UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Nested models and model uncertainty

Kriwoluzky, A.; Stoltenberg, C.A.

DOI
10.1111/sjoe.12134

Publication date
2016

Document Version
Author accepted manuscript

Published in
The Scandinavian Journal of Economics

Link to publication

Citation for published version (APA):
Kriwoluzky, A., & Stoltenberg, C. A. (2016). Nested models and model uncertainty. *The Scandinavian Journal of Economics*, *118*(2), 324-353. https://doi.org/10.1111/sjoe.12134

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:28 Jan 2022

Nested models and model uncertainty*

Alexander Kriwoluzky,† Christian A. Stoltenberg‡

October 10, 2014

Abstract

Uncertainty about the appropriate choice among nested models is a 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 sub-models, e.g., those resulting from zero restrictions. Following Sims (2008), we treat nested sub-models as "probability models" and we formalize a procedure that ensures that sub-models are not discarded too easily and do matter for optimal policy. For the United States, we find that optimal policy based on our procedure leads to substantial welfare gains compared to the standard procedure.

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

*We are especially thankful to our editor Kjetil Storesletten, two anonymous referees, Fabio Canova, Wouter Den Haan, Martin Ellison, Tom Krebs, Bartosz Maćkowiak, Peter McAdam, Morten Ravn, and Harald Uhlig. We further thank seminar participants at the Sveriges Riksbank, at the annual conferences of the Royal Economic Society and the European Economic Society as well as at the European workshop on Macroeconomics in Glasgow. Previous versions circulated under the title 'Optimal Policy under Model Uncertainty: A Structural-Bayesian Estimation Approach'.

†Department of Economics, Martin-Luther-Universität Halle-Wittenberg, 06099 Halle (Salle), Germany, email: alexander.kriwoluzky@wiwi.uni-halle.de, tel: +49 345 55233-80.

‡Department of Economics, Universiteit van Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands, email: C.A.Stoltenberg@uva.nl, tel: +31 20 525-3913.

1 Introduction

The empirical evaluation of monetary Dynamic Stochastic General Equilibrium (DSGE) models employing Bayesian methods (Castelnuovo, 2012; Smets and Wouters, 2003, 2007) has made substantial progress. Policymakers nowadays correspondingly employ estimated DSGE models, including various features and frictions, in their policy analysis. However, a central concern for policy makers is the uncertainty about the correct model which has led to the practice of using several models at the same time. The latter practice has encouraged research to find policy recommendations that perform well over a set of distinct models.

Correspondingly, model uncertainty in non-nested models has been extensively studied in the literature (see e.g. Levin and Williams, 2003; Levin, Wieland, and Williams, 2003; Kuester and Wieland, 2010). For the U.S., Levin and Williams (2003) and Levin, Wieland, and Williams (2003) study optimal policies in case of competing reference models. For the Euro area, Kuester and Wieland (2010) compare the performance of worst-case or minimax policy versus Bayesian policies over a range of different models. It is well understood that optimal policies *between* different models can be conflicting. Little attention however has been given to the uncertainty nested *within* a given model that results from the inclusion of various features and frictions.

In this paper, we seek to make two contributions. We document quantitatively that uncertainty about nested models is an important issue for policymakers in practice. Furthermore, we propose a procedure to guard against uncertainty about the appropriate choice of nested models. The procedure follows the spirit of Sims (2008) who argues that accounting for uncertainty in nested models requires a policymaker who treats nested sub-models as "probability models". Probability models are models that characterize the uncertainty incorporated in them, i.e., the probability that the models indeed generated a given set of time series (model uncertainty) and the corresponding probability distribution of the structural parameters (parameter uncertainty). A policymaker should take into account both types of uncertainty when choosing an optimal course of action. This is exactly the approach we formalize in this paper.

Starting with a baseline model, we subsequently estimate a set of competing and nested models including one model that comprises all features and frictions. This information puts us into a position to separately evaluate the gain in explanatory power of each extension. We compare two approaches to compute optimal simple rules under model uncertainty. The first approach computes optimal policy in the model that nests all features and frictions and ignores the set of nested models. As a methodological contribution, we propose a second approach that takes into account the whole set of nested models. The approach computes optimal policies over the set of nested models by weighting each model by its posterior probability. By weighting over the set of models, the policymaker can construct reasonable extensions of the baseline model and thereby insure against the pitfalls of only employing one potentially misspecified model.

In our application, we ask whether there is a quantitatively important welfare difference between the two approaches to model uncertainty for monetary policy in the United States as one example. To answer this question, we employ U.S. data and choose as a baseline model one of the most popular models employed in monetary policy: a standard cashless New Keynesian economy with staggered price-setting (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).

The baseline model and these extensions resemble a situation in which the main focus of the policymaker is on the trade-off between stabilizing fluctuations in inflation and output gap measures. In the course of the financial crisis, the stock of the monetary aggregate M1 increased from 1,375 Billions U.S. Dollars at the end of 2007 to 2,629 Billions Dollars at the end of 2013. As displayed in Figure 1, the increase was substantially higher than the increase in real GDP and the M1 velocity of income dropped by 40 percent. These changes renewed interest in the role of money for monetary policy. As an additional extension of the baseline model, we therefore consider an explicit transaction role of money captured by a money-in-utility-function specification. 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.

We compare the welfare implications of our approach with the welfare implications of policy derived from a model that nests all these features. In this environment, we find that our procedure leads to welfare gains of approximately 36 per cent compared to the optimal policy derived in a model that nests all features and frictions. We conclude that uncertainty over the choice of nested models is a quantitatively important concern for a policymaker especially in the light of the recent events.

The issue of model uncertainty in estimated DSGE models has been addressed before. Levin, Onatski, Williams, and Williams (2005) estimate a medium-scale New Keynesian Model with staggered price setting using U.S. data and determine the optimal monetary policy in that

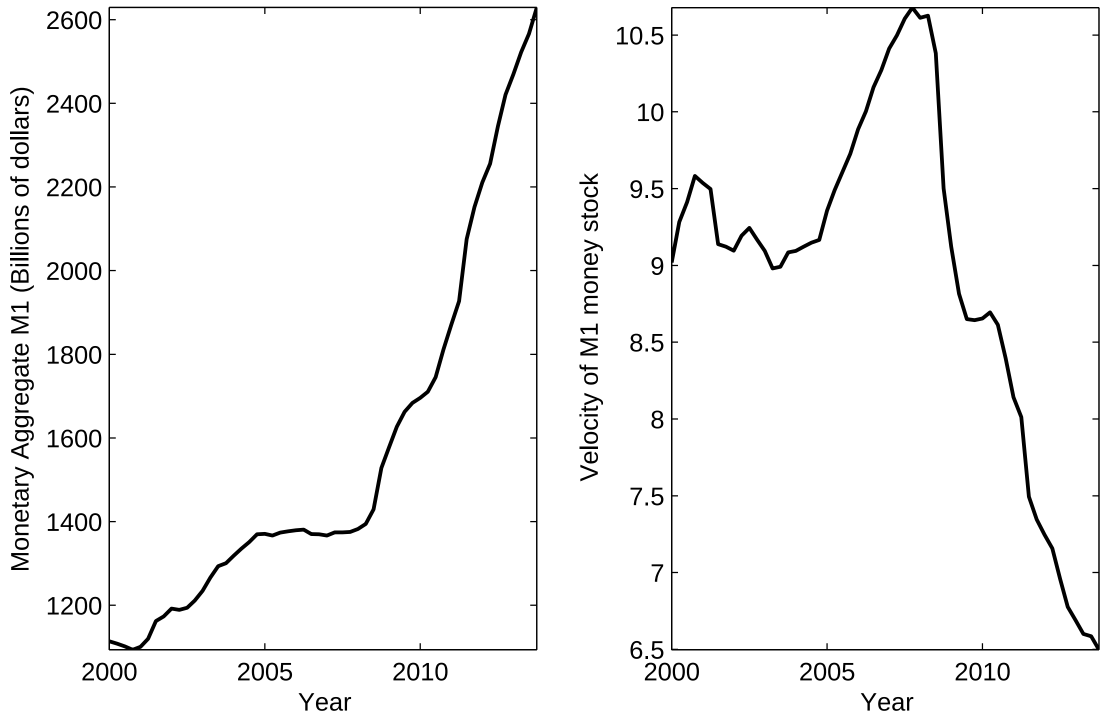

Figure 1: Money stock M1 and income velocity of M1 ($PY/M1$) for the United States.

model under parameter uncertainty. Using data for the United Kingdom, Cogley, De Paoli, Matthes, Nikolov, and Yates (2011) compute an optimal Taylor rule that performs best over a range of models. For the Euro area, Levine, McAdam, and Pearlman (2012) analyze the role of uncertainty about the degree of indexation in wages and prices for the determination of optimal interest-rate rules.

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, we start with a short description of our general setup and then we present two approaches to cope with model uncertainty in nested models. In the remainder of this section, we describe the training sample method for model comparison and how we assess policy performance under model uncertainty. The first approach to model uncertainty 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.[1] The state space form of the solution of model i is given by[2]:

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

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

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:

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

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

[1] 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.

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

equilibrium. We assume that the policymaker can credibly commit to a policy rule ϕ: when 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 equilibrium 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 is to develop and estimate one single model that nests all features and frictions and employ the model in determining optimal policy. This 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. Thus, 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 is motivated by Sims (2008). The approach starts with a stylized baseline model and then treats each extension by an additional feature or friction as a distinct and competing model. By averaging across models, this approach allows us to take not only parameter uncertainty into account but also uncertainty about model specification. 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 the corresponding posterior distribution of its structural parameters by $p_{\mathcal{M}_{\mathrm{C}}}(\theta_{\mathrm{C}}|Y)$, where Y is the set of time series used in the estimation. The optimal policy ($\phi_{\mathrm{C}}^\star$) is defined by:

$$\phi^\star_{\text{C}} = \arg\min_{\phi} \mathbb{E}_{\Theta_{\text{C}}} L(\theta_{\text{C}}, \widehat{x}) \tag{4}$$
$$s.t.\ \widehat{x}_t = T(\theta_{\text{C}}, \phi)\widehat{x}_{t-1} + R(\theta_{\text{C}}, \phi) z_t, \quad \forall \theta_{\text{C}},$$

where $\mathbb{E}_{\Theta_{\text{C}}} L(\theta_{\text{C}}, \widehat{x})$ is the expected loss when the structural parameters are a random vector and the expectation is taken with respect to the posterior distribution in Model C. 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, $\phi^\star_{\text{C}}$.

The second approach explicitly addresses specification uncertainty and averages over different models. We separately estimate a discrete set of nested models $\mathcal{M} = \{\mathcal{M}_1, ..., \mathcal{M}_{\text{C}}\}$, where $\mathcal{M}_1$ denotes the baseline model, $\mathcal{M}_{\text{C}}$ the complete model and $(\text{C} - 2)$ 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 $p_{\mathcal{M}_i}(Y) = \mathbb{E}[p_{\mathcal{M}_i}(Y|\theta_i)]$, where $p_{\mathcal{M}_i}(Y|\theta_i)$ denotes the data likelihood and the expectation is taken with respect to the prior distribution of the structural parameters in Model i, $p_{\mathcal{M}_i}(\theta_i)$.

Thus, the marginal likelihood (and the resulting posterior probability) of a model depends on its prior distribution. Since each Model i has a unique set of parameters the joint prior distribution differs across models. To take this into account and to adjust the marginal likelihoods, we follow the suggestion of Sims (2003) and use a training sample which is in more detail described in the next Subsection 2.3.

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

$$\phi^\star_a = \arg\min_{\phi} \mathbb{E}_{\mathcal{M},\Theta} L(\theta_i, \widehat{x}) \tag{5}$$
$$s.t.\ \widehat{x}_t = T(\theta_i, \phi)\widehat{x}_{t-1} + R(\theta_i, \phi) z_t, \quad \forall \theta_i, \quad i = 1, ..., n,$$

where the expectation operator averages not only across parameters using the posterior distribution in a given model but also across models using their posterior model probabilities.

The model-averaging approach assigns positive prior probability to zero restrictions on parameters that have low or even zero prior probability under the complete-model approach. Thus, the model-averaging approach can also be thought of as defining a bimodal prior distribution for the parameter that represents the additional feature or friction in the complete model. One part of the distribution is centered around a positive mode of the parameter, and the other part is centered around zero as the second mode. The complete-model approach is a limiting case of the model-averaging approach; they are equivalent only if the complete model exhibits a posterior probability of unity.

2.3 The training sample method

The weight of each model is given by its marginal likelihood. The marginal likelihood of a model also depends on its prior distribution. To take this into account, we employ the training sample method suggested by Sims (2003). The training sample method divides the whole sample Y_T into two subsamples: $Y_{T,1}$ and the training sample $Y_{T,0}$. Correspondingly, the product of the likelihood and the prior distribution of each model i, $p_{\mathcal{M}_i}(Y_T|\theta_i)p_{\mathcal{M}_i}(\theta_i)$, can be written as

$$p_{\mathcal{M}_i}(Y_T|\theta_i)p_{\mathcal{M}_i}(\theta_i) = p_{\mathcal{M}_i}(Y_{T,1}|Y_{T,0},\theta_i)p_{\mathcal{M}_i}(Y_{T,0}|\theta_i)p_{\mathcal{M}_i}(\theta_i).$$

We divide both sides by the marginal data density of the training sample:

$$p_{\mathcal{M}_i}(Y_{T,0}) = \int_{\theta_i \in \Theta_i} p_{\mathcal{M}_i}(Y_{T,0}|\theta_i)p_{\mathcal{M}_i}(\theta_i)\mathrm{d}\theta_i, \tag{6}$$

to obtain

$$\frac{p_{\mathcal{M}_i}(Y_T|\theta_i)p_{\mathcal{M}_i}(\theta_i)}{p_{\mathcal{M}_i}(Y_{T,0})} = p_{\mathcal{M}_i}(Y_{T,1}|Y_{T,0},\theta)\left[\frac{p_{\mathcal{M}_i}(Y_{T,0}|\theta_i)p_{\mathcal{M}_i}(\theta_i)}{p_{\mathcal{M}_i}(Y_{T,0})}\right].$$

The training sample method interprets the term $\left[\frac{p_{\mathcal{M}_i}(Y_{T,0}|\theta_i)p_{\mathcal{M}_i}(\theta_i)}{p_{\mathcal{M}_i}(Y_{T,0})}\right]$ as the prior distribution for the likelihood $p_{\mathcal{M}_i}(Y_{T,1}|Y_{T,0}\theta_i)$. Note that

$$\int_{\theta_i \in \Theta_i}\left[\frac{p_{\mathcal{M}_i}(Y_{T,0}|\theta)p_{\mathcal{M}_i}(\theta_i)}{p_{\mathcal{M}_i}(Y_{T,0})}\right]\mathrm{d}\theta = 1. \tag{7}$$

According to (7), each model therefore exhibits the same prior probability after the training sample. Hence, applying the training sample method prevents the manipulation of the marginal data density of each model by choosing prior distributions. Consequently, the marginal data density of each model is not influenced by the fact that the prior distribution allocates different prior probabilities to different models. Instead, the resulting posterior probabilities of the models are only driven by the additional evidence provided by the data $Y_{T,1}$. In our analysis, we report the marginal data density of model i corrected by the marginal data density of the training sample. The corrected marginal data density is given by the following expression

$$\tilde{p}_{\mathcal{M}_i}(Y_{T,1}) = \frac{\int_{\theta_i \in \Theta_i} p_{\mathcal{M}_i}(Y_T|\theta_i) p_{\mathcal{M}_i}(\theta_i) \mathrm{d}\theta_i}{p_{\mathcal{M}_i}(Y_{T,0})} \tag{8}$$

The posterior probability of model i is our key statistic to assess the explanatory power of each model. It is defined as the ratio of the (corrected) logged marginal data densities of the model over the sum of the logged marginal data densities of all models:

$$\pi_{\mathcal{M}_i} = \frac{\log(\tilde{p}_{\mathcal{M}_i}(Y_{T,1}))}{\sum_{j=1}^{\mathrm{C}} \log(\tilde{p}_{\mathcal{M}j}(Y_{T,1}))}. \tag{9}$$

2.4 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. In the next step, we express utility as reduction in certainty consumption equivalents by setting the latter expression 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 $\tilde{\phi}$.

3 Economic environment

To demonstrate our main result, we create a set of distinct monetary models including one model that nests all features and frictions. Starting with a standard 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

Our baseline economy is a textbook cashless New Keynesian economy with representative households and Calvo (1983) staggered price setting as laid down in Woodford (2003a).[3] The dynamics in the economy are described by an Euler equation and an aggregate supply curve:

$$\sigma(\mathbb{E}_t\,\widehat{y}_{t+1} - \mathbb{E}_t\,\widehat{y}^n_{t+1}) = \sigma(\widehat{y}_t - \widehat{y}^n_t) + \widehat{R}_t - \mathbb{E}_t\,\widehat{\pi}_{t+1} - \widehat{R}^n_t \tag{10}$$

$$\widehat{\pi}_t = \beta\,\mathbb{E}_t\,\widehat{\pi}_{t+1} + \kappa(\widehat{y}_t - \widehat{y}^n_t), \tag{11}$$

where $\sigma = -u_{cc}y/(u_c) > 0$, and $\omega = v_{ll}l/v_l > 0$ are the coefficient of relative risk aversion and the inverse of the Frisch elasticity of labor supply in the deterministic steady state with zero inflation. Further, $\kappa = (1-\alpha)(1-\alpha\beta)(\omega+\sigma)/\alpha$, where $\alpha > 0$ is the Calvo parameter and $0 < \beta < 1$ is the subjective discount factor of the representative household. 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, $\widehat{y}^n_t, \widehat{R}^n_t$, i.e the values for output and real interest under flexible prices, are given by the following expressions

$$\widehat{y}^n_t = \frac{(1+\omega)\widehat{a}_t + \sigma\tilde{g}_t - \widehat{\mu}_t}{\omega+\sigma}, \quad \widehat{R}^n_t = \sigma[(\tilde{g}_t - \widehat{y}^n_t) - \mathbb{E}_t(\tilde{g}_{t+1} - \widehat{y}^n_{t+1})],$$

[3] A detailed description can be also found in an accompanying working paper version.

where $\tilde{g}_t = (g_t - g)/y$ is a shock to government expenditures, $\widehat{a}_t$ is a shock to the productivity of labor as the only production factor, and $\widehat{\mu}_t$ denotes a wage-markup shock. The model is closed by a simple interest rate feedback rule:

$$\widehat{R}_t = \phi_\pi \widehat{\pi}_t + \phi_y \widehat{y}_t + \widehat{r}_t, \tag{12}$$

where $\widehat{r}_t$ is a monetary policy shock.[4] Other variables like for example real wages $\widehat{w}_t$ can be determined residually by static relationships

$$\widehat{w}_t = \widehat{\mu}_t + \omega(\widehat{y}_t - \widehat{a}_t) - \widehat{\lambda}_t,$$

where $\widehat{\lambda}_t$ is the Lagrange multiplier associated with households' budget constraint. In the baseline economy, the multiplier is given by $\sigma(\widehat{y}_t - \tilde{g}_t)$. The general system (1) in the baseline model then is the fundamental locally stable and unique solution that satisfies (10)-(12) for a certain vector of constant policy coefficients $\phi = (\phi_\pi, \phi_y)$.

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

$$L(\theta_1, \widehat{x}) = \frac{u_c y \zeta(\omega + \sigma)}{2\kappa} \{\mathrm{var}(\widehat{\pi}_t) + \lambda_d \,\mathrm{var}(\widehat{y}_t - \widehat{y}_t^e)\}, \tag{13}$$

where $\lambda_d = \kappa/\zeta$ with ζ as the elasticity of substitution between differentiated goods, 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.

[4] In Appendix A.3.1, we also analyze optimal simple rules that respond to deviations of output from its level under flexible prices and the efficient level.

[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' consumption. The constituting equations for (1) are the policy rule (12) and the modified versions of the Euler equation and the New Keynesian Phillips curve:

$$\varphi[d_t - \eta d_{t-1}] - \varphi\beta\eta\,\mathbb{E}_t[d_{t+1} - \eta d_t] = \mathbb{E}_t\,\widehat{\pi}_{t+1} + \widehat{R}_t^n - \widehat{R}_t ...$$
$$+ \mathbb{E}_t\,\varphi[d_{t+1} - \eta d_t] - \varphi\beta\eta\,\mathbb{E}_t[d_{t+2} - \eta d_{t+1}] \tag{14}$$

$$\widehat{\pi}_t = \kappa_h[(d_t - \delta^* d_{t-1}) - \beta\delta^*\,\mathbb{E}_t(d_{t+1} - \delta^* d_t)] + \beta\,\mathbb{E}_t\,\widehat{\pi}_{t+1}, \tag{15}$$

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

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

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

$$L(\theta_2, \widehat{x}) = \frac{(1-\beta\eta)\eta\varphi u_c^h y^h \zeta}{2\kappa_h\delta^*}\{\mathrm{var}(\widehat{\pi}_t) + \lambda_{d,h}\,\mathrm{var}(\widehat{y}_t - \widehat{y}_t^e - \delta^*(\widehat{y}_{t-1} - \widehat{y}_{t-1}^e))\}, \tag{16}$$

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

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

Like habit formation, the indexation of prices to past inflation induces the economy to evolve in a history-dependent way. We assume that the differentiated goods prices that are not reconsidered adjust 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. With indexation, the economy is characterized by a modified aggregate supply curve

$$\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1} = \beta\,\mathbb{E}_t(\widehat{\pi}_{t+1} - \gamma\widehat{\pi}_t) + \kappa(\widehat{y}_t - \widehat{y}_t^n), \tag{17}$$

[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.

(10) and (12). The corresponding loss function of the central bank reads (Woodford, 2003a):

$$L(\theta_3, \widehat{x}) = \frac{u_c y \zeta(\omega+\sigma)}{2\kappa}\{\mathrm{var}(\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1}) + \lambda_d \,\mathrm{var}(\widehat{y}_t - \widehat{y}_t^e)\}, \tag{18}$$

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)

To capture the increasing role of money in the course of the financial crisis, 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 m_t is additively augmented by the amount $z(m_t)$ and a demand equation for real money balances enters the set of equilibrium conditions. The utility associated with real money balances, $z(m_t)$, is assumed to be increasing and strictly concave. In log-linearized form this additional equilibrium condition is given by:

$$\widehat{m}_t = -\eta_R \widehat{R}_t - \frac{1}{\sigma_m}\widehat{\lambda}_t, \tag{19}$$

where $\sigma_m = -z_{mm}m/z_m$ and $\eta_R = \sigma_m\beta/(1-\beta)$ is the gross-interest rate elasticity. The stabilization loss in Model 4 is given by:

$$L(\theta_4, \widehat{x}) = \frac{u_c y \zeta(\omega+\sigma)}{2\kappa}\{\mathrm{var}(\widehat{\pi}_t) + \lambda_d \,\mathrm{var}(\widehat{y}_t - \widehat{y}_t^e) + \lambda_{1R}\,\mathrm{var}(\widehat{R}_t)\}, \tag{20}$$

where $\lambda_d = \kappa/\zeta$, $\lambda_{1R} = \lambda_d\eta_R/[v(\omega+\sigma)]$ and $v = y/m$. The general form (1) has to satisfy the equations (10)-(12) and (19).

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: (14), (12), (19) and

$$\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1} = \beta\,\mathrm{E}_t(\widehat{\pi}_{t+1} - \gamma\widehat{\pi}_t) + \kappa_h[(d_t - \delta^* d_{t-1}) - \beta\delta^*\,\mathrm{E}_t(d_{t+1} - \delta^* d_t)]. \tag{21}$$

In appendix A.1, we show that the stabilization loss for Model C is given by the following expression

$$L(\theta_{\mathrm{C}}, \widehat{x}) = \frac{(1-\beta\eta)\eta\varphi u_c^h y^h \zeta}{2\kappa_h \delta^*}\{\mathrm{var}(\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1}) + \lambda_{d,h}\,\mathrm{var}(\widehat{y}_t - \widehat{y}_t^e - \delta^*(\widehat{y}_{t-1} - \widehat{y}_{t-1}^e)) + \lambda_{2R}\,\mathrm{var}(\widehat{R}_t)\}, \tag{22}$$

$$\lambda_{d,h} = \kappa_h/\zeta, \quad \lambda_{2R} = \frac{\lambda_{d,h}\delta^*\eta_R}{\upsilon\eta\varphi}, \tag{23}$$

and $\upsilon = y/m > 0$.

4 Results

In this section, we first present the estimation results. In the second part, 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, prior distribution and calibrated parameters

The choice of observable variables is guided by the structural shocks of the models. Due to the wage mark-up shock we choose real wages. Since our models further include a demand, a supply, and a monetary policy shock, we employ output, inflation and the federal funds rate as observable variables. The data consists of quarterly values of these variables for the U.S. from the first quarter 1983 until the second quarter of 2008.[7] In our estimation, we de-trend the data before estimating the models by removing a linear quadratic time trend.

The observation equation linking the de-trended observable variables $\left(\widehat{y}_t^d, \widehat{\pi}_t^d, \widehat{R}_t^d, \widehat{w}_t^d\right)$ to the variables of the model is given by:

$$\begin{bmatrix} \widehat{y}_t^d \\ \widehat{\pi}_t^d \\ \widehat{R}_t^d \\ \widehat{w}_t^d \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \widehat{y}_t \\ \widehat{\pi}_t \\ \widehat{R}_t \\ \widehat{w}_t \end{bmatrix}. \tag{24}$$

The specification of the prior distribution of the estimated parameters closely follows Smets and Wouters (2007) and is displayed in Table 1. We employ the same prior distributions for

[7] A detailed description of the data is given in Appendix A.2. We choose a training sample of 50 quarters.

parameters common to all model extensions. For α, η, γ, we employ a relatively uninformative beta-distribution with mean 0.5 and standard distribution of 0.2. Relative risk aversion σ_c is assumed to be distributed around 1, i.e. a logarithmic specification. For the inverse of the Frisch elasticity (ω) we assume the same prior distribution. While we assume the disturbances $\tilde{g}_t$, $\widehat{a}_t$ and $\widehat{r}_t$ to follow stationary $AR(1)$ processes, $\widehat{\mu}_t$ is supposed to be *i.i.d.*. In particular, the prior distributions for the set of coefficients that describe the shock processes, ψ_g, ψ_a, ψ_r and $\sigma_g, \sigma_a, \sigma_\mu, \sigma_r$, do not change across models, and they are specified according to the procedure explained in Section 2.4. Two of the parameters are not identifiable given the employed dataset. These are the the elasticity of substitution between differentiated goods (ζ) and the money-demand parameter (σ_m). Nevertheless, we specify a distribution for each parameter to capture the uncertainty of the policymaker. We define the distribution of ζ as a normal distribution around mean 6, which is a value typically assumed in the literature (see Woodford, 2003a). For the money-demand parameter, we specify a normal distribution around the mean 1.92, implying a net-interest rate elasticity, $-\partial \log m_t / \partial \log(R_t - 1) = \eta_R / R$ equal to 0.5 at the mean. This is the value estimated by Lucas (2000) for U.S. data ranging from 1900 to 1994.[8]

The remaining parameters are calibrated to match steady-state values, which are directly observable from the data. The discount factor is calibrated to $\beta = 0.99$ to fit a quarterly real-interest rate of 1.01%. We set the steady-state fraction of private consumption relative to GDP to its mean in the raw data, $c/y = 0.8$. The income velocity of money, v, is calibrated to a value of 5.88 which is the average post-war-income velocity based on $M1$ for the U.S. (see e.g., Schmitt-Grohé and Uribe, 2007). Our data set does not comprise the years of the financial crisis. Nevertheless, we want to capture the decreasing trend of the velocity observed after 2008. For this reason, we employ the average post-war velocity of M1 which is lower than the average velocity found if we were to restrict attention to our data set.

[8] In Appendix A.3.2, we also report results for a distribution around a lower net-interest rate elasticity as estimated by Ireland (2009).

Table 1: Prior distribution of the structural parameters

Parameter	distribution	mean	std
ϕ_π	normal	1.7	0.3
ϕ_y	normal	0.125	0.05
ω	gamma	1	0.25
σ_c	gamma	1	0.25
α	beta	0.5	0.2
η	beta	0.5	0.2
γ	beta	0.5	0.2
σ_m	normal	1.92	0.2
ζ	normal	6	0.5
ψ_g	beta	0.5	0.2
ψ_a	beta	0.5	0.2
ψ_r	beta	0.5	0.2
σ_g	invgamma	0.02	inf
σ_a	invgamma	0.02	inf
σ_μ	invgamma	0.02	inf
σ_r	invgamma	0.02	inf

For each model, we employ the following procedure. First, we obtain the estimation results by approximating the posterior mode. Afterwards we employ a random walk Metropolis-Hastings algorithm to evaluate the posterior distribution. Each time, we run two chains, each chain consisting of 500,000 draws. We discard the first 400,000 draws of each chain and compute the reported statistics based on the remaining ones.[9] The marginal data density of each model is then computed based on 100,000 draws using the modified harmonic mean estimator as proposed by Geweke (1999a).

4.2 Model comparison and estimation results

The estimation results are displayed in Table 2. The estimation results are comparable to the ones found by related studies (see e.g., Smets and Wouters, 2007). In line with conventional wisdom, we find that U.S. monetary policy was aggressively responding to deviations of inflation from target. The estimate of the Calvo parameter implies that prices are reset roughly every fourth quarter. Furthermore, the estimates of σ_c and ω imply that the slope of the Philipps curve is well below unity for all models under consideration. While the posterior mean of the indexation parameter is above the prior mean, the habit parameter is estimated close to zero. As expected, the parameters ζ and σ_m cannot be identified and the posterior distribution is

[9] To check for convergence, we apply the convergence statistics suggested by Brooks and Gelman (1998). The statistics indicate convergence after 100,000 draws for all estimated regions and sample periods.

Table 2: Posterior estimates of the structural parameters in each model

	$Model_{1,4}$		$Model_2$		$Model_3$		$Model_C$	
Parameter	Mean	Std	mean	std	Mean	Std	Mean	Std
ϕ_π	3.2858	0.2267	3.2716	0.2274	3.2054	0.2182	3.1889	0.2186
ϕ_y	0.0864	0.0247	0.0829	0.0244	0.0948	0.0274	0.0905	0.0271
ω	1.1315	0.1813	1.0675	0.1749	1.0176	0.1618	0.9608	0.1568
σ_c	1.5331	0.2264	1.6053	0.2325	1.7457	0.2468	1.8221	0.2531
α	0.6693	0.0213	0.6708	0.0212	0.5962	0.0292	0.5976	0.0290
η	-	-	0.0437	0.0219	-	-	0.0505	0.0249
γ	-	-	-	-	0.7528	0.1348	0.7580	0.1335
ψ_g	0.8772	0.0210	0.8726	0.0213	0.8683	0.0223	0.8637	0.0226
ψ_a	0.9876	0.0057	0.9871	0.0058	0.9878	0.0056	0.9876	0.0056
ψ_r	0.6221	0.0412	0.6217	0.0410	0.7343	0.0479	0.7319	0.0480
σ_g	0.0085	0.0008	0.0083	0.0007	0.0081	0.0007	0.0080	0.0007
σ_a	0.0030	0.0002	0.0030	0.0002	0.0033	0.0002	0.0034	0.0003
σ_μ	0.0100	0.0009	0.0099	0.0009	0.0089	0.0008	0.0088	0.0008
σ_r	0.0031	0.0002	0.0031	0.0002	0.0028	0.0002	0.0028	0.0002

Notes: $\mathcal{M}_1$ is the benchmark model, $\mathcal{M}_2$ benchmark model plus habit, $\mathcal{M}_3$ is the benchmark model plus indexation, $\mathcal{M}_4$ is the benchmark model plus money, $\mathcal{M}_C$ is the benchmark plus habit, indexation and money.

equal to the prior distribution. We therefore do not report the distribution in Table 2 but employ their prior distribution in the analysis of optimal policy.

To assess the explanatory power of each model, we compute marginal likelihoods and the corresponding posterior probabilities using the training-sample method described in detail in Section 2.3. Table 3 presents the results on model comparison.[10] The key result here 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 (Model 4) does not increase the marginal likelihood: real money balances do not help to predict the observable variables. The posterior probability of Model 1 and Model 4 is therefore identical. Second, although the prior for the habit parameter is chosen in favor of Model 2, a habit in consumption does not improve the fit to the data. Instead, the additional parameter introduces additional uncertainty into the estimation and thus reduces the marginal likelihood slightly. Third, history dependence in inflation improves the fit of the model. With approximately 82 per cent, Model 3 exhibits the highest posterior probability. Thus, the complete model incorporates features that help 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.

[10] As a robustness test, we also derive and compute Savage–Dickey density ratios as proposed by Verdinelli and Wasserman (1995). The results for this exercise can be found in Appendix A.3.3.

Table 3: Log marginal data densities and model probabilities

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_{\text{C}}$
$\log(\tilde{p}_{\mathcal{M}_i}(Y_{T,1}))$	806.60	804.95	816.99	806.60	815.46
$\pi_{\mathcal{M}_i}$	0.00%	0.00%	82.10%	0.00%	17.90%

Notes: $\mathcal{M}_1$ is the benchmark model, $\mathcal{M}_2$ benchmark model plus habit, $\mathcal{M}_3$ is the benchmark model plus indexation, $\mathcal{M}_4$ is the benchmark model plus money, $\mathcal{M}_{\text{C}}$ is the benchmark plus habit, indexation and money.

In the next section, we analyze optimal policies in and across models.

4.3 Optimal policy at the posterior mean

To establish a standard and to explain the stabilization trade-off, we determine the optimal policy $\bar{\phi}_i^\star = (\phi_\pi^\star, \phi_y^\star)_i$ at the posterior mean $\bar{\theta}_i$ for each Model i, $i = 1, 2, .., \text{C}$. The optimal coefficients and the resulting business cycles costs ($\mathcal{BC}$) expressed as equivalent reductions in steady-state consumption are displayed in Table 4.[11]

Table 4: Optimal policy at the posterior mean ($\bar{\phi_i}^\star$)

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_{\text{C}}$
$\phi_\pi^\star$	20.87	21.93	15.65	4.72	3.62
$\phi_y^\star$	0.07	0.07	0.16	0.09	0.09
$\mathcal{BC}(\bar{\theta}_i, \phi_i^\star)$	0.0017%	0.0016%	0.0012%	0.0122%	0.0084%

Notes: $\mathcal{M}_1$ is the benchmark model, $\mathcal{M}_2$ benchmark model plus habit, $\mathcal{M}_3$ is the benchmark model plus indexation, $\mathcal{M}_4$ is the benchmark model plus money, $\mathcal{M}_{\text{C}}$ is the benchmark plus habit, indexation and money.

The optimal coefficients of the interest-rate rule displayed in Table 4 reflect the relative importance of the different stabilization goals in each model.[12] In the first three models inflation stabilization is the predominant aim. This in turn implies that if the uncertainty about nested models were just about the different cashless models, possible welfare losses of it would be small. In Models 4 and C, households value real money balances as a medium for transactions. The resulting demand for cash introduces the stabilization of the nominal interest rate as an additional and conflicting aim to price stability (see equations (20) and (26)). To see why fluctuations in interest and in inflation are conflicting, suppose that ϕ_y is small and that the

[11] The loss functions are relatively flat at the optimum such that optimal coefficients have to move long way without a significant change in loss. For this reason, we imposed a limit of 25 on the response coefficients on inflation and output.

[12] When computing the optimal policy rule, we treat the variance σ_ϵ as exogenous to the policymaker's decision.

economy in Model 4 is hit by a wage-markup shock. To fight inflationary tendencies the output gap must decrease according to the aggregate supply curve (11). This in turn requires a strong increase in the nominal interest rate to fulfill the Euler equation (10). While fluctuations in the nominal interest rate are not welfare relevant in Models 1,2 and 3, they are welfare reducing in the models with a demand for cash (compare e.g, (13) to (20)). Therefore, optimal policies in models with a transaction friction exhibit 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.

Table 5: Stabilization weights λ_d and λ_R at the posterior mean

Weights	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_{\text{C}}$
λ_d	0.09	0.09	0.15	0.09	0.15
λ_R	–	–	–	0.24	0.40

Notes: $\mathcal{M}_1$ is the benchmark model, $\mathcal{M}_2$ benchmark model plus habit, $\mathcal{M}_3$ is the benchmark model plus indexation, $\mathcal{M}_4$ is the benchmark model plus money, $\mathcal{M}_{\text{C}}$ is the benchmark plus habit, indexation and money.

Table 5 illustrates the importance of stabilizing the output gap and the nominal interest rate relative to inflation.[13] Inflation stabilization is the predominant aim in all models: even in the Model with the highest weight on stabilizing something other than inflation (λ_R in Model C), price stability is more than twice as important as interest-rate stabilization. Furthermore, Table 5 shows that stabilization aims for households change across models. For example, the stabilization of the output gap is more important in Model C than in Model 1. The exact gap that policy should stabilize to maximize welfare differs as well (see, e.g., equations (13) and (26)). 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

[13] The weight on output-gap fluctuations is higher than calibrated in Woodford (2003a). The reason is that we estimate a steeper Phillips curve, i.e. a higher κ.

important in Model C than in Model 4.[14]

4.4 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 equation (4), which yields

$$\phi^{\star}_{\mathrm{C}}: \quad \phi_{\pi} = 3.66; \quad \phi_{y} = 0.07.$$

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, the real wage and the federal funds rate (see Table 3). 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 (5):

$$\phi^{\star}_{a}: \quad \phi_{\pi} = 5.99; \quad \phi_{y} = 0.11.$$

Comparing the characteristics of the two rules reveals one similarity and one difference. Both rules put a similar 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. As a consequence, the optimal rule resulting from a model-averaging approach reacts stronger to inflation to avoid welfare losses in Model 3 than the optimal rule derived in Model C.

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. $\mathcal{BC}(\Theta_i, \phi^{\star}_{\mathrm{C}})$ and $\mathcal{BC}(\Theta_i, \phi^{\star}_{a})$ for $i = 1, 2, 3, 4, \mathrm{C}$.

[14] The net interest-rate elasticity of money demand is an important factor for the magnitude of λ_R (see its definition in equation (27)). To account for this, we also employ an alternative calibration which is based on money-demand estimates by Ireland (2009). The corresponding stabilization weights and the results for the optimal rules can be found in Appendix A.3.2.

Table 6: Relative performance of $\phi_{\mathrm{C}}^{\star}$ and $\phi_a^{\star}$

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_{\mathrm{C}}$	$\overline{\mathcal{BC}}$
$\mathcal{BC}(\Theta_i, \phi_{\mathrm{C}}^{\star})/\mathcal{BC}(\Theta_i, \phi_a^{\star})$	1.83	1.84	1.46	1.02	0.91	1.36

Notes: $\mathcal{M}_1$ is the benchmark model, $\mathcal{M}_2$ benchmark model plus habit, $\mathcal{M}_3$ is the benchmark model plus indexation, $\mathcal{M}_4$ is the benchmark model plus money, $\mathcal{M}_{\mathrm{C}}$ is the benchmark plus habit, indexation and money.

As can be seen from Table 6, the optimal rule $\phi_a^{\star}$ performs substantially better than $\phi_{\mathrm{C}}^{\star}$ 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 4), it avoids high welfare losses in Models C and 4. On average, optimal policy derived from the model-averaging approach leads to welfare gains of 36 per cent relative to the optimal policy rule derived by the complete-model approach.

Alternatively, the policymaker could decide to follow policy prescriptions derived in the most likeliest model (Model 3). But this cannot be recommended either: this policy rule fails to prevent high stabilization losses in Model C, and it leads to an increase in business cycle costs of over 20 percent on average compared to the optimal policy under model uncertainty. This means that even if a certain friction does not add any explanatory power it may be nevertheless relevant for welfare considerations and for the assessment of optimal policies. Summing up, optimal policy should take into account both, possible high stabilization losses and their probability of their occurrence.

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 a 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 U.S. data, we find that the model-averaging approach to optimal monetary policy leads to welfare gains of approximately

36 per cent compared to the optimal policy using one single model only.

A Appendix

A.1 Approximation of households'utility in Model c

In the following, we collect the exogenous disturbances in the vector $\xi_t = [a_t, g_t, \mu_t]$.

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:*

$$U_{t_0} = -E_{t_0} \sum_{t=t_0}^{\infty} \beta^{t-t_0} L(\theta_{\mathrm{C}}, \widehat{x}) + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t, (R-1)/R\|^3), \tag{25}$$

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

$$L(\theta_{\mathrm{C}}, \widehat{x}) = \frac{(1-\beta\eta)\eta\varphi u_c^h y^h \zeta}{2\kappa_h \delta^*} \{var(\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1}) + \lambda_{d,h} var(\widehat{y}_t - \widehat{y}_t^e - \delta^*(\widehat{y}_{t-1} - \widehat{y}_{t-1}^e)) + \lambda_{2R} var(\widehat{R}_t)\}, \tag{26}$$

$$\lambda_{d,h} = \kappa_h/\zeta, \quad \lambda_{2R} = \frac{\lambda_{d,h}\beta\delta^*}{v\sigma_m(1-\beta)\eta\varphi}, \tag{27}$$

and $v = y/m > 0$.

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

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

which is assumed to be increasing in consumption and real money balances, decreasing in labor time, strictly concave, twice differentiable, and to fulfill the Inada conditions. To derive (25) 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:

$$\begin{aligned} u(y_t - g_t - \eta(y_{t-1} - g_{t-1})) &= u_c y(1-\beta\eta)[\widehat{y}_t + \frac{1}{2}(1 - \varphi(1+\eta^2\beta))\widehat{y}_t^2 + \varphi\eta\widehat{y}_t\widehat{y}_{t-1} \\ &+ \varphi\widehat{y}_t(-\eta\tilde{g}_{t-1} - \beta\eta\,\mathbb{E}_t\,\tilde{g}_{t+1} + (1+\eta^2\beta)\tilde{g}_t)] + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3), \end{aligned} \tag{28}$$

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, $\sigma = -yu_{11}/u_1$, and $\tilde{g}_t = (g_t - g)/y$. The term $\widehat{y}_t\,\mathbb{E}\,g_{t+1}$ appears because households maximize the discounted sum of their period utility function such that current income appears in the current and in the first future period.

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

$$v(l_t) = u_c(1-\beta\eta)[\widehat{y}_t + \frac{1+\omega}{2}\widehat{y}_t^2 - (1+\omega)\widehat{a}_t\widehat{y}_t + \widehat{\Delta}_t] + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3), \tag{29}$$

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 (28) and (29), employ $\breve{g}_t = -\eta\tilde{g}_{t-1} - \beta\eta E_t\tilde{g}_{t+1} + (1+\eta^2\beta)\tilde{g}_t$, and obtain:

$$\begin{aligned} u(y_t - g_t - \eta(y_{t-1} - g_{t-1})) - \int_0^1 v(l_{jt})dj = u_c y(1-\beta\eta)[\frac{1}{2}(-\varphi(1+\eta^2\beta) - \omega)\widehat{y}_t^2 \\ + \varphi\eta\widehat{y}_t\widehat{y}_{t-1} + \varphi\widehat{y}_t\breve{g}_t + (1+\omega)\widehat{a}_t\widehat{y}_t - \widehat{\Delta}_t] + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3). \end{aligned} \tag{30}$$

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

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

If we use this expression to rewrite (30), we obtain the following:

$$\begin{aligned} \mathbb{E}_{t_0}\sum_{t=t_0}^{\infty}\beta^{t-t_0}\{u(\bullet) - \int_0^1 v(l_{jt})dj\} = -\,\mathbb{E}_{t_0}\sum_{t=t_0}^{\infty}\beta^{t-t_0}u_c y(1-\beta\eta)\{\frac{1}{2}(\varphi(1+\eta^2\beta)+\omega)\widehat{y}_t^2 ... \\ -\varphi\eta\widehat{y}_t\widehat{y}_{t-1} - [\omega + \varphi(1+\beta\eta^2)]\widehat{y}_t^e\widehat{y}_t + \varphi\eta\widehat{y}_{t-1}^e\widehat{y}_t + \varphi\eta\beta\,\mathbb{E}_t\,\widehat{y}_{t+1}^e\widehat{y}_t + \widehat{\Delta}_t\} \\ + t.i.s.p. + \mathcal{O}(\|\widehat{\xi}_t\|^3). \end{aligned} \tag{31}$$

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 (31) and

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

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

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

the quadratic approximation in (31) can be written as:

$$
\begin{aligned}
-\mathbb{E}_{t_0}\sum_{t=t_0}^{\infty}\beta^{t-t_0}\frac{u_c y(1-\beta\eta)}{2}[\frac{\eta\varphi}{\delta^*}(\widehat{y}_t-\widehat{y}_t^e-\delta^*(\widehat{y}_{t-1}-\widehat{y}_{t-1}^e))^2+\frac{\zeta\alpha}{(1-\alpha)(1-\alpha\beta)}(\widehat{\pi}_t-\gamma\widehat{\pi}_{t-1})^2]\\
+t.i.s.p.+\mathcal{O}(\|\widehat{\xi}_t\|^3).
\end{aligned}
$$

The last approximation needed is the one involving the utility of real money balances. Applying similar techniques we get

$$z(m_t)=z+yu_c(s_m(\widehat{m}_t+\frac{1}{2}s_m(1-\sigma_m)\widehat{m}_t^2)+t.i.s.p+\mathcal{O}(\|\widehat{\xi}_t\|^3), \tag{32}$$

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

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

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 (19) yields

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

where

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

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

$$z(m_t) = -\frac{\eta_i y u_c}{2v}(1-\beta\eta)(\widehat{R}_t^2 + 2\frac{R-1}{R}\widehat{R}_t) + t.i.s.p + \mathcal{O}(\|\widehat{\xi}_t, (R-1)/R\|^3). \tag{33}$$

We assume for simplicity that $[(R-1)/R-0]$ is of second order, and sum the results in expression (25) in the text. ■

A.2 Data appendix

The frequency of all data used is quarterly.

Real GDP: This series is *BEA NIPA table 1.1.6 line 1.*

Nominal GDP: This series is: *BEA NIPA table 1.1.5 line 1.*

Implicit GDP Deflator: The implicit GDP deflator is calculated as the ratio of Nominal GDP to Real GDP.

Civilian noninstitutional population: This series is taken from:

http://research.stlouisfed.org/fred2/series/CNP16OV?cid=104.

Nominal hourly wages total private industry: This series is $AHETPI$ obtained from Fred:

http://research.stlouisfed.org/fred2/series/ahetpi/10.

Interest rates: This series is the effective federal funds rate obtained from Fred:

http://research.stlouisfed.org/fred2/series/FEDFUNDS.

A.3 Robustness

In this section, we provide robustness tests for our main result that optimal policies determined by our model-averaging approach are preferable to optimal policies derived in the complete model. First, we consider simple rules that respond to deviations of output from the level under flexible prices or from the efficient level instead of responding to trend output. Second, we analyze how changes in the interest-rate elasticity of money demand affect the welfare comparison between the two approaches to model uncertainty. The third subsection investigates robustness of the model comparison by using Savage–Dickey density ratios instead of a model comparison exercise based on the marginal livelihoods estimated by the modified harmonic mean estimator.

A.3.1 Feedback on natural and efficient output

In our baseline specification, we employ interest-rate rules that respond to deviations of inflation and output from their long-run targets. In this robustness exercise, we allow the monetary authority to respond to deviations from output under flexible prices and from efficient output. Instead of (12), we consider

$$\widehat{R}_t = \phi_{\pi,n}\widehat{\pi}_t + \phi_{y,n}(\widehat{y}_t - \widehat{y}_t^n) + \widehat{r}_t, \tag{34}$$

and in case of deviations from efficient output

$$\widehat{R}_t = \phi_{\pi,e}\widehat{\pi}_t + \phi_{y,e}(\widehat{y}_t - \widehat{y}_t^e) + \widehat{r}_t, \tag{35}$$

where the definitions of the output under flexible prices and the efficient output are functions of the fundamental shocks given in the main text. Compared to the baseline specification, the central bank must now observe the fundamental shocks and adjust the definition of the relevant deviations according to the model under consideration.

When the central bank responds to deviations from the natural rate of output, the coefficients of the optimal rules $\phi^{\star}_{\mathrm{C},n}$ and $\phi^{\star}_{a,n}$ are

$$\phi^{\star}_{\mathrm{C},n}: \quad \phi_{\pi,n} = 3.62; \quad \phi_{y,n} = 0.00, \quad \phi^{\star}_{a,n}: \quad \phi_{\pi,n} = 5.91; \quad \phi_{y,n} = 0.00.$$

Comparing the policy coefficients to the coefficients found in the benchmark reveals that the optimal policies are similar. The optimal response to inflation and to deviations of output from the natural level are slightly lower than the response to output found in the main text for both approaches to model uncertainty. These similarities are also reflected in the relative performance of the two approaches across the set of models displayed in the first row of Table 7. While in the benchmark the optimal policy $\phi^{\star}_{a}$ yields welfare gains of 36 percent on average (see Table 6), the welfare gains now amount to 34 percent on average.

When deviations from efficient output are targeted then the optimal rules $\phi^{\star}_{\mathrm{C},n}$ and $\phi^{\star}_{a,n}$ are given by the following set of coefficients

$$\phi^{\star}_{\mathrm{C},e}: \quad \phi_{\pi,e} = 18.56; \quad \phi_{y,e} = 25.00, \quad \phi^{\star}_{a,e}: \quad \phi_{\pi,e} = 25.00; \quad \phi_{y,e} = 13.23.$$

The main difference to the benchmark specification (and to reactions on deviations from output under flexible prices) is that the policymaker can now directly respond to the arguments in the loss function – inflation and the deviation of output from its efficient level. This allows the policymaker to act less cautiously than in the cases analyzed so far, resulting in higher optimal policy coefficients and lower business cycle costs in absolute terms.

The relative performance of the two approaches is shown in the second row of Table 7 and reveals even higher relative gains of the model-averaging approach than in the cases considered before.

Table 7: Performance of $\phi^{\star}_{\mathrm{C}}$ and $\phi^{\star}_{a}$ with feedback on deviations from natural and efficient output

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_{\mathrm{C}}$	$\overline{\mathcal{BC}}$
$\mathcal{BC}(\Theta_i, \phi^{\star}_{\mathrm{C},n})/\mathcal{BC}(\Theta_i, \phi^{\star}_{a,n})$	1.93	1.93	1.44	1.03	0.91	1.34
$\mathcal{BC}(\Theta_i, \phi^{\star}_{\mathrm{C},e})/\mathcal{BC}(\Theta_i, \phi^{\star}_{a,e})$	1.55	1.54	1.64	0.79	0.84	1.49

Notes: $\mathcal{M}_1$ is the benchmark model, $\mathcal{M}_2$ benchmark model plus habit, $\mathcal{M}_3$ is the benchmark model plus indexation, $\mathcal{M}_4$ is the benchmark model plus money, $\mathcal{M}_{\mathrm{C}}$ is the benchmark plus habit, indexation and money.

A.3.2 Alternative money-demand estimates

In our benchmark estimation, we use money-demand estimates based on results found by Lucas (2000), for U.S. data ranging from 1900 to 1994. In this section, we employ the net-interest rate elasticity estimated by Ireland (2009) as an alternative to calibrate the mean of the distribution of the preference parameter σ_m. Ireland (2009) estimated a lower value of the elasticity than Lucas (2000) using U.S. data from 1980 to 2006 (0.09 instead of 0.5), which corresponds to a mean of $\sigma_m = 10.67$. In the following, we report the stabilization weights and the implications for optimal policy under model uncertainty.

Table 8: Optimal policy at the posterior mean ($\bar{\phi_i}^\star$) for Ireland (2009) estimates

	$\mathcal{M}_4$	$\mathcal{M}_C$
$\phi_\pi^\star$	9.65	6.32
$\phi_y^\star$	0.08	0.13
λ_R	0.04	0.07
$\mathcal{BC}(\bar{\theta}_i, \phi_i^\star)$	0.0043	0.0029

Notes: $\mathcal{M}_4$ is the benchmark model $\mathcal{M}_1$ plus money, $\mathcal{M}_C$ is the benchmark plus habit, indexation and money.

Comparing the welfare costs from Table 4 and 8 reveals that the absolute welfare costs in the models with a monetary transaction friction decrease with a lower value of the interest-rate elasticity. Intuitively, because the demand function becomes less elastic, fluctuations in the interest rate lead to less distortions in households' optimal decisions. Furthermore, a lower stabilization weight reduces the severeness of the conflict to stabilize inflation, the output gap and the nominal interest rate simultaneously. The latter is also reflected in a more pronounced reaction of the optimal rule to changes in inflation because inflation stabilization is now relatively more important than in the calibration based on Lucas (2000). The decreasing importance of conflicting policy aims is also reflected in the coefficients of the optimal rules $\phi_C^\star$ and $\phi_a^\star$

$$\phi_C^\star: \quad \phi_\pi = 6.41; \quad \phi_y = 0.11, \quad \phi_a^\star: \quad \phi_\pi = 10.98; \quad \phi_y = 0.13.$$

Thus, both rules now feature a stronger response to inflation than in the baseline calibration. The relative performance of the two rules is summarized in Table 9. The gains of the optimal rule in the model-averaging approach compared to the optimal rule in the the large model are still quantitatively important (12 percent) but smaller than under the Lucas (2000) calibration

(36 percent).

Table 9: Relative performance of $\phi_{\mathrm{C}}^{\star}$ and $\phi_a^{\star}$ for Ireland (2009) estimates

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_{\mathrm{C}}$	$\overline{\mathcal{BC}}$
$\mathcal{BC}(\Theta_i, \phi_{\mathrm{C}}^{\star})/\mathcal{BC}(\Theta_i, \phi_a^{\star})$	1.47	1.54	1.16	1.10	0.94	1.12

Notes: $\mathcal{M}_1$ is the benchmark model, $\mathcal{M}_2$ benchmark model plus habit, $\mathcal{M}_3$ is the benchmark model plus indexation, $\mathcal{M}_4$ is the benchmark model plus money, $\mathcal{M}_{\mathrm{C}}$ is the benchmark plus habit, indexation and money.

A.3.3 Model comparison

In the paper, we estimate the posterior model probabilities based on the estimates of the marginal likelihood of each model $p_{\mathcal{M}_i}(Y)$. The marginal likelihoods are estimated using the modified harmonic mean estimator suggested by Geweke (1999b). Since the models are nested, it is also possible to compute the posterior odds ratio of models i and j in a different way making use of the Savage–Dickey density ratio as suggested by Verdinelli and Wasserman (1995). The Savage–Dickey density ratio has the advantage that it does not require the computation of the separate marginal likelihoods. In this section we employ the Savage–Dickey density ratio to cross-check our model comparison results. Since this is not common in the DSGE literature, we start by briefly deriving the Savage–Dickey density ratio for DSGE models and explaining how we evaluate the Savage–Dickey density ratio numerically.

Denote the parameter vector of model i by θ_i. The parameter vector of model j is denoted by $\tilde{\theta}_j = [\theta_i \quad \theta_j]$, where θ_i is the vector of parameters shared by both models and θ_j the vector of parameters specific to Model j. Model j nests Model i and is equal to Model i for the case $\theta_j = \theta^*$.

Using Bayes' law, the ratio of the marginal likelihoods of Models i and j can be written as:

$$\frac{p_{\mathcal{M}_i}(Y)}{p_{\mathcal{M}_j}(Y)} = \frac{p_{\mathcal{M}_i}(Y|\theta_i)p_{\mathcal{M}_i}(\theta_i)}{p_{\mathcal{M}_j}(Y|\theta_j,\theta_i)p_{\mathcal{M}_j}(\theta_j,\theta_i)} \frac{p_{\mathcal{M}_j}(\theta_j,\theta_i|Y)}{p_{\mathcal{M}_i}(\theta_i|Y)},$$

where $p_{\mathcal{M}}(\theta)$ denotes the prior distribution and $p_{\mathcal{M}}(\theta|Y)$ the posterior distribution. In the case $\theta_j = \theta_j^*$, both models are identical and consequently $p_{\mathcal{M}_i}(Y|\theta_i) = p_{\mathcal{M}_j}(Y|\theta_j^*,\theta_i)$ which yields:

$$\frac{p_{\mathcal{M}_i}(Y)}{p_{\mathcal{M}_j}(Y)} p_{\mathcal{M}_i}(\theta_i|Y) = \frac{p_{\mathcal{M}_i}(\theta_i)}{p_{\mathcal{M}_j}(\theta_j^*,\theta_i)} p_{\mathcal{M}_j}(\theta_j^*,\theta_i|Y)$$

We further use the relationship $p_{\mathcal{M}_j}(\theta_j^*, \theta_i) = p_{\mathcal{M}_j}(\theta_i|\theta_j^*)p_{\mathcal{M}_j}(\theta_j^*)$ and rearrange terms to derive:

$$\frac{p_{\mathcal{M}_i}(Y)}{p_{\mathcal{M}_j}(Y)} \frac{p_{\mathcal{M}_j}(\theta_i|\theta_j^*)}{p_{\mathcal{M}_i}(\theta_i)} p_{\mathcal{M}_i}(\theta_i|Y) = \frac{p_{\mathcal{M}_j}(\theta_j^*, \theta_i|Y)}{p_{\mathcal{M}_j}(\theta_j^*)}.$$

In the next step, we integrate over the vector of parameters shared by both models Θ_i:

$$\frac{p_{\mathcal{M}_i}(Y)}{p_{\mathcal{M}_j}(Y)} \int_{\theta_i \in \Theta_i} \frac{p_{\mathcal{M}_j}(\theta_i|\theta_j^*)}{p_{\mathcal{M}_i}(\theta_i)} p_{\mathcal{M}_i}(\theta_i|Y) \mathrm{d}\theta_i = \frac{1}{p_{\mathcal{M}_j}(\theta_j^*)} \int_{\theta_i \in \Theta_i} p_{\mathcal{M}_j}(\theta_j^*, \theta_i|Y) \mathrm{d}\theta_i$$

Since we specify the same prior distribution for θ_i in model $\mathcal{M}_j$ and $\mathcal{M}_i$ which is also independent of θ_j^* it holds that $p_{\mathcal{M}_j}(\theta_i|\theta_j^*) = p_{\mathcal{M}_i}(\theta_i)$. We further re-write $p_{\mathcal{M}_i}(\theta_i|Y)$ as $\frac{p_{\mathcal{M}_i}(Y,\theta_i)}{p_{\mathcal{M}_i}(Y)}$ and use the fact that $\frac{1}{p_{\mathcal{M}_i}(Y)} \int_{\theta_i \in \Theta_i} p_{\mathcal{M}_i}(Y, \theta_i) \mathrm{d}\theta_i = 1$ to arrive at the Savage-Dickey density ratio:

$$\frac{p_{\mathcal{M}_i}(Y)}{p_{\mathcal{M}_j}(Y)} = \frac{p_{\mathcal{M}_j}(\theta_j^*|Y)}{p_{\mathcal{M}_j}(\theta_j^*)} \tag{36}$$

Equation (36) implies that ratios larger than unity are evidence in favor of the smaller Model i, while ratios smaller than unity represent evidence in favor of Model j. The computation of the right hand side does not involve the estimation of the marginal data density of each model anymore, but only the evaluation of the prior distribution and the posterior distribution of θ_j^*, i.e. $p_j(\theta_j^*)$ and $p_j(\theta_j^*|Y)$ respectively.

In practice, we compute the Savage–Dickey density ratios in the following way. For given draws from the posterior and prior distribution of model j, we determine how many draws satisfy $\theta_j = \theta_j^*$. We find support for the model-comparison results reported in Table 3. In particular, the benchmark model exhibits a higher marginal likelihood and is thus more likely than the model with habit but less likely than the model with indexation and the complete model. The complete model, however, is less likely than the model with indexation.

Even though the computation of the Savage–Dickey density ratio is straightforward, we employ measures based on the harmonic mean estimator to compute marginal likelihoods. We do so because the standard prior distributions we employ for η and for γ imply probabilities close to zero for these parameters to be exactly zero. Therefore, the numerical approximation of $p_j(\theta_j^*)$ and $p_j(\theta_j^*|Y)$ is not without difficulties either.

References

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

BROOKS, S. P., AND A. GELMAN (1998): "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7(4), 434–455.

CALVO, G. A. (1983): "Staggered price setting in a utility-maximizing framework.," *Journal of Monetary Economics*, 12(3), 383–398.

CASTELNUOVO, E. (2012): "Estimating the Evolution of Moneys Role in the U.S. Monetary Business Cycle," *Journal of Money, Credit and Banking*, 44(1), 23–52.

COGLEY, T., B. DE PAOLI, C. MATTHES, K. NIKOLOV, AND T. YATES (2011): "A Bayesian approach to optimal monetary policy with parameter and model uncertainty," *Journal of Economic Dynamics and Control*, 35(12), 2186–2212.

GEWEKE, J. (1999a): "Computational experiments and reality," Discussion paper, University of Minnesota and Federal Reserve Bank of Minneapolis.

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

IRELAND, P. N. (2009): "On the Welfare Cost of Inflation and the Recent Behavior of Money Demand," *American Economic Review*, 99(3), 1040–52.

KUESTER, K., AND V. WIELAND (2010): "Insurance Policies for Monetary Policy in the Euro Area," *Journal of the European Economic Association*, 8(4), 872–912.

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

LEVIN, A., V. WIELAND, AND J. C. WILLIAMS (2003): "The Performance of Forecast-Based Monetary Policy Rules under Model Uncertainty," *American Economic Review*, 93(3), 622–645.

LEVIN, A. T., AND J. C. WILLIAMS (2003): "Robust monetary policy with competing reference models," *Journal of Monetary Economics*, 50(5), 945–975.

Levine, P., P. McAdam, and J. Pearlman (2012): "Probability models and robust policy rules," *European Economic Review*, 56(2), 246–262.

Lucas, R. E. (2000): "Inflation and welfare," *Econometrica*, 68(2), 247–274.

Schmitt-Grohé, S., and M. Uribe (2007): "Optimal simple and implementable monetary and fiscal rules," *Journal of Monetary Economics*, 54(6), 1702–1725.

Sims, C. A. (2003): "Probability Models for Monetary Policy Decisions," Discussion paper, Princeton University.

Sims, C. A. (2008): "Improving monetary policy models," *Journal of Economic Dynamics and Control*, 32(8), 2460–2475.

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(430), 614–618.

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