}$. Determining the model $\mathcal{M}_{i(j)}$ that minimizes this prediction risk is equivalent to solving the minimization problem in Equation (6) with respect to λ . We restrict the model weights to be zero or one. The fit of mixtures of structural DSGE models is not explored in this paper.

The calculations of expected L_p losses and the Kullback-Leibler distances involve estimates of posterior densities $p(\varphi|Y_T, Y_0, \mathcal{M}_i)$. We use non-parametric Kernel

estimates of the form

$$\hat{p}(\varphi|Y_T, Y_0, \mathcal{M}_i) = \frac{1}{n_{sim} h^m} \sum_{s=1}^{n_{sim}} K((\varphi - \varphi^{(s)})/h) \quad (41)$$

where $K(x) = (2\pi)^{-m/2} \exp(-x'x/2)$ (cf. Silverman, 1986). It can be shown that the optimal bandwidth h' for the smoothing of normally distributed data with unit variance is

$$h' = \left(\frac{4}{m+2}\right)^{1/(m+4)} n_s^{-1/(m+4)}$$

which we scale by the average marginal variance of the components of φ .⁶

Table 4 and Figure 1 summarize the evaluation results with respect to correlation patterns between output growth and inflation. Results for individual moments as well as jointly for $\sigma(\Delta \ln gdp_t)$, $\sigma(\Delta \ln p_t)$, $corr(\Delta \ln gdp_t, \Delta \ln p_t)$ and the four first order autocorrelations are provided. We report the overall posterior means and standard deviations, posterior mean predictions of $\mathcal{M}_{i(j)}$, as well as expected L_p and L_{χ^2} prediction losses. The L_q prediction risks for an identity weight matrix can be obtained as squared difference between overall posterior mean and the posterior mean predictions of models $\mathcal{M}_{i(j)}$. The Kullback-Leibler distance serves as a general measure of discrepancy between the $p(\varphi|Y_T, Y_0, \mathcal{M}_{i(j)})$'s and the overall posterior $p(\varphi|Y_T, Y_0)$. Figure 1 depicts posterior density plots for three moments, Gaussian approximations to these posterior densities, and posterior mean predictions from models $\mathcal{M}_{i(j)}$.

[Table about here](#)

Table 4: Model Evaluation Statistics for Second Moments.

⁶All our posterior calculations are subject to numerical approximation errors which can, at least in principle, be made arbitrarily small by increasing the number of simulation draws. For practical purposes, the number of draws might often be limited to a few thousands and asymptotic distribution theory could be used to calculate approximation errors, e.g., Geweke (1992). In this paper, however, we do not pursue the computation of numerical errors.

It was argued in Section 2 that informal moment comparisons or the evaluation of the model predictions under a quadratic loss function is equivalent to ranking the models according to the difference between posterior mean predictions $\hat{\varphi}_{b,i(j)}$ and the overall posterior mean. For $\sigma(\Delta \ln gdp_t)$ the difference is minimized by model $\mathcal{M}_{1(1)}$. The other criteria support the conclusion that $\mathcal{M}_{1(1)}$ is the preferred specification.

Now consider $\sigma(\Delta \ln p_t)$. The ranking according to the quadratic loss function is $\mathcal{M}_{2(1)}$, $\mathcal{M}_{1(2)}$, $\mathcal{M}_{1(1)}$ and $\mathcal{M}_{2(2)}$. However, the difference between the posterior mean predictions of the first two models is only 0.0001. To avoid the difficulty of judging the magnitude of this difference it has to be normalized. In a GMM-framework this is achieved by p -values for tests based on overidentifying moment restrictions. The discrepancy is normalized by its sampling variance under the hypothesis that the structural model under consideration is actually true. We argued above that this approach is conceptually unappealing. Comparisons across models are difficult to interpret if there is a significant chance that all structural models under consideration are misspecified. Our L_{χ^2} loss function essentially standardizes the discrepancy by the variance of the overall posterior distribution. The expected L_{χ^2} is always between zero and one and can be loosely interpreted as a p -value. This statistic suggest that $\mathcal{M}_{2(1)}$ is clearly preferable to $\mathcal{M}_{1(2)}$. In Figure 1 the expected L_{χ^2} loss corresponds to the area under the Gaussian approximation of the posterior density for which $p(\varphi|Y_T)$ is larger than $p(\hat{\varphi}_{b,i(j)}|Y_T)$.

According to the L_p loss the ranking of models $\mathcal{M}_{1(2)}$ and $\mathcal{M}_{2(1)}$ is reversed. Moreover, model $\mathcal{M}_{1(1)}$ now appears better than $\mathcal{M}_{2(1)}$ and almost as good as $\mathcal{M}_{1(2)}$. It is apparent from the second plot of Figure 1 that the actual posterior distribution of $\sigma(\Delta \ln p_t)$ is skewed. The L_p loss indicates that the posterior modes of $\mathcal{M}_{1(1)}$ and $\mathcal{M}_{1(2)}$ are closer to the mode of the overall posterior distribution than the posterior mode predictions of $\mathcal{M}_{2(1)}$ and $\mathcal{M}_{2(2)}$.

The Kullback-Leibler distance calculations for $\sigma(\Delta \ln p_t)$ favor model $\mathcal{M}_{2(2)}$ which is ranked last according the other loss functions. Table 4 indicates that conditional on the other three specifications of the structural models the posterior distributions of $\sigma(\Delta \ln p_t)$ are sharply peaked around the mode. The standard deviations are

	Posterior	$\mathcal{M}_{1(1)}$	$\mathcal{M}_{1(2)}$	$\mathcal{M}_{2(1)}$	$\mathcal{M}_{2(2)}$
$\sigma(\Delta \ln gdp_t)$	Mean	0.0104	0.0103	0.0098	0.0101
	(Stdd)	(0.0006)	(0.0006)	(0.0005)	(0.0005)
	L_p	0.0917	0.1307	0.6957	0.3295
	L_{χ^2}	0.1101	0.1859	0.6737	0.4620
KL	0.3241	0.5049	2.0867	0.7873	
$\sigma(\Delta \ln p_t)$	Mean	0.0067	0.0070	0.0077	0.0085
	(Stdd)	(0.0004)	(0.0006)	(0.0006)	(0.0011)
	L_p	0.1478	0.1377	0.5408	0.7012
	L_{χ^2}	0.3911	0.2233	0.1233	0.5544
KL	62.399	13.342	11.435	2.0752	
$corr(\Delta \ln gdp_t, \Delta \ln p_t)$	Mean	-0.5216	-0.3813	-0.5346	-0.4248
	(Stdd)	(0.0341)	(0.0414)	(0.0365)	(0.0548)
	L_p	0.9716	0.7839	0.9848	0.6468
	L_{χ^2}	0.9834	0.6831	0.9903	0.8228
KL	332.54	93.480	299.95	25.891	
Joint, lag 0 and 1	L_p	0.6500	0.3548	0.5663	0.0000
	L_{χ^2}	0.9875	0.9772	0.9616	0.2290
	KL	423.38	338.03	228.59	46.633

Table 4: Model Evaluation Statistics for Second Moments. First three panels contain univariate statistics. Last Panel contains multivariate statistics jointly for $\sigma(\Delta \ln gdp_t)$, $\sigma(\Delta \ln p_t)$, $corr(\Delta \ln gdp_t, \Delta \ln p_t)$, and the first order autocorrelations.

0.0006 or less. Under $\mathcal{M}_{2(2)}$ this standard deviation is almost twice as large. The overall posterior distribution of $\sigma(\Delta \ln p_t)$ is also quite dispersed (standard deviation of 0.0021). The Kullback-Leibler distance not only takes into account the location of the overall posterior distribution versus the model prediction but also whether the posteriors conditional on the structural model resemble the overall posterior in terms of dispersion and shape which explains the small value for $\mathcal{M}_{2(2)}$.

The posterior distribution of $\text{corr}(\Delta \ln gdp_t, \Delta \ln p_t)$ is bimodal. Despite its low posterior probability, $\mathcal{M}_{2(2)}$ introduces a second mode into the posterior density due to a more concentrated posterior density for $\text{corr}(\Delta \ln gdp_t, \Delta \ln p_t)$ than the diffuse reference model. For this reason the expected L_p loss of $\mathcal{M}_{2(2)}$ is smaller than the loss for $\mathcal{M}_{1(2)}$ despite $\hat{\varphi}_{1(2)}$ being closer than $\hat{\varphi}_{2(2)}$ to mean and mode of the overall posterior density.

The last panel of Table 4 summarizes the prediction losses for seven moments jointly. The zero L_p loss indicates that the structural models introduce highly peaked modes into the overall posterior distribution. According to L_{χ^2} loss and the Kullback-Leibler measure, the liquidity model with Prior 2 predicts the covariation patterns more accurately than the simple CIA model.

The results demonstrate that the model evaluation is potentially sensitive to the choice of loss function. A simple comparison between model prediction and posterior mean can lead to a distorted ranking and incomplete assessment.

In terms of policy implications of the two models it is more interesting to examine their predictions with respect to the effects of a discretionary change in the growth rate of money supply, as measured by the impulse responses $\{\partial \ln gdp_{t+h} / \partial \eta_t\}_{h=0}^H$ and $\{\partial \Delta \ln p_{t+h} / \partial \eta_t\}_{h=0}^H$ to a standardized money growth shock η_t . Figure 2 depicts the posterior distribution and the Bayes predictors $\hat{\varphi}_{b,i(1)}$ for the impulse response functions under Prior 1. The posterior distribution of $\{\partial \ln gdp_{t+h} / \partial \eta_t\}_{h=0}^H$ is dominated by the positive hump-shaped response of output to the expansion of money supply. The liquidity model is able to generate a positive response of output, but initially smaller and in the long-run more persistent than the one predicted by the

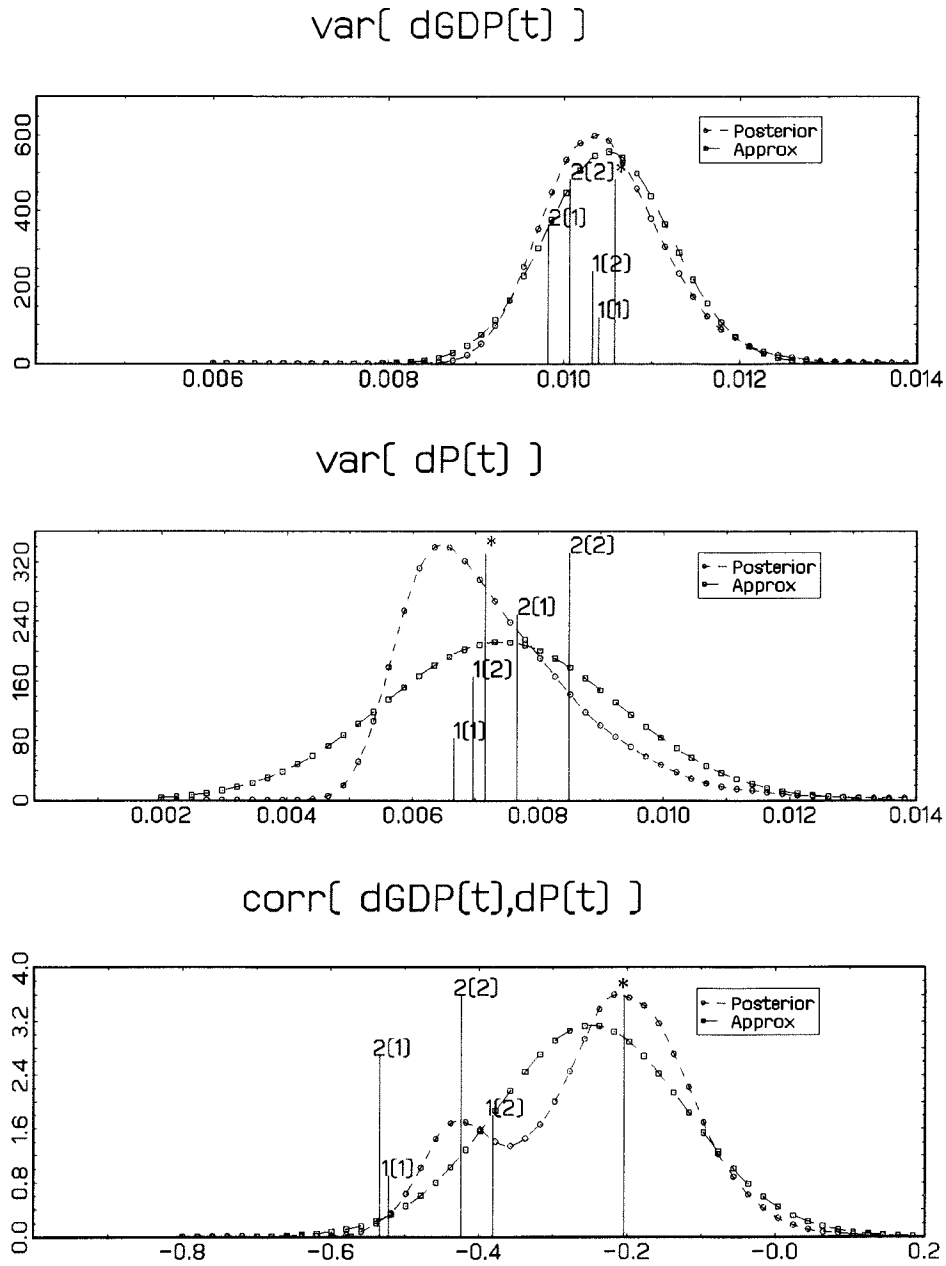


Figure 1: Posterior densities for the moments $\sigma(\Delta \ln gdp_t)$, $\sigma(\Delta \ln p_t)$, and $\text{corr}(\Delta \ln gdp_t, \Delta \ln p_t)$. *Posterior* refers to a Kernel density estimate based on the output of the Metropolis Algorithm. *Approx* refers to a Gaussian approximation with same mean and variance as the simulated moments. Vertical lines signify posterior mean predictions $\hat{\varphi}_{b,i(j)}$ for structural models \mathcal{M}_i with Prior j .

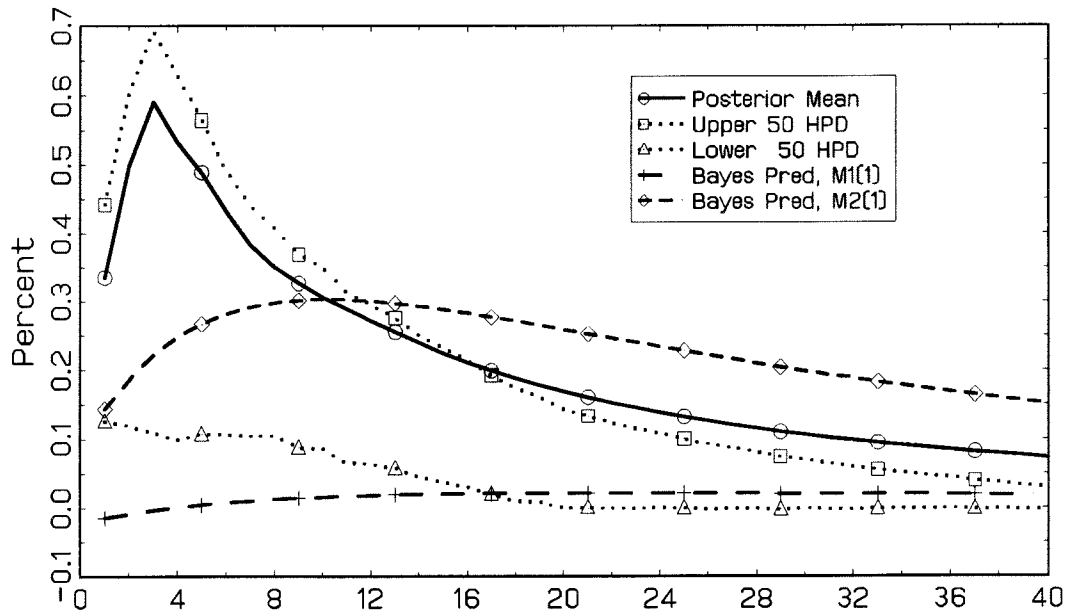
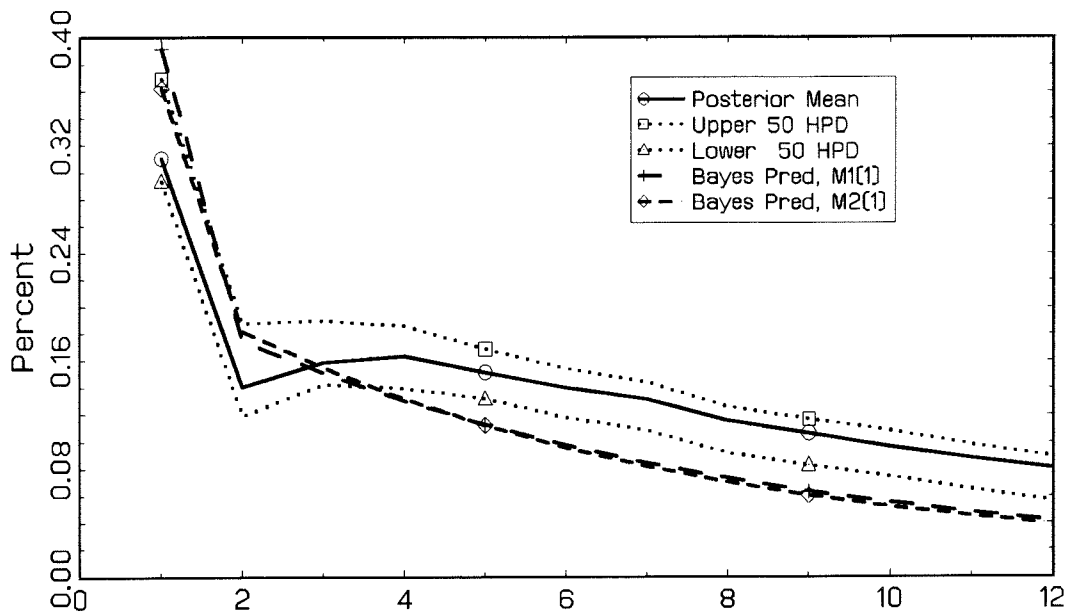
Response of Output to $\eta(t)$ Response of Inflation to $\eta(t)$ 

Figure 2: Posterior Distribution and Bayes Predictors.

VAR. The response of the standard cash-in-advance model is dominated by the expected inflation effect which causes a slight decrease of output initially. Figure 2 also depicts the shortest connected fifty percent confidence intervals. The fact that the posterior mean is at some horizons not contained in these sets indicates that the predictions of the structural model $\mathcal{M}_{2(2)}$ skew the overall posterior distribution. Both structural models predict a sudden increase in the price level which results in a substantially larger inflation rate than under \mathcal{M}_* .

3.3.3 Evaluation Based on Loss Function Estimation

In this section loss function estimates for the liquidity model \mathcal{M}_2 are computed. The loss function estimation corresponds to solving the minimization problem (6) of Section 2 for $\lambda = [0, 1]'$ and F_i being the set of point mass functions $\delta_{\{\theta_i=\theta_{i,0}\}}$, $\theta_{i,0} \in \Theta_i$. We focus on the liquidity model \mathcal{M}_2 because the standard CIA model cannot generate a positive response of output to a discretionary increase in the money growth rate. Thus, loss function estimation of \mathcal{M}_1 is not very interesting. Let $\varphi = [\varphi^{(1)}, \varphi^{(2)}]'$ where $\varphi^{(1)}$ corresponds to the impulse response sequences $\{\partial \ln gdp_{t+h}/\partial \eta_t\}_{h=0}^H$ and $\{\partial \Delta \ln p_{t+h}/\partial \eta_t\}_{h=0}^H$ and $\varphi^{(2)}$ to the structural parameters θ_2 of the liquidity model. A modified quadratic loss function of the form

$$L(\varphi, \hat{\varphi}) = (\varphi - \hat{\varphi}^{(1)})' \kappa W (\varphi - \hat{\varphi}^{(1)}) - \ln p(\hat{\varphi}^{(2)} | \mathcal{M}_2) \quad (42)$$

is used to obtain the estimates.

The weight matrix W is diagonal, discounting h -step ahead predictions by 0.99^h . The second term penalizes strong deviations from *a priori* “plausible” parametrizations of the structural model. If κ is large the loss is dominated by the accuracy of the impulse response predictions. If $\kappa = 0$, the loss is minimized by the prior mode parameters. Alternatively, $\varphi^{(2)}$ could be composed of unconditional moments to ensure that the loss function estimates do not have very unrealistic implications about correlation patterns. Instead of using $\varphi^{(1)}$ to estimate or calibrate the model parameters and $\varphi^{(2)}$ to test or evaluate the models, our approach treats all model predictions symmetrically.

Table 5 summarizes the estimation results. Figure 3 depicts the corresponding impulse response functions. For $\kappa = 1$ the estimates are similar to the prior mean. In the liquidity model 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 labor supply in response to the money growth shock. The labor-leisure trade-off determines aggregate real output in the initial period since the capital stock is also predetermined. For the prior mean values of the structural parameters the incentive for the household to increase its labor supply is not strong enough to generate such a large increase in output as suggested by the overall posterior distribution. Instead the structural model attributes a large fraction of the rise in nominal output to an increased price level.

	Prior Mean	Prior Std.Error	$\kappa = 1$	$\kappa = 1E2$	$\kappa = 1E3$	$\kappa = 1E4$
α	0.345	0.050	0.343	0.431	0.612	0.635
β	0.993	0.005	0.996	0.956	0.900	0.900
γ	0.003	0.001	0.003	0.003	0.002	1.0E-5
m^*	1.011	0.002	1.011	1.011	1.011	1.018
ρ	0.728	0.050	0.736	0.724	0.760	0.778
ϕ	0.773	0.050	0.782	0.787	0.810	0.946
δ	0.022	0.005	0.021	0.015	0.010	0.002
σ_ϵ	0.035	∞	0.016	0.016	0.016	0.016
σ_η	0.009	∞	0.004	0.005	0.004	0.004
$\sqrt{\alpha_1}$	0.007	0.001	0.007	0.006	0.004	0.003
$\sqrt{\alpha_2}$	31.62	2.000	31.62	30.15	25.81	21.00

Table 5: Loss Function Based Parameter Estimates, $\mathcal{M}_{2(1)}$.

As κ is increased and more weight is placed on matching the impulse response characteristics the estimate of the capital share parameter rises to $\hat{\alpha} = 0.431$ and the discount factor drops to $\hat{\beta} = 0.956$. This trend continues as κ is increased to 10000. Overall the propagation mechanism for the nominal shock is not strong

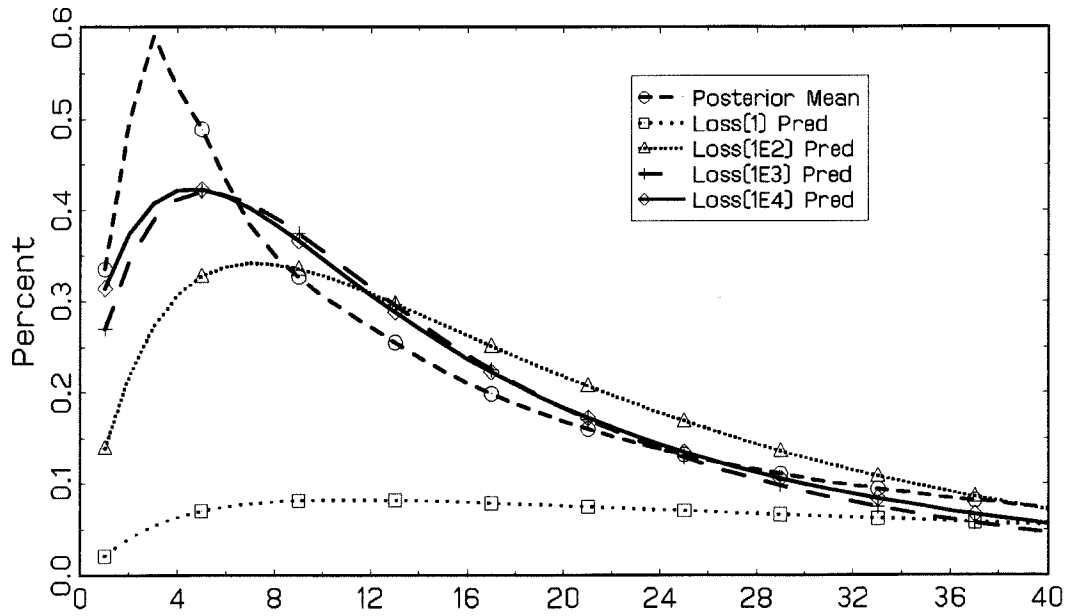
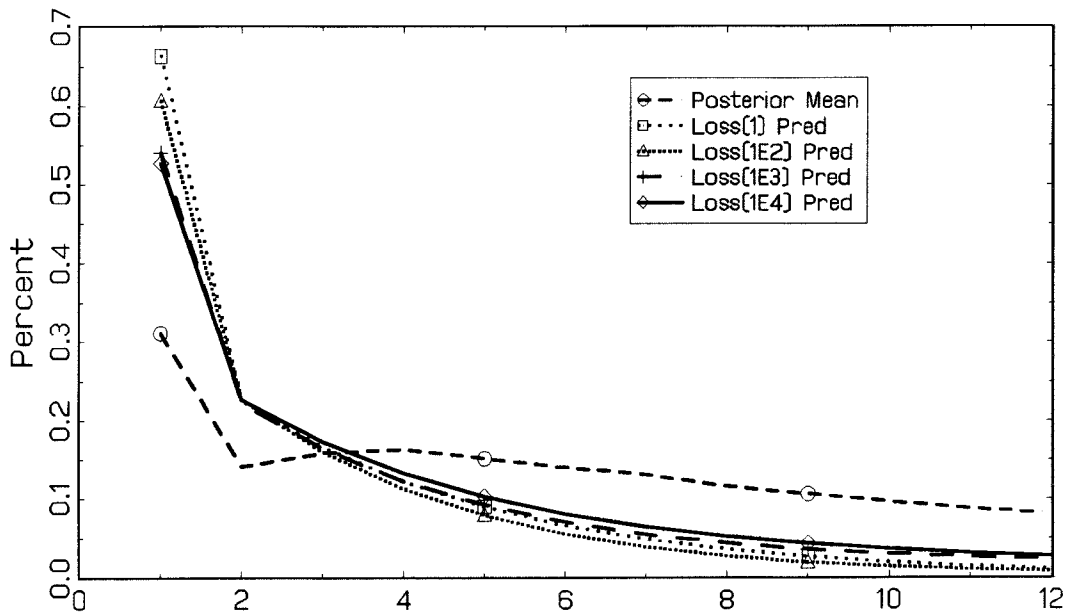
Response of Output to $\eta(t)$ Response of Inflation to $\eta(t)$ 

Figure 3: Posterior Mean and Loss Function Predictors.

enough to reproduce the posterior mean response of output.

The empirical illustration suggests that despite some overall deficiencies, the liquidity model \mathcal{M}_2 is preferable to the standard cash-in-advance model \mathcal{M}_1 in terms of posterior probabilities, its implications with respect to autocorrelations of order zero and one, and its impulse response patterns. Our assessment of the liquidity model is to some extent more favorable than Nason and Cogley's (1994) assessment. Using classical hypothesis tests, Nason and Cogley reject the null hypotheses that the impulse response functions $\{\partial \ln gdp_{t+h}/\partial \eta_t\}_{h=0}^H$ and $\{\partial \Delta \ln p_t/\partial \eta_t\}_{h=0}^H$ were generated from either of the structural models by a wide margin.

4 Conclusion

This paper introduced a unified framework for the evaluation of small DSGE models. It combines features of many existing evaluation techniques in a coherent procedure. A variety of measures are provided that are useful for comprehensive model comparisons. Under special circumstances, the evaluation procedure is similar in spirit to the popular informal moment comparisons. However, the illustrative application demonstrated that informal moment comparisons might lead to a distorted assessment and only yield an incomplete picture of relative model performance. The framework also builds a link to the traditional measures of fit for macroeconomic models, namely, their ability to track historic data and forecast future observations. This is measured by the posterior probabilities of the structural models, which determine how much weight should be given to the predictions of the DSGE models relative to the predictions of the reference model.

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 cases in which it is possible to generate predictions $p(\varphi|\mathcal{M}_*)$ from the reference model. The attractiveness of structural economic models, however, rests in their ability to generate predictions about aspects of economic activity that cannot be analyzed within the context of a reduced form

reference model. For instance, suppose φ is composed of the effects of permanent changes in the monetary policy rule. The vector autoregression cannot generate predictions with respect to such a φ .

If the posterior probability of the reference model \mathcal{M}_* were very small then one could regard the posterior mixture of structural models as “true” for practical purposes and base the decisions on predictions obtained from the \mathcal{M}_1 and \mathcal{M}_2 mixture. Of course, one has to keep in mind that there is always the possibility that despite high posterior probability, \mathcal{M}_1 and \mathcal{M}_2 do not capture the policy effects correctly. On the other hand, if the reference model has the highest posterior probability, it becomes more difficult to determine how to base decisions on \mathcal{M}_1 or \mathcal{M}_2 . Implicitly, macroeconomists seem to proceed as follows: use a vector of characteristics φ_* for which it is possible to generate predictions from the reference model, say responses to a discretionary increase in the growth rate of money supply, and weight the DSGE models according to their ability to predict φ_* . For this selection procedure to be persuasive, the vector φ_* and the loss function $L_*(\varphi_*, \hat{\varphi}_*)$ have to be carefully selected and might depend on various aspects of model fit. Choosing a φ_* and $L_*(\varphi_*, \hat{\varphi}_*)$ that is informative about the actual φ and $L(\varphi, \hat{\varphi})$ is beyond the scope of econometric analysis and the framework presented in this paper. The advantage of our framework is that it can be adapted to a variety of characteristics φ_* and loss functions $L_*(\varphi_*, \hat{\varphi}_*)$, to document in which respects the structural models \mathcal{M}_1 and \mathcal{M}_2 fit the data.

While the implementation of the framework is conceptually straightforward, the non-linear structure of the DSGE models might cause practical problems with respect to optimizations and simulation from the posterior distributions of the structural parameters. The reference model could be equipped with a hierarchical prior distribution based on the structural models and controlled by hyperparameters as in Ingram and Whiteman (1994) and Schorfheide (1998). The evaluation of L_p -losses requires non-parametric density estimation, which is likely to cause non-negligible approximation errors for high dimensional vectors φ . In this case, a careful asymptotic analysis of the numerical errors is desirable.

References

- Altug, Sumru (1989): "Time-to-Build and Aggregate Fluctuations: Some new evidence". *International Economic Review*, **30-4**, 889-920.
- Blanchard, Oliver Jean, and Danny Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances". *American Economic Review*, **79**, 655-673.
- Burnside, Craig, Martin Eichenbaum and Sergio Rebelo (1993): "Labor Hoarding and the Business Cycle". *Journal of Political Economy*, **101**, 245-273.
- Box, George E.P. (1980): "Sampling and Bayes' Inference in Scientific Modelling and Robustness". *Journal of the Royal Statistical Society A*, **143**, 383-430.
- Canova, Fabio (1994): "Statistical Inference in Calibrated Models". *Journal of Applied Econometrics*, **9**, S123-144.
- Canova, Fabio, Mary Finn and Adrian R. Pagan (1994): "Evaluating a Real Business Cycle Model". In: Colin P. Hargreaves (ed.) *"Nonstationary Time Series Analysis and Cointegration"*, Oxford University Press.
- Christiano, Lawrence J. (1991): "Modeling the Liquidity Effect of a Money Shock". *Federal Reserve Bank of Minneapolis - Quarterly Review*, Winter 1991, 3-34.
- Christiano, Lawrence J. and Martin Eichenbaum (1992a): "Current Real-Business Cycle Theory and Aggregate Labor-Market Fluctuations". *American Economic Review*, **82**, 430-450.
- (1992b): "Liquidity Effects and the Monetary Transmission Mechanism". *American Economic Review*, **82**, 346-353.
- Cogley, Timothy and James M. Nason (1995): "Output Dynamics in Real Business Cycle Models". *American Economic Review*, **85**, 492-511.
- DeJong, David N., Beth F. Ingram, and Charles H. Whiteman (1996): "A Bayesian Approach to Calibration". *Journal of Business Economics and Statistics*, **14**, 1-9.
- (1997): "A Bayesian Approach to Dynamic Macroeconomics". *Mimeographed*, University of Iowa.
- Diebold, Francis X., Lee E. Ohanian, and Jeremy Berkowitz (1998): "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data". *Review of Economic Studies*, **65**, 433-452.

- Fair, Ray (1994): *"Testing Macroeconometric Models"*. Harvard University Press, Cambridge.
- Fuerst, Timothy (1992): "Liquidity, loanable funds, and real activity". *Journal of Monetary Economics*, **29**, 3-24.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (1995): *"Bayesian Data Analysis"*, Chapman & Hall, New York.
- Geweke, John (1992): "Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments". In: J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (Eds) *"Bayesian Statistics 4"*, 169-194.
- (1995): "Bayesian Comparison of Econometric Models". *Mimeographed*, University of Minnesota.
- (1999): "Computational Experiments and Reality". *Mimeographed*, University of Minnesota.
- Good, I.J. (1952): "Rational Decisions". *Journal of the Royal Statistical Society B*, **14**, 107-114.
- Hansen, Lars Peter and James J. Heckman (1996): "The Empirical Foundations of Calibration". *Journal of Economic Perspectives*, **10**, 87-104.
- Ingram, Beth F. and Charles H. Whiteman (1994): "Supplanting the Minnesota prior – Forecasting macroeconomic time series using real business cycle model priors". *Journal of Monetary Economics*, **34**, 497-510.
- Ireland, Peter N. (1998): "A Method for Taking Models to the Data". *Mimeographed*, Rutgers University.
- Kass, Robert E. and Adrian E. Raftery (1995): "Bayes Factors". *Journal of the American Statistical Association*, **90**, 773-795.
- Kydland Finn and Edward Prescott (1996): "The Computational Experiment: An Econometric Tool". *Journal of Economic Perspectives*, **10**, 69-85.
- Leeper, Eric M. and Christopher A. Sims (1994): "Toward a Modern Macroeconomic Model Usable for Policy Analysis". *NBER Macroeconomics Annual 1994*, Stanley Fischer and Julio J. Rotemberg (eds.), MIT Press, Cambridge, MA.

- Min Chung-ki and Arnold Zellner (1993): "Bayesian and non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates". *Journal of Econometrics*, **56**, 89-118.
- Nasan, James M. and Timothy Cogley (1994): "Testing the Implications of Long-Run Neutrality for Monetary Business Cycle Models". *Journal of Applied Econometrics*, **9**, S37-70.
- Poirier, Dale (1997): "Comparing and Choosing between two Models with a third Model in the Background". *Journal of Econometrics*, **78**, 139-151.
- Phillips, Peter C.B. (1996): "Econometric Model Determination". *Econometrica*, **64**, 763-812.
- Rotemberg, Julio, and Michael Woodford (1996): "Real Business Cycle Models and the Forecastable Movements in Output, Hours, and Consumption". *American Economic Review*, **86**, 71-89.
- Schorfheide, Frank (1998): "Econometric Modeling of Macroeconomic Aggregates". *Ph.D. Dissertation*, Yale University.
- Silverman, B.W. (1986): "*Density Estimation for Statistics and Data Analysis*", Chapman & Hall, New York.
- Sims, Christopher A. (1982): "Policy Analysis with Econometric Models". *Brookings Paper of Economic Activity*, 107-64.
- (1986): "Are Forecasting Models Useable for Policy Analysis?". *Federal Reserve Bank of Minneapolis, Quarterly Review*, 1-16.
- (1995): "Solving Rational Expectations Models". *Manuscript*, Yale University.
- (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, Paul (1994): "Cyclical Properties of a Real Business Cycle Model". *Journal of Applied Econometrics*, **9**, S113-122.
- Zellner, Arnold (1971): "*Introduction to Bayesian Inference in Econometrics*". John Wiley & Sons, New York.