# Monetary Policy Analysis with Potentially Misspecified Models


Marco Del Negro — *Federal Reserve Bank of Atlanta*

Frank Schorfheide* — *University of Pennsylvania*



This Draft: September, 2005



**Abstract**

The paper proposes a novel method for conducting policy analysis with potentially misspecified dynamic stochastic general equilibrium (DSGE) models and applies it to a New Keynesian DSGE model along the lines of Christiano, Eichenbaum, and Evans (JPE 2005) and Smets and Wouters (JEEA 2003). We first quantify the degree of model misspecification and then illustrate its implications for the performance of different interest-rate feedback rules.



JEL CLASSIFICATION: C32

KEY WORDS: Bayesian Analysis, DSGE Models, Model Misspecification


---


*Correspondence: Marco Del Negro: Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St NE, Atlanta GA 30309-4470. Email: marcodelnegro@frbatlanta.org. Frank Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu.


We would like to thank Kosuke Aoki, David Arsenau, Jesus Fernandez-Villaverde, John Geweke, Lars Hansen, Ramon Marimon, Tom Sargent, Peter Summers, Charles Whiteman, as well as the participants of the 2004 EEA-ESEM session on "Empirical Models for Monetary Policy Analysis," the 2004 ECB Conference on "Monetary Policy and Imperfect Knowledge," the Fall 2004 Macro System Committee Meetings, the 2004 Southern Economic Association Meetings, the 2005 Conference on "25 Years of Macroeconomics and Reality," and seminar participants at the Bank of England, the Kansas City Fed, and the University of Miami for helpful comments. Schorfheide gratefully acknowledges financial support from the Alfred P. Sloan Foundation. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

# 1 Introduction

The quantitative evaluation of monetary policy rules plays an important role in the design of stabilization policies. Much of the recent debate about optimal monetary policy rules has been carried out with New Keynesian dynamic stochastic general equilibrium (DSGE) models (see Woodford 2003). DSGE models have the advantage that one can explicitly assess the effect of policy regime changes on expectation formation and decision rules of private agents. Yet, until recently, these models were scarcely used by central banks because they were perceived as being inferior to less structural model in terms of fit. Recent work by Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2003) has changed this perception. These papers show that the fit of large scale New Keynesian DSGE models is comparable to that of more heavily parameterized models. As a consequence, a number of central banks have begun to show interest in these models as tools for quantitative policy analysis. Despite the success in improving the empirical performance of DSGE models, misspecification remains a concern, as shown in Del Negro, Schorfheide, Smets, and Wouters (2004).

Accounting for misspecification and model uncertainty in general is an important aspect of the assessment of monetary policies. After all, it has long been recognized that model and parameter uncertainty affects optimal policies, e.g., Brainard (1967), Chow (1975), and Craine (1979).[1] This paper proposes a new method for taking model misspecification into account when assessing the performance of alternative interest-rate feedback rules. We apply this approach to evaluate different policies in the context of a New Keynesian DSGE model, including the estimated Volcker-Greenspan rule.

A natural approach in the presence of model uncertainty is to evaluate policy rules within all the model specifications that are under consideration. In choosing the best performing rule one can either follow a Bayesian route, assigning probabilities to models and minimizing the overall posterior expected loss. Alternatively, one can follow a minimax strategy by adopting a policy that minimizes the worst-case loss across models. The literature contains numerous applications of these ideas, e.g., McCallum (1988), Levin, Wieland, and Williams (1999, 2003), Rudebusch (2001, 2003), Onatski and Stock (2002), Onatski and Williams (2003), Brock, Durlauf, and West (2004), Cogley and Sargent (2005), and Hansen and Sargent (2005). All these papers differ with regard to the type of models included in the

[1]The distinction between model and parameter uncertainty is somewhat artificial. By enlarging the parameter space appropriately, we can always absorb different model specifications.

model set, and the formulation of the decision problem that leads to the choice of a preferred policy.

Our analysis differs from previous work with respect to the method that is used to construct a model set and assign probabilities to these models so that a Bayesian calculation of posterior expected losses is possible. We start from a DSGE model in which monetary policy is modelled through an interest-rate feedback rule. The DSGE model imposes cross-coefficient restrictions on a vector autoregressive representation of the data. While we are assuming that the data obey a vector autoregressive law of motion, we allow for potential deviations from the cross-coefficient restrictions. At the same time we maintain the assumption that monetary policy follows the same interest-rate rule as in the DSGE model. This collection of identified vector autoregressions forms our model set.

We place a prior distribution on the structural parameters of the DSGE model and on parameters that characterize the discrepancies between the DSGE model restrictions and the vector autoregressive law of motion. In constructing this prior we follow our earlier work in Del Negro and Schorfheide (2004). Roughly speaking, the prior for the discrepancies is centered at zero and has a covariance matrix that is scaled by a hyperparameter which is denoted by $\lambda$. If $\lambda$ is large then the prior for the discrepancies concentrates near zero reflecting the belief that a potential misspecification of the DSGE model restrictions is small. Vice versa, small values of $\lambda$ induce a prior that implies large misspecification.

Our procedure works in two steps. First, we construct a posterior for the hyperparameter $\lambda$, determining the overall extent of misspecification on the basis of the available sample. Conditional on the estimated $\lambda$ we compute a posterior for the DSGE model and the discrepancy parameters. Next, we evaluate the outcome of alternative policy rules under the simplifying assumption that the public believes the new policy to be in place indefinitely after being announced credibly. Policy evaluation is based on a loss function defined in terms of expected squared deviations of the output gap, inflation, and interest rates from their respective target levels. An important feature of our analysis is that we use the historical observations to learn about the degree of DSGE model misspecification. In other words, the extent to which our policy analysis relies on the DSGE model depends on the estimated degree of misspecification.

Following early work by McCallum (1988) a number of recent papers studies the performance of different interest-rate feedback rules across a variety of macroeconometric models including models that are currently use by the Board of Governors and the European Central Bank to analyze monetary policy. Examples of this line of research include Levin, Wieland,

and Williams (1999, 2003), Taylor (1999), Coenen (2003), Levin and Williams (2003), Brock, Durlauf, and West (2004), and Adalid, Coenen, McAdam, and Siviero (2005). As in our paper, policy loss functions are given by a weighted average of unconditional variances of output, inflation, and interest rates. While the class of models considered in these papers is arguably broad and contains both forward-looking as well as backward-looking specifications, with the exception of Brock, Durlauf, and West (2004) little or no attention is paid to fit and forecasting performance when weighting predictions from various models. Hence, potentially too much weight is given to specifications that are clearly at odds with the data. By using likelihood-based measures of fit, the approach proposed in this paper guarantees that model specifications that have performed well historically receive a lot of weight in the policy loss calculation.

Our framework is rich enough to encompass existing approaches to policy analysis such as the evaluation of policy rules directly based on DSGE models or by replacing the policy rule in identified vector autoregressions as in Sims (1999). Specifically, we consider four assumptions about the policy invariance of the misspecification parameters and calculate posterior expected losses as a function of the policy parameters. The first scenario – used as a benchmark – simply ignores misspecification, and computes the loss assuming the DSGE model correctly describes the data. The first scenario corresponds to the type of analysis conducted by Laforte (2003), and Levin, Onatski, Williams, and Williams (2005). These papers estimate New Keynesian DSGE models similar to the one used here with Bayesian methods to study the effects of uncertainty about structural parameters on optimal monetary policy. If uncertainty about taste and technology parameters is certainly important, we find that the uncertainty about the model specification is as at least equally important. Our estimates of $\lambda$ indeed suggest that policymakers cannot afford to ignore misspecification in the New Keynesian model.[2]

The remaining three scenarios are an attempt to incorporate the concern about model specification in the policy recommendations derived from DSGE models. The second scenario assumes that the policy maker is willing to learn from historical data about the overall degree of model misspecification, but not about its precise nature. In computing the expected loss from a given policy she therefore draws the misspecification parameters from her prior distribution conditional on $\lambda$. In the remaining two scenarios the policy maker learns from the data about the misspecification parameters, that is, uses the posterior dis-

[2]The out-of-sample forecasting results obtained by Del Negro, Schorfheide, Smets, and Wouters (2004) are consistent with this conclusion.

tribution of the misspecification parameters in computing the loss. In the third scenario the policy maker assumes that misspecification is policy invariant, while in the forth scenario she uses the conditional distribution of the misspecification and policy parameters to let the misspecification vary with policy.

We find that the outcomes associated with different policies change as misspecification is taken into account.[3] In particular, the loss associated with policies that deviate from the DSGE model prescriptions may not be nearly as large as policy analysis under the DSGE model would suggest. At the same time, our results indicate that following those prescriptions, even under misspecification, leads to outcomes that are not substantially inferior to those of the best-performing rule. In summary, a fairly robust policy recommendation emerges from our analysis: the central bank should avoid strong responses to output gap movements and not react weakly to inflation fluctuations. Moreover, we find that the best-performing rules are those with a substantial degree of inertia in the policy instrument. An implication of these findings is that the gains associated with deviating from the historical Volcker-Greenspan policy, whenever positive, are generally not very large. This suggests that the historical rule, if not always optimal among those we consider, has been reasonably good at least from the perspective of this New Keynesian DSGE model, even taking misspecification into account.

The literature is fairly divided between Bayesian and minimax or robust approaches to resolve model uncertainty. Rather than placing a prior distribution on the misspecification parameters, the robustness literature specifies either a static or dynamic two-player zero-sum game in which a malevolent "nature" chooses the misspecification parameters to harm the policy maker. Examples are Tetlow and von zur Muehlen (2001), Giannoni (2002), Onatski and Stock (2002), some of the analysis in Levin and Williams (2003), Onatski and Williams (2003), and the robust control approach developed in the monograph by Hansen and Sargent (2005). The disadvantage of this approach is that the resulting policy performs well in the worst-case but possibly poorly on average. A key difficulty in the use of minimax rules is to constrain the model set and to bound the worst case. In most formulations of the minimax problem the policy maker does not use historical data to learn about the extent of model misspecification. An exception is Onatski and Williams (2003) who bound deviations from the Rudebusch-Svenson (1999) model in a minimax calculation based on a confidence

[3] Preliminary empirical results based on a simple three-equation New Keynesian model without capital accumulation and variable factor utilization were reported in the 2005 Proceedings Volume of the Journal of the European Economic Association, Del Negro and Schorfheide (2005).

set derived from a posterior distribution. In Hansen and Sargent's approach the relevant model set is constructed by bounding the Kullback-Leibler discrepancy between the models contained in the set and a reference model, so that it becomes difficult to discriminate among them based on statistical methods. In fact, the minimax approach is arguably most compelling if there is little or no empirical evidence available that can discriminate between model specifications.

While our paper emphasizes a Bayesian resolution of uncertainty, our framework is general enough to enable a risk-sensitive analysis. More specifically, using a result from Jacobsen (1973) we compute posterior expected losses for an exponential transformation of our loss function. The resulting risk can be interpreted as the Nash-equilibrium of a zero-sum game in which "nature" distorts the probability distribution of the misspecification parameters subject to a penalty that is a function of the Kullback-Leibler discrepancy between the distorted and the non-distorted probabilities.

The paper is organized as follows. The DSGE model is presented in Section 2. This model is based on work by Altig, Christiano, Eichenbaum, and Linde (2002), Smets and Wouters (2003), and Christiano, Eichenbaum, and Evans (2005). Compared to the benchmark New Keynesian models discussed, for instance, in Woodford (2003), our model has been subjected to a number of modifications, all designed to improve its empirical fit. Section 3 discusses the estimation of potentially misspecified DSGE models. Bayesian inference is implemented through Markov Chain Monte Carlo methods described in the Appendix. The framework for policy analysis is introduced in Section 4. Section 5 describes the data set and discusses our empirical findings, and Section 6 concludes.

# 2 Model

This section describes the DSGE model, which is based on work by Altig, Christiano, Eichenbaum, and Linde (2002), Smets and Wouters (2003), and Christiano, Eichenbaum, and Evans (2005). The model contains nominal price rigidities, capital accumulation subject to adjustment costs, variable factor utilization, and habit formation.

## 2.1 Final goods producers

The final good $Y_t$ is a composite made of a continuum of intermediate goods $Y_t(i)$, indexed by $i \in [0, 1]$:

$$Y_t = \left[ \int_0^1 Y_t(i)^{\frac{1}{1+\lambda_f}} di \right]^{1+\lambda_f}. \tag{1}$$

The final goods producers are perfectly competitive firms that buy intermediate goods, combine them to the final product $Y_t$, and resell the final good to consumers. The firms maximize profits

$$P_t Y_t - \int P_t(i) Y_t(i) di$$

subject to (1). Here $P_t$ denotes the price of the final good and $P_t(i)$ is the price of intermediate good $i$. From their first order conditions and the zero-profit condition we obtain that:

$$Y_t(i) = \left( \frac{P_t(i)}{P_t} \right)^{-\frac{1+\lambda_f}{\lambda_f}} Y_t \quad \text{and} \quad P_t = \left[ \int_0^1 P_t(i)^{\frac{1}{\lambda_f}} di \right]^{\lambda_f}. \tag{2}$$

We define aggregate inflation as $\pi_t = P_t / P_{t-1}$.

## 2.2 Intermediate goods producers

Good $i$ is made using the technology:

$$Y_t(i) = \max \left\{ K_t(i)^\alpha [\gamma(t) Z_t L_t(i)]^{1-\alpha}, 0 \right\}, \tag{3}$$

where the technology shock $Z_t$ (common across all firms) follows a stationary autoregressive process

$$\ln(Z_t / Z^*) = \rho_z \ln(Z_{t-1} / Z^*) + \sigma_z \epsilon_{z,t} \tag{4}$$

and the function $\gamma(t)$ induces a trend into productivity. All firms face the same prices for their inputs, labor and capital. Hence cost minimization implies that the capital/labor ratio is the same for all firms, and equal to:

$$\frac{K_t}{L_t} = \frac{\alpha}{1-\alpha} \frac{W_t}{R_t^k}, \tag{5}$$

where $W_t$ is the nominal wage and $R_t^k$ is the rental rate of capital. Following Calvo (1983) we assume that in every period a fraction of firms $\zeta_p$ is unable to re-optimize their prices $P_t(i)$. These firms adjust their prices mechanically according to

$$P_t(i) = \pi^* P_{t-1}(i), \tag{6}$$

where $\pi^*$ is the steady state inflation rate of the final good.[4] Firms that are able to re-optimize prices choose the price level $\tilde{P}_t(i)$ by solving

$$\begin{aligned}&\max_{\tilde{P}_t(i)} I\!\!E_t \textstyle\sum_{s=0}^{\infty} \zeta_p^s \beta^s \Xi_{t+s}^p \left(\tilde{P}_t(i)(\pi_*)^s - MC_{t+s}\right) Y_{t+s}(i)\\ &\text{s.t. } Y_{t+s}(i) = \left(\frac{\tilde{P}_t(i)\,(\pi_*)^s}{P_{t+s}}\right)^{-\frac{1+\lambda_f}{\lambda_f}} Y_{t+s}, \quad MC_{t+s} = \frac{\alpha^{-\alpha} W_{t+s}^{1-\alpha} R_{t+s}^{k\ \alpha}}{(1-\alpha)^{(1-\alpha)} Z_{t+s}^{1-\alpha}}.\end{aligned} \tag{7}$$

where $\beta^s \Xi_{t+s}^p$ is today's value of a future dollar for the consumers and $MC_t$ reflects marginal costs. We consider only the symmetric equilibrium where all firms will choose the same $\tilde{P}_t(i)$. Hence from (2) we obtain the following law of motion for the aggregate price level:

$$P_t = \left[(1-\zeta_p)\tilde{P}_t^{\frac{1}{\lambda_f}} + \zeta_p(\pi_* P_{t-1})^{\frac{1}{\lambda_f}}\right]^{\lambda_f}. \tag{8}$$

## 2.3 Households

There is a continuum of households, indexed by $j \in [0,1]$. The objective function for each household is given by:

$$I\!\!E_t \sum_{s=0}^{\infty} \beta^s \left[\log(C_{t+s}(j) - hC_{t+s-1}(j)) - \frac{\varphi}{1+\nu_l} L_{t+s}(j)^{1+\nu_l} + \frac{\chi}{1-\nu_m}\left(\frac{M_{t+s}(j)}{\gamma(t+s)P_{t+s}}\right)^{1-\nu_m}\right] \tag{9}$$

where $C_t(j)$ is consumption, $L_t(j)$ is labor supply, and $M_t(j)$ are money holdings. Households' preferences display habit-persistence. Real money balances enter the utility function deflated by the trend growth of the economy to make real money demand stationary.

Household $j$'s budget constraint written in nominal terms is given by:

$$\begin{aligned}&P_{t+s}C_{t+s}(j) + P_{t+s}I_{t+s}(j) + B_{t+s}(j) + M_{t+s}(j) \le R_{t+s}B_{t+s-1}(j) + M_{t+s-1}(j) + A_{t+s-1}(j)\\ &+ \Pi_{t+s} + W_{t+s}(j)L_{t+s}(j) + \left(R_{t+s}^k u_{t+s}(j)\bar{K}_{t+s-1}(j) - P_{t+s}a(u_{t+s}(j))\bar{K}_{t+s-1}(j)\right),\end{aligned} \tag{10}$$

where $I_t(j)$ is investment, $B_t(j)$ is holdings of government bonds, $R_t$ is the gross nominal interest rate paid on government bonds, $A_t(j)$ is the net cash inflow from trading state-contingent securities, $\Pi_t$ is the per-capita profit the household gets from owning firms (households pool their firm shares, and they all receive the same profit), and $W_t(j)$ is the nominal wage earned by household $j$. The term within parenthesis represents the return to owning $\bar{K}_t(j)$ units of capital. Households choose the utilization rate of their own capital, $u_t(j)$. Households rent to firms in period $t$ an amount of effective capital equal to:

$$K_t(j) = u_t(j)\bar{K}_{t-1}(j), \tag{11}$$

[4] We also estimated a version of the DSGE model with dynamic price indexation $P_t(i) = \pi_{t-1}P_{t-1}(i)$ but found that the time series fit did not improve.

and receive $R_t^k u_t(j)\bar{K}_{t-1}(j)$ in return. They however have to pay a cost of utilization in terms of the consumption good equal to $a(u_t(j))\bar{K}_{t-1}(j)$. Households accumulate capital according to the equation:

$$\bar{K}_t(j) = (1-\delta)\bar{K}_{t-1}(j) + \left(1 - S\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\right) I_t(j), \tag{12}$$

where $\delta$ is the rate of depreciation, and $S(\cdot)$ is the cost of adjusting investment, with $S'(\cdot) > 0, S''(\cdot) > 0$.

Since we do not use wage data in our empirical analysis it is difficult to disentangle price rigidity and wage rigidity. We therefore assume that the labor market is perfectly competitive and wages are flexible. Finally, we assume there is a complete set of state contingent securities in nominal terms, which implies that the Lagrange multiplier associated with (10) must be the same for all households in all periods and across all states of the world. This in turn implies that in equilibrium households will make the same choices and can be aggregated into a representative household.

## 2.4 Government policies

The central bank follows a nominal interest rate rule by adjusting its instrument in response to deviations of inflation and output from their respective target levels:

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1} \left(\frac{Y_t}{Y_{t-1}^p}\right)^{\psi_2}\right]^{1-\rho_R} \sigma_R e^{\epsilon_{R,t}}, \tag{13}$$

where $R^*$ is the steady state nominal rate, $Y_t^p$ is a measure of potential output, and the parameter $\rho_R$ determines the degree of interest rate smoothing. The government budget constraint is of the form

$$P_t G_t + R_{t-1}B_{t-1} + M_{t-1} = T_t + M_t + B_t, \tag{14}$$

where $T_t$ are nominal lump-sum taxes (or subsidies) that also appear in household's budget constraint. Government spending is given by:

$$G_t = (1 - 1/g_t)Y_t, \tag{15}$$

where $g_t$ follows the process:

$$\ln g_t = (1-\rho_g)\ln g + \rho_g \ln g_{t-1} + \sigma_g \epsilon_{g,t}. \tag{16}$$

## 2.5 Resource constraint

The aggregate resource constraint

$$C_t + I_t + a(u_t)\bar{K}_{t-1} = \frac{1}{g_t} Y_t \tag{17}$$

can be derived by integrating the budget constraint (10) across households, and combining it with the government budget constraint (14) and the zero profit conditions of the final good producers.

## 2.6 Model solution and State-Space Representation

Our model economy evolves along a balanced growth path, generated by the trend in technology. Output, consumption, investment, real wage, physical capital and effective capital all grow according to $\gamma(t)$. Nominal interest rates, inflation, and hours worked are stationary. The model can be rewritten in terms of detrended variables. We find the steady states for the detrended variables and use the method in Sims (2002) to construct a log-linear approximation of the model around the steady state.

Our empirical analysis is based on data on nominal interest rates $R_t^a$ (annualized percentages), inflation rates $\pi_t^a$ (annualized percentages), and the log output gap. Hence, we define the vector of observations as $y_t = [R_t^a, \pi_t^a, 100 \ln Y_t^g]'$. The relationships between the percentage deviations from steady state $\widetilde{R}_t$, $\widetilde{\pi}_t$, $\widetilde{Y}_t$ derived from the DSGE model and the observables $y_t$ are given by the following measurement equation:

$$\begin{aligned} y_{1,t} &= r_a^* + \gamma_a + \pi_a^* + 4\widetilde{R}_t, \\ y_{2,t} &= \begin{bmatrix} \pi_a^* + 4\widetilde{\pi}_t \\ \widetilde{Y}_t \end{bmatrix}. \end{aligned} \tag{18}$$

Here, we have partitioned $y_t$ such that $y_{1,t}$ corresponds to the policymaker's instrument (the interest rate), and $y_{2,t}$ is a vector that includes the remaining two observables. The steady state (net) real interest rate in our model is given by $r_a^* + \gamma_a$, where $\gamma_a$ is the average annual growth rate of the economy. The parameter $r_a^*$ is related to the discount rate $\beta$ according to $\beta = 1/(1 + r_a^*/400)$. Within the model, $\widetilde{Y}_t$ denotes the percentage deviation of output from its trend path $\gamma(t)Y^*$. We interpret the potential output series published by the Congressional Budget Office (CBO) as a measure of $\gamma(t)Y^*$. Hence, the output gap, computed as log difference of real and potential GDP provides us with a measure of $\widetilde{Y}_t$. The monetary policy rule can be rewritten in terms of observables as follows:

$$y_{1,t} = (1-\rho_R)[(r_a^* + \gamma_a + \pi_a^*) - \psi_1 \ln \pi_a^*] + y_{1,t-1}\rho_R + y_{2,t}' \begin{bmatrix} (1-\rho_R)\psi_1 \\ 4(1-\rho_R)\psi_2 \end{bmatrix} + \sigma_R \epsilon_R \quad (19)$$

We collect all the DSGE model parameters in the vector $\theta$ and stack the structural shocks in the vector $\epsilon_t$.

# 3 Setup and Inference

In the subsequent analysis it is assumed that the DSGE model generates a covariance-stationary distribution of the sequence $\{y_t\}$ for all $\theta \in \Theta$. Expectations under this distribution are denoted by $I\!E_\theta^D[\cdot]$. We will derive an (approximate) vector autoregressive representation for the DSGE model and introduce model misspecifications as deviations from this representation.[5] Unlike in Del Negro and Schorfheide (2004) and Del Negro, Schorfheide, Smets, and Wouters (2004), we assume that the interest rate feedback rule in the DSGE model is correctly specified and do not relax the restriction generated by the policy rule. Finally, we specify a prior distribution for these model misspecifications and discuss posterior inference and policy analysis.

## 3.1 A VAR Representation of the DSGE Model

Let us rewrite Equation (13), which describes the policymaker's behavior, in more general form as:

$$y_{1,t} = x_t' M_1 \beta_1(\theta) + y_{2,t}' M_2 \beta_2(\theta) + \epsilon_{1,t}, \quad (20)$$

where $y_t = [y_{1,t}, y_{2,t}']'$ and the $k \times 1$ vector $x_t = [y_{t-1}', \dots, y_{t-p}', 1]'$ is composed of the first $p$ lags of $y_t$ and an intercept. The shock $\epsilon_{1,t}$ corresponds to the monetary policy shock $\sigma_R \epsilon_{R,t}$ in the DSGE model. The matrices $M_1$ and $M_2$ select the appropriate elements of $x_t$ and $y_{2,t}$ to generate the policy rule. In our application the vector $M_1$ selects the intercept and the lagged nominal interest rate and the matrix $M_2$ extracts inflation, and the output gap. The functions $\beta_1(\theta)$ and $\beta_2(\theta)$ are implicitly provided in Equation (19). Considering forecast-based policy rules in this framework would require significant modifications. However, according to the findings reported by Levin, Wieland, and Williams (2003) forecast-based rules do not provide substantial gains in stabilization performance compared

[5]We are working with vector autoregressive approximations rather than with state-space models to simplify the simulation of the posterior distributions.

with simple outcome-based rules. Hence, we decided not to pursue these modifications at this point.

The remainder of the system for $y_t$ is given by the following reduced form equations:

$$y'_{2,t} = x'_t \Psi^*(\theta) + u'_{2,t}. \tag{21}$$

In general, the VAR representation (21) is not exact if the number of lags $p$ is finite. We define $\Gamma_{XX}(\theta) = \mathbb{E}^D_\theta[x_t x'_t]$ and $\Gamma_{XY_2}(\theta) = \mathbb{E}^D_\theta[x_t y'_{2,t}]$ and let

$$\Psi^*(\theta) = \Gamma^{-1}_{XX}(\theta)\Gamma_{XY_2}(\theta). \tag{22}$$

Since the system is covariance stationary, the VAR approximation of the autocovariance sequence of $y_{2,t}$ can be made arbitrarily precise by increasing the number of lags $p$. If in addition, the moving-average (MA) representation of the DSGE model in terms of the structural shocks $\epsilon_t$ is invertible, then $u_{2,t}$ can also be expressed as a function of $\epsilon_t$ for large $p$. Conditions for invertibility and results on the accuracy of this VAR approximation can be found in Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2004).

The equation for the policy instrument (20) can be rewritten by replacing $y_{2,t}$ with expression (21):

$$y_{1,t} = x'_t M_1 \beta_1(\theta) + x'_t \Psi^*(\theta) M_2 \beta_2(\theta) + u_{1,t}, \tag{23}$$

where $u_{1,t} = u'_{2,t} M_2 \beta_2(\theta) + \epsilon_{1,t}$. Define $u'_t = [u_{1,t}, u'_{2,t}]$, $B_1(\theta) = [M_1\beta_1(\theta), 0_{k\times(n-1)}]$, $B_2(\theta) = [M_2\beta_2(\theta), I_{(n-1)\times(n-1)}]$, and let

$$\Phi^*(\theta) = B_1(\theta) + \Psi^*(\theta) B_2(\theta). \tag{24}$$

Hence, we obtain a restricted VAR for $y_t$

$$y'_t = x'_t \Phi + u'_t, \quad \mathbb{E}[u_t u'_t] = \Sigma^*(\theta) \tag{25}$$

with

$$\Phi = \Phi^*(\theta), \quad \Sigma = \Sigma^*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)\Gamma^{-1}_{XX}(\theta)\Gamma_{XY}(\theta).$$

Here the population covariance matrices are $\Gamma_{YY}(\theta) = \mathbb{E}^D_\theta[y_t y'_t]$ and $\Gamma_{XY}(\theta) = \Gamma'_{YX}(\theta) = \mathbb{E}^D_\theta[x_t y'_t]$.

Since the monetary policy rule (13) in the DSGE model is specified so that it can be exactly reproduced by the VAR, see Equation (20), $\Phi^*(\theta)$ equals the population least squares coefficients associated with (25), and the covariance matrix of $x_t$ under the DSGE model and its VAR approximation are identical. We will subsequently ignore the error made by approximating the state space representation of the DSGE model with the finite-order VAR or, in other words, treat (25) as the structural model that imposes – potentially misspecified – cross-equation restrictions on the matrices $\Phi$ and $\Sigma$. We can do so because we have checked that the impulse responses from the VAR representation of the DSGE model – obtained using the identification scheme discussed in Section 3.3 – match almost exactly those of the DSGE model even when the lag length is four (results are available upon request).

## 3.2 Misspecification and Bayesian Inference

We make the following assumptions about misspecification of the DSGE model. There is a vector $\theta$ and matrices $\Psi^{\Delta}$ and $\Sigma^{\Delta}$ such that the data are generated from the VAR in (25)

$$\Phi = B_1(\theta) + (\Psi^*(\theta) + \Psi^{\Delta})B_2(\theta), \quad \Sigma = \Sigma^*(\theta) + \Sigma^{\Delta} \tag{26}$$

and there does not exist a $\tilde{\theta} \in \Theta$ such that

$$\Phi = B_1(\tilde{\theta}) + \Psi^*(\tilde{\theta})B_2(\tilde{\theta}), \quad \Sigma = \Sigma^*(\tilde{\theta}).$$

We refer to the resulting specification as DSGE-VAR. A stylized graphical representation of our notion of misspecification can be found in Figure 1. Our econometric analysis is casted in a Bayesian framework in which initial beliefs about the DSGE model parameter $\theta$ and the model misspecification matrices $\Psi^{\Delta}$ and $\Sigma^{\Delta}$ are summarized in a prior distribution. In order to compare a Bayesian approach to model misspecification to a minimax approach, the reader might find it helpful to think of a fictitious other, "nature", that draws the misspecification matrices $\Psi^{\Delta}$ and $\Sigma^{\Delta}$ from a distribution – the prior – rather than maximizing the loss function to harm the policy maker. The remainder of this section describes the choice of this prior.

Our prior is based on the idea that "nature" is more likely to draw smaller than larger misspecification matrices, reflecting the belief that the DSGE model provides a good albeit not perfect approximation of reality. Specifically, we assume that the prior density decreases the larger the size of the discrepancies $\Psi^{\Delta}$ and $\Sigma^{\Delta}$. In the spirit of Hansen and Sargent's (2005) approach to model misspecification and robust control, the size of the discrepancies is determined by the ease with which they can be detected via likelihood ratios. This metric determines the shape of the prior contours (see Figure 1). The mass placed on these contours is determined by the parameter $\lambda$. Large values of $\lambda$ imply that large discrepancies are less likely to occur. Hence, the parameter $\lambda$ measures the overall degree of misspecification. We will now further motivate and explain the prior distribution using a thought experiment, where for ease of exposition we set $\Sigma^\Delta = 0$ and fix the DSGE model parameter vector $\theta$.

Suppose that a sample of $\lambda T$ observations is generated from the DSGE model (that is, from Equation (25), where $\Phi = \Phi^*$). Here $T$ denotes the size of the actual sample used in the estimation. We will construct a prior that has the property that its density is proportional to the expected likelihood ratio of $\Psi$ evaluated at its (misspecified) restricted value $\Psi * (\theta)$ versus the true value $\Psi = \Psi^*(\theta) + \Psi^\Delta$. The log-likelihood ratio is

$$\begin{aligned}\ln\left[\frac{\mathcal{L}(\Psi^*,\Sigma^*,\theta|Y,X)}{\mathcal{L}(\Psi^*+\Psi^\Delta,\Sigma^*,\theta|Y,X)}\right] &= -\frac{1}{2}tr\Big[\Sigma^{*-1}\Big(B_2'\Psi^{*'}X'X\Psi^*B_2 - 2B_2'\Psi^{*'}X'(Y-XB_1)\\ &\quad -B_2'(\Psi^*+\Psi^\Delta)'X'X(\Psi^*+\Psi^\Delta)B_2\\ &\quad +2B_2'(\Psi^*+\Psi^\Delta)'X'(Y-XB_1)\Big)\Big].\end{aligned}$$

Here $Y$ denotes the $\lambda T \times n$ matrix with rows $y_t'$ and $X_t$ is the $\lambda T \times k$ matrix with rows $x_t'$. After replacing $Y$ by $X(B_1 + (\Psi^* + \Psi^\Delta)B_2) + U$ the log likelihood ratio simplifies to

$$\begin{aligned}&\ln\left[\frac{\mathcal{L}(\Psi^*,\Sigma^*,\theta|Y,X)}{\mathcal{L}(\Psi^*+\Psi^\Delta,\Sigma^*,\theta|Y,X)}\right] \qquad (27)\\ &= -\frac{1}{2}tr\Big[\Sigma^{*-1}\Big(B_2'\Psi^{\Delta'}X'X\Psi^\Delta B_2 - 2B_2'\Psi^{\Delta'}X'U\Big)\Big]\end{aligned}$$

Taking expectations over $X$ and $U$ using the distribution induced by the data generating process yields (minus) the Kullback-Leibler discrepancy between the data generating process and the DSGE model:

$$I\!E^{VAR}_{\Psi^*,\Sigma^*}\left[\ln\frac{\mathcal{L}(\Psi^*,\Sigma^*,\theta|Y,X)}{\mathcal{L}(\Psi^*+\Psi^\Delta,\Sigma^*,\theta|Y,X)}\right] = -\frac{1}{2}tr\Big[\Sigma^{*-1}\Big(\lambda T B_2'\Psi^{\Delta'}\Gamma_{XX}\Psi^\Delta B_2\Big)\Big]. \qquad (28)$$

Here we have used the relationship $I\!E^{VAR}_{\Psi^*(\theta),\Sigma^*(\theta)}[x_t x_t'] = I\!E^D_\theta[x_t x_t'] = \Gamma_{XX}(\theta)$, where the first term refers to expectation taken with respect to the probability distribution generated by the VAR. We choose a prior density for $\Psi^\Delta$ that is proportional ($\propto$) to the Kullback-Leibler discrepancy:

$$p(\Psi^\Delta|\Sigma^*,\theta) \propto \exp\left\{-\frac{\lambda T}{2}tr\Big[\Sigma^{*-1}\Big(B_2'\Psi^{\Delta'}\Gamma_{XX}\Psi^\Delta B_2\Big)\Big]\right\} \qquad (29)$$

The hyperparameter $\lambda$ determines the length of the hypothetical sample as a multiple of the actual sample size $T$. This hyperparameter "scales" the overall degree of misspecification. For high values of $\lambda$, it is easy to tell the misspecified model and the DSGE model apart even for small values of the misspecification $\Psi^\Delta$. Hence the prior density places most of its mass near the restrictions imposed by the DSGE model when $\lambda$ is large, and for $\lambda = \infty$ the misspecification disappears altogether. On the contrary, if $\lambda$ is close to zero the Kullback-Leibler discrepancy can be small even for relatively large values of the discrepancy $\Psi^\Delta$.

Hence the prior is fairly diffuse. For computational reasons it is convenient to transform this prior into a prior for $\Psi$. Using standard arguments we deduce that this prior is multivariate normal

$$\Psi|\Sigma^*, \theta \sim \mathcal{N}\left(\Psi^*(\theta), \frac{1}{\lambda T}\left[(B_2(\theta)\Sigma^{*-1}B_2(\theta)') \otimes \Gamma_{XX}(\theta)\right]^{-1}\right). \tag{30}$$

In practice we also have to take potential misspecification of the covariance matrix $\Sigma^*(\theta)$ into account. Hence, we will use the following, slightly modified, prior distribution conditional on $\theta$ in the empirical analysis:

$$\begin{aligned} \Psi|\Sigma, \theta &\sim \mathcal{N}\left(\Psi^*(\theta), \frac{1}{\lambda T}\left[(B_2(\theta)\Sigma^{-1}B_2(\theta)') \otimes \Gamma_{XX}(\theta)\right]^{-1}\right) \\ \Sigma|\theta &\sim \mathcal{IW}\Big(\lambda T\Sigma^*(\theta), \lambda T - k, n\Big), \end{aligned} \tag{31}$$

where $\mathcal{IW}$ denotes the inverted Wishart distribution. The latter induces a distribution for the discrepancy $\Sigma^{\Delta} = \Sigma - \Sigma^*$.

The Appendix provides a characterization of the following conditional posterior densities:

$$p(\Psi|\Sigma, \theta, Y), \quad p(\Sigma|\Psi, \theta, Y), \quad \text{and} \quad p(\theta|\Psi, \Sigma, Y).$$

Unfortunately, it is not possible to give a characterization of all conditional distributions in terms of well-known probability distributions. To implement the Gibbs sampler we have to introduce two Metropolis steps that generate draws from the conditional distributions $p(\Sigma|\Psi, \theta, Y)$ and $p(\theta|\Psi, \Sigma, Y)$. The resulting Markov-Chain-Monte-Carlo (MCMC) algorithm is known as Metropolis-within-Gibbs sampler and allows us to generate draws from the joint posterior distribution of $\theta$, $\Psi$, and $\Sigma$. In addition to the posterior distribution of the parameters we are also interested in evaluating marginal data densities of the form

$$p(Y) = \int p(Y|\theta, \Sigma, \Phi)p_\lambda(\theta, \Sigma, \Phi)d(\theta, \Sigma, \Phi) \tag{32}$$

for various choices of the hyperparameter $\lambda$ and restrictions on the parameter space of the DSGE model. Based on the marginal data densities we can compute Bayes factors and posterior probabilities for the various specifications of our model. Under the assumption of equal prior probabilities, ratios of marginal likelihoods can be interpreted as model odds.

### 3.3 Identification

The model developed in the preceding subsections is of the form

$$\begin{aligned} y_{1,t} &= x_t' M_1 \beta_1 + x_t' \Psi M_2 \beta_2 + u_{1,t} \\ y_{2,t}' &= x_t' \Psi + u_{2,t}', \end{aligned} \tag{33}$$

where $u_{1,t} = u'_{2,t}M_2\beta_2 + \epsilon_{1,t}$. The policy rule coefficients $\beta_1$ and $\beta_2$, and hence the monetary policy shock $\epsilon_{1,t} = \sigma_R \epsilon_{R,t}$ are identifiable. Unlike in a standard VAR, identification is achieved through exclusion restrictions: lagged inflation and output gap do not enter the monetary policy rule by assumption. According to the underlying structural model, the one-step ahead forecast errors $u_{2,t}$ are a function of the monetary policy shock $\epsilon_{1,t}$ and the two other structural shocks $\epsilon_{2,t} = [\epsilon_{g,t}, \epsilon_{z,t}]'$:

$$u'_{2,t} = \epsilon_{1,t}A_1 + \epsilon'_{2,t}A_2.$$

Straightforward matrix algebra leads to the following formulas for the effect of the structural shocks on $u'_{2,t}$:

$$\begin{aligned} A_1 &= \left[\Sigma_{11} - \beta_2' M_2' \Sigma_{22} M_2 \beta_2 - 2(\Sigma_{12} - \beta_2' M_2' \Sigma_{22}) M_2 \beta_2\right]^{-1} (\Sigma_{12} - \beta_2' M_2' \Sigma_{22}) \\ A_2' A_2 &= \Sigma_{22} - A_1' \left[\Sigma_{11} - \beta_2' M_2' \Sigma_{22} M_2 \beta_2 - 2(\Sigma_{12} - \beta_2' M_2' \Sigma_{22}) M_2 \beta_2\right] A_1 \end{aligned}$$

Here, $\Sigma$ denotes the covariance matrix of $u_t$, and the partitions of $\Sigma$ conform with the partition of $u'_t = [u_{1,t}, u'_{2,t}]$. Once $A_1$ is determined, the impulse response function with respect to the monetary policy shock can be calculated.

In order to identify $A_2$ an additional assumption is needed. We follow the approach taken in Del Negro and Schorfheide (2004). Let $A'_{2,tr}A_{2,tr} = A'_2A_2$ be the Cholesky decomposition of $A'_2A_2$. The relationship between $A_{2,tr}$ and $A_2$ is given by $A'_2 = A'_{2,tr}\Omega$, where $\Omega$ is an orthonormal matrix that is not identifiable based on the estimates of $\beta(\theta)$, $\Psi$, and $\Sigma$. However, we are able to calculate an initial effect of $\epsilon_{2,t}$ on $y_{2,t}$ based on the DSGE model, denoted by $A_2^D(\theta)$. This matrix can be uniquely decomposed into a lower triangular matrix and an orthonormal matrix:

$$A_2^{D'}(\theta) = A_{2,tr}^{D'}(\theta)\Omega^*(\theta).$$

To identify $A_2$ above, we combine $A'_{2,tr}$ with $\Omega^*(\theta)$.[6] Loosely speaking, the rotation matrix is constructed such that in the absence of misspecification the DSGE's and the DSGE-VAR's impulse responses to $\epsilon_{2,t}$ would coincide. To the extent that misspecification is mainly in the dynamics as opposed to the covariance matrix of innovations, the identification procedure can be interpreted as matching, at least qualitatively, the short-run responses of the VAR with those from the DSGE model.

[6]The calculation is easily implementable in a Markov Chain Monte Carlo analysis. For every draw of $(\theta, \Psi^\Delta, \Sigma^\Delta)$ from their joint posterior distribution we compute $\Omega^*(\theta)$ and $A_2$.

# 4 Policy Analysis

Between time $t = T$ and $t = T + 1$ the policymaker seeks to replace the existing policy rule with one that minimizes a loss function that will be defined subsequently. We make the simplifying assumption that the public believes the new policy to be in place indefinitely after being announced credibly. The policymaker does not exploit the fact that the public has formed its time $T$ expectations based on the $T$ policy rule. This assumption is a short-cut to a more realistic scenario in which there are two types of policy shifts - normal policy making and rare regime shifts (using the terminology of Sims, 1982).

## 4.1 Loss Function

To simplify the exposition we begin by abstracting from parameter uncertainty. Suppose that prior to the policy the economy operates according to the parameters $\theta_0$, $\Psi_0^\Delta$, and $\Sigma_0^\Delta$. We assume that under this parameterization the VAR is non-explosive with long-run mean $\bar{y}$. Define $\widetilde{y}_t = y_t - \bar{y}$. Let $\widetilde{M}$ be the $(k-1) \times k$ matrix with zeros in the last column and a $(k-1) \times (k-1)$ identity matrix in the remaining columns. Moreover, define $\widetilde{x}_t = [\widetilde{y}_{t-1}, \dots, \widetilde{y}_{t-p}]'$. Then the VAR can be rewritten in terms of deviations from the mean as follows:

$$\widetilde{y}_t' = \widetilde{x}_t' \widetilde{M} [B_1(\theta) + (\Psi^*(\theta) + \Psi^\Delta) B_2(\theta)] + u_t'. \tag{34}$$

We assume the mean $\bar{y}$ is invariant to changes in the policy parameters[7] and that the policy maker considers the following loss function

$$\mathcal{L}_T(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) = (1-\delta) I\!E_T \left[ \sum_{t=T+1}^{T+h} \delta^{t-T-1} tr[\mathcal{W} \widetilde{y}_t \widetilde{y}_t'] \right], \tag{35}$$

where the law of motion of $\tilde{y}$ is given by (34). $\delta$ is a discount factor, $\theta$ is partitioned into policy rule parameters $\theta_p$ and taste-and-technology parameters $\theta_s$, and $tr[\cdot]$ denotes the trace operator. The expectation in (35) is taken conditional on post-intervention parameters $\theta$, $\Psi^\Delta$, and $\Sigma^\Delta$ and the pre-intervention observations $\widetilde{y}_{T-p+1}, \dots, \widetilde{y}_T$.

The loss function can be rewritten as

$$\mathcal{L}_T(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) = (1-\delta) \sum_{t=T+1}^{T+h} \delta^{t-T-1} \Big( tr[\mathcal{W} \mathbf{V}_T(\widetilde{y}_t)] + tr[\mathcal{W} I\!E_T[\widetilde{y}_t] I\!E_T[\widetilde{y}_t]'] \Big). \tag{36}$$

[7]This assumption is consistent with the DSGE model, in which the policy parameters $\psi_1$, $\psi_2$, and $\rho_R$ do not affect the steady state output gap, inflation, and interest rates.

Here $\mathbf{V}_T(\cdot)$ denotes the conditional covariance matrix of $\tilde{y}_t$. The loss function $\mathcal{L}_T(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta)$ is well defined, even if the post intervention VAR is explosive, as long as the horizon $h$ is finite or the reciprocal of the discount factor exceeds the largest eigenvalue of the vector autoregressive system. Since the loss is obtained by taking a conditional expectation it depends on the state of the economy in time $T$, summarized by $\widetilde{x}_{T+1}$. To remove this time dependence a common approach in the literature, see for instance Woodford (2003), is to integrate over $\widetilde{x}_{T+1}$ using the distribution implied by the VAR, provided the system is stationary. Hence we define

$$\mathcal{L}(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) = I\!\!E\left[\mathcal{L}_T(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta)\right] = (1-\delta^h) tr[\mathcal{W}\mathbf{V}(\widetilde{y}_{T+1})], \tag{37}$$

where $\mathbf{V}(\cdot)$ is now the unconditional variance.

We truncate the loss function $\mathcal{L}$ at the level $\mathcal{B}$. This truncation ensures that the expected loss is well defined, even if some of the parameter configurations in the support of the posterior imply explosive behavior of the vector autoregressive system. Let[8]

$$L(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) = \min\left\{\mathcal{B}, \mathcal{L}(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta)\right\}. \tag{38}$$

In the empirical analysis the weighting matrix $\mathcal{W}$ is diagonal with elements $\frac{1}{4}$ (interest rates, annualized), 1 (inflation, annualized), and $\frac{1}{4}$ (output gap, percentage deviations from potential output). Our weight on the output gap is somewhat larger than in Woodford (2003, Table 6.1) reflecting a smaller estimate of the price stickiness. Moreover, we place considerable weight on the nominal interest rate, which could be justified by a large interest elasticity of money demand and an important role of real money balances for transactions. The upper bound $\mathcal{B}$ of the loss is set to 50, which is more than 20 times larger than the weighted sample variance of the three series. As a basis for comparison, the sample variances of the output gap, inflation, and interest rates are approximately 2.9, 2.1, and 6 percent respectively.

We interpret monetary policy shocks as discretionary deviations from the interest-rate feedback rule, reflecting normal policy making in Sims' (1982) terminology. The monetary policy shocks add variability to the endogenous variables in the model and increase the loss. Hence, deviations $\epsilon_{R,t}$ from the policy rule are clearly suboptimal. In the empirical

[8]In our empirical analysis (not reported in this paper) we also calculated posterior expected losses for the loss function $(1-\delta)\sum_{t=T+1}^{T+h}\delta^{t-T-1}tr[\mathcal{W}\mathbf{V}_T(\widetilde{y}_t)]$ for $h = 80$ (20 years) and $\delta = 0.99$. Even though the expected losses are strictly speaking finite, the posterior risk was greater than $10^{10}$ for those values of $\theta_p$ that with some probability lead to explosive behavior of the resulting VAR, and less than 20 for the other values of $\theta_p$.

analysis we therefore use the procedure described in Section 3.3 to identify the monetary policy shocks and then set their standard deviation to zero in the calculation of the policy losses.

## 4.2 Taking Misspecification into Account

The policymaker minimizes the loss $L(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta)$ as a function of the policy parameter $\theta_p$. She has imperfect knowledge about: (i) the policy invariant private sectors' taste and technology parameters $\theta_s$; and (ii) the degree of model misspecification captured by $\lambda$, $\Psi^\Delta$ and $\Sigma^\Delta$. The uncertainty is summarized in the posterior distribution.

We consider four different scenarios for the policy invariance of the misspecification matrices $\Phi^\Delta$ and $\Sigma^\Delta$. Then we calculate the posterior expected loss associated with different policies according to each scenario. If the DSGE model does not suffer from serious misspecification all scenarios collapse to Scenario 1. The goal of the subsequent empirical analysis is to illustrate the sensitivity of policy predictions when the DSGE model is embedded in a collection of identified VARs.

The challenge in the evaluation of monetary policy rules is to predict the private sectors' behavioral responses to policy regime changes. Standard VAR analysis proceeds under the assumption that the behavioral changes are negligible. Our framework creates a link between VAR and DSGE model parameter. Hence, it becomes possible to re-calculate the private sector decision rules with the DSGE model and exploit the link to make non-trivial predictions about the private sector behavior with the identified VAR. If the data contain substantial evidence against the DSGE model restrictions, that is, $\lambda$ is small, less weight is placed on the DSGE model predictions about the behavior of the private sector.

*Scenario 1 – Ignore Misspecification:* The DSGE model is estimated directly and its potential misspecification is ignored. The policymaker does, however, take the uncertainty with respect to the non-policy parameters into account when calculating the expected loss. This scenario is explored in detail by Laforte (2003) and Levin, Onatski, Williams, and Williams (2005). If no deviations from the DSGE model restrictions are contemplated then the ad-hoc loss function introduced in the previous subsection could be replaced by the households' utility function. Unfortunately, the empirical evidence points toward misspecifications. Once we allow for deviations from $\Psi^*(\theta)$ and $\Sigma^*(\theta)$, it is not clear whether the DSGE model-based welfare function remains appropriate. Hence, we conduct our policy analysis based on the ad-hoc loss function in all the scenarios.

*Scenario 2 – Acknowledge Misspecification, Discard the Past:* The policymaker believes that the sample (hence the posterior) provides no information about potential misspecification after a regime shift has been implemented. This skepticism about the relevance of sample information is shared by the robust control approaches of Onatski and Stock (2002), Onatski and Williams (2003), and Hansen and Sargent (2005). However, instead of using a minimax calculation, our Bayesian policymaker relies on her prior distribution $p(\Psi^\Delta, \Sigma^\Delta|\theta, \lambda)$ to cope with uncertainty about model misspecification. The sample is only used to learn about the non-policy parameters $\theta_s$ and the overall degree of misspecification $\lambda$.

*Scenario 3 – Learn about Misspecification (Policy Invariant):* $\Psi^\Delta$ and $\Sigma^\Delta$ are assumed to be invariant to changes in policy. The sample information is used to learn about the model misspecification via the posterior distribution. Looking forward, the information is used to adjust the policy predictions derived from the DSGE model by the estimated discrepancies. To implement the analysis, we generate draws from the marginal posterior distribution of $\theta_s$, $\Psi^\Delta$, and $\Sigma^\Delta$, combine $\tilde{\theta} = [\tilde{\theta}_p', \theta_s']'$, and calculate $\Psi^*(\tilde{\theta}) + \Psi^\Delta$ and $\Sigma^*(\tilde{\theta}) + \Sigma^\Delta$. Here, $\tilde{\theta}_p$ is the new set of policy parameters. Since the choice of $\tilde{\theta}_p$ does not affect beliefs about the misspecification matrices we refer to the treatment of misspecification as *policy invariant.* This rather mechanical post-intervention adjustment of $\Psi^*(\theta)$ and $\Sigma^*(\theta)$ has some undesirable properties which will become evident in the empirical analysis.

*Scenario 4 – Learn about Misspecification (Conditional):* "Nature" generates a new set of draws from the posterior distribution of $\Psi^\Delta$ and $\Sigma^\Delta$ conditional on the post-intervention DSGE model parameters $\tilde{\theta}$. To implement the risk calculation we take a draw from the marginal posterior distribution of $\theta_s$, combine it with the policy parameter to obtain $\tilde{\theta} = [\tilde{\theta}_p', \theta_s']'$, and generate a draw from $p(\Psi^\Delta, \Sigma^\Delta|Y^T, \tilde{\theta}, \lambda)$ by iterating between the conditional distributions of $\Psi^\Delta$ and $\Sigma^\Delta$ provided in the Appendix (see Equations (A.5) and (A.7)). As before, we then calculate $\Psi^*(\tilde{\theta}) + \Psi^\Delta$ and $\Sigma^*(\tilde{\theta}) + \Sigma^\Delta$. In this scenario, the policy maker revises her beliefs about the misspecification matrix as she contemplates different values of the policy parameters. Hence, we use the term *conditional.* Roughly speaking, the calculation can be interpreted as follows: based on the data there is uncertainty about the historical policy rule coefficients. Now suppose one fixes the policy parameters at a particular value $\tilde{\theta}_p$: what information do the data provide about the misspecification parameters? This information is summarized in $p(\Psi^\Delta, \Sigma^\Delta|Y^T, \tilde{\theta}, \lambda)$. If the estimated value of $\lambda$ is small and hence the perceived DSGE model misspecification is large, the DSGE model's prediction of the effect of policy rule changes on the reduced form equations for the private sector become less credible. In our analysis the conditional posterior distribution of $\Psi$ and $\Sigma$ given $\theta$ will effectively become insensitive to $\theta$. As $\lambda$ tends to zero we effectively analyze monetary policy with a VAR by simply changing the coefficients in the policy rule, ignoring any changes in private-sector behavior that the policy shift might induce as in Sims (1999).

### 4.3 Risk-Sensitivity

So far, we placed a probability distribution over the misspecification parameters and minimized posterior expected loss. There is a growing literature in economics[9] that studies the robustness of decision rules to model misspecification. Underlying this robustness analysis is typically a static or dynamic two-person zero-sum game. The decision maker, in our case the central bank, is minimizing her loss function while a malevolent fictitious other, "nature", chooses the misspecification to harm the decision maker. "Nature's" choice, in our notation $\Psi^\Delta$ and $\Sigma^\Delta$, is either limited to a bounded set or it is subject to a penalty function that is increasing in the size of the misspecification. The policy maker's decision is robust, if it corresponds to a Nash equilibrium in the two-person game.

In the Bayesian framework the risk sensitivity that is inherent in the robust control approach can be introduced by transforming the loss function. Instead of minimizing the expected value of $L(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta)$, the policy maker is equipped with an exponential utility function. She considers the transformed loss $e^{\tau L}$, and solves

$$\min_{\theta_p} \quad \frac{1}{\tau} \ln \int \exp\{\tau L(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta)\} p(\theta_s, \Psi^\Delta, \Sigma^\Delta) d(\theta_s, \Psi^\Delta, \Sigma^\Delta), \tag{39}$$

where $p(\theta_s, \Psi^\Delta, \Sigma^\Delta)$ denotes the joint density of $\theta_s$, $\Psi^\Delta$, $\Sigma^\Delta$. A positive $\tau$ makes the policy maker risk averse. It can be shown that the optimization of (39) is the solution to the following zero-sum game

$$\begin{aligned} \min_{\theta_p} \max_{q(\theta_s, \Psi^\Delta, \Sigma^\Delta)} \quad & \int L(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) p(\theta_s, \Psi^\Delta, \Sigma^\Delta) q(\theta_s, \Psi^\Delta, \Sigma^\Delta) d(\theta_s, \Psi^\Delta, \Sigma^\Delta) \\ & -\frac{1}{\tau} \int \Big( \ln q(\theta_s, \Psi^\Delta, \Sigma^\Delta) \Big) p(\theta_s, \Psi^\Delta, \Sigma^\Delta) q(\theta_s, \Psi^\Delta, \Sigma^\Delta) d(\theta_s, \Psi^\Delta, \Sigma^\Delta). \end{aligned} \tag{40}$$

The maximization with respect to $q(\cdot)$ is subject to the constraints

$$\int p(\theta_s, \Psi^\Delta, \Sigma^\Delta) q(\theta_s, \Psi^\Delta, \Sigma^\Delta) d(\theta_s, \Psi^\Delta, \Sigma^\Delta) = 1, \quad q(\theta_s, \Psi^\Delta, \Sigma^\Delta) \geq 0.$$

[9]See for instance, the monograph by Hansen and Sargent (2005) or the February 2002 special issue of *Macroeconomic Dynamics.*

The interpretation of this game is that "nature" chooses the function $q(\cdot)$ to distort the probabilities from which the model (misspecification) parameters are drawn. Notice that

$$\begin{aligned} &\int [\ln q(\cdot)] p(\cdot) q(\cdot) d(\theta_s, \Psi^\Delta, \Sigma^\Delta) \\ &= \int [\ln p(\cdot) q(\cdot)] p(\cdot) q(\cdot) d(\theta_s, \Psi^\Delta, \Sigma^\Delta) - \int [\ln p(\cdot)] p(\cdot) q(\cdot) d(\theta_s, \Psi^\Delta, \Sigma^\Delta) \end{aligned}$$

is the Kullback-Leibler discrepancy between the distorted and the undistorted probabilities. The larger $\tau$, the larger the penalty for deviating from $p(\cdot)$. The link between the exponential transformation of the loss function and the zero-sum game representation was pointed out by Jacobsen (1973) in one of the first studies of optimization under a risk-sensitive criterion.

In the subsequent empirical analysis we will also compute posterior expected losses under Scenarios 1 and 2 for the risk sensitive version of the policy problem. Under Scenario 1 the DSGE model itself is assumed to be correctly specified but "nature" is allowed to distort the believes about the structural parameters. The larger the uncertainty about a parameter, the easier it is to shift probability mass to create havoc. This analysis is similar in spirit to Giannoni (2002) who assess the robustness of monetary policy to changes in non-policy parameters of a simple three-equation New Keynesian DSGE model. Under Scenario 2 nature can also distort believes about the misspecification matrices, that is, the deviation of the "true" law of motion from the DSGE model restrictions and hence our analysis becomes comparable to Onatski and Stock (2002) and Onatski and Williams (2003).

# 5 Policy Evaluation Under DSGE-VAR

Our empirical analysis is based on interest rate, inflation, and output gap time series. We use the CBO's potential output series and obtain all other series from Haver Analytics (Haver mnemonics are in italics). The output gap is defined as the log difference of real GDP (nominal *GDP* divided by the chained-price deflator *JGDP*) and real potential output. The log differences are scaled by 100 to convert them in percentages. Inflation is computed using quarter-to-quarter log-differences of the GDP deflator, scaled by 400 to obtain annualized percentages. The nominal rate corresponds to the effective Federal Funds rate (*FFED*), also in percent. The results reported below are based on a sample from 1983:Q3 to 2004:Q1. We begin with the estimation of the state-space representation of the DSGE model and of the DSGE-VAR for different values of $\lambda$. We document the degree of misspecification of the DSGE model. Next, we discuss the estimates of the "deep" parameters and the extent to which they are identified. Finally, we proceed with the policy analysis.

## 5.1 Estimation

We begin the section by discussing our assumptions on the priors for the "deep" parameters. Since we do not use observations on consumption and investment, it is difficult to identify the capital share and the depreciation rate. Therefore, we let $\alpha = 0.25$ and $\delta = 0.025$. Moreover, in a log-linear approximation the price markup parameter $\lambda_f$, which we fix at 0.3, is typically not identifiable. The parameters $\chi$ and $\nu_m$ only affect the dynamics of the money stock, which is not included in the set of observables. The parameter $\varphi$ determines the steady state labor supply and does not influence the dynamics of interest rates, inflation, and the output gap. We set the average annualized growth rate of the economy $\gamma_a$ equal to 1.5%.

Priors for the remaining DSGE model parameters are provided in Table 1. All intervals reported in the text are meant to be 90% intervals. The distribution for $\psi_1$ and $\psi_2$ is approximately centered at Taylor's (1993) values, whereas the smoothing parameter lies in the range from 0.18 to 0.83. The prior mean for the growth adjusted real interest rate, $r_a^* + \gamma_a$, is 2.5% and annualized steady state inflation ranges from 0 to 6.25%, which is consistent with pre-1982 long-run historical averages. The prior mean of $g^*$ implies that that the government share of GDP is 15%. According to our prior the habit persistence parameter $h$ lies between 0.55 and 0.85. Boldrin, Christiano, and Fisher (2001) found that a value of 0.7 enhances the ability of a standard DSGE model to account for key asset market statistics. The interval for $\nu_l$ implies that the Frisch labor supply elasticity lies between 0.3 and 1.3, reflecting the micro-level estimates at the lower end, and the estimates of Kimball and Shapiro (2003) and Chang and Kim (2005) at the upper end. According to the prior for $\zeta_p$ firms re-optimize their prices on average every 2.5 to 12.5 quarters. This interval encompasses findings in micro-level studies of price adjustments such as Bils and Klenow (2004). The prior for the adjustment cost parameter $s''$ is consistent with the values that Christiano, Eichenbaum, and Evans (2005) when matching consumption and investment DSGE impulse response functions, among others, to VAR responses. Our prior for $a'$ implies that in response to a 1% increase in the return to capital utilization rates rise by 0.1 to 0.3%. These numbers are considerably smaller than the one used by Christiano, Eichenbaum, and Evans (2005). Finally, the priors for $\rho_z$ and $\rho_g$ are centered around 0.8. The priors for the standard deviation parameters are chosen to obtain realistic magnitudes for the implied volatility of the output gap, inflation, and interest rates. These priors are by and large similar to the ones that have been used elsewhere in the literature, e.g., Smets and Wouters (2003), Del Negro, Schorfheide, Smets, and Wouters (2004), and Levin, Onatski, Williams, and Williams (2005).

We estimate the state-space representation of the DSGE model using the Bayesian techniques described in Schorfheide (2000) and the DSGE-VARs with the Gibbs sampler discussed in the Appendix for various values of $\lambda$. Although $\lambda$ is in principle a continuous parameter, for computational reasons we consider only 8 values on a grid ranging from 0.25, i.e., large prior variance of the misspecification matrices $\Psi^{\Delta}$ and $\Sigma^{\Delta}$, to 10, which implies small misspecification. The DSGE-VAR analysis is based on $p = 4$ lags.

Table 2 summarizes the posterior of the hyperparameter $\lambda$. Log marginal data densities are reported in column 2 of the table. Differences of log marginal densities across $\lambda$'s can be interpreted as log posterior odds, under the assumption that the prior odds are equal to one. The odds reported in the last column of Table 2 are relative to $\lambda = 0.5$, which is the specification with the largest marginal data density and, according to this likelihood-based criterion, the best fit. The posterior of $\lambda$ has an inverted $U$-shape. There is little variation in the marginal data densities for $\lambda$ values between 0.50 and 1, whereas values outside of this interval lead to a substantial deterioration in fit. In the first row of the Table 2 we report the marginal data density for the DSGE model, which is about 30 points lower than the density of the DSGE-VAR($\lambda = 0.5$) on a log scale. We conclude that over the range of the historical sample the DSGE model is strongly dominated by DSGE-VARs with fairly low values of $\lambda$, indicating that the structural model is to some extent misspecified and that its policy predictions should be interpreted with care. However, the results of the next section indicate that even when $\lambda$ is as low as .5 much of the mechanics of the DSGE model carries over to the DSGE-VAR. All DSGE-VAR results reported subsequently are based on $\lambda = 0.5$.

Parameter estimation results for the DSGE-VAR and the state-space representation of the DSGE model are reported in Table 3. There is a growing literature highlighting parameter identification problems associated with New Keynesian DSGE models, e.g. Beyer and Farmer (2004), Canova and Sala (2005), and Lubik and Schorfheide (2004, 2005). In some cases the rational expectations solution of the DSGE model implies that a subset of structural parameters disappear from the reduced form law of motion of the observables, in other cases the estimation objective function may have little curvature in some directions. Straightforward manipulations of Bayes Theorem can be used to show that priors are not updated in directions of the parameter space in which the likelihood function is flat, e.g. Poirier (1998). Hence, to enable a careful comparison of priors and posteriors and an assessment of the information extracted from the sample we also report means and confidence intervals for the prior in Table 3. The summary statistics for the prior reflect the truncation at the boundary of the determinacy region. For the purpose of this study we are mostly interested in the estimation of the degree of misspecification $\lambda$ and of policy loss differentials. With respect to $\lambda$, lack of identification of the deep parameters is not a concern. Therefore, we are only concerned about lack of sample information with respect to those parameters that significantly affect the ranking of policies.

The policy parameter estimates obtained from the DSGE-VAR can be viewed as Bayesian instrumental variable estimates. Since inflation and the output gap are endogenous variables, the estimator of $\psi_1$ and $\psi_2$ has to be adjusted for the non-zero conditional expectation of the monetary policy shock. Both the likelihood function associated with the state-space representation of the DSGE model and the DSGE-VAR likelihood generate such an adjustment. The former imposes all cross-coefficient restrictions of the DSGE model whereas the latter relaxes the restrictions. The estimates of $\psi_1$ are 2.19 for the DSGE-VAR and 1.90 for the DSGE model itself, implying a strong response of the Fed to inflation in the post-Volcker era. The estimated degree of interest rate smoothing is about 0.8. A comparison of prior and posterior means for $\psi_2$ indicates that the location shift is fairly small. The 90% posterior probability interval is however much tighter than the prior interval reflecting the information about the central bank's response to the output gap contained in the data. The estimated interest-feedback rate rule is admittedly a stylized description of monetary policy in the Volcker-Greenspan years. Yet the $R^2$ of the policy rule equation, obtained using the parameters that maximize the DSGE model posterior, is 94% for our sample period. This number is not too far from the 97% percent obtained by Blinder and Reis (2005) for the 1987:Q3 to 2005:Q1 period using a more sophisticated rule. We have experimented with a four-quarter moving average of inflation as an argument of the rule. We found that this does not affect the estimates of the other parameters, and slightly worsens the overall fit of the model.

To understand the remaining parameter estimates obtained from the DSGE-VAR it is instructive to consider the likelihood function, which is given by

$$
\begin{aligned}
&p(Y|\Psi,\Sigma,\theta) \\
&\propto \quad |\Sigma|^{-T/2}\mathrm{etr}\left[\Sigma^{-1}\Big(Y - X(B_1(\theta)+\Psi B_2(\theta))\Big)'\Big(Y - X(B_1(\theta)+\Psi B_2(\theta))\Big)\right].
\end{aligned} \tag{41}
$$

We show in the Appendix that the mode of the posterior distribution of $\Psi$ conditional on $\Sigma$ and $\theta$ is of the form

$$
\tilde{\Psi}(\Sigma,\theta) = (\lambda T\Gamma_{XX} + X'X)^{-1}(\lambda T\Gamma_{X\tilde{Y}} + X'\tilde{Y})\Sigma^{-1}B_2'(B_2\Sigma^{-1}B_2')^{-1},
$$

where $\tilde{Y}(\theta) = Y - XB_1(\theta)$. Since we are unable to compute a marginal likelihood function $p(Y|\theta)$ analytically, we consider the concentrated likelihood function instead $p(Y|\tilde{\Psi}(\Sigma,\theta),\Sigma,\theta)$. Notice that as $\lambda$ approaches zero the non-policy parameters, which do not affect $B_1(\theta)$ and $B_2(\theta)$ vanish from $p(Y|\tilde{\Psi}(\Sigma,\theta),\Sigma,\theta)$ because $\tilde{\Psi}(\Sigma,\theta)$ approaches $(X'X)^{-1}X'\tilde{Y}\Sigma^{-1}B_2'(B_2\Sigma^{-1}B_2')^{-1}$. Hence, as the value of $\lambda$ decreases we would expect the posterior distribution of the non-policy parameters to resemble more closely the prior distribution. Indeed, we found this to be the case.

The DSGE model estimates for the non-policy parameter point toward a large degree of habit persistence, $\hat{h} = 0.92$, and a small elasticity of investment with respect to the value of installed capital $1/\widehat{s''} = 0.12$ implying fairly large capital adjustment costs. The estimate of the Calvo parameter $\zeta_p$ is 0.59, indicating a fairly low degree of price-stickiness: Agents re-optimize their prices on average every 2.5 quarters. We find that all shocks are fairly persistent, but our autocorrelation estimates are not as large as in other studies, e.g. Smets and Wouters (2003). The upper bound of the 90% posterior intervals for $\rho_z$ and $\rho_g$ are well below 1. The likelihood function of the DSGE model provides little information on $g^*$, and $a'$. While our values for $g^*$ can be justified by the historical government share, the determination of $a'$ is more difficult. We decided to conduct a robustness analysis by comparing DSGE model-based loss differentials for various values of $a'$ and we found that the losses and the ranking of policies were insensitive to this parameter.

For many parameters, the estimates obtained from the DSGE-VAR are by and large similar to the the DSGE model estimates. One exception is the standard deviation of demand shocks $\sigma_g$, which is much lower according to the DSGE-VAR. Other exceptions are $h$ and $s''$, for which the DSGE-VAR likelihood function contains little information. We have compared the loss differentials under Scenario 1 for both the DSGE model and the DSGE-VAR, and found that the the differences in the estimates of the deep parameters between the two are quantitatively not important.

## 5.2 Policy Outcomes Under Misspecification

This section studies how policy outcomes change when misspecification is taken into account. We proceed to analyze the loss from different policies under the four different assumptions regarding misspecification described in Section 4.2. Specifically, we evaluate the loss (38) under each scenario as a function of all the parameters characterizing the interest rate rule (13): $\psi_1$ and $\psi_2$, the central bank's response to inflation and output, respectively, and

$\rho_R$, the interest rate smoothing parameter. We compute this loss for each point of a three dimensional grid, where: $\psi_1$ takes nine values ranging from 1.001 to 3 in intervals of 0.25; $\psi_2$ takes six different values, computed taking the Taylor's (1993) value $\psi_2^T = 0.125$ as a reference, namely 0, $\frac{1}{2}\psi_2^T = 0.062$, $\psi_2^T = 0.125$, $\frac{3}{2}\psi_2^T = 0.188$, $2\psi_2^T = 0.250$, $3\psi_2^T = 0.375$; and $\rho_R$ takes three different values, namely 0.7, 0.8, and 0.95.[10] Finally, we compute the loss differential relative to the benchmark $\psi_1 = 2$, $\psi_2 = 0.188$, $\rho_R = 0.8$, and take expectations. The benchmark is chosen by selecting the point in the grid that roughly corresponds to the estimated values for those parameters, i.e., the historical Volcker-Greenspan rule. Negative differentials therefore indicate an improvement relative to the Volcker-Greenspan rule. The loss function has a coefficient of 1 on inflation, implying that the loss-differentials can be easily translated into an "inflation-equivalent" number. That is, a policy with an associated loss differential of 10 implies that under it the policymaker is as worse of as if she suffered an increase in the variance of inflation from 2 to 12, and no change in the variance of the other two variables.

The results are summarized in Figures 2 and 3. Figure 2 contains four charts, one for each scenario. Each chart shows three surfaces describing the expected loss as a function of $\psi_1$ and $\psi_2$. The three surfaces correspond to the different values for $\rho_R$ (0.7, 0.8, and 0.95), with darkness of the surface being directly proportional to $\rho_R$. Figure 3 depicts a slice of this three-dimensional plot, obtained by further fixing $\psi_2$ to its benchmark value of 0.188.

The first of these scenarios, which we refer to as "Ignore Misspecification," amounts to evaluating policy as if the New Keynesian model correctly described the data. This first scenario is the natural benchmark for comparing the results obtained once we allow for the presence of misspecification. Although the model considered here has a number of additional rigidities relative to the stylized model considered in Woodford (2003), the policy recommendation emerging from the analysis of the expected loss differentials are in line with those in Woodford, at least qualitatively. First, a high response of the interest rate to inflation (high $\psi_1$) is preferred to a low one, regardless of the value taken by $\psi_2$ and $\rho_R$. The drop in the loss is particularly steep as $\psi_1$ increases from 1.001 to 1.5, but flattens thereafter, as can be appreciated from the two dimensional plot in Figure 3. The mechanism underlying this result is well known: due to the forward looking nature of the model an increase in $\psi_1$ results in a drop of inflation variability, which in turn implies – for given $\rho_R$ – a lower volatility of interest rates.

[10] We have experimented with a finer grid, especially for $\psi_1$ and $\rho_R$, and we have found the loss to be a smooth function of these parameters. Hence, for ease of exposition we focus on the coarser grid.

The top panel of Figure 4, which shows the impulse responses for the DSGE model as a function of $\psi_1$, illustrates this mechanism. The panel displays the impulse responses with respect to the technology shock $\epsilon_{z,t}$ and the demand shock $\epsilon_{g,t}$ for three different values of $\psi_1$, 1.25, 2, and 2.75, keeping $\psi_2$ and $\rho_R$ at their historical values. Recall that these shocks are the only ones that matter for the loss calculation since we set $\sigma_R = 0$ and exclude the monetary policy shocks. The figure shows that the response of inflation and the interest rate to both shocks is less strong the higher the value of $\psi_1$. Quantitatively, it is the response to technology shocks that makes the difference, with inflation responding roughly twice as strongly for $\psi_1$ equal to 1.25 than for $\psi_1$ equal to 2.75. The response of interest rates to both shocks is slightly stronger on impact the higher the value of $\psi_1$, but becomes weaker soon thanks to the fact that inflation is kept under control. The output gap is more volatile for high $\psi_1$ with respect to technology shocks, but quantitatively the difference is not large. With respect to demand shocks the response of the output gap is strong on impact, but is not affected by the size of $\psi_1$.

The expected loss differential as a function of $\rho_R$ depends on the values of $\psi_1$ and $\psi_2$, as shown in Figures 2. For low values of $\psi_1$ and high values of $\psi_2$, a higher interest rate inertia is preferred. As the value of $\psi_2$ decreases and, especially, that of $\psi_1$ increases, the expected loss differential as a function of $\rho_R$ becomes flatter. Figure 3 shows that for values of $\psi_1$ higher than 1.25 the optimal value of $\rho_R$ is 0.8. The impulse responses for the DSGE model as a function of $\rho_R$, shown in the top panel of Figure 5, shows the trade-off implied by the choice of $\rho_R$ for given $\psi_1$ and $\psi_2$. A value of $\rho_R$ equal to 0.95 implies a drastic reduction in the variability of interest rates relative to the baseline $\rho_R = 0.8$, but a slightly increased variability of inflation. Given the specific coefficients of our loss function, this trade-off is resolved by choosing a value of $\rho_R$ that is neither too high (here, 0.95) or too low (0.7). It is clear however that the outcome depends on the specific choice of the loss function.

Figure 2 shows that the expected loss differential as a function of $\psi_2$ is sharply decreasing for low values of $\psi_1$ and $\rho_R$: Targeting the output gap in presence of a weak interest rate response to inflation results in high inflation and interest rate variability. The slope of the loss function with respect to $\psi_2$ decreases as $\rho_R$, but especially $\psi_1$, increase. For high values of $\psi_1$ the expected loss differential appears from Figure 2 almost invariant with respect to $\psi_2$. In fact, it is slightly U-shaped, with a minimum for the historical value of $\psi_2 = 0.188$. Table 4 shows that the best performing policy when the analysis is conducted ignoring misspecification as $\psi_1$ at its highest value in the grid, 3.00, while $\psi_2$ and $\rho_R$ are at their estimated values of 0.188 and 0.8, respectively. The inference about the misspecification parameter $\lambda$ in Table 2 casts some doubts on the reliability of DSGE model predictions, however. Hence we move to take potential misspecification into account.

In Scenario 2 the policymaker still uses the DSGE model to compute the mean response of the endogenous variables to the change in the policy parameters, but recognizes that "nature" may be injecting noise around these mean responses using the prior distribution. Under this scenario the policymaker learns from the data about the overall amount of noise ($\lambda$) but refuses to learn about the precise nature of the misspecification. We therefore refer to this scenario as "Acknowledge Misspecification – Discard the Past," where "Discard the Past" refers to the fact that under this scenario the policymaker refuses to use the posterior information regarding $\Psi^{\Delta}$ and $\Sigma^{\Delta}$ on the ground that it contains no useful information once policy changes. Rather, she uses the prior distribution to generate draws of $\Psi^{\Delta}$ and $\Sigma^{\Delta}$ in evaluating policy outcomes. In both Figures 2 and 3 we focus on the expected loss differentials computed for $\lambda = .5$, which represents the "best-fitting" model. We have computed the expected loss differentials for the entire sequence of $\lambda$ values included in Table 2 and have found that the policy outcomes for intermediate values of $\lambda$ conform with our expectations. That is, the results are pretty similar to those for the DSGE model when $\lambda$ equals 5 or 10, and as $\lambda$ decreases they become more similar to those obtained for $\lambda = 0.5$. Moreover, we find that for $\lambda$ values between 0.50 and 1 – that is, for all the model that would receive non-neglible weights in a Bayesian posterior averaging calculation – the loss function is both qualitatively and quantitatively very similar. Therefore we focus on $\lambda = 0.5$.

Under Scenario 2 the qualitative policy implications obtained under the DSGE model remain valid. Figures 2 and 3 show that a high response of the interest rate to inflation is preferred to a low one, and that for high enough $\psi_1$ the best choice of $\rho_R$ equals the historical value of 0.8. A small difference is that under Scenario 2 the policy maker should not respond at all to the output gap. Indeed, Table 4 shows that the best performing policy under Scenario 2 is identical to that under Scenario 1, except that $\psi_2$ equals 0. The cost of increasing $\psi_2$ is fairly small, as long as $\psi_1$ is reasonably high. Quantitatively, the expected loss differentials also appear very similar under Scenarios 1 and 2.[11] Under Scenario 2

[11]This is a bit misleading, however, as the underlying draw of the deep parameters $\theta$ is different under the two scenarios – one corresponds to the DSGE model and the other to DSGE-VAR (the values are given in Table 3). Whenever we computed the loss under Scenario 1 using the DSGE-VAR draws for $\theta$, we find that the shape of the loss functions is exactly the same, but the expected loss differentials is smaller, largely because the estimated standard deviation of demand shocks is smaller. When we use the same set of draws for the deep parameters under both scenarios, we see that allowing for misspecification actually enhances the loss differentials, particularly those associated with small values of $\psi_1$. Quantitatively, the difference is

(as under the prior) the misspecification matrices $\Psi^{\Delta}$ and $\Sigma^{\Delta}$ have zero mean. Because of non-linearities in the impulse-response functions, the zero mean in the misspecification matrices does not imply that the expected loss differentials are the same with and without misspecification. But in our application they are roughly the same. This does not imply that introducing misspecification under Scenario 2 bears no consequences for the policy maker. The impact of misspecification can be felt in the uncertainty surrounding a given policy outcome. This uncertainty averages out in Figures 2 and 3, but can play a substantial role in a risk-sensitive analysis. We will return to this point later.

In Scenario 3 the policymaker uses sample information to learn about the precise nature of misspecification, unlike in the previous scenario. In addition, she believes that the historically observed discrepancies $\Psi^{\Delta}$ and $\Sigma^{\Delta}$ are policy invariant. We refer to this scenario as "Learn about Misspecification (Policy Invariant)." Figure 2 shows that under this scenario the loss surface is quite different than under Scenarios 1 and 2. First of all, there is much more curvature with respect to $\psi_2$. Second, the loss profile is no longer strictly decreasing in $\psi_1$, at least for values of $\rho_R$ less than 0.95, as can be seen from Figure 3. For small $\rho_R$s the loss differential is a U-shaped function of $\psi_1$, with the minimum attained at the value of 1.5 for $\rho_R = 0.8$.

The contours of the loss differentials under Scenario 3 are almost exclusively driven by the presence of explosive roots, at least for values of $\rho_R$ less than 0.95. Explosiveness is not a concern at all for Scenarios 1 (mechanically), and 2 (as is non-uniqueness of the rational expectation equilibrium). For Scenario 4 the fraction of explosive draws is very small. For Scenario 3 explosiveness is ubiquitous. Hence, the policy recommendation under Scenario 3 is largely driven by the desire to avoid explosiveness.

We view this policy recommendation with suspicion, because it results from ignoring the correlation between the policy parameters and the size of misspecification in models that have a backward-looking component. We now elaborate on this point for the case of DSGE-VAR. Recall that the estimated VAR parameters $\hat{\Psi}$ can be decomposed into the sum of the parameters implied by the DSGE restrictions $\Psi^*(\theta)$ and of the misspecification $\Psi^{\Delta}$. Roughly speaking, under Scenario 3, the new set of VAR parameters is computed as the sum of $\Psi^*(\tilde{\theta})$, which changes with policy, and $\Psi^{\Delta}$, which is assumed to be invariant. If the policy parameters are close to estimated ones, the sum of $\Psi^*(\tilde{\theta})$ and $\Psi^{\Delta}$ returns the estimated VAR parameters, which do not have explosive roots. But as we move away from the estimated policy parameters, or as we ignore the correlation between the policy parameters and $\Psi^{\Delta}$,

not very large.

the sum of $\Psi^*(\tilde{\theta})$ and $\Psi^{\Delta}$ can deliver new VAR parameters whose roots are explosive. Here, this is particularly true for low values of $\rho_R$.[12] Inserting a policy rule different from the estimated one into a backward-looking model can often produce explosiveness.[13] This explosiveness is driven by the fact that the backward looking component of the system (here, $\Psi^{\Delta}$) is not allowed to change with policy. Here, for instance, we find that a number of policies that are not too different from the historical one (say, $\psi_1$ between 1.75 and 2.5, $\psi_2 = 0.188$, and $\rho_R = 0.8$) delivers a non-negligible probability of explosiveness. We see this as a warning against having the backward-looking component of the model invariant with policy, rather than a warning against these policies.

Scenario 4 is an attempt to address this concern. Under this Scenario the policymaker again uses sample information to learn about potential model misspecification. As in Scenario 3, the misspecification is backward looking. But unlike in Scenario 3, the policymaker takes into account the in-sample correlation between the policy parameters and the misspecification matrices $\Psi^{\Delta}$ and $\Sigma^{\Delta}$. Specifically, the policymaker now asks the question: What would the estimates of the discrepancies $\Psi^{\Delta}$ and $\Sigma^{\Delta}$ be if the new policy had been in place during the sample period? Explosiveness is no longer an overriding concern in Scenario 4, as it was in Scenario 3. The reason for this result is that now as $\tilde{\theta}$ changes both $\Psi^*(\tilde{\theta})$ and $\Psi^{\Delta}(\tilde{\theta})$ change in such a way that the sum of the two is not too different from the estimated VAR parameters. Indeed, for $\lambda$ close to zero the dynamics for all equations other than the interest rate rule are approximately independent of the policy parameters.

Figure 2 shows that under Scenario 4 the loss profile becomes flatter. In particular, the large drop in the expected loss differential that characterized the increase in $\psi_1$ from 1.001 to 1.5 under the DSGE model nearly disappears under Scenario 4, as can be appreciated from Figure 3. Under the DSGE model the mechanics of the rational expectations equilibrium imply that high values of $\psi_1$ help to anchor inflationary expectations. Since in equilibrium inflation moves less than under high values of $\psi_1$, interest rates need to move less as well. The presence of substantial misspecification changes these dynamic responses, as shown in the bottom panel of Figure 4. As $\psi_1$ increases the response of inflation to both technology and demand shocks is more subdued, as in the DSGE model. Unlike in the DSGE model,

[12] Whenever $\lambda$ is large, however, $\Psi^{\Delta}$ is negligible and the new VAR parameters roughly coincide with $\Psi^*(\tilde{\theta})$, which is non explosive.

[13] There are several examples in the literature. For instance, Levin and Williams (2003) find that many of the rules they consider generate explosiveness in the backward-looking Rudebusch-Svenson model. Likewise, Cogley and Sargent (2005) find that the optimal policy computed under their rational expectations "Lucas" model leads to explosive behavior when plugged into the backward-looking "Solow-Samuelson" model.

the interest rate becomes more volatile as $\psi_1$ increases, not only on impact but also in the medium run. A decomposition of the loss into its three components indeed confirms that interest rate variability rises as the central bank responds more strongly to inflation. If the policymaker only cared about inflation, she would still choose a high response to inflation even under misspecification. But the fact that according to our policy rule she cares about interest rate variability as well leads to a trade-off between the two, and hence to a flatter loss contour. Figure 3 shows that as $\psi_1$ rises past 1.75, the loss starts to increase, albeit very moderately.

The presence of misspecification also affects the policy prescription with respect to the degree of inertia $\rho_R$. Figure 5 shows that under the DSGE model high interest rate inertia leads to slightly higher inflation variability, but quantitatively the effect is small. This is because in the DSGE model agents are aware that the eventual increase in interest rates following a sustained increase in inflation would be large, even if today's response is small. Under misspecification, this mechanism does not come into play: The milder response of interest rates to shocks under high inertia leads to a substantial increase in the volatility of inflation. Finally, Scenario 4 shares with Scenario 2 an aversion against high responses to the output gap. Indeed, the best performing policy under Scenario 4 has $\psi_1 = 1.75$, $\psi_2 = 0.06$ and $\rho_R = 0.8$ (Table 4).

The results in Figures 2 and 3 depend on the somewhat arbitrary choice of the bound. For this reason, we have recomputed all figures using a bound that is double (100) or ten times larger (500) than the one used so far. Although the loss differentials change substantially with the bound, particularly in Scenario 3, we find that the overall shape of the contours, and hence the gist of our conclusions, stay roughly the same.

Finally, Figure 6 compares the loss differentials that we just analyzed with those obtained under the risk-sensitive version of our problem. We focus in particular on Scenario 2, which is the closest in spirit to the robustness literature, but we also show the benchmark, Scenario 1. For both scenarios the Figure shows the risk-sensitive loss (black) as well as the risk-neutral loss (light grey).[14] For Scenario 1, where the risk-sensitivity is only with respect to the deep parameters $\theta_s$, we find that the risk sensitive loss is generally not too different from the plain-vanilla one. The only difference is quantitative: The loss stemming from a weak response to inflation is much larger under the risk-sensitive calculations. In

[14] A caveat of our analysis so far is that we do not distinguish between uncertainty in the deep parameters $\theta$ and in the misspecification parameters $\Psi^\Delta$ and $\Sigma^\Delta$. In principle we want to be robust gains the latter, but not necessarily the former.

Scenario 2, risk-sensitivity would induce the policymaker to avoid not only low values of $\psi_1$, but also high values of $\psi_2$ as long as $\psi_1$ is not above 2. Recall that misspecification alone had little effect on the expected loss differential in Scenario 2. However, when combined with a concern for robustness it leads to a starker recommendation relative to the DSGE model: Avoid the combination of moderate interest rate responses to inflation and strong responses to the output gap.

In Table 5 we report loss differentials for alternative policy rules that have been considered in the literature. Equation (19) provides the basis for the conversion of policy rule coefficients into our setup. Taylor considers five benchmark policy rules described in Table 1 of his 1999 volume on monetary policy rules. Taylor I and II are interest rate smoothing rules with a coefficient $\rho_R = 1$. These rules were favored in the simulations by Levin, Wieland, and Williams (1999) who studied the performance of interest-rate rules in four different structural macroeconometric models of the U.S. economy: the Federal Reserve Board staff model, the Monetary Studies Research model, the Fuhrer-Moore model, and Taylor's multi-country model. In our DSGE model $\rho_R = 1$ rules make interest rates non-stationary and the calculation of the model-based autocovariance matrices that is needed for the construction of the prior distribution would have to be adjusted. We decided to replace the coefficient of unity by $\rho_R = 0.95$.[15] Taylor III is Taylor's (1993) rule which does not entail interest rate smoothing, whereas under the Taylor IV rule the central bank responds more strongly to the output gap. The fifth rule in Taylor (1999) is a super-inertia rule ($\rho_R = 1.3$) which we omit from our analysis. Instead, Rules 5 and 6 considered in Table 5 correspond to the robust Bayesian and minimax rules reported in Levin and Williams (2003).[16] Levin and Williams' (2003) calculation of robust rules is based on three models, a New Keynesian DSGE model, the Rudebusch-Svenson (1999) model, and Fuhrer's (2000) habit persistence model.

Based on our previous discussion it is not surprising that the $\rho_R = 0$ rules Taylor III and Taylor IV perform poorly across all four scenarios. The rules Taylor I and Taylor II, on the other hand, perform very well under Scenarios 1 and 2 since they imply a strong response to inflation movements. In fact, the Taylor I rule performs better than our preferred rules for Scenarios 1 and 2 because we restrict the value of $\psi_1$ to be less than 3 in our calculation of the preferred rule. Quantitatively the difference is not large. Levin and Williams' (2003) robust rules also perform well under Scenarios 1 and 2. This may stem from the fact that

[15]The interest rate and output gap coefficients are converted according to $\psi_1 = g_\pi/0.05$ and $\psi_2 = g_y/(4 \cdot 0.05)$.

[16]Taken from the $\lambda = 0.5$ and $\phi = 0.1$ entry in Tables 4 and 7 of Levin and Williams (2003) and converted according to $\psi_1 = 1 + \alpha/(1-\rho)$ and $\psi_2 = \beta/(4 \cdot (1-\rho))$.

the authors included a New Keynesian DSGE model in their analysis of monetary policy rules. However, Rules 5 and 6 are dominated by Taylor I across all scenarios. All the rules considered in Table 5 perform somewhat poorly in the "learn about misspecification" scenarios compared to our preferred rules presented in Table 4. This is the case even if we discount Scenario 3 for the reasons discussed above. One important exception is the Taylor I rule, which performs reasonably well under Scenario 4. A strong response of interest rates to inflation leads to policy outcomes that are not too far from that attained by the optimal rule even under Scenarios 4.

We find that the particular values of the loss differentials, that is, the risks and the gains associated with deviating from the historical Volcker-Greenspan policy, are sensitive to the misspecification assumptions considered. In particular, when the policymaker chooses to learn from the historical data about misspecification (Scenario 4), she finds that the loss associated with policies that deviate from the DSGE model prescriptions are not as large as policy analysis under the DSGE model would suggest. At the same time, we find that following those prescriptions, even under misspecification, leads to outcomes that are not substantially inferior to those of the best-performing rule. In summary, a fairly robust policy recommendation emerges from our analysis: the central bank should avoid strong responses to output gap movements and not react weakly to inflation fluctuations. Moreover, we find that the best-performing rules are those with a substantial degree of inertia in the policy instrument. Also, we find that the gains associated with deviating from the historical Volcker-Greenspan policy, whenever positive, are generally not very large. This suggests that the historical rule, if not always optimal among those we consider, has been reasonably good at least from the perspective of this sticky-prices DSGE model, even taking misspecification into account.

# 6 Conclusion

Encouraged by the work of Smets and Wouters (2003) many central banks are in the process of developing and estimating DSGE models usable for quantitative monetary policy analysis. The empirical results in this paper and in Del Negro, Schorfheide, Smets, and Wouters (2004) document that model misspecification remains a concern as less restrictive vector autoregressive specifications attain a better time series fit than the DSGE model itself. In this paper we developed and applied techniques to conduct quantitative monetary policy analysis with DSGE models while explicitly taking account of their potential misspecification.

## References

Adalid, Ramon, Günter Coenen, Peter McAdam, and Stefano Siviero (2005): "The Performance and Robustness of Interest-Rate Rules in Models of the Euro Area," *ECB Working Paper*, **479**.

Altig, David, Lawrence J. Christiano, Martin Eichenbaum, and Jesper Linde (2002): "Technology Shocks and Aggregate Fluctuations," *Manuscript*, Northwestern University.

Bils, Mark and Peter Klenow (2004): "Some Evidence on the Importance of Sticky Prices". *Journal of Political Economy*, **112**, 947-985.

Blinder, Alan and Ricardo Reis (2005): "Understanding the Greenspan Standard," *Manuscript*, Princeton University.

Boldrin, Michele, Lawrence Christiano, and Jonas Fisher (2001): "Habit Persistence, Asset Returns, and the Business Cycle," *American Economic Review*, **91**, 149-166.

Brainard, William (1967): "Uncertainty and the Effectiveness of Policy," *American Economic Review*, **57**, 411-25.

Brock, William A., Steven N. Durlauf, and Kenneth D. West, 2004, "Model Uncertainty and Policy Evaluation: Some Theory and Empirics." Mimeo. University of Wisconsin.

Calvo, Guillermo (1983): " Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, **12**, 383-398.

Chang, Yongsung and Sun-Bin Kim (2005): "From Individual to Aggregate Labor Supply: A Quantitative Analysis based on Heterogeneous Agent Macroeconomy," *International Economic Review*, forthcoming.

Chow, Gregory C. (1975): *"Analysis and Control of Dynamic Economic Systems,"* John Wiley & Sons, New York.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, **113**, 1-45.

Coenen, Günter (2003): "Inflation Persistence and Robust Monetary Policy Design," *ECB Working Paper*, **290**.

Cogley, Timothy and Thomas Sargent (2005): "The Conquest of U.S. Inflation: Learning and Robustness to Model Uncertainty," *Review of Economic Dynamics*, forthcoming.

Craine, Roger (1979): "Optimal Monetary Policy with Uncertainty," *Journal of Economic Dynamics and Control*, **1**, 59-83.

Del Negro, Marco and Frank Schorfheide (2004): "Priors from Equilibrium Models for VARs," *International Economic Review*, **45**, 643-673.

Del Negro, Marco and Frank Schorfheide (2005): "Policy Predictions if the Model Doesn't Fit," *Journal of the European Economic Association*, Papers and Proceedings, forthcoming.

Del Negro, Marco, Frank Schorfheide, Frank Smets, and Raf Wouters (2004): "On the Fit and Forecasting Performance of New Keynesian Models," *Manuscript*, University of Pennsylvania.

Fernandez-Villaverde, Jesus, Juan Rubio-Ramirez, and Thomas J. Sargent (2005): "The A, B, C (and D's) of VARs," *Manuscript*, New York University.

Fuhrer, Jeffrey (2000): "Habit Formation in Consumption and its Implications for Monetary Policy Models," *American Economic Review*, **90**, 367-390.

Hansen, Lars P and Thomas J. Sargent (2005): "Misspecification in Recursive Macroeconomic Theory," *Manuscript*, New York University, Department of Economics.

Jacobsen, David H. (1973): "Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Differential Games," *IEEE Transactions on Automatic Control*, **AC-18**, 124-131.

Laforte, Jean-Phillipe (2003): "Comparing Monetary Policy Rules in an Estimated General Equilibrium Model of the U.S. Economy," *Manuscript*, Federal Reserve Board of Governors, Washington, DC.

Kimball, Miles and Matthew Shapiro (2003): "Labor Supply: Are the Income and Substitution Effects Both Large or Both Small?" *Manuscript*, University of Michigan, Department of Economics.

Levin, Andrew, Volker Wieland, and John C. Williams (1999): "Robustness of Simple Monetary Policy Rules under Model Uncertainty," in: John Taylor (ed): *Monetary Policy Rules*, NBER Business Cycle Series, Vol 31.

Levin, Andrew, Volker Wieland, and John C. Williams (2003): "The Performance of Forecast-Based Monetary Policy Rules Under Model Uncertainty," *American Economic Review*, **93**, 622-645.

Levin, Andrew, and John C. Williams (2003): "Robust Monetary Policy with Competing Reference Models," *Journal of Monetary Economics*, **50**, 945-975.

Levin, Andrew, Alexei Onatski, John Williams, and Noah Williams (2005): "Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models," *NBER Macroeconomics Annual*, forthcoming.

McCallum, Bennett T. (1988): "Robustness Properties of a Rule for Monetary Policy," *Carnegie-Rochester Conference Series on Public Policy*, **29**, 173-204.

Onatski, Alexei and James H. Stock (2002): "Robust Monetary Policy under Model Uncertainty in a Small Model of the US Economy," *Macroeconomic Dynamics*, **6**, 85-110.

Onatski, Alexei and Noah Williams (2003): "Modeling Model Uncertainty," *Journal of the European Economic Association*, **1**, 1087-1022.

Onatski, Alexei and Noah Williams (2004): "Empirical and Policy Performance of a Forward-Looking Monetary Model," *Manuscript*, Princeton University, Department of Economics.

Rudebusch, Glenn (2001): "Is the Fed Too Timid? Monetary Policy in an Uncertain World," *The Review of Economics and Statistics*, **83**, 203-217.

Rudebusch, Glenn (2002): "Assessing Nominal Income Rules for Monetary Policy with Model and Data Uncertainty," *The Economic Journal*, **112**, 402-432.

Rudebusch, Glenn and Lars Svenson (1999): "Policy Rules for Inflation Targeting," in: John Taylor (ed): *Monetary Policy Rules*, NBER Business Cycle Series, Vol 31.

Schorfheide, Frank (2000): "Loss-Function Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, **15**, 645-70.

Sims, Christopher A. (1982): "Policy Analysis with Econometric Econometric Models," *Brookings Papers on Economic Activity*, **1**, 107-64.

Sims, Christopher A. (1999): "The Role of Interest Rate Policy in the Generation and Propagation of Business Cycles: What has Changed Since the 30's?", *Proceedings* of the 1998 Boston Federal Reserve Bank Annual Research Conference.

Sims, Christopher A. (2002): "Solving Rational Expectations Models." *Computational Economics*, **20**(1-2), 1-20.

Smets, Frank and Raf Wouters (2003): "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, **1**, 1123-75.

Taylor, John B. (1993): "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, **39**, 195-214.

Taylor, John B. (1999): *"Monetary Policy Rules,"* NBER Business Cycles Series, **31**, Chicago.

Tetlow, Robert, and Peter von zur Muehlen (2001): "Robust Monetary Policy with Misspecified Models: Does Uncertainty always Call for Attenuated Policy?" *Journal of Economic Dynamics and Control*, **25**, 911-949.

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

# A Implementation of the Posterior Simulation

## A.1 Draws from the Posterior

We adopt the notation that $\tilde{Y}(\theta) = Y - XB_1(\theta)$ which leads to the definitions

$$\Gamma_{\tilde{Y}\tilde{Y}} = \Gamma_{YY} - \Gamma_{YX}B_1(\theta) - B_1(\theta)'\Gamma_{XY} - B_1(\theta)'\Gamma_{XX}B_1(\theta), \quad \Gamma_{X\tilde{Y}} = \Gamma_{XY} - \Gamma_{XX}B_1(\theta).$$

Let $\text{etr}[A] = \exp[-\frac{1}{2}tr[A]]$. The likelihood function for the VAR representation is given by

$$\begin{aligned} & p(Y|\Psi,\Sigma,\theta) \\ \propto \quad & |\Sigma|^{-T/2}\text{etr}\left[\Sigma^{-1}\Big(Y - X(B_1(\theta)+\Psi B_2(\theta))\Big)'\Big(Y - X(B_1(\theta)+\Psi B_2(\theta))\Big)\right]. \end{aligned} \tag{A.1}$$

Using Lemma 1(i) we can rewrite the prior mean of $\Psi$ as

$$\Psi^*(\theta) = \bar{\Psi}(\Sigma,\theta) = \Gamma^{-1}_{XX}(\theta)\Gamma_{X\tilde{Y}}(\theta)\Sigma^{-1}B_2'(\theta)[B_2(\theta)\Sigma^{-1}B_2'(\theta)]^{-1}.$$

The prior density for $\Psi$ conditional on $\Sigma$ is of the form

$$p(\Psi|\Sigma,\theta) \quad \propto \quad \text{etr}\left[\Sigma^{-1}\lambda T\Big(-2B_2'\Psi'\Gamma_{X\tilde{Y}}(\theta) + B_2'\Psi'\Gamma_{XX}(\theta)\Psi B_2\Big)\right]. \tag{A.2}$$

The prior density for $\Sigma$ is given by

$$p(\Sigma|\theta) \propto |\Sigma|^{-\frac{1}{2}(\lambda T-k+n+1)}\text{etr}\Big[\Sigma^{-1}\lambda T\Sigma^*(\theta))\Big] \tag{A.3}$$

To simplify notation the $(\theta)$-argument of the functions $B_1$, $B_2$, $\tilde{Y}$, $\Gamma_{XY}$, $\Gamma_{XX}$, and $\Gamma_{YY}$ is omitted.

**Conditional Posterior of $\Psi$:** Combining the the prior density (A.2) with the likelihood function (A.1) yields

$$\begin{aligned} & p(\Psi|\Sigma,\theta,Y) \\ \propto \quad & p(Y|\Psi,\Sigma,\theta)p(\Psi|\Sigma,\theta) \qquad\qquad (A.4) \\ \propto \quad & \text{etr}\left[\Sigma^{-1}\lambda T\Big(\Gamma_{YY} - 2B_2'\Psi'\Gamma_{X\tilde{Y}} + B_2'\Psi'\Gamma_{XX}(\theta)\Psi B_2\Big) + (\tilde{Y} - X\Psi B_2)'(\tilde{Y} - X\Psi B_2)\right] \\ \propto \quad & \text{etr}\left[\Sigma^{-1}\Big(-2B_2'\Psi'(\lambda T\Gamma_{X\tilde{Y}} + X'\tilde{Y}) + B_2'\Psi'(\lambda T\Gamma_{XX} + X'X)\Psi B_2\Big)\right] \end{aligned}$$

Define

$$\tilde{\Psi}(\Sigma,\theta) = (\lambda T\Gamma_{XX} + X'X)^{-1}(\lambda T\Gamma_{X\tilde{Y}} + X'\tilde{Y})\Sigma^{-1}B_2'(B_2\Sigma^{-1}B_2')^{-1}.$$

The previous calculations show that

$$\Psi|\Sigma,\theta,Y \sim \mathcal{N}\left(\tilde{\Psi}(\Sigma,\theta), \Big[(B_2\Sigma^{-1}B_2') \otimes (\lambda T\Gamma_{XX} + X'X)\Big]^{-1}\right). \tag{A.5}$$

**Conditional Posterior of $\Sigma$:** Combining the the prior densities (A.2) and (A.3) with the likelihood function (A.1) yields

$$
\begin{aligned}
p(\Sigma|\Psi,\theta,Y) &\propto p(Y|\Psi,\Sigma,\theta)p(\Psi|\Sigma,\theta)p(\Sigma|\theta) \qquad (A.6)\\
&\propto |\Sigma|^{-\frac{1}{2}((\lambda+1)T-k+n+1)}|(B_2\Sigma^{-1}B_2')^{-1}|^{-\frac{k}{2}}\\
&\quad \mathrm{etr}\Big[\Sigma^{-1}\Big(\lambda T(\Gamma_{\tilde{Y}\tilde{Y}}-\Gamma_{\tilde{Y}X}\Gamma_{XX}^{-1}\Gamma_{X\tilde{Y}})+(\tilde{Y}-X\Psi B_2)'(\tilde{Y}-X\Psi B_2)\Big)\\
&\quad +\lambda T(B_2\Sigma^{-1}B_2')(\Psi-\bar{\Psi})'\Gamma_{XX}(\Psi-\bar{\Psi})\Big].
\end{aligned}
$$

Using the definition of $\bar{\Psi}$, the last term can be manipulated as follows

$$
\begin{aligned}
&\mathrm{etr}\Big[\lambda TB_2\Sigma^{-1}B_2'(\Psi-\bar{\Psi})'\Gamma_{XX}(\Psi-\bar{\Psi})\Big]\\
&= \mathrm{etr}\Big[\lambda T\Sigma^{-1}\Big(B_2'\Psi'\Gamma_{XX}\Psi B_2-2B_2'\Psi'\Gamma_{X\tilde{Y}}\Big)\\
&\quad +\lambda T\Sigma^{-1}B_2'(B_2\Sigma^{-1}B_2')^{-1}B_2\Sigma^{-1}\Gamma_{X\tilde{Y}}'\Gamma_{XX}^{-1}\Gamma_{X\tilde{Y}}\Big]
\end{aligned}
$$

Hence,

$$
\begin{aligned}
p(\Sigma|\Psi,\theta,Y) &\propto |\Sigma|^{-\frac{1}{2}((\lambda+1)T-k+n+1)}|(B_2\Sigma^{-1}B_2')^{-1}|^{-\frac{k}{2}} \qquad (A.7)\\
&\quad \times\mathrm{etr}\Big[\Sigma^{-1}\Big(\lambda T\Gamma_{\tilde{Y}\tilde{Y}}+\tilde{Y}'\tilde{Y}-2B_2'\Psi'(\lambda T\Gamma_{X\tilde{Y}}+X'\tilde{Y})\\
&\quad +B_2'\Psi'(\lambda T\Gamma_{XX}+X'X)\Psi B_2\Big)\Big]\\
&\quad \times\mathrm{etr}\Big[\lambda T(\Sigma^{-1}B_2'(B_2\Sigma^{-1}B_2')^{-1}B_2\Sigma^{-1}-\Sigma^{-1})\Gamma_{X\tilde{Y}}'\Gamma_{XX}^{-1}\Gamma_{X\tilde{Y}}\Big].
\end{aligned}
$$

If the DSGE model satisfies Equation (20) and the error $u_{1,t}$ is orthogonal to $x_t$ then

$$\Gamma_{X\tilde{Y}}=\Gamma_{XX}\Psi_0(\theta)B_2$$

and

$$(\Sigma^{-1}B_2'(B_2\Sigma^{-1}B_2')^{-1}B_2\Sigma^{-1}-\Sigma^{-1})\Gamma_{X\tilde{Y}}'\Gamma_{XX}^{-1}\Gamma_{X\tilde{Y}}=0. \qquad (A.8)$$

While the conditional posterior distribution of $\Sigma$ given our prior distribution is not of the $\mathcal{IW}$ form use an $\mathcal{IW}$ distribution as proposal distribution in a Metropolis-Hastings step. Define

$$
\begin{aligned}
\tilde{S}(\Psi,\theta) &= \lambda T\Gamma_{\tilde{Y}\tilde{Y}}+\tilde{Y}'\tilde{Y}-(\lambda T\Gamma_{X\tilde{Y}}+X'\tilde{Y})'\Psi B_2-B_2'\Psi'(\lambda T\Gamma_{X\tilde{Y}}+X'\tilde{Y}) \qquad (A.9)\\
&\quad +B_2'\Psi'(\lambda T\Gamma_{XX}+X'X)\Psi B_2
\end{aligned}
$$

Our proposal distribution for $\Sigma$ is

$$\mathcal{IW}(\tilde{S}(\Psi,\theta),(\lambda+1)T,n).$$

**Conditional Posterior of $\theta$:** The posterior distribution of $\theta$ is irregular. Its density is proportional to the joint density of $Y$, $\Psi$, $\Sigma$, and $\theta$, which we can evaluate numerically since the normalization constants for $p(\Psi|\Sigma,\theta)$ and $p(\Sigma|\theta)$ are readily available.

$$p(\theta|\Psi,\Sigma,Y) \propto p(Y,\Psi,\Sigma,\theta) = p(Y|\Psi,\Sigma,\theta)p(\Psi|\Sigma,\theta)p(\Sigma|\theta)p(\theta). \tag{A.10}$$

To obtain a proposal density for $p(\theta|\Psi,\Sigma,Y)$ we (i) maximize the posterior density of the DSGE model with respect to $\theta$ and (ii) calculate the inverse Hessian at the mode, denoted by $V_{\tilde{\theta},DSGE}$. (iii) We then use a random-walk Metropolis step with proposal density

$$\mathcal{N}(\theta_{(s-1)}, cV_{\tilde{\theta},DSGE})$$

where $\theta_{(s-1)}$ is the value of $\theta$ drawn in iteration $s-1$ of the MCMC algorithm, and $c$ is a scaling factor that can be used to control the rejection rate in the Metropolis step.

Table 1: PRIOR DISTRIBUTION

| Name | Domain | Density | P(1) | P(2) |
|---|---|---|---|---|
| $\psi_1$ | $I\!R^+$ | Gamma | 1.500 | 0.400 |
| $\psi_2$ | $I\!R^+$ | Gamma | 0.200 | 0.100 |
| $\rho_R$ | $[0,1)$ | Beta | 0.500 | 0.200 |
| $r_a^*$ | $I\!R^+$ | Gamma | 1.000 | 0.400 |
| $\pi_a^*$ | $I\!R$ | Normal | 3.000 | 2.000 |
| $g^*$ | $I\!R^+$ | Beta | 0.150 | 0.050 |
| $h$ | $[0,1)$ | Beta | 0.700 | 0.100 |
| $\nu_l$ | $I\!R^+$ | Gamma | 2.000 | 0.750 |
| $\zeta_p$ | $[0,1)$ | Beta | 0.750 | 0.100 |
| $s'$ | $I\!R^+$ | Gamma | 4.000 | 1.500 |
| $a''$ | $I\!R^+$ | Gamma | 0.200 | 0.075 |
| $\rho_z$ | $[0,1)$ | Beta | 0.800 | 0.050 |
| $\rho_g$ | $[0,1)$ | Beta | 0.800 | 0.050 |
| $\sigma_z$ | $I\!R^+$ | InvGamma | 0.400 | 2.000 |
| $\sigma_g$ | $I\!R^+$ | InvGamma | 0.300 | 2.000 |
| $\sigma_R$ | $I\!R^+$ | InvGamma | 0.200 | 2.000 |

*Notes:* P(1) and P(2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; $s$ and $\nu$ for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. See Section 2 for a definition of the DSGE model's parameters. We are reporting annualized values for $\pi^*$ and $r^*$ ($a$-subscript). The following parameters were fixed: $\alpha = 0.25$, $\delta = 0.025$, $\gamma = 1.5$, $\lambda_f = 0.3$. The effective prior is truncated at the boundary of the determinacy region.

Table 2: Log Marginal Data Densities and Posterior Odds

| Specification | $\ln p(Y)$ | Post Odds |
|---|---|---|
| DSGE Model | -305.44 | 5E-13 |
| DSGE-VAR, $\lambda = 10.0$ | -297.16 | 2E-09 |
| DSGE-VAR, $\lambda = 5.0$ | -291.28 | 7E-07 |
| DSGE-VAR, $\lambda = 2.0$ | -283.70 | 0.001 |
| DSGE-VAR, $\lambda = 1.5$ | -281.23 | 0.017 |
| DSGE-VAR, $\lambda = 1.0$ | -278.46 | 0.278 |
| DSGE-VAR, $\lambda = 0.75$ | -277.22 | 0.968 |
| DSGE-VAR, $\lambda = 0.50$ | -277.18 | 1.000 |
| DSGE-VAR, $\lambda = 0.25$ | -283.34 | 0.002 |

*Notes:* The marginal data densities are obtained by integrating the likelihood function with respect to the model parameters, weighted by the prior density conditional on $\lambda$. The difference of log marginal data densities can be interpreted as log posterior odds under the assumption of that the two specifications have equal prior probabilities.

Table 3: Parameter Estimation Results

| | Prior | | Posterior | | | |
|---|---|---|---|---|---|---|
| | | | DSGE-VAR ($\lambda = 0.5$) | | DSGE (State Sp) | |
| | Mean | 90% Interval | Mean | 90% Interval | Mean | 90% Interval |
| $\psi_1$ | 1.56 | [ 1.00 , 2.07 ] | 2.19 | [ 1.68 , 2.71 ] | 1.90 | [ 1.43 , 2.37 ] |
| $\psi_2$ | 0.20 | [ 0.05 , 0.35 ] | 0.15 | [ 0.07 , 0.23 ] | 0.18 | [ 0.11 , 0.26 ] |
| $\rho_R$ | 0.50 | [ 0.18 , 0.83 ] | 0.83 | [ 0.79 , 0.88 ] | 0.80 | [ 0.75 , 0.85 ] |
| $r_a^*$ | 1.00 | [ 0.37 , 1.61 ] | 0.86 | [ 0.36 , 1.34 ] | 0.88 | [ 0.47 , 1.27 ] |
| $\pi_a^*$ | 3.00 | [ -0.31 , 6.26 ] | 3.10 | [ 2.46 , 3.71 ] | 3.02 | [ 2.44 , 3.59 ] |
| $g^*$ | 0.15 | [ 0.07 , 0.23 ] | 0.15 | [ 0.07 , 0.23 ] | 0.13 | [ 0.06 , 0.20 ] |
| $h$ | 0.70 | [ 0.54 , 0.86 ] | 0.73 | [ 0.59 , 0.89 ] | 0.92 | [ 0.87 , 0.96 ] |
| $\nu_l$ | 2.00 | [ 0.78 , 3.12 ] | 1.55 | [ 0.72 , 2.35 ] | 1.19 | [ 0.76 , 1.60 ] |
| $\zeta_p$ | 0.75 | [ 0.59 , 0.92 ] | 0.66 | [ 0.54 , 0.78 ] | 0.59 | [ 0.46 , 0.71 ] |
| $s''$ | 4.00 | [ 1.61 , 6.30 ] | 4.39 | [ 1.95 , 6.77 ] | 8.15 | [ 5.09 , 11.15 ] |
| $a'$ | 0.20 | [ 0.08 , 0.32 ] | 0.22 | [ 0.10 , 0.34 ] | 0.23 | [ 0.11 , 0.35 ] |
| $\rho_z$ | 0.80 | [ 0.72 , 0.88 ] | 0.83 | [ 0.76 , 0.90 ] | 0.83 | [ 0.78 , 0.88 ] |
| $\rho_g$ | 0.80 | [ 0.72 , 0.88 ] | 0.83 | [ 0.75 , 0.90 ] | 0.89 | [ 0.86 , 0.93 ] |
| $\sigma_z$ | 0.71 | [ 0.16 , 1.24 ] | 0.83 | [ 0.47 , 1.18 ] | 0.85 | [ 0.59 , 1.09 ] |
| $\sigma_g$ | 0.53 | [ 0.12 , 0.93 ] | 0.37 | [ 0.27 , 0.47 ] | 0.66 | [ 0.56 , 0.76 ] |
| $\sigma_R$ | 0.36 | [ 0.08 , 0.63 ] | 0.13 | [ 0.10 , 0.16 ] | 0.16 | [ 0.13 , 0.18 ] |

Table 4: Comparison of Preferred Policy Rules

| Best in Scenario | Parameters $\psi_1$ | $\psi_2$ | $\rho_R$ | Loss in Scenario 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|
| 1 | 3.00 | 0.19 | 0.80 | -0.77 | -0.86 | 18.5 | 0.62 |
| 2 | 3.00 | 0.00 | 0.80 | -0.64 | -1.17 | 4.83 | -0.72 |
| 3 | 3.00 | 0.19 | 0.95 | 0.05 | -0.40 | -17.57 | 4.79 |
| 4 | 1.75 | 0.06 | 0.80 | 0.37 | -0.34 | -13.71 | -1.46 |

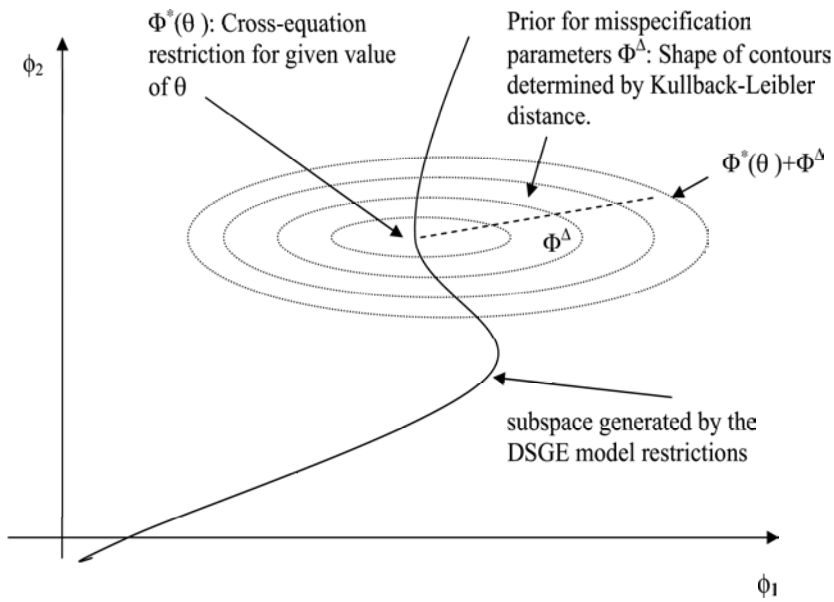
*Notes:* Columns 5 to 8 contain mean policy loss differentials relative to baseline policy rule $\psi_1 = 2.0$, $\psi_2 = 0.188$, $\rho_R = 0.80$. The baseline losses for the four scenarios are 2.46, 2.54, 20.13, and 4.77 respectively.

Table 5: Loss Differentials for Alternative Policy Rules

| Rule | Parameters $\psi_1$ | $\psi_2$ | $\rho_R$ | Loss in Scenario 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|
| 1 – Taylor I (modified) | 60.00 | 4.00 | 0.95 | -1.36 | -1.31 | 29.87 | 0.35 |
| 2 – Taylor II (modified) | 24.00 | 5.00 | 0.95 | -0.81 | -0.61 | 29.86 | 6.85 |
| 3 – Taylor III | 1.50 | 0.125 | 0.00 | 5.21 | 4.78 | 29.81 | 2.51 |
| 4 – Taylor IV | 1.50 | 0.25 | 0.00 | 6.00 | 7.50 | 29.86 | 5.10 |
| 5 – Levin - Williams (Bayes) | 4.20 | 0.80 | 0.40 | -0.02 | 0.93 | 29.87 | 4.64 |
| 6 – Levin - Williams (Minimax) | 6.50 | 1.375 | 0.80 | -0.53 | -0.07 | 29.86 | 6.54 |

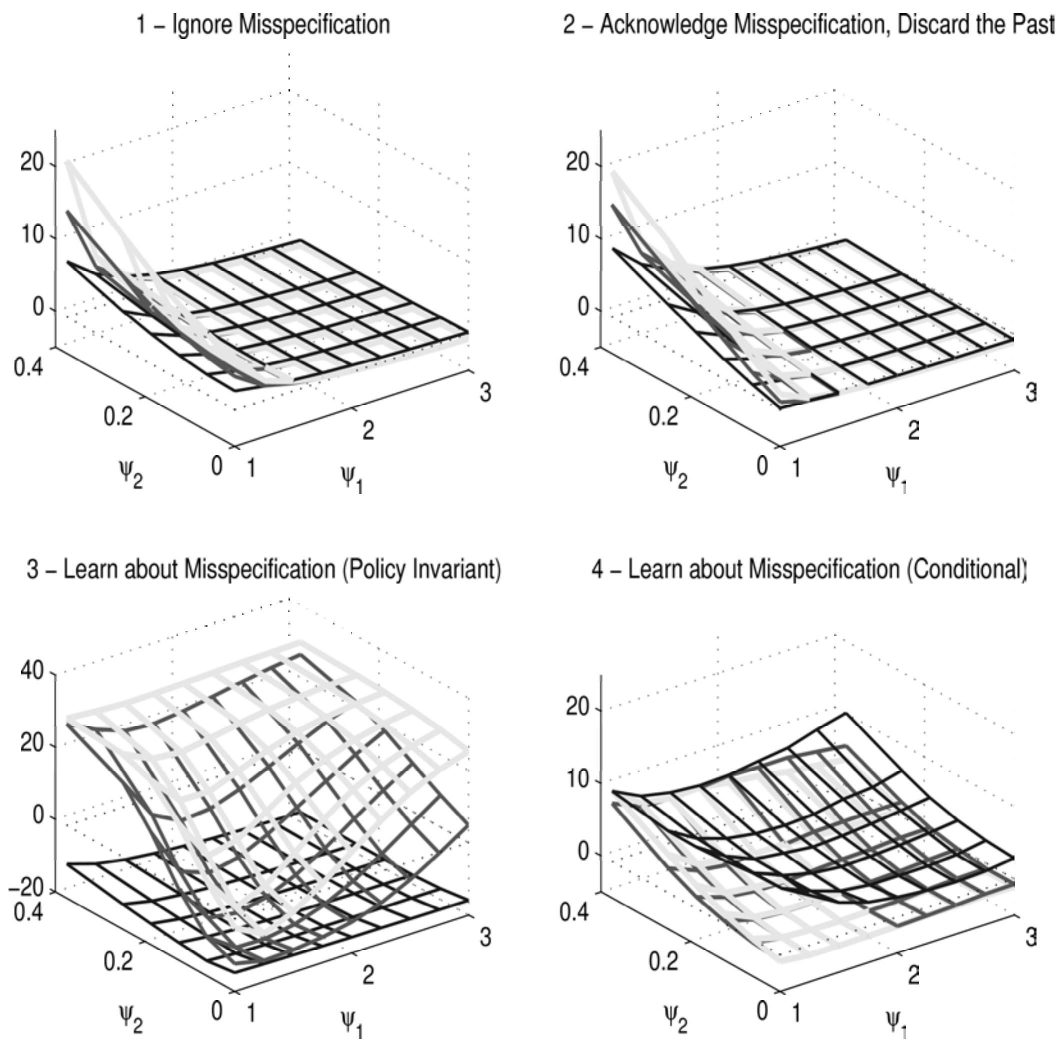
*Notes:* Columns 5 to 8 contain mean policy loss differentials relative to baseline policy rule $\psi_1 = 2.0$, $\psi_2 = 0.188$, $\rho_R = 0.80$. The baseline losses for the four scenarios are 2.46, 2.54, 20.13, and 4.77 respectively.

Figure 1: Stylized View of DSGE Model Misspecification



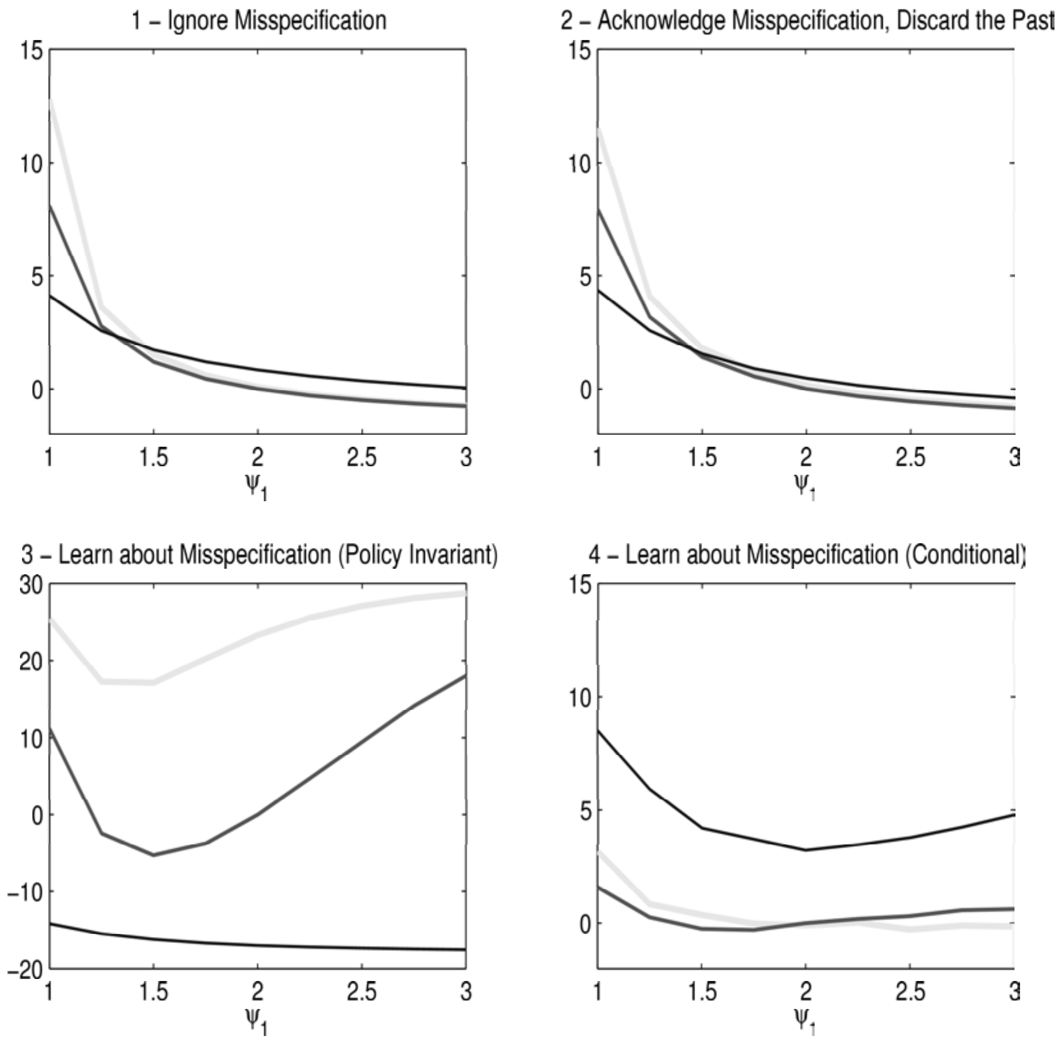

*Notes:* $\Phi = [\phi_1, \phi_2]'$ can be interpreted as the VAR parameters, and $\Phi^*(\theta)$ is the restriction function implied by the DSGE model.

Figure 2: Expected Policy Loss Differentials




*Notes:* Mean policy loss differentials as a function of $\psi_1$ and $\psi_2$ relative to baseline policy rule $\psi_1 = 2.0$, $\psi_2 = 0.188$, $\rho_R = 0.80$. The baseline losses for the four scenarios are 2.46, 2.54, 20.13, and 4.77 respectively. Negative differentials signify an improvement relative to baseline rule. Surfaces' color ranges from very light grey ($\rho = 0.7$) to dark grey ($\rho = 0.95$), with the darkness of the surface being directly proportional to $\rho$.

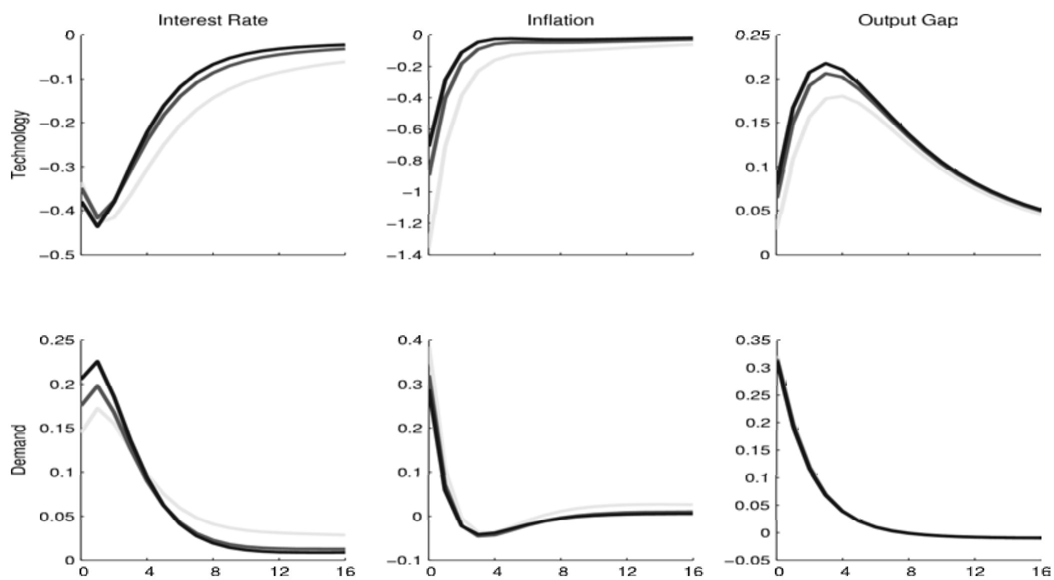
Figure 3: EXPECTED POLICY LOSS DIFFERENTIALS



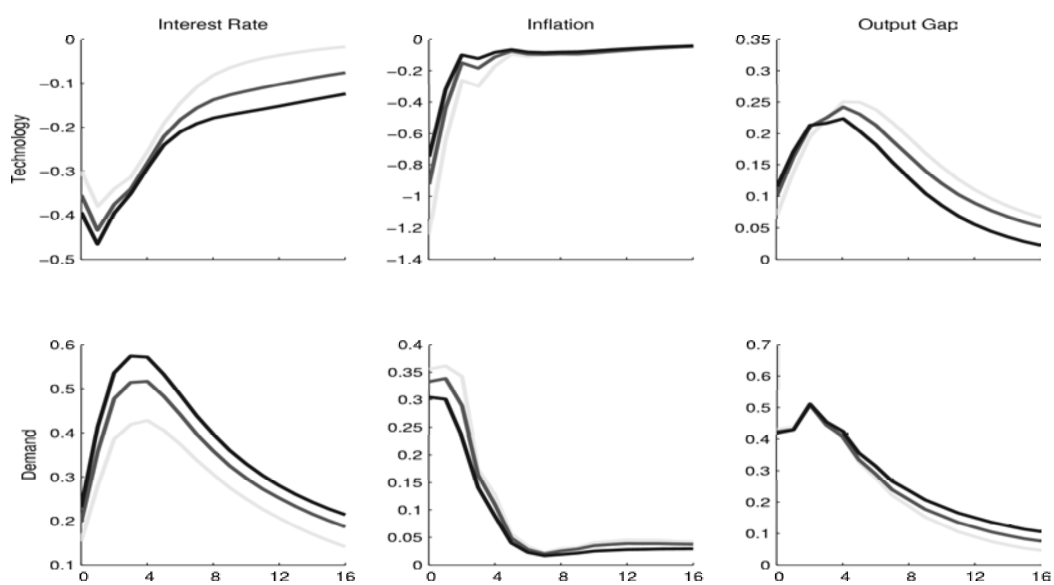

*Notes:* Mean policy loss differentials for $\psi_2 = 0.188$ as a function of $\psi_1$ and relative to baseline policy rule $\psi_1 = 2.0$, $\psi_2 = 0.188$, $\rho_R = 0.80$. Negative differentials signify an improvement relative to baseline rule. Lines' color ranges from very light grey ($\rho = 0.7$) to dark grey ($\rho = 0.95$), with the darkness of the line being directly proportional to $\rho$.

Figure 4: Impulse Responses as Function of $\psi_1$

Scenario 1 – Ignore Misspecification



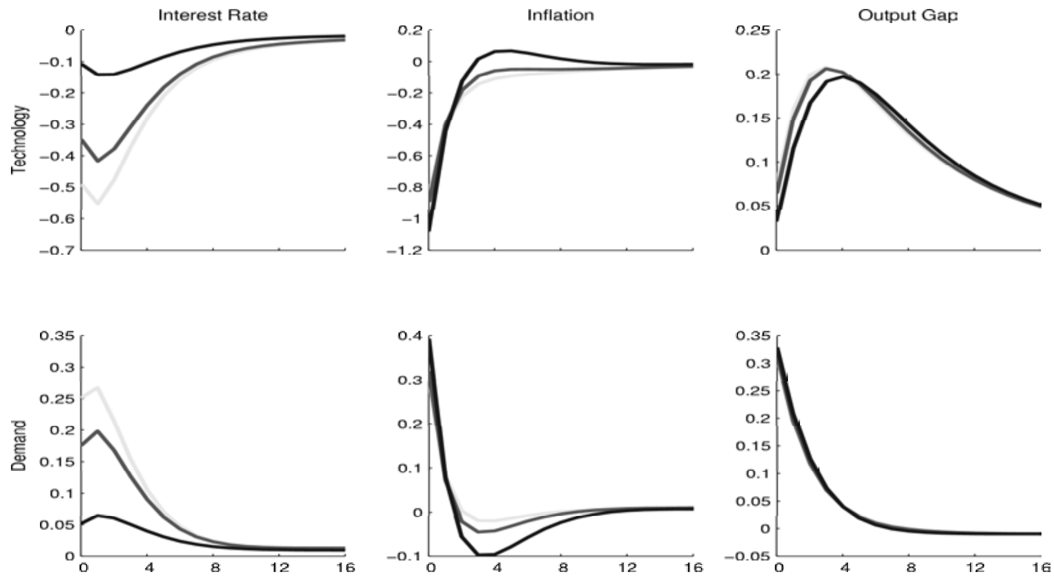

Scenario 4 – Learn about Misspecification (Conditional)



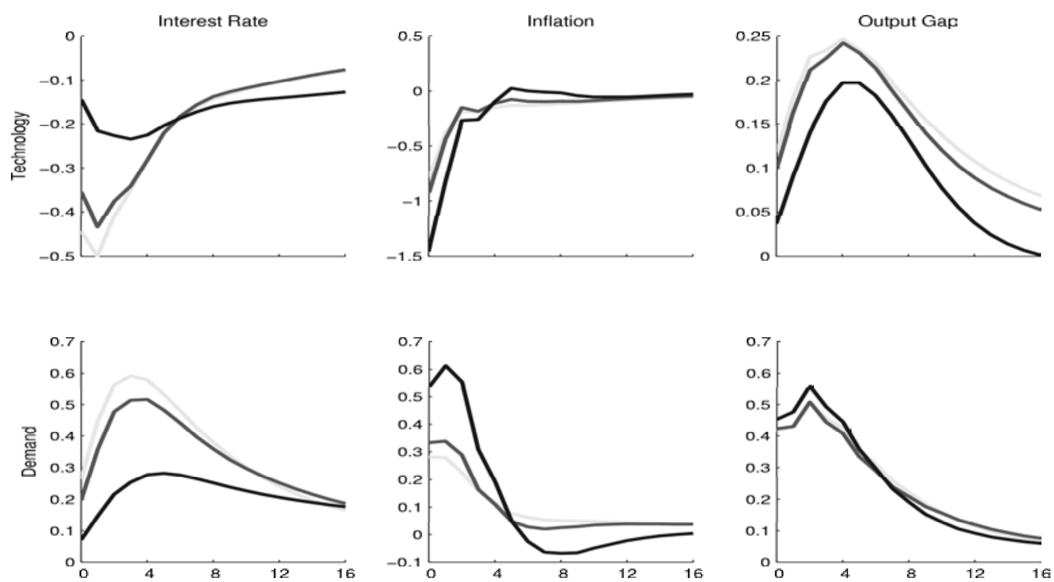

*Notes:* Lines' color ranges from very light grey ($\psi_1 = 1.25$) to dark grey ($\psi_1 = 2.75$), with the darkness of the line being directly proportional to $\psi_1$.

Figure 5: Impulse Responses as Function of $\rho_R$

Scenario 1 – Ignore Misspecification

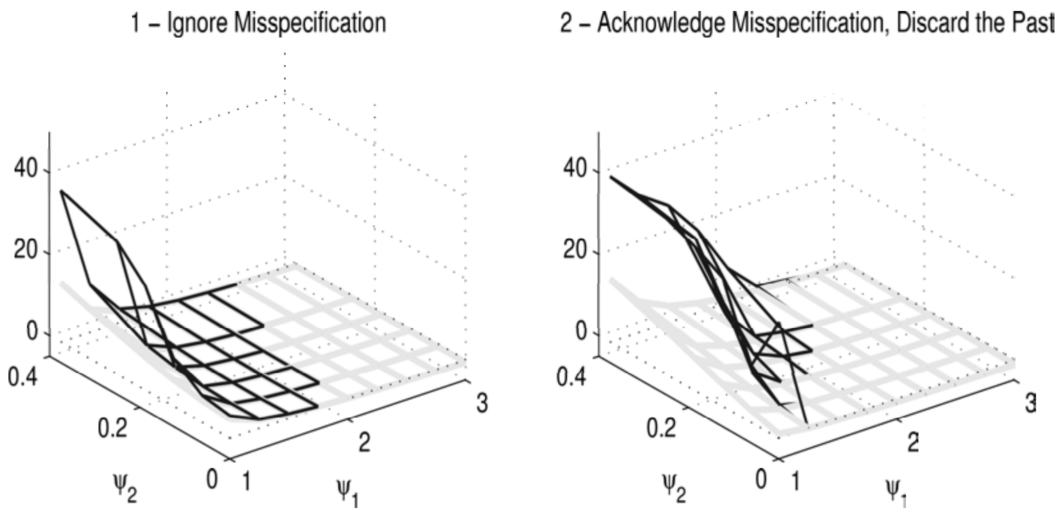



Scenario 4 – Learn about Misspecification (Conditional)




*Notes:* Lines' color ranges from very light grey ($\rho_R = .7$) to dark grey ($\rho_R = .95$), with the darkness of the line being directly proportional to $\rho_R$.

Figure 6: Expected Policy Loss Differentials - Risk-Neutral versus Risk-Sensitive




*Notes:* Policy loss differentials relative to baseline policy rule $\psi_1 = 2.0$, $\psi_2 = 1.88$, $\rho_R = 0.80$. All numbers are computed fixing the value of $\rho_R = 0.8$. The baseline losses for the four scenarios are 2.46, 2.54, 20.13, and 4.77 respectively. Negative differentials signify an improvement relative to baseline rule. For each scenario, the expected loss differential is shown in light grey, and the risk-sensitive loss differential in black.