Discussion Papers in Economics

DESIGNING ROBUST MONETARY POLICY USING PREDICTION POOLS

By

Szabolcs Deák
(University of Surrey),

Paul Levine
(University of Surrey),

Afrasiab Mirza
(University of Birmingham),

&

Joseph Pearlman
(City University).

DP 12/19

School of Economics
University of Surrey
Guildford
Surrey GU2 7XH, UK
Telephone +44 (0)1483 689380
Facsimile +44 (0)1483 689548
Web https://www.surrey.ac.uk/school-economics

ISSN: 1749-5075

Designing Robust Monetary Policy Using Prediction Pools*

Szabolcs Deák, University of Surrey, s.deak@surrey.ac.uk
Paul Levine, University of Surrey, p.levine@surrey.ac.uk
Afrasiab Mirza, University of Birmingham, a.mirza@bham.ac.uk
Joseph Pearlman, City University London, Joseph.Pearlman.1@city.ac.uk,

June 5, 2019

Abstract

How should a forward-looking policy maker conduct monetary policy when she has a finite set of models at her disposal, none of which are believed to be the true data generating process? In our approach, the policy maker first assigns weights to models based on relative forecasting performance rather than in-sample fit, consistent with her forward-looking objective. These weights are then used to solve a policy design problem that selects the optimized Taylor-type interest-rate rule that is robust to model uncertainty across a set of well-established DSGE models with and without financial frictions. We find that the choice of weights has a significant impact on the robust optimized rule which is more inertial and aggressive than either the non-robust single model counterparts or the optimal robust rule based on backward-looking weights as in the common alternative Bayesian Model Averaging. Importantly, we show that a price-level rule has excellent welfare and robustness properties, and therefore should be viewed as a key instrument for policy makers facing uncertainty over the nature of financial frictions.

JEL codes: D52, D53, E44, G18, G23.

Keywords: Bayesian estimation, DSGE models, Financial frictions, Forecasting, Prediction Pools, Optimal Simple Rules.

*Corresponding author: Paul Levine, School of Economics, University of Surrey, Guildford, GU2 7XH, UK. Email: p.levine@surrey.ac.uk. We acknowledge financial support from the ESRC, grant reference ES/K005154/1 and from the EU Framework Programme 7, grant agreement number 612796.

1 Introduction

"All models are wrong, but some are useful."

— George Box

We study the problem of designing simple policy rules when all models are wrong yet *every* model can be useful. We consider an environment with three forms of uncertainty. The first is standard and derives from uncertain future shocks; the second is parameter uncertainty within each competing model, which we refer to as "within-model uncertainty"; the third source of uncertainty is the existence of multiple competing models, referred to as "across-model uncertainty."

The novelty of our paper lies in the way we handle this third form of uncertainty in the design of optimized Taylor-type monetary policy rules. Specifically, we follow the procedure of Geweke and Amisano (2012) to form prediction pools where weights are assigned to models on the basis of their forecasting accuracy, rather than in-sample data fit as in the common alternative Bayesian Model Averaging (BMA). These weights are then used to solve a policy design problem that selects the optimized Taylor-type interest-rate rule that is robust to all three forms of uncertainty. Unlike BMA which assumes that one of the models is the true data generating process, prediction pools allow us to consider that all models among a comparative set are misspecified, but they all may be useful at different periods of time.

We apply our methodology to investigate the welfare consequences of alternative monetary policy rules using three medium-scale new Keynesian DSGE models. The Smets and Wouters (2007) model (henceforth, SW) is the workhorse model in policy-making institutions for forecasting and policy analysis. The other two models build financial frictions into the SW model along the lines of Gertler *et al.* (2012) (henceforth, GKQ) and Bernanke *et al.* (1999) (henceforth, BGG), respectively. These two models represent the leading theories in modelling financial frictions in the macroeconomic literature. Hence, our model pool can be motivated by considering a policy maker who is uncertain how to incorporate financial frictions into a DSGE model, or if they should be incorporated at all. We estimate the models with Bayesian methods using US data on seven key macroeconomic variables over the sample period 1966:1-2017:4.

Our main results are as follows. First and foremost, the choice of model weighting strategy matters for the optimized rule. Using the optimal prediction pool weights leads to a more persistent policy rule that also responds more aggressively to changes in inflation and output growth compared to the policy rules obtained using the BMA weights. Second, within-model uncertainty has only a small impact on the optimized rule. This provides support for the practice of using only the mean or the mode of estimated posterior mode or mean of the parameters for policy design. Third, we find that both within-model and across-model robustness results in more inertial and more aggressive rules than the non-robust single model counterparts. Fourth, a special case of our interest-rate rule is a simple price rule and this has particularly good robustness features.

We restrict our attention to optimized simple rules, that is to find the optimal parameter values in a Taylor-type monetary policy rule. There are several reasons for this. First, simple rules are transparent and easy to implement and to communicate. Second, a large literature has arrived at a consensus that they mimic the Ramsey-optimal policy. Finally, the same literature suggests that they already contain good robustness properties compared with Ramsey-optimal policy.

Overall, our methodology is suited to a policy maker who does not believe any one model to be correct. She would then be advised to use a forward-looking procedure such as prediction pools to weight models rather than a backward-looking one such as Bayesian model averaging (henceforth BMA) which is based solely on in-sample model fit. Since the former is less likely to see one model completely dominate, she is then more cautious in rejecting models that appear to be inferior. The intuition is that such models, while not particularly accurate most of the time, may well be very useful some of the time. Optimal policy should take this into consideration rather than rejecting the use of models that perform poorly simply from an in-sample likelihood perspective (BMA). Moreover, we find that prediction pooling is a far more robust method for attaching weight to models as these are less sensitive to outliers and therefore evolve more smoothly over time than corresponding weights from BMA.

The rest of the paper is organized as follows. Section 2 locates our contribution in the extensive literature on model selection and robust policy design. We survey three strands of literature that our paper draws upon. The first is related to the extensive statistical literature on Bayesian predictive methods for assessing, comparing and selecting models. The second is the current generation of Bayesian-estimated micro-founded dynamic stochastic general equilibrium (DSGE) models, which are frequently employed in Central Banks for forecasting and for the computation of optimal policy. The third is the literature on robust policy. Section 3 outlines the weighting strategy based on model forecasting performance and how the model weights and the posterior distribution are combined to design policy that is robust with respect to both within-model and between-model uncertainty. Section 4 describes the models in our application and summarizes the characteristics of the posterior distribution that are most important for optimal policy design. Section 5 discusses the optimal weights in our prediction pool while Section 6 reviews the optimal policy rules that are robust within- and across-models. Full details of the models and their estimation are provided in Online Appendices along with additional results.

2 Related Literature

This paper is related to three strands of literature. First, it is related to the extensive statistical literature on Bayesian predictive methods for assessing, comparing and selecting models (see Vehtari and Ojanen, 2012, for a survey). Within this literature model selection (including more than one model) proceeds via maximization of an expected utility/loss function using the predictive distribution. A broad range of loss functions and various types of mis-specification errors have been considered in the literature. Following Bernardo and Smith (1994), all methods can be classified in accordance with the type of mis-specification error that the method seeks to address. In their terminology, M-closed or M-completed refer to methods that assume the true data generating process to be within the set of models that are considered. Techniques that fall into these categories include BMA, and using an encompassing model. The latter can be viewed as a more general version of the former with a continuous rather than a discrete distribution over priors. On the other hand, our method which is based on prediction pooling as in Geweke (2010a) and Geweke and Amisano (2011) falls into the M-open category where the true data generating process is not assumed to be among the candidate models.[1]

One particular criterion used in the literature is a scoring rule that measures forecast

[1]In the language of Geweke (2010b), for BMA the model space is 'complete', i.e., the space includes the DGP whereas for prediction pools the space is 'incomplete'. See Section 5.1 for a rigorous treatment of this point.

accuracy. A particular form of selection then amounts to combining density forecast estimates as a means of improving forecasting accuracy as measured by a scoring rule (see for example Gneiting and Raftery, 2007; Hall and Mitchell, 2007). In Geweke and Amisano (2011), the utility/loss function is a scoring rule that maximizes forecast accuracy, and they compare BMA with linear combinations of predictive densities, so-called 'opinion pools', where the weights on the component density forecasts are optimized to maximize the score, typically the logarithmic score, of the density combination as suggested in Hall and Mitchell (2007).

Kapetanios *et al.* (2015) develop an extension of this method whereby the weights can vary by region of the density to allow additional focus on the variable one is attempting to forecast. We use the method proposed by Geweke and Amisano (2011) to combine the forecasts from different models as it allows us to be agnostic about the variables that need to be forecast, and also as it is straightforward to implement. Our paper is most closely related to Del Negro *et al.* (2016) who develop a methodology for combining forecasts across two DSGE models, one with and one without financial frictions. The dynamic nature of their procedure, whereby the relative weights placed on forecasts across models vary over time, leads to an improvement over real-time forecasts produced by alternative methods. Importantly, they use the improved forecasts to carry out a novel counter factual analysis that re-examines how policy makers should respond to labour market conditions following a financial shock. Our paper departs from Del Negro *et al.* (2016) in that we seek to use the forecasts to design robust optimal policy across models rather than focus on a counter factual exercise. As such we take the analysis in Del Negro *et al.* (2016) a step further by asking how policy ought be designed when the policymaker is aware that models are mis-specified.

Second, our paper is also related to the current generation of Bayesian-estimated micro-founded dynamic stochastic general equilibrium (DSGE) models. These models are frequently employed in Central Banks and used for forecasting and for the computation of optimal policy in the form of optimized Taylor-type rules (see, for example Christiano *et al.*, 2005; Smets and Wouters, 2007; Schmitt-Grohe and Uribe, 2007; Levine *et al.*, 2007). Optimized constrained simple rules were first proposed by Levine and Currie (1987) in a linear-quadratic framework. Woodford (2003, Chapter 7) discussed and modified the welfare loss criterion in that paper so as to minimize only the *stochastic component* leading to a time-consistent policy choice. We follow this approach in our computation of robust optimized rules.

Third, this paper is also related to a large literature on robust policy. Sims (2002, 2007, 2008) in particular has argued that policy makers are still very far from exploiting the full richness of the Bayesian (or "probability models") approach.[2] A related literature compares optimized constrained simple rules with their optimal unconstrained counterparts (see, for example Levine and Currie (1987), Schmitt-Grohe and Uribe 2007; Brock *et al.* 2007b; Orphanides and Williams. 2008; a review is provided by Taylor and Williams 2010). A common finding in this literature is that simple rules can closely mimic optimal policies and perform well in a wide variety of models. By contrast optimal policy can perform very poorly if the policymaker's reference model is mis-specified. The reason for this is that optimal polices can be overly fine tuned to the particular assumptions of the reference model. If the model is the correct one all is well; but if not, the costs can be high. In contrast, our chosen simple monetary policy rules are designed to take account of only the most basic principle of monetary policy of leaning against the wind of inflation and output movements. Because they are not fine tuned to specific model assumptions, they are more robust to mistaken assumptions regarding the parameters of the model ('within-model robustness') or to basic modelling features ('between-model robustness').

Our approach differs from the existing literature in several important respects. A recent literature draws on Hansen and Sargent (2007) in assuming uncertainty is unstructured, with malign Nature 'choosing' exogenous disturbances to minimize the policymaker's welfare criterion ("robust control").[3] Robust control may be appropriate if little information is available on the uncertainty facing the policymaker. But are policymakers ever in such a "Knightian" world? CBs devote considerable resources to assessing the forecasting

[2]Formally, a probability model is a mathematical representation of a stochastic phenomenon, defined by its sample space (i.e., the set of all possible outcomes), events within the sample space, and probabilities associated to each event, Ross (2006). He views the probability-models approach as reflecting policymaking in practice by committees comprising individuals with separate views (models) of how the economy works and of the likely outlook (in the context of that model). Each model (or outlook) has an estimated parameter probability distribution which embodies its own measure of within-model uncertainty. Aggregating those views mirrors and substantiates the probability-models approach. Although any model is imperfect, the greater the uncertainty the more policymakers may benefit from pooling information *across and within* models, as we do in this paper. Our paper follows Levin and Williams (2003); Orphanides and Williams (2007); Ilbas *et al.* (2012) and Tetlow (2015) in focusing on simple, robust optimized Taylor-type rules.

[3]See, for example, Dennis *et al.* (2009) and Ellison and Sargent (2012).

properties of the approximating model, those of rival models, and estimates of parameter uncertainty gleaned from various estimation methods. To then fail to fully utilize the fruits of such exercises seems incongruous and a counsel of despair. Also, robust control pursues fully optimal rather than simple optimal rules. Yet Levine and Pearlman (2010) show if one designs simple operational rules, that mimic the fully optimal but complex one, then they take the form of highly unconventional Taylor Rules which must respond to Nature's malign interventions. Furthermore, robust control in general satisfies a supremum condition rather than a maximum condition; this implies that the supremum may well on be on the edge of being an unstable solution. Rules with these properties may be very hard to sell to policymakers.[4]

Our paper also differs from studies that design robust rules across competing models, but attach probabilities to models under the assumption that one of the models is the true data generating process. For instance, the 'rival models' approach (e.g. Côté *et al.*, 2004; Levin *et al.*, 2003; Adalid *et al.*, 2005) arbitrarily calibrate the relative probabilities of alternative models being true. They define a robust rule as one that "works well" across several (though not necessarily all) models. However, without accounting for how well different models fit the data, it is difficult to assess the value of implementing a rule which performs well in $M-1$ models but poorly in the M^{th} most data-compatible one.

Bayesian model averaging (e.g. Brock *et al.*, 2007a; Cogley and Sargent, 2005; Levine *et al.*, 2012; Binder *et al.*, 2017, 2018) promotes models with good in-sample fit over models with good forecasting performance by using estimated model probabilities. However, modern monetary policy practices among the inflation-targeting countries are forward-looking and rely heavily on forecasts. This is reflected in our approach which uses a forecasting accuracy criterion to pool models. The main contribution of our paper then is to exploit both within-model and across-model uncertainty as in Levine *et al.* (2012) and Cogley *et al.* (2011), but using a forward-looking perspective based on prediction pools, rather than a backward-looking perspective based on Bayesian model averaging.

[4]As Svensson (2000) and Sims (2001) comment, the worst-case outcome is likely to represent a low probability event and, from the Bayesian perspective, it would be inappropriate to design policy heavily conditioned by it. Further Chamberlain (2000) shows the conditions under which a Bayesian and worst-case policymaker would correspond are highly restrictive.

3 Methodology: Designing Robust Rules

We restrict our attention to optimized simple rules, that is to find the optimal parameter values in a Taylor-type monetary policy rule. The main reason is tractability since a Ramsey-optimal policy would involve the complete state vector in the model. In medium-scale estimated DSGE models, like those in our empirical analysis, finding the Ramsey-optimal policy can be a very challenging task numerically. Moreover, a large literature has arrived at a consensus that optimized simple Taylor-type monetary policy rules mimic the Ramsey-optimal policy and they already contain good robustness properties compared with Ramsey-optimal policy.

The goal of the policy maker is to choose the parameters of a Taylor-type monetary policy rule to maximize welfare that are robust to both within- and across-model uncertainty. Suppose the parameters the policy rule are collected in the vector ρ. We use the expected lifetime utility of households

$$\Omega_i(\rho,\psi) = \mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^t U_t(\rho,\psi)\right] \qquad \psi \in \Psi_i \tag{1}$$

in model $\mathcal{M}_i$ as our welfare measure, where β is the discount factor, Ψ_i is the parameter space for $\mathcal{M}_i$ and $U_t(\rho,\psi)$ denotes utility in period t given the vector of estimated parameters $\psi \in \Psi_i$ and policy rule ρ. We allow the parameter space Ψ_i to differ, but require the policy rule ρ to be the same across models.

We use the estimated posterior distribution from the Bayesian estimation of the model to account for within-model uncertainty. For a policy rule ρ, welfare in model $\mathcal{M}_i$ is given by

$$\Omega_i(\rho) = \int_{\Psi_i}\Omega(\psi,\rho)p(\psi|\mathbf{Y}^o_{i,T},\mathcal{M}_i)d\psi \tag{2}$$

where $p(\psi|\mathbf{Y}^o_{i,T},\mathcal{M}_i)$ is the joint posterior probability distribution of the model parameters estimated for model $\mathcal{M}_i$ given observations $\mathbf{Y}^o_{i,T} = \{\mathbf{y}^o_{i,1},\ldots,\mathbf{y}^o_{i,T}\}$. Notice that, unlike BMA, prediction pools do not require the models to have the same vector of observed variables.

We attach weights to each model to account for across-model uncertainty. Given weights $w = \{w_i\}_{i=1}^m$, the policy maker seeks a common rule ρ^* across every model that maximizes

$$\bar{\Omega}(\rho, w) = \sum_{i=1}^{m} w_i \Omega_i(\rho),$$

a welfare measure that incorporates both within- and across-model uncertainty.

The novelty of our paper lies in the way the weights are constructed for the above policy problem. We use forecasting performance as a criterion for assessing the value of different models. Specifically, we follow the procedure of Geweke and Amisano (2012) to form prediction pools where weights are assigned to models on the basis of the accuracy of their k-period ahead forecasts. Unlike in the case of Bayesain model averaging which assumes that one of the models is the true data generating process, prediction pools allow us to consider that all models among a comparative set are misspecified, but they all may be useful at different periods of time.

Let $p(\mathbf{y}_{T+k}^f | \mathbf{Y}_{i,T}^o, \mathcal{M}_i)$ be the k-period ahead predictive density of model $\mathcal{M}_i$ for a vector of model variables $\mathbf{y}_{T+k}^f$ given observations$\mathbf{Y}_{i,T}^o$:

$$p(\mathbf{y}_{T+k}^f | \mathbf{Y}_{i,T}^o, \mathcal{M}_i) = \int_{\Psi_i} p(\mathbf{y}_{T+k}^f | \mathbf{Y}_{i,T}^o, \psi, \mathcal{M}_i) p(\psi | \mathbf{Y}_{i,T}^o, \mathcal{M}_i) d\psi, \tag{3}$$

where $p(\mathbf{y}_{T+k}^f | \mathbf{Y}_{i,T}^o, \psi, \mathcal{M}_i)$ is the density of k-period ahead predictions of the model given a parameter vector $\psi \in \Psi_i$. Notice that we require all models to share the same vector of forecast variables $\mathbf{y}_{T+k}^f$, but not the observables used for estimation. The predictive density characterizes out of sample observations that have not been used to estimate the posterior density of the parameter vector ψ. Furthermore, the predictive density is independent of the parameter vector ψ which we have integrated over using the posterior. As such this provides predictions about future observations that fully incorporate the information regarding within-model uncertainty in the data.

We assess each model using the log predictive score function. Given a sample $\mathbf{Y}_T^f = \{\mathbf{y}_1^f, \ldots, \mathbf{y}_T^f\}$ of forecast variables, the log predictive score of model $\mathcal{M}_i$ is given by

$$LS(\mathbf{Y}_T^f, \mathcal{M}_i) = \sum_{t=h}^{T-K} \sum_{k=1}^{K} \log p(\mathbf{y}_{t+k}^f | \mathbf{Y}_{i,t}^o, \mathcal{M}_i) \tag{4}$$

where $1 \leq h \leq T$ ensures that there are enough observations in the first subsample to estimate the model. The log predictive score function measures the track record of out-of-sample predictive performance of a model.

We use linear prediction pools to assess the predictive performance of a combination of models.[5] Given a sample $\mathbf{Y}_T^f$ and a model pool $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_m\}$, the log predictive score of the pool is given by

$$LS(\mathbf{Y}_T^f, \mathcal{M}) = \sum_{t=h}^{T-K} \sum_{k=1}^{K} \log \left[\sum_{i=1}^{m} w_i p(\mathbf{y}_{t+k}^f | \mathbf{Y}_{i,t}^o, \mathcal{M}_i) \right]; \qquad \sum_{i=1}^{m} w_i = 1; \qquad w_i \geq 0. \tag{5}$$

The log predictive score function measures the out-of-sample predictive performance of a convex linear combination of the models in the pool. The optimal prediction pool has weights chosen such that the log predictive score of the pool is maximized[6]

$$w_i^* = \arg\max_{w_i} LS(\mathbf{Y}_T^f, \mathcal{M}) \tag{6}$$

Before we turn to our empirical analysis to demonstrate the methodology in practice, let us highlight the differences between prediction pools and BMA (Table 2). First, BMA attaches weights to each model based on their marginal data density. These weights can be interpreted as the posterior probability that a given model is the true data generating process. Prediction pools however, assume that all models are misspecified and attach weights to each model by choosing the prediction pool with the best forecasting accuracy out of all possible convex linear combinations of these models. Second, BMA requires all models to have the same set of observable variables while prediction pools require them only to share the same set of forecast variables. Finally, it is unlikely that a single model $\mathcal{M}_i \in \mathcal{M}$ will consistently produce the best forecasts. Thus, non-zero weights are typically assigned to several models since there will be less tendency for one model to dominate all the others (some $w_i^* \to 1$) as in the case of BMA.

[5] Del Negro *et al.* (2016) use the terminology static pools to reflect the fact that weights are time invariant.

[6] Logs are used in general since they make the densities globally concave, making the maximization easier.

Table 1: BMA versus Optimal Pooling

BMA	**Prediction Pools**
Attaches weights to each model based on their marginal data density.	Attach weights to each model by choosing the prediction pool with the best forecasting accuracy.
Assumes a complete model space - one of the models is the true DGP.	Assumes an incomplete model space - all models are misspecified.
Same set of observable variables	Same set of forecast variables only
Tendency to assign all weights to a single model	Less of a tendency that a single model dominates

4 Models

To illustrate our method, we investigate the welfare consequences of alternative monetary policy rules using three medium-scale new Keynesian DSGE models. The first is the Smets and Wouters (2007) model which is the workhorse model in policy-making institutions for forecasting and policy analysis. The other two models build financial frictions into the SW model along the lines of Gertler *et al.* (2012) and Bernanke *et al.* (1999), respectively. These two models represent the leading theories in modelling financial frictions in the macroeconomic literature. Hence, our model pool can be motivated by considering a policy maker who is uncertain how to incorporate financial frictions into a DSGE model or if they should be incorporated at all.

We estimate the models with Bayesian methods. For all three models we use the same seven time series as observable variables as in Smets and Wouters (2007): the log difference of real GDP, real consumption, real investment and the real wage, log hours worked, the log difference of the GDP deflator, and the federal funds rate. We use US quarterly data over the sample period 1966:1-2017:4.[7] The parameter prior and posterior distributions are reported in Table D1 in the Appendix, while further details on the estimation procedure are given in Appendix D.

We calibrate several parameters in the estimation procedure that are hard to identify in the models. We estimate the posterior distribution of the remaining 26 parameters, which are common in all three models. The reason that all three models have the same parameter vector estimated is to accommodate BMA in our exercise. BMA requires the

[7] We use the same sample as Smets and Wouters (2007) extended to the last full year observations are available for. We have also estimated the model with a sample that stops just before the financial crisis, but parameter estimates are robust to the sample size.

models to share the same set of observable variables, but the SW model has no empirical implications for financial variables. Since parameters specific to the banking sector in the GKQ or the BGG model are hard to identify without including financial variables among the observable variables, we had to fix those parameters and estimate only parameters common in all three models.

4.1 The workhorse New Keynesian model

Our first model follows closely Smets and Wouters (2007). It is a stochastic neoclassical growth model augmented with price and nominal wage stickiness, price and nominal wage indexation, habit persistence and investment adjustment costs. Household welfare in the model is defined by their expected lifetime utility

$$\Omega = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \frac{[C_t - \chi C_{t-1}]^{1-\sigma_c}}{1-\sigma_c} \exp\left[(\sigma_c - 1) \frac{H_t^{1+\psi}}{1+\psi} \right] \right] \tag{7}$$

where C_t is real consumption, H_t is hours supplied, β is the discount factor, χ controls habit formation, σ_c is the inverse of the elasticity of intertemporal substitution (for constant labour), and ψ is the inverse of the Frisch labour supply elasticity. The monetary policy rule for the nominal interest rate $R_{n,t}$ in the model is given by the Taylor-type rule

$$\begin{aligned} \log\left(\frac{R_{n,t}}{R_n}\right) &= \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) \\ &+ (1-\rho_r)\left(\theta_\pi \log\left(\frac{\Pi_t}{\Pi}\right) + \theta_y \log\left(\frac{Y_t}{Y}\right) + \theta_{dy} \log\left(\frac{Y_t}{Y_{t-1}}\right)\right) + MPS_t, \end{aligned} \tag{8}$$

where Π_t is inflation, Y_t is real output, and MPS_t is a monetary policy shock that follows an AR(1) process. The rest of the model is described in detail in Appendix A.

We review here the estimates for the parameters that are important for our policy problem; the estimated posterior distribution for the rest of the parameters can be found in Table D1 in the Appendix. The estimated Calvo parameter for price setting implies that prices are updated on average every 2.07 quarters, but price indexation is rather weak given that the indexation parameter is estimated to be lower than the prior mean. Nominal wages are updated more frequently, on average every 1.67 quarters, but they are also indexed more strongly to inflation. The Taylor rule parameters are very close to the original estimates of Smets and Wouters (2007) and correspond to the typical estimates in the DSGE literature.

4.2 The banking model with outside equity

Our second model extends the SW model with a banking sector along the lines of Gertler *et al.* (2012). In the model banks raise deposits and issue outside equity to finance loans to firms. To motivate an endogenous constraint on the bank's ability to raise funds, Gertler *et al.* (2012) assume that the banker managing the bank may transfer a Θ fraction of assets to their family. Hence, only $1 - \Theta$ fraction of assets can be pledged as collateral. Recognizing this possibility, risk averse households limit the funds they lend to banks. Our setup for the banking sector follows closely Gertler *et al.* (2012), and embeds in our SW model in a similar fashion to Gertler and Karadi (2011). For a detailed description of the model see Appendix B.

The posterior distributions of the estimated parameters are close to the estimates in the SW model. The only exceptions that are important for our policy problem are the parameters characterizing intermediate firms' price setting behaviour. Prices are updated less frequently, on average every 2.91 quarters, but they are indexed to inflation more weakly than in the SW model.

4.3 The financial accelerator model

Our third model extends the SW model with a banking sector along the lines of Bernanke *et al.* (1999). In the model banks collect deposits and lend to entrepreneurs who are subject to idiosyncratic shocks that affect their ability to repay their loans. The financial market friction in this model, which is between the entrepreneur and the bank, is driven by private information. Banks pool loans to protect themselves against credit risk and charge a spread over the deposit rate. The household welfare function and the Taylor-rule is the same as in the SW model. For a detailed description of the model see Appendix C.

The posterior distributions of the estimated parameters resemble the estimates of the SW model, but there a few notable exceptions that are important for our policy problem. Prices are even less flexible as in the GK model and updated on average every 3.34 quarters.

Nominal wages are more flexible and are updated on average every 1.76 quarters.

5 Model Weights

We estimate our models repeatedly with an increasing window of data, and compute log predictive scores (4) and (5) for predictions made by our estimated models. Each estimation sample starts at 1966:1. The first sample ends at 1970:4 ($h = 20$). We assess our models based on how well they predict all seven observable variables jointly up to four quarters ahead ($K = 4$).[8] We increase the sample size by four quarters each time and repeat the same steps.[9] Our last sample ends at 2016:4 ($T = 208$) as to allow for the computation of predictive densities using data up to 2017:4.

The top panel of Figure 1 shows the log predictive score function for each model. These are similar across models throughout most of our sample period, indicating that the predictive performance of all models is similar. There are exceptions to this at various periods, for example from the mid 1990's until the mid 2000's (where the financial frictions models dominate the SW model), and in the years around 1980 (where the SW model does particularly well relative to the financial frictions models). Importantly, we employ prediction pooling to aggregate the relative predictive performance differences over time.

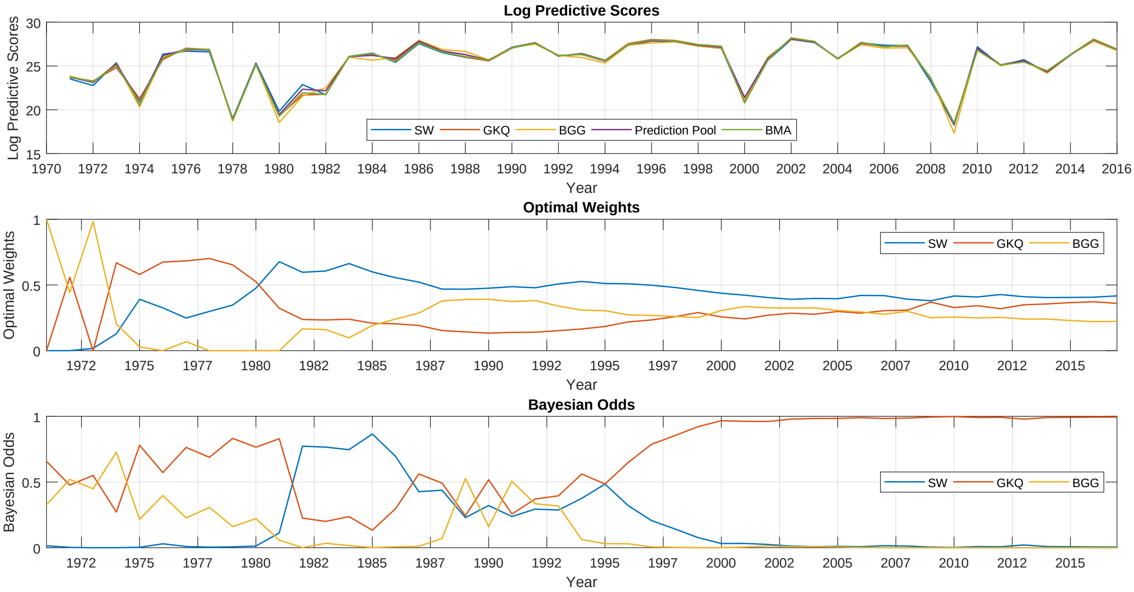
The middle panel of Figure 1 shows the optimal prediction pool weights over the sample period 1970:4-2017:4. To obtain these weights we solve the optimization problem (6) recursively. At each point in time we use the log predictive scores up to that point to determine the weights as if our full sample ended there. As is clear from the figure, the BGG model predicts the best in the earlier part of the sample while in the later part of the sample a larger weight is assigned to the SW model. Weights of approximately 42, 36 and 22 percent are assigned to the SW, BGG and GKQ models, respectively, by the end of the sample period.

The bottom panel of Figure 1 shows an interesting contrast to the middle panel. It shows how the Bayesian odds evolve over our sample, given a uniform prior belief of the policy maker over the competing models. While for most of the sample period the GKQ

[8] We modify Dynare's estimation routine to calculate the predictive densities.

[9] We reestimate the models only every four quarters to reduce the computational complexity of the task. This way we need to estimate each model only 47 times, and our forecasting periods do not overlap each other.

Figure 1: Pooling figure

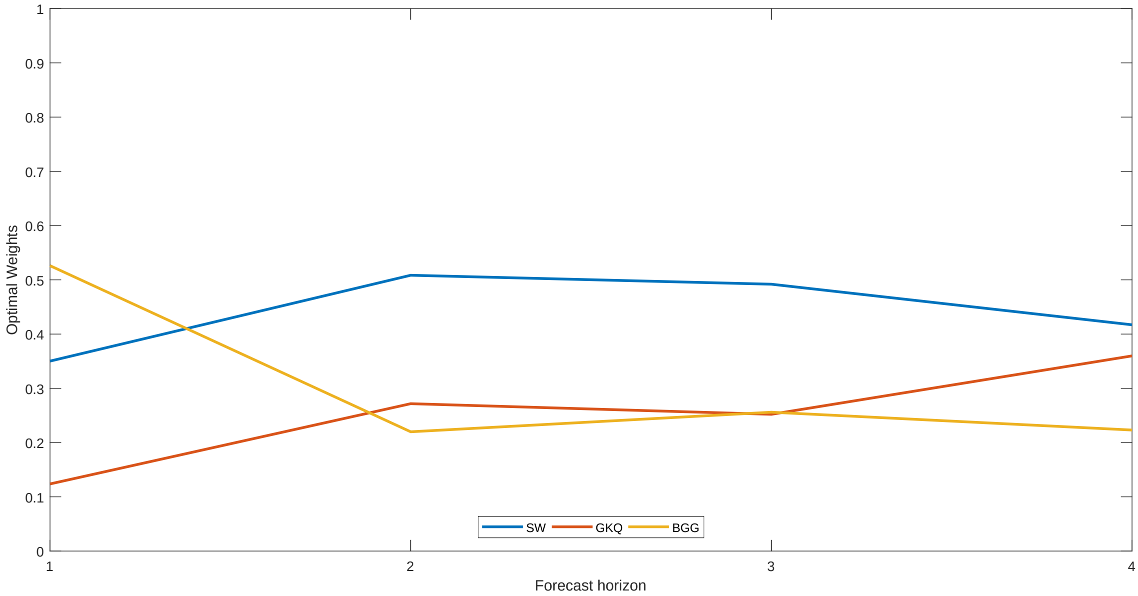


model provides the best fit and gets the highest odds, it receives much lower weight in the optimal prediction pool. Had the policy maker used BMA to attach weights to the models, he would have put most of her faith in the GKQ model while ignoring the other two models entirely. In fact, with the exception of the years in the early 1990's, BMA have the tendency to assign almost zero weight to at least one model in our model pool. Moreover, the optimal prediction pool weights seem to change slowly over time while large changes in Bayesian odds can be brought by adding only a handful of observations to the sample.

The difference in rankings across models between prediction pooling and BMA is due to the standard trade-off between in-sample fit and out-of-sample predictive performance. The financial frictions models fit the in-sample data quite well relative to the SW model (and therefore have higher marginal log-likelihoods) yet they tend to predict poorly (especially in non-crises periods) in comparison to the SW model as they tend to over-fit the data. As a result, the prediction pools assign a significant weight to the SW model. Nevertheless, the situation is reversed at particular times (e.g. crisis times) when model complexity may be beneficial for forecasting and therefore prediction pools should also attach significant weight to the financial frictions models.

Figure 2 shows how the weights assigned to the SW, BGG and GKQ models at the

Figure 2: Optimal prediction pool weights at the end of our sample period using different forecast horizons



end of our sample period depend on the forecast horizons used. Irrespective of the forecast horizon used, all three models are useful in terms of predictability and ought to be employed by a policy maker when designing policy even though the forecasting performance of the GKQ and BGG models are inferior to the SW model at most horizons.

Overall, our results suggest that if the policy maker does not believe any model to be correct, she must nevertheless be much more cautious in rejecting models that appear to be inferior. The intuition here is that such models while not particular accurate most of the time, may well be very useful some of the time, and thus the optimal policy should take this into consideration rather than rejecting the use of such models. Moreover, prediction pooling is a far more robust method for attaching weight to models as these are more robust to outliers and therefore evolve more smoothly over time than corresponding weights from BMA.

6 Optimal Robust Rules

This section analyses the optimal simple rules for our model economies. First, we define the different types of optimal simple rules we compute. Then, we examine optimal simple rules designed for individual models. Then, we proceed to study optimal simple rules using different ways to attach weights to models. Finally, we examine the welfare cost of suboptimal policy using two different sets of criteria: i) the cost of implementing the rule designed for model i in the environment of model $k \neq i$ and ii) the cost of implementing a rule with deviations from the optimal parameter values.

6.1 Computation and Welfare Measures

We seek policy parameters $\rho = \begin{bmatrix} \rho_r & \alpha_\pi & \alpha_y & \alpha_{dy} \end{bmatrix}$ in the Taylor-rule

$$\log\left(\frac{R_{n,t}}{R_n}\right) = \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) + \alpha_\pi \log\left(\frac{\Pi_t}{\Pi}\right) + \alpha_y \log\left(\frac{Y_t}{Y}\right) + \alpha_{dy} \log\left(\frac{Y_t}{Y_{t-1}}\right) \tag{9}$$

that maximize the unconditional lifetime utility of households. We estimate the model using the same form of the Taylor-rule (8) and the same priors on its parameters as in Smets and Wouters (2007). However, we reparametrize the feedback coefficients by setting $\alpha_\pi = (1-\rho_r)\theta_\pi$, $\alpha_y = (1-\rho_r)\theta_y$, and $\alpha_{dy} = (1-\rho_r)\theta_{dy}$ to allow for the possibility of a price level rule ($\rho_r = 1$) when computing optimized simple rules.

We evaluate policies by computing the welfare in model i associated with a particular policy rule. Let

$$\Omega_i(\rho, \psi_i) = \mathbb{E}_0\left[\sum_{t=0}^{\infty} \beta^t U(C_t(\rho, \psi_i), H_t(\rho, \psi_i))\right] \qquad \psi_i \in \Psi_i \tag{10}$$

denote the unconditional lifetime utility of households from a second-order approximation around a deterministic steady state in model $\mathcal{M}_i$ conditional on a particular rule ρ and a parameter vector ψ_i. Then given weights $\{w_i\}_{i=1}^m$, the policy maker seeks a rule ρ^* common to every model in the pool that maximizes

$$\bar{\Omega}(\rho, w) = \sum_{i=1}^{m} w_i \int_{\Psi_i} \Omega_i(\rho, \psi_i) p(\psi_i|\mathbf{Y}_T^o, \mathcal{M}_i) d\psi_i \approx \sum_{i=1}^{m} w_i \sum_{j=1}^{N} \Omega_i(\rho, \psi_{i,j}). \tag{11}$$

where $\psi_{i,j}$ is a draw from parameter space Ψ_i. The optimization problem defined by maximizing (11) can incorporate both a *within-model* robustness, by averaging of N draws for each model $\mathcal{M}_i$ and an *across-model* robustness by using a weighted average of model specific welfare measures. Calculation of (11) is numerically facilitated by MCMC methods that allow the sampling of a given distribution by simulating an appropriately-constructed

Markov chain. Thus, the integral can be approximated by a sum.

We compute two types of optimal simple rules which we will refer to as robust optimal simple rules and mean optimal simple rules. Robust optimal simple rules incorporate both within- and across-model uncertainty. For each model i, we draw $N = 500$ parameter vectors $\psi_{i,j}, j \in [1, N]$ from the estimated posterior distributions of each model to approximate the integral above. Mean optimal simple rules on the other hand are based on a single parameter vector ($N = 1$). They are computed by fixing the parameter vector at the estimated posterior mean of each model. Comparing the two types of rules allows us to assess what happens if the policy maker ignores within-model uncertainty in the policy design.

We compare alternative policies in terms of consumption equivalent welfare changes. Consider two alternative policies ρ_1 and ρ_2. The consumption equivalent welfare cost of adopting ρ_2 in model $\mathcal{M}_i$ is the fraction ω^i of the consumption stream households are willing to give up to be as well off under ρ_2 as under ρ_1. It is implicitly defined by the indifference condition

$$\Omega_i(\rho_2, \psi_i) = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t U((1 - \omega^i)C_t(\rho_1, \psi_i), H_t(\rho_1, \psi_i)) \right] \tag{12}$$

This welfare measure allows us to compare the effect of the same policy in different model economies directly or aggregate the effect of the policy across models. To compute the welfare cost of a policy rule common to all models in a pool, we calculate the weighted average of the associated welfare cost from each model using the weights assigned to each model

$$\omega = \sum_{i=1}^{3} w_i \omega_i \tag{13}$$

6.2 Optimal Within-model Robust Rules

Table 2 compares three policy rules and their outcomes for each of the three models in our prediction pool. The first is the estimated rule which are very similar across models. The second is the robust optimal simple rule in the 'Robust' column and the third in the 'Mean' column is the mean optimal simple rule.

Both types of optimal simple rules are considerably different from the estimated rule. The sub-optimality of the estimated rules is well-established in the literature. Our consumption-equivalent measure ω^i indicates that using either the robust or the mean optimized rules provides a welfare gain over the estimated rule ranging from around 0.05 percent for the GKQ model to 0.10 percent for the SW model.

The two types of optimal simple rules are very similar. The small difference between them suggests that ignoring within-model uncertainty has only a small impact. The difference between the welfare gains of the robust and the mean optimal simple rules over the estimated rules shows that if the policy maker ignores within-model uncertainty, then she underestimates the gains from optimal policy for each model. Moreover, when we compare the robust and the mean optimal simple rules to each other directly we find that the welfare cost associated with the mean optimal simple rule is about ten-thousandth of one percent of consumption.[10] Comparing the welfare gains associated with the robust and mean optimal simple rules reported in Table 2 directly is misleading. The welfare gain in the 'Mean' columns shows the fraction of the consumption stream households need to be given to be indifferent between the mean optimal simple rule and the rule fixed at the estimated posterior mean. The welfare gain of the robust optimal simple rule, on the other hand, is computed as the average welfare gain of the robust optimal simple rule across the $N = 500$ parameter vector draws from the estimated posterior distributions of each model. Hence, the differences in the welfare gains reported in Table 2 reflect the effect of within-model uncertainty on the estimated rule and not on the optimal rule.[11]

Across-model differences of optimal rules are considerable. The optimal simple rules, both robust and mean, designed for the SW model are close to a pure inflation targeting rule. They have almost zero interest rate smoothing parameter ρ_r but see a very aggressive immediate response to changes in inflation captured by the feedback parameter α_π.[12] Consequently, the welfare gains are closely related to the large decrease in inflation volatility implied by the optimal rules: the standard deviation of inflation decreases by a factor of three from 0.00584 to 0.00187. By contrast the optimal rules for the GKQ and BGG models have considerable interest rate smoothing but far more muted response

[10]The welfare cost associated with the mean optimal simple rule is 0.16, 0.54 and 1.87 ten-thousandth of one percent of consumption in the SW, GKW and BGG models, respectively.

[11]This difference is due to Jensen's inequality given that the welfare function is concave.

[12]We do not impose an upper bound on the policy parameter α_π during optimization. Imposing the upper bound $\alpha_\pi \leq 3$ as in Schmitt-Grohe and Uribe (2007) for example would reduce the welfare gain for the SW model slightly, but would be binding for this optimal simple rule only throughout our paper.

to inflation. The welfare gains from decreased inflation volatility, which works through agents' inflation expectations, are much smaller.

The lack of interest rate smoothing in the optimal policy for the SW model results in a higher nominal interest rate volatility than under the estimated rule. An important consequence of the wider interest rate distribution is that the unconditional probability of hitting the zero lower bound p_{ZLB} is much higher than with the estimated policy rule– over 9 percent per quarter or about once every 3 years.[13] This contrasts with about 2 percent per quarter for the optimal rules in the GKQ and BGG models, or about once every 12.5 years. Available data suggest that zero lower bound episodes are rare but long-lived (Dordal-i-Carreras *et al.*, 2016a). The U.S. post-WWII experience (seven years at the zero lower bound over seventy years) indicates that unconditional probabilities below 10 percent are empirically plausible. However, a policy maker may want to set an upper bound on this probability as part of the policy mix.

6.3 Optimal Across-model Robust Rules

We now turn to optimal across-model optimal rules. We compare two sets of policy rules computed using two different set of pooling weights. The first uses BMA and the second we call 'Prediction Pool' uses optimal prediction pool weights.[14] Table 3 shows the results for robust optimal simple rules which incorporate both across-model and within-model uncertainty. The mean optimal simple rules, which are only across-model robust with parameters in each model set at their estimated posterior means and are very similar to the robust simple rules, are reported in Table E3 in the Appendix.

The contribution of within-model robustness, just like in the case of individual models in the previous section, seems small regardless of which sets of weights we use to account for across-model uncertainty. Comparing the welfare gains from robust and mean optimal simple rules shows again that the policy maker underestimates the gains from optimal policy if within-model uncertainty is ignored. But the resulting optimal rules are very

[13]The unconditional probability of hitting the zero lower bound is computed from a normal approximation of the gross nominal interest rate's ergodic distribution. Let R_N and σ_{R_N} denote the deterministic steady state and the unconditional standard deviation of the gross nominal interest rate, respectively, computed from a second-order approximation around a deterministic steady state. Then the probability of hitting the zero lower bound, p_{ZLB} is given by the probability that the gross nominal interest rate is below one in the normal distribution $\mathcal{N}(R_N, \sigma_{R_N})$.

[14]We have also experimented with an equally weighted pool, but the results are very close to those obtained using the prediction pool weights.

Table 2: Optimized simple rules

	SW			GKQ			BGG		
	Estimated	Robust	Mean	Estimated	Robust	Mean	Estimated	Robust	Mean
ρ_r	0.668	0.001	0.001	0.665	0.707	0.704	0.706	0.797	0.782
α_π	0.663	8.621	8.195	0.653	1.176	1.156	0.565	1.270	1.170
α_y	0.002	0.042	0.048	0.002	0.003	0.004	-0.002	0.004	0.004
α_{dy}	0.060	1.062	0.985	0.060	0.188	0.178	0.064	0.156	0.124
$\omega^i(\%)$	0.000	0.102	0.096	0.000	0.050	0.041	0.000	0.082	0.066
$\sigma_{R_N}(\times 100)$	0.695	1.008	1.016	0.788	0.652	0.659	0.757	0.640	0.665
$\sigma_\pi(\times 100)$	0.584	0.187	0.192	0.620	0.351	0.356	0.599	0.281	0.300
σ_Y	0.178	0.178	0.169	0.179	0.178	0.170	0.135	0.137	0.132
$\sigma_{dy}(\times 100)$	0.862	0.832	0.830	0.835	0.808	0.803	0.841	0.841	0.839
$p_{ZLB}(\%)$	2.818	9.274	9.486	4.573	2.102	2.165	3.962	1.938	2.267

Note: The simple rules in the 'Robust' column are optimal robust simple rules based on 500 draws from the estimated posterior distribution of each model. Every variable in the column is computed for each draw then averaged across draws. The simple rules in the 'Mean' column are optimal simple rules computed at the estimated posterior mean of each model. ω_i shows the fraction of the consumption stream households need to give up to have the same welfare as under the estimated rule. σ_{R_N}, σ_Π, σ_Y, and σ_{dY} denote the unconditional standard deviation of the gross nominal interest rate, the inflation rate, output and output growth computed from a second-order approximation around a deterministic steady state. Some of the standard deviations are scaled by 100 for ease of presentation. p_{ZLB} is the probability that the gross nominal interest rate is below one in the normal distribution $\mathcal{N}(R_N, \sigma_{R_N})$, where R_N is the deterministic steady state of the nominal interest rate.

similar in Tables 3 and E3 and ignoring within-model uncertainty has only a small impact. A direct comparison of the robust and mean optimal simple rules to each other reveals that that the welfare cost associated with the mean optimal simple rule is again about ten-thousandth of one percent of consumption.[15]

The choice of pooling weights matters for the optimized rule. The optimal simple rule obtained using the optimal prediction pool weights implies a higher welfare gain over the estimated rule compared to the policy rules obtained using the BMA weights. This higher welfare gain is explained by a combination of two factors: differences in i) the weights and ii) in the rules. The first is a composition effect since the weights used in (13) to aggregate the welfare gains from the individual models are different. If we aggregated the welfare gains of the prediction pool optimal rule from the individual models using the BMA weights, then the aggregate welfare gain would be 0.045 only. Hence, differences in

[15]The welfare cost associated with the mean optimal simple rule is 0.49 and 0.90 ten-thousandth of one percent of consumption using the BMA and prediction pool weights, respectively.

the optimal rules would imply a higher difference in welfare gains were the weights the same.

The second factor is the differences in the rules themselves. Using the optimal prediction pool weights implies a more persistent policy rule that also responds more aggressively to changes in inflation and output growth compared to the policy rules obtained using the BMA weights. The welfare gains from the individual models in the ω^i row reveal that the prediction pool rule trades-off welfare gains from the GKQ model for welfare gains from the SW model given that it attaches more weight to the latter. Interestingly, inflation volatility is reduced in all three models the most by the prediction pool optimal rule, but the likely welfare gains from this reduced volatility are partly negated by the interest rate smoothing parameter being above its optimal level in all three models.

Finally, across-model robustness removes the high probability of hitting the nominal interest rate lower bound seen in Table 2 for the SW model with only within-model robustness. Since the optimal values for the interest rate smoothing parameter are well above the estimated values regardless of the weights we use, interest rate volatility decreases for every optimal policy and so does the probability of hitting the zero lower bound.

6.4 The Welfare Cost of Suboptimal Policy

In this section we examine the welfare cost of suboptimal policy using two different sets of criteria. First, we quantify the cost of implementing the rule designed for model i in the environment of model $j \neq i$. Then, we analyse the cost of implementing a rule with deviations from the optimal parameter values.

Table 4 shows the welfare cost of using a rule optimized for a specific model in another model. This is a counterfactual exercise that shows the cost of incorrectly identifying the data generating process. For example, if we use the robust simple rule optimized for the SW model in the BGG model, then the welfare loss is 0.052 percent of consumption relative to the robust simple rule optimized for the BGG model itself. The results confirm that that the robust simple rule optimized for the GKQ and BGG models are very similar, and they both are very different from the robust simple rule optimized for the SW model. Using the rule optimized for the GKQ model in the BGG model, or vice versa, imply very small welfare losses. Using the robust simple rule optimized for the SW model in the other

Table 3: Optimized robust simple rules

	BMA			Prediction pool		
	NK	GK	BGG	NK	GK	BGG
$w(\%)$	1.72	98.27	0.00	41.71	35.98	22.31
ρ_r		0.708			0.800	
α_π		1.185			1.634	
α_y		0.003			0.005	
α_{dy}		0.189			0.269	
$\omega^i(\%)$	0.068	0.050	0.079	0.083	0.044	0.080
$\omega(\%)$		0.050			0.068	
$\sigma_{R_N}(\times 100)$	0.617	0.651	0.687	0.597	0.628	0.646
$\sigma_\pi(\times 100)$	0.360	0.348	0.333	0.286	0.277	0.261
σ_Y	0.178	0.178	0.136	0.178	0.178	0.137
$\sigma_{dy}(\times 100)$	0.815	0.808	0.819	0.810	0.809	0.828
$p^i_{ZLB}(\%)$	1.612	2.090	2.676	1.356	1.751	2.027
$p_{ZLB}(\%)$		2.082			1.648	

Note: The robust optimal simple rules in the table are based on 500 draws from the estimated posterior distribution of each model. Every variable in the table is computed for each draw then averaged across draws. ω_i shows the fraction of the consumption stream households need to give up to have the same welfare as under the estimated rule in ach model. σ_{R_N}, σ_Π, σ_Y, and σ_{dY} denote the unconditional standard deviation of the gross nominal interest rate, the inflation rate, output and output growth computed from a second-order approximation around a deterministic steady state. Some of the standard deviations are scaled by 100 for ease of presentation. p_{ZLB} is the probability that the gross nominal interest rate is below one in the normal distribution $\mathcal{N}(R_N, \sigma_{R_N})$, where R_N is the deterministic steady state of the nominal interest rate.

two models, however, implies much larger welfare losses than the other way around. This explains why the robust simple rule optimized for the prediction pool is so close to the rule optimized for the GKQ and BGG models. The final row shows the welfare cost of using the robust rule optimized for the prediction pool in Table 3 relative to the model specific robust optimal rules reported in Table 2. These now avoid the large costs of the single model optimized rules and the costs are generally small relative to the gains from using optimal rules.

Figure 3 shows how sensitive welfare is to changes in the coefficients in the robust optimal simple rule in each model. Each panel in the graph shows the consumption equivalent variation in welfare when changing a single parameter in the robust optimal simple rule at a time in a given model. Each column shows the welfare consequences

Table 4: Welfare cost of robust optimal policy i in model $j \neq i$

	SW	GKQ	BGG
SW	0.000	0.140	0.052
GKQ	0.035	0.000	0.003
BGG	0.028	0.002	0.000
Pool	0.019	0.006	0.001

Note: The table shows what happens when an optimal simple rules optimized for model i is used in model $j \neq i$. The first column shows the consumption equivalent welfare loss in the SW model relative to the welfare attained using the robust simple rule optimized for the SW model if, for example, we use the robust simple rules optimized for the SW, GKQ, BGG models, respectively. The last row shows the welfare cost incurred in model i when instead of using the robust simple rule optimized for model i we use the robust optimal simple rule obtained with the optimal prediction pool weights.

of changing a parameter in the Taylor rule, while each row shows the welfare costs of deviating from the optimal parameter values in a given model.

In the SW model welfare is quite flat around the optimal point along every dimension. Deviations have small effects and welfare is most sensitive to changes in the output feedback parameter α_y.[16] In the financial friction models deviations from the optimal value α_y still dominate the policy maker's objective: deviations can have large welfare consequences that can reach as much as a 5% consumption equivalent. However, welfare in those models is also very sensitive to deviations of the interest rate smoothing parameter, ρ_r, from its optimal value. The panels in the last row show the welfare cost of deviating from the optimal parameter values of the common robust rule in all three models at the same time optimized for the prediction pool. These panels are essentially weighted averages of the three panels above them using the weights assigned to each model in the prediction pool. The numbers here are considerably smaller than in the second and third rows due to the fact that the SW model receives the largest weight in the prediction pool.

6.5 A Price Level Rule

Our results then show that even small deviations of ρ_r and α_y have serious welfare consequences with losses far greater than any gains relative to the estimated rule reported so far. So suppose that the monetary policymaker commits to a rule with $\rho_r = 1$ and

[16] Schmitt-Grohe and Uribe (2007) termed this result the 'importance of not responding to output'.

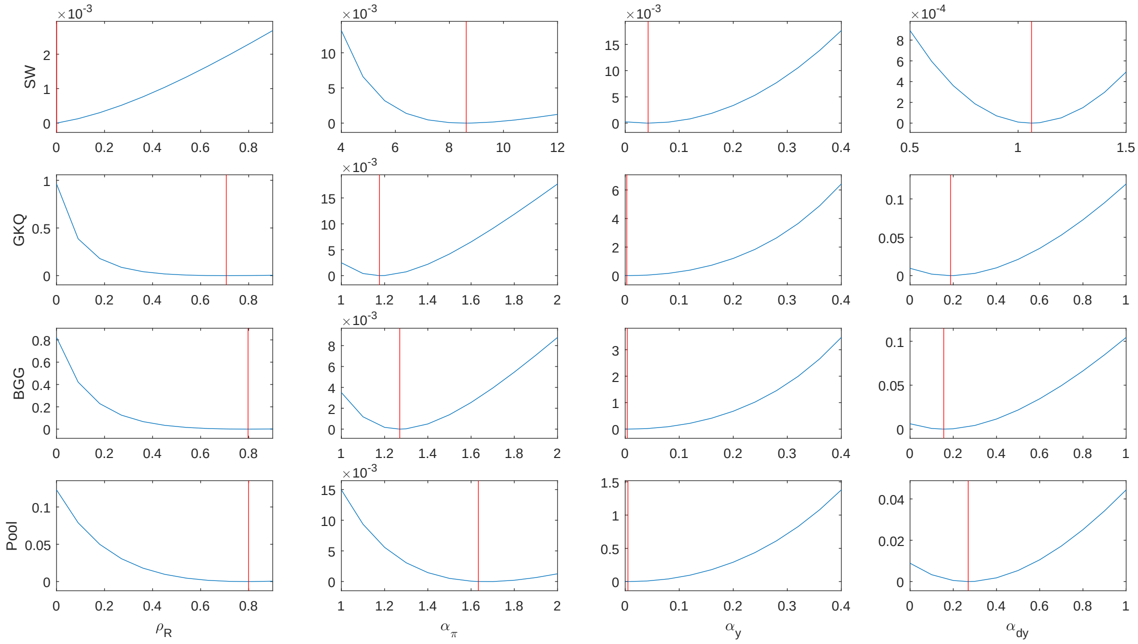


Figure 3: The cost of bad policy I

Note: The graph shows the welfare cost of deviating from the optimal Taylor-rule parameters in the optimal robust rule. Each column shows the welfare cost of changing a single parameter in the Taylor-rule while keeping all other parameters constant. The first three rows show the welfare cost of deviating from the robust rule optimized for that particular model. The last row shows the welfare cost of deviating from the common robust rule in all three models at the same time optimized for the prediction pool. Welfare costs are measured in consumption equivalent welfare changes. The vertical red line shows the optimal value of the parameter that is changed in each column.

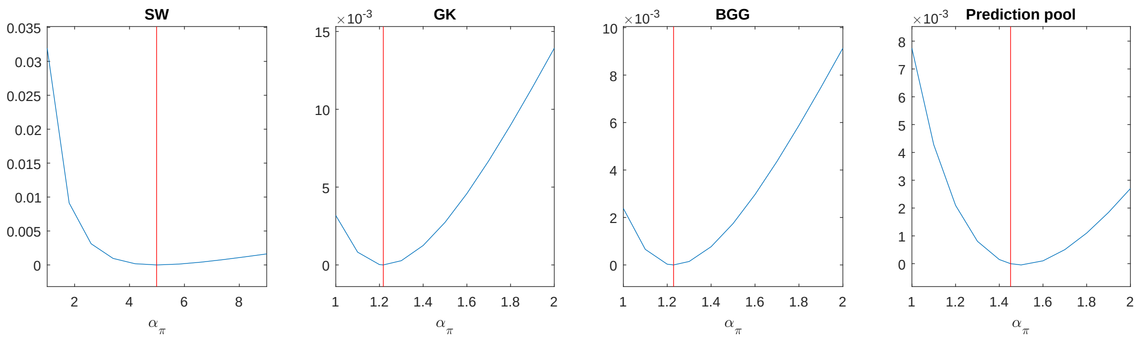


Figure 4: The cost of bad policy II

Note: The graph shows the welfare cost of deviating from the optimal value of α_π in in the optimal robust price level rule. Price level rules use the restriction $\rho_r = 1$ and $\alpha_y = \alpha_{dy} = 0$. The first three panel show the welfare cost of deviating from the robust price level rule optimized for that particular model. The last panel shows the welfare cost of deviating from the common robust price level rule in all three models at the same time optimized for the prediction pool. Welfare costs are measured in consumption equivalent welfare changes. The vertical red line shows the optimal value of the parameter that is changed in each column.

$\alpha_y = \alpha_{dy} = 0$. Then integrating (9) and putting $\frac{\Pi_t}{\Pi} = \frac{P_t/P_{t-1}}{\bar{P}_t/\bar{P}_{t-1}}$ where $\bar{P}_t$ is the price trend in the constant inflation rate steady state, we arrive at the rule

$$\frac{R_{n,t}}{R_n} = \alpha_\pi \frac{P_t}{\bar{P}_t} \tag{14}$$

which is a *price-level rule* that adjusts the deviation of the nominal interest rate to changes in the price level relative to its long-run trend. The benefits of price-level targeting versus inflation targeting have been studied in the literature now for some time. (See, for example, Svensson (1999), Schmitt-Grohe and Uribe (2000), Vestin (2006), Gaspar *et al.* (2010), Giannoni (2014)). These papers examine the good determinacy and stability properties of price-level targeting. Holden (2016) shows these benefits extend to a ZLB setting. The intuition for the benefits of price-targeting is as follows: faced with of an unexpected temporary rise in inflation, price-level stabilization commits the policymaker to bring inflation below the target in subsequent periods. In contrast, with inflation targeting, the drift in the price level is accepted.

Our results in Table 5 and Figure 4 indicate a further benefit of price-level targeting: when the robust rule is implemented, even with departures from its optimal setting of the single feedback parameter α_π that defines the policy, it remains robust across models with and without financial frictions.

Table 5: Optimized simple robust price level rules

	SW	GKQ	BGG	BMA			Prediction Pool		
				SW	GKQ	BGG	SW	GKQ	BGG
ρ_r	1.000	1.000	1.000		1.000			1.000	
α_π	4.988	1.217	1.229		1.223			1.452	
α_y	0.000	0.000	0.000		0.000			0.000	
α_{dy}	0.000	0.000	0.000		0.000			0.000	
$\omega^i(\%)$	0.092	0.022	0.064	0.075	0.021	0.063	0.077	0.020	0.063
$\omega(\%)$					0.023			0.053	
$\sigma_{R_N}(\times 100)$	0.738	0.535	0.565	0.520	0.553	0.582	0.526	0.559	0.588
$\sigma_\pi(\times 100)$	0.116	0.269	0.232	0.257	0.246	0.213	0.249	0.239	0.207
σ_Y	0.179	0.180	0.142	0.179	0.180	0.143	0.179	0.181	0.143
$\sigma_{dy}(\times 100)$	0.902	0.875	0.919	0.876	0.876	0.924	0.876	0.876	0.926
$p_{ZLB}(\%)$	3.630	0.689	0.983	0.577	0.854	1.166	0.626	0.918	1.240

Note: Simple price level rules use the restriction $\rho_r = 1$ and $\alpha_y = \alpha_{dy} = 0$ and only the value of α_π is optimized. All rules in the table are based on 500 draws from the estimated posterior distribution of each model. Every variable in the column is computed for each draw then averaged across draws. ω_i shows the fraction of the consumption stream households need to give up to have the same welfare as under the estimated rule in each model. σ_{R_N}, σ_Π, σ_Y, and σ_{dY} denote the unconditional standard deviation of the gross nominal interest rate, the inflation rate, output and output growth computed from a second-order approximation around a deterministic steady state. Some of the standard deviations are scaled by 100 for ease of presentation. p_{ZLB} is the probability that the gross nominal interest rate is below one in the normal distribution $\mathcal{N}(R_N, \sigma_{R_N})$, where R_N is the deterministic steady state of the nominal interest rate.

7 Conclusion

This paper studies the problem of designing robust simple rules when the policy maker has at her disposal a finite set of models, none of which are believed to be the true data generating process. We assign weights to models on the basis of the accuracy of their 4-period ahead forecasts rather than their in-sample fit, consistent with the forward-looking viewpoint of the policy maker. We study the robust optimal policy problem in the form of an optimized Taylor-type nominal interest rate rule under this weighting scheme using three estimated models exemplifying the policy makers' uncertainty about how to incorporate financial frictions into the canonical DSGE model of Smets and Wouters (2007). In comparison with Bayesian model averaging, we find that our prediction pool choice has a significant impact on the robust optimized rule.

Our approach provides a very general framework for the combination of models in a policy design problem. It only requires models to share the same policy instrument, to provide a k-period ahead predictive density given macro-economic data, and to have a welfare criterion to rank alternative policies. The models in the pool do not need to share the estimated parameter vector, nor even the observables; they can be nested as well as non-nested. Thus, the methodology can be applied to a wide range of macroeconomic models from mainstream DSGE, behavioural to agent-based, and indeed to other non-macroeconomic settings as long as these three requirements are met.

Several open questions offer possible avenues for further research. First, our policy problem can be extended to accommodate macro-prudential instruments where banks are subjected to a common capital, or equivalently leverage-ratio, requirement.

Second, whereas the results we have obtained in the paper examine the possibility of a hitting a zero-lower bound for monetary policy, they are not designed to minimize this outcome as in Levine *et al.* (2012). In a recent paper Dordal-i-Carreras *et al.* (2016b) demonstrate that a regime switching representation of risk premium shocks can generate a realistic distribution of zero-lower bound durations in a standard New Keynesian model. Incorporating zero-lower bound considerations into our policy problem may reveal policy trade-offs that are not present in our results.

Third, we have confined our study to one standard form of interest rate rule responding with interest rate inertia to inflation, output and output growth. This nested a price level rule which many papers in the literature have found to have good robustness properties. Other rules, such as wage inflation targeting, need to be explored.

Fourth an aspect of the robustness approach, largely ignored by the literature, is the scope for expectation differences between private and public sectors. There is a need to consider the possibility that the private sector may also believe there are competing models. By assuming that each model has RE with model-consistent expectations our application has ruled out this possibility. This case of *model-inconsistent expectations* needs to be factored into truly robust Bayesian rules. An alternative is to pursue a behavioural approach and drop the RE assumption altogether.[17]

Fifth, the use of real-time data when analysing the out-of-sample forecast performances of competing models to compute weights would be an interesting exercise. An application

[17]Candidates for such an exercise can be found in the survey Calvert Jump and Levine (2019).

of density forecasting using real time data, but without model pooling, is provided by McAdam and Warne (2019). Another important technical issue that may arise is that the state space over which the two models are stable may not coincide. This is a problem as then in both the determination of the optimal pool and the robust optimal policy, one model may be favoured over the other.

References

Adalid, R., Coenen, G., McAdam, P., and Siviero, S. (2005). The Performance and Robustness of Interest-Rate Rules in Models of the Euro Area. *International Journal of Central Banking*, **1**(1), 95–132.

Bernanke, B., Gertler, M., and Gilchrist, S. (1999). The Financial Accelerator in Quantitative Business Cycles. In M. Woodford and J. B. Taylor, editors, *Handbook of Macroeconomics, vol. 1C*. Amsterdam: Elsevier Science.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory*. Wiley, Chichester, UK.

Binder, M., Lieberknecht, P., Quintana, J., and Wieland, V. (2017). Model Uncertainty in Macroeconomics: On the Implications of Financial Frictions. C.E.P.R. Discussion Papers.

Binder, M., Lieberknecht, P., Quintana, J., and Wieland, V. (2018). Robust Macroprudential Policy Rules under Model Uncertainty. Annual Conference 2018, German Economic Association.

Brock, W. A., Durlauf, S. N., and West, K. D. (2007a). Model uncertainty and policy evaluation: Some theory and empirics. *Journal of Econometrics*, **127**(2), 629–664.

Brock, W. A., Durlauf, S. N., Nason, J. M., and Rondina, G. (2007b). Simple versus optimal rules as guides to policy. *Journal of Monetary Economics*, **54**(5), 1372–1396.

Calvert Jump, R. and Levine, P. (2019). Behavioural New Keynesian Models. *Journal of Macroeconomics*, **59**(C), 59–77.

Calvo, G. (1983). Staggered Prices in a Utility-Maximizing Framework. *Journal of Monetary Economics*, **12**(3), 383–398.

Chamberlain, G. (2000). Econometric Applications of Maximin Expected Utility. *Journal of Applied Econometrics*, **15**, 625–644.

Christiano, L. L., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, **113**(1), 1–45.

Cogley, T. and Sargent, T. (2005). The conquest of U.S. inflation: learning and robustness to model uncertainty. *Review of Economic Dynamics*, **8**(2), 528–563.

Cogley, T., De Paoli, B., Matthes, C., Nikolov, K., and Yates, T. (2011). A Bayesian approach to optimal monetary policy with parameter and model uncertainty. *Journal of Economic Dynamics and Control*, **35**(12), 2186–2212.

Côté, D., Kuszczak, J., Lam, J.-P., Liu, Y., and St-Amant, P. (2004). The Performance and Robustness of Simple Monetary Policy Rules in Models of the Canadian Economy. *Canadian Journal of Economics*, **37**(4), 978–998.

Del Negro, M., Hasegawa, R., and Schorfheide, F. (2016). Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance. *Journal of Econometrics*, **192**, 391–405.

Dennis, R., Leitemo, K., and Soderstrom, U. (2009). Methods for Robust Control. *Journal of Economic Dynamics and Control*, **33**(8), 1604–1616.

Dordal-i-Carreras, M., Coibion, O., Gorodnichenko, Y., and Wieland, J. (2016a). Infrequent but Long-Lived Zero-Bound Episodes and the Optimal Rate of Inflation. NBER Working Papers 22510, National Bureau of Economic Research, Inc.

Dordal-i-Carreras, M., Coibion, O., Gorodnichenko, Y., and Wieland, J. (2016b). Infrequent but Long-Lived Zero-Bound Episodes and the Optimal Rate of Inflation. NBER Working Papers 22510, National Bureau of Economic Research, Inc.

Ellison, M. and Sargent, T. J. (2012). A Defense Of The FOMC. *International Economic Review*, **53**(4), 1047–1065.

Gaspar, V., Smets, F., and Vestin, D. (2010). Is Time Ripe for Price Level Path Stability? In P. L. Siklos, M. T. Bohl, and M. E. Wohar, editors, *Challenges in central banking:*

the current institutional environment and forces affecting monetary policy. Cambridge University Press.

Gertler, M. and Karadi, P. (2011). A Model of Unconventional Monetary Policy. *Journal of Monetary Economics*, **58**(1), 17– 34.

Gertler, M., Kiyotaki, N., and Queralto, A. (2012). Financial crises, bank risk exposure and government financial policy. *Journal of Monetary Economics*, **59**(S), 17–34.

Geweke, J. (2010a). *Complete and Incomplete Econometric Models*. Princeton University Press.

Geweke, J. (2010b). *Contemporary Bayesian Econometrics and Statistics*. New Jersey: Princeton University Press.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, **164**(1), 130–141.

Geweke, J. and Amisano, G. (2012). Prediction with misspecified models. *American Economics Review*, **102**(3), 482–486.

Giannoni, M. (2014). Optimal interest-rate rules and inflation stabilization versus price-level stabilization. *Journal of Economic Dynamics and Control*, **41**(C), 110–129.

Gneiting, T. and Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**(477), 359–378.

Hall, S. and Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, **23**(1), 1–13.

Hansen, L. and Sargent, T. J. (2007). *Robustness*. Princeton University Press.

Holden, T. (2016). Existence and uniqueness of solutions to dynamic models with occasionally binding constraints. EconStor Preprints 130142, ZBW - German National Library of Economics.

Ilbas, P., Roisland, O., and Sveen, T. (2012). Robustifying Optimal Monetary Policy using Simple Rules as Cross-checks. Working Paper 2012/22, Norges Bank.

Kapetanios, G., Fawcett, N., Mitchell, J., and Price, S. (2015). Generalised density forecast combinations. *Journal of Econometrics*, **188**(1), 150–165.

Levin, A., Wieland, V., and Williams, J. C. (2003). The Performance of Inflation Forecast-Based Rules Under Model Uncertainty. *American Economic Review*, **93**, 622–45.

Levin, A. T. and Williams, J. C. (2003). Robust Monetary Policy with Competing Reference Models. *Journal of Monetary Economics*, **50**(5), 945–975.

Levine, P. and Currie, D. A. (1987). The design of feedback rules in linear stochastic rational expectations models. *Journal of Economic Dynamics and Control*, **11**, 1–28.

Levine, P. and Pearlman, J. (2010). Robust Monetary Rules under Unstructured and Structured Uncertainty. *Journal of Economic Dynamics and Control*, **34**(3), 456–471.

Levine, P., McAdam, P., and Pearlman, J. (2007). Quantifying and Sustaining Welfare Gains from Monetary Commitment. ECB Working Paper No. 709 and *Journal of Monetary Economics*, 55(7), 2008.

Levine, P., McAdam, P., and Pearlman, J. (2012). Probability Models and Robust Policy Rules. *European Economic Review*, **56**(2), 246 – 262.

McAdam, P. and Warne, A. (2019). Euro area real-time density forecasting with financial or labor market frictions. *Journal of International Forecasting*. Forthcoming.

Orphanides, A. and Williams, J. C. (2007). Robust Monetary Policy with Imperfect Knowledge. *Journal of Monetary Economics*, **54**, 1406–1435.

Orphanides, A. and Williams., J. C. (2008). Learning, Expectations Formation, and the Pitfalls of Optimal Control Monetary Policy. *Journal of Monetary Economics*, **55S**, S80–S96.

Ross, S. M. (2006). *Introduction to Probability Models*. Academic Press, 9th Edition.

Schmitt-Grohe, S. and Uribe, M. (2000). Price level determinacy and monetary policy under a balanced-budget requirement. *Journal of Monetary Economics*, **45**, 211–246.

Schmitt-Grohe, S. and Uribe, M. (2007). Optimal Simple and Implementable Monetary and Fiscal Rules. *Journal of Monetary Economics*, **54**(6), 1702–1725.

Sims, C. A. (2001). Pitfalls of a Minimax Approach to Model Uncertainty. *American Economic Review*, **91**(2), 51–54.

Sims, C. A. (2002). The role of models and probabilities in the monetary policy process. *Brooking Papers on Economic Activity*, **2**, 1–62.

Sims, C. A. (2007). Monetary Policy Rules. CEPS Working Paper No. 155.

Sims, C. A. (2008). Improving Monetary Policy Models. *Journal of Economic Dynamics and Control*, **32**(8), 2460–2475.

Smets, F. and Wouters, R. (2007). Shocks and Frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, **97**(3), 586–606.

Svensson, L. E. (1999). Price Level Targeting Versus Inflation Targeting. *Journal of Money, Credit and Banking*, **31**, 277–295.

Svensson, L. E. O. (2000). Robust control made simple. Unpublished manuscript, Princeton University.

Taylor, J. B. and Williams, J. C. (2010). Simple and Robust Rules for Monetary Policy. In B. M. Friedman and M. Woodford, editors, *Handbook of Monetary Economics*, volume 3 of *Handbook of Monetary Economics*, chapter 15, pages 829–859. Elsevier.

Tetlow, R. J. (2015). Real-Time Model Uncertainty in the United States: "Robust" Policies Put to the Test. *International Journal of Central Banking*, **11**(1), 113–154.

Townsend, R. M. (1979). Optimal Contracts and Competitive Markets with Costly State Verification. *Journal of Economic Theory*, **21**(2), 265 – 293.

Vehtari, A. and Ojanen, J. (2012). A survey of bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, **6**, 142–228.

Vestin, D. (2006). Price Level Targeting Versus Inflation Targeting. *Journal of Monetary Economics*, **53**, 1361–1376.

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

How should a forward-looking policy maker conduct monetary policy when she has a finite set of models at her disposal, none of which are believed to be the true data generating process? In our approach, the policy maker first assigns weights to models based on relative forecasting performance rather than in-sample fit, consistent with her forward-looking objective. These weights are then used to solve a policy design problem that selects the optimized Taylor-type interest-rate rule that is robust to model uncertainty across a set of well-established DSGE models with and without financial frictions. We find that the choice of weights has a significant impact on the robust optimized rule which is more inertial and aggressive than either the non-robust single model counterparts or the optimal robust rule based on backward-looking weights as in the common alternative Bayesian Model Averaging. Importantly, we show that a price-level rule has excellent welfare and robustness properties, and therefore should be viewed as a key instrument for policy makers facing uncertainty over the nature of financial frictions.How should a forward-looking policy maker conduct monetary policy when she has a finite set of models at her disposal, none of which are believed to be the true data generating process? In our approach, the policy maker first assigns weights to models based on relative forecasting performance rather than in-sample fit, consistent with her forward-looking objective. These weights are then used to solve a policy design problem that selects the optimized Taylor-type interest-rate rule that is robust to model uncertainty across a set of well-established DSGE models with and without financial frictions. We find that the choice of weights has a significant impact on the robust optimized rule which is more inertial and aggressive than either the non-robust single model counterparts or the optimal robust rule based on backward-looking weights as in the common alternative Bayesian Model Averaging. Importantly, we show that a price-level rule has excellent welfare and robustness properties, and therefore should be viewed as a key instrument for policy makers facing uncertainty over the nature of financial frictions.

Appendices

A The SW Model

Final good producers The representative final good producer uses a continuum of intermediate goods $Y_t(m)$ to produce a homogeneous final good

$$Y_t = \left(\int_0^1 Y_t(m)^{(\zeta_p - 1)/\zeta_p} dm \right)^{\zeta_p/(\zeta_p - 1)} \tag{A.1}$$

Final goods producers are perfectly competitive and choose output to maximize profits $P_t Y_t - \int_0^1 P_t(m) Y_t(m) dm$, where $P_t(m)$ is the price of intermediate good m and P_t is the aggregate price index. This implies the standard demand function

$$Y_t(m) = \left(\frac{P_t(m)}{P_t} \right)^{-\zeta_p} Y_t. \tag{A.2}$$

Intermediate good producers Intermediate good m is produced using the technology

$$Y_t(m) = (A_t H_t^d(m))^{\alpha} K_{t-1}(m)^{1-\alpha} \tag{A.3}$$

where $H_t^d(m)$ and $K_t(m)$ are labour and capital demand, respectively. Labour productivity is decomposed into a cyclical component and a deterministic trend $A_t = \bar{A}_t A_t^c$, where the cyclical component follows the process

$$\log A_t^c - \log A^c = \rho_A (\log A_{t-1}^c - \log A^c) + \epsilon_{A,t}. \tag{A.4}$$

Intermediate producer m maximizes profits $P_t(m) Y_t(m) - W_t H_t^d(m) - r_t^K K_{t-1}(m)$, where W_t and r_t^K are the real wage and return on capital, respectively. This implies the marginal cost

$$MC_t = \alpha^{-\alpha} (1-\alpha)^{-(1-\alpha)} W_t^{1-\alpha} (r_t^K)^{\alpha} \tag{A.5}$$

which is the same for all intermediate producers.

Following Calvo (1983) we assume that each period the price of retail good m is set optimally to $P_t^0(m)$ with probability $1-\xi_p$. If the price is not re-optimized, then prices are indexed to last period's aggregate inflation, with indexation parameter γ_p. Each retail producer m chooses $P_t^0(m)$ to maximize discounted profits

$$\mathbb{E}_t \sum_{k=0}^{\infty} \xi_p^k \Lambda_{t,t+k} Y_{t+k}(m) \left[\frac{P_t^0(m)}{P_{t+k}} \left(\frac{P_{t+k-1}}{P_{t-1}} \right)^{\gamma_p} - MC_{t+k} \right] \tag{A.6}$$

subject to demand (A.2) given the stochastic discount factor $\Lambda_{t,t+k}$ of the households.

The solution to the above problem is the first-order condition

$$\mathbb{E}_t \sum_{k=0}^{\infty} \xi_p^k \Lambda_{t,t+k} Y_{t+k}(m) \left[\frac{P_t^0(m)}{P_{t+k}} \left(\frac{P_{t+k-1}}{P_{t-1}} \right)^{\gamma_p} - \frac{1}{(1-1/\zeta_p)} MC_{t+k} MS_{t+k} \right] = 0 \tag{A.7}$$

where we have introduced a mark-up shock MS_t that follows the process

$$\log MS_t - \log MS = \rho_{MS}(\log MS_{t-1} - \log MS) + \epsilon_{MS,t}. \tag{A.8}$$

Note that this is a small departure from the original SW model where the elasticity of substitution ζ_p is the sum of its long-run mean and a mark-up shock the follows an AR(1) process; i.e. the shock is hitting the elasticity of substitution directly.

Labour markets Households supply their homogeneous labour to trade unions that differentiate the labour services. A labour packer buys the differentiated labour from the trade unions and aggregate them into a composite labour using the Dixit-Stigliz aggregator $H_t = \left(\int_0^1 H_t(j)^{(\zeta_w-1)/\zeta_w} dj \right)^{\zeta_w/(\zeta_w-1)}$, where H_t is aggregate labour supply, ζ_w is the elasticity of substitution among different types of labour, and we index trade unions by j. The labour packer minimizes the cost $\int_0^1 W_{n,t}(j) H_t(j) dj$ of producing the composite labour service, where $W_{n,t}(j)$ denote the nominal wage set by union j. This leads to the standard demand function

$$H_t(j) = \left(\frac{W_{n,t}(j)}{W_{n,t}} \right)^{-\zeta_w} H_t^d \tag{A.9}$$

where $W_{n,t}$ is the aggregate nominal wage.

Sticky wages are introduced through Calvo contracts supplemented with indexation. At each period there is a probability $1-\xi_w$ that trade union j can choose $W_{n,t}^O(j)$ to maximize

$$\mathbb{E}_t \sum_{k=0}^{\infty} \xi_w^k \Lambda_{t,t+k} H_{t+k}(j) \left[\frac{W_{n,t}^O(j)}{P_{t+k}} \left(\frac{P_{t+k-1}}{P_{t-1}} \right)^{\gamma_w} - W_{h,t+k} \right] \tag{A.10}$$

subject to the demand function (A.9), where γ_w is the indexation parameter.

The solution to the above problem is the first-order condition

$$\mathbb{E}_t \sum_{k=0}^{\infty} \xi_p^k \Lambda_{t,t+k} H_{t+k}(j) \left[\frac{W_{n,t}^O(j)}{P_{t+k}} \left(\frac{P_{t+k-1}}{P_{t-1}} \right)^{\gamma_w} - \frac{1}{(1 - 1/\zeta_w)} W_{h,t+k} \right] = 0 \tag{A.11}$$

where we have introduced a mark-up shock $MRSS_t$ that follows the process

$$\log MRSS_t - \log MRSS = \rho_{MRSS}(\log MRSS_{t-1} - \log MRSS) + \epsilon_{MRSS,t} \tag{A.12}$$

Note that this is a small departure from the original SW model where the elasticity of substitution ζ_w is the sum of its long-run mean and a mark-up shock the follows an AR(1) process; i.e. the shock is hitting the elasticity of substitution directly.

Households Household j maximizes its expected lifetime utility

$$\Omega(j) = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \frac{[C_t(j) - \chi C_{t-1}(j)]^{1-\sigma_c}}{1 - \sigma_c} \exp \left[(\sigma_c - 1) \frac{H_t(j)^{1+\psi}}{1 + \psi} \right] \right] \tag{A.13}$$

where C_t is real consumption, H_t is hours supplied, β is the discount factor, χ controls habit formation, σ_c is the **NEW: inverse of** the elasticity of intertemporal substitution (for constant labour), and ψ is the inverse of the Frisch labour supply elasticity. Note that, unlike in the original SW model, we use internal instead of external habit formation.

The household's budget constraint in period t is given by

$$C_t(j) + I_t(j) + \frac{B_t(j)}{RPS_t R_{n,t} P_t} + T_t = \frac{B_{t-1}(j)}{P_t} + r_t^K K_{t-1}(j) + W_{h,t} H_t(j) + \Gamma_t \tag{A.14}$$

where I_t is investment into physical capital, B_t is government bonds held at the end of period t, $R_{n,t-1}$ is the nominal interest rate paid on government bonds held at the beginning of period t, RPS_t is an exogenous premium in the return on bonds that follows the process

$$\log RPS_t - \log RPS = \rho_{RPS}(\log RPS_{t-1} - \log RPS) + \epsilon_{RPS,t}, \tag{A.15}$$

T_t is lump-sum taxes, r_t^K is the real rental rate, $W_{h,t}$ is the real wage rate, and Γ_t is profits of intermediate firms distributed to households. Notice that we deviate from the original SW model and do not allow for variable capital utilization in the model. The capital stock K_t accumulates according to

$$K_t(j) = (1-\delta)K_{t-1}(j) + (1 - S(X_t(j)))I_t(j)IS_t \tag{A.16}$$

where IS_t is an investment specific technological shock that follows the process

$$\log IS_t - \log IS = \rho_{IS}(\log IS_{t-1} - \log IS) + \epsilon_{IS,t}, \tag{A.17}$$

$X_t(j) = I_t(j)/I_{t-1}(j)$ is the growth rate of investment, and $S(\cdot)$ is an adjustment cost function such that $S(X) = 0$, $S'(X) = 0$, and $S''(\cdot) = 0$ where X is the steady state value of investment growth.

The solution to the household's problem imply the arbitrage condition

$$\mathbb{E}_t[\Lambda_{t,t+1}R_{t+1}] = \mathbb{E}_t[\Lambda_{t,t+1}R_{t+1}^K] = 1. \tag{A.18}$$

where

$$R_t = \frac{R_{n,t-1}}{\Pi_t} \qquad\qquad R_t^K = \frac{[r_t^K + (1-\delta)Q_t]}{Q_{t-1}}$$

are the real gross returns on government bonds and physical capital, and Q_t is the price of capital (Tobin's q). The modelling strategy when we introduce financial frictions into the model will be to break the above no arbitrage condition and introduce a wedge between the two returns.

The Government Problem A monetary policy rule for the nominal interest rate is given by the Taylor-type rule

$$\begin{aligned}\log\left(\frac{R_{n,t}}{R_n}\right) &= \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) \\ &\quad + (1-\rho_r)\left(\theta_\pi \log\left(\frac{\Pi_t}{\Pi}\right) + \theta_y \log\left(\frac{Y_t}{Y}\right) + \theta_{dy}\log\left(\frac{Y_t}{Y_{t-1}}\right)\right) + MPS_t,\end{aligned} \tag{A.19}$$

where Π_t is inflation, and MPS_t is a monetary policy shock that follows the process

$$\log MPS_t - \log MPS = \rho_{MPS}(\log MPS_{t-1} - \log MPS) + \epsilon_{MPS,t} \tag{A.20}$$

The government budget constraint is

$$P_t G_t + B_{t-1} = T_t + \frac{B_t}{R_{n,t}} \tag{A.21}$$

where G_t is government spending that follow the process

$$\log G_t - \log G = \rho_G(\log G_{t-1} - \log G) + \epsilon_{G,t} \tag{A.22}$$

B The GKQ Model

The GKQ model extends the SW model with a banking sector to introduces a wedge between the expected ex ante cost of loans from households, R_t and the return on capital r_t^K. Our setup follows closely Gertler *et al.* (2012) but embeds in our SW model in a similar fashion to Gertler and Karadi (2011). Apart from the arbitrage condition (A.18), the details of the model are unchanged, so we concentrate here on the banking sector only.

The financial market friction in this model is driven by the costs of enforcing contracts. Financial frictions affect real activity via the impact on funds available to banks but there is no friction in transferring funds between banks and nonfinancial firms. Given a certain deposit level a bank can lend frictionlessly to nonfinancial firms against their future profits. In this regard, firms offer to banks a perfect state-contingent security.

Banks The flow of funds constraint of an individual bank is given by

$$Q_t s_t = n_t + q_t e_t + d_t \tag{B.23}$$

where s_t denotes claims on non-financial firms to finance capital acquired at the end of period t for use in period $t+1$, and Q_t is the price of a unit of capital. Therefore $Q_t s_t$ is the value of loans funded in period t, which equals the sum of bank net worth n_t, household deposits d_t and outside equity raised from households $q_t e_t$.

The net worth of the bank accumulates according to

$$n_t = R_t^K Q_{t-1} s_{t-1} - R_t d_{t-1} - R_{e,t} q_{t-1} e_{t-1} \tag{B.24}$$

where the real return equity is given by $R_{e,t} = [r_t^K + (1-\delta)q_t]/q_{t-1}$.

Banks exit with probability $1-\sigma_B$ per period and pay dividends only when they exit. The banker's objective is to maximize expected discounted terminal wealth

$$V_t = \mathbb{E}_t \sum_{i=1}^{\infty} (1-\sigma_B)\sigma_B^{i-1} \Lambda_{t,t+i} n_{t+i} \tag{B.25}$$

subject to an incentive constraint for households to be willing to supply funds to the banker.

To motivate an endogenous constraint on the bank's ability to obtain funds, we introduce the following simple agency problem. We assume that after a bank obtains funds, the bank's manager may transfer a fraction of assets to her family. In recognition of this possibility, households limit the funds they lend to banks.

Divertible assets consists of total gross assets $Q_t s_t$. If a bank diverts assets for its personal gain, it defaults on its debt and shuts down. The creditors may re-claim the remaining fraction $1-\Theta$ of funds. Because its creditors recognize the bank's incentive to divert funds, they will restrict the amount they lend. In this way a borrowing constraint may arise. In order to ensure that bankers do not divert funds the following incentive constraint must therefore hold:

$$V_t \geq \Theta(x_t) Q_t s_t \tag{B.26}$$

where $x_t \equiv \frac{q_t e_t}{Q_t s_t}$ is the fraction of bank assets financed by outside equity, $\Theta_t' >, \Theta_t'' > 0$ captures the idea that it is easier to divert assets funded by outside equity than by households. As before, the incentive constraint states that for households to be willing to supply funds to a bank, the bank's franchise value V_t must be at least as large as its gain from diverting funds.

The optimization problem for the bank is to choose a path for loans, $\{s_{t+i}\}$ to maximize V_t subject to (B.23) and (B.24).

Aggregation At the aggregate level the banking sector balance sheet is:

$$Q_t S_t = N_t + q_t E_t + D_t$$

At the aggregate level net worth is the sum of existing (old) bankers and new bankers:

$$N_t = N_{o,t} + N_{n,t}$$

Net worth of existing bankers equals earnings on assets held in the previous period net cost of deposit finance, multiplied by a fraction σ_B, the probability that they survive until the current period:

$$N_{o,t} = \sigma_B \{(r_t^K + (1-\delta)Q_t)S_{t-1} - (r_t^K + (1-\delta)q_t)E_{t-1} - R_t D_{t-1}\}$$

Since new bankers cannot operate without any net worth, we assume that the family transfers to each one the fraction $\xi_B/(1-\sigma_B)$ of the total value assets of exiting bankers. This implies:

$$N_{n,t} = \xi_B [r_t^K + (1-\delta)Q_t] S_{t-1} \tag{B.27}$$

Links to the SW model In the absence of credit policy by the authorities the model is closed by household arbitrage conditions

$$\mathbb{E}_t[\Lambda_{t,t+1} R_{t+1}] = \mathbb{E}_t[\Lambda_{t,t+1} R_{e,t+1}] = 1. \tag{B.28}$$

Moreover, the market for loans clears implying that $S_t = K_t$.

C The BGG Model

In a 'costly state verification model' due originally to Townsend (1979), the modelling strategy is once again to replace $\mathbb{E}_t[\Lambda_{t,t+1} R_{t+1}] = \mathbb{E}_t[\Lambda_{t,t+1} R_{t+1}^K]$ with a wedge that arises from the friction between a risk neutral entrepreneur and a financial intermediary. The former borrows from the latter to purchase capital from capital producers at a price Q_t and combines it with labour to produce an intermediate good. In order to ensure they cannot grow out of the financial constraint, entrepreneurs exit with probability σ_E. As we shall see this setup introduces a wedge between the expected ex post (non-riskless) rate, $\mathbb{E}_t[R_{t+1}]$ and the expected return on capital $\mathbb{E}_t[R^K_{t+1}]$.

The entrepreneur seeks loans l_t to bridge the gap between its net worth $n_{E,t}$ and the expenditure on new capital $Q_t k_t$, all end-of-period. Thus

$$l_t = Q_t k_t - n_{E,t} \tag{C.29}$$

where the entrepreneur's *real* net worth accumulates according to

$$n_{E,t} = R^K_t Q_{t-1} k_{t-1} - \frac{R_{l,t-1}}{\Pi_t} l_{t-1}$$

where $R_{l,t}$ is the nominal loan rate to be decided in the contract.

In each period an idiosyncratic capital quality shock, ψ_t results in a return $R^K_t \psi_t$ which is the entrepreneur's private information. Default in period $t+1$ occurs when net worth becomes negative, i.e., when $n_{E,t+1} < 0$ and shock falls below a threshold $\bar{\psi}_{t+1}$ given by

$$\bar{\psi}_{t+1} = \frac{R_{l,t} l_t}{\Pi_{t+1} R^K_{t+1} Q_t k_t} \tag{C.30}$$

With the idiosyncratic shock, ψ_t drawn from a density $f(\psi_t)$ with a lower bound ψ_{min}, the probability of default is then given by

$$p(\bar{\psi}) = \int_{\psi_{min}}^{\bar{\psi}_{t+1}} f(\psi) d\psi$$

In the event of default the bank receives the assets of the firm and pays a proportion μ of monitoring costs to observe the realized return. Otherwise the bank receives the full payment on its loans, $R_{l,t} l_t / \Pi_{t+1}$ where $R_{l,t}$ is the agreed loan rate at time t.

At the heart of the model is the bank's *incentive compatibility (IC) constraint* given by

$$\mathbb{E}_t \left[R^K_{t+1} Q_t k_t \left((1-\mu) \int_{\psi_{min}}^{\bar{\psi}_{t+1}} \psi f(\psi) d\psi + \bar{\psi}_{t+1}(1 - p(\bar{\psi}_{t+1})) \right) \geq R_{t+1} l_t \right] \tag{C.31}$$

The LHS of (C.31) is the expected return to the bank from the contract averaged over all realizations of the shock, the RHS is the return from a riskless bond. To be incentive compatible, the expected return from the contract must be equal or greater than the intermediary's opportunity cost, which is the rate R_{t+1}.

Now define $\Gamma(\bar{\psi}_{t+1})$ to be the expected fraction of net capital received by the lender (the bank) and $\mu G(\bar{\psi}_{t+1})$ to be expected monitoring costs where

$$\Gamma(\bar{\psi}_{t+1}) \equiv \int_{\psi_{min}}^{\bar{\psi}_{t+1}} \psi f(\psi) d\psi + \bar{\psi}_{t+1}(1 - p(\bar{\psi}_{t+1})) \tag{C.32}$$

$$G(\bar{\psi}_{t+1}) \equiv \int_{\psi_{min}}^{\bar{\psi}_{t+1}} \psi f(\psi) d\psi \tag{C.33}$$

Then the optimal contract for the risk neutral entrepreneur solves

$$\max_{\bar{\psi}_{t+1}, k_t} \mathbb{E}_t \left[\left(1 - \Gamma(\bar{\psi}_{t+1})\right) R^K_{t+1} Q_t k_t \right]$$

given initial net worth $n_{E,t}$, subject to the IC constraint (C.31).

So far we have set out the optimizing decision of the representative entrepreneur. We now aggregate assuming that entrepreneurs exit with fixed probability $1 - \sigma_E$. To allow new entrants start up we assume exiting entrepreneurs transfer a proportion ξ_E of their wealth to new entrants. Aggregate net worth then accumulates according to

$$N_{E,t} = (\sigma_E + \xi_E)(1 - \Gamma(\bar{\psi}_t)) R^K_t Q_{t-1} K_{t-1}$$

and on exiting the entrepreneur consumes

$$C_{E,t} = (1 - \sigma_E)(1 - \xi_E)(1 - \Gamma(\bar{\psi}_t)) R^K_t Q_{t-1} K_{t-1}$$

The resource constraint becomes

$$Y_t = C_t + C_{E,t} + G_t + I_t + \mu G(\bar{\psi}_t)) R^K_t Q_{t-1} K_{t-1}$$

D Bayesian Estimation

We estimate the three models in our pool with Bayesian methods. For all three models we use the same seven time series as observable variables as in Smets and Wouters (2007): the log difference of real GDP, real consumption, real investment and the real wage, log hours worked, the log difference of the GDP deflator, and the federal funds rate. The corresponding measurement equations are

$$\begin{bmatrix} dlGDP_t \\ dlCONS_t \\ dlINV_t \\ dlWAG_t \\ lHOURS_t \\ dlP_t \\ FEDFUNDS_t \end{bmatrix} = \begin{bmatrix} \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{l} \\ \bar{\pi} \\ \bar{r} \end{bmatrix} + \begin{bmatrix} y_t - y_{t-1} \\ c_t - c_{t-1} \\ i_t - t_{t-1} \\ w_t - w_{t-1} \\ l_t \\ \pi_t \\ r_t \end{bmatrix} \tag{D.34}$$

where $\bar{\gamma}$ is the common quarterly trend growth rate of real GDP, consumption, investment and wages, π is the steady state inflation rate, $\bar{l}$ is steady-state hours worked and $\bar{r}$ is the steady state nominal interest rate. The trend growth rate and all three steady state values entering the measurement equations are not estimated, but set to their corresponding sample means.

We fix several parameters in the estimation procedure that are hard to identify in the model. The steady state value of the government spending-GDP ratio is set to 18 percent. The depreciation rate is fixed at 2.5 percent per quarter and the output elasticity of labour is set to 0.67. The elasticity of substitution among different types of intermediate goods ζ_p, and among labour types ζ_w are fixed at 7 and 3, respectively.

The parameters of the banking sector in the GKQ model are calibrated the following way. We choose the functional form $\Theta(x_t) = \theta(1+\epsilon x_t + \kappa x_t^2/2)$ for the fraction of divertible funds. We follow Gertler *et al.* (2012) and set $\epsilon = -2$ which, together with the steady state relationship $\Theta'(x) = 0$, implies that we can calibrate κ to imply a target of 0.15 for the steady state outside equity ratio x.[18] We choose the values for θ and ξ_B in order to match an economy wide leverage ratio of four and an average credit spread of 276 basis points per year observed in the data. Finally, we choose $\sigma_B = 0.942$ using micro data from Bankscope, implying that bankers survive on average for 4.3 years.

The financial parameters of the BGG model to calibrate are σ_ψ, σ_E, ξ_E and μ. These four parameters are calibrated jointly to hit the following four targets: a default probability

[18]Note that only the ratio $\frac{\epsilon}{\kappa}$ is pinned down in the deterministic steady state, so ϵ remains undetermined using this calibration strategy.

of $p(\bar{\psi}) = 0.045$, $\rho(\bar{\psi}) = 1.0069$ corresponding to a credit spread of 276 basis points per year as in GKQ, an entrepreneur leverage $\frac{QK}{N_E} = 2$ as in Bernanke *et al.* (1999), and an entrepreneurial consumption to GDP ratio of $\frac{C_E}{Y} = 0.075$.

The remaining 26 parameters, which are common in all three models, are estimated using data over the sample period 1966:1-2017:4.[19] We use the same priors as in Smets and Wouters (2007), which are all reported in Table D1.[20] To compute the log predictive scores we estimate our models repeatedly with an increasing window of data. Each estimation sample starts at 1966:1. The first sample ends at 1970:4 and increase the sample size by four quarters each time. Our last sample ends at 2016:4. We estimate each model a total of 47 times.

Table D1 reports the parameter priors and their estimated posterior distributions and the loglikelihood values. Table D2 compares the moments of our observable variables implied by our model at the estimated posterior mean to the data.

E Additional Results

This sections presents some additional results that are omitted from the main text to conserve space.

Table E3 shows the results for mean optimal simple rules which are only across-model robust with parameters in each model set at their estimated posterior means. We compare three sets of policy rules computed using three different set of pooling weights. The first termed 'Naive' weights the three models equally. The second uses Bayesian Model Averaging (BMA) and the third we call 'Prediction Pool' uses optimal prediction pool weights. These are to be compared with robust optimal simple rules in Table 3 which incorporate both across-model and within-model uncertainty.

Figure 5 shows how sensitive welfare is to changes in the coefficients in the mean optimal simple rule in each model. Each panel in the graph shows the consumption equivalent variation in welfare when changing a single parameter in the mean optimal

[19] We use the same sample as Smets and Wouters (2007) extended to the last full year observations are available for. We have also estimated the model with a sample that stops just before the financial crisis, but parameter estimates are robust to the sample size.

[20] We implement the models in Dynare. Instead of log-lineariznig the equations by hand as Smets and Wouters (2007), we include them in their nonlinear form and let Dynare do the log-linearization. Consequently, we rescaled the priors on the standard deviations of the shock processes accordingly.

Table D1: Parameter prior and posterior distributions

	Prior distribution			SW	GKQ	BGG
Parameter	Density	Mean	Std			
σ_c	Normal	1.500	0.375	1.216 [1.047; 1.390]	1.357 [1.185; 1.524]	1.766 [1.443; 2.087]
ψ	Normal	2.000	0.750	2.440 [1.635; 3.270]	1.858 [1.080; 2.611]	0.986 [0.403; 1.543]
χ	Beta	0.500	0.100	0.397 [0.318; 0.475]	0.422 [0.352; 0.493]	0.356 [0.287; 0.427]
ϕ_X	Normal	2.000	0.750	0.564 [0.341; 0.782]	1.533 [0.889; 2.173]	2.512 [1.595; 3.420]
ξ_p	Beta	0.500	0.100	0.605 [0.547; 0.663]	0.657 [0.610; 0.704]	0.701 [0.646; 0.755]
ξ_w	Beta	0.500	0.100	0.518 [0.450; 0.590]	0.508 [0.429; 0.588]	0.433 [0.326; 0.540]
γ_p	Beta	0.500	0.100	0.357 [0.210; 0.499]	0.277 [0.157; 0.394]	0.310 [0.170; 0.449]
γ_w	Beta	0.500	0.100	0.612 [0.462; 0.761]	0.608 [0.461; 0.757]	0.598 [0.449; 0.752]
ρ_R	Beta	0.750	0.100	0.671 [0.609; 0.732]	0.665 [0.597; 0.733]	0.709 [0.659; 0.759]
θ_π	Normal	1.500	0.250	1.998 [1.771; 2.231]	1.953 [1.727; 2.183]	1.939 [1.711; 2.166]
θ_y	Normal	0.120	0.050	0.005 [-0.015; 0.026]	0.008 [-0.012; 0.027]	-0.007 [-0.032; 0.017]
θ_{dy}	Normal	0.120	0.050	0.186 [0.119; 0.256]	0.180 [0.113; 0.249]	0.217 [0.147; 0.287]
ρ_A	Beta	0.500	0.200	0.978 [0.968; 0.987]	0.987 [0.981; 0.994]	0.976 [0.965; 0.987]
ρ_G	Beta	0.500	0.200	0.961 [0.951; 0.970]	0.962 [0.951; 0.972]	0.965 [0.953; 0.977]
ρ_{MCS}	Beta	0.500	0.200	0.912 [0.881; 0.943]	0.925 [0.892; 0.959]	0.883 [0.835; 0.932]
ρ_{MRSS}	Beta	0.500	0.200	0.968 [0.952; 0.985]	0.967 [0.952; 0.982]	0.954 [0.933; 0.976]
ρ_{MPS}	Beta	0.500	0.200	0.294 [0.173; 0.412]	0.308 [0.188; 0.430]	0.238 [0.132; 0.346]
ρ_{RPS}	Beta	0.500	0.200	0.599 [0.476; 0.720]	0.552 [0.431; 0.674]	0.847 [0.770; 0.927]
ρ_{IS}	Beta	0.500	0.200	0.939 [0.906; 0.973]	0.861 [0.830; 0.892]	0.727 [0.642; 0.817]
σ_A	InvGamma	0.001	0.020	0.009 [0.009; 0.010]	0.009 [0.009; 0.010]	0.009 [0.008; 0.010]
σ_G	InvGamma	0.001	0.020	0.032 [0.030; 0.035]	0.031 [0.028; 0.033]	0.029 [0.027; 0.031]
σ_{MCS}	InvGamma	0.001	0.020	0.015 [0.012; 0.017]	0.016 [0.013; 0.019]	0.020 [0.015; 0.025]
σ_{MRSS}	InvGamma	0.001	0.020	0.034 [0.026; 0.042]	0.030 [0.022; 0.038]	0.023 [0.016; 0.030]
σ_{MPS}	InvGamma	0.001	0.020	0.003 [0.002; 0.003]	0.003 [0.002; 0.003]	0.003 [0.002; 0.003]
σ_{RPS}	InvGamma	0.001	0.020	0.005 [0.003; 0.007]	0.006 [0.004; 0.009]	0.003 [0.001; 0.004]
σ_{IS}	InvGamma	0.001	0.020	0.017 [0.013; 0.020]	0.037 [0.029; 0.045]	0.050 [0.033; 0.066]
Loglik				5294.4	5299.0	5286.7

Note: All estimations are done with Dynare version 4.4.3. A sample of 500,000 draws was created and the first 125,000 draws were discarded. The numbers in the brackets show the 90% HPD intervals.

simple rule at a time in a given model. Each column shows the welfare consequences of changing a parameter in the Taylor rule, while each row shows the welfare costs of deviating from the optimal parameter values in a given model. The results in Figure 4 incorporate only across-model uncertainty, which is compared to the results in Figure 1 which incorporate both across-model and within-model uncertainty.

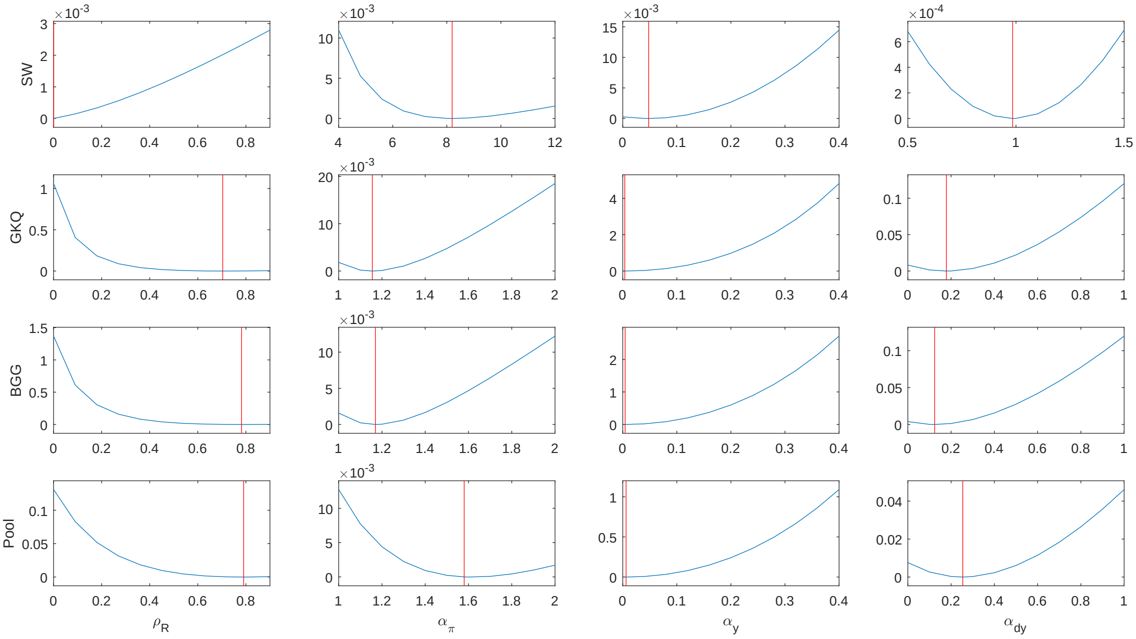


Figure 5: The cost of bad policy

Note: The graph shows the welfare cost of deviating from the optimal Taylor-rule parameters in the optimal rule computed at the estimated posterior mean. Each column shows the welfare cost of changing a single parameter in the Taylor-rule while keeping all other parameters constant. The first three rows show the welfare cost of deviating from the rule optimized for that particular model. The last row shows the welfare cost of deviating from the common rule in all three models at the same time optimized for the prediction pool. Welfare costs are measured in consumption equivalent welfare changes. The vertical red line shows the optimal value of the parameter that is changed in each column.

Table D2: Data and model moments

Observable	Data	SW		GKQ		BGG	
		Mode	Mean	Mode	Mean	Mode	Mean
				Second moments			
$dlGDP_t$	0.8017	0.8505	0.8554	0.8281	0.8274	0.8525	0.8336
$dlCONS_t$	0.6791	0.8120	0.8310	0.8012	0.8218	0.8224	0.8400
$dlINV_t$	2.1127	2.0545	2.0730	2.3075	2.2799	2.1389	2.0979
$dlWAG_t$	0.7531	1.0807	1.1089	1.0515	1.0869	1.0420	1.1002
$lHOURS_t$	3.7669	4.2799	4.2745	4.7095	4.6099	5.1979	4.6158
dlP_t	0.5855	0.5602	0.5743	0.5919	0.6031	0.5693	0.5913
$FEDFUNDS_t$	0.9493	0.6658	0.6966	0.7599	0.7820	0.7278	0.7583
				Correlation with $dlGDP_t$			
$dlGDP_t$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$dlCONS_t$	0.6797	0.5187	0.5140	0.5434	0.5447	0.5906	0.5867
$dlINV_t$	0.6791	0.7228	0.7128	0.6713	0.6555	0.6492	0.6175
$dlWAG_t$	0.0167	0.4726	0.4622	0.4285	0.4185	0.4318	0.4268
$lHOURS_t$	0.1383	0.1225	0.1239	0.1280	0.1281	0.1283	0.1267
dlP_t	-0.1772	-0.1298	-0.1461	-0.0585	-0.0881	-0.1498	-0.2020
$FEDFUNDS_t$	-0.0675	-0.1555	-0.1660	-0.0984	-0.1154	-0.1706	-0.1960

Note: Data moments computed at the estimated posterior mean with second-order approximation using Dynare version 4.4.3. The sample used for computing the data moments is 1966:1-2017:4.

Table E3: Optimized simple rules at the estimated mean

	BMA			Prediction pool		
	NK	GK	BGG	NK	GK	BGG
$w(\%)$	1.72	98.27	0.00	41.71	35.98	22.31
ρ_r		0.704			0.791	
α_π		1.161			1.580	
α_y		0.003			0.006	
α_{dy}		0.178			0.254	
$\omega^i(\%)$	0.061	0.041	0.063	0.076	0.036	0.064
$\omega(\%)$		0.041			0.059	
$\sigma_{R_N}(\times 100)$	0.629	0.659	0.709	0.608	0.635	0.667
$\sigma_\pi(\times 100)$	0.366	0.355	0.345	0.292	0.283	0.270
σ_Y	0.169	0.170	0.130	0.169	0.171	0.131
$\sigma_{dy}(\times 100)$	0.813	0.803	0.815	0.807	0.804	0.824
$p^i_{ZLB}(\%)$	1.710	2.162	3.010	1.420	1.802	2.295
$p_{ZLB}(\%)$		2.154			1.753	

Note: The mean optimal simple rules in the table are computed at the estimated posterior mean of each model. σ_{R_N}, σ_Π, σ_Y, and σ_{dY} denote the unconditional standard deviation of the gross nominal interest rate, the inflation rate, output and output growth computed from a second-order approximation around a deterministic steady state. Some of the standard deviations are scaled by 100 for ease of presentation. p_{ZLB} is the probability that the gross nominal interest rate is below one in the normal distribution $\mathcal{N}(R_N, \sigma_{R_N})$, where R_N is the deterministic steady state of the nominal interest rate.