EUI Working Papers

NESTED MODELS AND MODEL UNCERTAINTY

Alexander Kriwoluzky and Christian A. Stoltenberg

Nested Models and Model Uncertainty

ALEXANDER KRIWOLUZKY
and
CHRISTIAN A. STOLTENBERG

Nested models and model uncertainty[*]

Alexander Kriwoluzky,[†]Christian A. Stoltenberg[‡]

October 8, 2009

Abstract

Uncertainty about the appropriate choice among nested models is a central concern for optimal policy when policy prescriptions from those models differ. The standard procedure is to specify a prior over the parameter space ignoring the special status of some sub-models, e.g. those resulting from zero restrictions. This is especially problematic if a model's generalization could be either true progress or the latest fad found to fit the data. We propose a procedure that ensures that the specified set of sub-models is not discarded too easily and thus receives no weight in determining optimal policy. We find that optimal policy based on our procedure leads to substantial welfare gains compared to the standard practice.

JEL classification: E32, C51, E52.
Keywords: Optimal monetary policy, model uncertainty, Bayesian model estimation.

1   Introduction

Recently, the empirical evaluation of Dynamic Stochastic General Equilibrium (DSGE) models employing Bayesian methods has made substantial progress (Smets and Wouters, 2003, 2007; An and Schorfheide, 2007; Lubik and Schorfheide, 2004). Policymakers nowadays correspondingly employ relatively large estimated DSGE models, including various features and frictions, in their policy analysis more and more. This practice is based on the implicit idea that by capturing many aspects of the economy in one single model, policy prescriptions derived from this model should guard against the risks of an uncertain economic environment.[1] However, as it ignores the special status of sub-models that are defined by zero restrictions, the recent practice is prone to uncertainty about the appropriate choice of nested models. We show that this source of uncertainty is a central concern for optimal policy and propose a procedure that insures against it by assigning a non-zero weight to the set of sub-models.

To fix ideas, consider the following situation. After some process of theorizing and data analysis, a policymaker has arrived at a baseline model, Model A. One day, a researcher proposes to extend this model by adding a new feature or friction, replacing it with Model B, in which Model A is nested. At a first glance, this seems to be a win-win situation because the new model nests all the advantages of Model A and moreover may improve the understanding of the economy and lead the policymaker to make better policy decisions. However, the gain in explanatory power may be relatively small, i.e. the posterior odds may not indicate substantial evidence against Model A. Discarding Model A is further problematic because instead of true improvement, Model B may be just the latest fad found to fit the data. When Model B introduces a conflicting stabilization aim into the decision about policy, optimal policy prescriptions from the two models differ. In this situation, the policymaker risks welfare losses by ignoring Model A and putting all her eggs in one basket. In this paper, we develop an approach that takes into account both Models A and B to determine optimal policy.

Starting with a baseline model, we subsequently estimate a set of competing and nested models. This bottom-up approach puts us into a position to separately evaluate the gain in explanatory power of each extension. Optimal policy is then computed by weighting each model with its posterior probability. Weighting over the set of nested models allows the policymaker to make reasonable extensions of the baseline model but also insure against the pitfalls of only employing one potentially misspecified model.

Using Euro-13 area data, we illustrate our approach to deal with model uncertainty in

---

[1]Exemplary papers that fall in this category are Levin, Onatski, Williams, and Williams (2005), Christiano, Trabandt, and Walentin (2007), and more recently Adolfson, Laseen, Linde, and Svensson (2008).

nested models by choosing as a baseline model one of the most popular models employed in monetary analysis nowadays: a standard cashless New Keynesian economy with staggered price-setting without indexation (Woodford, 2003a). As examples of uncertainty linked to the choice between nested models, we subsequently allow for more lags in endogenous variables (indexation and habit formation) and omitted variables (money). To represent the standard practice, we also consider one model that nests all these features. While the predominant principle of optimal policy in cashless models is price stability, a demand for money introduces a conflicting policy aim, namely the stabilization of the nominal interest rate. In this environment, we find that our procedure leads to welfare gains of approximately 70 percent compared to the standard practice.

The remainder of the paper is organized as follows. In the next section we introduce our approach to analyze the optimal conduct of policy under model uncertainty. In Section 3 we describe the baseline model and its extensions. In Section 4 we present our estimation results and its consequences for optimal monetary policy. The last section concludes.

2   Analyzing optimal policy under model uncertainty

In this section, after a short description of the general setup we present two approaches to cope with model uncertainty and describe how we assess the policy performance under model uncertainty. The first approach is set to represent the standard practice: without paying special attention to the set of sub-models, the policymaker determines optimal policy by maximizing households' utility within one single model that nests all features and frictions. The second approach takes uncertainty about the appropriate choice of nested models into account and weights over the set of nested models to derive optimal policy prescriptions.

2.1   General setup

Consider a system of linear equations that represent log-linear approximations to the non-linear equilibrium conditions under rational expectations around a deterministic steady state of a particular Model i. Let x_t be the vector of state variables, z_t the vector of structural shocks and y_t the vector of observable variables. Furthermore, let Θ_i denote the random vector of deep parameters and θ_i a particular realization from the joint posterior distribution in Model i. Policy influences the equilibrium outcome through simple feedback rules. The link between the set of policy instruments as a subset of x is characterized by the vector of

constant policy coefficients ϕ, i.e. by definition we consider steady state invariant policies.[2] The state space form of the solution of model i is given by[3]:

$$\widehat{x}_t = T(\theta_i, \phi)\widehat{x}_{t-1} + R(\theta_i, \phi)z_t \tag{1}$$

$$\widehat{y}_t = G\widehat{x}_t, \tag{2}$$

where $T(\theta_i, \phi)$ and $R(\theta_i, \phi)$ are matrices one obtains after solving a DSGE model with standard solution techniques. The matrix G is a picking matrix that equates observable and state variables.

We assess the performance of a particular policy ϕ by its effects on households' unconditional expected utility, i.e. before any uncertainty has been resolved. In Model i and for a particular realization θ_i, this unconditional expectation up to second order is represented by:

$$E\sum_{t=t_0}^{\infty} \beta^t U(x_t, \theta_i) \approx \frac{U(\bar{x}, \theta_i)}{1-\beta} - E\sum_{t=t_0}^{\infty} \beta^t A(\theta_i)\widehat{x}_t\widehat{x}_t' = \frac{U(\bar{x}, \theta_i)}{1-\beta} - \frac{L(\theta_i, \widehat{x})}{1-\beta}. \tag{3}$$

This approximation decomposes households' utility in two parts. The first part is utility in the steady state, and the second part comprises welfare-reducing fluctuations around the long-run equilibrium. We assume that the policymaker can credibly commit to a policy rule ϕ: if a policymaker decides to follow a certain policy rule ϕ once and forever, agents believe indeed that the policymaker will. Given a particular value θ_i, the optimal steady state invariant policy $\phi_i^\star(\theta_i)$ maximizes (3) by minimizing short-run fluctuations captured in $L(\theta_i, \widehat{x})$. Since the specification of households' preferences is independent of policy choices, the policymaker can only indirectly influence households' loss by shaping the dynamics of the endogenous variables $\widehat{x}$ as defined by (1).

2.2 Two approaches to model uncertainty

We now turn to the optimal conduct of policy if the policymaker faces uncertainty about the economic environment. We consider two approaches to cope with this uncertainty.

Specifying a marginal prior distribution with a positive unique mode for each parameter, the first approach or the standard practice is to develop and estimate one single model that nests all features and frictions and employ the model in determining optimal policy. This

---

[2]A steady state-invariant policy is a policy which affects the dynamic evolution of the endogenous variables around a steady state, but not the steady state itself.

[3]$\widehat{x}_t$ denotes the percentage deviation of the generic variable x_t from a deterministic steady state x chosen as approximation point.

is based on the idea that by capturing many aspects of the economy in one single model, policy prescriptions derived from this model should guard against the risks of an uncertain economic environment. The only source of uncertainty for the policymaker is uncertainty about the structural parameters of the model. We refer to this approach as the *complete-model* approach.

The second approach starts with a stylized baseline model and treats each extension by an additional feature or friction as a distinct and competing model. By averaging across models, this approach allows to take not only parameter uncertainty but also uncertainty about model specification into account. In the following we refer to this approach as the *model-averaging* approach.

When pursuing the first approach to deal with model uncertainty, the relevant uncertainty that a policymaker faces when she makes her decision about ϕ is given by the joint posterior distribution in the model that nests all features and frictions. We denote this 'complete' model by Model C and its corresponding posterior distribution of its structural parameters by $f(\theta_c) \equiv f(\theta|Y, \mathcal{M}_C)$, where Y is the set of time series used in the estimation. The optimal policy ($\phi_C^\star$) is defined by:

$$\phi_C^\star = \arg\min_{\phi} E_{\Theta_C} L(\Theta_C, \widehat{x}) \tag{4}$$
$$s.t. \ \widehat{x}_t = T(\theta_C, \phi)\widehat{x}_{t-1} + R(\theta_C, \phi)z_t, \quad \forall \theta_C,$$

where $E_{\Theta_C} L(\Theta_C, \widehat{x})$ is the expected loss when the structural parameters are a random vector. Due to parameter uncertainty the policymaker has to average the loss over all possible realizations of Θ_C to find the optimal vector of constant policy coefficients in Model C, $\phi_C^\star$.

The second approach explicitly addresses specification uncertainty and averages over different models. We separately estimate a discrete set of nested models $\mathcal{M} = \{\mathcal{M}_1, ..., \mathcal{M}_C\}$, where $\mathcal{M}_1$ denotes the baseline model, $\mathcal{M}_C$ the complete model and (C − 2) possible one-feature extensions of the baseline model. Employing the same data and prior specification of shocks and common parameters, we calculate marginal data densities $p(Y|\mathcal{M}_i) = E(f(Y|\Theta_i))$, where $f(Y|\Theta_i)$ denotes the data likelihood and the expectation is taken with respect to the prior distribution of the structural parameters. Since all models are nested in Model C, the marginal data density for Model i satisfies:

$$p(Y|\mathcal{M}_i) \equiv p(Y|\theta_{C \notin i} = \mathbf{0}, \mathcal{M}_C), \tag{5}$$

where $\theta_{C \notin i}$ denotes the vector of structural parameters for Model C that are not contained in

the set of structural parameters of Model $i = 1, 2, ... c$. We employ the harmonic mean estimator to compute the data likelihood in a certain model as proposed by Geweke (1999) and more recently applied among others by An and Schorfheide (2007). To compare the explanatory power of each model relative to the other models, we compute posterior probabilities which are defined as

$$P(M_i|Y) = \frac{P(M_i)p(Y|\mathcal{M}_i)}{\sum_{j=1}^{c} P(M_j)p(Y|\mathcal{M}_j)}, \tag{6}$$

where $P(M_i)$ denotes the prior probability for each model.[4] To ensure that sub-models are not discarded too easily and to facilitate competition between the nested models, we assign positive and equal prior weights to each model. The posterior probability of each model is then solely determined by its relative success to explain a given set of time series, i.e. it takes a value close to zero when the predictive density of a model relative to the others is neglectable.

In nested models, the second approach can also be thought of as defining a bimodal prior distribution for the parameter that represents the additional feature or friction. One part of the distribution is centered around the assumed positive modulus of the parameter, and the other modulus is centered around zero. The idea of paying special attention to this zero restriction and giving this possibility relatively more weight reflects the natural scepticism every researcher and policymaker has when extending a reasonable model.

Our approach however is more general than specifying bimodal prior distributions because it can also be applied when models are not nested. In particular, it avoids a discontinuity problem in the parameter space that arises when models are not nested. To see this, suppose that the baseline model $\mathcal{M}_1$ is replaced by a very similar model $\mathcal{M}_1^*$ that is not nested in the complete Model c. In other words, there is at least one parameter that is not included in the prior specification of the complete model. In this case, it seems to be reasonable to weight over all models, also including $\mathcal{M}_1^*$. The complete-model approach – even if it includes a bimodal prior specification – gives zero weight to the parameter included in $\mathcal{M}_1^*$ but not in Model c. The model-averaging approach weights over models independent whether they are nested or not, and thereby avoids this discontinuity. In addition, the formulation of a bimodal prior distribution in standard Bayesian model estimation is not straightforward and estimating a model extension to zero might cause serious troubles when approximating the posterior mean.

---

[4]An alternative approach to compute posterior model probabilities in nested models involves calculating Savage-Dickey density ratios as proposed by Verdinelli and Wasserman (1995).

The optimal policy for the model-averaging approach ($\phi_a^\star$) is defined by

$$\phi_a^\star = \arg\min_\phi E_{\mathcal{M},\Theta} L(\Theta_i, \widehat{x}) \tag{7}$$

$$s.t. \ \widehat{x}_t = T(\theta_i, \phi)\widehat{x}_{t-1} + R(\theta_i, \phi)z_t, \quad \forall \theta_i, \quad i = 1, ..., n.$$

The complete-model approach is a limiting case of model-averaging approach; they are equivalent if the complete model exhibits a posterior probability of unity.

2.3 Assessing policy performance within and across models

We compare the performance of the two approaches by computing the average costs of welfare relevant short-run fluctuations over all draws and models. This allows us to assess the pitfalls of employing only one model that nests all features and frictions in the policy analysis, i.e. focussing on parameter uncertainty in the complete model and thereby ignoring the issue of specification uncertainty about nested models. Throughout the paper we express the resulting business cycle costs ($\mathcal{BC}$) as the percentage loss in certainty (steady state) equivalent consumption. First we compute the loss of a certain policy $\tilde{\phi}$ given a particular parameter vector $\tilde{\theta}$ in model i to derive overall utility:

$$U\left(c(\tilde{\theta}_i), x_{\backslash c}(\tilde{\theta}_i), \tilde{\theta}_i\right) - L(\tilde{\theta}_i, \tilde{\phi}),$$

where the first term is steady state utility and $x_{\backslash c}$ denotes the variables vector excluding consumption. Since we want to express utility as reduction in certainty consumption equivalents we set this expression to be equal to:

$$U\left(c(\tilde{\theta}_i)(1 - \mathcal{BC}), x_{\backslash c}(\tilde{\theta}_i), \tilde{\theta}_i\right)$$

and solve for $\mathcal{BC}$ in percentage terms. Under parameter uncertainty this results in a distribution for $\mathcal{BC}(\tilde{\theta}_i, \tilde{\phi})$ over Θ_i. Taking the expectation of this expression yields a measure of the average losses in certainty consumption equivalents under a particular policy $\tilde{\phi}$.

As can be seen from (3), theoretical unconditional second moments derived from the DSGE model are relevant for households' utility losses due to short fluctuations – and thus for the computation of business-cycle costs under different policies. As Del Negro and Schorfheide (2008) point out, whether the theoretical unconditional moments relevant for policy assessment and the ones observed in the data coincide depends in particular on the specification of the prior distribution of standard deviations and autoregressive coefficients for the driving exogenous disturbances. We choose the prior distribution for the standard

deviations of the $i.i.d.$ terms in z_t and the autoregressive coefficients of the shocks contained in $T(\bullet)$ such that the relevant theoretical unconditional second moments at the posterior mean in each model are in line the ones computed directly from the stationary times series. This in turn yields welfare costs of short run fluctuations consistent with the limit put forward by Lucas (2003).

3   Optimal monetary policy: the economic environment

To demonstrate our main result, we create a set of monetary models including one model that nests all features and frictions. Starting with a plain-vanilla cashless new Keynesian economy as our baseline model (Woodford, 2003a), we subsequently introduce two additional features (indexation and habit formation) and a transaction friction (money in the utility function). While optimal policy in the baseline model and in the models that feature indexation and habit formation seeks to stabilize fluctuations in inflation and in the output gap, a transaction friction adds the stabilization of the nominal interest rate as an additional and conflicting policy aim. In this section we describe the models, derive the equations characterizing the equilibrium and the relevant policy objectives as the unconditional expectation of households' utility for each model.

3.1   The baseline economy: Model 1

The baseline economy consists of a continuum of infinitely-lived households indexed with $j \in [0, 1]$ that have identical initial asset endowments and identical preferences. Household j acts as a monopolistic supplier of labor services l_j. Lower (upper) case letters denote real (nominal) variables. At the beginning of period t, households' financial wealth comprises a portfolio of state contingent claims on other households yielding a (random) payment Z_{jt}, and one-period nominally non-state contingent government bonds B_{jt-1} carried over from the previous period. Assume that financial markets are complete, and let $q_{t,t+1}$ denote the period t price of one unit of currency in a particular state of period $t + 1$ normalized by the probability of occurrence of that state, conditional on the information available in period t. Then, the price of a random payoff Z_{t+1} in period $t + 1$ is given by $E_t[q_{t,t+1} Z_{jt+1}]$. The budget constraint of the representative household reads

$$B_{jt} + E_t[q_{t,t+1} Z_{jt+1}] + P_t c_{jt} \leq R_{t-1} B_{jt-1} + Z_{jt} + P_t w_{jt} l_{jt} + \int_0^1 D_{jit} di - P_t T_t, \qquad (8)$$

where c_t denotes a Dixit-Stiglitz aggregate of consumption with elasticity of substitution ζ, P_t the aggregate price level, w_{jt} the real wage rate for labor services l_{jt} of type j, T_t a lump-sum tax, R_t the gross nominal interest rate on government bonds, and D_{it} dividends from monopolistically competitive firms. The objective of the representative household is

$$E_{t_0} \sum_{t=t_0}^{\infty} \beta^t \{u(c_{jt}) - v(l_{jt})\}, \quad \beta \in (0,1), \tag{9}$$

where β denotes the subjective discount factor. The instantaneous utility function is assumed to be non-decreasing in consumption, decreasing in labor time, strictly concave, twice differentiable, and to fulfill the Inada conditions. Households are wage-setters supplying differentiated types of labor l_j, which are transformed into aggregate labor l_t with $l_t^{(\epsilon_t-1)/\epsilon_t} = \int_0^1 l_{jt}^{(\epsilon_t-1)/\epsilon_t} dj$. We assume that the elasticity of substitution between different types of labor, $\epsilon_t > 1$, varies exogenously over time. Cost minimization implies that the demand for differentiated labor services l_{jt}, is given by $l_{jt} = (w_{jt}/w_t)^{-\epsilon_t} l_t$, where the aggregate real wage rate w_t is given by $w_t^{1-\epsilon_t} = \int_0^1 w_{jt}^{1-\epsilon_t} dj$. The transversality condition is given by

$$\lim_{i \to \infty} E_t \beta^i \lambda_{jt+i} (B_{jt+i} + Z_{jt+1+i})/P_{jt+i} = 0 \tag{10}$$

.

The final consumption good Y_t is an aggregate of differentiated goods produced by monopolistically competitive firms indexed with $i \in [0,1]$ and defined as $y_t^{\frac{\zeta-1}{\zeta}} = \int_0^1 y_{it}^{\frac{\zeta-1}{\zeta}} di$, with $\zeta > 1$. Let P_{it} and P_t denote the price of good i set by firm i and the price index for the final good. The demand for each differentiated good is $y_{it}^d = (P_{it}/P_t)^{-\zeta} y_t$, with $P_t^{1-\zeta} = \int_0^1 P_{it}^{1-\zeta} di$. A firm i produces good y_i using a technology that is linear in the labor bundle $l_{it} = [\int_0^1 l_{jit}^{(\epsilon_t-1)/\epsilon_t} dj]^{\epsilon_t/(\epsilon_t-1)}$: $y_{it} = a_t l_{it}$, where $l_t = \int_0^1 l_{it} di$ and a_t is a productivity shock with mean 1. Labor demand satisfies: $mc_{it} = w_t/a_t$, where $mc_{it} = mc_t$ denotes real marginal costs independent of the quantity that is produced by the firm. We allow for a nominal rigidity in form of a staggered price setting as developed by Calvo (1983). Each period firms may reset their prices with probability $1 - \alpha$ independently of the time elapsed since the last price setting. A fraction $\alpha \in [0,1)$ of firms are assumed to keep their previous period's prices, $P_{it} = P_{it-1}$. In each period a measure $1 - \alpha$ of randomly selected firms set new prices $\widetilde{P}_{it}$ as the solution to

$$\max_{\widetilde{P}_{it}} E_t \sum_{T=t}^{\infty} \alpha^{T-t} q_{t,T} (\widetilde{P}_{it} y_{iT} (1-\tau) - P_T mc_T y_{iT}), \quad s.t. \ y_{iT} = (\widetilde{P}_{it})^{-\zeta} P_T^{\zeta} y_T, \tag{11}$$

where τ denotes an exogenous sales tax. We assume that firms have access to contingent

claims.

The aggregate resource constraint is given by

$$y_t = a_t l_t / \Delta_t, \tag{12}$$

where $\Delta_t = \int_0^1 (P_{it}/P_t)^{-\zeta} di \geq 1$ and thus $\Delta_t = (1-\alpha)(\widetilde{P}_t/P_t)^{-\zeta} + \alpha \pi_t^\zeta \Delta_{t-1}$. The dispersion measure Δ_t captures the welfare decreasing effects of staggered price setting. Goods' market clearing requires

$$c_t + g_t = y_t. \tag{13}$$

The central bank as the monetary authority is assumed to control the short-term interest rate R_t with a simple feedback rule contingent on past interest rates, inflation and output:

$$R_t = f(R_{t-1}, \pi_t, y_t). \tag{14}$$

The consolidated government budget constraint reads: $R_{t-1}B_{t-1} + P_t G_t = B_t + P_t T_t + \int_0^1 P_{it} y_{it} \tau di$. The exogenous government expenditures g_t evolve around a mean $\bar{g}$, which is restricted to be a constant fraction of output, $\bar{g} = \bar{y}(1-sc)$. We assume that tax policy guarantees government solvency, i.e., ensures $\lim_{i \to \infty} (B_{t+i}) \prod_{v=1}^i R_{t+v}^{-1} = 0$.

We collect the exogenous disturbances in the vector $\xi_t = [a_t, g_t, \mu_t]$, where $\mu_t = \frac{\epsilon_t}{\epsilon_t - 1}$ is a wage mark-up shock. It is assumed that the percentage deviations of the first two elements of the vector from their means evolve according to autonomous AR(1)-processes with autocorrelation coefficients $\rho_a, \rho_g \in [0, 1)$. The process for $\log(\mu_t/\bar{\mu})$ and all innovations, $z_t = [\varepsilon_t^a, \varepsilon^{\tilde{g}_t}, \varepsilon_t^\mu]$, are assumed to be i.i.d..

The recursive equilibrium is defined as follows:

Definition 1 *Given initial values $P_{t_0-1} > 0$ and $\Delta_{t_0-1} \geq 1$, a monetary policy and a Ricardian fiscal policy $T_t \ \forall \ t \geq t_0$, and a sales tax τ, a rational expectations equilibrium (REE) for $R_t \geq 1$, is a set of sequences $\{y_t, c_t, l_t, mc_t, w_t, \Delta_t, P_t, \widetilde{P}_{it}, R_t\}_{t=t_0}^\infty$ for $\{\xi_t\}_{t=t_0}^\infty$*

   (i) that solve the firms' problem (11) with $\widetilde{P}_{it} = \widetilde{P}_t$,

   (ii) that maximize households' utility (9) s.t. their budget constraints (8),

   (iii) that clear the goods market (13),

   (iv) and that satisfy the aggregate resource constraint (12) and the transversality condition (10).

In the next step, we seek to estimate the model by employing Bayesian methods. To do so, we log-linearize the structural equations around the deterministic steady state under

10

zero inflation. Thus, the dynamics in the baseline economy are described by the following two structural equations:

$$\sigma(E_t\widehat{y}_{t+1} - E_t\widehat{y}_{t+1}^n) = \sigma(\widehat{y}_t - \widehat{y}_t^n) + \widehat{R}_t - E_t\widehat{\pi}_{t+1} - \widehat{R}_t^n \tag{15}$$

$$\widehat{\pi}_t = \beta E_t\widehat{\pi}_{t+1} + \kappa(\widehat{y}_t - \widehat{y}_t^n), \tag{16}$$

where $\sigma = -u_{cc}c/(u_csc)$, $\omega = v_{ll}l/v_l$ and $\kappa = (1-\alpha)(1-\alpha\beta)(\omega+\sigma)/\alpha$. Furthermore, $\widehat{k}_t$ denotes the percentage deviation of a generic variable k_t from its steady-state value k. The natural rates of output and interest, i.e the values for output and real interest under flexible prices, are given by the following expressions

$$\widehat{y}_t^n = \frac{(1+\omega)\widehat{a}_t + \sigma\widetilde{g}_t - \widehat{\mu}_t}{\omega + \sigma}, \quad \widehat{R}_t^n = \sigma[(\widetilde{g}_t - \widehat{y}_t^n) - E_t(\widetilde{g}_{t+1} - \widehat{y}_{t+1}^n)],$$

where $\widetilde{g}_t = (g_t - g)/y$. The model is closed by a simple interest rate feedback rule as an approximation to (14):

$$\widehat{R}_t = \rho_R\widehat{R}_{t-1} + \phi_\pi\widehat{\pi}_t + \phi_y\widehat{y}_t. \tag{17}$$

The general system (1) in the baseline model then is the fundamental locally stable and unique solution that satisfies (15)-(17) for a certain vector of constant policy coefficients $\phi = (\rho_R, \phi_\pi, \phi_y)$.

Our welfare measure is the unconditional expectation of representative households' utility. Building on Woodford (2003a), after averaging over all households, a second-order approximation to (9) results in the following quadratic loss function (for a given realization θ_1)[5]:

$$L(\theta_1, \widehat{x}) = \frac{u_cy\zeta(\omega+\sigma)}{2\kappa}\{var(\widehat{\pi}_t) + \lambda_d var(\widehat{y}_t - \widehat{y}_t^e)\}, \tag{18}$$

where $\lambda_d = \kappa/\zeta$ and the efficient rate of output is given by

$$\widehat{y}_t^e = \widehat{y}_t^n + \widehat{\mu}_t/(\omega + \sigma).$$

In the next subsection we consider habit formation and indexation to past inflation as examples of missing lags in consumption and inflation.

---

[5]Throughout we assume that the steady state is rendered efficient by an appropriate setting of the sales tax rate.

3.2 Habit formation (Model 2) and indexation (Model 3)

One example of a missing lag in an endogenous variable is to allow for an internal habit (e.g. Boivin and Giannoni (2006); Woodford (2003a)) in households' total consumption. The constituting equations for (1) are the policy rule (17) and the modified versions of the Euler equation and the New Keynesian Philips curve:

$$\varphi[d_t - \eta d_{t-1}] - \varphi\beta\eta E_t[d_{t+1} - \eta d_t] = E_t\widehat{\pi}_{t+1} + \widehat{R}_t^n - \widehat{R}_t...$$
$$+ E_t\varphi[d_{t+1} - \eta d_t] - \varphi\beta\eta E_t[d_{t+2} - \eta d_{t+1}] \quad (19)$$

$$\widehat{\pi}_t = \kappa_h[(d_t - \delta^* d_{t-1}) - \beta\delta^* E_t(d_{t+1} - \delta^* d_t)] + \beta E_t\widehat{\pi}_{t+1}, \quad (20)$$

where $d_t = \widehat{y}_t - \widehat{y}_t^n$, $\kappa_h = \eta\varphi\kappa[\delta^*(\omega + \sigma)]^{-1}$, $\varphi = \sigma/(1 - \eta\beta)$, and the natural rate of output follows[6]

$$[\omega + \varphi(1 + \beta\eta^2)]\widehat{y}_t^n - \varphi\eta\widehat{y}_{t-1}^n - \varphi\eta\beta E_t\widehat{y}_{t+1}^n = \varphi(1 + \beta\eta^2)\tilde{g}_t - \varphi\eta\tilde{g}_{t-1} - \varphi\eta\beta E_t g_{t+1}...$$
$$+ (1 + \omega)\widehat{a}_t - \widehat{\mu}_t.$$

Approximating households' utility to second order results in the following loss function:

$$L(\theta_2, \widehat{x}) = \frac{(1 - \beta\eta)\eta\varphi u_c^h y^h \zeta}{2\kappa_h \delta^*}\{var(\widehat{\pi}_t) + \lambda_{d,h}var(\widehat{y}_t - \widehat{y}_t^e - \delta^*(\widehat{y}_{t-1} - \widehat{y}_{t-1}^e))\}, \quad (21)$$

where $\lambda_{d,h} = \kappa_h/\zeta$ and the efficient rate of output is characterized by

$$[\omega + \varphi(1 + \beta\eta^2)]\widehat{y}_t^e - \varphi\eta\widehat{y}_{t-1}^e - \varphi\eta\beta E_t\widehat{y}_{t+1}^e = \varphi(1 + \beta\eta^2)\tilde{g}_t - \varphi\eta\tilde{g}_{t-1} - \varphi\eta\beta E_t g_{t+1}...$$
$$+ (1 + \omega)\widehat{a}_t.$$

Like habit formation, the indexation of prices to past inflation induces the economy to evolve in a history-dependent way. We assume that the fraction of prices that are not reconsidered, α, adjusts according to $\log P_{it} = \log P_{it-1} + \gamma\log\pi_{t-1}$ with $0 \leq \gamma \leq 1$ as the degree of indexation. This implies that price dispersion evolves according to $\Delta_t = (1-\alpha)(\frac{\tilde{P}_t}{P_t})^{-\zeta} + \alpha\pi_{t-1}^{-\zeta\gamma}\Delta_{t-1}\pi_t^\zeta$. Correspondingly, the economy with indexation is characterized by a modified aggregate supply curve

$$\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1} = \beta E_t(\widehat{\pi}_{t+1} - \gamma\widehat{\pi}_t) + \kappa(\widehat{y}_t - \widehat{y}_t^n), \quad (22)$$

---

[6]The parameter δ^*, $0 \leq \delta^* \leq \eta$, is the smaller root of the quadratic equation $\eta\varphi(1 + \beta\delta^2) = [\omega + \varphi(1 + \beta\eta^2)]\delta$. This root is assigned to past values of the natural and efficient rate of output in their stationary solutions.

(15) and (17). The corresponding loss function of the central bank reads Woodford (2003a):

$$L(\theta_3, \widehat{x}) = \frac{u_c y \zeta(\omega + \sigma)}{2\kappa} \{var(\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1}) + \lambda_d var(\widehat{y}_t - \widehat{y}_t^e)\}, \qquad (23)$$

where λ_d and the efficient rate of output are defined as in the baseline economy.

3.3 Money in the utility function (Model 4)

We introduce a transaction friction by letting real money balances enter households' utility in a separable way. More precisely, households' utility of holding real money balances is augmented by the amount $z(m_t)$ and a demand equation for real money balances enters the set of equilibrium conditions. In log-linearized form this additional equilibrium condition is given by:

$$\widehat{m}_t = -\frac{1}{\sigma_m(R-1)}\widehat{R}_t - \frac{1}{\sigma_m}\widehat{\lambda}_t, \qquad (24)$$

where $\sigma_m = -z_{mm}m/z_m$ and $\hat{\lambda}_t$ denotes the Lagrangian multiplier on the budget constraint of the household. The stabilization loss in Model 4 is given by:

$$L(\theta_4, \widehat{x}) = \frac{u_c y \zeta(\omega + \sigma)}{2\kappa} \{var(\widehat{\pi}_t) + \lambda_d var(\widehat{y}_t - \widehat{y}_t^e) + \lambda_{1R} var(\widehat{R}_t)\}, \qquad (25)$$

where $\lambda_d = \kappa/\zeta$, $\lambda_{1R} = \lambda_d \beta[v(\omega + \sigma)(1 - \beta)\sigma_m]^{-1}$ and $v = y/m$. The general form (1) has to satisfy the (15)-(17) and (??).

3.4 The complete model (Model c)

The complete model builds on the baseline model and comprises habit formation, indexation and money in the utility function. The equilibrium conditions in this case are: (19), (17), (??) and

$$\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1} = \beta E_t(\widehat{\pi}_{t+1} - \gamma\widehat{\pi}_t) + \kappa_h[(d_t - \delta^* d_{t-1}) - \beta\delta^* E_t(d_{t+1} - \delta^* d_t)]. \qquad (26)$$

In the following proposition we state the loss function for Model c.

Proposition 1 *If the fluctuations in y_t around y, R_t around R, ξ_t around ξ, π_t around π are small enough, $(R - 1)/R$ is small enough, and if the steady state distortions ϕ vanish due to the existence of an appropriate subsidy τ, the utility of the average household can be*

approximated by:

$$U_{t_0} = -E_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} L(\theta_{\mathrm{C}}, \widehat{x}) + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t, (R-1)/R\|^3), \tag{27}$$

where t.i.s.p. indicate terms independent of stabilization policy,

$$L(\theta_{\mathrm{C}}, \widehat{x}) = \frac{(1-\beta\eta)\eta\varphi u_c^h y^h \zeta}{2\kappa_h \delta^*} \{var(\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1}) + \lambda_{d,h} var(\widehat{y}_t - \widehat{y}_t^e - \delta^*(\widehat{y}_{t-1} - \widehat{y}_{t-1}^e)) + \lambda_{2R} var(\widehat{R}_t)\}, \tag{28}$$

$$\lambda_x = \lambda_{d,h} = \kappa_h/\zeta, \quad \lambda_{2R} = \frac{\lambda_{d,h}\beta\delta^*}{v\sigma_m(1-\beta)\eta\varphi},$$

and $v = y/m > 0$.

Proof: see appendix A.1.

4   Results

In this section we first present and interpret the estimation results. These results will be key for the assessment of the relevant model uncertainty faced by the policymaker. In the second part we compute optimal simple rules along with the procedures laid out in section 2. As a standard, we determine optimal monetary policy at the posterior mean, i.e. optimal policy in the absence of any model uncertainty. Then we analyze optimal policy when there is uncertainty about the appropriate choice of nested models.

4.1   Data and estimation results

We treat the variables real wage, output and consumer price inflation as observable. The data consists of HP filtered quarterly values of these variables for the EU 13 countries from 1970-2006.[7]

We calibrate the discount factor to $\beta = 0.99$, the steady-state fraction of private consumption relative to GDP $c/y = 0.8$ and the elasticity of substitution between differentiated goods to $\zeta = 6$ (see Woodford, 2003a). The specification of the prior distributions of the estimated deep parameters closely follows Negro and Schorfheide (2009), Smets and Wouters (2003) and Smets and Wouters (2007).[8] While we assume the disturbances $\tilde{g}_t$ and $\widehat{a}_t$ to follow stationary $AR(1)$ processes, $\widehat{\mu}_t$ is supposed to be *i.i.d.*. Since we are interested in evaluating

---

[7]The dataset we use was kindly provided by the Euro Area Business Cycle Network (EABCN). For a description of how this data is constructed see Fagan, Henry, and Mestre (2001).

[8]See Appendix A.2 Table 6 for a detailed description.

the explanatory power of each extension of the baseline model separately, common parameters in the set of models need to exhibit the same sufficient prior statistics. In particular, the marginal prior distributions for the set of coefficients that describe the shock processes, $\psi_{\tilde{g}}, \psi_a$ and $\sigma_{\tilde{g}}, \sigma_a, \sigma_\mu$, do not change across models, and they are specified according to the procedure explained in Section 2.3.

We approximate the joint posterior distribution of structural parameters by drawing $100,000$ times employing a standard MCMC-algorithm as described in An and Schorfheide (2007) and discard the first $80,000$ draws. The estimation results are displayed in Table 7 and the posterior and prior distributions are plotted in Figure 1-5 in Appendix A.2. The estimates of the posterior mean of the degree of relative risk aversion with respect to consumption (σ_c), the degree of indexation (γ), and the degree of price stickiness (α) correspond almost one-for-one to the findings by Smets and Wouters (2003). In line with Woodford (2003a) we find the labor supply decision with respect to changes in the real wage ($1/\omega$) to be elastic, i.e. values for ω vary between 0.3 and 0.4. Our estimate of the internal habit parameter (η) is comparable to Negro and Schorfheide (2009). Real money balances contribute only separately to households' utility in Model 4 and Model c and do not influence the equilibrium dynamics of output, inflation and the real wage. The parameter of relative risk aversion with respect to real money balances (σ_m) cannot be identified and thus the prior distribution and the posterior distribution are alike (see Figures 4 and 5).

In order to assess the explanatory power of each model, we compute marginal likelihoods and the corresponding posterior probabilities. The results are presented in Table 1. Here the key result is, that adding frictions and features to the baseline model does not necessarily increase the posterior probability. First, enriching the baseline model with a demand for cash does not increase the marginal likelihood for Model 4: real money balances do not help to predict the observable variables. Second, although the posterior distribution of the habit parameter (η) in Model 2 indicates a positive posterior mean of this parameter, a habit in consumption does not improve the fit to the data. This points to a well-known problem in Bayesian model estimation: The informative prior on the habit parameter introduces curvature into the posterior density surface (as pointed out by Poirier (1998) and An and Schorfheide (2007)). Third, history dependence in inflation improves the fit of the model. With approximately 81% Model 3 exhibits the highest posterior probability. Thus, the complete model incorporates features that helps to predict the data (indexation) and others that do not (habit and money). It therefore exhibits a marginal likelihood higher than Model 1 but lower than Model 3.

Table 1: Posterior probabilities and marginal data densities

	M_1	M_2	M_3	M_4	M_{C}
$p(Y\|\mathcal{M}_i)$	1683.98	1682.69	1696.83	1683.57	1695.39
$P(M_i\|Y)$	0.00	0.00	0.81	0.00	0.19

The welfare-assessment of optimal and sub-optimal policies in and across models depends on the magnitude of the resulting stabilization losses, i.e. the welfare relevant unconditional variances or standard deviations. In our context, these are the unconditional fluctuations in inflation and consumption (expressed in terms of a welfare-relevant output gap) for the models without a transaction friction (see e.g. (18)), and additionally fluctuations in interest rates, when money enters the utility function (see e.g (28)). As can be verified in Table 2, our estimated theoretical moments at the posterior mean are consistent with the corresponding ones directly estimated from the stationary times series.

Table 2: Welfare-relevant standard deviations: models vs. data

	M_1	M_2	M_3	M_4	M_{C}	$Data$
$std(c,\bar{\theta}_i)$	0.0070	0.0090	0.0068	0.0070	0.0078	0.0073
$std(\pi,\bar{\theta}_i)$	0.0020	0.0020	0.0023	0.0020	0.0022	0.0020
$std(R,\bar{\theta}_i)$	0.0028	0.0027	0.0031	0.0028	0.0031	0.0028

In the next section we begin the analysis of optimal policies in and across models.

4.2   Optimal policy at the posterior mean

To establish a standard and to explain the stabilization trade-off, we determine the optimal policy $\phi_i^\star = (\rho_R^\star, \phi_\pi^\star, \phi_y^\star)_i$ at the posterior mean $\bar{\theta}_i$ for each Model i, $i = 1, 2, .., \mathrm{C}$. To ease the numerical computation and to exclude unreasonably high policy responses, we assume the following bounds for the policy coefficients of the simple interest rate rule:

$$\rho_R \in [0, 20], \quad \rho_\pi \in [0, 20], \text{ and } \rho_y \in [0, 20].$$

The optimal coefficients and the resulting business cycles costs ($\mathcal{BC}$) expressed as equivalent reductions in steady-state consumption are displayed in Table 3.

Table 3: Optimal policy at the posterior mean $(\phi_i^\star)$

$$\widehat{R}_t = \rho_R \widehat{R}_{t-1} + \phi_\pi \widehat{\pi}_t + \phi_y \widehat{y}_t$$

	M_1	M_2	M_3	M_4	M_{C}
ρ_R	0.81	1.05	0.62	1.26	1.36
ϕ_π	20.00	20.00	20.00	2.42	1.01
ϕ_y	0.00	0.02	0.00	0.00	0.00
$\mathcal{BC}(\bar{\theta}_i, \phi_i^\star)$	0.0014%	0.0014%	0.0020%	0.0194%	0.0178%

Optimal policies are characterized by drawing on past interest rates. Put differently, optimal policy is history-dependent (Woodford, 2003a,b). In the first three models inflation stabilization is the predominant aim. Correspondingly, optimal policies feature a strong reaction on inflation.[9] In Models 4 and C, households value real money balances as a medium for transactions. This introduces stabilization of the nominal interest rate as a conflicting aim to price stability (see (25) and (27)) in the presence of fluctuations in the natural rate of interest. For intuition on this, suppose that ϕ_y is small and that the economy in Model 1 is hit by a wage-markup shock. To fight inflationary tendencies the output gap must decrease according to the aggregate supply curve (16). This in turn requires a strong increase in the nominal interest rate to fulfil the Euler equation (15), since the cost-push shock affects the natural rate of interest. Therefore, optimal policies in models with a demand for cash exhibit a higher coefficient ρ_R to smooth interest rates and a less aggressive response to inflation.

Welfare costs in models that feature a transaction friction are substantially higher. This increase is due to two effects. First, the stabilization of the interest rate adds a new component to the welfare-relevant stabilization loss, which accounts for over fifty percent of the increase in business cycle costs in Model 4 relative to Model 1. The second effect relates to the conflict of stabilizing interest rates, inflation and the output gap simultaneously, as apparent in the muted response to inflation in the optimal rules for Models 4 and C. The resulting increase in the unconditional weighted variances of inflation and the output gap accounts for the remaining increase in the costs of business cycle fluctuations.

---

[9]The optimal policy response on inflation in these models always corresponds to its upper bound. However, the welfare comparison between the two approaches to model uncertainty is independent of the particular choice of the upper bound on the inflation response.

Table 4: The weights λ_d and λ_R at the posterior mean

Weights	M_1	M_2	M_3	M_4	M_C
λ_d	0.0063	0.0231	0.0079	0.0057	0.0328
λ_R	-	-	-	0.0602	0.0728

Table 4 shows how the importance of stabilization aims relative to inflation for households changes across models. For example, the stabilization of the output gap is five times more important in Model C than in Model 1. In addition, the exact gap that policy should stabilize to maximize welfare differs (see (18) and (27)). Furthermore, comparing the two models that feature a demand for cash reveals that the optimal response to changes in inflation is larger in Model 4 than in Model C. Although both specifications incorporate stabilizing the nominal interest rate as a policy aim, this aim is relatively more important in Model C than in Model 4.

4.3 Evaluating two approaches to model uncertainty

In this section we quantitatively compare the two approaches to model uncertainty, the complete-model and the model-averaging approach. We start by determining the set of policy coefficients for the former approach according to (5), which yields

$$\phi_C^\star : \quad \rho_R = 1.34; \quad \phi_\pi = 1.17; \quad \phi_y = 0.00.$$

However, Model C is not the likeliest model since it also contains features which do not help to explain the given time series of GDP, inflation and the real wage (see Table 1). A policymaker pursuing a model-averaging approach to model uncertainty weights welfare losses in a particular model with its posterior probability, i.e. derives an optimal policy over all draws and models according to (7):

$$\phi_a^\star : \quad \rho_R = 1.39; \quad \phi_\pi = 3.36; \quad \phi_y = 0.00.$$

Comparing the characteristics of the two rules reveals two similarities and one difference. Both rules draw heavily on past interest rates to avoid welfare-reducing fluctuations in the interest rate in Models 4 and C, and put no emphasize on stabilizing the output gap. The main difference between both rules is the preference to stabilize inflation. While there is a conflict in stabilizing inflation and the nominal interest rate jointly in Model C, this trade-off is absent in the likeliest model, Model 3.

To evaluate the performance of the two approaches as a guard against model uncertainty we compute the business cycle cost for both policy rules in each Model i, i.e. $\mathcal{BC}(\Theta_i, \phi_{\mathrm{C}}^{\star})$ and $\mathcal{BC}(\Theta_i, \phi_a^{\star})$ for $i = 1, 2, 3, 4, \mathrm{C}$.

Table 5: Relative performance of $\phi_{\mathrm{C}}^{\star}$ and $\phi_a^{\star}$

	M_1	M_2	M_3	M_4	M_{C}	WA
$\mathcal{BC}(\Theta_i, \phi_{\mathrm{C}}^{\star})/\mathcal{BC}(\Theta_i, \phi_a^{\star})$	2.22	2.16	1.90	1.26	0.74	1.68

WA denotes the posterior-model probability average of business cycle costs.

As can be seen from Table 5, the optimal rule $\phi_a^{\star}$ performs twice as good as $\phi_{\mathrm{C}}^{\star}$ in Models 1, 2 and 3 where inflation stabilization is the predominant principle. Nevertheless, by reacting less harshly to inflation than the optimal rules from those models (see Table 3), it avoids high welfare losses in Model C. On average, optimal policy derived from the model-averaging approach leads to welfare gains of 68% relative to the optimal policy rule derived by the complete-model approach.

5   Conclusion

In this paper we have analyzed how to optimally conduct policy from a Bayesian perspective when the policymaker faces uncertainty about the appropriate choice among nested models. In particular, we have compared two approaches to model uncertainty. The complete-model approach is set to represent the standard practice: without paying special attention to the set of sub-models, the policymaker determines optimal policy by maximizing households' utility within one single model that nests all features and frictions. The model-averaging approach takes uncertainty about the appropriate choice of nested models into account and weights over the set of nested models to derive optimal policy prescriptions. Using EU-13 data, we find that the model-averaging approach leads to welfare gains of approximately 70 percent compared to the standard practice.

References

ADOLFSON, M., S. LASEEN, J. LINDE, AND L. E. SVENSSON (2008): "Optimal Monetary Policy in an Operational Medium-Sized DSGE Model," NBER Working Papers 14092, National Bureau of Economic Research, Inc.

AN, S., AND F. SCHORFHEIDE (2007): "Bayesian Analysis of DSGE Models by S. An and F. Schorfheide," *Econometric Reviews*, 26(2), 187–192.

BOIVIN, J., AND M. GIANNONI (2006): "Has monetary policy become more effective?," *Review of Economics and Statistics*, 3(88), 445–462.

CALVO, G. A. (1983): "Staggered price setting in a utility-maximizing framework.," *Journal of Monetary Economics*, 12, 383–398.

CHRISTIANO, L. J., M. TRABANDT, AND K. WALENTIN (2007): "Introducing Financial Frictions and Unemployment into a Small Open Economy Model," Working Paper Series 214, Sveriges Riksbank (Central Bank of Sweden).

DEL NEGRO, M., AND F. SCHORFHEIDE (2008): "Forming Priors for DSGE Models (and How it Affects the Assessment of Nominal Rigidities)," NBER Working Papers 13741, National Bureau of Economic Research.

FAGAN, G., J. HENRY, AND R. MESTRE (2001): "An area-wide model (AWM) for the euro area," Discussion Paper 42, ECB.

GEWEKE, J. (1999): "Using Simulation Methods for Bayesian Econometric Models," Computing in economics and finance 1999, Society for Computational Economics.

LEVIN, A., A. ONATSKI, J. WILLIAMS, AND N. WILLIAMS (2005): "Monetary Policy under Uncertainty in Microfounded Macroeconometric Models," *NBER Macroeconomics Annual*.

LUBIK, T. A., AND F. SCHORFHEIDE (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94(1), 190–217.

LUCAS, R. E. (2003): "Macroeconomic Priorities," *American Economic Review*, 93(1), 1–14.

NEGRO, M. D., AND F. SCHORFHEIDE (2009): "Monetary policy analysis with potentially misspecified models," *American Economic Review, forthcoming*.

POIRIER, D. J. (1998): "Revising beliefs in nonidentified models," *Econometric Theory*, 14(04), 483–509.

SMETS, F., AND R. WOUTERS (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

SMETS, F., AND R. WOUTERS (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97(3), 586–606.

VERDINELLI, I., AND L. A. WASSERMAN (1995): "Computing Bayes Factors by Using a Generalization of the Savage-Dickey density ratio," *Journal of the American Statistical Association*, 90, 614–618.

WOODFORD, M. (2003a): *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.

———— (2003b): "Optimal interest rate smoothing," *Review of Economic Studies*, 70, 861–886.

A Appendix

A.1 Proof of proposition 1

The period utility function of the average household in equilibrium is given by:

$$\int_0^1 [u(\bullet) - v(l_{jt}) + z(m_t)]dj = u(y_t - g_t - \eta(y_{t-1} - g_{t-1})) - \int_0^1 v(l_{jt})dj + z(m_t).$$

To derive (27) we need to impose that, in the optimal steady state, real money balances are sufficiently close to being satiated (see Woodford, 2003a, Assumption 6.1) such that we can treat $(R-1)/R$ as an expansion parameter.

The first summand can be approximated to second order by:

$$u(y_t - g_t - \eta(y_{t-1} - g_{t-1})) = u_c y(1 - \beta\eta)[\widehat{y}_t + \frac{1}{2}(1 - \varphi(1 + \eta^2\beta))\widehat{y}_t^2 + \varphi\eta\widehat{y}_t\widehat{y}_{t-1}$$
$$+ \varphi\widehat{y}_t(-\eta g_{t-1} - \beta\eta g_{t+1} + (1 + \eta^2\beta)g_t)] + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3), \quad (29)$$

where we used $(x_t - x) = x(\widehat{x}_t + 0.5\widehat{x}_t^2) + \mathcal{O}(\|\widehat{x}_t\|^3)$, $\varphi = \frac{\sigma}{1 - \beta\eta}$, t.i.s.p denotes terms independent of stabilization policy, $\sigma = -y u_{11}/u_1$, and $g_t = (G_t - G)/y$.

Since $y_t = a_t l_t / \Delta_t$, the second term can be approximated by

$$v(l_t) = u_c(1 - \beta\eta)[\widehat{y}_t + \frac{1 + \omega}{2}\widehat{y}_t^2 - (1 + \omega)\widehat{a}_t\widehat{y}_t + \widehat{\Delta}_t] + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3), \quad (30)$$

where we posited that in the equilibrium under flexible wages each household supplies the same amount of labor, $l = y$, $\omega = \frac{v_{ll}}{v_l}l$, and that due to the existence of an output subsidy the steady state is rendered efficient with $v_l = u_c(1 - \beta\eta)$. In the next step we combine (29) and (30), employ $\tilde{g}_t = -\eta g_{t-1} - \beta\eta E_t g_{t+1} + (1 + \eta^2\beta)g_t$, and obtain:

$$u(y_t - g_t - \eta(y_{t-1} - g_{t-1})) - \int_0^1 v(l_{jt})dj = u_c y(1 - \beta\eta)[\frac{1}{2}(-\varphi(1 + \eta^2\beta) - \omega)\widehat{y}_t^2$$
$$+ \varphi\eta\widehat{y}_t\widehat{y}_{t-1} + \varphi\widehat{y}_t\tilde{g}_t + (1 + \omega)\widehat{a}_t\widehat{y}_t - \widehat{\Delta}_t] + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3). \quad (31)$$

The efficient rate of output is defined by the following difference equation:

$$[\omega + \varphi(1 + \beta\eta^2)]\widehat{y}_t^e - \varphi\eta\widehat{y}_{t-1}^e - \varphi\eta\beta E_t\widehat{y}_{t+1}^e = \varphi\tilde{g}_t + (1 + \omega)\widehat{a}_t + \mathcal{O}(\|\widehat{\xi}_t\|^2).$$

If we use this expression to rewrite (31), we obtain the following:

$$E_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} \{u(\bullet) - \int_0^1 v(l_{jt})dj\} = -E_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} u_c y(1-\beta\eta)\{\frac{1}{2}(\varphi(1+\eta^2\beta)+\omega)\widehat{y}_t^2$$

$$- \varphi\eta\widehat{y}_t\widehat{y}_{t-1} - [\omega+\varphi(1+\beta\eta^2)]\widehat{y}_t^e\widehat{y}_t + \varphi\eta\widehat{y}_{t-1}^e\widehat{y}_t + \varphi\eta\beta E_t\widehat{y}_{t+1}^e\widehat{y}_t + \widehat{\Delta}_t\} + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3). \tag{32}$$

We seek to rewrite this expression in purely quadratic terms of the welfare-relevant gaps for inflation and output. To do so we apply the method of undetermined coefficients to reformulate the first part (all but $\widehat{\Delta}_t$), i.e. we seek to find the coefficient δ_0, such that (32) and

$$-\frac{1}{2}\delta_0(\widehat{y}_t - \widehat{y}_t^e - \delta^*(\widehat{y}_{t-1} - \widehat{y}_{t-1}^e))^2$$

$$= -\frac{1}{2}\delta_0[\widehat{y}_t^2 - 2\widehat{y}_t\widehat{y}_t^e + (\widehat{y}_t^e)^2 - 2\delta^*(\widehat{y}_t - \widehat{y}_t^e)(\widehat{y}_{t-1} - \widehat{y}_{t-1}^e) + (\delta^*)^2(\widehat{y}_{t-1}^2 - 2\widehat{y}_{t-1}\widehat{y}_{t-1}^e + (\widehat{y}_{t-1}^e)^2)]$$

$$= -\frac{1}{2}\delta_0[\widehat{y}_t^2 - 2\widehat{y}_t\widehat{y}_t^e - 2\delta^*\widehat{y}_t\widehat{y}_{t-1} + 2\delta^*\widehat{y}_t\widehat{y}_{t-1}^e + 2\delta^*\widehat{y}_t^e\widehat{y}_{t-1} + (\delta^*)^2\widehat{y}_{t-1}^2 - 2(\delta^*)^2\widehat{y}_{t-1}\widehat{y}_{t-1}^e]$$

$$= -\frac{1}{2}\delta_0[((\delta^*)^2\beta + 1)\widehat{y}_t^2 - 2\delta^*\widehat{y}_t\widehat{y}_{t-1} + 2\delta^*\widehat{y}_t\widehat{y}_{t-1}^e + 2\delta^*\beta\widehat{y}_{t+1}^e\widehat{y}_t - (2(\delta^*)^2\beta + 2)\widehat{y}_t\widehat{y}_t^e]$$

$$= -\frac{1}{2}\delta_0((\delta^*)^2\beta + 1)\widehat{y}_t^2 + \delta_0\delta^*\widehat{y}_t\widehat{y}_{t-1} - \delta_0\delta^*\widehat{y}_t\widehat{y}_{t-1}^e - \delta_0\delta^*\beta\widehat{y}_{t+1}^e\widehat{y}_t + \delta_0((\delta^*)^2\beta + 1)\widehat{y}_t\widehat{y}_t^e$$

are consistent. We use that $\widehat{y}_{t0-1}$ is $t.i.s.p.$. The parameter δ^*, $0 \leq \delta^* \leq \eta$, is the smaller root of this quadratic equation: $\eta\varphi(1+\beta\delta^2) = [\omega+\varphi(1+\beta\eta^2)]\delta$. This root is assigned to past values of the natural and efficient rate of output in their stationary solutions. Comparing coefficients, δ_0 is

$$\delta_0 = \frac{u_c y(1-\beta\eta)\eta\varphi}{\delta^*}.$$

If firms are allowed to index with past inflation, such that

$$E_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} 2\widehat{\Delta}_t = E_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} \frac{\zeta\alpha}{(1-\alpha)(1-\alpha\beta)}(\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1})^2 + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3),$$

the quadratic approximation in (32) can be written as:

$$- E_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} \frac{u_c y(1-\beta\eta)}{2}[\frac{\eta\varphi}{\delta^*}(\widehat{y}_t - \widehat{y}_t^e - \delta^*(\widehat{y}_{t-1} - \widehat{y}_{t-1}^e))^2 + \frac{\zeta\alpha}{(1-\alpha)(1-\alpha\beta)}(\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1})^2]$$

$$+ t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3).$$

The last approximation needed is that involving the utility of real money balances. Applying

23

similar techniques we get

$$z(m_t) = z + yu_c(s_m(\widehat{m}_t + \frac{1}{2}s_m(1 - \sigma_m)\widehat{m}_t^2) + t.i.s.p + \mathcal{O}(\|\widehat{\xi}_t\|^3), \tag{33}$$

where we employ $s_m = z_m m/(u_c y) = (R - 1)(1 - \beta\eta)R$ and $\sigma_m = -z_{mm}m/z_m$. Since we treat $(R - 1)/R$ as an expansion parameter, s_m and $1/\sigma_m$ are of first order. However, $s_m\sigma_m$ approaches a finite limit for $(R - 1)/R \to 0$, which is given by

$$s_m\sigma_m = \frac{z_{mm}m^2}{yu_c}.$$

The interest elasticity of money demand is given by the following expression:

$$\eta_i = -\frac{u_c(1 - \beta_n)}{z_{mm}}\frac{1 - \frac{R-1}{R}}{m} = \frac{1}{\sigma_m(R - 1)}.$$

At the limit for $(R - 1)/R \to 0$, it follows that $\eta_i = -u_c(1 - \beta\eta)/(z_{mm}m)$ and therefore $s_m\sigma_m = (1 - \beta\eta)/(v\eta_i)$, with $v = y/m$. A first-order approximation of the money demand equation (24) yields

$$\widehat{m}_t = -\eta_i\widehat{R}_t - \frac{1}{\sigma_m}\widehat{\lambda}_t + \mathcal{O}(\|\widehat{\xi}_t\|^2),$$

where

$$\widehat{\lambda}_t = -\varphi(\widehat{y}_t - \eta\widehat{y}_{t-1}) + \beta\eta\varphi(\widehat{y}_{t+1} - \eta\widehat{y}_t) + \varphi(g_t - \eta g_{t-1}) - \beta\eta\varphi(g_{t+1} - \eta g_t) + \mathcal{O}(\|\widehat{\xi}_t\|^2).$$

Using all the above we can rewrite $z(m_t)$ in the following way:

$$z(m_t) = -\frac{\eta_i yu_c}{2v}(1 - \beta\eta)(\widehat{R}_t^2 + 2\frac{R - 1}{R}\widehat{R}_t) + t.i.s.p + \mathcal{O}(\|\widehat{\xi}_t, (R - 1)/R\|^3). \tag{34}$$

We assume for simplicity that $[(R - 1)/R - 0]$ is of second order, and sum the results in expression (27) in the text.

A.2 Estimation Results

Table 6: Prior distribution of the structural parameters

Parameter	Prior distribution		
	distribution	mean	std
ρ	beta	0.8	0.1
ϕ_π	normal	1.7	0.1
ϕ_y	normal	0.125	0.05
ω	gamma	1	0.5
σ_c	normal	1.5	0.375
α	beta	0.75	0.05
η	beta	0.7	0.1
γ	beta	0.75	0.15
σ_m	normal	1.25	0.375
ψ_g	beta	0.7	0.1
ψ_a	beta	0.7	0.1
σ_g	invgamma	0.04	0.026
σ_a	invgamma	0.04	0.026
σ_μ	invgamma	0.04	0.026

Table 7: Posterior estimates of the structural parameters in each model

Parameter	$Model_1$		$Model_2$		$Model_3$		$Model_4$		$Model_5$	
	mean	std	mean	std	mean	std	mean	std	mean	std
ρ	0.4379	0.0719	0.4582	0.0727	0.3578	0.0742	0.4477	0.0730	0.3756	0.0756
ϕ_π	1.6972	0.0983	1.7188	0.0985	1.6255	0.1009	1.6719	0.0963	1.6772	0.0972
ϕ_y	0.0964	0.0263	0.0701	0.0261	0.1154	0.0324	0.0939	0.0261	0.0821	0.0297
ω	0.4170	0.1104	0.2610	0.0547	0.4156	0.0984	0.3962	0.0903	0.2904	0.0644
σ_c	1.4202	0.1744	1.4296	0.1798	1.4142	0.1879	1.4117	0.1735	1.4158	0.1779
α	0.8810	0.0139	0.8988	0.0126	0.8672	0.0177	0.8855	0.0156	0.8777	0.0168
η	-	-	0.8368	0.0566	-	-	-	-	0.8307	0.0515
γ	-	-	-	-	0.4677	0.0842	-	-	0.4563	0.0796
σ_m	-	-	-	-	-	-	1.2685	0.3866	1.2230	0.3777
ψ_g	0.8126	0.0438	0.7880	0.0579	0.7992	0.0482	0.8099	0.0477	0.7950	0.0571
ψ_a	0.7733	0.0367	0.7929	0.0372	0.6766	0.0545	0.7836	0.0403	0.6780	0.0562
σ_g	0.0095	0.0007	0.0084	0.0005	0.0097	0.0007	0.0095	0.0007	0.0084	0.0005
σ_a	0.0168	0.0030	0.0206	0.0034	0.0178	0.0032	0.0174	0.0026	0.0201	0.0039
σ_μ	0.0092	0.0006	0.0090	0.0006	0.0097	0.0007	0.0092	0.0006	0.0094	0.0006

Figure 1: Deep parameters **prior** vs. **posterior** distribution in Model 1
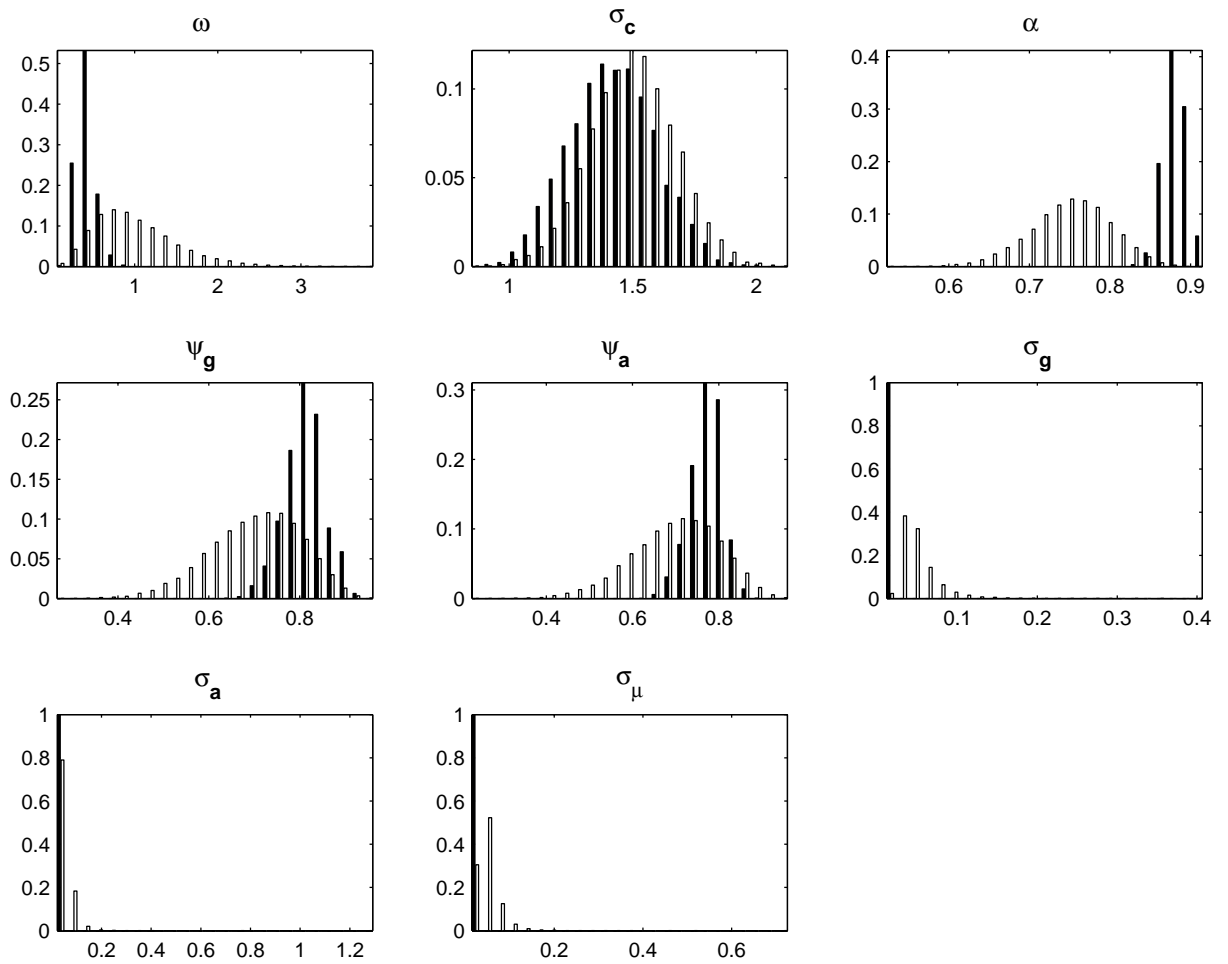
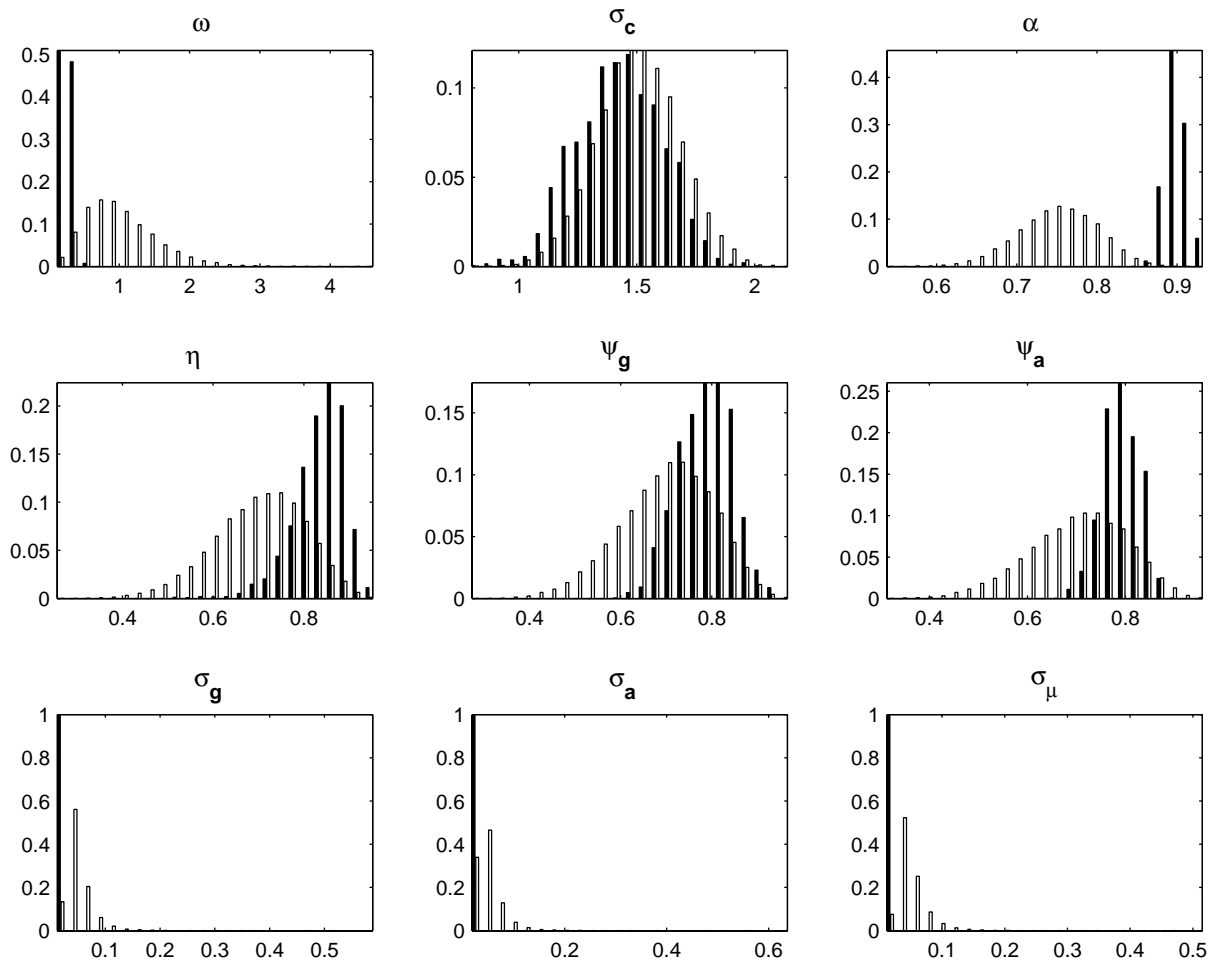Figure 2: Deep parameters prior vs. posterior distribution in Model 2

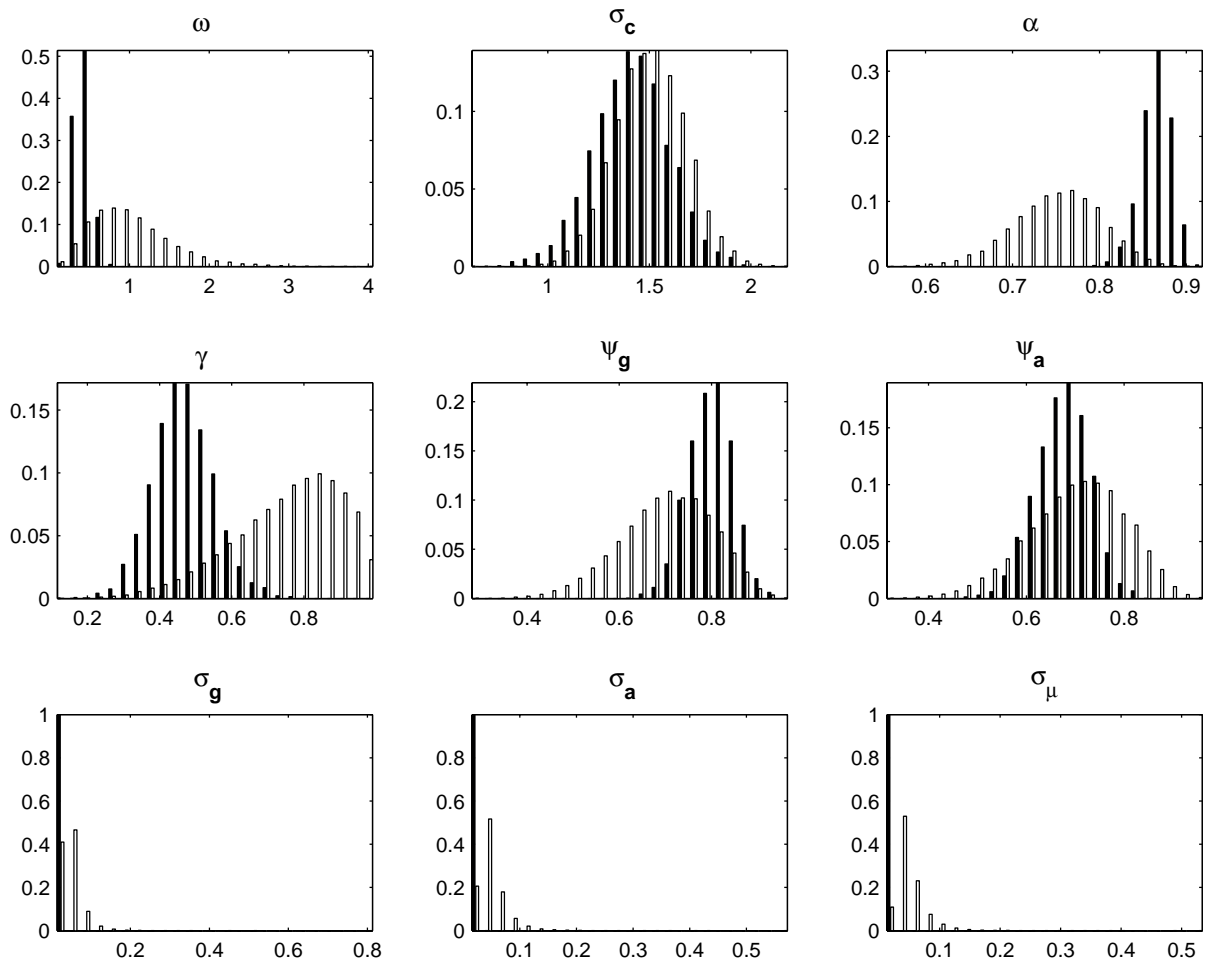Figure 3: Deep parameters **prior** vs. **posterior** distribution in Model 3

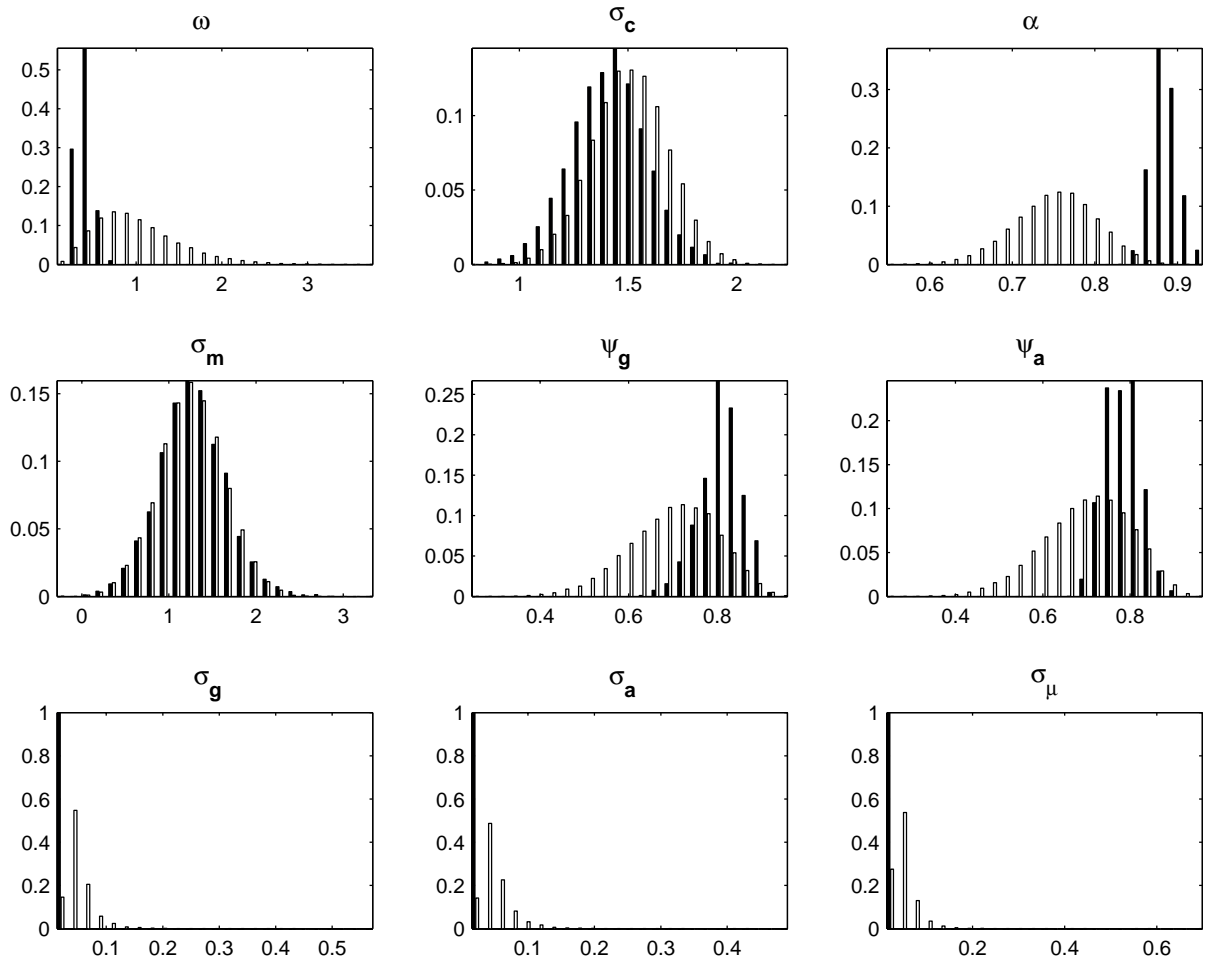Figure 4: Deep parameters prior vs. **posterior** distribution in Model 4

Figure 5: Deep parameters **prior** vs. **posterior** distribution in Model 5