Monetary Policy Analysis with Potentially Misspecified Models

Marco Del Negro
Federal Reserve Bank of Atlanta

Frank Schorfheide*
University of Pennsylvania

This Draft: March, 2004

Abstract

This 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). Specifically, we are studying the effects of coefficient changes in interest-rate feedback rules on the volatility of output growth, inflation, and nominal rates. The paper illustrates the sensitivity of the results to assumptions on the policy invariance of model misspecifications.

JEL CLASSIFICATION: C32

KEY WORDS: Bayesian Analysis, DSGE Models, Model Misspecification

---

*Correspondence: M. Del Negro: Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St NE, Atlanta GA 30309-4470. Email: Marco.DelNegro@atl.frb.org. F. 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, 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, and seminar participants at 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

Despite recent successes in improving the empirical performance of dynamic stochastic general equilibrium (DSGE) models, e.g., Smets and Wouters (2003), even large-scale DSGE models suffer to some extent from misspecification. In this paper misspecification means that the DSGE model potentially imposes invalid cross-coefficient restrictions on the moving-average representation of the macroeconomic time series that it aims to explain. As a consequence, one typically observes that the forecasting performance of DSGE models is worse than that of vector autoregressions (VARs) estimated with well-calibrated shrinkage methods. On the other hand, 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. Thus, policy analysis with DSGE models is robust to the Lucas critique and potentially more reliable than conclusions drawn from VARs. This trade-off poses a challenge to policymakers who want to use DSGE models in practice.

Del Negro and Schorfheide (2004) proposed a framework that combines VARs and DSGE models, extending earlier work by Ingram and Whiteman (1994). In this framework DSGE model restrictions are neither completely ignored as in the unrestricted estimation of VARs, nor are they dogmatically imposed as in the direct estimation of DSGE models. Instead the VAR estimates are tilted toward the restrictions implied by the DSGE model, where the degree of tilting is determined by a Bayesian data-driven procedure that trades off model fit against complexity. Starting from the same DSGE model that is used in this paper, Del Negro, Schorfheide, Smets, and Wouters (2004) show that relaxing the DSGE model restrictions leads to a substantial improvement of in-sample-fit (adjusted for model complexity) and more accurate pseudo-out-of-sample predictions.

This paper extends our earlier work and further develops procedures that are suitable to study the effects of rare regime shifts with potentially misspecified DSGE models. Monetary policy is modelled through an interest-rate feedback rule. While our earlier work relaxes all DSGE model restrictions simultaneously, this paper assumes that the monetary policy rule in the DSGE model is correctly specified and strictly imposes the associated restrictions. We consider the following policy experiment. Between time $t = T$ and $t = T + 1$ the monetary policy maker seeks to replace an existing interest-rate feedback rule with one that minimizes her loss function. We make the simplifying assumption that the public believes the new policy to be in place indefinitely after being announced credibly. The loss function is defined in terms of expected squared deviations of output growth, inflation, and interest rates from their respective target levels. In time $t = 0$ the policy maker places a prior distribution on the parameters that characterize the misspecification of the DSGE model restrictions and subsequently updates this prior based on time T information. We then consider a variety of assumptions on the policy invariance of the misspecification parameters and calculate posterior expected losses as a function of the policy parameters.

A key feature of our analysis is that we treat model misspecification as non-structured. Once the restrictions derived from the DSGE model are relaxed there is no formal interpretation of the resulting specification in terms of a fully specified general equilibrium model. We view this analysis as a complement to a structured misspecification analysis in which model misspecification is phrased in terms of omitted or wrongly specified structural components, such as, for instance, omitted capital adjustment costs or the use of Calvo-style nominal rigidities instead of state-dependent pricing rules. The DSGE model analyzed in this paper is based on work by Christiano, Eichenbaum, and Evans (2005), Altig, Christiano, Eichenbaum, and Linde (2002), and Smets and Wouters (2003). 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. Nevertheless, as documented in Del Negro, Schorfheide, Smets, and Wouters (2004) misspecification is still a concern and we believe that this concern should be reflected in policy recommendations derived from this model.[1]

The procedures developed in this paper can be viewed as a Bayesian alternative to robust control and minimax approaches that recently have been proposed to cope with model misspecification, e.g., Hansen and Sargent (2005), Giannoni (2002), Levin, Wieland, and Williams (1999), and Onatski and Stock (2002). 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 to harm the policy maker. The disadvantage of this approach is that the resulting policy performs well in the worst-case but possibly poorly on average. Moreover, in most formulations of the minimax problem, the policy maker does not use historical data to learn about the extent of the model misspecification. To compare our analysis to a risk-sensitive approach, 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

[1] In Del Negro and Schorfheide (2005) we applied the approach developed in this paper to a simple three-equation New Keynesian model without capital accumulation and wage rigidities.

a penalty that is a function of the Kullback-Leibler distance between the distorted and the non-distorted probabilities.

The empirical analysis is based on quarterly U.S. output, inflation, and interest data from 1983 to 2004. We estimate the state-space representation of the log-linearized DSGE model and its vector autoregressive approximation. To obtain a measure of the extent of its misspecification, parameters that capture the discrepancy between the restricted and unrestricted VAR approximation are introduced. The estimation is implemented with a Markov Chain Monte Carlo algorithm that allows us to generate draws from the joint posterior distribution of the DSGE model and the misspecification parameters. We then compute expected policy loss differentials relative to the estimated interest rate feedback rule. Four scenarios are considered that differ with respect to assumptions on the effect of policy changes on the beliefs about model misspecification. While the particular values of the loss differentials are sensitive to the misspecification assumptions considered, a fairly robust policy recommendation emerges from our analysis: the central bank should avoid strong responses to output growth movements and not react weakly to inflation fluctuations.

The paper is organized as follows. The DSGE model is presented in Section 2. Section 3 discusses the estimation of potentially misspecified DSGE models. The framework for policy analysis is introduced in Section 4. Section 5 describes the data set, Section 6 discusses our empirical findings, and Section 7 concludes. The posterior simulator that is used to implement the empirical analysis is described in the Appendix.

2 Model

This section describes the DSGE model, which is a slightly modified version of the DSGE model developed and estimated for the Euro area in Smets and Wouters (2003). In particular, we introduce stochastic trends into the model, so that it can be fitted to unfiltered time series observations. The DSGE model, largely based on the work of Christiano, Eichenbaum, and Evans (2005), contains a large number of nominal and real frictions. Next, we describe each of the agents that populate the model economy and the decision problems they face.

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 $i \in [0,1]$:

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

$\lambda_{f,t} \in (0, \infty)$ follows the exogenous process:

$$\ln \lambda_{f,t} = \ln \lambda_f + \sigma_{\lambda,f} \epsilon_{\lambda,t}, \tag{2}$$

where $\epsilon_{\lambda,t}$ is an exogenous shock with unit variance. 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:

$$Y_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-\frac{1+\lambda_{f,t}}{\lambda_{f,t}}} Y_t \quad \text{and} \quad P_t = \left[\int_0^1 P_t(i)^{\frac{1}{\lambda_{f,t}}} di \right]^{\lambda_{f,t}}. \tag{3}$$

2.2 Intermediate goods producers

Good i is made using the technology:

$$Y_t(i) = \max \left\{ Z_t^{1-\alpha} K_t(i)^\alpha L_t(i)^{1-\alpha} - Z_t \mathcal{F}, 0 \right\}, \tag{4}$$

where the technology shock Z_t (common across all firms) follows a unit root process, and where $\mathcal{F}$ represent fixed costs faced by the firm. We define technology growth $z_t = \log(Z_t / Z_{t-1})$ and assume that z_t follows the autoregressive process:

$$(z_t - \gamma) = \rho_z (z_{t-1} - \gamma) + \sigma_z \epsilon_{z,t}. \tag{5}$$

All firms face the same prices for their inputs, labor and capital. Hence profit's maximization implies that the capital/labor ratio is the same for all firms, and equal to:

$$\frac{K_t}{L_t} = \frac{\alpha}{1-\alpha} \frac{W_t}{R_t^k}., \tag{6}$$

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 ζ_p is unable to re-optimize their prices $P_t(i)$. These firms adjust their prices mechanically according to

$$P_t(i) = (\pi_{t-1})^{\iota_p} (\pi^*)^{1-\iota_p}, \tag{7}$$

where $\pi_t = P_t/P_{t-1}$ and π^* is the steady state inflation rate of the final good. In our empirical analysis we will restrict ι_p to be either zero or one. Those firms that are able to re-optimize prices choose the price level $\tilde{P}_t(i)$ that solves:

$$\max_{\tilde{P}_t(I)} I\!\!E_t \sum_{s=0}^{\infty} \zeta_p^s \beta^s \Xi^p_{t+s} \left(\tilde{P}_t(i) \left(\Pi_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p} \right) - MC_{t+s} \right) Y_{t+s}(i)$$
$$\text{s.t. } Y_{t+s}(i) = \left(\frac{\tilde{P}_t(i) \left(\Pi_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p} \right)}{P_{t+s}} \right)^{-\frac{1+\lambda_{f,t}}{\lambda_{f,t}}} Y_{t+s}, \; MC_{t+s} = \frac{\alpha^{-\alpha} W_{t+s}^{1-\alpha} R_{t+s}^{k\;\alpha}}{(1-\alpha)^{(1-\alpha)} Z_{t+s}^{1-\alpha}}. \tag{8}$$

where $\beta^s \Xi^p_{t+s}$ 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 (3) we obtain the following law of motion for the aggregate price level:

$$P_t = [(1-\zeta_p)\tilde{P}_t^{\frac{1}{\lambda_{f,t}}} + \zeta_p(\pi_{t-1}^{\iota_p}\pi_*^{1-\iota_p}P_{t-1})^{\frac{1}{\lambda_{f,t}}}]^{\lambda_{f,t}}. \tag{9}$$

2.3 Labor packers

There is a continuum of households, indexed by $j \in [0,1]$, each supplying a differentiated form of labor, $L(j)$. The "labor packers" are perfectly competitive firms that hire labor from the households and combine it to labor services L_t that are offered to the intermediate goods producers:

$$L_t = \left[\int_0^1 L_t(j)^{\frac{1}{1+\lambda_w}} di \right]^{1+\lambda_w}, \tag{10}$$

where $\lambda_w \in (0, \infty)$. From first-order and zero-profit conditions of the labor packers we obtain the labor demand function and an expression for the price of aggregated labor services L_t:

$$L_t(j) = \left(\frac{W_t(j)}{W_t} \right)^{-\frac{1+\lambda_w}{\lambda_w}} L_t \quad \text{and} \quad W_t = \left[\int_0^1 W_t(j)^{\frac{1}{\lambda_w}} di \right]^{\lambda_w}. \tag{11}$$

2.4 Households

The objective function for household j is given by:

$$I\!\!E_t \sum_{s=0}^{\infty} \beta^s b_{t+s} \left[\log(C_{t+s}(j) - hC_{t+s-1}(j)) - \frac{\varphi_{t+s}}{1+\nu_l} L_{t+s}(j)^{1+\nu_l} + \frac{\chi}{1-\nu_m} \left(\frac{M_{t+s}(j)}{Z_{t+s}P_{t+s}} \right)^{1-\nu_m} \right] \tag{12}$$

where $C_t(j)$ is consumption, $L_t(j)$ is labor supply, and $M_t(j)$ are money holdings. Household's preferences display habit-persistence. We depart from Smets and Wouters (2003) in assuming separability in the utility function for a reason that will be discussed later. The preference shifters φ_t, which affects the marginal utility of leisure, and b_t, which scales the overall period utility, are exogenous processes common to all households that evolve as:

$$\ln \varphi_t = (1-\rho_\varphi)\ln\varphi + \rho_\varphi \ln\varphi_{t-1} + \sigma_\varphi \epsilon_{\varphi,t}, \tag{13}$$

$$\ln b_t = \rho_b \ln b_{t-1} + \sigma_b \epsilon_{b,t}. \tag{14}$$

Real money balances enter the utility function deflated by the (stochastic) trend growth of the economy, so to make real money demand stationary.

The household's budget constraint written in nominal terms is given by:

$$\begin{aligned} &P_{t+s}C_{t+s}(j) + P_{t+s}I_{t+s}(j) + B_{t+s}(j) + M_{t+s}(j) \leq R_{t+s}B_{t+s-1}(j) + M_{t+s-1}(j) \\ &+ \Pi_{t+s} + W_{t+s}(j)L_{t+s}(j) + \left(R^k_{t+s}u_{t+s}(j)\bar{K}_{t+s-1}(j) - P_{t+s}a(u_{t+s}(j))\bar{K}_{t+s-1}(j)\right), \end{aligned} \tag{15}$$

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, Π_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:

$$K_t(j) = u_t(j)\bar{K}_{t-1}(j), \tag{16}$$

and receive $R^k_t u_t(j)\bar{K}_{t-1}(j)$ 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:

$$\bar{K}_t(j) = (1-\delta)\bar{K}_{t-1}(j) + \mu_t \left(1 - S\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\right) I_t(j), \tag{17}$$

where δ is the rate of depreciation, and $S(\cdot)$ is the cost of adjusting investment, with $S'(\cdot) > 0, S''(\cdot) > 0$. The term μ_t is a stochastic disturbance to the price of investment relative to consumption, which follows the exogenous process:

$$\ln\mu_t = (1-\rho_\mu)\ln\mu + \rho_\mu \ln\mu_{t-1} + \sigma_\mu \epsilon_{\mu,t}. \tag{18}$$

The households' wage setting is subject to nominal rigidities á la Calvo (1983). In each period a fraction ζ_w of households is unable to re-adjust wages. For these households, the wage $W_t(j)$ will increase at a geometrically weighted average of the steady state rate increase in wages (equal to steady state inflation π_* times the growth rate of the economy e^γ) and of last period's inflation times last period's productivity ($\pi_{t-1}e^{z_{t-1}}$). The weights are $1-\iota_w$ and ι_w, respectively. Those households that are able to re-optimize their wage solve the problem:

$$\begin{array}{ll} \max_{\tilde{W}_t(j)} & I\!\!E_t \sum_{s=0}^{\infty}(\zeta_w\beta)^s b_{t+s}\left[-\frac{\varphi_{t+s}}{\nu_l+1}L_{t+s}(j)^{\nu_l+1}\right] \\ \text{s.t.} & \text{(15) for } s=0,\ldots,\infty, \text{ (11a), and} \\ & W_{t+s}(j) = \left(\Pi_{l=1}^{s}(\pi_* e^{\gamma})^{1-\iota_w}(\pi_{t+l-1}e^{z^*_{t+l-1}})^{\iota_w}\right)\tilde{W}_t(j). \end{array} \tag{19}$$

We again consider only the symmetric equilibrium in which all agents solving (19) will choose the same $\tilde{W}_t(j)$. From (11b) it follows that:

$$W_t = [(1-\zeta_w)\tilde{W}_t^{\frac{1}{\lambda_w}} + \zeta_w((\pi_* e^{\gamma})^{1-\iota_w}(\pi_{t-1}e^{z^*_{t-1}})^{\iota_w}W_{t-1})^{\frac{1}{\lambda_w}}]^{\lambda_w}. \tag{20}$$

Finally, we assume there is a complete set of state contingent securities in nominal terms, which implies that the Lagrange multiplier $\Xi_t^p(j)$ associated with (15) 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 choice of consumption, money demand, investment and capital utilization. Since the amount of leisure will differ across households due to the wage rigidity, separability between labor and consumption in the utility function is key for this result.

2.5 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:

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R}\left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1}\left(\frac{Y_t}{e^{\gamma}Y_{t-1}}\right)^{\psi_2}\right]^{1-\rho_R}\sigma_R e^{\epsilon_{R,t}}, \tag{21}$$

where R^* is the steady state nominal rate, Y_t^s is the target level of output, and the parameter ρ_R determines the degree of interest rate smoothing. In our formulation of the policy rule, the central bank responds to output growth rather than some measure of the output gap.

The government budget constraint is of the form

$$P_t G_t + R_{t-1}B_{t-1} + M_{t-1} = T_t + M_t + B_t, \tag{22}$$

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

$$G_t = (1-1/g_t)Y_t, \tag{23}$$

where g_t follows the process:

$$\ln g_t = (1-\rho_g)\ln g + \rho_g \ln g_{t-1} + \sigma_g \epsilon_{g,t} \tag{24}$$

2.6 Resource constraint

The aggregate resource constraint:

$$C_t + I_t + a(u_t)\bar{K}_{t-1} = \frac{1}{g_t} Y_t. \tag{25}$$

can be derived by integrating the budget constraint (15) across households, and combining it with the government budget constraint (22) and the zero profit conditions of both labor packers and final good producers.

2.7 Model solution and State-Space Representation

As in Altig, Christiano, Eichenbaum, and Linde (2002) our model economy evolves along stochastic growth path. Output Y_t, consumption C_t, investment I_t, the real wage W_t/P_t, physical capital K_t and effective capital $\bar{K}_t$ all grow at the rate Z_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 (annualized), inflation rates (annualized), and quarterly output growth rates. Hence, let $y_t = [R_t^a, \pi_t^a, \Delta \ln Y_t]'$. The relationships between the steady-state deviations $\widetilde{R}_t$, $\widetilde{\pi}_t$, $\widetilde{Y}_t$ and the observables are given by the following measurement equations:

$$\begin{aligned} y_{1,t} &= \ln r_a^* + \ln \pi_a^* + 4\widetilde{R}_t, \\ y_{2,t} &= \begin{bmatrix} \ln \pi_a^* + 4\widetilde{\pi}_t \\ \ln \gamma + \Delta \widetilde{x}_t + \widetilde{z}_t \end{bmatrix}. \end{aligned} \tag{26}$$

Here, $y_{1,t}$ denotes the policy-maker's instrument (the interest rate), and $y_{2,t}$ is a vector that includes the remaining two endogenous variables. We collect all the DSGE model parameters in the vector θ and stack the structural shocks in the vector ϵ_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 $\mathbb{E}_\theta^D[\cdot]$. We will derive an (approximate) vector autoregressive representation for the DSGE model and introduce model misspecifications as deviations from this representation.[2] 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, a prior distribution for these model misspecifications is specified and posterior inference and policy analysis are discussed.

3.1 A VAR Representation of the DSGE Model

Let us rewrite Eq. (21), which describes the policy-maker's behavior, in more general form as:

$$y_{1,t} = x_t' M_1 \beta_1(\theta) + y_{2,t}' M_2 \beta_2(\theta) + \epsilon_{1,t}, \tag{27}$$

where $y_t = [y_{1,t}, y_{2,t}']'$ and the $k \times 1$ vector $x_t = [y_{t-1}', \ldots, y_{t-p}', 1]'$ 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 output growth. The functions $\beta_1(\theta)$ and $\beta_2(\theta)$ can be easily derived from (21) and the measurement equation (26) for R_t.

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

$$y_{2,t}' = x_t' \Psi^*(\theta) + u_{2,t}'. \tag{28}$$

In general, the VAR representation (28) is not exact if the number of lags p is finite. We define $\Gamma_{XX}(\theta) = I\!\!E_\theta^D[x_t x_t']$ and $\Gamma_{XY_2}(\theta) = I\!\!E_\theta^D[x_t y_{2,t}']$ and let

$$\Psi^*(\theta) = \Gamma_{XX}^{-1}(\theta) \Gamma_{XY_2}(\theta). \tag{29}$$

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 ϵ_t is invertible, then $u_{2,t}$ can also be expressed as a function of ϵ_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).

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

The equation for the policy instrument (27) can be rewritten by replacing $y_{2,t}$ with expression (28):

$$y_{1,t} = x_t' M_1 \beta_1(\theta) + x_t' \Psi^*(\theta) M_2 \beta_2(\theta) + u_{1,t}, \tag{30}$$

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

$$\Phi^*(\theta) = B_1(\theta) + \Psi^*(\theta) B_2(\theta). \tag{31}$$

Hence, we obtain a restricted VAR for y_t

$$y_t' = x_t' \Phi + u_t', \quad I\!E[u_t u_t'] = \Sigma^*(\theta) \tag{32}$$

with

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

Here the population covariance matrices are $\Gamma_{YY}(\theta) = I\!E_\theta^D[y_t y_t']$ and $\Gamma_{XY}(\theta) = \Gamma_{YX}'(\theta) = I\!E_\theta^D[x_t y_t']$. The following Lemma will be useful for the subsequent analysis and can be verified by straightforward matrix manipulations. Let $I\!E_{\Psi,\Sigma}^{VAR}[\cdot]$ denote expectations under the probability distribution generated by (32).

Lemma 1 *(i) The VAR coefficient matrix* $\Phi^*(\theta) = \Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta)$. *(ii)* $I\!E_{\Psi^*(\theta),\Sigma^*(\theta)}^{VAR}[x_t x_t'] = I\!E_\theta^D[x_t x_t'] = \Gamma_{XX}(\theta)$.

Since the monetary policy rule (21) in the DSGE model is specified so that it can be exactly reproduced by the VAR, see Eq. (27), $\Phi^*(\theta)$ equals the population least squares coefficients associated with (32), and the covariance matrix of x_t under the DSGE model and its VAR approximation are identical. For the ease of exposition 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 (32) as the structural model that imposes potentially misspecified restrictions on the matrices Φ and Σ.

3.2 Misspecification and Bayesian Inference

We make the following assumptions about misspecification of the DSGE model. There is a vector θ and matrices Ψ^Δ and Σ^Δ such that the data are generated from the VAR in Eq. (32)

$$\Phi = B_1(\theta) + (\Psi^*(\theta) + \Psi^\Delta) B_2(\theta), \quad \Sigma = \Sigma^*(\theta) + \Sigma^\Delta \tag{33}$$

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 θ and the model misspecification matrices Ψ^Δ and Σ^Δ are summarized in a prior distribution. In order to compare the Bayesian approach to model misspecification pursed in this paper to minimax and robust control approaches, the reader might find it helpful to think of a fictitious other, "nature", that draws the misspecification matrices Ψ^Δ and Σ^Δ from a distribution rather than maximizing the loss function to harm the policy maker.

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 Ψ^Δ and Σ^Δ. Discrepancies are defined to be "large" when they are easily detectable with likelihood ratios. The parameter λ measures the overall degree of misspecification. Large values of λ imply that large discrepancies are less likely to occur. We will now motivate the prior distribution using a thought experiment, where for ease of exposition we set $\Sigma^\Delta = 0$ and fix the DSGE model parameter vector θ.

Suppose that a sample of λT observations is generated from the DSGE model (that is, from equation (32), where $\Phi = \Phi^*$). Suppose that you use these observations to tell the DSGE ($\Psi = \Psi^*$) and the misspecified model ($\Psi = \Psi^* + \Psi^\Delta$) apart using a likelihood ratio. Since the likelihood ratio is decreasing in the number of observations λT for fixed Ψ^Δ, the misspecification is re-scaled as follows. Let

$$\Psi^\Delta = \frac{1}{\sqrt{\lambda T}} \widetilde{\Psi}^\Delta.$$

The log-likelihood ratio is

$$\begin{aligned} \ln\left[\frac{\mathcal{L}(\Psi^*, \Sigma^*, \theta | Y, X)}{\mathcal{L}(\Psi^* + \Psi^\Delta, \Sigma^*, \theta | Y, X)}\right] \quad = \quad & -\frac{1}{2} tr\Bigg[\Sigma^{*-1}\Bigg(B_2'\Psi^{*'}X'X\Psi^* B_2 - 2B_2'\Psi^{*'}X'(Y - XB_1) \\ & -B_2'(\Psi^* + (\lambda T)^{-1/2}\widetilde{\Psi}^\Delta)'X'X(\Psi^* + (\lambda T)^{-1/2}\widetilde{\Psi}^\Delta)B_2 \\ & +2B_2'(\Psi^* + (\lambda T)^{-1/2}\widetilde{\Psi}^\Delta)'X'(Y - XB_1)\Bigg)\Bigg]. \end{aligned}$$

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

$$\begin{aligned} & \ln\left[\frac{\mathcal{L}(\Psi^*, \Sigma^*, \theta|Y, X)}{\mathcal{L}(\Psi^* + \Psi^\Delta, \Sigma^*, \theta|Y, X)}\right] \\ = & -\frac{1}{2} tr\left[\Sigma^{*-1}\left(B_2'\widetilde{\Psi}^{\Delta\prime}(\lambda T)^{-1}X'X\widetilde{\Psi}^\Delta B_2 - 2B_2'\widetilde{\Psi}^{\Delta\prime}(\lambda T)^{-1/2}X'U\right)\right] \end{aligned} \tag{34}$$

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

$$I\!E^{VAR}_{\Psi^*,\Sigma^*}\left[\ln\frac{\mathcal{L}(\Psi^*, \Sigma^*, \theta|Y, X)}{\mathcal{L}(\Psi^* + \Psi^\Delta, \Sigma^*, \theta|Y, X)}\right] = -\frac{1}{2} tr\left[\Sigma^{*-1}\left(B_2'\widetilde{\Psi}^{\Delta\prime}\Gamma_{XX}\widetilde{\Psi}^\Delta B_2\right)\right]. \tag{35}$$

Here we have used Lemma 1(ii). We choose a prior density for Ψ^Δ that is proportional ($\propto$) to the Kullback-Leibler discrepancy:

$$p(\Psi^\Delta|\Sigma^*, \theta) \propto \exp\left\{-\frac{\lambda T}{2} tr\left[\Sigma^{*-1}\left(B_2'\widetilde{\Psi}^{\Delta\prime}\Gamma_{XX}\widetilde{\Psi}^\Delta B_2\right)\right]\right\} \tag{36}$$

The hyperparameter λ 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 λ, it is easy to tell the misspecified model and the DSGE model apart even for small values of the misspecification Ψ^Δ. Hence the prior density places most of its mass near the restrictions imposed by the DSGE model when λ is large, and for $\lambda = \infty$ the misspecification disappears altogether. On the contrary, if λ is close to zero the Kullback-Leibler distance can be small even for relatively large values of the discrepancy Ψ^Δ. Hence the prior is fairly diffuse. For computational reasons it is convenient to transform this prior into a prior for Ψ. Using standard arguments we deduce that this prior is multivariate normal

$$\Psi|\Sigma^*, \theta \sim \mathcal{N}\left(\Psi^*(\theta), \frac{1}{\lambda T}\left[(B_2(\theta)\Sigma^{*-1}B_2(\theta)') \otimes \Gamma_{XX}(\theta)\right]^{-1}\right). \tag{37}$$

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 θ in the empirical analysis:

$$\Psi|\Sigma, \theta \sim \mathcal{N}\left(\Psi^*(\theta), \frac{1}{\lambda T}\left[(B_2(\theta)\Sigma^{-1}B_2(\theta)') \otimes \Gamma_{XX}(\theta)\right]^{-1}\right) \tag{38}$$

$$\Sigma|\theta \sim \mathcal{IW}\Big(\lambda T\Sigma^*(\theta), \lambda T - k, n\Big), \tag{39}$$

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 θ, Ψ, and Σ. In addition to the posterior distribution of the parameters we are also interested in evaluating marginal data densities of the form

$$p(Y) = \int p(Y|\theta,\Sigma,\Phi)p_\lambda(\theta,\Sigma,\Phi)d(\theta,\Sigma,\Phi) \tag{40}$$

for various choices of the hyperparameter λ 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.

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 θ_0, Ψ_0^Δ, and Σ_0^Δ. 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

$\widetilde{x}_t = [\widetilde{y}_{t-1}, \ldots, \widetilde{y}_{t-p}]'$. Then the VAR can be rewritten in terms of deviations from the mean as follows:

$$\widetilde{y}_t' = \widetilde{x}_t' \widetilde{M}[B_1(\theta) + (\Psi^*(\theta) + \Psi^\Delta) B_2(\theta)] + u_t'. \tag{41}$$

We assume the mean $\bar{y}$ is invariant to changes in the policy parameters[3] and that the policy maker considers the following loss function

$$\mathcal{L}_T(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) = (1-\delta) I\!\!E_T \left[\sum_{t=T+1}^{T+h} \delta^{t-T-1} tr[\mathcal{W} \widetilde{y}_t \widetilde{y}_t'] \right], \tag{42}$$

where the law of motion of $\tilde{y}$ is given by (41). δ is a discount factor, θ is partitioned into policy rule parameters θ_p and taste-and-technology parameters θ_s, and $tr[\cdot]$ denotes the trace operator. The expectation in (42) is taken conditional on post-intervention parameters θ, Ψ^Δ, and Σ^Δ and the pre-intervention observations $\widetilde{y}_{T-p+1}, \ldots, \widetilde{y}_T$.

The loss function can be rewritten as

$$\mathcal{L}_T(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) = (1-\delta) \sum_{t=T+1}^{T+h} \delta^{t-T-1} \Big(tr[\mathcal{W} \mathbf{V}_T(\widetilde{y}_t)] + tr[\mathcal{W} I\!\!E_T[\widetilde{y}_t] I\!\!E_T[\widetilde{y}_t]'] \Big). \tag{43}$$

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

$$\mathcal{L}(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) = I\!\!E \left[\mathcal{L}_T(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) \right] = (1-\delta^h) tr[\mathcal{W} \mathbf{V}(\widetilde{y}_{T+1})], \tag{44}$$

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

Since the policy maker faces uncertainty with respect to the post-intervention parameters θ_s, Ψ^Δ, and Σ^Δ, and seeks to minimize posterior expected losses, 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[4]

$$L(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) = \min \left\{ \mathcal{B}, \mathcal{L}_T(\theta_p, \theta_s, \Psi^\Delta, \Sigma^\Delta) \right\}. \tag{45}$$

[3]This assumption is consistent with the DSGE model, in which the policy parameters ψ_1, ψ_2, and ρ_R do not affect steady state output growth, inflation, and interest rates.

[4]In our empirical analysis (not reported in this paper) we also calculated posterior expected losses for

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 θ_p. She has imperfect knowledge about: (i) the policy invariant private sectors' taste and technology parameters θ_s; and (ii) the degree of model misspecification captured by λ, Ψ^Δ and Σ^Δ. The uncertainty is summarized in the posterior distribution.

We consider four different scenarios for the policy invariance of the misspecification matrices Φ^Δ and Σ^Δ. 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. At this point we have no theory that lets us determine which of the scenarios will provide the most accurate prediction of the policy effects. The goal of the subsequent empirical analysis is to illustrate the sensitivity of policy predictions to assumptions on (un-structured) model misspecification.

Scenario 1: 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 dates back at least to Brainard (1967) and serves as a benchmark. More recent examples in the context of DSGE models include Laforte (2003) and Onatski and Williams (2004).

Scenario 2: 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 Hansen and Sargent (2005) and Onatski and Stock (2002). Here, instead of using a minimax argument, our Bayesian policymaker relies on her prior distribution $p(\Psi^\Delta, \Sigma^\Delta | \theta, \lambda)$ to cope with uncertainty about model misspecification. She still uses the sample to learn about θ_s and λ, however.

Scenario 3: Ψ^Δ and Σ^Δ 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. To implement the analysis, we generate draws from the marginal posterior distribution of θ_s, Ψ^Δ, and Σ^Δ, combine $\tilde{\theta} = [\tilde{\theta}_p', \theta_s']'$, and calculate $\Psi^*(\tilde{\theta}) + \Psi^\Delta$ and $\Sigma^*(\tilde{\theta}) +$

---

the loss function $(1-\delta)\sum_{t=T+1}^{T+h} \delta^{t-T-1} tr[\mathcal{W}\mathbf{V}_T(\tilde{y}_t)]$ 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 θ_p that with some probability lead to explosive behavior of the resulting VAR, and less than 20 for the other values of θ_p.

Σ^{Δ}. Here, $\tilde{\theta}_p$ is the new set of policy parameters. The choice of $\tilde{\theta}_p$ does not affect beliefs about the misspecification matrices.

Scenario 4: "Nature" generates a new set of draws from the posterior distribution of Ψ^{Δ} and Σ^{Δ} 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 θ_s, combine it with the policy parameter to obtain $\tilde{\theta} = [\tilde{\theta}_p', \theta_s']'$, and generate a draw from $p(\Psi^{\Delta}, \Sigma^{\Delta} | Y^T, \tilde{\theta}, \lambda)$. 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. For small values of λ the conditional posterior distribution of Ψ and Σ given θ is is effectively insensitive to θ. In this case Scenario 4 corresponds to analyzing policy effects 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.

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[5] 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, which we called "nature" previously, chooses the misspecification to harm the decision maker. "Nature's" choice, in our notation Ψ^{Δ} and Σ^{Δ}, 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

$$\min_{\theta_p} \quad \frac{1}{\tau} \ln \int \exp\{\tau L(\theta_p, \theta_s, \Psi^{\Delta}, \Sigma^{\Delta})\} p(\theta_s, \Psi^{\Delta}, \Sigma^{\Delta}) d(\theta_s, \Psi^{\Delta}, \Sigma^{\Delta}), \tag{46}$$

where $p(\theta_s, \Psi^{\Delta}, \Sigma^{\Delta})$ denotes the joint density of θ_s, Ψ^{Δ}, Σ^{Δ}. A positive τ makes the policy maker risk averse. It can be shown that the optimization of (46) is the solution to the

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

following zero-sum game

$$\min_{\theta_p} \max_{q(\theta_s,\Psi^\Delta,\Sigma^\Delta)} \int L(\theta_p,\theta_s,\Psi^\Delta,\Sigma^\Delta)p(\theta_s,\Psi^\Delta,\Sigma^\Delta)q(\theta_s,\Psi^\Delta,\Sigma^\Delta)d(\theta_s,\Psi^\Delta,\Sigma^\Delta) \quad (47)$$
$$-\frac{1}{\tau}\int \left(\ln q(\theta_s,\Psi^\Delta,\Sigma^\Delta)\right) p(\theta_s,\Psi^\Delta,\Sigma^\Delta)q(\theta_s,\Psi^\Delta,\Sigma^\Delta)d(\theta_s,\Psi^\Delta,\Sigma^\Delta).$$

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

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

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 distance between the distorted and the undistorted probabilities. The larger τ, 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 the four scenarios for the risk sensitive version of the policy problem. Hence, "nature" is not only drawing misspecification matrices Ψ^Δ and Σ^Δ, but at the same time also distorting the probabilities.[6]

5 The Data

In our empirical analysis we use observations on interest rates, inflation, and output growth. All data are obtained from Haver analytics (Haver mnemonics are in italics). Real output is obtained by dividing the nominal series (*GDP*) by population 16 years and older (*LF+LH*), and deflating using the chained-price GDP deflator (*JGDP*). Growth rates are computed using log-differences from quarter to quarter, and are in percent. Inflation is computed using log-differences of the GDP delflator, in percent. The nominal rate corresponds to the effective Federal Funds rate (*FFED*), also in percent. The results reported subsequently are based on a sample from 1983:Q3 to 2004:Q1.

[6] Strictly speaking only Ψ^Δ and Σ^Δ capture misspecification. This suggests that one should integrate out θ_s from $L(\theta_p,\theta_s,\Psi^\Delta,\Sigma^\Delta)$ with respect to $p(\theta_s|\theta_p,\Psi^\Delta,\Sigma^\Delta)$ prior to the exponential transformation. However, this integration is numerically cumbersome and we decided to apply the notion of robustness not only to the misspecification parameters but also to θ_s.

6 Empirical Application

We estimate the DSGE-VARs based on only three variables and hence we set most shocks equal to zero, except for the technology growth shock $\epsilon_{z,t}$, the monetary policy shock $\epsilon_{R,t}$, and the government spending shock $\epsilon_{g,t}$. Since the model is to some extent able to endogenously generate persistence in real variables, we impose that technology growth shocks are serially uncorrelated, that is, $\rho_z = 0$. In Del Negro, Schorfheide, Smets and Wouters (2004), henceforth DSSW, we did not find evidence in favor of price indexation. Therefore, we let $\iota_p = \iota_w = 0$. Moreover, we set the fixed costs $\mathcal{F} = 0$. Unlike in DSSW, we do not use observations on consumption and investment, which makes it difficult to identify the capital share and the depreciation rate. Therefore, we let $\alpha = 0.25$ and $\delta = 0.025$. Since we are not extracting information from wage and money data we fix the wage-markup parameter $\lambda_w = 0.3$, and the money demand elasticity $\nu_m = 2$. In a log-linear approximation the Calvo parameter is typically not separately identifiable from the price markup parameter λ_f, which we fix at 0.3.

6.1 Model Estimation

We begin with a direct estimation of the state-space representation of the DSGE model using Bayesian techniques described in Schorfheide (2000). Table 1 reports prior mean and standard deviations, as well as posterior means and 90% probability intervals for the structural parameters. The estimates for the inflation and output growth coefficients in the monetary policy rule are 1.43 and 0.36, respectively. Our estimate of the smoothing coefficient is fairly high compared to estimates reported elsewhere in the literature: $\hat{\rho}_r = 0.83$. The Calvo parameters for wages and prices are 0.72, and 0.79, respectively. Thus, agents change their prices on average every 4 quarters. We estimate a large degree of habit persistence, whereas the data seem to be fairly uninformative with respect to the labor supply elasticity μ_l and the cost of capital utilization a".

We proceed by estimating DSGE-VARs for values of λ between 0.25, i.e., large prior variance of the misspecification matrices Ψ^{Δ} and Σ^{Δ}, and 5, i.e., small potential misspecification. Specifically, the values of λ we consider are 0.25, .50, .75, 1, 1.50, 2, and 5. The subsequent results are based on $p = 4$ lags. Table 2 describes the posterior of the misspecification parameter λ. The table reports log marginal data densities for the directly estimated DSGE model and DSGE-VARs based on different values of λ. Differences of log marginal densities across model specifications 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 $\lambda = 0.75$, which is the specification with the largest marginal data density and, according to this likelihood-based criterion, the best fit. The posterior of λ has an inverted U-shape. There is little variation in the marginal data densities for λ values between 0.50 and 2, whereas values outside of this interval lead to a substantial deterioration in fit. We conclude that over the range of the historical sample the DSGE model is strongly dominated by DSGE-VARs with intermediate values of λ indicating that the structural model is to some extent misspecified and that its policy predictions should be interpreted with care.

Table 3 compares the posterior means of the structural parameters obtained from the estimation of DSGE-VAR specifications for various values of λ. The DSGE-VAR generates Bayesian instrumental variable estimates of the policy rule parameters ψ_1, ψ_2, and ρ_R. For large values of λ the instruments are very similar to the scores of the DSGE models' likelihood function. For values of λ near zero the instruments essentially correspond to lagged values of interest rates, inflation, and output growth. While the estimates of ψ_2 and ρ_R are fairly insensitive to the choice of λ, the estimate of the inflation coefficient rises from 1.43 to 1.99 as the prior variance of the discrepancies Ψ^Δ and Σ^Δ increases. As shown in Del Negro and Schorfheide (2004) the estimates of the remaining DSGE model coefficients can be interpreted as minimum distance estimates, in which the estimator of Ψ is projected onto the restricted subspace generated by $\Psi^*(\theta)$.

6.2 Policy Analysis

Based on the parameter estimates we calculate expected policy losses. The loss is based on Eq. (45) where the weighting matrix $\mathcal{W}$ is diagonal with elements $\frac{1}{4}$ (interest rates, annualized), 1 (inflation, annualized), and $\frac{1}{4}$ (output growth, quarter-to-quarter). Our weight on output growth is somewhat larger than in Woodford (2003, Table 6.1) reflecting a larger estimate of κ. 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 variances of annualized output growth, inflation, and interest rates are approximately 5, 2, and 6 percent respectively.

We evaluate the expected loss as a function of all the parameters characterizing the

Taylor rule (21): ψ_1 and ψ_2, the central bank's response to inflation and output, respectively, and ρ_R, the interest rate smoothing parameter. Specifically, we compute the expected loss for each point of a three dimensional grid. In this grid, ψ_1 takes 11 different values obtained by equally subdividing the interval $[1.1, 2.5]$; ψ_2 takes 6 different values obtained by equally subdividing the interval $[0, 1]$, and ρ_r takes 5 different values obtained by equally subdividing the interval $[0, 0.8]$. The results are summarized in Figures 2 and 3. Both figures depict expected loss differentials relative to the benchmark $\psi_1 = 1.8$, $\psi_2 = 0.4$, $\rho_R = 0.8$. The benchmark is chosen by selecting the point in the grid that are, roughly, closest to the sample estimates of the parameters. Negative differentials indicate an improvement relative to the benchmark. Figure 2 is a three-dimensional plot showing the expected loss as a function of ψ_1 and ψ_2 with ρ_R fixed at the benchmark value 0.8. Figure 3 depicts a slice of this three-dimensional plot, obtained by further fixing ψ_2 to its benchmark value of 0.4.

Every figure contains four charts, one for each of the scenarios described in Section 4.2. Scenario 1 shows the loss differential computed according to the state-space representation of the DSGE model. For Scenarios 2 trough 4, the charts display the results obtained from the DSGE-VAR for values of λ equal to 0.25, .50, .75, 1, 1.50, 2, and 5. Surfaces (or lines, in case of Figure 3) have colors ranging from very light grey for λ equal to .25 to dark grey for λ equal to 5, with the darkness of the surface being directly proportional to λ, that is, inversely proportional to the amount of misspecification.

In Scenario 1 the policymaker calculates the policy loss with the DSGE model, ignoring misspecification. Figure 2 shows that the loss decreases as the value of ψ_1 increases, regardless of the value of ψ_2. The increase in the response of interest rates to inflation in the Taylor rule results in a drop of inflation variability, which in turn implies a lower volatility of interest rates as well. The drop in the loss is particularly steep as ψ_1 increases from 1.1 to 1.5, but flattens thereafter, as can be appreciated from the two dimensional plot in Figure 2. While Figure 2 reports only the expected loss differential, we also computed (but did not display) 90 percent probability bands in order to characterize the dispersion in the distribution. These bands show that there is substantial uncertainty regarding the magnitude of the drop in the loss when ψ_1 increases from 1.1 to 1.5, but very little uncertainty in the loss differential for values of ψ_1 larger than 1.5.

The loss differential increases as ψ_2 varies from 0 to 1. The higher variance of output resulting from a low response in the Taylor rule is more than out-weighted by the lower variability of inflation and the interest rate. Overall, the changes in loss as a function of ψ_2 are roughly one order of magnitude smaller than those resulting from changes in ψ_1: in

Figure 2 the surface looks nearly flat along the ψ_2 dimension. For the sake of brevity we do not show how the loss differential changes as a function of ρ_R. We find that for all Scenarios the optimal value of ρ_R is essentially determined by the value of ψ_1, and independent from ψ_2. For low values of ψ_1 it is optimal to have very low interest rate persistence ($\rho_R = 0$). For high values of ψ_1 it is optimal to make the interest rate process very persistent ($\rho_R = 0.8$). In conclusion, the policy analysis conducted using the DSGE model suggests that the policy maker should choose high values of ψ_1 and low values of ψ_2, although quantitatively the differences in the expected loss are small as long as ψ_1 is larger than 1.5. The inference about the misspecification parameter λ 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 (λ) but refuses to learn about the precise nature of the misspecification: She therefore uses the prior to generate draws of Ψ^{Δ} and Σ^{Δ} rather than the posterior. Again, dark shades of grey in Figure 2 correspond to a small amount of noise (large values of λ), whereas lighter shades of grey are associated with a larger variance of the noise (small values of λ). Not surprisingly when the misspecification is small (say, $\lambda = 5$) the shape of the loss does not change substantially relative to Scenario 1. But as the amount of misspecification increases the loss profile becomes flatter. In particular, the large drop in the expected loss differential that characterized the increase in ψ_1 from 1.1 to 1.5 under the DSGE model nearly disappears under Scenario 2 with low values of λ, as can be easily appreciated from Figure 3.

Under the DSGE model the mechanics of the rational expectations equilibrium imply that high values of ψ_1 help to anchor inflationary expectations. Since in equilibrium inflation moves less than under low values of ψ_1, interest rates need to move less as well. The presence of substantial misspecification changes these dynamic responses. A decomposition of the loss into its three components indicates that for small values of λ interest rate variability actually rises as the central bank responds more strongly to inflation, albeit slightly. However this rise is roughly off-set by the drop in inflation variability. The overall minimum for the loss differential is still achieved for high values of ψ_1 and low values in ψ_2, as in Scenario 1. But the most important policy implication under Scenario 1, namely to stay away from low values of ψ_1, loses much of its strength under the misspecification considered in Scenario 2.

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 Ψ^{Δ} and Σ^{Δ} are policy invariant. Figure 2 shows that under this Scenario the loss surface is quite different than under Scenarios 1 and 2 whenever the amount of misspecification is non-negligible. First of all, there is much more curvature with respect to ψ_2. Second, the loss profile is no longer strictly decreasing in ψ_1 for values of λ less than 1.5, as can be seen from Figure 3. For small λs the loss differential is a U-shaped function of ψ_1, with the minimum attained at the value of roughly 1.8.

In order to understand this finding, it is useful to look at Figure 4, which shows under each scenario the fraction of draws for which either the linearized DSGE model does not have a unique stable rational expectations solution (indeterminacy) or the resulting vector autoregressive system is explosive (explosiveness). The figure shows that this fraction is virtually zero under Scenario 1 (of course here only indeterminacy is possible), very small under Scenarios 2 and 4, but quite large under Scenario 3 particularly for small values of λ. This is particularly true for large values of ψ_2, but also for either small or large values of ψ_1. Specifically, the fraction increases the further we move away from the estimated policy parameters, which roughly coincide with the benchmark. In Scenario 3 this fraction is mainly composed by draws that generate explosiveness, rather than indeterminacy.

To see why this happens, 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 Ψ^{Δ}. 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 Ψ^{Δ}, which is assumed to be invariant. If the policy parameters are close to estimated ones, the sum of $\Psi^*(\tilde{\theta})$ and Ψ^{Δ} returns the estimated VAR parameters, which generally do not have explosive roots. But as we move away from the estimated policy parameters, for small values of λ the sum of $\Psi^*(\tilde{\theta})$ and Ψ^{Δ} often delivers new VAR parameters whose roots are explosive. Whenever λ is large, however, Ψ^{Δ} is negligible and the new VAR parameters roughly coincide with $\Psi^*(\tilde{\theta})$, which is non explosive. Hence, for small λ the policy recommendation under Scenario 3 implies that the policymaker should not stray away too much from the estimated policy parameters, for this increases the risk of encountering explosive behavior when the misspecification is large. (Cogley and Sargent (2005) also present results where the policy recommendation coming from their model is largely driven by the concern for explosiveness). However, not all deviations from the estimated policy parameters matter equally: i) changes in ψ_2 matter more than changes in ψ_1, and ii) decreases in ψ_2 are far less harmful than increases.

Finally, under Scenario 4 the policymaker again uses sample information to learn about potential model misspecification. Unlike in Scenario 3, the policymaker now asks the question: What would the estimates of the discrepancies Ψ^Δ and Σ^Δ be if the new policy had been in place during the sample period? Figure 2 shows that under this Scenario the loss surface is similar to that computed under Scenarios 2. Namely, when the misspecification is small (say, $\lambda = 5$) the shape of the loss does not change substantially relative to Scenario 1. But as the amount of misspecification increases the loss profile becomes flatter. As we see from Figure 4, 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 very small values of λ the dynamics for all equations other than the Taylor rule are approximately independent of the policy parameters. This explains why the loss surface is nearly flat whenever the misspecification is high.

At this point we have no theory that lets us determine which of the scenarios will provide the most accurate prediction of the policy effects. We show that the results of the policy analysis depend on: (i) whether the policymaker relies on the data to assess the degree of misspecifications, i.e., learns about λ; and (ii) the assumption she makes on the process driving the discrepancies between the DSGE model and the data in the aftermath of the policy intervention. According to our analysis, the risks associated with straying away from the historical policy parameters are very different depending on both the overall size of the misspecification and the assumptions on how the nature of the misspecification changes with policy. Nevertheless, a fairly robust policy recommendation emerges from our analysis: the central bank should avoid strong responses to output growth movements and not react weakly to inflation fluctuations.

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 the overall shape of the contours, and hence the gist of our conclusions, stay roughly the same.

Finally, Figure 5 compares the loss differentials that we just analyzed with those obtained under the risk-sensitive version of our problem. For each scenario the Figure shows the risk-sensitive loss (dark grey) as well as the risk-neutral loss (light grey). For Scenarios 2 through 4 the loss-differentials are computed for λ equal to 0.75 (best-fitting model).

A caveat of our analysis so far is that we do not distinguish between uncertainty in the deep parameters θ and in the misspecification parameters Ψ^{Δ} and Σ^{Δ}. In principle we want to be robust gains the latter, but not necessarily the former.

For Scenario 1, where the risk-sensitivity is only with respect to the deep parameters θ_s, we find that the risk sensitive loss is generally not too different from the plain-vanilla one. The only major difference is that risk-sensitivity suggests to stay away from low values of ψ_2. These are the values that minimize the risk-neutral loss . In Scenario 2, risk-sensitivity would induce the policymaker to avoid low values of ψ_1 (as well as ψ_2). Interestingly, risk-sensitivity re-instates the large drop in loss related to increasing ψ_1 from 1.1 to 1.5, which in absence of risk-sensitivity is no longer there for $\lambda = .75$. In Scenario 3 the surface is much flatter under the risk-sensitive loss than under the expected loss. Tentatively, we explain this finding as follows: for most grid-points there is a non-negligible probability of encountering explosive draws. A perverse "nature" would tilt the distribution precisely toward this outcome. Hence, under Scenario 3, the policy maker is "doomed no-matter-what." Lastly, also under Scenario 4 the slope is flatter under risk-sensitivity than under the expected loss. While so far Scenarios 2 and 4 have looked fairly similar, under risk-sensitivity the two scenarios deliver very different implications. In particular, Scenario 4 suggests that the risk-sensitive policymaker should choose low values of ψ_1 (and ψ_2), although the loss differential is not very large.

7 Conclusion

Current DSGE models are to some extent misspecified, even large-scale models such as the one in Smets and Wouters (2003). While they allow policymakers to assess the effects of rare policy changes on the expectation formation and decision rules of private agents, their fit is typically worse than the fit of alternative econometric models, such as VARs estimated with well-calibrated shrinkage methods. The DSGE-VARs studied in Del Negro and Schorfheide (2004) and Del Negro, Schorfheide, Smets, and Wouters (2004) provide a framework that allows researchers to account for model misspecification. In this paper we developed techniques to conduct policy analysis with potentially misspecified DSGE models and applied them to a New Keynesian DSGE model with capital accumulation and several real and nominal frictions. We studied the effect of changing the response to inflation under an ad-hoc loss function that penalizes inflation, output growth, and interest rate variability.

We view our framework as an attractive alternative to robust control approaches to model misspecification that deserves to be explored in future research.

References

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

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

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

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.

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

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.

Hansen, Lars P and Thomas J. Sargent (2005): "Robust Control and Economic Model Uncertainty," *Manuscript*, New York University, Department of Economics.

Ingram, Beth F. and Charles H. Whiteman (1994): "Supplanting the Minnesota prior – Forecasting macroeconomic time series using real business cycle model priors." *Journal of Monetary Economics*, **34**, 497-510.

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.

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.

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

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

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 Ψ as

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

The prior density for Ψ conditional on Σ 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 Σ is given by

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

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

Conditional Posterior of Ψ: Combining the the prior density (A.2) with the likelihood function (A.1) yields

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

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

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

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

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

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

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

Hence,

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

If the DSGE model satisfies Eq. (27) and the error $u_{1,t}$ is orthogonal to x_t then

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

and

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

While the conditional posterior distribution of Σ 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

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

Our proposal distribution for Σ is

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

Conditional Posterior of θ: The posterior distribution of θ is irregular. Its density is proportional to the joint density of Y, Ψ, Σ, and θ, 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 θ and (ii) calculate the inverse Hessian at the mode, denoted by $V_{\bar{\theta},DSGE}$. (iii) We then use a random-walk Metropolis step with proposal density

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

where $\theta_{(s-1)}$ is the value of θ 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: DSGE Model – Parameter Estimation Results

Parameter	Prior		Posterior		
	Mean	Stdd	Mean	90% Interval	
ψ_1	1.500	0.250	1.433	1.131	1.770
ψ_2	0.125	0.100	0.361	0.102	0.596
ρ_r	0.500	0.200	0.834	0.800	0.869
r_a^*	1.000	1.000	0.577	0.000	1.298
π_a^*	3.000	2.000	4.602	3.085	6.073
γ_a	2.000	1.000	1.945	1.358	2.518
h	0.800	0.100	0.987	0.979	0.997
ν_l	2.000	0.750	2.464	1.131	3.684
ζ_w	0.750	0.100	0.721	0.538	0.957
ζ_p	0.750	0.100	0.794	0.725	0.868
s'	4.000	1.500	6.274	3.734	8.725
$a'\,'$	0.200	0.075	0.225	0.109	0.332
g^*	0.150	0.050	0.131	0.057	0.200
ρ_g	0.800	0.050	0.904	0.867	0.943
σ_z	0.400	2.000	2.086	1.234	2.958
σ_g	0.300	2.000	0.551	0.470	0.634
σ_r	0.200	2.000	0.142	0.121	0.162

Notes: The table reports prior means and standard deviations, and posterior means and 90 percent probability intervals for the estimated parameters. See Section 2 for a definition of the DSGE model's parameters, and Section 5 for a description of the data. We are reporting annualized values for π^*, r^*, and γ (a-subscript). The following parameters were fixed: $\alpha = 0.25$, $\delta = 0.025$, $\iota_p = \iota_w = 0$, $\mathcal{F} = 0$, $\lambda_f = \lambda_w = 0.3$, $\chi = 0$, $\nu_m = 2$, $\rho_z = 0$. All shocks other than ϵ_z, ϵ_R, ϵ_g are equal to zero.

Table 2: LOG MARGINAL DATA DENSITIES AND POSTERIOR ODDS

Specification	$\ln p(Y)$	Post Odds
DSGE Model	-313.58	4E-11
DSGE-VAR, $\lambda = 5.0$	-297.01	6E-04
DSGE-VAR, $\lambda = 2.0$	-293.96	0.012
DSGE-VAR, $\lambda = 1.5$	-292.83	0.039
DSGE-VAR, $\lambda = 1.0$	-290.88	0.270
DSGE-VAR, $\lambda = .75$	-289.58	1.000
DSGE-VAR, $\lambda = .50$	-289.78	0.816
DSGE-VAR, $\lambda = .25$	-298.50	1E-04

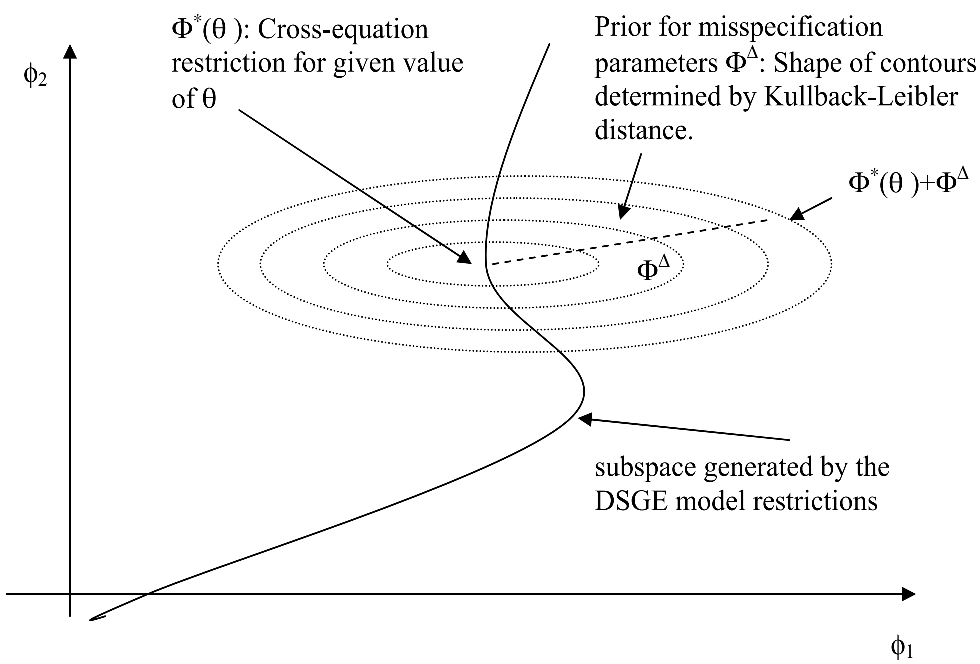
Notes: The marginal data densities are obtained by integrating the likelihood function with respect to the model parameters, weighted by the prior density conditional on λ. The difference of log marginal data densities can be interpreted as log posterior odds under the assumption of that the two specifications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

Table 3: DSGE-VAR – Parameter Estimation Results

Parameter	Prior Mean	Posterior Mean (λ)							
		0.25	0.5	0.75	1.0	1.5	2.0	5.0	DSGE
ψ_1	1.500	1.990	1.788	1.774	1.714	1.824	1.669	1.650	1.433
ψ_2	0.125	0.275	0.278	0.263	0.259	0.285	0.296	0.315	0.361
ρ_r	0.500	0.836	0.845	0.849	0.857	0.861	0.855	0.856	0.834
r_a^*	1.000	0.537	0.378	0.346	0.378	0.498	0.419	0.515	0.577
π_a^*	3.000	3.596	3.392	3.442	3.431	3.782	3.704	3.980	4.602
γ_a	2.000	1.925	1.879	2.081	1.943	2.214	2.044	2.218	1.945
h	0.800	0.944	0.882	0.919	0.970	0.982	0.984	0.987	0.987
ν_l	2.000	2.043	2.161	2.097	2.245	2.326	2.501	2.451	2.464
ζ_w	0.750	0.726	0.728	0.732	0.755	0.727	0.739	0.745	0.721
ζ_p	0.750	0.699	0.618	0.640	0.688	0.739	0.746	0.773	0.794
a''	0.200	0.204	0.203	0.220	0.207	0.197	0.214	0.208	0.225
s'	4.000	4.296	4.429	4.503	4.565	4.540	4.500	5.091	6.274
g^*	0.150	0.149	0.158	0.139	0.142	0.141	0.142	0.136	0.131
ρ_g	0.800	0.813	0.823	0.822	0.826	0.823	0.831	0.836	0.904
σ_z	0.400	0.956	0.912	1.033	1.322	1.689	1.837	2.139	2.086
σ_g	0.300	0.303	0.339	0.365	0.369	0.376	0.390	0.424	0.551
σ_r	0.200	0.123	0.129	0.132	0.128	0.132	0.134	0.137	0.142

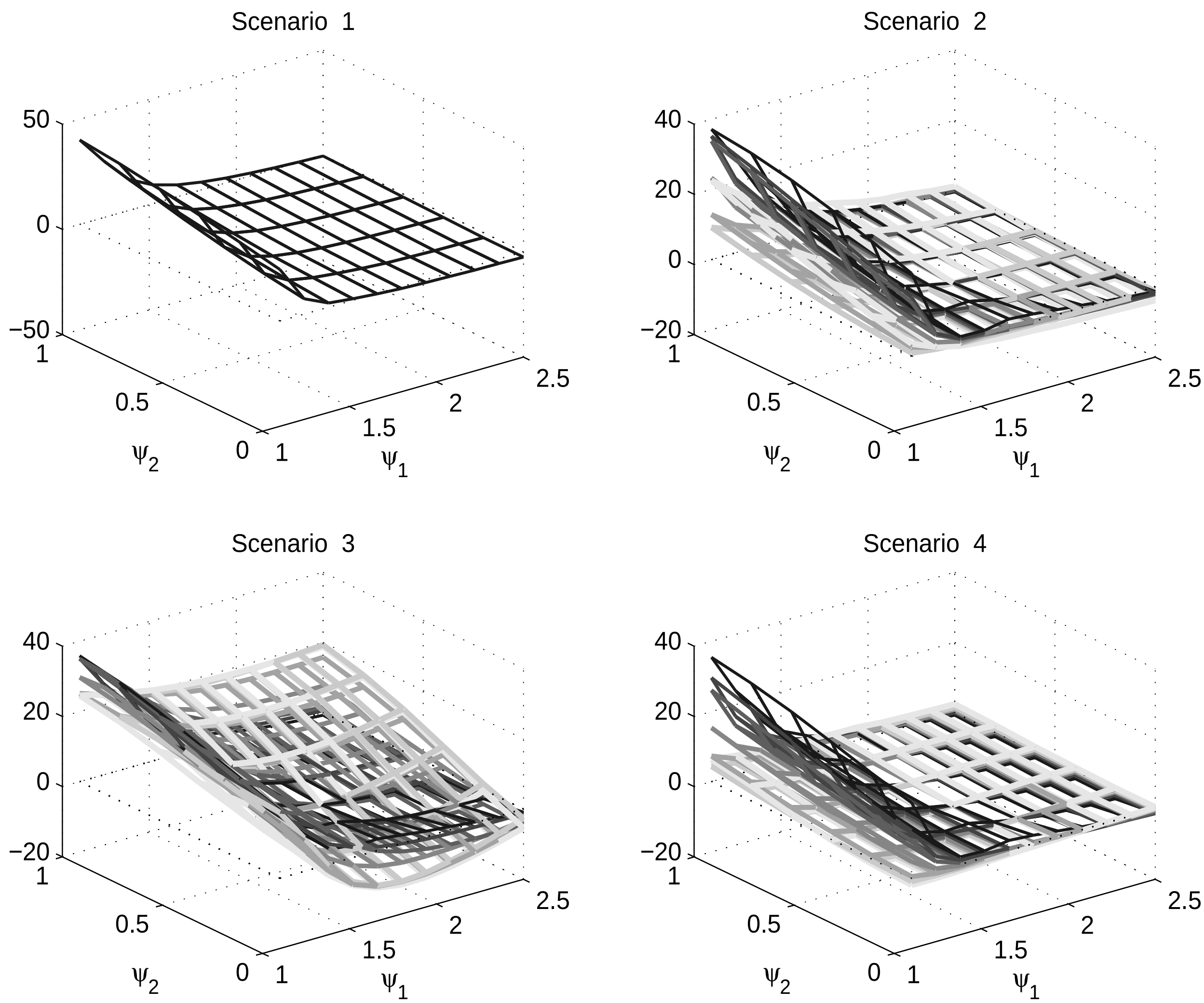
Notes: The table reports prior and posterior means for the DSGE-VAR as a function of λ and the DSGE model. See Section 2 for a definition of the DSGE model's parameters, and Section 5 for a description of the data. We are reporting annualized values for π^*, r^*, and γ (a-subscript). The following parameters were fixed: $\alpha = 0.25$, $\delta = 0.025$, $\iota_p = \iota_w = 0$, $\mathcal{F} = 0$, $\lambda_f = \lambda_w = 0.3$, $\chi = 0$, $\nu_m = 2$, $\rho_z = 0$. All shocks other than ϵ_z, ϵ_R, ϵ_g are equal to zero.

Figure 1: Stylized View of DSGE Model Misspecification



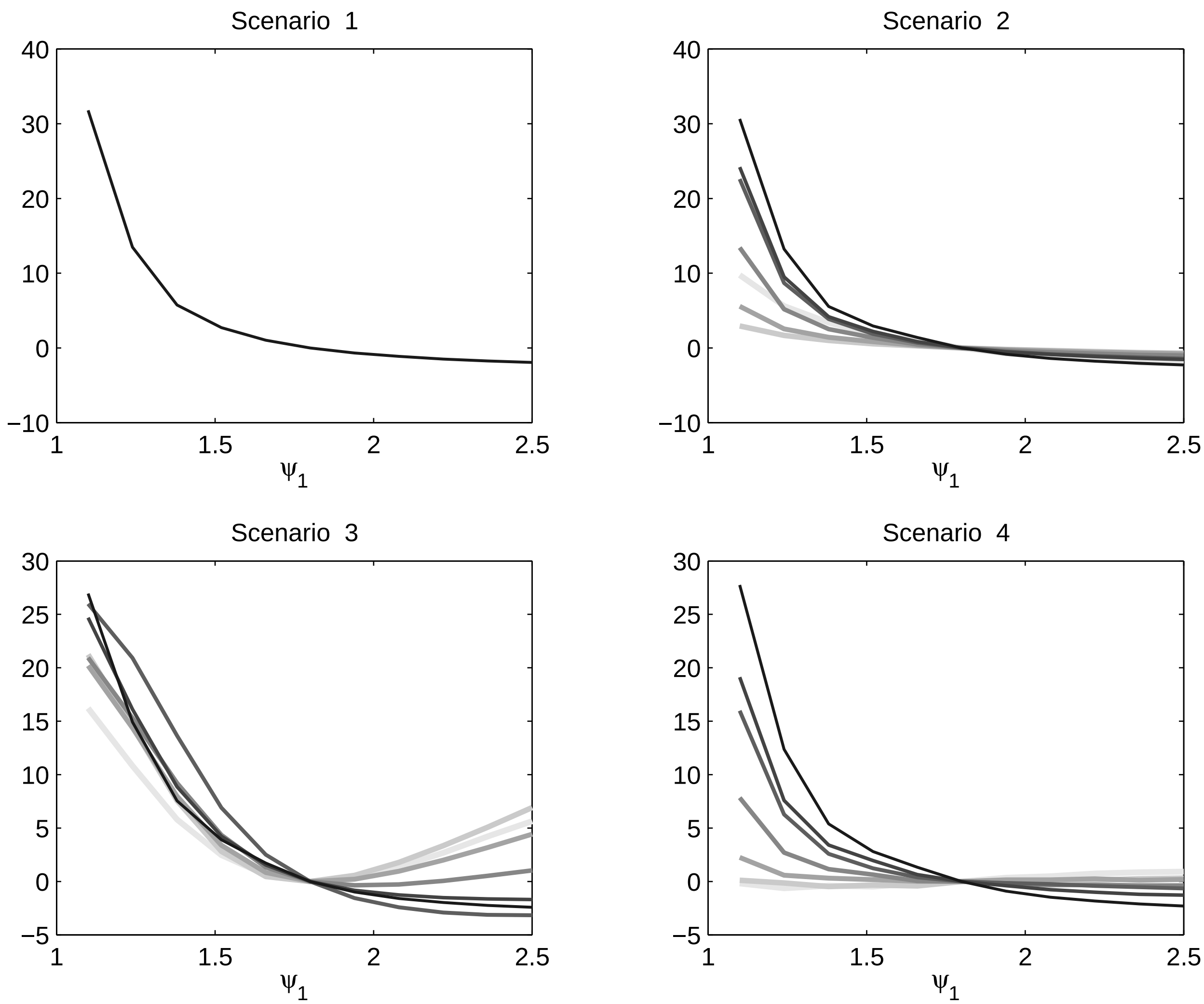
Notes: $\Phi = [\phi_1, \phi_2]'$ can be interpreted as the VAR parameters, and $\Phi^*(\theta)$ is the restriction function implied by the DSGE model.

Figure 2: EXPECTED POLICY LOSS DIFFERENTIALS - As a function of ψ_1 and ψ_2



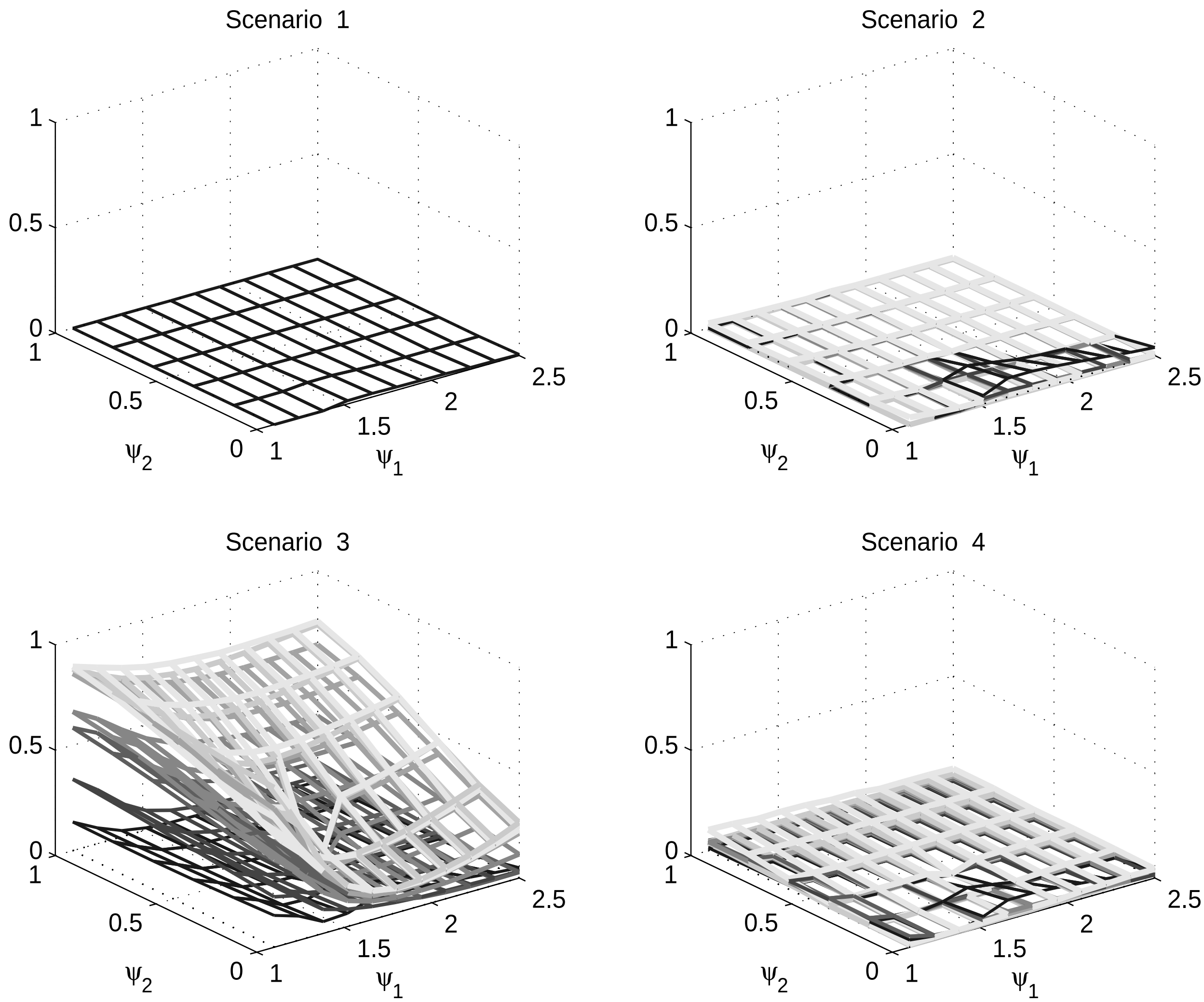
Notes: Mean policy loss differentials as a function of ψ_1 and ψ_2 relative to baseline policy rule $\psi_1 = 1.8$, $\psi_2 = 0.4$, $\rho_R = 0.80$. All numbers are computed fixing the value of ρ_R at the benchmark. Negative differentials signify an improvement relative to baseline rule. Scenario 1 shows the loss differential computed according to the DSGE model. Scenarios 2, 3 , and 4 (see section 4.2) take misspecification into account. For Scenarios 2 through 4, the charts display the results obtained for values of λ equal to 0.25, .50, .75, 1, 1.50, 2, and 5. Surfaces' color ranges from very light grey ($\lambda = .25$) to dark grey ($\lambda = 5$), with the darkness of the surface being directly proportional to λ.

Figure 3: Expected Policy Loss Differentials - As a function of ψ_1



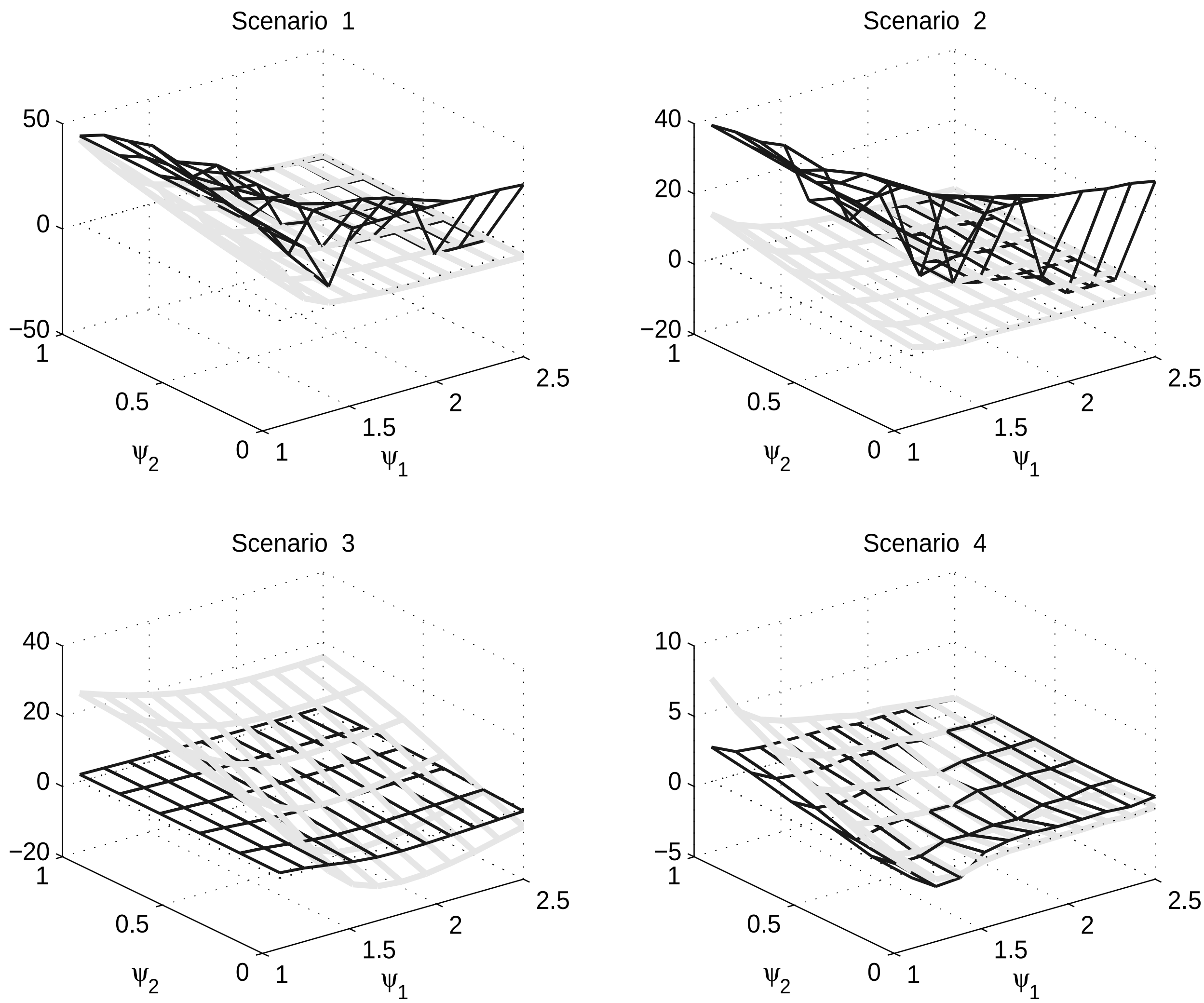
Notes: Mean policy loss differentials as a function of ψ_1 relative to baseline policy rule $\psi_1 = 1.8$, $\psi_2 = 0.4$, $\rho_R = 0.80$. All numbers are computed fixing the values of ψ_2 and ρ_R at the benchmark. Negative differentials signify an improvement relative to baseline rule. Scenario 1 shows the loss differential computed according to the DSGE model. Scenarios 2, 3, and 4 (see section 4.2) take misspecification into account. For Scenarios 2 through 4, the charts display the results obtained for values of λ equal to 0.25, .50, .75, 1, 1.50, 2, and 5. Surfaces' color ranges from very light grey ($\lambda = .25$) to dark grey ($\lambda = 5$), with the darkness of the surface being directly proportional to λ.

Figure 4: Fraction of Draws Featuring Explosiveness/Indeterminacy



Notes: Fraction of draws for which either the linearized DSGE model does not have a unique stable rational expectations solution (indeterminacy) or the resulting vector autoregressive system is explosive (explosiveness), computed as a function of ψ_1 and ψ_2. All numbers are computed fixing the value of ρ_R at the benchmark. at the benchmark. Negative differentials signify an improvement relative to baseline rule. Scenario 1 shows the loss differential computed according to the DSGE model. Scenarios 2, 3, and 4 (see section 4.2) take misspecification into account. For Scenarios 2 through 4, the charts display the results obtained for values of λ equal to 0.25, .50, .75, 1, 1.50, 2, and 5. Surfaces' color ranges from very light grey ($\lambda = .25$) to dark grey ($\lambda = 5$), with the darkness of the surface being directly proportional to λ.

Figure 5: Expected Policy Loss Differentials - Risk-Neutral versus Risk-Sensitive



Notes: Policy loss differentials relative to baseline policy rule $\psi_1 = 1.8$, $\psi_2 = 0.4$, $\rho_R = 0.80$. All numbers are computed fixing the value of ρ_R at the benchmark. Negative differentials signify an improvement relative to baseline rule. Scenario 1 shows the loss differential computed according to the DSGE model. Scenarios 2, 3 , and 4 (see section 4.2) take misspecification into account. For each scenario, the expected loss differential is shown in light grey, and the risk-sensitive loss differential in dark grey.