UNIVERSITY OF SURREY

Discussion Papers in Economics

Negotiating the Wilderness of Bounded Rationality through Robust Policy

By

Szabolcs Deak
(University of Exeter),

Paul Levine
(University of Surrey),

Afrasiab Mirza
(University of Birmingham),

&
Son Pham
(University of Hamburg),

DP 02/23

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

Negotiating the Wilderness of Bounded Rationality through Robust Policy

Szabolcs Deak, University of Exeter, s.deak@exeter.ac.uk

Paul Levine, University of Surrey, p.levine@surrey.ac.uk

Afrasiab Mirza, University of Birmingham, A.Mirza@bham.ac.uk

Son Pham, University of Hamburg, thanh.son.pham@uni-hamburg.de

March 1, 2023

Abstract

We show how the “wilderness of non-rationality" posed for the policymaker may be negotiated by designing a robust Taylor-type monetary rule across a RE NK model and competing behavioural alternatives. The latter consist of a model with “Euler learning" and a bounded rational one with myopia due to Gabaix (2020). For the former expectations of endogenous variables take the form of a general heuristic rule, encompassing simple adaptive expectations, that is supported by an experiment study. This gives four competing NK models, the benchmark one with rational expectations (**model RE**), Euler learning with a simple adaptive expectations heuristic rule (**model EL-SAE**), Euler learning with the general rule (**model EL-GAE**) and the Gabaix bounded rational model (**model BR**). In our novel forward-looking approach, policymakers weight models based on relative forecasting performance rather than Bayesian model averaging. Our main results are: first, three models completely dominate model EL-SAE with weights $w_{RE} = 0.4$, $w_{EL-GAE} = 0.32$ and $w_{EL-BR} = 0.28$. Second, whereas Bayesian model averaging would design a welfare-optimized rule that hits the ZLB with a probability solely based on the Gabaix model, we find that our prediction pool using these weights choice has a significant impact on the robust optimized rule. Third, there are significant differences between the optimized rules for each model separately highlighting the need for seeking a robust rule. Fourth, we find that robust optimized rule found using optimal pooling weights is very close to the price level rule. This confirms good robustness properties of such a rule found in other studies. Finally to achieve a probability of hitting the ZLB constraint on the nominal interest rate of 5% per quarter, the robust optimal rule requires a target (steady-state) net inflation annual rate of between 3% and 4%.

JEL Classification: C11; C18; C32; E32

Keywords: New Keynesian Model, Behavioural Macroeconomics, Optimized Rules, Zero Lower Bound Constraint, Optimal Trend Inflation.

Contents

1 Introduction

It was Sims (1980) who warned of the dangers that leaving the rational expectations (RE) equilibrium concept sends us into a "wilderness". Sargent (2008) followed this up by pointing to a " bewildering variety of ways to imagine discrepancies between objective and subjective distributions and the "infinite number of ways to be wrong, but only one way to be correct." He recommended cautious modifications of rational expectations theories and rational expectations econometrics by virtue of the ways that we allow our adaptive agents to use economic theory, statistics, and dynamic programming.

Models

The challenge posed by the wilderness is clearly demonstrated by the sheer size of literature on behavioural macroeconomics and the huge number of equilibria proposed. Surveys include Evans and Honkapohja (2009), Eusepi and Preston (2016), Branch and McGough (2018) and Calvert Jump and Levine (2019). We focus on a small sub-set of this literature and study three 'cautious' departures from RE NK models where agents are individually rational given expectations. The latter then take the form of adaptive learning, and one model of inattention due to Gabaix (2020). In the first learning model, "**Euler Learning (EL)**", RE $\mathbb{E}_t y_{t+1}$, where y_t is a vector of endogenous variables in the model, is replaced with $\mathbb{E}_t^* y_{t+1}$ which are expectations of the perceived law of motion (of whatever form). In the seminal contribution by Evans and Honkapohja (2001) these expectations take the form of statistical OLS, but subsequent forms of EL model then as heuristic forecasting rules; see, for example, De Grauwe (2011), De Grauwe (2012b) and De Grauwe (2012a).

To formulate possible heuristic rules that encompass those in these papers, we draw upon the general form studied Anufriev et al. (2019) in an experimental setting that takes the log-linear form.

$$E_t^*(y_{t+1}) = [E_{t-1}^*(y_t)]^{1-\lambda_y^1}[y_t]^{\lambda_y^1+\lambda_y^2}[y_{t-1}]^{-\lambda_y^2}, \quad 0 < \lambda_y^1 < 1, \quad -1 < \lambda_y^2 < 1 \tag{1}$$

This encompasses simple adaptive expectations ($\lambda_y^2 = 0$), 'trend extrapolation' ($\lambda_y^1 = 0$), and a 'fundamentalist' rule ($\lambda_y^2 = \lambda_y^1 = 0$) for which $E_t^*(y_{t+1}) = E_{t-1}^*(y_t) =$ the model's steady state. In Anufriev et al. (2019) parameters λ_y^1, λ_y^2 and λ_y^1 are modelled as changing over time, as the agents repeatedly fine-tune the rule to adapt to the specific market conditions. In their paper, this learning is embodied as a heuristic optimization with a Genetic Algorithm procedure, and introduces the individual heterogeneity to the model. In our paper (as in much of the behavioural macro-literature) we embody the rules with fixed parameters into a representative agent DSGE NK model and allow the data to pin down their values in the estimation of the model.

Our third bounded rationality model is from Gabaix (2020) and is a model where agents are **myopic** with respect to future events concerning current outcomes. This is closely related to a NK model with finite-time planning due to Woodford (2018).

To summarize: we restrict ourselves to four NK models, the benchmark one with rational expectations (**model RE**), Euler learning with a simple adaptive expectations heuristic rule (**model EL-SAE**), Euler learning with the general rule (**model EL-GAE**) and the Gabaix bounded rational model (**model BR**).

Pooling and optimized Rules.

We follow the general methodology of Deak et al. (2020) to address the problem of designing simple policy rules when all models are wrong yet *every* model could 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 methodology lies in the way we handle this third form of uncertainty in the design of simple policy rules. Specifically, following the procedure of Geweke and Amisano (2011) and Geweke and Amisano (2012) we 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 for the simple policy 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.

Road-Map

Section 3 first sets out the micro-foundations of a RE NK model, the consumption and price-setting behaviour in particular, by deriving the decision rules. We then proceed from rational expectations to a number of bounded rationality models in stages. Section 4 gives the structure of a monetary mandate with ZLB. Section 5 illustrates the algorithm to derive the optimal pools. Section 6 provides details of estimation, pooling and results for the robust monetary rule; Section 7 concludes the paper.

2 Other Related Literature

Apart from the area of behavioural macroeconomics already covered, this paper is related to four further 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 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 two types 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 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

[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.

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.

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 policymakers 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. 2007a; Orphanides and Williams. 2008; a review is provided by Taylor and Williams 2010). A common finding in this literature is that optimized 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

[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. (2013) and Tetlow (2015) in focusing on simple, robust optimized Taylor-type rules.

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 methodological 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 properties of the approximating model, those of rival models, and estimates of parameter uncertainty gleaned from various estimation methods. In our optimal pooling approach, policymakers fully utilize the fruits of such exercises. 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 approach 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

[3] See, for example, Dennis et al. (2009) and Ellison and Sargent (2012). Variants of the Hansen-Sargent approach are developed in Adam and Woodford (2012, 2020).

[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.

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., 2007b; Cogley and Sargent, 2005; Levin et al., 2006; Reiss, 2009; 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.

The final strand of literature relates to the benefits of price-level targeting; (see, for example, Svensson, 1999; Schmitt-Grohe and Uribe, 2000; Vestin, 2006; Reiss, 2009; 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. A very recent literature describes these benefits in terms of “make-up" strategies for central banks and in particular the Federal Reserve; see Powell (2020), Svensson (2020). Under such strategies policymakers seek to redress past deviations of inflation from its target. Assuming a make-up rule enjoys credibility, undershooting (overshooting) the target will raise (lower) inflation expectations, lower (raise) the real interest rate and help to stabilize the economy. Inertial Taylor rules have by design the make-up feature as they commitment to a response of the nominal interest rate to a weighted average of past inflation with the weights increasing with the degree of inertia. “Average inflation targeting” is a variant that sets a rolling window of cumulative past deviations; a further variant sets an asymmetric target whereby the central bank responds to average inflation above and below the long-run target in a different way. Hebden et al. (2020) provide details of these different makeup strategies and analyze their effectiveness using the Federal Reserve US macroeconomic model. In our paper we study optimized inertial Taylor rules that are parameterized so as to encompass a simple form of price-level targeting.

3 Models

3.1 Rational Expectation Model (RE)

We first consider a standard NK workhorse model which consists of four sets of representative agents: households, final goods producers, intermediate goods producers and a monetary authority. The intermediate goods producers produce differentiated goods respectively and, in each period of time, consist of a group that is locked into an existing contract and another group that can re-optimize (price rigidity assumption).

3.1.1 Households

Household j chooses between work and leisure and therefore how much labour they supply. Let $C_t(j)$ and $H_t(j)$ denote consumption and labour supply, respectively. The single-period utility is given by

$$U_t(j) = U(C_t(j), H_t(j)) = \log(C_t(j)) - \kappa \frac{H_t(j)^{1+\phi}}{1+\phi} \tag{2}$$

In a stochastic environment, the value function of the representative household at time t is given by

$$V_t(j) = \mathbb{E}_t \left[\sum_{s=0}^{\infty} \beta^s U_{t+s}(j) \right] \tag{3}$$

The household's problem at time t is to choose paths for consumption $\{C_t(j)\}$, labour supply $\{H_t(j)\}$ and holdings of financial assets $\{B_t(j)\}$ to maximize $V_t(j)$ given by (35) given its budget constraint in period t

$$B_t(j) = R_t B_{t-1}(j) + W_t H_t(j) + \Gamma_t - C_t(j) - T_t \tag{4}$$

where $B_t(j)$ is holdings of financial assets at the end of period t, W_t is the real wage rate, R_t is the interest rate paid on assets held at the beginning of period t, Γ_t are profits from wholesale and retail firms owned by households and T_t denote taxes. W_t, R_t, Γ_t and T_t are all exogenous to household j.

To solve the household problem we form a Lagrangian

$$\mathcal{L} = \mathbb{E}_t\left[\sum_{s=0}^{\infty}\beta^s\left\{U_{t+s}(j)\right.\right.$$
$$\left.\left.+\lambda_{t+s}(j)\left[R_{t+s}B_{t+s-1}(j)+W_{t+s}H_{t+s}(j)+\Gamma_{t+s}-C_{t+s}(j)-T_{t+s}-B_{t+s}(j)\right]\right\}\right] \quad (5)$$

The first-order conditions with respect to $\{C_{t+s}(j)\}$, $\{B_{t+s}(j)\}$ and $\{H_{t+s}(j)\}$ are

$$\{C_{t+s}(j)\}: \quad \mathbb{E}_t\beta^s U_{C,t+s}(j)+\beta^s\lambda_{t+s}(j)=0$$
$$\{B_{t+s}(j)\}: \quad \mathbb{E}_t\left[\beta^{s+1}\lambda_{t+s+1}(j)R_{t+s+1}\right]-\beta^s\lambda_{t+s}(j)=0$$
$$\{H_{t+s}(j)\}: \quad \mathbb{E}_t\left[\beta^s U_{H,t+s}(j)+\beta^s\lambda_{t+s}(j)W_{t+s}\right]=0$$

Rearranging the first-order conditions we get:

$$1=\mathbb{E}_t\left[\Lambda_{t,t+1}(j)R_{t+1}\right] \quad (6)$$
$$W_t=-\frac{U_{H,t}(j)}{U_{C,t}(j)} \quad (7)$$

where

$$\Lambda_{t,t+1}(j)=\beta\frac{U_{C,t+1}(j)}{U_{C,t}(j)} \quad (8)$$
$$U_{C,t}=\frac{1}{C_t} \quad (9)$$
$$U_{H,t}=-\kappa H_t^{\phi} \quad (10)$$

is the real stochastic discount factor for household j over the interval $[t,t+1]$.

3.1.2 Firms in the Wholesale

Wholesale firms employ a Cobb-Douglas production function to produce a homogeneous output

$$Y_t^W=F(A_t,H_t)=A_tH_t^{\alpha} \quad (11)$$

where A_t is total factor productivity. Profit-maximizing demand for labour results in the first order condition

$$W_t = \frac{P_t^W}{P_t} F_{H,t} = \alpha \frac{P_t^W}{P_t} \frac{Y_t^W}{H_t} \tag{12}$$

3.1.3 Firms in the Retail Sector

The retail sector uses a homogeneous wholesale good to produce a basket of differentiated goods for aggregate consumption

$$C_t = \left(\int_0^1 C_t(m)^{(\zeta-1)/\zeta} dm \right)^{\zeta/(\zeta-1)} \tag{13}$$

where ζ is the elasticity of substitution. For each m, the consumer chooses $C_t(m)$ at a price $P_t(m)$ to maximize (13) given total expenditure $\int_0^1 P_t(m)C_t(m)dm$. This results in a set of consumption demand equations for each differentiated good m with price $P_t(m)$ of the form

$$C_t(m) = \left(\frac{P_t(m)}{P_t} \right)^{-\zeta} C_t \Rightarrow Y_t(m) = \left(\frac{P_t(m)}{P_t} \right)^{-\zeta} Y_t \tag{14}$$

where $P_t = \left[\int_0^1 P_t(m)^{1-\zeta} dm \right]^{\frac{1}{1-\zeta}}$. P_t is the aggregate price index. C_t and P_t are Dixit-Stigliz aggregates – see Dixit and Stiglitz (1977).

For each variety m the retail good is produced from wholesale production according to an iceberg technology

$$Y_t(m) = Y_t^W = A_t H_t(m)^\alpha \tag{15}$$

Following Calvo (1983), we now assume that there is a probability of $1 - \xi$ at each period that the price of each retail good m is set optimally to $P_t^0(m)$. If the price is not re-optimized, then it is held fixed.[5] For each retail producer m, given its real marginal cost MC_t, the objective is at time t to choose $\{P_t^0(m)\}$ to maximize discounted profits

$$\mathbb{E}_t \sum_{k=0}^{\infty} \xi^k \frac{\Lambda_{t,t+k}}{P_{t+k}} Y_{t+k}(m) \left[P_t^0(m) - P_{t+k} MC_{t+k} MS_{t+k} \right] \tag{16}$$

subject to (14). Where MS_t is a mark-up shock which follows a AR(1) process. The

[5] Thus we can interpret $\frac{1}{1-\xi}$ as the average duration for which prices are left unchanged.

solution to this is

$$\mathbb{E}_t \sum_{k=0}^{\infty} \xi^k \frac{\Lambda_{t,t+k}}{P_{t+k}} Y_{t+k}(m) \left[P_t^0(m) - \frac{1}{(1-1/\zeta)} P_{t+k} MC_{t+k} MS_{t+k} \right] = 0$$

which leads to

$$\frac{P_t^0(m)}{P_t} = \frac{1}{1-1/\zeta} \frac{\mathbb{E}_t \sum_{k=0}^{\infty} \xi^k \Lambda_{t,t+k} (\Pi_{t,t+k})^{\zeta} Y_{t+k} MC_{t+k} MS_{t+k}}{\mathbb{E}_t \sum_{k=0}^{\infty} \xi^k \Lambda_{t,t+k} (\Pi_{t,t+k})^{\zeta-1} Y_{t+k}} \tag{17}$$

where k periods ahead inflation is defined by

$$\Pi_{t,t+k} \equiv \frac{P_{t+k}}{P_t} = \frac{P_{t+1}}{P_t} \frac{P_{t+2}}{P_{t+1}} \cdot \cdot \frac{P_{t+k}}{P_{t+k-1}} = \Pi_{t+1} \Pi_{t+2} \cdot \cdot \Pi_{t+k}$$

Note that $\Pi_{t,t+1} = \Pi_{t+1}$ and $\Pi_{t,t} = 1$.

Let us define

$$\begin{aligned} J_t &= \frac{1}{1-\frac{1}{\zeta}} \mathbb{E}_t \sum_{k=0}^{\infty} \xi^k \Lambda_{t,t+k} \Pi_{t,t+k}^{\zeta} Y_{t+k} MC_{t+k} MS_{t+k} \\ &= \frac{1}{1-\frac{1}{\zeta}} Y_t MC_t MS_t + \xi \mathbb{E}_t \Lambda_{t,t+1} \Pi_{t,t+1}^{\zeta} J_{t+1} \qquad (18) \\ JJ_t &= \mathbb{E}_t \sum_{k=0}^{\infty} \xi^k \Lambda_{t,t+k} \Pi_{t,t+k}^{\zeta-1} Y_{t+k} \\ &= Y_t + \xi \mathbb{E}_t \Lambda_{t,t+1} \Pi_{t,t+1}^{\zeta-1} JJ_{t+1} \qquad (19) \end{aligned}$$

Then (17) can be written as

$$\frac{P_t^0(m)}{P_t} = \frac{J_t}{JJ_t} \tag{20}$$

By the law of large numbers the evolution of the price index is given by

$$P_{t+1}^{1-\zeta} = \xi P_t^{1-\zeta} + (1-\xi)(P_{t+1}^0)^{1-\zeta}$$

which can be written as

$$1 = \xi \Pi_t^{\zeta-1} + (1-\xi) \left(\frac{J_t}{JJ_t} \right)^{1-\zeta} \tag{21}$$

Price dispersion is defined as $\Delta_t = \int (P_t(m)/P_t)^{-\zeta}$. Assuming that the number of firms is large, we obtain the following dynamic relationship:

$$\begin{aligned}\Delta_t &= \xi \int_{not\ optimize} \left(\frac{P^0_{t-1}(m)}{P_{t-1}} \frac{P_{t-1}}{P_t} \right)^{-\zeta} + (1-\xi) \int_{optimize} \left(\frac{P^0_t(m)}{P_t} \right)^{-\zeta_p} \\ &= \xi \Pi_t^{\zeta} \Delta_{t-1} + (1-\xi) \left(\frac{P^0_t(m)}{P_t} \right)^{-\zeta} \\ &= \xi \Pi_t^{\zeta} \Delta_{t-1} + (1-\xi) \left(\frac{J_t}{JJ_t} \right)^{-\zeta} \end{aligned} \tag{22}$$

3.1.4 Profits

Total profits from retail and wholesale firms, Γ_t, are remitted to households. This is given in real terms by

$$\Gamma_t = \underbrace{Y_t - \frac{P_t^W}{P_t} Y_t^W}_{\text{retail}} + \underbrace{\frac{P_t^W}{P_t} Y_t^W - W_t H_t}_{\text{Wholesale}} = Y_t - \alpha \frac{P_t^W}{P_t} Y_t^W \tag{23}$$

using the first-order condition (12).

3.1.5 Closing the Model

The model is closed with a resource constraint

$$Y_t = C_t + G_t \tag{24}$$

and the government's budget constraint

$$G_t = T_t$$

Market clearing in the goods market requires

$$\int_0^1 Y_t(m) dm = \int_0^1 \left(\frac{P_t(m)}{P_t} \right)^{-\zeta} Y_t dm = Y_t \Delta_t \tag{25}$$

using (14). Hence in a symmetric equilibrium

$$Y_t^W = Y_t \Delta_t \tag{26}$$

A monetary policy rule for the nominal interest rate is given by the following Taylor-type rule

$$\log\left(\frac{R_{n,t}}{R_n}\right) = \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) + (1-\rho_r)\left(\theta_\theta \log\left(\frac{\Pi_t}{\Pi}\right) + \theta_y \log\left(\frac{Y_t}{Y}\right)\right) + \log MPS_t, \quad (27)$$

where MPS_t is a monetary policy shock. The ex ante nominal gross interest rate $R_{n,t}$ set at time t and the ex post real interest rate, R_t are related by the Fischer equation

$$R_t = \frac{R_{n,t-1}}{\Pi_t} \quad (28)$$

Exogenous processes evolve according to:

$$\begin{aligned}
\log A_t - \log A &= \rho_A(\log A_{t-1} - \log A) + \epsilon_{A,t} \\
\log MPS_t - \log MPS &= \rho_{MPS}(\log MPS_{t-1} - \log MPS) + \epsilon_{MPS,t} \\
\log G_t - \log G &= \rho_G(\log G_{t-1} - \log G) + \epsilon_{G,t} \\
\log MS_t - \log MS &= \rho_m s(\log MS_{t-1} - \log MS) + \epsilon_{MS,t}
\end{aligned}$$

3.2 Euler Learning Model (EL)

We follow Evans and Honkapohja (2009) and adopt a statistical rational learning. This introduces a specific form of bounded rationality in which utility-maximizing agents make forecasts in each period based on standard econometric techniques called the generalized adaptive expectation.

Because we study mandate framework under welfare analysis, it is insufficient to linearize the model. Therefore, Euler learning in this paper will be implemented within a non-linear setup. The model represented above is written in a compacted from as follows:

$$E_t^*[f(y_{t+1}, y_t, y_{t-1}, w_t)] = 0 \quad (29)$$

where y_t is a vector of the endogenous variables. And w_t is a vector of 4 exogenous variables: MS_t, A_t^c, MPS_t, G_t^c. So that:

$$w_t = g(w_{t-1}, \epsilon_t) \quad (30)$$

We assume that agents are boundedly rational and use a generalized adaptive expectation approach agents to forecast the forward-looking variables:

$$E_t^*(y_{t+1}) = [E_{t-1}^*(y_t)]^{1-\lambda_y^1}[y_t]^{\lambda_y^1+\lambda_y^2}[y_{t-1}]^{-\lambda_y^2}, \quad 0 < \lambda_y^1 < 1, \quad -1 < \lambda_y^2 < 1 \qquad (31)$$

This encompasses simple adaptive expectations ($\lambda_y^2 = 0$) giving us two competing models EL-GAE and EL-SAE. y_{t+1} is partitioned into 2 sets of endogenous variables: the first set is household's subjective forward-looking variables, namely the marginal utility, UC_{t+1}^c, and the inflation, Π_{t+1}. Secondly, price-setting firms have their own subjective forward-looking variables to forecast, Π_{t+1}, J_{t+1}^c, and JJ_{t+1}^c. Finally, $E_t^*(y_{t+1})$ is the subjective forecast. In particular, household subjective forecast of marginal utility and inflation are:

$$\begin{aligned} E_t^*(UC_{t+1}^c) &= [E_{t-1}^*(UC_t^c)]^{1-\lambda_{h,uc}^1}[UC_t^c]^{\lambda_{h,uc}^1+\lambda_{h,uc}^2}[UC_{t-1}^c]^{-\lambda_{h,uc}^2} \\ E_t^*(\Pi_{t+1}) &= [E_{t-1}^*(\Pi_t)]^{1-\lambda_{h,\pi}^1}[\Pi_t]^{\lambda_{h,\pi}^1+\lambda_{h,\pi}^2}[\Pi_{t-1}]^{-\lambda_{h,\pi}^2} \end{aligned}$$

and similarly, we have subjective forecast for firms:

$$\begin{aligned} E_t^*(\Pi_{t+1}) &= [E_{t-1}^*(\Pi_t)]^{1-\lambda_{f,\pi}^1}[\Pi_t]^{\lambda_{f,\pi}^1+\lambda_{f,\pi}^2}[\Pi_{t-1}]^{-\lambda_{f,\pi}^2} \\ E_t^*(J_{t+1}^c) &= [E_{t-1}^*(J_t^c)]^{1-\lambda_J^1}[J_t^c]^{\lambda_J^1+\lambda_J^2}[J_{t-1}^c]^{-\lambda_J^2} \\ E_t^*(JJ_{t+1}^c) &= [E_{t-1}^*(JJ_t^c)]^{1-\lambda_{JJ}^1}[JJ_t^c]^{\lambda_{JJ}^1+\lambda_{JJ}^2}[JJ_{t-1}^c]^{-\lambda_{JJ}^2} \end{aligned}$$

3.3 Myopia Formation Model (BR)

As indicated above there are a large number of different ways of modelling bounded rationality in NK macroeconomic models.[6] In this and the final chapter of the thesis we choose to focus on the model of Gabaix (2020) for a number of reasons. First it is a parsimonious generalization of the widely used work-horse NK model as for example set out in the Gali (2015) recent text-book. Second, it is encompassed by another recent and influential paper, Woodford (2018). Finally, two important paradoxes are resolved: forward guidance is much less powerful than in the standard RE NK model resolving the "forward guidance puzzle" and a permanent rise in the nominal interest rate cases inflation to fall in the short-run, and rise in the long-run resolving the "Fisher paradox".

[6] See Calvert Jump and Levine (2019) for a recent survey.

The rest of this section first describes the idea of a sparse agent that lies at the centre of the Gabaix model, subsection 3.3.1. Then in sub-sections 3.3.2 and 3.3.3 we use this concept to derive the behavioural household decisions of the household and the price-setting firm. Sub-section B.3 then sets out the linearized model about a zero net inflation steady state that recovers the linear model of Gabaix (2020). It is important here to emphasize that we employ a non-linear set-up with a non-zero net inflation rate in the steady state with a ZLB constraint, features that are essential for the optimized simple rules that follow.

3.3.1 The Sparse Agent

Gabaix (2020) generalizes the max operator in economics by assuming less than fully attentive agents. The general idea is as follows: The traditional agent will solve a standard maximization problem: i.e.:

$$Max_a u(a, z) \quad subject \quad to \quad b(a, z) \geq 0 \tag{32}$$

where u is the utility function and b is a constraint.

The *"sparse agent"* will then solve an attention augmented maximization problem as following:

$$SMax_a u(a, z, m) \quad subject \quad to \quad b(a, z, m) \geq 0 \tag{33}$$

where $m \in [0, 1]$ is a vector of agent's attention degree. The idea of a *"sparse agent"* is that she has a low-dimensional sub-model of the world. Hence, first she pays attention only to a few dimensions of the world - which is usually endogenously determined by assuming that attention creates a psychic cost function - or the attention vector is sparse, and second she takes decisions by optimizing her sub-model of the world.

In the concept of this chapter, we assume that the agent's attention degree level is exogenously determined, or the attention parameter vector, m, is given, which means the Sparsemax operator is simplified as the standard maximization operator while the only difference is in the attention vector of parameter m. This attention parameter vector will then be matched with the data by standard Bayesian estimation.

In the Section B.3 we show that with an exact households' utility function and firms' production function, we can derive log-linearized version of the IS and Phillips curves by solving the model forward and then directly apply this inattentive vector m into households' and firms' decision rules.

3.3.2 Household Decisions

In this section, there are three distinct features compared to the one in Gabaix (2020). First, we employ the log form in consumption of the utility function (or $\gamma = 1$). Second, instead of assuming that households are only inattentive to their total income, we distinguish between wage's and government transfer's incomes. Therefore, households' inattentive levels to these different income sources would end up being unequal. Third, we employ a non-linear set-up (which is essential for the computation of the optimized rule) with a non-zero net trend inflation rate.

Household j chooses between work and leisure and therefore how much labour they supply and how much she consumes today. Let $C_t(j)$ and $H_t(j)$ denote consumption and labour supply, respectively. The single-period utility is given by

$$U_t(j) = U(C_t(j), H_t(j)) = \log(C_t(j)) - \kappa \frac{H_t(j)^{1+\phi}}{1+\phi} \tag{34}$$

In a stochastic environment, the value function of the representative household at time t is given by

$$V_t(j) = \mathbb{E}_t \left[\sum_{s=0}^{\infty} \beta^s U_{t+s}(j) \right] \tag{35}$$

The household's problem at time t is to choose paths for consumption $\{C_t(j)\}$, labour supply $\{H_t(j)\}$ and holdings of financial assets $\{B_t(j)\}$ to maximize $V_t(j)$ given by (35) given its budget constraint in period t

$$B_t(j) = R_t B_{t-1}(j) + W_t H_t(j) + \Gamma_t - C_t(j) - T_t \tag{36}$$

where $B_t(j)$ is holdings of financial assets at the end of period t, W_t is the real wage rate, R_t is the interest rate paid on assets held at the beginning of period t, Γ_t are profits from wholesale and retail firms owned by households and T_t denote taxes. W_t, R_t, Γ_t and T_t are all exogenous to household j.

To solve the household problem we form a Lagrangian which is presented in the Appendix where the full rational expectation model is also solved and presented.

For households, aggregating over j, we stationarize the non-stationary variables as follows:

$$\frac{B_t}{\bar{A}_t} = R_t \frac{B_{t-1}}{\bar{A}_{t-1}} \frac{\bar{A}_{t-1}}{\bar{A}_t} + \frac{W_t H_t}{\bar{A}_t} + \frac{\Gamma_t}{\bar{A}_t} - \frac{T_t}{\bar{A}_t} - \frac{C_t}{\bar{A}_t}$$

Solving it forward in time and imposing the transversality condition we can write:

$$\begin{aligned} B^c_{t-1} = {} & \mathbb{E}^{BR}_t \sum_{i=0}^{\infty} \frac{(1+g)^{i+1} C^c_{t+i}}{R_{t,t+i}} + \mathbb{E}^{BR}_t \sum_{i=0}^{\infty} \frac{(1+g)^{i+1} T^c_{t+i}}{R_{t,t+i}} \\ & - \mathbb{E}^{BR}_t \sum_{i=0}^{\infty} \frac{(1+g)^{i+1} W^c_{t+i} H_{t+i}}{R_{t,t+i}} - \mathbb{E}^{BR}_t \sum_{i=0}^{\infty} \frac{(1+g)^{i+1} \Gamma^c_{t+i}}{R_{t,t+i}} \end{aligned} \tag{37}$$

where $R_{t,t+1} \equiv R_t R_{t+1} R_{t+2} \cdots R_{t+i}$ is the real interest rate over the interval $[t, t+i]$. And the variables with superscript c are the stationary version of the endogenous variables, $X^c_t = \frac{X_t}{\bar{A}_t}$.

The forward-looking budget constraint (37) holds for the representative household. In aggregate there is no net debt so $B_{t-1} = 0$. Then in a symmetric equilibrium, substituting for

$$W^c_{t+i} H_{t+i} = \frac{(W^c_{t+i})^{1+\frac{1}{\phi}}}{(\kappa C^c_{t+i})^{\frac{1}{\phi}}}$$

which is the first order condition on the household's supply decision with the following utility function:

$$U(C_t, H_t) = log(C_t) - \kappa \frac{H_t^{1+\phi}}{1+\phi}$$

From (37), substituting (3.3.2) and multiplying both sides by $R_t/(1+g)$ we have

$$\begin{aligned} \mathbb{E}^{BR}_t \sum_{i=0}^{\infty} \frac{(1+g)^i C^c_{t+i}}{R_{t+1,t+i}} = {} & \frac{(W^c_t)^{1+\frac{1}{\phi}}}{(\kappa C^c_t)^{\frac{1}{\phi}}} + \mathbb{E}^{BR}_t \sum_{i=1}^{\infty} \frac{(1+g)^i (W^c_{t+i})^{1+\frac{1}{\phi}}}{(\kappa C^c_{t+i})^{\frac{1}{\phi}} R_{t+1,t+i}} \\ & + \Gamma^c_t + \mathbb{E}^{BR}_t \sum_{i=1}^{\infty} \frac{(1+g)^i \Gamma^c_{t+i}}{R_{t+1,t+i}} - \mathbb{E}^{BR}_t \sum_{i=0}^{\infty} \frac{(1+g)^i T^c_{t+i}}{R_{t+1,t+i}} \end{aligned} \tag{38}$$

Solving the Euler equation $\frac{1}{C^c_t} = \left(\frac{\beta}{1+g}\right) \mathbb{E}^{BR}_t \left[\frac{R_{t+1}}{C^c_{t+1}}\right]$ forward in time we have for $i \geq 1$

$$\frac{1}{C^c_t} = \left(\frac{\beta}{1+g}\right)^i \mathbb{E}^{BR}_t \left[\frac{R_{t+1,t+i}}{C^c_{t+i}}\right] \; ; \; i \geq 1 \tag{39}$$

We assume point expectations, i.e. $E_t f(X_t) \approx f(E_t(X_t))$ and $E_t f(X_t Y_t) \approx f(E_t(X_t)E_t(Y_t))$. For instance, agents are only able to make single variable expectation rather than the expectation of the complicated functions. The concert of BR studied in this chapter is about the limited cognitive capacities of the agents. This additional point expectation assumption is crucial to the result of the non-linear set-up, but it is in line with the agents' cognitive discounting assumption. Notice that, up to the first order Taylor approximation, this assumption is equivalent to using linear approximation as shown in the appendices, where the linear approximation set-up is equivalent to the linear set-up in Gabaix (2020).

We now rearrange (39) to obtain

$$\mathbb{E}_t^{BR} C_{t+i}^c = C_t^c \left(\frac{\beta}{1+g}\right)^i \mathbb{E}_t^{BR} R_{t+1,t+i} \; ; \; i \geq 1 \tag{40}$$

Using it on the the left-hand side of (38) we get

$$\begin{aligned} \sum_{i=0}^{\infty} \frac{(1+g)^i \mathbb{E}_t^{BR} C_{t+i}^c}{\mathbb{E}_t^{BR} R_{t+1,t+i}} &= \sum_{i=0}^{\infty} \frac{(1+g)^i C_t^c \left(\frac{\beta}{1+g}\right)^i \mathbb{E}_t^{BR} R_{t+1,t+i}}{\mathbb{E}_t^{BR} R_{t+1,t+i}} \\ &= \frac{C_t^c}{1-\beta} \end{aligned} \tag{41}$$

Using it on the right-hand side of (38) we get

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{(1+g)^i (\mathbb{E}_t^{BR} W_{t+i}^c)^{1+\frac{1}{\phi}}}{(\kappa \mathbb{E}_t^{BR} C_{t+i}^c)^{\frac{1}{\phi}} \mathbb{E}_t^{BR} R_{t+1,t+i}} &= \sum_{i=1}^{\infty} \frac{(1+g)^i (\mathbb{E}_t^{BR} W_{t+i}^c)^{1+\frac{1}{\phi}}}{\left(\kappa C_t^c \left(\frac{\beta}{1+g}\right)^i \mathbb{E}_t^{BR} R_{t+1,t+i}\right)^{\frac{1}{\phi}} \mathbb{E}_t^{BR} R_{t+1,t+i}} \\ &= \left(\frac{1}{\kappa C_t^c}\right)^{\frac{1}{\phi}} \sum_{i=1}^{\infty} \frac{\beta^{-\frac{i}{\phi}} (1+g)^{i\left(1+\frac{1}{\phi}\right)} (\mathbb{E}_t^{BR} W_{t+i}^c)^{1+\frac{1}{\phi}}}{(\mathbb{E}_t^{BR} R_{t+1,t+i})^{1+\frac{1}{\phi}}} \end{aligned} \tag{42}$$

Substituting back into the forward-looking household budget constraint we arrive at

$$\begin{aligned} \frac{C_t^c}{1-\beta} &= \frac{(W_t^c)^{1+\frac{1}{\phi}}}{(\kappa C_t^c)^{\frac{1}{\phi}}} + \left(\frac{1}{\kappa C_t^c}\right)^{\frac{1}{\phi}} \sum_{i=1}^{\infty} \frac{\beta^{-\frac{i}{\phi}} (1+g)^{i\left(1+\frac{1}{\phi}\right)} (\mathbb{E}_t^{BR} W_{t+i}^c)^{1+\frac{1}{\phi}}}{(\mathbb{E}_t^{BR} R_{t+1,t+i})^{1+\frac{1}{\phi}}} \\ &\quad + \Gamma_t^c + \sum_{i=1}^{\infty} \frac{(1+g)^i \mathbb{E}_t^{BR} \Gamma_{t+i}^c}{\mathbb{E}_t^{BR} R_{t+1,t+i}} - T_t^c - \sum_{i=1}^{\infty} \frac{(1+g)^i \mathbb{E}_t^{BR} T_{t+i}^c}{\mathbb{E}_t^{BR} R_{t+1,t+i}} \\ &= \frac{(W_t^c)^{1+\frac{1}{\phi}} + \sum_{i=1}^{\infty} \left(\frac{(1+g)^{1+\frac{1}{\phi}}}{\beta^{\frac{1}{\phi}}}\right)^i \left(\frac{\mathbb{E}_t^{BR} W_{t+i}^c}{\mathbb{E}_t^{BR} R_{t+1,t+i}}\right)^{1+\frac{1}{\phi}}}{[\kappa C_t^c]^{\frac{1}{\phi}}} \end{aligned} \tag{43}$$

$$+\Gamma_t^c+\sum_{i=1}^{\infty}\frac{(1+g)^i\mathbb{E}_t^{BR}\Gamma_{t+i}^c}{\mathbb{E}_t^{BR}R_{t+1,t+i}}-T_t^c-\sum_{i=1}^{\infty}\frac{(1+g)^i\mathbb{E}_t^{BR}T_{t+i}^c}{\mathbb{E}_t^{BR}R_{t+1,t+i}} \tag{44}$$

Notice that from the (43) to (44), we use the notation such that $R_{t+1,t}=1$.

Employing the Fisher relation, we obtains:

$$\mathbb{E}_t^{BR}R_{t+1,t+i}=\mathbb{E}_t^{BR}\left[R_{t+1}R_{t+2}\cdots R_{t+i}\right]=\mathbb{E}_t^{BR}\left[\frac{R_{n,t,t+i-1}}{\Pi_{t+1,t+i}}\right] \tag{45}$$

Substituting equation (45) into equation (44) to yields:

$$\frac{C_t^c}{1-\beta}=\frac{(W_t^c)^{1+\frac{1}{\phi}}+\sum_{i=1}^{\infty}\left(\frac{(1+g)^{1+\frac{1}{\phi}}}{\beta^{\frac{1}{\phi}}}\right)^i\left(\frac{\mathbb{E}_t^{BR}W_{t+i}^c}{\mathbb{E}_t^{BR}\left[\frac{R_{n,t,t+i-1}}{\Pi_{t+1,t+i}}\right]}\right)^{1+\frac{1}{\phi}}}{[\kappa C_t^c]^{\frac{1}{\phi}}}$$
$$+\Gamma_t^c+\sum_{i=1}^{\infty}\frac{(1+g)^i\mathbb{E}_t^{BR}\Gamma_{t+i}^c}{\mathbb{E}_t^{BR}\left[\frac{R_{n,t,t+i-1}}{\Pi_{t+1,t+i}}\right]}-T_t^c-\sum_{i=1}^{\infty}\frac{(1+g)^i\mathbb{E}_t^{BR}T_{t+i}^c}{\mathbb{E}_t^{BR}\left[\frac{R_{n,t,t+i-1}}{\Pi_{t+1,t+i}}\right]} \tag{46}$$

Hence, we can rewrite equation (46) as follows:

$$\frac{C_t^c}{1-\beta}=\frac{Z_t}{[\kappa C_t^c]^{\frac{1}{\phi}}}+ZZ_t \tag{47}$$

Where Z_t and ZZ_t are expressed as follows:

$$Z_t \;=\; (W_t^c)^{1+\frac{1}{\phi}}+\sum_{i=1}^{\infty}\left(\frac{(1+g)^{1+\frac{1}{\phi}}}{\beta^{\frac{1}{\phi}}}\right)^i\left(\frac{\mathbb{E}_t^{BR}W_{t+i}^c}{\mathbb{E}_t^{BR}\left[\frac{R_{n,t,t+i-1}}{\Pi_{t+1,t+i}}\right]}\right)^{1+\frac{1}{\phi}} \tag{48}$$

$$ZZ_t \;=\; \Gamma_t^c-T_t^c+\sum_{i=1}^{\infty}\frac{(1+g)^i\mathbb{E}_t^{BR}\Gamma_{t+i}^c}{\mathbb{E}_t^{BR}\left[\frac{R_{n,t,t+i-1}}{\Pi_{t+1,t+i}}\right]}-\sum_{i=0}^{\infty}\frac{(1+g)^i\mathbb{E}_t^{BR}T_{t+i}^c}{\mathbb{E}_t^{BR}\left[\frac{R_{n,t,t+i-1}}{\Pi_{t+1,t+i}}\right]} \tag{49}$$

We can now write equations (48) and (49) in a recursive form as follows:

$$Z_t \;=\; (W_t^c)^{1+\frac{1}{\phi}}+\left(\frac{(1+g)^{1+\frac{1}{\phi}}}{\beta^{\frac{1}{\phi}}}\right)\left(\frac{\mathbb{E}_t^{BR}Z_{t+1}}{\left[\frac{R_{n,t}}{\mathbb{E}_t^{BR}\Pi_{t+1}}\right]^{1+\frac{1}{\phi}}}\right) \tag{50}$$

$$ZZ_t \;=\; \Gamma_t^c-T_t^c+(1+g)\frac{\mathbb{E}_t^{BR}ZZ_{t+1}}{\left[\frac{R_{n,t}}{\mathbb{E}_t^{BR}\Pi_{t+1}}\right]} \tag{51}$$

We now follow Gabaix (2020) to assume that the behavioural agent perceives reality with some myopia, which is associated with deviations from the steady state, say $E_t^{BR}X_{t+1} = \bar{m}_h f(X_t)$. Note that, we assume that households do not have inattentive level on the individual aggregate variables because the parameter $0 < \bar{m}_h < 1$ already takes into account for the inattention. Hence, rewriting equations (50) and (51) yields:

$$Z_t = (W_t^c)^{1+\frac{1}{\phi}} + \left(\frac{(1+g)^{1+\frac{1}{\phi}}}{\beta^{\frac{1}{\phi}}}\right)\left(\frac{\mathbb{E}_t\left(Z + \bar{m}_h \hat{Z}_{t+1}\right)}{\mathbb{E}_t\left(\frac{R_{n,t}}{\Pi + \bar{m}_h \hat{\Pi}_{t+1}}\right)^{1+\frac{1}{\phi}}}\right) \tag{52}$$

$$ZZ_t = (\Gamma_t^c - T_t^c) + (1+g)\left(\frac{\mathbb{E}_t\left(ZZ + \bar{m}_h \hat{ZZ}_{t+1}\right)}{\mathbb{E}_t\left(\frac{R_{n,t}}{\Pi + \bar{m}_h \hat{\Pi}_{t+1}}\right)}\right) \tag{53}$$

The three equations (47),(52), and (53) then constitute a nonlinear behavioural *consumption function* which also nests the fully rational expectation one when the vector of inattentive parameters, m, is equal to a vector of ones. In general, equations (52), and (53) are the future discounted values of the proportioned wages and net transfers from the government to households. Hence, equation (47) shows that consumption today is an increasing function of future discounted values of household's incomes, but it is decreasing in the central bank's policy rates, $\frac{R_{n_t}}{\Pi + \bar{m}_h \hat{\Pi}_{t+1}}$. Since these variables are exogenous to the atomistic household we therefore have an 'anticipated utility' form of household behaviour suitable for either our behavioural or RE models.

3.3.3 Price-setting Firms

The homogeneous production technology in the economy is:

$$Y_t^c = A_t^c H_t^\alpha \tag{54}$$

There is a probability of $1 - \xi$ at each period that the price of each retail good i is set optimally to $P_t^0(i)$; otherwise it is held fixed.

Retail behavioural producer i, given the common real marginal cost $MC_t(i) = MC_t$ chooses $\{P_t^0(i)\}$ to maximize discounted real profits

$$\mathbb{E}_t^{BR} \sum_{k=0}^{\infty} \xi^k \frac{\Lambda_{t,t+k}}{P_{t+k}} Y_{t+k}^c(i) \left[P_t^0(i) - P_{t+k} MC_{t+k}\right] \tag{55}$$

where $\Lambda_{t,t+k} \equiv \beta^k \frac{U_{C,t+k}}{U_{C,t}}$ is the stochastic discount factor over the interval $[t, t+k]$, subject to

$$Y^c_{t+k}(i) = \left(\frac{P^O_t(i)}{P_{t+k}} \right)^{-\zeta} Y^c_{t+k} \tag{56}$$

The solution to this is

$$\mathbb{E}^{BR}_t \sum_{k=0}^{\infty} \xi^k \Lambda_{t,t+k} Y^c_{t+k}(i) \left[\frac{P^0_t(i)}{P_{t+k}} - \frac{1}{(1-1/\zeta)} MC_{t+k} \right] = 0 \tag{57}$$

which leads to

$$\frac{P^0_t(m)}{P_t} = \frac{1}{1-1/\zeta} \frac{\mathbb{E}^{BR}_t \sum_{k=0}^{\infty} \xi^k \Lambda_{t,t+k} (\Pi_{t,t+k})^{\zeta} Y^c_{t+k} MC_{t+k}}{\mathbb{E}^{BR}_t \sum_{k=0}^{\infty} \xi^k \Lambda_{t,t+k} (\Pi_{t,t+k})^{\zeta-1} Y^c_{t+k}} \tag{58}$$

where k periods ahead inflation is defined by

$$\Pi_{t,t+k} \equiv \frac{P_{t+k}}{P_t} = \frac{P_{t+1}}{P_t} \frac{P_{t+2}}{P_{t+1}} \cdot \cdot \frac{P_{t+k}}{P_{t+k-1}} = \Pi_{t+1} \Pi_{t+2} \cdot \cdot \Pi_{t+k}$$

Note that $\Pi_{t,t+1} = \Pi_{t+1}$ and $\Pi_{t,t} = 1$.

Let us define

$$\begin{aligned} J^c_t &= \frac{1}{1-\frac{1}{\zeta}} \mathbb{E}^{BR}_t \sum_{k=0}^{\infty} \xi^k \Lambda_{t,t+k} \Pi^{\zeta}_{t,t+k} Y^c_{t+k} MC_{t+k} \\ &= \frac{1}{1-\frac{1}{\zeta}} Y^c_t MC_t + \xi \mathbb{E}^{BR}_t \Lambda_{t,t+1} \Pi^{\zeta}_{t,t+1} J^c_{t+1} \end{aligned} \tag{59}$$

$$\begin{aligned} JJ^c_t &= \mathbb{E}^{BR}_t \sum_{k=0}^{\infty} \xi^k \Lambda_{t,t+k} \Pi^{\zeta-1}_{t,t+k} Y^c_{t+k} \\ &= Y^c_t + \xi \mathbb{E}^{BR}_t \Lambda_{t,t+1} \Pi^{\zeta-1}_{t,t+1} JJ^c_{t+1} \end{aligned} \tag{60}$$

Then (58) can be written as

$$\frac{P^0_t(m)}{P_t} = \frac{J_t}{JJ_t} \tag{61}$$

By the law of large numbers the evolution of the price index is given by

$$P^{1-\zeta}_{t+1} = \xi P^{1-\zeta}_t + (1-\xi)(P^0_{t+1})^{1-\zeta} \tag{62}$$

which can be written as

$$1 = \xi \Pi_t^{\zeta-1} + (1-\xi)\left(\frac{J_t^c}{JJ_t^c}\right)^{1-\zeta} \tag{63}$$

We first transform the equations (59), (60) to the expectations of the **behavioural** agents, where we also utilise the relation $\mathbb{E}_t^{BR}\Lambda_{t,t+k} = \frac{1}{\mathbb{E}_t^{BR} R_{t+1,t+k}} = \frac{1}{\mathbb{E}_t^{BR}\left[\frac{R_{n,t,t+k-1}}{\Pi_{t+1,t+k}}\right]}$, and employing the assumption about the firms' myopia about the future state such that: $\mathbb{E}_t^{BR}(X_{t+1} - X) = \bar{m}_f \mathbb{E}_t(X_{t+1} - X)$. In addition, firms are inattentive to the market's variables which are exogenous to them. Hence, we can re-write the equations (59) and (60) in a recursive form as follows:

$$J_t^c = \frac{1}{1-\frac{1}{\zeta}} Y_t^c MC_t + \xi \mathbb{E}_t \frac{(\Pi + \bar{m}_f \hat{\Pi}_{t+1})^{\zeta}}{\mathbb{E}_t\left(\frac{R_{n,t}}{\Pi + \bar{m}_f \hat{\Pi}_{t+1}}\right)} (J^c + \bar{m}_f \hat{J}_{t+1}^c) \tag{64}$$

$$JJ_t^c = Y_t^c + \xi \mathbb{E}_t \frac{(\Pi + \bar{m}\hat{\Pi}_{t+1})^{\zeta-1}}{\mathbb{E}_t\left(\frac{R_{n,t}}{\Pi + \bar{m}_f \hat{\Pi}_{t+1}}\right)} (JJ^c + \bar{m}_f \hat{JJ}_{t+1}^c) \tag{65}$$

As for the household, price-setting is now expressed in terms of real marginal cost and aggregate demand, variables that are exogenous to the atomistic firm. Again we therefore have an 'anticipated utility' form of firm behaviour suitable for either our behavioural or RE models.

4 A Mandate Framework for Imposing the ZLB

Recall the nominal interest rate rule in 'implementable form':

$$\log\left(\frac{R_{n,t}}{R_n}\right) = \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) + (1-\rho_r)\left(\theta_\pi \log\left(\frac{\Pi_t}{\Pi}\right) + \theta_y \log\left(\frac{Y_t}{Y}\right) + \theta_{dy} \log\left(\frac{Y_t}{Y_{t-1}}\right)\right) \tag{66}$$

which for optimal policy purposes we re-parameterize as

$$\log\left(\frac{R_{n,t}}{R_n}\right) = \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) + \alpha_\pi \log\left(\frac{\Pi_t}{\Pi}\right) + \alpha_y \log\left(\frac{Y_t}{Y}\right) + \alpha_{dy} \log\left(\frac{Y_t}{Y_{t-1}}\right) \tag{67}$$

which allows for the possibility of an integral rule with $\rho_r = 1$

Let $\rho \equiv [\rho_r, \alpha_\pi, \alpha_y, \alpha_{dy}]$ be the policy choice of feedback parameters that defines the form of the rule. The equilibrium is solved by backward induction in the following two-stage delegation game.

1. **Stage 1**: The policymaker (the leader) chooses a per period probability of hitting the ZLB, a trend inflation rate and designs the optimal loss function in the mandate.
2. **Stage 2**: The CB receives the mandate in the form of a welfare criterion and rule of the form (67). Welfare is then optimized with respect to ρ resulting in an optimized rule.

The equilibrium of this ZLB delegation mandate is solved by backward induction in the following two-stage game.

This delegation game is solved by backwards induction as follows:

4.1 Stage 2: The CB Choice of Rule

Given a steady state inflation rate target, Π, the Central Bank (CB) receives a mandate to implement the rule (67) and to maximize with respect to ρ a modified welfare criterion

$$\begin{aligned}\Omega_t^{mod} &\equiv \mathbb{E}_t\left[\sum_{\tau=0}^{\infty}\beta^{\tau}\left(U_{t+\tau} - w_r\left(R_{n,t+\tau}-R_n\right)^2\right)\right] \\ &= \left(U_t - w_r\left(R_{n,t}-R_n\right)^2\right) + \beta(1+g)^{1-\sigma}\mathbb{E}_t\left[\Omega_{t+1}^{mod}\right] \qquad (68)\end{aligned}$$

One can think of this as a mandate with a penalty function $P = w_r\left(R_{n,t}-R_n\right)^2$, penalizing the variance of the nominal interest rate with weight w_r.[7]

Following Den Haan and Wind (2012), an alternative mandate that only penalizes the zero interest rate in an asymmetric fashion is $P = P(a_t)$ where the OBC is $a_t \equiv R_{n,t} - 1 \geq 0$ with

$$P = P(a_t) = \frac{\exp(-w_r a_t)}{w_r} \qquad (69)$$

and chooses a large w_r. $P(a_t)$ then has the property

$$\begin{aligned}\lim_{w_r\to\infty} P(a_t) &= \infty \text{ for } a_t < 0 \\ &= 0 \text{ for } a_t > 0\end{aligned}$$

[7] This closely follows the approximate form of the ZLB constraint of Woodford (2003) and Levine et al. (2008).

Thus $P(a_t)$ enforces the ZLB approximately but with more accuracy as w_r becomes large. Stages 2–1 then proceed as before, but we now confine ourselves to a large w_r which will enable Π to be close to unity.

Both the symmetric and asymmetric forms of a ZLB mandate result in a probability of hitting the ZLB

$$p = p(\Pi, \rho^*(\Pi, w_r)) \tag{70}$$

where $\rho^*(\Pi, w_r)$ is the optimized form of the rule given the steady state target Π and the weight on the interest rate volatility, w_r.

4.2 Stage 1: Choice of the Steady State Inflation Rate Π and Design of the Mandate

The policymaker first chooses a per period probability $\bar{p}$ of the nominal interest rate hitting the ZLB (which defines the tightness of the ZLB constraint). Then given a target low probability $\bar{p}$ and given w_r, $\Pi = \Pi^*$ is chosen so satisfy

$$p(R_{n,t} \leq 1) \equiv p(\Pi^*, \rho^*(\Pi^*, w_r)) \leq \bar{p} \tag{71}$$

This then achieves the ZLB constraint

$$R_{n,t} \geq 1 \text{ with high probability } 1 - \bar{p} \tag{72}$$

where $R_{n,t}$ is the nominal interest rate.

The mandate is then designed to maximize the *actual* household intertemporal welfare

$$\Omega_t = \mathbb{E}_t \left[\sum_{\tau=0}^{\infty} \beta^\tau U_{t+\tau} \right] = U_t + \beta(1+g)^{1-\sigma} \mathbb{E}_t \left[\Omega_{t+1} \right] \tag{73}$$

with respect to w_r.

This two-stage delegation game defines an equilibrium in choice variables w_r^*, ρ^* and Π^* that maximizes the true household welfare subject to the ZLB constraint (72).

5 Methodology: Designing a Robust Mandate

The goal of the policymaker is to choose the parameters of a Taylor-type monetary policy rule ρ^*, an optimal long-run inflation level $\bar{\Pi}^*$ and an optimal delegated mandate w_r^* to maximize the actual welfare and satisfy the ZLB on the nominal interest rate, we denote these parameters as the parameter set defining the mandate $\delta = [\rho, \bar{\Pi}, w_r]$, that are robust to both within- and across-model uncertainty. We use the expected lifetime utility of households

$$\Omega_i(\delta, \psi) = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t U_t(\delta, \psi) \right] \qquad \psi \in \Psi_i \tag{74}$$

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 $U_t(\delta, \psi)$ denotes utility in period t given the vector of estimated parameters $\psi \in \Psi_i$, policy rule ρ, long-term inflation level and the optimal formation of the delegated mandate w_r. We allow the parameter space Ψ_i to differ, but require the policy rule ρ, long-term inflation level and the optimal formation of the delegated mandate w_r 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. We denote the mandate parameter set is δ, welfare in model $\mathcal{M}_i$ is given by

$$\Omega_i(\delta) = \int_{\Psi_i} \Omega(\psi, \delta) p(\psi | \mathbf{Y}^o_{i,T}, \mathcal{M}_i) d\psi \tag{75}$$

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 policymaker seeks a common mandate δ^* across every model that maximizes

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

a welfare measure that incorporates both within- and across-model uncertainty and subjected to the ZLB constraint on the nominal interest rate

$$P_{zlb}(\delta, w) = \sum_{i=1}^{m} w_i P_{zlb,i}(\delta) \leq \bar{P}$$

where $\bar{P}$ is a given unconditional probability of hitting the ZLB of the nominal interest rate. Given the weights the chosen optimal mandate with $\delta = \delta^*$ is our *optimal robust mandate.*

The novelty of our methodology 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 the case of Bayesian 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 $p(\mathbf{y}^f_{T+k}|\mathbf{Y}^o_{i,T}, \mathcal{M}_i)$ be the k-period ahead predictive density of model $\mathcal{M}_i$ for a vector of model variables $\mathbf{y}^f_{T+k}$ given observations$\mathbf{Y}^o_{i,T}$:

$$p(\mathbf{y}^f_{T+k}|\mathbf{Y}^o_{i,T}, \mathcal{M}_i) = \int_{\Psi_i} p(\mathbf{y}^f_{T+k}|\mathbf{Y}^o_{i,T}, \psi, \mathcal{M}_i) p(\psi|\mathbf{Y}^o_{i,T}, \mathcal{M}_i) d\psi, \tag{76}$$

where $p(\mathbf{y}^f_{T+k}|\mathbf{Y}^o_{i,T}, \psi, \mathcal{M}_i)$ 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 $\mathbf{y}^f_{T+k}$, 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 $\mathbf{Y}^f_T = \{\mathbf{y}^f_1, \ldots, \mathbf{y}^f_T\}$ of forecast variables, the log predictive score of model $\mathcal{M}_i$ is given by

$$LS(\mathbf{Y}^f_T, \mathcal{M}_i) = \sum_{t=h}^{T-K} \sum_{k=1}^{K} \log p(\mathbf{y}^f_{t+k}|\mathbf{Y}^o_{i,t}, \mathcal{M}_i) \tag{77}$$

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.[8] 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

$$LS(\mathbf{Y}_T^f, \mathcal{M}) = \sum_{t=h}^{T-K} \sum_{k=1}^{K} \log \left[\sum_{i=1}^{m} w_i p(\mathbf{y}_{t+k}^f | \mathbf{Y}_{i,t}^o, \mathcal{M}_i) \right]; \quad \sum_{i=1}^{m} w_i = 1; \quad w_i \geq 0. \tag{78}$$

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[9]

$$w_i^* = \arg\max_{w_i} LS(\mathbf{Y}_T^f, \mathcal{M}) \tag{79}$$

Before we turn to our empirical analysis to demonstrate the methodology in practice, let us highlight the differences between prediction pools and BMA (Table 1). 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.

[8] Del Negro et al. (2016) use the terminology static pools to reflect the fact that weights are time invariant.
[9] Logs are used in general since they make the densities globally concave, making the maximization easier.

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 weight to a single model	Less of a tendency that a single model dominates

Table 1: BMA versus Optimal Pooling

6 Results

This section reports our numerical results starting with a description of the data and the measurement equations, then proceeding to identification tests, Bayesian estimation and the computation of the optimal pooling weights. Finally the latter are used to compute the optimal robust rule with a soft ZLB constraint.

6.1 Measurement Equations and Data

Our observables used in the estimation are: GDP per capita growth (dyobs), percentage deviation of hours worked per capita from mean (labobs), monetary policy rate (robs) which we employ both the nominal FED short-term rate and the shadow rate from inflation rate (pinfobs). The corresponding measurement equations expressed in terms of stationarized variables are:

$$\begin{aligned} \text{dyobs} &= \log\left((1+g)\frac{Y_t}{Y_{t-1}}\right) \\ \text{labobs} &= \frac{H_t^d - H^d}{H^d} \\ \text{robs} &= R_{n,t} - 1 \\ \text{pinfobs} &= \log(\Pi_t) \end{aligned}$$

The steady state values of the observables are dyobs=log$(1+g)$, labobs=0, robs=$R_n - 1$, and pinfobs=log(Π).

The original data are taken from the FRED Database available through the Federal Reserve Bank of St.Louis. The data consists of 7 quarterly time series, namely log output growth (*dyobs*), labour hours supply (*labobs*), the net inflation (*pinfobs*), and finally the policy rate measurement (*robs*). Since our focus on the ZLB we also provide a new estimation with the Wu-Xia Shadow interest rate replacing the FED rate, *robs* - see Wu and Xia (2016). The sample period is 1958:1-2017:4. There is a pre-sample period of 4 quarters so the observations actually used for the estimation go from 1959:1-2017:4, 240 observations.

6.2 Identification

We use DYNARE implementation to check for the identification of the models. Iskrev (2010) follows Ratto (2008) in using the information matrix derived from the Jacobian (mean and covariance matrix), Komunjer and Ng (2011) provide a difference route to the local identification of a linear state space, they examine directly the relationship between the coefficients of the state-space representation of the DSGE model and the parameter vector θ. In addition, the setup also accounts for the condition of left-invertablity (or the number of structural shocks is different from the that of the observables). However, in our setup, we always have the "squared matrix", so the full-rank condition on the coefficients matrix and on the Jacobian matrix as in Ratto (2008) is sufficient for local identification.

Qu and Tkachenko (2012) work in the frequency domain, i.e., whether the mean and spectrum of observables is uniquely determined by the deep parameters at all frequencies? Using a frequency domain approximation of the likelihood function and utilizing the information matrix equality, they express the Hessian as the outer product of the Jacobian matrix of derivatives of the spectral density with respect to θ. However, this approach has to be implemented numerically. For each conjectured θ_0 we have to compute the rank of the spectral density matrixy. Because in a typical implementation the computation of the matrix relies on numerical differentiation (and integration) over the subset frequency domains, there might arise discordant results in the matrix rank. For instance, if two parameters jointly enter the model and play a very similar role in the model after linearization (i.e., stickiness level of price parameter and the rate of substitution jointly determine the speed of adjustment of prices through the Calvo probability), thus they are separately unidentifiable. Qu and Tkachenko (2012) procedure tests the identification over a subset of estimated parameters, so the model fails to pass the test over each subset of parameters.

Identification Criteria	RE	BR	EL (SAE)	EL (GAE)
REDUCED-FORM	satisfied	satisfied	satisfied	satisfied
MINIMAL SYSTEM (KOMUNJER AND NG, 2011)	satisfied	satisfied	satisfied	satisfied
SPECTRUM (QU AND TKACHENKO, 2012)	unsatisfied	unsatisfied	satisfied	satisfied
MOMENTS (ISKREV, 2010)	satisfied	satisfied	satisfied	satisfied

Table 2: Identification results

6.3 Bayesian Estimation

Parameters	**Prior**			**Post. RE**		**Post. BR**		**Post. EL(SAE)**		**Post. EL(GAE)**	
	pdf	Mean	Std	Mean	s.d	Mean	s.d	Mean	s.d	Mean	s.d
(ϵ_a)	**IG**	0.001	0.02	0.0065	0.0003	0.0065	0.0003	0.0066	0.0003	0.0065	0.0003
(ϵ_{ms})	**IG**	0.001	0.02	0.0370	0.0036	0.0297	0.0035	0.0889	0.0369	0.0285	0.0033
(ϵ_{mps})	**IG**	0.001	0.02	0.005	0.0004	0.0050	0.0004	0.0023	0.0001	0.0035	0.0003
(ϵ_g)	**IG**	0.001	0.02	0.0518	0.0045	0.0434	0.0037	0.0246	0.0011	0.0456	0.0040
(ρ_a)	**IG**	0.50	0.20	0.9897	0.0050	0.9919	0.0039	0.9854	0.0065	0.9840	0.0072
(ρ_{ms})	**IG**	0.50	0.20	0.9560	0.0124	0.9633	0.0120	0.7667	0.0892	0.9728	0.0108
(ρ_{mps})	**IG**	0.50	0.20	0.5994	0.0377	0.6175	0.0364	0.3871	0.0681	0.7941	0.0433
(ρ_g)	**IG**	0.50	0.20	0.9088	0.0106	0.9475	0.0132	0.9843	0.0061	0.9761	0.0096
(ξ)	**B**	0.50	0.10	0.7552	0.0175	0.6962	0.0290	0.7955	0.0194	0.5503	0.0468
(ϕ)	**N**	2	0.75	4.6626	0.4918	3.7243	0.5046	1.2257	0.4632	3.7331	0.5167
(α)	**B**	0.50	0.10	0.8861	0.0369	0.9092	0.0319	0.9734	0.0101	0.8882	0.0358
	pdf	Mean	Std	Mean	s.d	Mean	s.d	Mean	s.d	Mean	s.d
(ρ_r)	**B**	0.75	0.10	0.3015	0.0513	0.3343	0.0531	0.8808	0.0239	0.4400	0.0618
(θ_π)	**N**	1.50	0.25	2.4474	0.1551	2.6075	0.1567	1.4193	0.1724	2.0188	0.1848
(θ_y)	**N**	0.12	0.05	0.0597	0.0189	0.0417	0.0236	0.1132	0.0307	0.0480	0.0286
(θ_{dy})	**N**	0.12	0.05	0.1092	0.0300	0.1739	0.0348	0.1905	0.0485	0.0601	0.0369
Myopic parameters											
$(\bar{m}_h)$	**B**	0.50	0.20			0.9379	0.0156				
$(\bar{m}_f)$	**B**	0.50	0.20			0.5405	0.2801				
Euler Learning parameters											
$(\lambda^1_{h,uc})$	**B**	0.50	0.20					0.1195	0.0393	0.0344	0.0166
$(\lambda^1_{h,\pi})$	**B**	0.50	0.20					0.0678	0.0059	0.0556	0.0247
$(\lambda^2_{h,uc})$	**B**	0.0	0.25							0.8036	0.0296
$(\lambda^2_{h,\pi})$	**B**	0.0	0.25							-0.7315	0.0872
$(\lambda^1_{f,\pi})$	**B**	0.50	0.20					0.2608	0.0036	0.2386	0.0690
(λ^1_{j})	**B**	0.50	0.20					0.2285	0.0476	0.8861	0.0593
(λ^1_{jj})	**B**	0.50	0.20					0.5448	0.0579	0.6522	0.1584
$(\lambda^2_{f,\pi})$	**B**	0.0	0.25							0.1345	0.0739
(λ^2_{j})	**B**	0.0	0.25							0.6223	0.1068
(λ^2_{jj})	**B**	0.0	0.25							0.1073	0.1996

Table 3: Estimation results - Parameters

Estimated results show that there is a significant difference in the general myopia levels about the future state between households and firms. Moreover, the estimated learning parameters of the EL model are also statistically significant.

Smets and Wouters (2003) and Smets and Wouters (2007) have shown that rational models with a rich set of frictions and a general stochastic structure can explain the data relatively well. However, these models require an implausibly high level of price and wage stickiness and exogenous shocks to explain the observed persistence in the data.[10] My estimated results show that the boundedly rational expectation models reduce the scale of structural price-stickiness friction, ξ, and the magnitude of estimated shocks, which, most importantly, improves the marginal log likelihood relative to the RE model.

	RE	Gabaix	EL (SAE)	EL (GAE)
LogDataDensity (Nominal rate)	3813.00	3805.60	3766.64	3817.98
LogDataDensity (Shadow rate)	3750.70	3750.80	3715.47	3772.25

Table 4: Log data density, based on the Modified Harmonic Mean Estimator (Full sample Q1-1958 to Q4-2017 with Shadow rate). The preference parameter, β, is calibrated of 0.9995 at the data sample. The usual procedure before was that we calibrated the mean of inflation and nominal interest rate over the data sample then we calculated the preference parameter, β, accordingly.

- Based on the marginal log likelihood, the estimated rational expectation model outperform the models under BR and EL at fitting with the data if we use the nominal interest rate as the observable.[11]

[10] Smets and Wouters (2007) resolve this problem by introducing Kimball rather than Dixit-stiglitz preferences. However, for Kimball preference to have a significant impact requires a huge super-price elasticity which is inconsistent with empirical evidence, Deak et al. (2020). Hence, BR and EL are the alternative approaches to explain the persistence in observed data.

[11] Log data density. based on the Modified Harmonic Mean Estimator (Full sample 1959 to 2017 with Fed interest rate as nominal interest rate).

6.4 Optimal Pools

We estimate our models repeatedly with an increasing window of data, and compute log predictive scores (77) and (78) 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 eight quarters ahead ($K = 8$).[12] We increase the sample size by four quarters each time and repeat the same steps.[13] Our last sample ends at 2016:4 ($T = 208$) as to allow for the computation of predictive densities using data up to 2017:4.

First consider a pool of all four models **model RE**,**model EL-SAE**), **model EL-GAE** and **model BR** in Figure 1. The top panel 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 roughly the same in most periods.

Importantly, we employ prediction pooling to aggregate these 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 (79) 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. It is clear that the RE, EL-GAE and BR models provide far better predictions than EL-SAE.

Weights of approximately 0.40, 0.32 and 0.28 are assigned to these dominant models, by the end of the sample period.

The bottom panel of Figure 1 shows an interesting and important contrast to the middle panel. It shows how the Bayesian odds evolve over our sample, given a uniform prior belief of the policymaker over the competing models. Had the policymaker used BMA to attach weights to the models, she would have put most of her faith in the BR 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 change slowly over time while large changes in Bayesian odds can be brought by adding only a handful of observations to

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

[13]We re-estimate 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.

the sample.

Figure 2 increases the prediction period to 8 periods with a the similar result that model EL-SAE is dominated by the other two. We therefore focus on the three empirically relevant models **model RE**, **model EL-GAE** and **model BR** in figures 3 and 4. For the optimal policy exercise we choose the 8-period end-of-sample weights of approximately 0.40, 0.32 and 0.28.

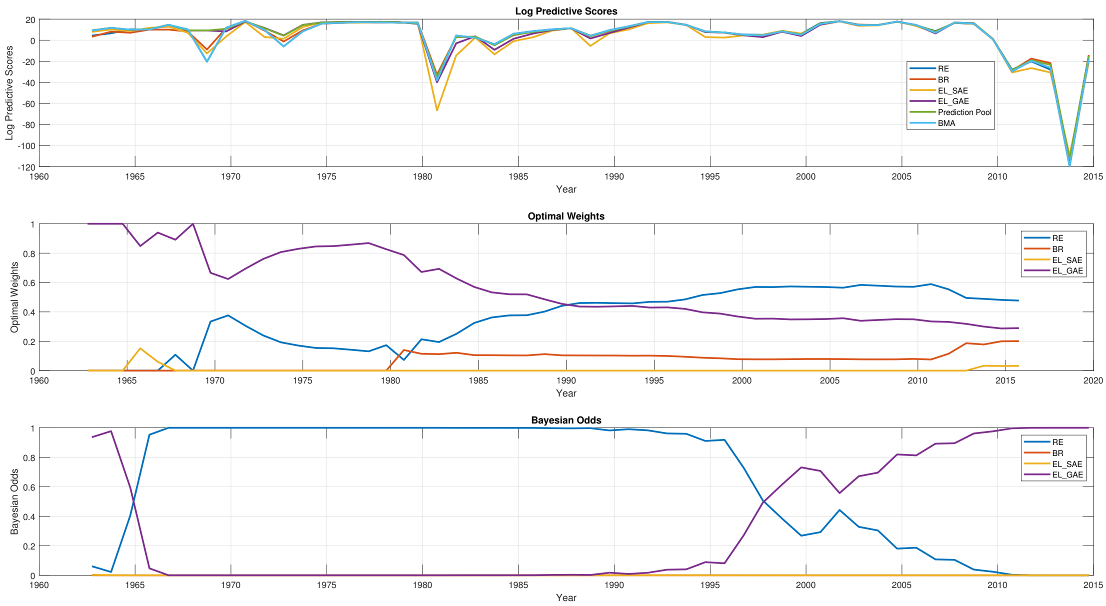

Figure 1: (4 models) 4-period ahead forecasting optimal weights (Shortest sample from 1958Q1 to 1963Q1, longest sample from 1958Q1 to 2014Q4)

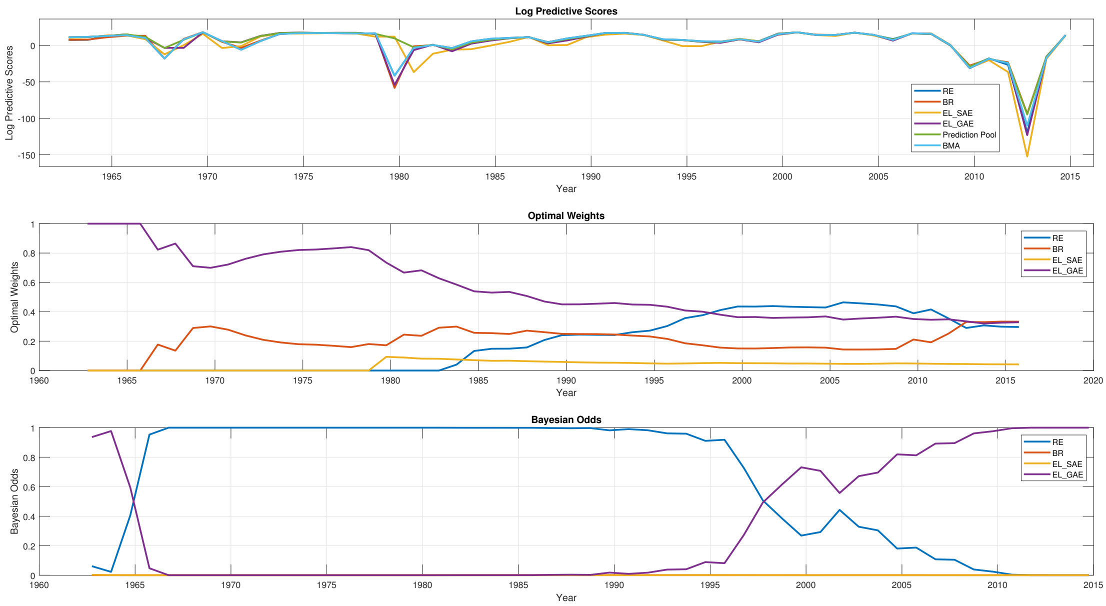

Figure 2: **(4 models) 8-period ahead forecasting optimal weights(Shortest sample from 1958Q1 to 1963Q1, longest sample from 1958Q1 to 2014Q4)**

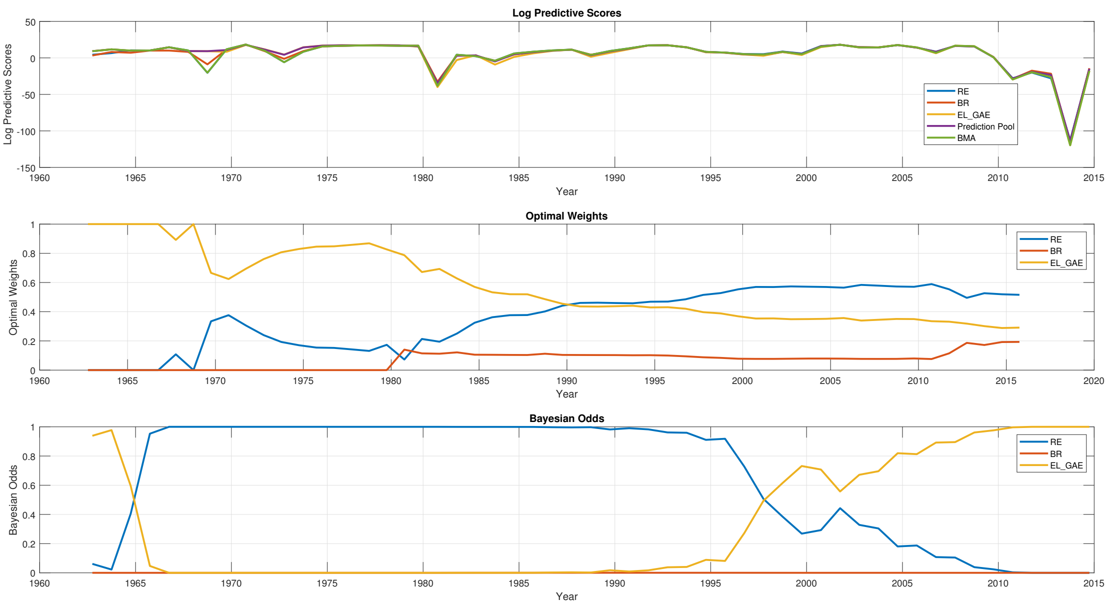

Figure 3: **(3 models) 4-period ahead forecasting optimal weights (Shortest sample from 1958Q1 to 1963Q1, longest sample from 1958Q1 to 2014Q4)**

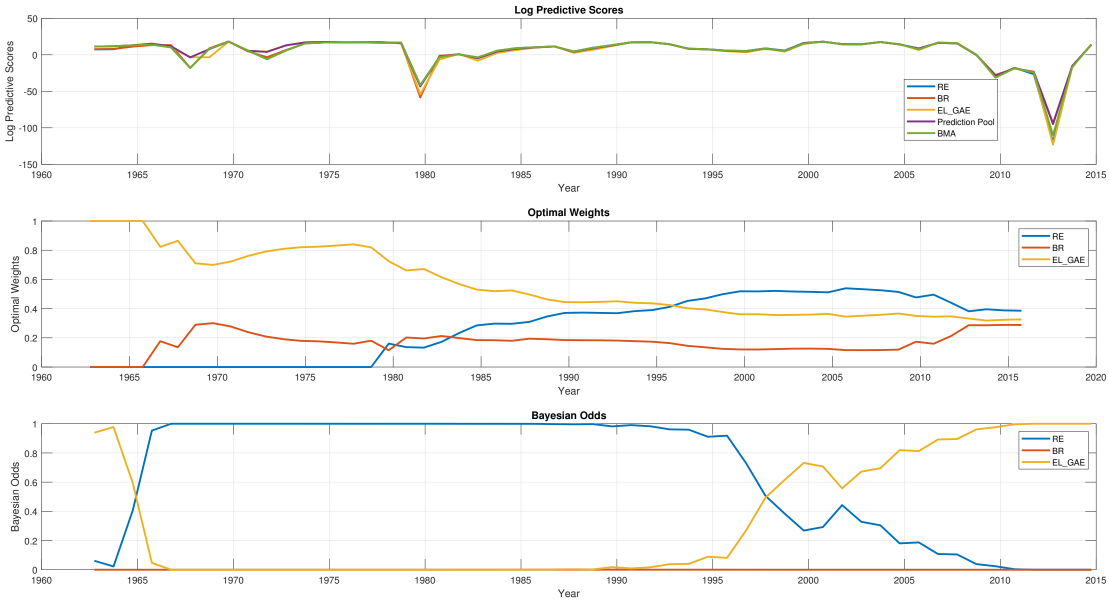

Figure 4: (3 models) 8-period ahead forecasting optimal weights(Shortest sample from 1958Q1 to 1963Q1, longest sample from 1958Q1 to 2014Q4)

6.5 The Robust ZLB Mandate

6.5.1 Optimized Simple Rules

Before turning to the ZLB mandate, Table 5 computes the optimized rules $\rho^* = [\rho_r^*, \alpha_\pi^*, \alpha_y^*, \alpha_{dy}^*]$ in the absence of any ZLB considerations. First the steady-state gross inflation rate is set at its welfare-optimal $\Pi = 1$ (zero net inflation) and then at the empirical average level of the period of estimation $\Pi = 1.008$ (an annual rate 3.24%).

Clearly there is a significant difference of the equilibrium points between models in terms of ρ^* for all elements except $\alpha_y^* = 0$. For $\Pi = 1$ we see a very high probability of hitting the nominal interest rate ZLB for each model individually and for the optimal pool as well. A higher steady-state empirical gross inflation rate mitigates the problem for rules chosen for each model individually and more so for the pool.[14]

[14]The consumption equivalent variations (CEV) is calculated as follows:$CEV = \frac{\Omega^*_{M_i} - \Omega^*_{M_i,OSR,\Pi=1}}{CE}$, where $CE = 19.9$ is the estimated steady state value of the consumption equivalence in the common deterministic steady of the three models.

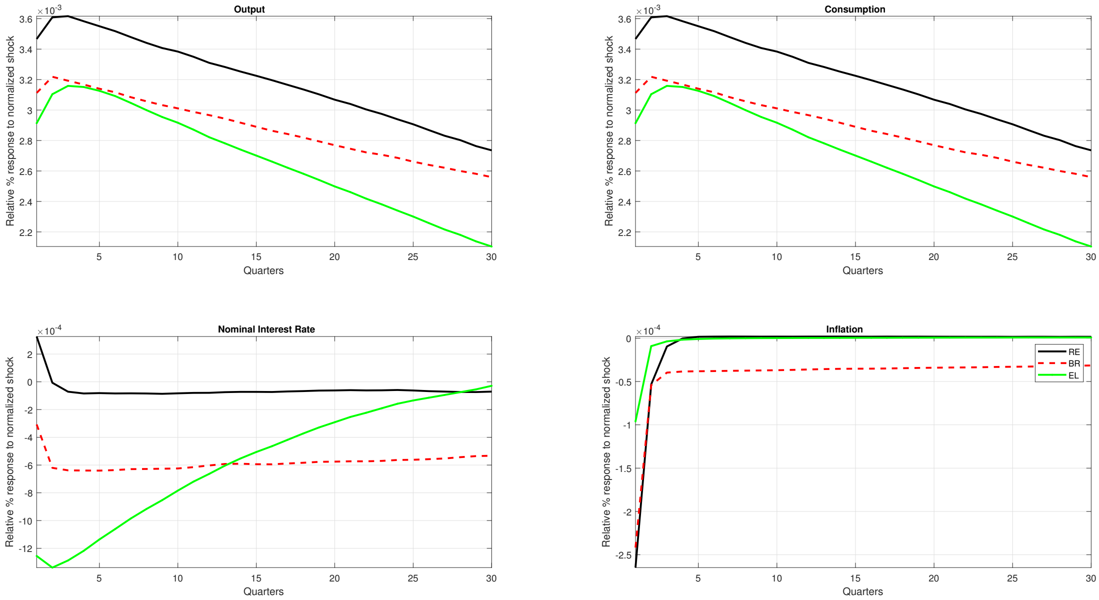

Figure 5: The impulse-response to the technology shock of the key variables at the optimized simple rule

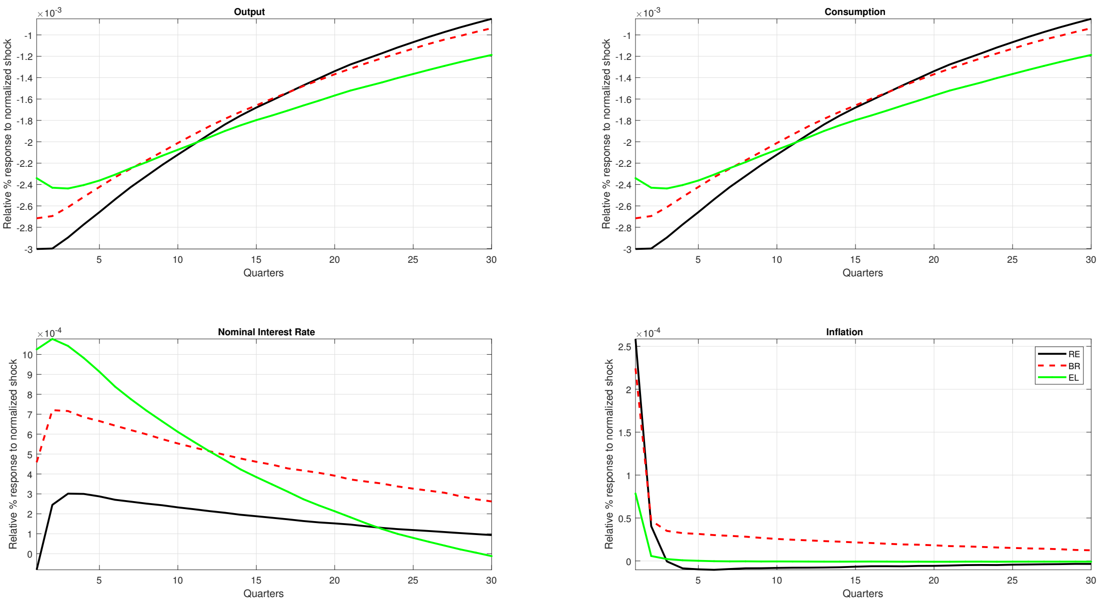

Figure 6: The impulse-response to the Marginal cost shock of the key variables at the optimized simple rule

	Optimized Simple Rule Across Models ($\bar{\Pi} = 1.0$)								
Models	ρ_r^*	α_π^*	α_y^*	α_{dy}^*	Π^*	Act welfare	CEV	p_zlb	w_r^*
RE	0.99	8.4	0.00	0.46	1.0	-2312.45	0	0.194	0
BR	0.34	16.0	0.04	0.6	1.0	-2624.06	0	0.308	0
EL(GAE)	0.88	65.9	0.0	0.98	1.0	-2601.12	0	0.344	0
Pool of models	0.8	14.06	0.01	0.6	1.0	-2456.57	0	0.262	0
	Optimized Simple Rule Across Models ($\bar{\Pi} = 1.005$)								
Models	ρ_r^*	α_π^*	α_y^*	α_{dy}^*	Π^*	Act welfare	CEV	p_zlb	w_r^*
RE	0.47	65.18	0.07	2.67	1.005	-2315.04	-0.1302	0.0431	0
BR	1.0	18.6	0.12	0.62	1.005	-2625.56	-0.0754	0.1385	0
EL(GAE)	0.99	38.65	0.35	0.48	1.005	- 2601.59	-0.0236	0.198	0
Pool of models	0.96	38.89	0.1	1.46	1.005	-2458.33	-0.0884	0.105	0
	Optimized Simple Rule Across Models ($\bar{\Pi} = 1.00799$)								
Models	ρ_r^*	α_π^*	α_y^*	α_{dy}^*	Π^*	Act welfare	CEV	p_zlb	w_r^*
RE	0.73	12.98	0.02	0.68	1.00799	-2319.77	-0.3678	0.0115	0
BR	1.0	27.17	0.19	0.59	1.00799	-2628.19	-0.2075	0.0775	0
EL(GAE)	0.99	60.42	0.85	0.69	1.00799	-2602.38	-0.0633	0.133	0
Pool of models	0.95	77.8	0.24	2.89	1.00799	-2461.52	-0.2487	0.06	0

Table 5: Optimized Rules without a ZLB

6.5.2 The ZLB Mandate

Now consider the imposition of a ZLB constraint that sets the probability of hitting the lower bound first at $\bar{p} = 0.05$ and then at the empirical value over the sample of $\bar{p} = 0.096$. The $\bar{p} = 0.096$ is calibrated from the sample of the shadow rate, by using the sample from 1957Q1 to 2017Q4, the ZLB episode is from 2010Q1 to 2015Q4 when the shadow rate experienced negative values. Figures 8, 9 and 10 show Stage 1 of the optimal robust ZLB mandate that imposes a probability of hitting the ZLB of $\bar{p} = 0.05$ for the RE, Gabaix bounded rationality (BR), Euler Learning (EL) models respectively.[15] Each figure shows how the choice of the penalty weight on nominal interest rate variability w_r drives down the steady-state target gross inflation rate Π from Stage 1 necessary to hit the probability constraint by lowering the standard deviation of the $R_{n,t}$ in the stochastic steady state (computed in a second-order perturbation solution). Also shown is actual welfare Ω_t converted to consumption equivalent variations (CEV) and the optimized parameter feedback on the inflation α_π from Stage 2. The optimal robust rule then picks the maximum value of CEV.

The full results for the optimized are set out in Table rule 6. Again there is a significant

[15]The corresponding results for the Anticipated Utility learning model (AU) is work in progress.

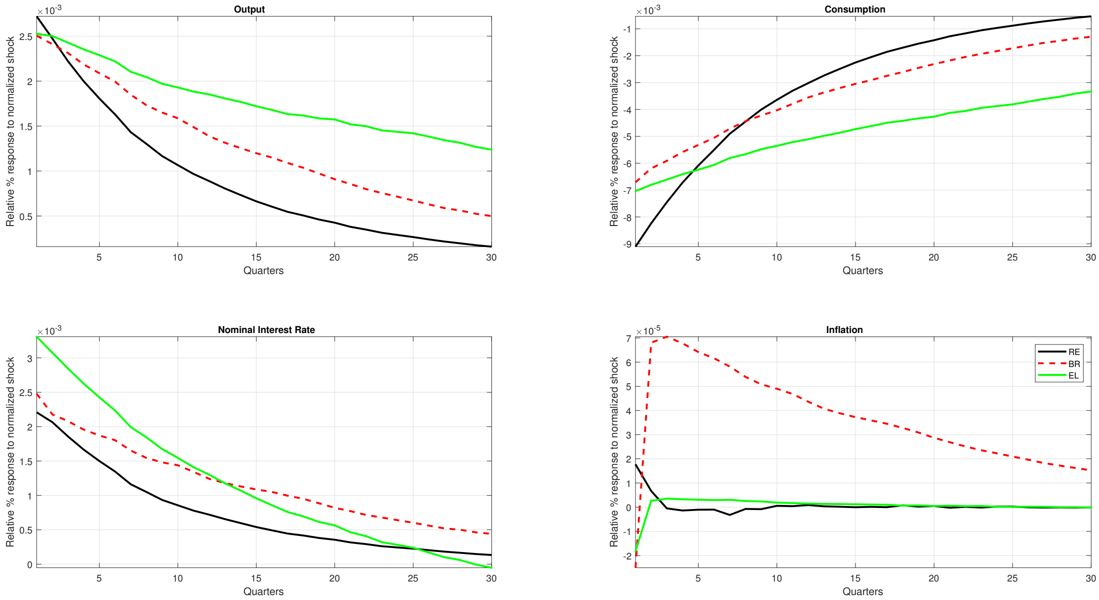

Figure 7: The impulse-response to the Government spending shock of the key variables at the optimized simple rule

difference of the equilibrium points between rules for individual models and the robust rule, but they all share the feature that $\rho^* = 1$.

	Optimal ZLB Mandate Across Models ($\bar{p} = 0.05$)								
Models	ρ_r^*	α_π^*	α_y^*	α_{dy}^*	Π^*	Act welfare	CEV	p_zlb	w_r^*
RE	1.0	0.8317	0.00	0.016	1.0026	-2313.38	-0.0467	0.05	50
BR	1.0	0.50	0.0	0.0	1.0083	-2629.27	-0.2618	0.05	70
EL(GAE)	1.0	98.87	1.1	0.02	1.014	-2605.14	-0.2020	0.05	30
Pool of models	1.0	2.17	0.0	0.0	1.0084	-2462.26	-0.2859	0.05	50
	Optimal ZLB Mandate Across Models ($\bar{p} = 0.096$)								
Models	ρ_r^*	α_π^*	α_y^*	α_{dy}^*	Π^*	Act welfare	CEV	p_zlb	w_r^*
RE	1.0	1.59	0.00	0.06	1.0016	-2312.76	-0.0156	0.096	20
BR	1.0	0.965	0.0	0.03	1.006	-2626.49	-0.1221	0.096	30
EL(GAE)	1.0	97.15	1.22	0.24	1.01	-2602.89	-0.0889	0.096	10
Pool of models	1.0	1.56	0.0	0.0	1.0048	-2458.29	-0.0864	0.096	70

Table 6: Optimized Rules with a ZLB

This leads us to consider the case where the monetary policymaker commits to a rule with $\rho_r = 1$ and $\alpha_y = \alpha_{dy} = 0$. Then integrating (67) and putting $\frac{\Pi_t}{\Pi} = \frac{P_t/P_{t-1}}{\bar{P}_t/\bar{P}_{t-1}}$ where $\bar{P}_t$

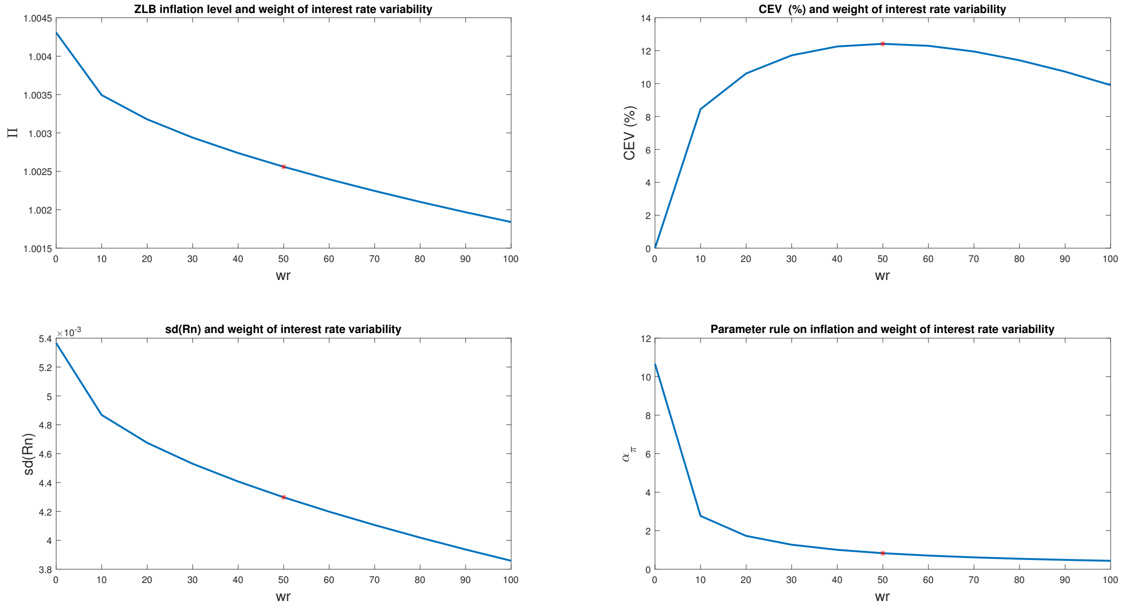

Figure 8: ZLB results for a single RE model. $\bar{p} = 0.05$

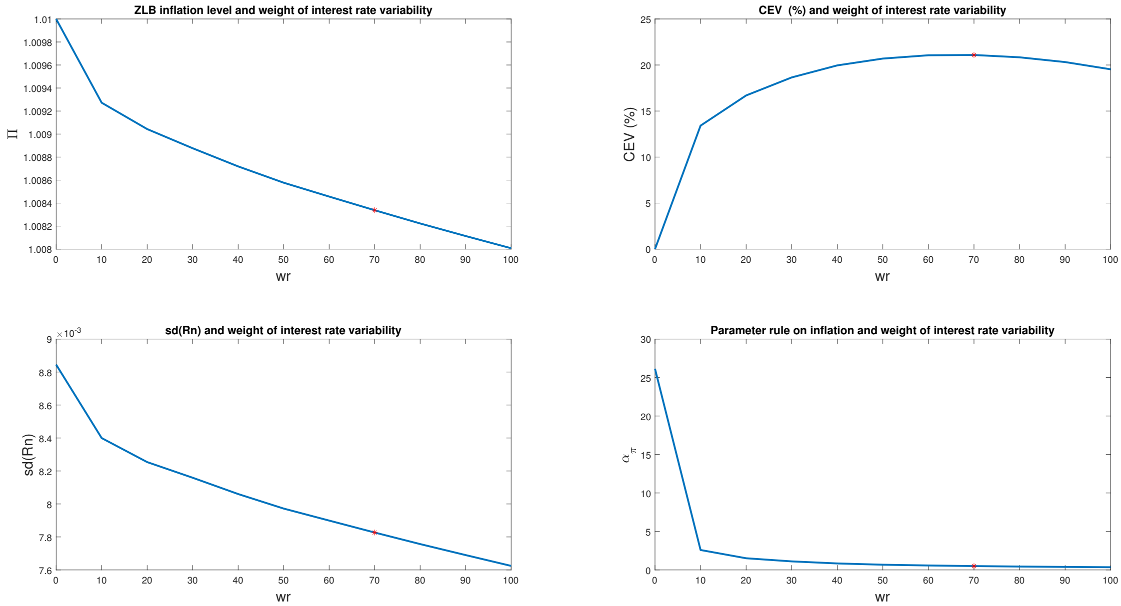

Figure 9: ZLB results for a single BR model. $\bar{p} = 0.05$

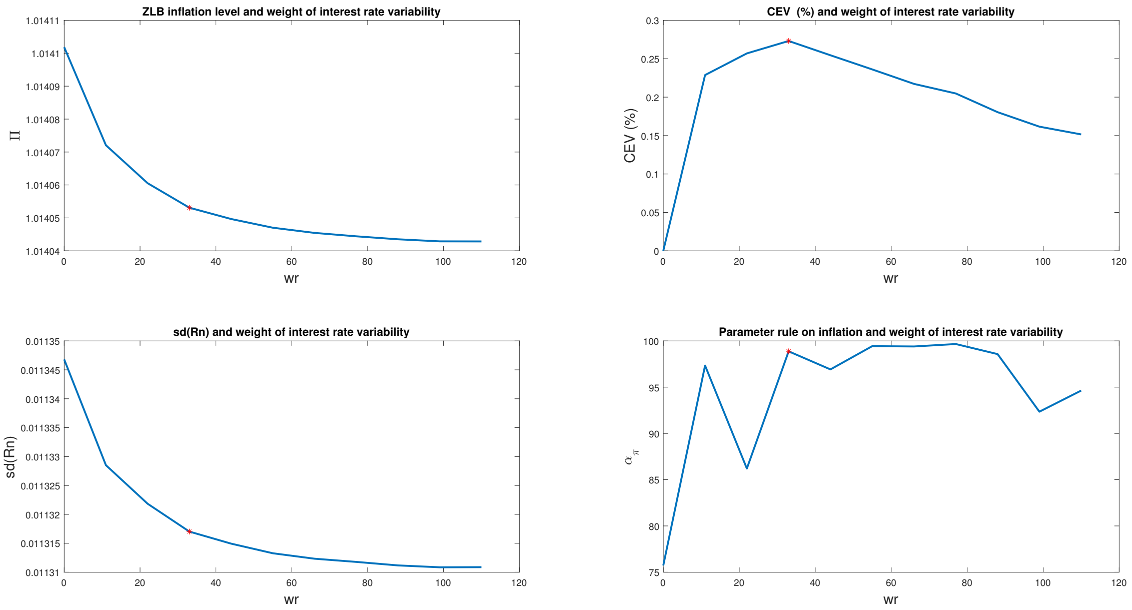

Figure 10: ZLB results for a single EL model with GAE. $\bar{p} = 0.05$

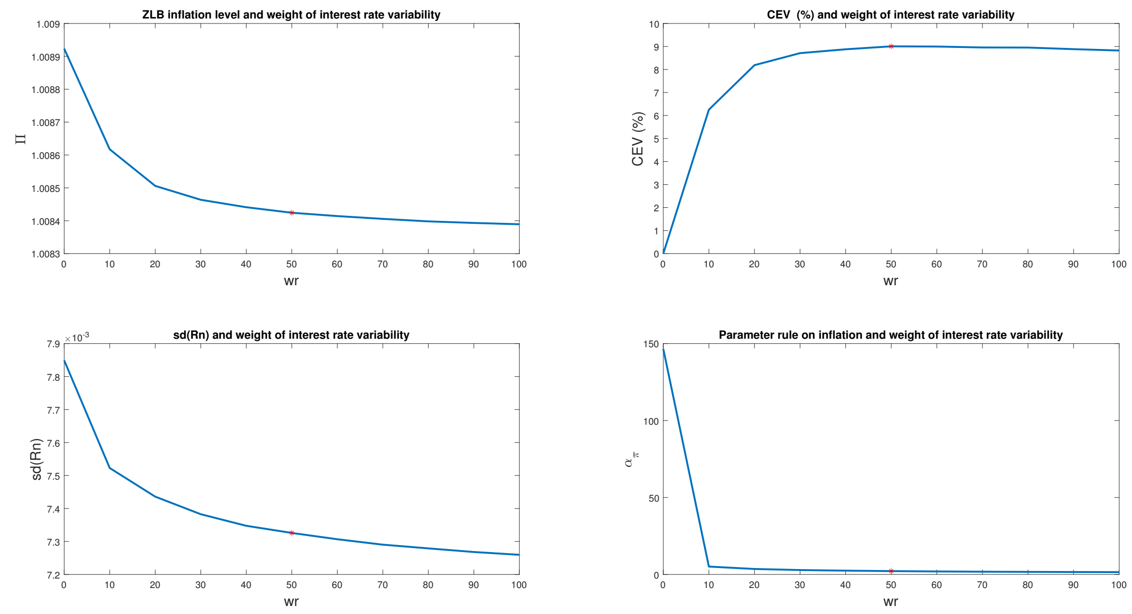

Figure 11: ZLB results for the pooled model (with four-period-ahead forecast. $\bar{p} = 0.05$

is the price trend in the constant inflation rate steady state, we arrive at the rule

$$\frac{R_{n,t}}{R_n} = \left(\frac{P_t}{\bar{P}_t}\right)^{\alpha_\pi} \tag{80}$$

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.

We have described the literature on the benefits of price-level targeting versus inflation targeting in Section 2 where we describe price-targeting (and indeed any inertial Taylor rule) as possible makeup strategy as follows. It anchors expectations 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.

Table 6 shows that to avoid the ZLB optimized rules must have $\rho_r^* = 1$ and $\alpha_y \approx 0$ and $\alpha_{dy}^* \approx 0$; i.e., they are close to a price-level rule. Indeed for the empirical probability $\bar{p} = 0.096$ the optimized rule is up to two decimal points exactly a price-level rule.

Table 7 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, the first row shows that if we use the robust simple rule optimized for the RE model in the BR and EL models, then the welfare actually increases to 0.23 and 1.42 percent of consumption respectively relative to that from the robust simple rules optimized for the latter two models themselves. However this comes at the expense of a frequent occurrence of the ZLB. The results show that incorrectly identifying the EL model as the data generating process implies the largest welfare costs. Using the rule optimized any one model imply rather large welfare losses. The final row shows the welfare cost of using the robust rule optimized for the prediction pool in Table 6 relative to the model specific robust optimal rules reported in Table 7. These now avoid these large costs of the single model optimized rules and the costs are generally small relative to the gains from using optimal rules.

	RE	BR	EL (GAE)
Opt_RE	0.00 ($p_{zlb} = 0.05$)	0.23 ($p_{zlb} = 0.158$)	1.42 ($p_{zlb} = 0.267$)
Opt_BR	-2.41 ($p_{zlb} = 0.0012$)	0.00 ($p_{zlb} = 0.05$)	1.33 ($p_{zlb} = 0.14$)
Opt_EL (GAE)	-1.39 ($p_{zlb} = 0.00028$)	-0.49 ($p_{zlb} = 0.015$)	0.00 ($p_{zlb} = 0.05$)
Opt_Pool	-0.48 ($p_{zlb} = 0.0034$)	0.026 ($p_{zlb} = 0.047$)	1.36 ($p_{zlb} = 0.127$)

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 RE model relative to the welfare attained using the robust simple rule optimized for the RE model if, for example, we use the robust simple rules optimized for the RE, BR, EL 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.

Table 7: Welfare *gains* of robust optimal ZLB mandate i (first column) in model j (first row), $j \neq i$

6.5.3 Impulse Responses with Optimized Rules

Our next comparison of rules, Figures 12–14, are the impulse responses to a monetary shock which compare the optimal policy for each of the three models and the robust counterpart, all taking into account the soft ZLB constraint. As before these emphasize the effect and importance of robustness especially for the model EL-GAE where the very aggressive monetary rule chosen for that assumed model is totally inappropriate for the other two.

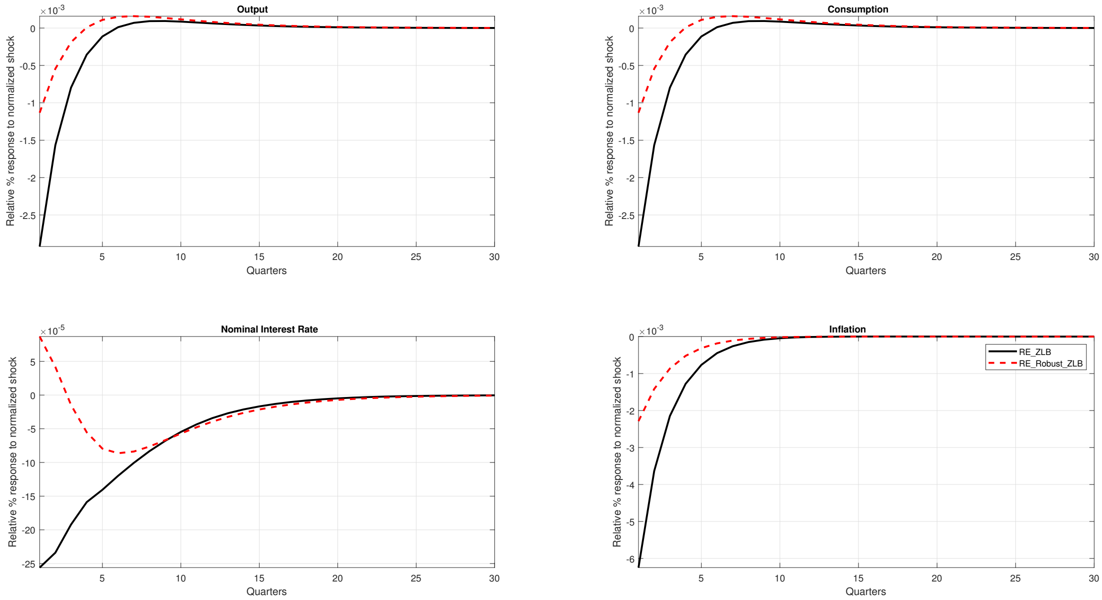

Figure 12: Impulse responses to the monetary policy shock comparison between the ZLB mandate produced by the individual RE model and robust ZLB mandate. $\bar{p} = 0.05$

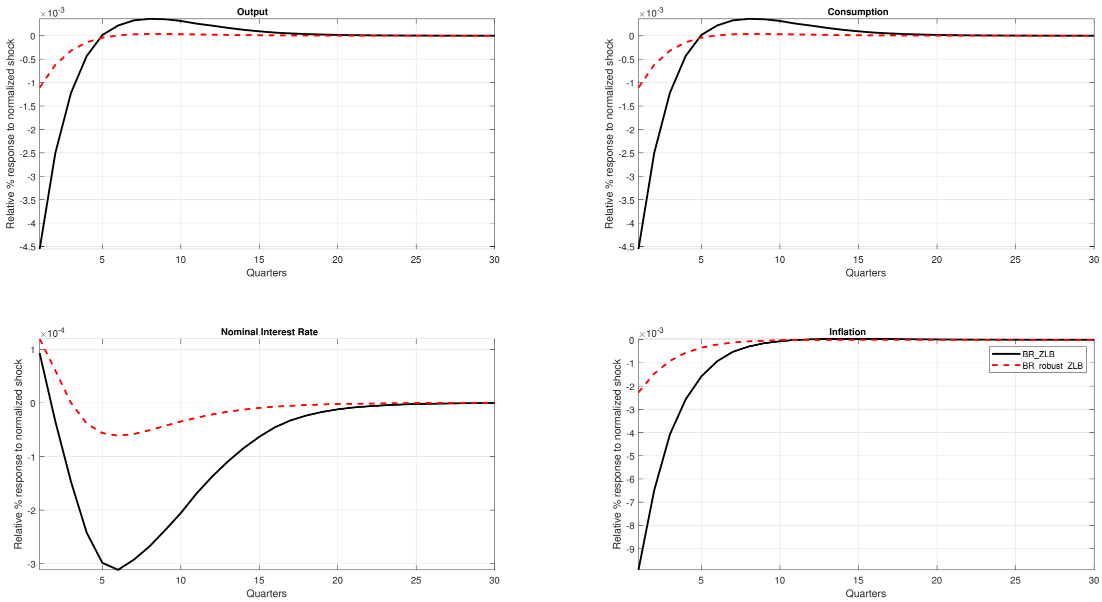

Figure 13: Impulse responses to the monetary policy shock comparison between the ZLB mandate produced by the individual BR model and robust ZLB mandate. $\bar{p} = 0.05$

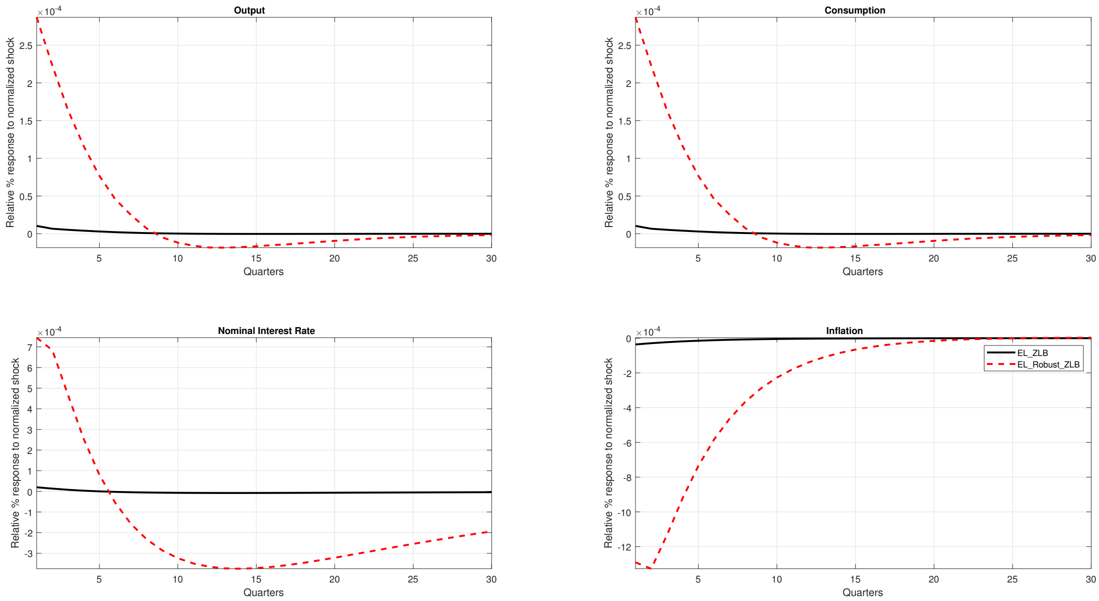

Figure 14: Impulse responses to the monetary policy shock comparison between the ZLB mandate produced by the individual EL model and robust ZLB mandate. $\bar{p} = 0.05$

Finally in Appendix E Figures 19–21 in the show how the responses to the cost-push, government spending and technology differs in the three states of the world under the robust rule with the ZLB mandates. These emphasize the fact that a very simple price-level rule responds in a flexible way to all three of these shocks.

7 Conclusions

This paper studies the problem of designing robust mandate framework when the policymaker 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 policymaker. We study the robust optimal policy problem in the form of an optimal mandate with optimized Taylor-type inertial nominal interest rate rule and the ZLB on the nominal interest rate under this weighting scheme using three estimated models exemplifying the policymakers' uncertainty about the expectation formations of the economic agents.

Our main results are: first, three models completely dominate model EL-SAE with weights $w_{RE} = 0.4$, $w_{EL-GAE} = 0.32$ and $w_{EL-BR} = 0.28$. Second, whereas Bayesian model averaging would design a welfare-optimized rule that hits the ZLB with a probability solely based on the Gabaix model, we find that our prediction pool using these weights choice has a significant impact on the robust optimized rule. Third, there are significant differences between the optimized rules for each model separately highlighting the need for seeking a robust rule. Fourth, we find that robust optimized rule found using optimal pooling weights is very close to the price level rule. This confirms good robustness properties of such a rule found in other studies. Finally to achieve a probability of hitting the ZLB constraint on the nominal interest rate of 5% per quarter, the robust optimal rule requires a target (steady-state) net inflation annual rate of between 3% and 4%.

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.

Regarding the wilderness we have alluded to the large number of competing behavioural models of which we have focused on only those with Euler learning and myopia. Future work could add a model with the **Anticipated Utility**(AU) approach aka **Infinite Horizon Learning**. AU assumes that agents follow an optimal decision rule conditional on their beliefs over aggregate states and prices. This takes into account all information available to the agent, and involves forecasts of variables external to them. See Eusepi and Preston (2011), Deak et al. (2015), Eusepi and Preston (2018) and Calvert Jump et al. (2019). Beliefs affect the data-generating process which in turn feeds back on beliefs. The fixed point of this process has been called self-confirming (unlike the beliefs in the form of heuristic rules in our EL approach). Self-confirming learning equilibria in the form of parsimonious first-order VAR to fit mean and persistence of each state variable to data are also studied by Hommes and Zhu (2014), Hommes and Zhu (2015), Hommes et al. (2022): Anticipated Utility learning is similar to, but distinct from, the **Internal Rationality** approach in which agents, "maximize utility under uncertainty, given their constraints and given a consistent set of probability beliefs about payoff-relevant variables that are beyond their control or external" (Adam and Marcet, 2011). The approach of Adam and Marcet (2011) requires a fully Bayesian plan for beliefs, as opposed to the anticipated utility approach, in which households do not consider the possibility that their beliefs might change in the future. The latter is obviously more straightforward than the former, although Cogley and Sargent (2008) demonstrate that the anticipated utility approach can be seen as a good approximation to the fully Bayesian approach. Finally an alternative approach is the k-level learning of Woodford (2013), Garcia-Schmidt and Woodford (2019), Farhi and Werning (2019) where beliefs are updated iteratively with observed temporary equilibrium over n stages. All these models of non-rational beliefs are candidates for a pooling and robust policy exercise of the type offered in our paper.

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.

Adam, K. and Marcet, A. (2011). Internal Rationality, Imperfect Market Knowledge and Asset Prices. *Journal of Economic Theory*, 146(3):1224–1252.

Adam, K. and Woodford, M. (2012). Robustly Optimal Monetary Policy in a Micro-Founded New Keynesian Model. *Journal of Monetary Economics*, 59:468–487.

Adam, K. and Woodford, M. (2020). Robustly Optimal Monetary Policy in a New Keynesian Model with Housing. CESifo Working Paper No. 8127.

Anufriev, M., Hommes, C., and Makarewicz, T. (2019). Simple Forecasting Heuristics that Make us Smart: Evidence from Different Market Experiments. *Journal of the European Economic Association* , 17:1538–1584.

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.

Branch, W. A. and McGough, B. (2018). Heterogeneous Expectations and Micro-Foundations in Macroeconomics. In *Handbook of Computational Economics 4*. Elsevier.

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

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

Calvert Jump, R., Hommes, C., and Levine, P. (2019). Learning, Heterogeneity, and Complexity in the New Keynesian model. *Journal of Economic Behavior and Organization*, 166(C):446–470.

Calvert Jump, R. and Levine, P. (2019). Behavioural New Keynesian Models. *Journal of Macroeconomics*, 59(C):58–77.

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., 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.

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. and Sargent, T. J. (2008). Anticipated utility and rational expectations as approximations of bayesian decision making. *International Economic Review*, 49(1):185–221.

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.

De Grauwe, P. (2011). Animal spirits and monetary policy. *Economic Theory*, 47(2-3):423–457.

De Grauwe, P. (2012a). Booms and Busts in Economic Activity: A Behavioral Explanation. *Journal of Economic Behavior and Organization*, 83(3):484–501.

De Grauwe, P. (2012b). *Lectures on Behavioral Macroeconomics*. Princeton University Press.

Deak, S., Levine, P., Mirza, A., and Pearlman, J. (2020). Is Price Level Targeting a Robust Monetary Rule? Discussion Papers 20-27, Department of Economics, University of Birmingham.

Deak, S., Levine, P., and Yang, B. (2015). A New Keynesian Behavioural Model with Internal Rationality and Heterogeneous Agents. Presented at the 2015 Annual CEF Conference, June, Taiwan.

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.

Den Haan, W. and Wind, J. D. (2012). Nonlinear and Stable Perturbation Based Approximation. *Journal of Economic Dynamics and Control*, 36(10):1477–1479.

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

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimal product diversity. *American Economic Review*, 67(3):297–308.

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

Eusepi, S. and Preston, B. (2011). Expectations, learning, and business cycle fluctuations. *American Economic Review*, 101(6):2844–2872.

Eusepi, S. and Preston, B. (2016). The science of monetary policy: An imperfect knowledge perspective. Federal Reserve Bank of New York Satff Reports, no. 782.

Eusepi, S. and Preston, B. (2018). Fiscal Foundations of Inflation. *American Economic Review* , 108(9):2551–2589.

Evans, G. W. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press.

Evans, G. W. and Honkapohja, S. (2009). Learning and Macroeconomics. *Annual Review of Economics*, 1:421–449.

Farhi, E. and Werning, I. (2019). Monetary Policy, Bounded Rationality, and Incomplete Markets. *American Economic Review* , 109(11):3887–3928.

Gabaix, X. (2020). A Behavioral New Keynesian Model. *American Economic Review*, 110(8):2271–2327.

Gali, J. (2015). *Monetary Policy, Inflation and the Business Cycle*. Princeton University Press, second edition edition.

Garcia-Schmidt, M. and Woodford, M. (2019). Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis. *American Economic Review* , 109(1):86–120.

Gaspar, V., Smets, F., and Vestin, D. (2010). Is Time Ripe for Price Level Path Stability? In Siklos, P. L., Bohl, M. T., and Wohar, M. E., editors, *Challenges in central banking: the current institutional environment and forces affecting monetary policy*. Cambridge University Press.

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*, 23:1–13.

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

Hebden, J., Herbst, E. P., Tangm J. anf Topa, G., and Winkler, F. (2020). How robust are make-up strategies to key alternative assumptions? Finance and economics discussion series 2020-069, Federal Reserve Board.

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

Hommes, C., Mavromatis, C., Tolga Ozden, T., and Zhu, M. (2022). Behavioral Learning Equilibria in New Keynesian Models. Bank of Canada Staff Working Paper number 2022-42.

Hommes, C. and Zhu, M. (2014). Behavioural learning equilibria. *Journal of Economic Theory*, 150:778–814.

Hommes, C. and Zhu, M. (2015). Behavioral Learning Equilibria for the New Keynesian Model. Presented at the CIMS Workshop on Agent-Based and DSGE Macroeconomic Modelling: Bridging the Gap, November 206, 2015, University of Surrey .

Ilbas, P., Roisland, O., and Sveen, T. (2013). Robustifying Optimal Monetary Policy using Simple Rules as Cross-checks. Working Paper 2012/22, Norges Bank.

Iskrev, N. (2010). Local identification in dsge models. *Journal of Monetary Economics*, 57:189–202.

Iskrev, N. and Ratto, M. (2010). Analysing identification issues in DSGE models. mimeo.

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

Komunjer, I. and Ng, S. (2011). Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica*, 79(6):1995–2032.

Levin, A., Onatski, A., Williams, J. C., and Williams, N. (2006). Monetary Policy Under Uncertainty in Micro-Founded Macroeconomic Models. in M. Gertler and K. Rogoff (eds.), *NBER Macroeconomics Annual*, 2005, pp 229–387 .

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., 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. (2008). Quantifying and Sustaining Welfare Gains from Monetary Commitment. *Journal of Monetary Economics*, 55(7):1253–1276.

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

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

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.

Powell, J. H. (2020). New Economic Challenges and the Fed's Monetary Policy Review. Speech given at "Navigating the Decade Ahead: Implications for Monetary Policy," an economic policy symposium sponsoredthe Federal Reserve Bank of Kansas City, Jackson Hole.

Qu, Z. and Tkachenko, D. (2012). Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models. *Quantitative Economics*, 3(1):95–132.

Ratto, M. (2008). Analyzing DSGE Models with Global Sensitivity Analysis. *Computational Economics*, 115:115–139.

Reiss, R. (2009). Optimal Monetary Policy Rules in an Estimated Sticky-Information Model. *American Economic Journal: Macroeconomics*, 1(2):1–28.

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

Sargent, T. (2008). Evolution and intelligent design. *American Economics Review*, 98(1):5–37.

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. (1980). Macroeconomics and Reality. *Econometrica*, 48:1–48.

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. (2003). An estimated Stochastic Dynamic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association*, 1(5):1123–1175.

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.

Svensson, L. E. O. (2020). Monetary policy strategies for the federal reserve.

Taylor, J. B. and Williams, J. C. (2010). Simple and Robust Rules for Monetary Policy. In Friedman, B. M. and Woodford, M., 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.

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.

Woodford, M. (2013). Macroeconomic Analysis Without the Rational Expectations Hypothesis. *Annual Review of Economics, Annual Reviews*, 5(1):303–346.

Woodford, M. (2018). Monetary Policy Analysis when Planning Horizons are Finite. In *NBER Macroeconomics Annual 2018*, volume 33. National Bureau of Economic Research, Inc.

Wu, J. C. and Xia, F. D. (2016). Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. *Journal of Money Credit and Banking*, 48(2-3):253–291.

Appendices

A The Rational Expectations Model

A.1 Equilibrium

A symmetric equilibrium is determined by the following equations:

$$U_t = \log(C_t) - \kappa \frac{H_t^{1+\phi}}{1+\phi} \tag{A.1}$$

$$V_t = \mathbb{E}_t \left[\sum_{s=0}^{\infty} \beta^s U_{t+s} \right] = U_t + \beta \mathbb{E}_t V_{t+1} \tag{A.2}$$

$$U_{C,t} = \frac{1}{C_t} \tag{A.3}$$

$$U_{H,t} = -\kappa H_t^{\phi} \tag{A.4}$$

$$\Lambda_{t,t+1} = \beta \frac{U_{C,t+1}}{U_{C,t}} \tag{A.5}$$

$$R_t = \frac{R_{n,t-1}}{\Pi_t} \tag{A.6}$$

$$1 = \mathbb{E}_t\left[\Lambda_{t,t+1} R_{t+1}\right] \tag{A.7}$$

$$W_t = -\frac{U_{H,t}}{U_{C,t}} \tag{A.8}$$

$$Y_t^W = A_t H_t^\alpha \tag{A.9}$$

$$W_t = \alpha \frac{P_t^W}{P_t} \frac{Y_t^W}{H_t} \tag{A.10}$$

$$MC_t = \frac{P_t^W}{P_t} \tag{A.11}$$

$$J_t = \frac{1}{1-\frac{1}{\zeta}} Y_t MC_t MS_t + \xi \mathbb{E}_t \Lambda_{t,t+1} \Pi_{t,t+1}^{\zeta} J_{t+1} \tag{A.12}$$

$$JJ_t = Y_t + \xi \mathbb{E}_t \Lambda_{t,t+1} \Pi_{t,t+1}^{\zeta-1} JJ_{t+1} \tag{A.13}$$

$$1 = \xi \Pi_t^{\zeta-1} + (1-\xi)\left(\frac{J_t}{JJ_t}\right)^{1-\zeta} \tag{A.14}$$

$$Y_t = \frac{Y_t^W}{\Delta_t} \tag{A.15}$$

$$\Delta_t = \xi \Pi_t^{\zeta} \Delta_{t-1} + (1-\xi)\left(\frac{J_t}{JJ_t}\right)^{-\zeta} \tag{A.16}$$

$$Y_t = C_t + G_t \tag{A.17}$$

$$\begin{aligned} \log\left(\frac{R_{n,t}}{R_n}\right) &= \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) \\ &\quad + (1-\rho_r)\left(\theta_\theta \log\left(\frac{\Pi_t}{\Pi}\right) + \theta_y \log\left(\frac{Y_t}{Y}\right)\right) + \log MPS_t \end{aligned} \tag{A.18}$$

$$\log A_t - \log A = \rho_A(\log A_{t-1} - \log A) + \epsilon_{A,t} \tag{A.19}$$

$$\log MS_t - \log MS = \rho_{MS}(\log MS_{t-1} - \log MS) + \epsilon_{MS,t} \tag{A.20}$$

$$\log MPS_t - \log MPS = \rho_{MPS}(\log MPS_{t-1} - \log MPS) + \epsilon_{MPS,t} \tag{A.21}$$

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

where we have introduced a mark-up shock MS_t.

A.2 Stationary equilibrium

Labour-augmenting technical progress parameter is decomposed into a cyclical component, A_t^c, and a deterministic trend $\bar{A}_t$:

$$A_t = \bar{A}_t A_t^c$$
$$\bar{A}_t = (1+g)\bar{A}_{t-1}$$

Rewrite the equilibrium conditions as

$$U_t - \log(\bar{A}_t) = \log(C_t/\bar{A}_t) - \kappa\frac{H_t^{1+\phi}}{1+\phi} \tag{A.23}$$
$$V_t = U_t + \beta\mathbb{E}_t V_{t+1} \tag{A.24}$$
$$\bar{A}_t U_{C,t} = \frac{1}{C_t/\bar{A}_t} \tag{A.25}$$
$$U_{H,t} = -\kappa H_t^{\phi} \tag{A.26}$$
$$\Lambda_{t,t+1} = \frac{\beta}{\bar{A}_{t+1}/\bar{A}_t}\frac{\bar{A}_{t+1}U_{C,t+1}}{\bar{A}_t U_{C,t}} \tag{A.27}$$
$$R_t = \frac{R_{n,t-1}}{\Pi_t} \tag{A.28}$$
$$1 = \mathbb{E}_t\left[\Lambda_{t,t+1}R_{t+1}\right] \tag{A.29}$$
$$\frac{W_t}{\bar{A}_t} = -\frac{U_{H,t}}{\bar{A}_t U_{C,t}} \tag{A.30}$$
$$\frac{Y_t^W}{\bar{A}_t} = \frac{A_t}{\bar{A}_t}H_t^{\alpha} \tag{A.31}$$
$$\frac{W_t}{\bar{A}_t} = \alpha\frac{P_t^W}{P_t}\frac{Y_t^W/\bar{A}_t}{H_t} \tag{A.32}$$
$$MC_t = \frac{P_t^W}{P_t} \tag{A.33}$$
$$\frac{J_t}{\bar{A}_t} = \frac{1}{1-\frac{1}{\zeta}}\frac{Y_t}{\bar{A}_t}MC_t MS_t + \xi\mathbb{E}_t\frac{\bar{A}_{t+1}}{\bar{A}_t}\Lambda_{t,t+1}\Pi_{t,t+1}^{\zeta}\frac{J_{t+1}}{\bar{A}_{t+1}} \tag{A.34}$$
$$\frac{JJ_t}{\bar{A}_t} = \frac{Y_t}{\bar{A}_t} + \xi\mathbb{E}_t\frac{\bar{A}_{t+1}}{\bar{A}_t}\Lambda_{t,t+1}\Pi_{t,t+1}^{\zeta-1}\frac{JJ_{t+1}}{\bar{A}_{t+1}} \tag{A.35}$$
$$1 = \xi\Pi_t^{\zeta-1} + (1-\xi)\left(\frac{J_t/\bar{A}_t}{JJ_t/\bar{A}_t}\right)^{1-\zeta} \tag{A.36}$$
$$\frac{Y_t}{\bar{A}_t} = \frac{Y_t^W/\bar{A}_t}{\Delta_t} \tag{A.37}$$

$$\Delta_t = \xi \Pi_t^{\zeta} \Delta_{t-1} + (1-\xi)\left(\frac{J_t/\bar{A}_t}{JJ_t/\bar{A}_t}\right)^{-\zeta} \tag{A.38}$$

$$\frac{Y_t}{\bar{A}_t} = \frac{C_t}{\bar{A}_t} + \frac{G_t}{\bar{A}_t} \tag{A.39}$$

$$\begin{aligned}\log\left(\frac{R_{n,t}}{R_n}\right) &= \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) \\ &+ (1-\rho_r)\left(\theta_\theta \log\left(\frac{\Pi_t}{\Pi}\right) + \theta_y \log\left(\frac{Y_t}{Y}\right)\right) + \log MPS_t\end{aligned} \tag{A.40}$$

$$\log A_t - \log A = \rho_A(\log A_{t-1} - \log A) + \epsilon_{A,t} \tag{A.41}$$

$$\log MS_t - \log MS = \rho_{MS}(\log MS_{t-1} - \log MS) + \epsilon_{MS,t} \tag{A.42}$$

$$\log MPS_t - \log MPS = \rho_{MPS}(\log MPS_{t-1} - \log MPS) + \epsilon_{MPS,t} \tag{A.43}$$

$$\log G_t - \log G = \rho_G(\log G_{t-1} - \log G) + \epsilon_{G,t} \tag{A.44}$$

Use change of variables to arrive to the following equilibrium conditions:[16]

$$U_t^c = \log(C_t^c) - \kappa \frac{H_t^{1+\phi}}{1+\phi} \tag{A.45}$$

$$V_t^c = U_t^c + \beta \mathbb{E}_t V_{t+1}^c \tag{A.46}$$

$$U_{C,t}^c = \frac{1}{C_t^c} \tag{A.47}$$

$$U_{H,t} = -\kappa H_t^{\phi} \tag{A.48}$$

$$\Lambda_{t,t+1} = \frac{\beta}{1+g} \frac{U_{C,t+1}^c}{U_{C,t}^c} \tag{A.49}$$

$$R_t = \frac{R_{n,t-1}}{\Pi_t} \tag{A.50}$$

$$1 = \mathbb{E}_t\left[\Lambda_{t,t+1} R_{t+1}\right] \tag{A.51}$$

$$W_t^c = -\frac{U_{H,t}}{U_{C,t}^c} \tag{A.52}$$

$$Y_t^{W,c} = A_t^c H_t^{\alpha} \tag{A.53}$$

$$W_t^c = \alpha \frac{P_t^W}{P_t} \frac{Y_t^{W,c}}{H_t} \tag{A.54}$$

$$MC_t = \frac{P_t^W}{P_t} \tag{A.55}$$

$$J_t^c = \frac{1}{1-\frac{1}{\zeta}} Y_t^c MC_t MS_t + \xi(1+g)\mathbb{E}_t \Lambda_{t,t+1} \Pi_{t,t+1}^{\zeta} J_{t+1}^c \tag{A.56}$$

[16] The first equation is based on a hunch. Since the normalization of utility is additive, we cannot have a different discount factor. However, we cannot derive the first equation above from (A.2). We can derive it starting from the definition $V_t^c = \mathbb{E}_t\left[\sum_{s=0}^{\infty} \beta^s U_{t+s}^c\right]$.

$$JJ_t^c = Y_t^c + \xi(1+g)\mathbb{E}_t\Lambda_{t,t+1}\Pi_{t,t+1}^{\zeta-1}JJ_{t+1}^c \tag{A.57}$$

$$1 = \xi\Pi_t^{\zeta-1} + (1-\xi)\left(\frac{J_t^c}{JJ_t^c}\right)^{1-\zeta} \tag{A.58}$$

$$Y_t^c = \frac{Y_t^{W,c}}{\Delta_t} \tag{A.59}$$

$$\Delta_t = \xi\Pi_t^{\zeta}\Delta_{t-1} + (1-\xi)\left(\frac{J_t^c}{JJ_t^c}\right)^{-\zeta} \tag{A.60}$$

$$Y_t^c = C_t^c + G_t^c \tag{A.61}$$

$$\begin{aligned}\log\left(\frac{R_{n,t}}{R_n}\right) &= \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) \\ &\quad + (1-\rho_r)\left(\theta_\theta \log\left(\frac{\Pi_t}{\Pi}\right) + \theta_y \log\left(\frac{Y_t^c}{Y^c}\right)\right) + \log MPS_t\end{aligned} \tag{A.62}$$

$$\log A_t^c - \log A^c = \rho_A(\log A_{t-1}^c - \log A^c) + \epsilon_{A,t} \tag{A.63}$$

$$\log MS_t - \log MS = \rho_{MS}(\log MS_{t-1} - \log MS) + \epsilon_{MS,t} \tag{A.64}$$

$$\log MPS_t - \log MPS = \rho_{MPS}(\log MPS_{t-1} - \log MPS) + \epsilon_{MPS,t} \tag{A.65}$$

$$\log G_t^c - \log G^c = \rho_G(\log G_{t-1}^c - \log G^c) + \epsilon_{G,t} \tag{A.66}$$

This is a system of 22 equation in the following 22 "variables" (in order of appearance): V^c, U^c, C^c, H, Λ, R, W^c, U_H^c, U_C^c, $Y^{W,c}$, A^c, $\frac{P^W}{P}$, J^c, Y^c, MC, MS, Π, JJ^c, Δ, G^c, R_n, MPS.

A.3 Steady State

The exogenous variables have steady states $A^c = MS = MPS = 1$. Given the steady state inflation rate Π and the steady state nominal interest rate Rn, the steady state values of the other variables can be computed as

$$\begin{aligned}
&\text{(A.49)} &&\Rightarrow && \Lambda = \frac{\beta}{1+g} \\
&\text{(A.51)} &&\Rightarrow && R = \frac{1}{\Lambda} \\
&\text{(A.58)} &&\Rightarrow && \frac{J^c}{JJ^c} = \left(\frac{1-\xi\Pi^{\zeta-1}}{1-\xi}\right)^{\frac{1}{1-\zeta}} \\
&\text{(A.56), (A.57)} &&\Rightarrow && MC = \left(1-\frac{1}{\zeta}\right)\frac{J^c}{JJ^c}\frac{1-\xi\beta\Pi^{\zeta}}{1-\xi\beta\Pi^{\zeta-1}} \\
&\text{(A.60)} &&\Rightarrow && \Delta = \frac{(1-\xi)\left(\frac{J^c}{JJ^c}\right)^{-\zeta}}{1-\xi\Pi^{\zeta}}
\end{aligned}$$

$$(A.54), \text{using } (A.47), \quad (A.48) \quad , (A.52), (A.55), (A.59), (A.61) \Rightarrow H = \left(\frac{\alpha \Delta MC}{\kappa(1-gy)}\right)^{\frac{1}{1+\phi}}$$

$$\begin{aligned}
(A.53) \quad & \Rightarrow \quad Y^{W,c} = (A^c H)^{\alpha} \\
(A.59) \quad & \Rightarrow \quad Y^c = \frac{Y^{W,c}}{\Delta} \\
& \qquad\quad G^c = gy * Y^c \\
(A.61) \quad & \Rightarrow \quad C^c = Y^c - G^c \\
(A.56) \quad & \Rightarrow \quad J^c = \frac{Y^c MCMS}{(1-\frac{1}{\zeta})(1-\xi\beta\Pi^{\zeta})} \\
(A.57) \quad & \Rightarrow \quad JJ^c = \frac{Y^c}{(1-\xi\beta\Pi^{\zeta-1})} \\
(A.45) \quad & \Rightarrow \quad U^c = \log(C^c) - \kappa\frac{H^{1+\phi}}{1+\phi} \\
(A.47) \quad & \Rightarrow \quad U^c_{C^c} = \frac{1}{C^c} \\
(A.48) \quad & \Rightarrow \quad U_H = -\kappa H^{\phi} \\
(A.55) \quad & \Rightarrow \quad \frac{P^W}{P} = MC \\
(A.54) \quad & \Rightarrow \quad W^c = \alpha\frac{P^W}{P}\frac{Y^{W,c}}{H} \\
(A.46) \quad & \Rightarrow \quad V^c = \frac{U^c}{1-\beta}
\end{aligned}$$

Finally we can define

$$\begin{aligned}
CEquiv_t &= \mathbb{E}_t\left[\sum_{t=s}^{\infty}\beta^s U(1.01C_{t+s}, H_{t+s})\right] - \mathbb{E}_t\left[\sum_{t=s}^{\infty}\beta^s U(C_{t+s}, H_{t+s})\right] \\
&= \mathbb{E}_t\left[\sum_{t=s}^{\infty}\beta^s\left\{\log(1.01C^c_{t+s}) - \kappa\frac{H^{1+\phi}_{t+s}}{1+\phi} - \log(C^c_{t+s}) - \kappa\frac{H^{1+\phi}_{t+s}}{1+\phi}\right\}\right] \\
&= \log(1.01)\sum_{t=s}^{\infty}\beta^s = \frac{\log(1.01)}{1-\beta}
\end{aligned}$$

A.4 Limits on Π and ξ in the Steady state

Non-negativity constraints imply the following conditions

$$\xi\Pi^{\zeta-1} < 1 \tag{A.67}$$

$$\xi\beta\Pi^{\zeta-1} < 1 \tag{A.68}$$

$$\xi\beta\Pi^{\zeta} < 1 \tag{A.69}$$

If we confine ourselves to a non-negative net inflation steady state ($\Pi \geq 1$) then a sufficient condition for (A.67)–(A.69) to hold is $\xi\Pi^{\zeta} < 1$. This places an upper-bound on steady-state inflation given by

$$\Pi < \left(\frac{1}{\xi}\right)^{\frac{1}{\zeta}} \tag{A.70}$$

With $\zeta = 7$, for quarterly settings $\xi = 0.5, 0.75, 0.8, 0.8$ these gives upper bounds $\Pi = 1.104, 1.042, 1.032, 1.015$. So the constraint is only important for very high degrees of price stickiness.

A.5 The Measurement Equations

Our 4 observables are: output growth (`dyobs`) defined in various ways, hours worked per capita (`labobs`), monetary policy rate (`robs`), inflation rate (`pinfobs`). The corresponding measurement equations are:

$$\texttt{dyobs} = \log\left((1+g)\frac{Y_t^c}{Y_{t-1}^c}\right) \tag{A.71}$$

$$\texttt{labobs} = \frac{H_t - H}{H} \tag{A.72}$$

$$\texttt{robs} = R_{n,t} - 1 \tag{A.73}$$

$$\texttt{pinfobs} = \Pi_t - 1 \tag{A.74}$$

The steady state values of the observables are $\texttt{dyobs} = \texttt{dcobs} = \texttt{dyobs} = \log(1+g)$, $\texttt{labobs} = H$, $\texttt{robs} = R_n - 1$, and $\texttt{pinfobs} = \Pi - 1$.

The estimated parameters $\bar{\Pi}$, $\bar{R}_n$ and $\bar{g}$ are related to the steady state variables of our model by

$$\Pi = \frac{\bar{\Pi}}{100} + 1$$
$$R_n = \frac{\bar{R}_n}{100} + 1$$
$$g = \frac{\bar{g}}{100}$$

From our non-zero-inflation-growth steady state this implies that we should impose the restrictions

$$R_n = \frac{\Pi}{\beta(1+g)^{-1}} = \frac{\bar{R}_n}{100} + 1 \tag{A.75}$$

on β rather than calibrating it at the usual $\beta = 0.99$. This implies that β is calibrated as

$$\beta = \frac{\frac{\bar{\Pi}}{100} + 1}{\left(\frac{\bar{R}_n}{100} + 1\right)\left(1 + \frac{\bar{g}}{100}\right)^{-1}} = 0.9995 \tag{A.76}$$

For the given empirical steady state inflation $\bar{\Pi} = 1.00799$.

B Gabaix Model

B.1 Model's equilibrium conditions

$$\frac{C_t^c}{1-\beta} = \frac{Z_t}{[\kappa C_t^c]^{\frac{1}{\phi}}} + ZZ_t \tag{B.1}$$

$$Z_t = (W_t^c)^{1+\frac{1}{\phi}} + \left(\frac{(1+g)^{1+\frac{1}{\phi}}}{\beta^{\frac{1}{\phi}}}\right)\left(\frac{\mathbb{E}_t\left(Z + \bar{m}_h \hat{Z}_{t+1}\right)}{\mathbb{E}_t\left(\frac{R_{n,t}}{\Pi + \bar{m}_h \hat{\Pi}_{t+1}}\right)^{1+\frac{1}{\phi}}}\right) \tag{B.2}$$

$$ZZ_t = (\Gamma_t^c - T_t^c) + (1+g)\left(\frac{\mathbb{E}_t\left(ZZ + \bar{m}_h \hat{ZZ}_{t+1}\right)}{\mathbb{E}_t\left(\frac{R_{n,t}}{\Pi + \bar{m}_h \hat{\Pi}_{t+1}}\right)}\right) \tag{B.3}$$

$$W_t^c = \kappa H_t^\sigma C_t^c \tag{B.4}$$

$$W_t^c = \alpha \frac{P_t^W}{P_t} \frac{Y_t^{W,c}}{H_t} \tag{B.5}$$

$$MC_t = \frac{P_t^W}{P_t} \tag{B.6}$$

$$Y_t^{W,c} = A_t^c H_t^\alpha \tag{B.7}$$

$$Y_t^c = \frac{Y_t^{W,c}}{\Delta_t} \tag{B.8}$$

$$Y_t^c = C_t^c + G_t^c \tag{B.9}$$

$$G_t^c = T_t^c \tag{B.10}$$

$$\Gamma_t^c = Y_t^c - \alpha \frac{P_t^W}{P_t} Y_t^{W,c} \tag{B.11}$$

$$\Delta_t = \xi \Pi_t^\zeta \Delta_{t-1} + (1-\xi)\left(\frac{J_t^c}{JJ_t^c}\right)^{-\zeta} \tag{B.12}$$

$$
\begin{aligned}
J_t^c &= \frac{1}{1-\frac{1}{\zeta}} Y_t^c MC_t \\
&+ \xi \mathbb{E}_t \frac{(\Pi + \bar{m}_f \hat{\Pi}_{t+1})^{\zeta}}{\mathbb{E}_t \left(\frac{R_{n,t}}{\Pi + \bar{m}_f \hat{\Pi}_{t+1}} \right)} (J^c + \bar{m}_f \hat{J}_{t+1}^c) && \text{(B.13)} \\
JJ_t^c &= Y_t^c + \xi \mathbb{E}_t \frac{(\Pi + \bar{m}_f \hat{\Pi}_{t+1})^{\zeta-1}}{\mathbb{E}_t \left(\frac{R_{n,t}}{\Pi + \bar{m}_f \hat{\Pi}_{t+1}} \right)} (JJ^c + \bar{m}_f \hat{J}J_{t+1}^c) && \text{(B.14)} \\
1 &= \xi \Pi_t^{\zeta-1} + (1-\xi) \left(\frac{J_t^c}{JJ_t^c} \right)^{1-\zeta} && \text{(B.15)} \\
log\left(\frac{Rn_t}{Rn} \right) &= \rho_r log\left(\frac{Rn_{t-1}}{Rn} \right) \\
&+ (1-\rho_r) \left[\alpha_\pi log\left(\frac{\Pi_t}{\Pi} \right) + \alpha_y log\left(\frac{Y_t}{Y} \right) + \alpha_{dy} log\left(\frac{Y_t}{Y_{t-1}} \right) \right] \\
&+ log(M_t) && \text{(B.16)} \\
log\left(\frac{M_t}{M} \right) &= \rho_m log\left(\frac{M_{t-1}}{M} \right) + \epsilon_t^m && \text{(B.17)} \\
log\left(\frac{G_t^c}{G^c} \right) &= \rho_g log\left(\frac{G_{t-1}^c}{G^c} \right) + \epsilon_t^g && \text{(B.18)} \\
log\left(\frac{MS_t}{MS} \right) &= \rho_{ms} log\left(\frac{MS_{t-1}}{MS} \right) + \epsilon_t^{ms} && \text{(B.19)} \\
log\left(\frac{A_t^c}{A^c} \right) &= \rho_{ma} log\left(\frac{A_{t-1}^c}{A^c} \right) + \epsilon_t^a && \text{(B.20)} \\
& && \text{(B.21)}
\end{aligned}
$$

B.2 Steady State

The exogenous variables have steady states $A^c = MS = M = 1$. Given the steady state inflation rate Π and the steady state nominal interest rate Rn, the steady state values of the other variables can be computed as

$$\Lambda = \frac{\beta}{1+g} \tag{B.22}$$

$$R = \frac{1}{\Lambda} \tag{B.23}$$

$$\frac{J^c}{JJ^c} = \left(\frac{1 - \xi \Pi^{\zeta-1}}{1-\xi} \right)^{\frac{1}{1-\zeta}} \tag{B.24}$$

$$MC = \left(1 - \frac{1}{\zeta} \right) \frac{J^c}{JJ^c} \frac{1 - \xi \beta \Pi^{\zeta}}{1 - \xi \beta \Pi^{\zeta-1}} \tag{B.25}$$

$$\Delta = \frac{(1-\xi) \left(\frac{J^c}{JJ^c} \right)^{-\zeta}}{1 - \xi \Pi^{\zeta}} \tag{B.26}$$

$$\frac{P^W}{P} = MC \tag{B.27}$$

$$H = \left(\frac{\alpha \Delta MC}{\kappa(1-gy)}\right)^{\frac{1}{1+\phi}} \tag{B.28}$$

$$Y^{W,c} = (A^c H)^\alpha \tag{B.29}$$

$$Y^c = \frac{Y^{W,c}}{\Delta} \tag{B.30}$$

$$\Gamma^c = Y^c - \alpha \frac{P^W}{P} Y^{W,c} \tag{B.31}$$

$$W^c = \alpha \frac{P^W}{P} \frac{Y^{W,c}}{H} \tag{B.32}$$

$$G^c = gy * Y^c \tag{B.33}$$

$$T^c = G^c \tag{B.34}$$

$$C^c = Y^c - G^c \tag{B.35}$$

$$J^c = \frac{Y^c MCMS}{(1-\frac{1}{\zeta})(1-\xi\beta\Pi^\zeta)} \tag{B.36}$$

$$JJ^c = \frac{Y^c}{(1-\xi\beta\Pi^{\zeta-1})} \tag{B.37}$$

Finally we can define

$$\begin{aligned}
CEquiv_t &= \mathbb{E}_t\left[\sum_{t=s}^{\infty} \beta^s U(1.01C_{t+s}, H_{t+s})\right] - \mathbb{E}_t\left[\sum_{t=s}^{\infty} \beta^s U(C_{t+s}, H_{t+s})\right] \\
&= \mathbb{E}_t\left[\sum_{t=s}^{\infty} \beta^s \left\{\log(1.01C^c_{t+s}) - \kappa\frac{H_{t+s}^{1+\phi}}{1+\phi} - \log(C^c_{t+s}) - \kappa\frac{H_{t+s}^{1+\phi}}{1+\phi}\right\}\right] \\
&= \log(1.01)\sum_{t=s}^{\infty} \beta^s = \frac{\log(1.01)}{1-\beta} && \text{(B.38)}
\end{aligned}$$

$$U^c = \log(C^c) - \kappa\frac{H^{1+\phi}}{1+\phi} \tag{B.39}$$

$$U^c_{C^c} = \frac{1}{C^c} \tag{B.40}$$

$$U_H = -\kappa H^\phi \tag{B.41}$$

$$V^c = \frac{U^c}{1-\beta} \tag{B.42}$$

B.3 Linearized model

Our set-up is non-linear (an essential feature for the computation of the optimized rules), but in order to compare our model with Gabaix (2020) we now perform a standard log-linearization of the consumption function and NK Phillips curve.

First log-linearizing (B.2) and (B.3) around the steady state $\frac{C}{Z} = 1 - \frac{1}{R}$ gives

$$\tilde{Z}_t = m_y(1 - \frac{1}{R})\tilde{C}_t + \frac{\bar{m}}{R}\tilde{Z}_{t+1} - m_r\frac{1}{R^2}\hat{R}_{t+1} \tag{B.43}$$

$$\tilde{ZZ}_t = \frac{\bar{m}}{R}(\beta R)^{\frac{1}{\gamma}}\tilde{ZZ}_{t+1} + m_r(\beta R)^{\frac{1}{\gamma}}(\frac{1}{\gamma} - 1)\frac{1}{R^2}\hat{R}_{t+1} \tag{B.44}$$

where $(\beta R)^{\frac{1}{\gamma}} = 1$ and log-linearising equation (B.1) yields $\tilde{C}_t = \tilde{Z}_t - \tilde{ZZ}_t$. hence, we can subtract equation (B.43) by equation (B.44) to get:

$$\tilde{C}_t = \tilde{Z}_t - \tilde{ZZ}_t = m_y(1 - \frac{1}{R})\tilde{C}_t + \frac{\bar{m}}{R}(\tilde{Z}_{t+1} - \tilde{ZZ}_{t+1}) - \frac{m_r}{\gamma R^2}\hat{R}_{t+1} \tag{B.45}$$

which is gives the linearised consumption function as in Gabaix (2020):

$$\tilde{C}_t = \frac{\bar{m}}{[R - m_y(R-1)]}\tilde{C}_{t+1} - \frac{m_r}{\gamma R[R - m_y(R-1)]}\hat{R}_{t+1} \tag{B.46}$$

Turning to the Phillips curve we log-linearise equations (64), (65), and (63) again conditional on the zero growth and net inflation steady state inflation to get:

$$\tilde{J}_t = (1 - \beta\xi\Pi^{\zeta})\left(\tilde{Y}_t + m_{fmc}\tilde{MC}_t\right) + \beta\xi\Pi^{\zeta}\mathbb{E}_t\left((1+\zeta)\bar{m}_f m_{f\pi}\tilde{\Pi}_{t+1} - m_{fr}\tilde{Rn}_t + \bar{m}_f\tilde{J}_{t+1}\right) \tag{B.47}$$

$$\tilde{JJ}_t = (1 - \beta\xi\Pi^{\zeta-1})\tilde{Y}_t + \beta\xi\Pi^{\zeta-1}\mathbb{E}_t\left(\zeta\bar{m}_f m_{f\pi}\tilde{\Pi}_{t+1} - m_{fr}\tilde{Rn}_t + \bar{m}_f\tilde{JJ}_{t+1}\right) \tag{B.48}$$

$$\tilde{\Pi}_t = \frac{1 - \xi\Pi^{\zeta-1}}{\xi\Pi^{\zeta-1}}(\tilde{J}_t - \tilde{JJ}_t) \tag{B.49}$$

Notice that the expectation terms here are fully rational, the vector of myopia parameters included in the set of equations above represents the behavioural element of the boundedly rational price-setting firms. When the steady state of inflation is zero (or the steady state gross inflation $\Pi = 1$), we can directly subtract equation (B.48) from equation (B.47) and substitute into equation (B.49) to eliminate $\tilde{Y}_t$ and $\tilde{Rn}_t$ to get a standard Phillips curve at the zero steady state level of inflation as follows:

$$\tilde{\Pi}_t = \frac{(1-\xi)(1-\beta\xi)m_{fmc}}{\xi}\tilde{MC}_t + \beta\left[(1-\xi)\bar{m}_f m_{f\pi} + \xi\bar{m}_f\right]\mathbb{E}_t\tilde{\Pi}_{t+1} \tag{B.50}$$

Again, we can retreat the Phillips curve of the fully rational price-setting firm if the vector of myopia parameters, $[\bar{m}, m_{fmc}, m_{f\pi}, m_{fr}]$, is equal to the vector of 1. Although my behavioural Phillips curve (with zero steady state inflation) is isomophic to that of Gabaix (2020), my behavioural Phillips curve also has the same property as Gabaix's which is less forward-looking compared to the fully rational case. In other word, when firms are more attentive to the macroeconomic outcomes, say, vector m is closer to one, then firms are more forward-looking because the slope on future inflation is higher.

C Euler Learning Model

The EL model have the same set of equations as the RE model, except we add the learning rule specified in the main text for the expectation variables in the system :

$$U_t^c = \log(C_t^c) - \kappa \frac{H_t^{1+\phi}}{1+\phi} \tag{C.51}$$

$$V_t^c = U_t^c + \beta \mathbb{E}_t V_{t+1}^c \tag{C.52}$$

$$U_{C,t}^c = \frac{1}{C_t^c} \tag{C.53}$$

$$U_{H,t} = -\kappa H_t^{\phi} \tag{C.54}$$

$$\Lambda_{t,t+1} = \frac{\beta}{1+g} \frac{\mathbb{E}_t^* U_{C,t+1}^c}{U_{C,t}^c} \tag{C.55}$$

$$R_{t+1} = \frac{R_{n,t}}{\mathbb{E}_t^* \Pi_t} \tag{C.56}$$

$$1 = \mathbb{E}_t \left[\Lambda_{t,t+1} R_{t+1} \right] \tag{C.57}$$

$$W_t^c = -\frac{U_{H,t}}{U_{C,t}^c} \tag{C.58}$$

$$Y_t^{W,c} = A_t^c H_t^{\alpha} \tag{C.59}$$

$$W_t^c = \alpha \frac{P_t^W}{P_t} \frac{Y_t^{W,c}}{H_t} \tag{C.60}$$

$$MC_t = \frac{P_t^W}{P_t} \tag{C.61}$$

$$J_t^c = \frac{1}{1 - \frac{1}{\zeta}} Y_t^c MC_t MS_t + \xi(1+g)\Lambda_{t,t+1} \mathbb{E}_t^{f*} \Pi_{t,t+1}^{\zeta} \mathbb{E}_t^* J_{t+1}^c \tag{C.62}$$

$$JJ_t^c = Y_t^c + \xi(1+g)\Lambda_{t,t+1} \mathbb{E}_t^{f*} \Pi_{t,t+1}^{\zeta-1} \mathbb{E}_t^* JJ_{t+1}^c \tag{C.63}$$

$$1 = \xi \Pi_t^{\zeta-1} + (1-\xi) \left(\frac{J_t^c}{JJ_t^c} \right)^{1-\zeta} \tag{C.64}$$

$$Y_t^c = \frac{Y_t^{W,c}}{\Delta_t} \tag{C.65}$$

$$\Delta_t = \xi \Pi_t^{\zeta} \Delta_{t-1} + (1-\xi)\left(\frac{J_t^c}{JJ_t^c}\right)^{-\zeta} \tag{C.66}$$

$$Y_t^c = C_t^c + G_t^c \tag{C.67}$$

$$\log\left(\frac{R_{n,t}}{R_n}\right) = \rho_r \log\left(\frac{R_{n,t-1}}{R_n}\right) + (1-\rho_r)\left(\theta_\theta \log\left(\frac{\Pi_t}{\Pi}\right) + \theta_y \log\left(\frac{Y_t^c}{Y^c}\right)\right) + \log MPS_t \tag{C.68}$$

$$\log A_t^c - \log A^c = \rho_A(\log A_{t-1}^c - \log A^c) + \epsilon_{A,t} \tag{C.69}$$

$$\log MS_t - \log MS = \rho_{MS}(\log MS_{t-1} - \log MS) + \epsilon_{MS,t} \tag{C.70}$$

$$\log MPS_t - \log MPS = \rho_{MPS}(\log MPS_{t-1} - \log MPS) + \epsilon_{MPS,t} \tag{C.71}$$

$$\log G_t^c - \log G^c = \rho_G(\log G_{t-1}^c - \log G^c) + \epsilon_{G,t} \tag{C.72}$$

$$E_t^*(UC_{t+1}^c) = [E_{t-1}^*(UC_t^c)]^{1-\lambda_{h,uc}^1}[UC_t^c]^{\lambda_{h,uc}^1+\lambda_{h,uc}^2}[UC_{t-1}^c]^{-\lambda_{h,uc}^2} \tag{C.73}$$

$$E_t^*(\Pi_{t+1}) = [E_{t-1}^*(\Pi_t)]^{1-\lambda_{h,\pi}^1}[\Pi_t]^{\lambda_{h,\pi}^1+\lambda_{h,\pi}^2}[\Pi_{t-1}]^{-\lambda_{h,\pi}^2} \tag{C.74}$$

$$E_t^{f*}(\Pi_{t+1}) = [E_{t-1}^{f*}(\Pi_t)]^{1-\lambda_{f,\pi}^1}[\Pi_t]^{\lambda_{f,\pi}^1+\lambda_{f,\pi}^2}[\Pi_{t-1}]^{-\lambda_{f,\pi}^2} \tag{C.75}$$

$$E_t^*(J_{t+1}^c) = [E_{t-1}^*(J_t^c)]^{1-\lambda_J^1}[J_t^c]^{\lambda_J^1+\lambda_J^2}[J_{t-1}^c]^{-\lambda_J^2} \tag{C.76}$$

$$E_t^*(JJ_{t+1}^c) = [E_{t-1}^*(JJ_t^c)]^{1-\lambda_{JJ}^1}[JJ_t^c]^{\lambda_{JJ}^1+\lambda_{JJ}^2}[JJ_{t-1}^c]^{-\lambda_{JJ}^2} \tag{C.77}$$

D Estimation Results

D.1 Identification

Assuming that a unique solution exists for each model, it can be cast in the following form

$$\mathbf{z}_t = \mathbf{A}(\boldsymbol{\theta})\mathbf{z}_{t-1} + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}_t \tag{D.78}$$

Some of the variables in z_t are not observed, so the transition equation (D.78) is complemented by a measurement equation

$$\mathbf{x}_t = \mathbf{C}\mathbf{z}_t + \mathbf{D}\mathbf{u}_t + \boldsymbol{\nu}_t \tag{D.79}$$

The unconditional first and second moments of $\mathbf{x}_t$ are given by

$$E\mathbf{x}_t \quad := \mu_{\mathbf{x}} = \mathbf{s} \tag{D.80}$$

$$\operatorname{cov}\left(\mathbf{x}_{t+i}, \mathbf{x}_t'\right) \quad := \Sigma_{\mathbf{x}}(i) = \begin{cases} \mathbf{C}\Sigma_{\mathbf{z}}(0)\mathbf{C}' & \text{if } i = 0 \\ \mathbf{C}\mathbf{A}^i\Sigma_{\mathbf{z}}(0)\mathbf{C}' & \text{if } i > 0 \end{cases} \tag{D.81}$$

where $\Sigma_z(0) := E\mathbf{z}_t\mathbf{z}_t'$ solves the matrix equation

$$\Sigma_z(0) = \mathbf{A}\Sigma_z(0)\mathbf{A}' + \mathbf{\Omega} \tag{D.82}$$

Denote τ collecting the non-constant elements of $\hat{\mathbf{z}}^*, \mathbf{A}$, and $\mathbf{\Omega}$, i.e. $\tau := [\tau_z', \tau_A', \tau_\Omega']'$. Denote the observed data with $\mathbf{X}_T := [\mathbf{x}_1', \ldots, \mathbf{x}_T']'$, and let $\mathbf{\Sigma}_T$ be its covariance matrix, i.e.

$$\mathbf{\Sigma_T} := \mathbb{E}\mathbf{X}_T\mathbf{X}_T' \tag{D.83}$$

$$\mathbf{\Sigma_T} \quad = \begin{pmatrix} \Sigma_{\mathbf{x}}(0), & \Sigma_{\mathbf{x}}(1)', & \ldots, & \Sigma_{\mathbf{x}}(T-1)' \\ \Sigma_{\mathbf{x}}(1), & \Sigma_{\mathbf{x}}(0), & \ldots, & \Sigma_{\mathbf{x}}(T-2)' \\ \cdots & \cdots & \cdots & \cdots \\ \Sigma_{\mathbf{x}}(T-1), & \Sigma_{\mathbf{x}}(T-2), & \ldots, & \Sigma_{\mathbf{x}}(0) \end{pmatrix} \tag{D.84}$$

We define $\mathbf{m}_T := [\boldsymbol{\mu}', \boldsymbol{\sigma}_T']'$, where

$$\sigma_T := \left[\operatorname{vech}\left(\Sigma_{\mathbf{x}}(0)\right)', \operatorname{vec}\left(\Sigma_{\mathbf{x}}(1)\right)', \ldots, \operatorname{vec}\left(\Sigma_{\mathbf{x}}(T-1)\right)'\right]'$$

$\mathbf{m}_T$ is a function of $\boldsymbol{\theta}$. If either $\mathbf{u}_t$ is Gaussian (which is true in our case), or there are no distributional assumptions about the structural shocks, the model-implied restrictions on $\mathbf{m}_T$ contain all information that can be used for the estimation of θ. The identifiability of θ depends on whether that information is sufficient or not.

Global identification: the Gaussian case. Suppose that the data $\mathbf{X}_T$ is generated by the model (D.78) and (D.79) with parameter vector θ_0. Then θ_0 is globally identified if

$$\mathbf{m}_T(\tilde{\boldsymbol{\theta}}) = \mathbf{m}_T\left(\boldsymbol{\theta}_0\right) \Leftrightarrow \tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 \tag{D.85}$$

for any $\tilde{\boldsymbol{\theta}} \in \Theta$. If (D.85) is true only for values $\tilde{\boldsymbol{\theta}}$ in an open neighborhood of θ_0, the identification of θ_0 is local.

Local identification: The rank condition. Suppose that $\mathbf{m}_T$ is a continuously differentiable function of $\boldsymbol{\theta}$. Then θ_0 is locally identifiable if the Jacobian matrix $J(q) := \frac{\partial \mathbf{m}_q}{\partial \boldsymbol{\theta}'}$ has a full column rank at θ_0 for $q \leq T$. This condition is both necessary and sufficient when $q = T$ if $\mathbf{u}_t$ is normally distributed.

Given the chain rule

$$J(T) = \frac{\partial \mathbf{m}_T}{\partial \boldsymbol{\tau}'} \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}'} \tag{D.86}$$

another necessary condition discussed in Iskrev and Ratto (2010). The point θ_0 is locally identifiable only if the rank of $J_2 = \frac{\partial \tau}{\partial \theta'}$ at θ_0 is equal to k (the number of estimated parameters). The condition is necessary because the distribution of $\mathbf{X}_T$ depends on θ only through τ, irrespectively of the distribution of $\mathbf{u}_t$. It is not sufficient since, unless all state variables are observed, τ may be unidentifiable.

Results: prior_mean - Identification using info from observables. Identification strength-plots are provided in the appendices.

Upper Panel: the bar charts depict the identification strength of the parameters based on the Fischer information matrix normalized by either the parameter at the prior mean (blue bars) or by the standard deviation at the prior mean (red bars). The weighting with the prior standard deviation is only available if priors have been specified. Intuitively, the bars represent the normalized curvature of the log likelihood function at the prior mean in the direction of the parameter.

Lower Panel: This panel further decomposes the effect shown in the upper panel. A weak identification can be due to either other parameters linearly compensating/replacing the effect of a parameter (i.e. parameters having exactly the same effect on the likelihood) or the fact that the likelihood does not change at all with the respective parameter. This latter effect is called sensitivity.

The identification results show that the BR model is weakly identified for 2 parameters α and ϕ. Other three models are identified.

```
BR Model
```

```
REDUCED-FORM:
    All parameters are identified in the Jacobian of steady state and reduced-form solutio

MINIMAL SYSTEM (KOMUNJER AND NG, 2011):
    All parameters are identified in the Jacobian of steady state and minimal system (rank

SPECTRUM (QU AND TKACHENKO, 2012):
    !!!WARNING!!!
    The rank of Gbar (Jacobian of mean and spectrum) is deficient!

    [phi,xi] are PAIRWISE collinear!
    [alp,xi] are PAIRWISE collinear!
    [alp,phi] are PAIRWISE collinear!

MOMENTS (ISKREV, 2010):
    All parameters are identified in the Jacobian of first two moments (rank(J) is full wi
```

```
RE Model
REDUCED-FORM:
    All parameters are identified in the Jacobian of steady state and reduced-form solutio

MINIMAL SYSTEM (KOMUNJER AND NG, 2011):
    All parameters are identified in the Jacobian of steady state and minimal system (rank

SPECTRUM (QU AND TKACHENKO, 2012):
    !!!WARNING!!!
    The rank of Gbar (Jacobian of mean and spectrum) is deficient!
```

```
    theta_dy is not identified!
    rhoMS is not identified!
    rhoG is not identified!
    [alp,phi] are PAIRWISE collinear!

MOMENTS (ISKREV, 2010):
    All parameters are identified in the Jacobian of first two moments (rank(J) is full wi
```

```
EL Model
REDUCED-FORM:
    All parameters are identified in the Jacobian of steady state and reduced-form solutio

MINIMAL SYSTEM (KOMUNJER AND NG, 2011):
    All parameters are identified in the Jacobian of steady state and minimal system (rank

SPECTRUM (QU AND TKACHENKO, 2012):
    All parameters are identified in the Jacobian of mean and spectrum (rank(Gbar) is full

MOMENTS (ISKREV, 2010):
    All parameters are identified in the Jacobian of first two moments (rank(J) is full wi
```

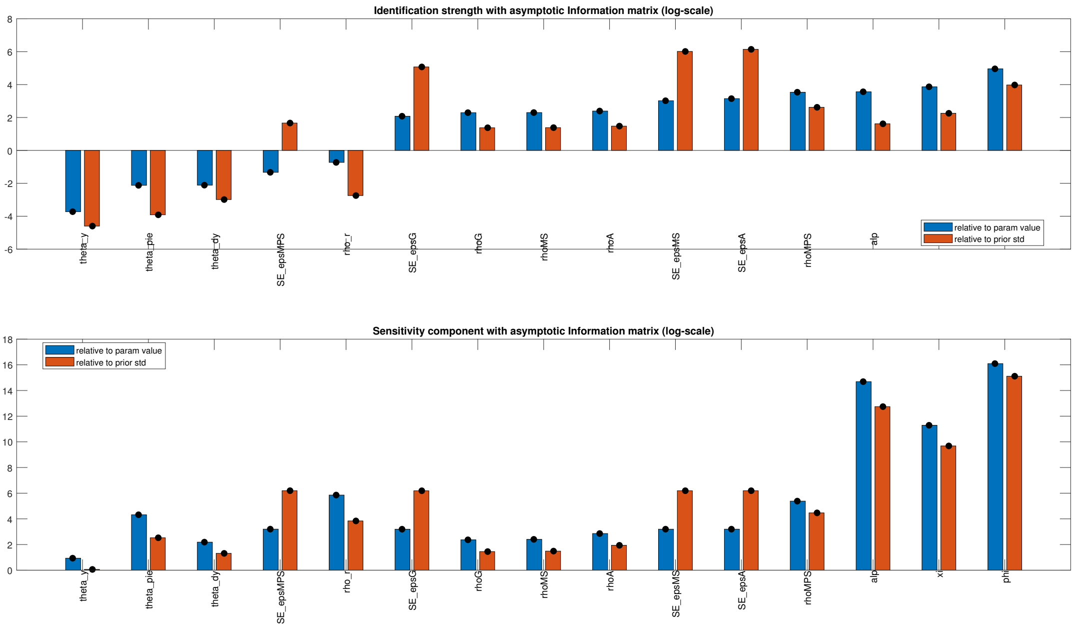

Figure 15: Identification of the RE model

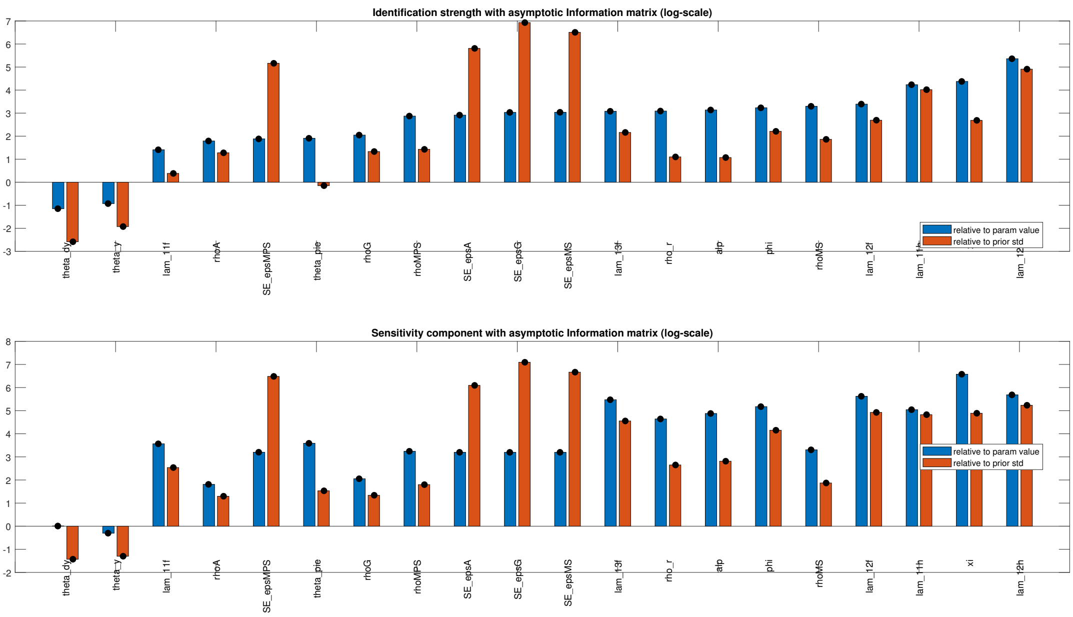

Figure 16: Identification of the Euler leaning with simple adaptive expectation

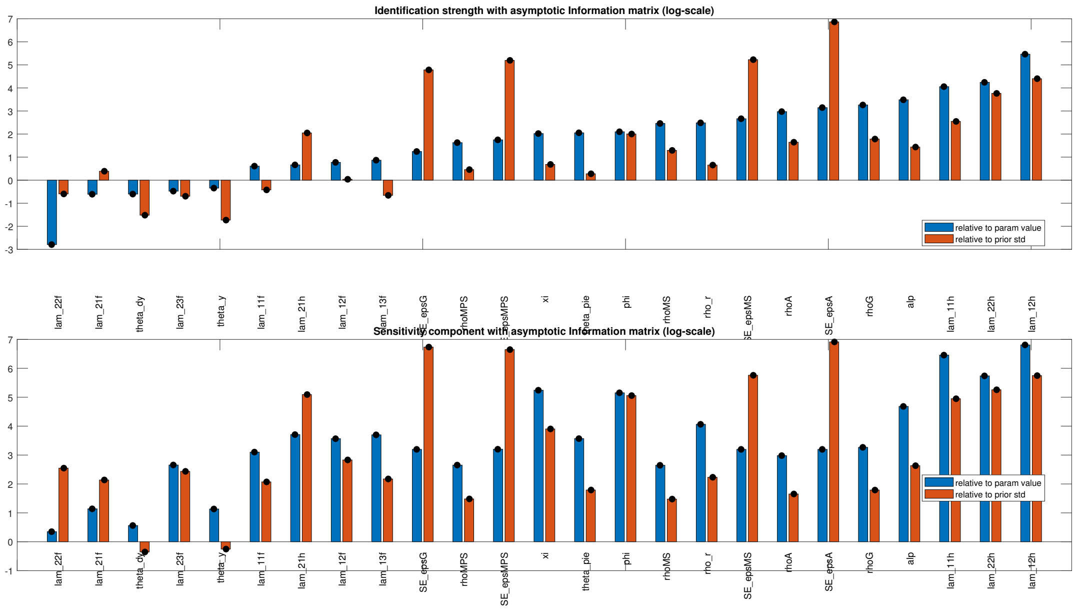

Figure 17: Identification of the Euler leaning with Generalized adaptive expectation

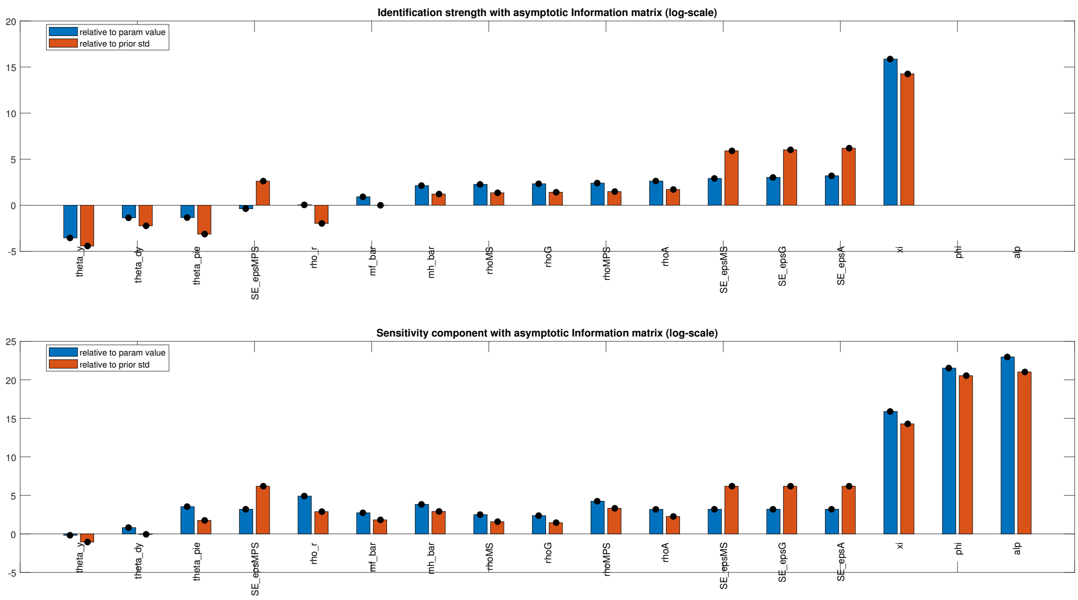

Figure 18: Identification of the Gabaix model

D.2 Estimation results

Estimation of the Rational expectation. Sample from 1958Q1 to 2017Q4 with shadow rate.

Log data density is 3750.748983.

parameters

	prior mean	post. mean	90% HPD interval		prior	pstdev
xi	0.500	0.7552	0.7256	0.7849	beta	0.1000
phi	2.000	4.6626	3.8633	5.4614	norm	0.7500
alp	0.700	0.8861	0.8278	0.9469	beta	0.1000
rho_r	0.750	0.3015	0.2164	0.3846	beta	0.1000
theta_pie	1.500	2.4474	2.1938	2.7060	norm	0.2500
theta_y	0.120	0.0597	0.0272	0.0919	norm	0.0500
theta_dy	0.120	0.1092	0.0601	0.1598	norm	0.0500
rhoA	0.500	0.9897	0.9820	0.9980	beta	0.2000
rhoMS	0.500	0.9560	0.9362	0.9763	beta	0.2000
rhoMPS	0.500	0.5994	0.5369	0.6635	beta	0.2000
rhoG	0.500	0.9088	0.8913	0.9268	beta	0.2000

standard deviation of shocks

	prior mean	post. mean	90% HPD interval		prior	pstdev
epsA	0.001	0.0065	0.0060	0.0070	invg	0.0200
epsMS	0.001	0.0370	0.0309	0.0428	invg	0.0200
epsMPS	0.001	0.0050	0.0043	0.0057	invg	0.0200
epsG	0.001	0.0518	0.0440	0.0592	invg	0.0200

Estimation of the Euler learning with simple adaptive expectation. Sample from 1958Q1 to 2017Q4 with shadow rate.

Log data density is 3714.418834.

parameters

	prior mean	post. mean	90% HPD interval		prior	pstdev
xi	0.500	0.7955	0.7521	0.8359	beta	0.1000
phi	2.000	1.2257	0.2960	2.0417	norm	0.7500
alp	0.700	0.9734	0.9579	0.9904	beta	0.1000
rho_r	0.750	0.8808	0.8276	0.9373	beta	0.1000
theta_pie	1.500	1.4193	1.1440	1.6953	norm	0.2500
theta_y	0.120	0.1132	0.0595	0.1674	norm	0.0500
theta_dy	0.120	0.1905	0.1067	0.2674	norm	0.0500
rhoA	0.500	0.9854	0.9746	0.9960	beta	0.2000
rhoMS	0.500	0.7667	0.5304	0.9516	beta	0.2000
rhoMPS	0.500	0.3871	0.2591	0.5145	beta	0.2000
rhoG	0.500	0.9843	0.9742	0.9958	beta	0.2000
lam_11h	0.500	0.1195	0.0527	0.1826	beta	0.2000
lam_11f	0.500	0.2608	0.0790	0.4741	beta	0.2000
lam_12h	0.500	0.0678	0.0070	0.1569	beta	0.2000
lam_12f	0.500	0.2285	0.0907	0.3401	beta	0.2000
lam_13f	0.500	0.5448	0.1028	0.8776	beta	0.2000

standard deviation of shocks

	prior mean	post. mean	90% HPD interval		prior	pstdev
epsA	0.001	0.0066	0.0061	0.0070	invg	0.0200
epsMS	0.001	0.0889	0.0461	0.1314	invg	0.0200
epsMPS	0.001	0.0023	0.0022	0.0025	invg	0.0200
epsG	0.001	0.0246	0.0227	0.0265	invg	0.0200

Estimation of the Gabaix Model. Sample from 1958Q1 to 2017Q4 with shadow rate.

```
Log data density is 3750.800048.


parameters
            prior mean   post. mean        90% HPD interval    prior      pstdev


xi               0.500       0.6962      0.6565      0.7351    beta       0.1000
phi              2.000       3.7243      2.8656      4.5382    norm       0.7500
alp              0.700       0.9092      0.8594      0.9606    beta       0.1000
rho_r            0.750       0.3343      0.2471      0.4192    beta       0.1000
theta_pie        1.500       2.6075      2.3458      2.8594    norm       0.2500
theta_y          0.120       0.0417      0.0039      0.0811    norm       0.0500
theta_dy         0.120       0.1739      0.1169      0.2312    norm       0.0500
rhoA             0.500       0.9919      0.9855      0.9986    beta       0.2000
rhoMS            0.500       0.9633      0.9446      0.9824    beta       0.2000
rhoMPS           0.500       0.6175      0.5562      0.6782    beta       0.2000
rhoG             0.500       0.9475      0.9260      0.9689    beta       0.2000
mh_bar           0.500       0.9379      0.9130      0.9623    beta       0.2000
mf_bar           0.500       0.5405      0.2211      0.8785    beta       0.2000

standard deviation of shocks
            prior mean   post. mean        90% HPD interval    prior      pstdev


epsA             0.001       0.0065      0.0060      0.0070    invg       0.0200
epsMS            0.001       0.0297      0.0236      0.0355    invg       0.0200
epsMPS           0.001       0.0050      0.0043      0.0057    invg       0.0200
epsG             0.001       0.0434      0.0374      0.0495    invg       0.0200
```

Estimation of the Euler learning Model with Generalized Adaptive Expectation. Sample from 1958Q1 to 2017Q4 with shadow rate.

Log data density is 3772.209739.

parameters

	prior mean	post. mean	90% HPD interval		prior	pstdev
xi	0.500	0.5449	0.4379	0.6542	beta	0.1000
phi	2.000	3.7491	2.8618	4.6466	norm	0.7500
alp	0.700	0.8851	0.8195	0.9518	beta	0.1000
rho_r	0.750	0.4400	0.3400	0.5395	beta	0.1000
theta_pie	1.500	2.0188	1.7185	2.3248	norm	0.2500
theta_y	0.120	0.0480	0.0003	0.0965	norm	0.0500
theta_dy	0.120	0.0601	-0.0004	0.1217	norm	0.0500
rhoA	0.500	0.9840	0.9729	0.9958	beta	0.2000
rhoMS	0.500	0.9729	0.9561	0.9903	beta	0.2000
rhoMPS	0.500	0.7959	0.7245	0.8667	beta	0.2000
rhoG	0.500	0.9764	0.9602	0.9921	beta	0.2000
lam_11h	0.500	0.0344	0.0104	0.0569	beta	0.2000
lam_11f	0.500	0.2386	0.0208	0.5157	beta	0.2000
lam_21h	0.000	0.8036	0.7618	0.8486	beta	0.2500
lam_21f	0.000	0.1345	-0.0505	0.3118	beta	0.2500
lam_12h	0.500	0.0556	0.0143	0.0930	beta	0.2000
lam_12f	0.500	0.8861	0.7913	0.9846	beta	0.2000
lam_22h	0.000	-0.7315	-0.8682	-0.5983	beta	0.2500
lam_22f	0.000	0.6223	0.4359	0.8167	beta	0.2500
lam_13f	0.500	0.6522	0.4069	0.9099	beta	0.2000
lam_23f	0.000	0.1073	-0.2696	0.4868	beta	0.2500

standard deviation of shocks

	prior mean	post. mean	90% HPD interval		prior	pstdev

epsA	0.001	0.0065	0.0060	0.0070	invg	0.0200
epsMS	0.001	0.0286	0.0230	0.0341	invg	0.0200
epsMPS	0.001	0.0035	0.0030	0.0040	invg	0.0200
epsG	0.001	0.0458	0.0387	0.0531	invg	0.0200

E Impulse Responses of Robust Policy to Shocks

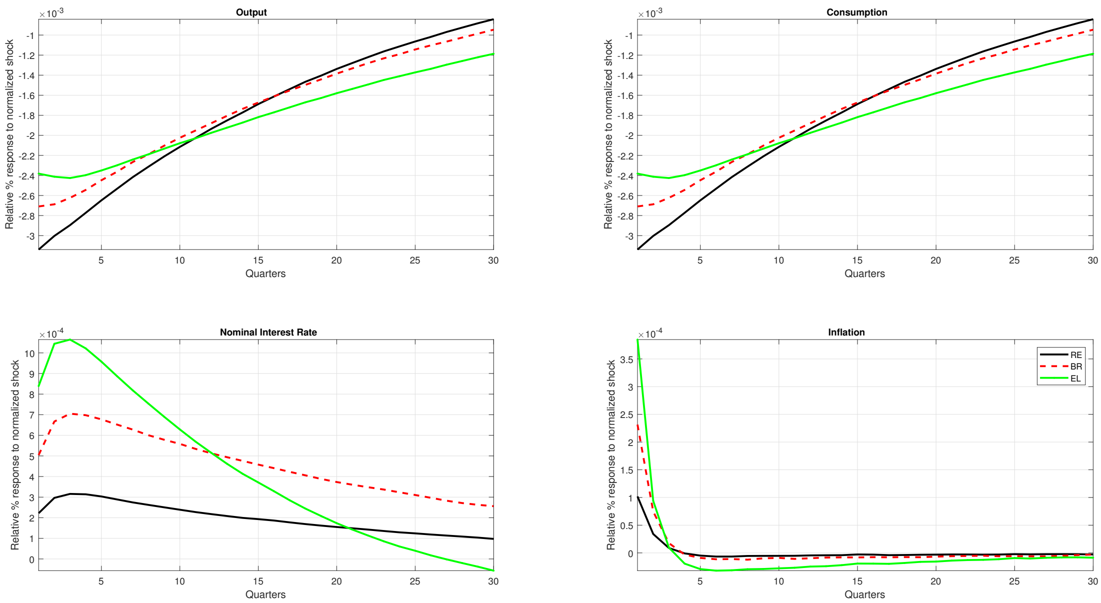

Figure 19: Impulse responses to the cost-push shock comparison between the models at the robust ZLB mandate. $\bar{p} = 0.05$

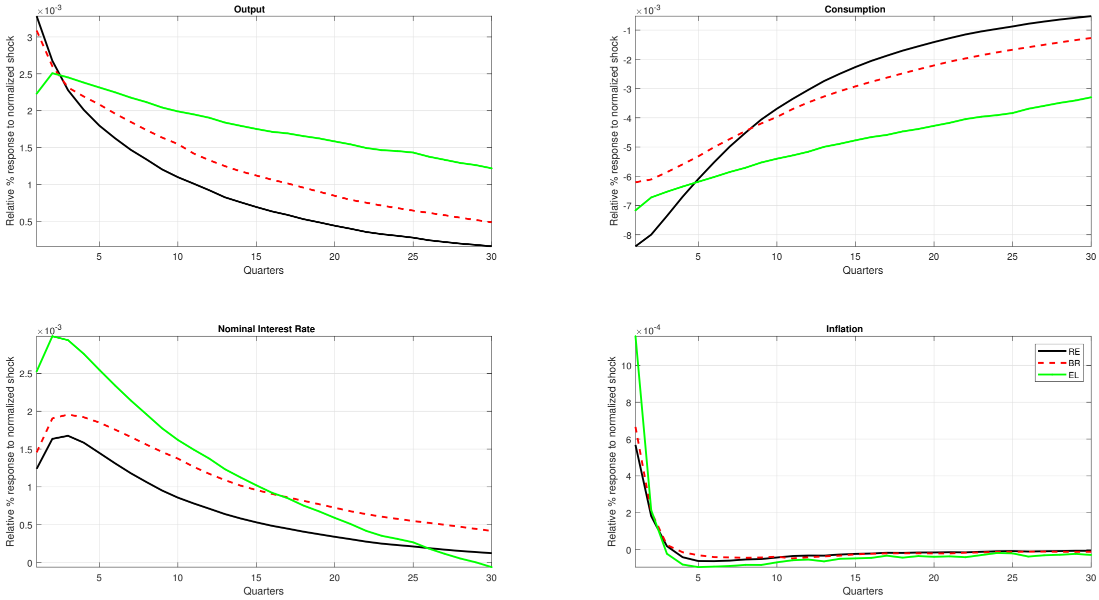

Figure 20: Impulse responses to the government spending shock comparison between the models at the robust ZLB mandate. $\bar{p} = 0.05$

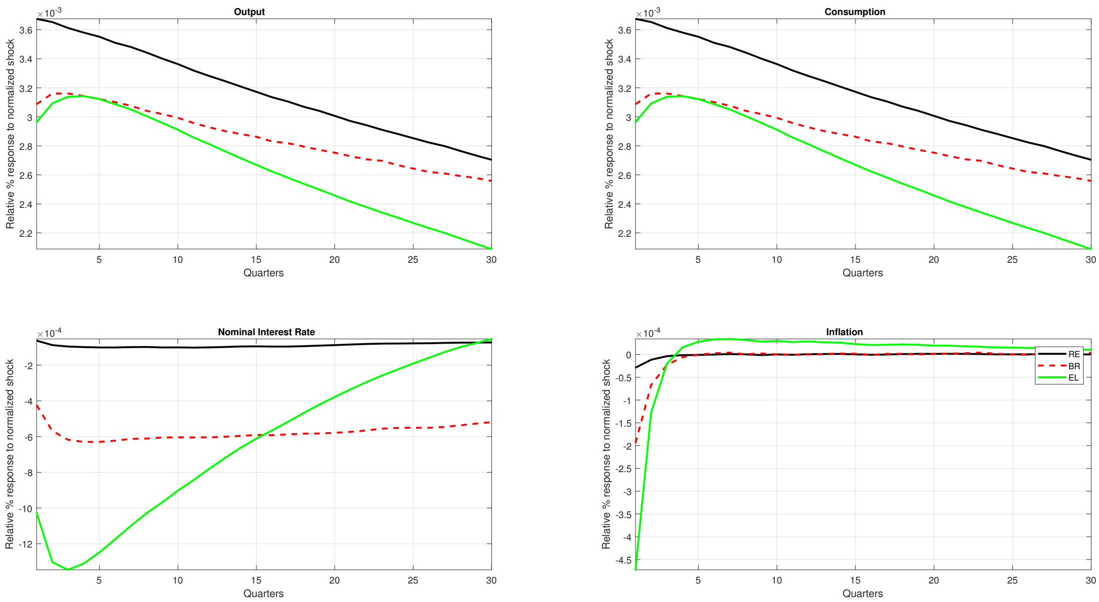

Figure 21: Impulse responses to the technology shock comparison between the models at the robust ZLB mandate. $\bar{p} = 0.05$