Sparse Restricted Perception Equilibrium

Volha Audzei and Sergey Slobodyan *

March, 2018

Preliminary draft

Abstract

In this work we consider a concept of sparse rationality (developed recently by Gabaix, 2014) as a selection tool in a model with multiple equilibria. Under sparse rationality paying attention to all possible variables is costly, and the agents could choose to over- or under-emphasize particular variables, even fully excluding some of them. Our main question is whether an initially mis-specified equilibrium (the Restricted Perceptions Equilibrium, or RPE) is compatible with the equilibrium choice of sparse weights, describing allocation of attention to different variables by the agents inhabiting this RPE. In a simple New Keynesian model, we find that the agents stick to their initial mis-specified AR(1) forecasting model choice when monetary policy is less aggressive or inflation becomes more persistent. We also identify a region in the parameter space where the agents find it advantageous to pay attention to no variable at all.

JEL Codes: D84, E31, E37.

Keywords: Bounded Rationality, Expectations, Learning.

Volha Audzei, Czech National Bank, Economic Research Division, Monetary Department, and CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Prague, Czech Republic, vaudzei@cerge-ei.cz

Sergey Slobodyan, National Research University Higher School of Economics, St. Petersburg, Russia, and CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Prague, Czech Republic, sergey.slobodyan@cerge-ei.cz

We thank to Filip Matejka, Xavier Gabaix, Raf Wouters, Pierpaolo Benigno, Alexis Dervitz and participants of various conferences and workshops. Authors are grateful for the financial support of the Czech Science Foundation project No. P402/12/G097 DYME Dynamic Models in Economics. The views expressed in this paper are those of the authors and not necessarily those of the Czech National Bank.

Nontechnical Summary

There has been a growing literature showing that economic agents form their expectations in a way that is often inconsistent with rational expectations hypothesis. Agents' expectations in turn influence efficiency of monetary policy and may even call for discussion of different policy choices (e.g. Gabaix 2016, García-Schmidt and Woodford 2015 or Hommes et al. 2017).

In this paper we contribute to the literature by studying how monetary policy itself influences the way the agents form their expectations. We consider a stylized model, where adaptively learning agents are restricted to use only a subset of the observables. This subset is chosen such that it results in the smallest forecast errors. We then allow our agents to reconsider their initial choice of forecasting rules subject to information constraints as in Gabaix (2014). We then study under which conditions agents stick to the initially mis-specified forecasting rule, move to another mis-specified rule or to a rule consistent with rational expectations.

Within a text-book New Keynesian model framework, we show that one of the crucial parameters, governing the agents' choice is the monetary policy reaction to inflation expectations in the Taylor rule. The stronger the monetary policy reacts to deviation of inflation expectations from the target, the larger is the parameter space where rational expectations consistent rule is selected. We consider two basic policy rules – with forward looking or with contemporaneous inflation expectations. The predictions are similar with both rules, except that the rule with contemporaneous expectations results in lower volatility and less tightening is needed for the agents to move to the rational expectations equilibrium.

We also find that as inflation becomes more persistent, the mis-specified AR(1) rule for inflation forecasting survives for larger parameter space. The same prediction holds for smaller correlation between inflation and output. These results are supported by some studies on professional forecasters' behavior. An example is Lopez-Perez (2017) where forecasters are found to pay more attention to inflation and less to output gap in recent years of high inflation persistence and low correlation between inflation and output gap.

1. Introduction

It has been understood for a long time that the hypothesis of Rational Expectations (RE), while delivering a theoretically elegant, model consistent, and typically unique solution for the agents' expectations, imposes on them cognitive and computational demands that might be incompatible with reality. As a result, deviations from RE have been studied in a growing stream of theoretical and empirical literature, including Bounded Rationality (Marcet and Sargent 1989), Adaptive Learning (Evans and Honkapohja 2001), Sticky Information (Mankiw and Reis 2007), Rational Inattention (Sims 2003), and Sparse Rationality (Gabaix 2014) approaches, among others.

Our paper contributes to the literature by studying the interaction between monetary policy and equilibrium forecasting rules, and making connections between two distinct concepts, adaptive learning and sparse rationality. We study an economy where adaptively learning agents choose a strict subset of variables from the rational expectation equilibrium set for their forecasting functions, thus inducing a restricted perception equilibrium (RPE). We then allow these agents, inhabiting the RPE, to reconsider their forecast rules, subject to the informational cost constraint modeled as in Gabaix (2014). We ask whether the model parameters which make the RPE stable are sufficient to ensure that the same subset of variables is selected by informationally constrained agents. In other words, we are interested in whether the initial mis-specification becomes self-perpetuating in case of informational constraints. One of the key parameters defining the survival of a mis-specified rule is expectational feedback, which is a function of monetary policy aggressiveness. We show that a stronger monetary policy increases the region in the parameter space where the agents move to the RE equilibrium. We consider two policy rules, with forward looking inflation expectations and with contemporaneous inflation expectations. Under policy rule with the contemporaneous inflation expectations, there is smaller region in the parameter space, where mis-specified rule survives in equilibrium.

Our paper is related to the literature on adaptive learning and bounded rationality. In an adaptive learning approach to modeling deviations from RE, agents are assumed to not possess prior knowledge of underlying structure of the economy and to be gradually learning the coefficients in their forecasting rules. The survey of this approach to learning in macroeconomics can be found in Evans and Honkapohja (2009). Several papers have shown that adaptive agents could persist in using the forecasting rules that are mis-specified relatively to the RE ones, cf., Molnar (2007) and Evans et al. (2012). In Molnar (2007) there is a class of agents who are learning what is the best forecasting model given the past data. Even if their forecasting rules are mis-specified, such learners can survive the competition with RE agents. In Evans et al. (2012) convergence to a mis-specified equilibrium happens when the expectation feedback is strong. Adam (2005) considers the economy where agents are restricted to process only a certain number of variables in the regression and thus to use underparametrized forecasting rules. As the agents' expectations affect the data generating process of the model and induce a Restricted Perception Equilibrium (RPE), the restricted rule can outperform the rational expectation rule in equilibrium. Similarly to Evans et al. (2012), this happens for the large enough feedback from the expectations to the outcome variable.

One of the ways to justify the agents using mis-specified forecasting rules is to assume that they have limited information processing capacity, as is done in the rational inattention literature, cf., Sims (2003), Mackowiak and Wiederholt (2009), and Matejka and McKay (2015). In this literature, attention allocation is based on the concept of entropy. Gabaix (2014) pursues a different, less computationally demanding, approach, with agents allocating attention weights to variables based on their relative importance.

There are now a number of papers, that show how bounded rationality influences monetary and fiscal policy effects. Among them are Gabaix (2016) and García-Schmidt and Woodford (2015). Some of their findings call for reconsideration of Taylor principle, and explain forward guidance puzzle. As agents do not see far in the future in Gabaix (2016), current policy has limited effect for the future agents' decisions. Likewise, the announcement of future policies has an altenuated effect on agents contemporaneous decision. Unlike these papers, however, we are interested in how monetary policy influences expectation formation, not in how expectation formation influences monetary policy.

Another strand of research on bounded rationality we relate to is experimental research on interaction of monetary policy and agents' expectations. Examples are Pfajfar and Žakelj (2017) and Assenza et al. (2013), who find that the monetary policy aggressiveness influences choice of the model. Studies by Hommes (2014) and Heemeijer et al. (2009) support the importance of expectational feedback parameter for the survival of mis-specified rules. In our paper, expectational feedback parameter is an inverse of monetary policy aggressiveness, and we study how it influences the agents' modeling choice.

The paper is organized as follows. In the second section we study an economy where agents learn about a simple exogenous process. We derive analytical results and provide economic intuition for them. In the third section we move to a simple New Keynesian model and study the interaction of monetary policy and agents forecasting rules. The last section concludes.

2. Simple Model

In this section we demonstrate main intuition behind our results for a simple process. Later in the paper, we generalize our findings for a 3-equation New Keynesian model, and discuss the possible policy implications. We start our analysis with a simple process

$$y_t = \alpha + \beta E_t y_{t+1} + \gamma_1 w_t^1 + \gamma_2 w_t^2 + \eta_t, \tag{1}$$

where w_t^1 and w_t^2 are persistent shocks such that

$$\begin{bmatrix} w_t^1 \\ w_t^2 \end{bmatrix} = \begin{bmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{bmatrix} \cdot \begin{bmatrix} w_{t-1}^1 \\ w_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix}.$$

The $\left(w_t^1, w_t^2\right)$ shocks are normally distributed around zero with variance-covariance matrix

$$\Sigma^w = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

with $\rho \in [-1,1]$ being the correlation coefficient between the shocks, defined as $\frac{Cov(w_t^1, w_t^2)}{\sigma_1\sigma_2}$. The RE minimum state variable solution (MSV) of this model is given by

$$y_t = a + g_1 w_t^1 + g_2 w_t^2 + \eta_t,$$

with MSV coefficients:

$$g_1 = \frac{\gamma_1}{1-\beta\rho_1}, \tag{2}$$

$$g_1 = \frac{\gamma_2}{1-\beta\rho_2}, \tag{3}$$

and η_t is a white noise.

We restrict the agents to using only one variable in their forecasting models, as in Adam (2005) and Adam (2007) framework.[1] In our model, their forecasting rule could use either w_t^1 or w_t^2:

$$\begin{aligned} y_t &= a_1 + b_1 w_t^1 + \eta_t, \qquad (4) \\ y_t &= a_2 + b_2 w_t^2 + \eta_t. \qquad (5) \end{aligned}$$

Without loss of generality, we assume that our agents use (4) as their Perceived Law of Motion (PLM), which, when the agents use this PLM to form expectations about y_{t+1}, induces the restricted perceptions equilibrium that we call RPE1. Substituting the forecast formed using (4) into (1), we obtain the actual law of motion (ALM):

$$y_t = \alpha + \beta a_1 + \bar{b}_1 w_t^1 + \bar{b}_2 w_t^2 + \eta_t, \qquad (6)$$

with

$$\begin{aligned} \bar{b}_1 &= \beta \rho_1 b_1 + \gamma_1, \qquad (7) \\ \bar{b}_2 &= \gamma_2. \qquad (8) \end{aligned}$$

We model our agents as econometricians, who do not have the prior knowledge about the underlying structure of the economy. They, however, do the best they can using the past data. In order for this learning process to converge, three conditions must hold. First, for the agents' PLM to be the equilibrium solution, the coefficient b_1 must be derived as a regression coefficient from ordinary least squares:

$$b_1 = \frac{Cov(y_t, w_t^1)}{Var\left(w_t^1\right)}. \qquad (9)$$

Second, the equilibrium (4) must be expectationally stable (E-stable). Finally, the forecast errors produced by the rule of their choice, (4), must be smaller than that of the alternative, (5). We measure forecast errors as mean squared forecast errors (MSFE)

Proposition 2.0.1. *In RPE1 (RPE2), where the agents use as forecasting rule equation 4 (5), the equilibrium coefficient b_1 (b_2) is given by*

$$\begin{aligned} b_1 &= \frac{\gamma_1 + \gamma_2 \cdot \rho \frac{\sigma_2}{\sigma_1}}{1 - \beta \rho_1}, \\ b_2 &= (\beta \rho_1 b_1 + \gamma_1) \cdot \rho \frac{\sigma_1}{\sigma_2} + \gamma_2. \end{aligned}$$

Both RPE1 and RPE2 are E-stable. The MSFE for an agent inhabiting RPE1 and using (4) as forecasting rule, $MSFE_1$, is smaller than MSFE of an agent using (5), $MSFE_2$, and thus RPE1 is an equilibrium if:

$$b_1^2 \sigma_1^2 > b_2^2 \sigma_2^2. \qquad (10)$$

[1] Such a restriction could be motivated by empirical and experimental evidence, cf., Branch and Evans (2006b), Adam (2007), Hommes (2014), and Pfajfar and Žakelj (2014), who show that very simple AR(1) rules might be used by subjects to forecast inflation in survey and experimental settings. Several papers that estimate DSGE models with adaptive expectations, for example, Slobodyan and Wouters (2012) and Ormeno and Molnar (2015), show that assuming the agents are using very simple forecasting rules leads to superior model fit.

Proof. Appendix A. □

We next allow the agents to challenge their equilibrium forecasting rules and possibly reconsider them. We conduct our analysis for the case when RPE1 is an equilibrium. That is, agents have initially chosen to use w^1 in their forecasting rule and have ex-post found out that it produces smaller forecast errors than a model with w^2, i.e. (10) holds. Note, that in our simple model, both forecasting rules are under-specified. Thus, the analysis for the case where the agents have initially started with RPE2 would be symmetric.

Our agents know that there exist other variables in the RPE1 (6), and w_t^2 is observable. They also know that using w_t^2 alone for forecasting is inferior to using only w_t^1, because (10) is true. However they may wonder whether *adding* w_t^2 to their forecasting rule is beneficial. The forecasting rule that includes both w_t^1 and w_t^2 would be clearly superior in this model, if the agents were allowed to learn its coefficients, coinciding with $\bar{b}_1$ and $\bar{b}_2$.[2] The agents, however, are subject to attention cost, modeled as in Gabaix (2014). They could attach weights to a variable according to its importance. The importance of a variable depends on its contribution to the variance of the variable of interest (y in our case) and to the agents' utility. The weights then determine how much attention is paid to a variable given the exogenously cost of attention, and the loss stemmimg from inattention, which is a reduced quality of the forecast. We let the agents to choose the attention vector by maximizing the precision of their forecast of y_t as in (6):

$$u = -\frac{1}{2}\left(\hat{y}_t - y_t\right)^2 . \tag{11}$$

For agents with rational expectations, the optimal forecast is equal to $\hat{y}_t = \widehat{\bar{b}}_1 w_t^1 + \widehat{\bar{b}}_2 w_t^2$, where $\widehat{\bar{b}}_1$ and $\widehat{\bar{b}}_2$ are OLS estimates of the coefficients in (6). That is, agents with rational expectations use both shocks in the forecasting rule. Sparse rational agents face a trade-off between attention cost and increase in forecast precision. That is why they choose to allocate attention between the variables and form their optimal forecast rule as $\hat{y}_t = \widehat{\bar{b}}_1(m_1 w_t^1) + \widehat{\bar{b}}_2(m_2 w_t^2)$, where $m_1, m_2 \in [0,1]$ are attention weights, and $\widehat{\bar{b}}_1$, $\widehat{\bar{b}}_2$ are OLS estimates of the coefficients in the selected forecasting rule. Denoting $x = \left(w_t^1, w_t^2\right)^T$, and $a = \hat{y}_t = \sum_{i=1}^{2} \widehat{\bar{b}}_i m_i x_i$ the agent's forecasting rule and action, we rewrite (11) as

$$\max_{m_1, m_2} u = -\frac{1}{2}\left(a - \bar{b}_1 w_t^1 - \bar{b}_2 w_t^2 - \eta_t\right)^2 .$$

Then, following Gabaix (2014), we use lasso estimator to derive the optimal attention vector m :

$$m = \arg\min_{m \in [0,1]^n} \frac{1}{2} \sum_{i,j=1\ldots n} (1-m_i)\Lambda_{ij}\left(1-m_j\right) + \kappa \sum_{i,j=1\ldots n} |m_i| \tag{12}$$

, s.t. $m_i \leq 1$ and for $n=2$. The loss from inattention is given by $\Lambda_{ij} = -\sigma_{ij} a_{w_i} u_{aa} a_{w_j}$, while the parameter κ governs the attention cost. Loss from inattention reflects how much of variation in the process we lose when neglecting the variable. In the formula (12), σ_{ij} denotes the (i,j) element

[2] If *all* agents start using both shocks in their forecasting rule, using least-squares learning from then onward, then their PLM coefficients will converge to those of MSV solution (2), (3), as was shown in Evans and Honkapohja (1994). However, we are asking whether an *atomistic* agent could find it advantageous to use both shocks for forecasting.

of shocks' variance-covariance matrix Σ^w, $a_{w_i} = -u_{aa}^{-1} u_{aw_i}$ determines by how much a change in a variable, w_i, changes the agent's action a, and

$$\begin{aligned} u_{aa} &= \frac{\partial^2 u_a}{\partial a^2} = \frac{\partial}{\partial a}\left(-(a - \bar{b}_1 w_t^1 - \bar{b}_2 w_t^2)\right) = -1, \\ u_{aw_i} &= \bar{b}_i. \end{aligned}$$

Then $a_{w_i} = \bar{b}_i$, and the cost of inattention is therefore given by $\Lambda_{ij} = \sigma_{ij} \bar{b}_i \bar{b}_j$. For the inner solution, taking the derivatives of (12) with respect to m_1 and m_2 gives the following expressions (details and corner solutions are in Appendix A.2):

$$m_1 = 1 - \frac{\kappa}{\bar{b}_1 \bar{b}_2 \sigma_1 \sigma_2 \left(1-\rho^2\right)} \frac{\bar{b}_2 \sigma_2 - \bar{b}_1 \rho \sigma_1}{\bar{b}_1 \sigma_1}, \tag{13}$$

$$m_2 = 1 - \frac{\kappa}{\bar{b}_1 \bar{b}_2 \sigma_1 \sigma_2 \left(1-\rho^2\right)} \frac{\bar{b}_1 \sigma_1 - \rho \bar{b}_2 \sigma_2}{\bar{b}_2 \sigma_2}. \tag{14}$$

One can immediately observe that both weights are falling with attention costs, κ. If attention costs are large enough, agents would choose not to pay attention to any variable (and use only constant term in their forecasts). Both weights are decreasing with the ALM coefficients on another variable: e.g., the weight on the first shock is lower if $\bar{b}_2$ is larger and vice versa.

Proposition 2.0.2. *The condition for $m_1 > m_2$ coincides with the condition for $MSFE_1 < MSFE_2$:*

$$\bar{b}_1^2 \sigma_1^2 > \bar{b}_2^2 \sigma_2^2,$$

which is equivalent to condition (10).

Proof. Appendix A.2. □

Proposition 2.0.2 states that as long as using consistent with RPE1 PLM (4) produces smaller forecast errors than using PLM (5), sparsely rational agents optimally pay more attention to the variable w_t^1 than to w_t^2.

For the agents to stick to the first mis-specified rule, the weight on the shock w_t^2 must be zero. In the RE MSV equilibrium, the weights on the both shocks must be significantly larger than zero.

Proposition 2.0.3. *A shock i gets non-zero weight in agents' forecast when for $i \neq j$:*

$$\kappa \leq \frac{\left(1-\rho^2\right) \bar{b}_i^2 \sigma_i^2}{1 - \rho \frac{\bar{b}_i \sigma_i}{\bar{b}_j \sigma_j}} \tag{15}$$

Proof. Re-arranging (14),

$$\frac{\kappa}{\bar{b}_j \bar{b}_i \sigma_j \sigma_i \left(1-\rho^2\right)} \frac{\bar{b}_j \sigma_j - \rho \bar{b}_i \sigma_i}{\bar{b}_i \sigma_i} \leq 1 \Rightarrow \kappa \leq \frac{\left(1-\rho^2\right) \bar{b}_i^2 \sigma_i^2}{1 - \rho \frac{\bar{b}_i \sigma_i}{\bar{b}_j \sigma_j}}.$$

Appendix A.2 shows that if a solution with non-zero weight exist, it outperforms corner solutions with zero weight. That is why, condition (15) is sufficient for a non-zero weight on the shock constituting PLM. □

Proposition 2.0.3 also states that if the agents use PLM consistent with RPE1 and costs of attention are large enough, agents choose not to pay attention to any variable.

We could interpret the condition (15) as follows. Consider the ALM (6). It includes two normally distributed variables, $\tilde{w}_t^1 = \bar{b}_1 w_t^1$ and $\tilde{w}_t^2 = \bar{b}_2 w_t^2$. The correlation coefficient between $\tilde{w}_t^1$ and $\tilde{w}_t^2$ is equal to ρ. Then $\rho \frac{\bar{b}_2 \sigma_2}{\bar{b}_1 \sigma_1}$ represents the coefficient in a regression of $\tilde{w}_t^2$ on $\tilde{w}_t^1$, while $\left(1-\rho^2\right) \bar{b}_2^2 \sigma_2^2$ is the variance of the conditional distribution of $\tilde{w}_t^2$ given $\tilde{w}_t^1$. Thus, if the cost of attention, κ, corrected for the information about $\tilde{w}_t^2$ already contained in (4), is larger than the variance of omitted information — variance of $\tilde{w}_t^2$ conditional on $\tilde{w}_t^1$, then the agents find it beneficial not to include the second shock into their forecasts. The condition (15), then, has an intuitive economic interpretation of equalizing costs and benefits of considering that portion of information, contained in the second shock w_t^2, which is above and beyond that which is already evaluated, given that w_t^1 is taken into account in the forecast.

With correlation close to zero, the second shock is included into the forecasting rule even for larger attention costs. Note, that this effect is more pronounced for positive correlation. This is because in (2) costs of innattention increase with larger correlation, forcing an increase in both m_2 and m_1. The same effect is observed for a decrease in κ.

For a more intuitive representation, we now introduce a measure which is not affected by the absolute values of the shock variances and persistences, but only by their ratio. For this purpose, we normalize the attention costs in (15) by $\bar{b}_2^2 \sigma_{\varepsilon 2}^2$, where $\sigma_{\varepsilon 2}^2$ is the variance of ε_t^2, the innovation to the shock w_t^2. We define the cost-to-variance ratio as $f \equiv \frac{\bar{\kappa}}{\bar{b}_2^2 \sigma_{\varepsilon 2}^2}$ where $\bar{\kappa}$ is the threshold for the second shock to be included into the agents' rule. For $\kappa < \bar{\kappa}$, agents use both shocks in their forecasting rules, and for $\kappa \geq \bar{\kappa}$ agents stick to the rule with the first shock only. We plot this ratio, f, in figure 2 in the coordinates $\left(\rho, \log(\Gamma\tilde{\Sigma})\right)$, where $\Gamma\tilde{\Sigma} \equiv \frac{\gamma_1 \sigma_{\varepsilon_1}}{\gamma_2 \sigma_{e_2}}$ is the relative importance of the first shock's innovation, ε_t^1, in the true data-generating process (1). We do this for different values of ρ_1 and ρ_2, autocorrelation coefficients of the shocks w_t^1 and w_t^2. The agents include w_t^2 into forecasting rule when the normalized attention cost, $\frac{\kappa}{\bar{b}_2^2 \sigma_{\varepsilon 2}^2}$, is smaller than f. Thus, larger f means that the range of costs consistent with w_t^2 used for forecasting is wider.

In figure 2, white area corresponds to the parameters' values where RPE1 is not selected as it produces larger forecast error than RPE2. Larger persistence of w_t^2, ρ_2, decreases the parameter space where $MSFE_1 < MSFE_2$, while increasing ρ_1 expands this area. Higher ρ_1 increases the share of the variance of y_t explained by the first shock, and thus decreases the value of taking into account the second one. Higher ρ_2 has the opposite effect.

Large correlation between the shocks, $|\rho|$, also contributes to better forecasting performance of RPE1-consistent PLM (A1) so that $MSFE_1 < MSFE_2$. As described in Appendix A.1, the $MSFE_1 < MSFE_2$ condition is satisfied for positive ρ if $\rho > \frac{1-\beta\rho_1-\Gamma\Sigma}{\beta\rho_1}$, which is the upper bound for white area in all the graphs. As one can see in the figure for $\rho_1 = 0.9$, in case of large persistence of the first shock so that $\beta\rho_1 > 1/2$, RPE1 PLM outperforms RPE2 also for very negative correlation such that $\rho < \frac{-(1-\beta\rho_1)-\Gamma\Sigma}{\beta\rho_1}$. When the MSFE criterion (10) is satisfied, large absolute correlation increases the variation in y_t explained by the first shock alone making right hand side of (15) smaller. When the threshold for cost-to-variance ratio is small, it means that the second shock is added to the PLM only for very small learning costs to variance ratio. The effect of correlation is illustrated on the Figure **??**, we we plot inattention costs and corresponding isocost line (attention costs) for different

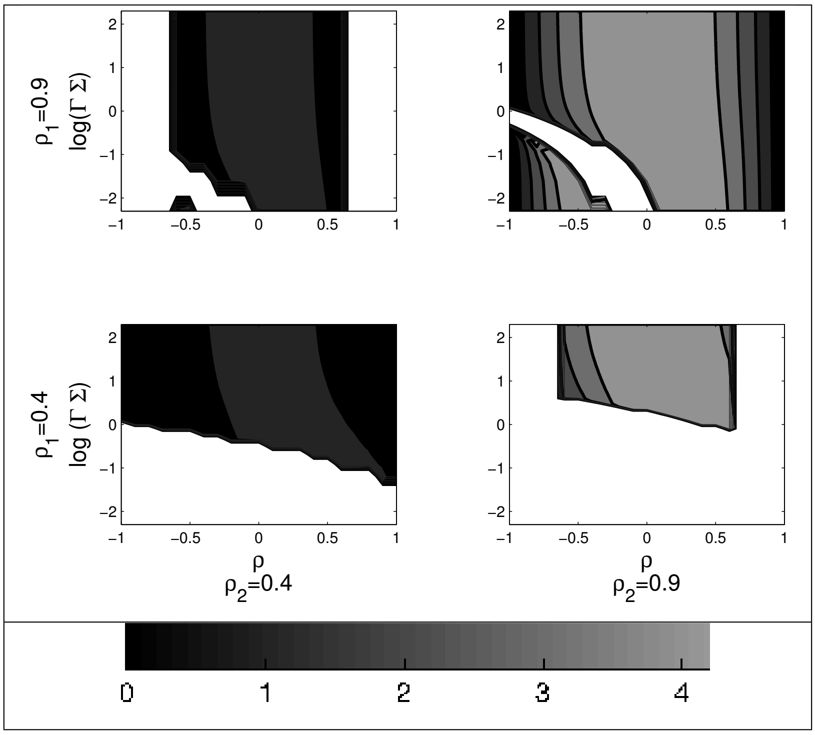


Figure 1: Threshold for the Cost-to-Variance Ratio

values of correlation, ρ. Finally, note that even under condition $\log(\Gamma\Sigma) > 0$, which means $\Gamma\Sigma > 1$, so that σ_t^1 plays more important role in the explaining variation in (1) than σ_t^2, and both ρ_1 and ρ are large so that taking into account w_t^1 is very informative (see lower left panel of figure 2), there could be still attention costs low enough for w_t^2 to be included into sparsely rational agents PLM. Therefore, unconstrained agents with $\kappa = 0$ will always find it beneficial to include w_t^2 into their forecasting rules.

In the next section we move to a three-equation New Keynesian model and study possible policy implications.

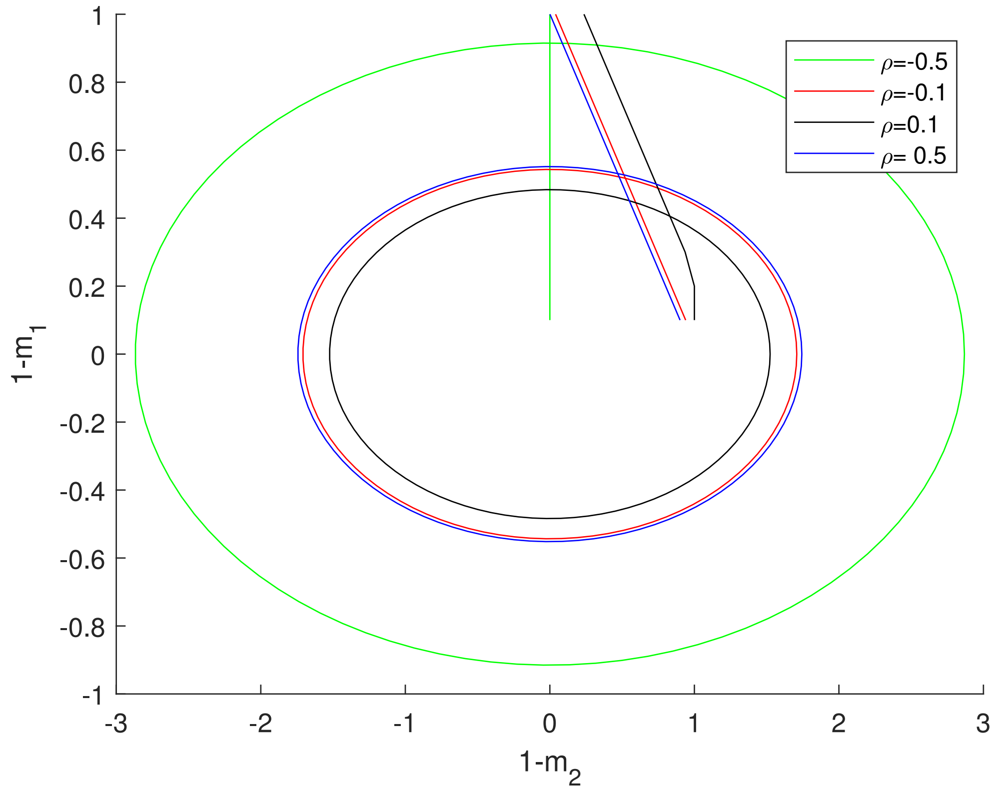


Figure 2: Inattention costs as a function of correlation, ρ.

The numerical values used for the plot $\kappa = 0.01$, $\rho_1 = \rho_2 = 0.9$, $\Gamma\Sigma = 3.15$

3. Three-Equation New Keynesian Model

In this section we employ the textbook three-equation New Keneysian Model (as in Galí 2015, section 3). We first show that the conditions derived in the previous section can be generalized to this model, and then discuss the policy implications.

The reduced, linearized around the deterministic steady state, solution of the model takes the form:

$$y_t = -\frac{1}{\sigma}(i_t - E_t\pi_{t+1}) + E_t y_{t+1} + g_t, \tag{16}$$

$$\pi_t = \beta E_t \pi_{t+1} + \omega y_t + u_t, \tag{17}$$

where $\pi_t = p_t - p_{t-1}$ is inflation and y_t is output gap. g_t and u_t are Gaussian shocks with certain variance. We assume that agents do not observe current shock realization, that is $E_t(g_t) = 0$ and $E_t(u_t) = 0$.

Parameter ω is a combination of the underlying model parameters :

$$\omega = \frac{(1-\theta)(1-\beta\theta)}{\theta}\frac{1-\alpha}{1-\alpha+\alpha\varepsilon}(\sigma + \frac{\phi+\alpha}{1-\alpha}), \tag{18}$$

with θ is Calvo probability, ε is demand elasticity and α governs the return to labor in the production function, σ is consumers risk-aversion, and ϕ is Inverse Frisch elasticity of labor supply. The production function takes the form:

$$Y_t = A_t N_t^{1-\alpha}. \tag{19}$$

If we assume constant return to scale in the production function, $\alpha = 0$, expression for ω collapses to

$$\omega = \frac{(1-\theta)(1-\beta\theta)}{\theta}(\sigma+\phi), \tag{20}$$

leaving us with Calvo probability and 3 standard parameters of utility function. We make another simplifying assumption that agents use naive forecast of output gap, so that our agents have to forecast only inflation:

$$E_t y_{t+1} = y_{t-1}. \tag{21}$$

In the framework of adaptive learning such an assumption can be viewed as an approximation of constant gain learning, where gain on output gap is larger then gain on inflation.[3]

The monetary authority sets interest rates in reaction to inflation expectations. One may consider alternative rules of the form:

$$i_t = \phi_\pi(E_t\pi_{t+1} - \bar{\pi}) + \bar{\pi}, \tag{22}$$
$$i_t = \phi_\pi(E_t\pi_t - \bar{\pi}) + \bar{\pi}, \tag{23}$$

where $\bar{\pi}$ is steady state inflation or inflation target. That is, monetary authority reacts either to deviation of future or of current inflation expectations from the target.

Rational expectations equilibria and their properties under different policy rules are formalized in the proposition 3.0.1.

Proposition 3.0.1. *Under forward looking policy rule, the solution to the model is:*

$$y_t = a_y + b_y E_t\pi_{t+1} + c_y y_{t-1} + g_t, \tag{24}$$
$$\pi_t = a_\pi + b_\pi E_t\pi_{t+1} + c_\pi y_{t-1} + \omega g_t + u_t, \tag{25}$$

with the coefficients defined in (B3) - (B8). The rational expectation equilibrium solution to (24) and (25) of the form $\hat{\pi}_t = \alpha + b\pi_{t-1} + cy_{t-1}$ *with the coefficients*

$$\alpha = -\frac{a_\pi}{b_\pi}, \tag{26}$$
$$b = \frac{1}{b_\pi}, \tag{27}$$
$$c = -\frac{c_\pi}{b_\pi}, \tag{28}$$

is not E-stable. The minimum state variable solution (MSV) of the form $\hat{\pi}_t = \alpha + cy_{t-1}$ *with the coefficients*

$$\alpha = \frac{a_\pi}{1-b_\pi}, \tag{29}$$
$$b = 0, \tag{30}$$
$$c = \frac{c_\pi}{1-b_\pi}, \tag{31}$$

[3] A simple constant gain learning was found to give the best fit for survey of professional forecasters in Branch and Evans (2006a).

is determinate and E-stable for $\phi_\pi > 1$.

Under the rule with contemporaneous inflation expectations, the solution to the model is:

$$y_t = a_y + b_y^c E_t \pi_t + b_y^f E_t \pi_{t+1} + y_{t-1} + g_t, \tag{32}$$
$$\pi_t = a_\pi + b_\pi^c E_t \pi_t + b_\pi^f E_t \pi_{t+1} + \omega y_{t-1} + \omega g_t + u_t, \tag{33}$$

with the coefficients defined in (B46) - (B53). The rational expectation equilibrium solution to (33) and (32) of the form $\hat{\pi}_t = \alpha + b\pi_{t-1} + cy_{t-1}$ *with the coefficients*

$$\alpha = -\frac{a_\pi}{b_\pi^f}, \tag{34}$$
$$b = \frac{1 - b_\pi^c}{b_\pi^f}, \tag{35}$$
$$c = -\frac{c_\pi}{b_\pi^f}, \tag{36}$$

is not E-stable. The minimum state variable solution (MSV) of the form $\hat{\pi}_t = \alpha + cy_{t-1}$ *with the coefficients:*

$$\alpha = \frac{a_\pi}{1 - b_\pi^f - b_\pi^c}, \tag{37}$$
$$b = 0, \tag{38}$$
$$c = \frac{c_\pi}{1 - b_\pi^f - b_\pi^c}, \tag{39}$$

is determinate and E-stable for $\phi_\pi > 1$.

Proof. Proof is in the Appendix B.1. □

As in the previous section, we restrict our agents to use only one variable in their forecasting rules. In the context of this model, however, one rule is mis-specified, while another would correspond to rational expectations minimum state variable solution (MSV). Similar to Adam (2005) we consider two RPE's, in which agents use either lagged inflation or lagged output gap in their forecasting rules. In M_π agents use the rule:

$$\hat{\pi}_t = \alpha_\pi + b_\pi \pi_{t-1}, \tag{40}$$

while in M_y it is:

$$\hat{\pi}_t = \alpha_y + c_y y_{t-1}. \tag{41}$$

M_y rule could nest the MSV solution, while M_π is mis-specified as the lag of inflation does not enter the law of motion (B2). Yet, if the agents stick to M_π, the lag of inflation affects the actual law of motion through expectational terms. Under each forecasting rule, the actual law of motion is different, because the way expectations are formed affects the actual process for inflation and output gap through $E_t\pi_{t+1}$. If the agents use M_y rule, the actual law of motion is that of MSV solution and the rule is an equilibrium choice. If, however, agents use M_π, there is a region in the parameter space where this mis-specified rule is preferred in equilibrium. For a mis-specified rule to be an equilibrium choice, it should result in smaller forecast errors than the alternative rule.

Proposition 3.0.2. *For the model described by (16), (17), and policy rule as in (22) or in (23), the condition for M_π to be the equilibrium, $MSFE_\pi$ <$MSFE_y$, is:*

$$\bar{b}_\pi^2 \sigma_\pi^2 > \bar{c}_\pi^2 \sigma_y^2, \tag{42}$$

where $\bar{b}_\pi$ and $\bar{c}_\pi$ are ALM coefficients on lag of inflation and output gap respectively, and σ_π^2 and σ_y^2 are corresponding variances.

Proof. Proof is in the Appendix B.2. □

Rewriting (42), we can see that M_π results in smaller mean forecast squared errors than M_y if:

$$\bar{\Gamma}^2 \Sigma^2 > 1 \tag{43}$$

where $\Sigma = \frac{\sigma_\pi}{\sigma_y}$ and $\bar{\Gamma} = \frac{\bar{b}_\pi}{\bar{c}_\pi}$. That is, if larger share of inflation variation (B68) is explained under M_π.

Proposition 3.0.2 also states that the criterion for a mis-specified PLM to be an equilibrium, (42), is similar to the condition (10) from the simple model.

3.1 Attention Weights

Suppose that the actual law of motion is induced by M_π rule given in (B68)-(B69), and the agents could reconsider their forecasting rules. Will they give a significant weight to the past output? In other words, would they stick to the mis-specified rule, or move to the RE solution?[4] Or, if attention costs are present, will the agents choose not to pay attention to any of the variables? To answer these questions, we let the agents select the sparse weights as in (11) by minimizing $\hat{\pi}_t - \pi_t$ subject to attention cost. The difference from the simple case in Section 2 is that the variables used for forecasting are endogenous now. Denoting $a = \hat{\pi}_t = m_y \bar{c}_\pi y_{t-1} + m_\pi \bar{b}_\pi \pi_{t-1}$ to be the agents forecasting rule , $x = (\pi_{t-1}, y_{t-1})^T$, we rewrite utility as:

$$u = -\frac{1}{2}(a - b_\pi \pi_{t-1} - c_\pi y_{t-1})^2. \tag{44}$$

To find the sparse weights, agents minimize (12), where the cost of inattention, $\Lambda_{ij} = -\sigma_{ij} a_{w_i} u_{aa} a_{w_j}$, is now modified to account for the new variables and covariance:

$$\begin{aligned}
u_{aa} &= \frac{\partial^2 u_a}{\partial a^2} = \frac{\partial}{\partial a}\left(-(a - \bar{b}_\pi \pi_{t-1} - \bar{c}_\pi y_{t-1})\right) = -1, \\
u_{a\pi_{t-1}} &= \bar{b}_\pi, \\
u_{ay_{t-1}} &= \bar{c}_\pi, \\
a_{\pi_{t-1}} &= -u_{aa} u_{a\pi_{t-1}} = -\bar{b}_\pi, \\
a_{y_{t-1}} &= -u_{aa} u_{ay_{t-1}} = -\bar{c}_\pi.
\end{aligned}$$

Then the cost of inattention is: $\Lambda = \begin{pmatrix} \sigma_\pi^2 \bar{b}_\pi^2 & \sigma_{\pi y} \bar{b}_\pi \bar{c}_\pi \\ \sigma_{\pi y} \bar{b}_{\pi \bar{c}_\pi} & \sigma_y^2 \bar{c}_\pi^2 \end{pmatrix}$, where σ_π^2, σ_y^2 and $\sigma_{\pi y}$ are variance of inflation, output and their covariance respectively. Under M_π they are derived in Appendix B.2.

[4] Again, if agents start to include lagged output in their forecasting rules and employ least squares learning onward, they eventually learn the coefficient on past π and y, with the former being zero.

Taking the first order conditions of (12) and solving for weights results in the following expressions:

$$m_y = 1 - \frac{\kappa}{\bar{c}_\pi^2 \sigma_y^2 \sigma_\pi \left(1 - R^2\right)} \frac{\bar{b}_\pi \sigma_\pi - \bar{c}_\pi \sigma_y R}{\bar{b}_\pi}, \tag{45}$$

$$m_\pi = 1 - \frac{\kappa}{\bar{b}_\pi^2 \sigma_\pi^2 \sigma_y \left(1 - R^2\right)} \frac{\bar{c}_\pi \sigma_y - \bar{b}_\pi \sigma_\pi R}{\bar{c}_\pi}, \tag{46}$$

where $R = \frac{\sigma_{\pi y}}{\sigma_\pi \sigma_y}$ is the correlation between y_t and π_t in RPE_π, equilibrium consistent with agents using M_π rule.

Proposition 3.0.3. *Whenever $MSFE_\pi < MSFE_y$, past inflation gets larger weight than past output:*

$$m_\pi > m_y.$$

Proof. Proof is in Appendix C. □

Note, that the result is identical to condition (43).

The next proposition summarizes the condition for output to get a positive weight in agents' forecast.

Proposition 3.0.4. *Lag of output gets positive weight in agents' forecast when*

$$\kappa \leq \frac{\left(1 - R^2\right) \bar{c}_\pi^2 \sigma_y^2}{1 - R \frac{\bar{c}_\pi \sigma_y}{\bar{b}_\pi \sigma_\pi}} \tag{47}$$

Proof. Proof is in Appendix C. □

Notice the similarity of (47) with the (15).

3.2 Policy Rules and Forecasting Rules

Below we consider the interaction between monetary policy and equilibrium selection of forecasting rules. For all numerical simulations we use textbook calibration with $\beta = 0.99$, coefficient of risk aversion $\sigma = 1$, Frish elasticity of labor supply $\phi = 1$, Calvo probability $\theta = 2/3$, and inflation target $\bar{\pi} = 3$ [5]. For some of the graphs we use ratio of shock deviations, $r \equiv \frac{\sigma_u}{\sigma_g}$, where u and g are innovations in the IS and Philips curves respectively, cf. (16) and (17). Without loss of generality, we set $\sigma_g = 1$ in our simulations.

Figure 3 shows the dimensionless threshold for the cost-to-variance ratio for both policy rules, defined as:

$$f \equiv \frac{\bar{\kappa}}{\bar{c}_\pi^2 \sigma_y^2} = \frac{\left(1 - R^2\right)}{1 - R \frac{\bar{c}_\pi \sigma_y}{\bar{b}_\pi \sigma_\pi}}. \tag{48}$$

Here $\bar{\kappa}$ is the maximum costs of attention at which the agents are willing to include output gap in their forecasts. Figure 3 shows that when attention costs are present, initially mis-specified forecasting rule can be supported under sparse-rationality. With a very aggressive policy (ϕ_π is close to 2),

[5] These values are taken from Galí 2015, section 3.

the mis-specified model is never the equilibrium choice as it produces larger MSFE than M_y. This region is colored blue in Figure 3. In the parameter region, where the mis-specified rule is chosen, the stronger is the policy reaction, the larger is the range of attention costs consistent with the lagged output gap being included in the forecasting rule: the agents choose to include both output gap and inflation even for larger attention costs. This is evident from the larger area under threshold for the cost-to-variance, represented by the blue solid line. The intuition for this can be found in ALM coefficients. The coefficient on the past inflation in the ALM for inflation, green dashed line, is decreasing as ϕ_π increases, meaning that inflation becomes predicted better by lagged output gap, and worse by its own past value. With an increasing importance of output gap in predicting inflation, it is not surprising that the agents agree to pay more to include the output gap in their forecasting rules.

Correlation between output gap and inflation (dotted line in Figure 3) is increasing with policy aggressiveness. As was discussed in Section 2, the effect of correlation on weights is not linear. When inflation is persistent (coefficient on its lag is large in the ALM) and volatile, larger correlation contributes to smaller weight on output gap. When inflation becomes less persistent and volatile, agents assign larger weight with larger correlation.

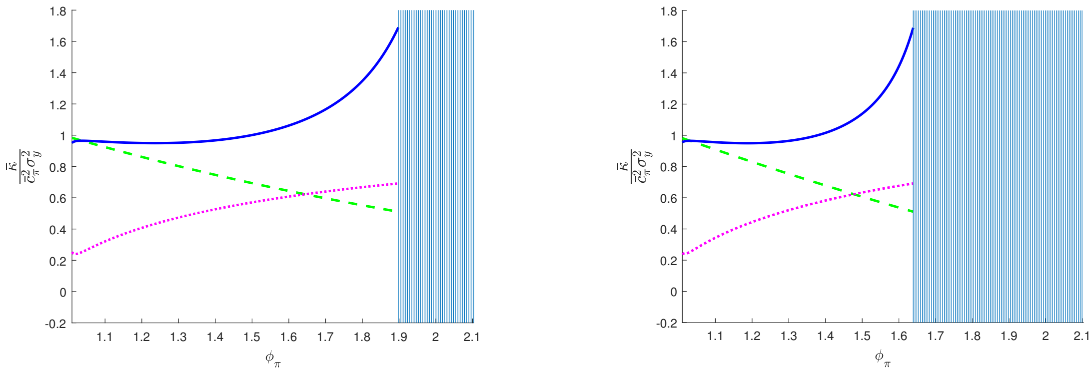
The two panels on Figure 3 are rather similar, demonstrating that there is no significant difference between contemporaneous data-based and expectations-based policy rules. Under contemporaneous policy rule, (23), the mis-specified equilibrium PLM, M_π, can be avoided with less aggressive monetary policy, as for any ϕ_π, the threshold attention cost-to-variance ratio is larger; therefore, the agents find it more important to pay attention to the output with the rule (23).

Figure 4 shows the weights m_i the agents choose for the different values of the learning costs, κ, and policy parameter, ϕ_π, while fixing the ratio of deviations in innovations to $r = 0.1$.[6] When the weight on output gap is zero but is positive on inflation, the region is defined as M_π, as it corresponds to the agents sticking to an initially mis-specified rule. For large attention costs, κ, and large ϕ_π, the agents choose to have zero weights on both variables and use only the constant in their forecasting rules. This region in the upper-right corner of Figure 4 is defined as (0,0). When both weights are positive, an increasing ϕ_π first decreases both weights, as aggressive policy decreases volatility of output and inflation, and the agents choose to pay less attention to both variables. Then, with even more aggressive policy, volatility of inflation falls more than volatility of output, weight on output starts to increase, while weight on inflation continues falling. Figure D1 in the Appendix D presents the graph, similar to the Figure 4, which plots m_π as a function of (ϕ_π, κ).

Figure 5 shows standard deviations of inflation and output gap under both policy rules. Note, that the U-shaped form of M_π region in Figure 4 depends on which variable has larger volatility in Figure 5. While volatility of the both variables declines with an increasing policy parameter, ϕ_π, volatility of inflation is initially larger. In this region, both weights are falling with the tighter policy. As the policy gets more aggressive and the output volatility becomes larger than that of inflation, weight on output starts to increase with the policy response, while weight on inflation starts to fall. For any strength of the policy response, the rule with contemporaneous inflation results in less overall volatility. This finding is consistent with the experimental results in Pfajfar and Žakelj (2017), where the contemporaneous rule is found to produce smaller volatility than the forward-looking rule. The large inflation volatility in Figure 5 can be linked to large forecast errors of output gap.

[6] With the standard deviation of output shock σ_g fixed, larger r means larger relative inflation volatility. This results in smaller estimates of coefficient on inflation, and smaller weights. With larger r the Figure 4 and Figure D1 look similar, but the agents move to REE (m_y>0) with smaller policy parameter, ϕ_π.

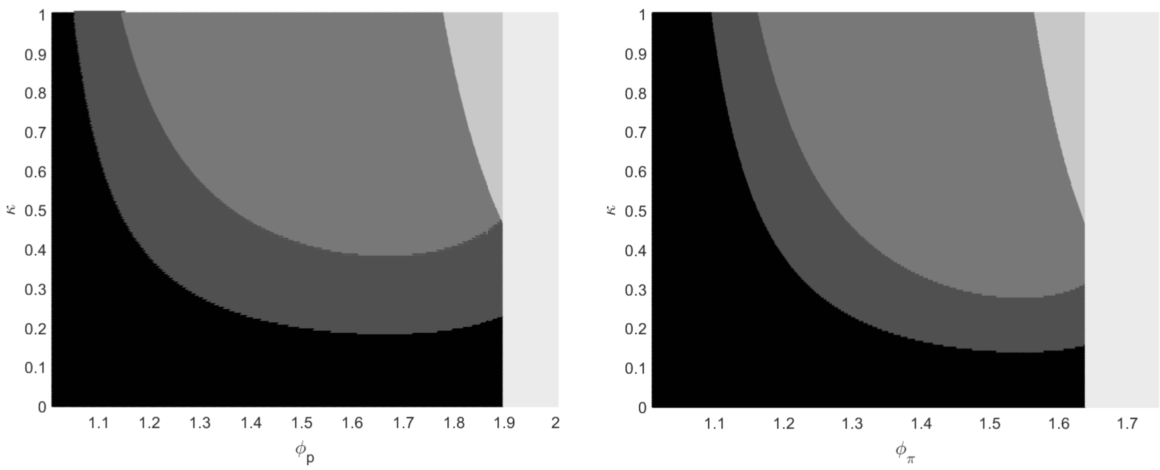
Figure 3: Threshold for the Cost-to-Variance Ratio



(a) Forward Looking Rule ***(b) Contemporaneous Rule***

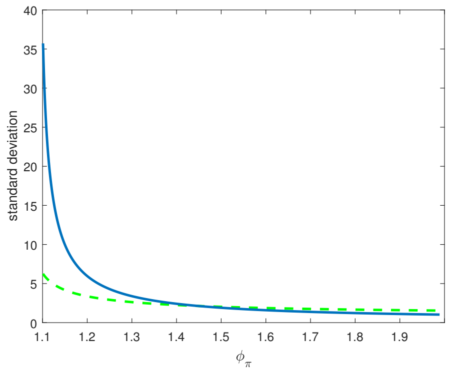
Note: dotted line corresponds to the correlation between output and inflation, dashed line - to the ALM coefficient on the past inflation, blue solid line - to the threshold for cost-to-variance ratio, blue-colored region shows the area where MSFE$_\pi$>MSFE$_y$, r=0.1.

Figure 4: Model Selection under Sparse Weights, M_y

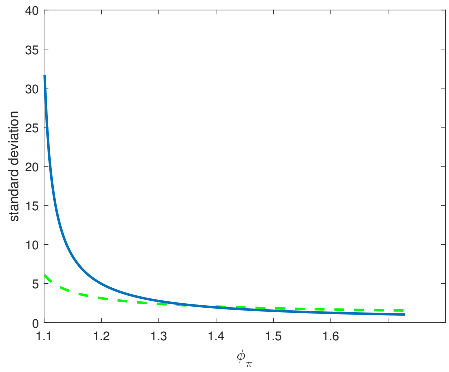


(a) Forward Looking Rule ***(b) Contemporaneous Rule***

Figure 5: Equilibrium dynamic of inflation and output



(a) Forward Looking Rule



(b) Contemporaneous Rule

Note: dashed line corresponds to output gap, solid line - to inflation, r=0.1

Our results contribute to the literature discussing how monetary policy affects not only the level of expectations, but the way the expectations are formed. In our model a more restrictive monetary policy decreases the parameter range of stability of M_π rule, both in favor of MSV-nesting rule and the intercept-only rule: the agents either start taking into account the output gap if the κ is small or disregard any variables if the κ is larger. The result is intuitive, since in the model we study the strength of the policy, ϕ_π, decreases the feedback from expectations in the actual law of motion.[7] The role of the feedback parameter is widely studied in the literature. In an experimental setting Hommes (2014) and Heemeijer et al. (2009) emphasize the importance of expectations feedback parameter. In Hommes (2014) when the expectations feedback parameter is negative, convergence to REE is observed, but with a positive parameter agents coordinate on a non-rational self-fulfilling equilibrium. Large feedback parameter results in convergence to a mis-specified equilibrium in Evans et al. (2012). In the laboratory experiments of Pfajfar and Žakelj (2014), Pfajfar and Žakelj (2017), and Assenza et al. (2013), agents choose forecasting rules of inflation under alternative monetary policy regimes, differing by the aggressiveness of response to inflation in the Taylor rule. As monetary policy becomes more aggressive, more agents switch to using forecasting rules compatible with rational expectations. As Pfajfar and Žakelj (2017) discusses, the aggressiveness of monetary policy response plays the role of the expectation feedback. This supports our finding that for low κ and large ϕ_π agents increase their weight on lagged output gap.

Another prediction of the paper is that agents stick to AR(1) model for inflation forecasting when persistence of inflation is high and/or the correlation between output and inflation is low, see Figure 3. This result is in line with the observed behavior of professional forecasters after the recent financial crisis. There are number of studies, examples being Fendel et al. (2011), Lopez-Perez (2017), and Frenkel et al. (2011), showing that professional forecasters' predictions behave as if they are using Phillips curve. After the financial crisis, inflation became more persistent, cf. Watson (2014), while Phillips curve got flatter.[8] In terms of our model, this means lower threshold for cost-

[7] To see this, consider the coefficient on inflation expectations in the actual law of motion for inflation in (B2) and (B45) for forward looking and contemporaneous rules, respectively.

[8] Although, the evidence on flattening of the Phillips curve is mixed due to different specifications and time horizons considered, there are studies showing the decline in slope, examples being IMF (2013) and Kuttner and Robinson (2010) Donayre and Panovska (2016) document breaks in the wage Phillips curve during the recessions and the subsequent recoveries.

to-variance ratio in Figure 3. Lopez-Perez (2017) [9] shows that forecasters' predictions started to react much less to the unemployment after the financial crisis 2007-2009, consistent with our model predictions.

4. Conclusion

In this paper we study if the initially mis-specified forecasting rule can be an equilibrium choice under sparse-rationality. We first consider a simple process consisting only of exogenous variables, and then generalize our results to the three-equation New Keynesian model with lagged endogenous variables, arising due to the agents' expectation.

For both models we find regions in the parameter space, where mis-specified RPE is selected by both minimum squared forecast error condition and the sparse-weights. Thus, the agents who are re-considering their initial choice of a mis-specified RPE, could select to continue using an initial rule. If learning costs are very large, there is a region of parameter space where agents choose not to allocate attention to any of the variables. If learning costs are small, agents tend to switch to the REE, and do this certainly when costs of attention are zero. For medium range of the learning costs, an initial mis-specified forecasting rule prevails for large persistence of the variable used in the rule and for large correlation between the included and the omitted variables, especially if the omitted variable has low persistence. This behavior is explained by the amount of additional information that is contained in the omitted variable.

The prediction from a New Keynesian model is that when inflation persistence increases, the survival of the mis-specified AR(1) rule for inflation forecasting is achieved for broader area of the parameter space. The same is true for the small correlation between inflation and output gap. This prediction is supported by the professional forecasters' behavior, whose predictions are found to be consistent with paying less attention to output gap after the financial crisis (Lopez-Perez 2017), when inflation became more persistent and the correlation between output and inflation might have changed.

We find that aggressiveness of the monetary policy rule, and to some extent, the rule itself, determine the survival of a mis-specified forecasting rule. Strong monetary policy reaction reduces inflation volatility and persistence, making adding output gap to the forecasting rule more attractive. With even larger monetary policy reaction, agents do not consider the mis-specified forecasting rule at all, because the REE-consistent one results in smaller forecast errors. In line with the previous literature, our study supports the importance of expectation feedback parameter for survival of a mis-specified rule. In our model the feedback parameter is decreasing with policy rule parameter. If this expectation feedback is large enough, the mis-specified forecasting rule prevails in equilibrium.

[9] Frenkel et al. (2011) uses data up to 2010Q and does not find evidence for a change in forecasters behavior, while Lopez-Perez (2017) uses longer data set and includes a forward looking inflation term in the Philips curve.

References

ADAM, K. (2005): "Learning To Forecast And Cyclical Behavior Of Output And Inflation." *Macroeconomic Dynamics,* 9(01):1–27.

ADAM, K. (2007): "Experimental Evidence on the Persistence of Output and Inflation." *Economic Journal,* 117(520):603–636.

ASSENZA, T., P. HEEMEIJER, C. HOMMES, AND D. MASSARO (2013): "Individual Expectations and Aggregate Macro Behavior." Tinbergen Institute Discussion Papers 13-016/II, Tinbergen Institute

BRANCH, W. A. AND G. W. EVANS (2006): "A simple recursive forecasting model." *Economics Letters,* 91(2):158 – 166.

BRANCH, W. A. AND G. W. EVANS (2006): "Intrinsic heterogeneity in expectation formation." *Journal of Economic Theory,* 127(1):264–295.

BULLARD, J. AND K. MITRA (2002): "Learning about monetary policy rules." *Journal of Monetary Economics,* 49(6):1105 – 1129.

DONAYRE, L. AND I. PANOVSKA (2016): "Nonlinearities in the U.S. wage Phillips curve." *Journal of Macroeconomics,* 48(C):19–43.

EVANS, G. W. AND S. HONKAPOHJA (1994): "Learning, convergence, and stability with multiple rational expectations equilibria." *European Economic Review,* 38(5):1071–1098.

EVANS, G. W. AND S. HONKAPOHJA (2009): "Learning and Macroeconomics." *Annual Review of Economics,* 1(1):421–451.

EVANS, G. W. AND S. HONKAPOHJA (2001): *Learning and Expectations in Macroeconomics.* Princeton University Press, Princeton.

EVANS, G. W., S. HONKAPOHJA, T. SARGENT, AND N. WILLIAMS (2012): "Bayesian Model Averaging, Learning and Model Selection." CDMA Working Paper Series 201203, Centre for Dynamic Macroeconomic Analysis

FENDEL, R., E. M. LIS, AND J. C. RÜLKE (2011): "Do professional forecasters believe in the Phillips curve? Evidence from the G7 countries." *Journal of Forecasting,* 30(2):268–287.

FRENKEL, M., E. M. LIS, AND J. C. RÜLKE (2011): "Has the economic crisis of 2007-2009 changed the expectation formation process in the Euro area?." *Economic Modelling,* 28(4): 1808–1814.

GABAIX, X. (2014): "A Sparsity-Based Model of Bounded Rationality." *The Quarterly Journal of Economics,* 129(4):1661–1710.

GABAIX, X. (2016): "A Behavioral New Keynesian Model." Working Paper 22954, National Bureau of Economic Research

GALÍ, J. (2015): *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications Second edition.* Princeton University Press, 2 edition.

GARCÍA-SCHMIDT, M. AND M. WOODFORD (2015): "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis." NBER Working Papers 21614, National Bureau of Economic Research, Inc

HEEMEIJER, P., C. HOMMES, J. SONNEMANS, AND J. TUINSTRA (2009): "Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation." *Journal of Economic Dynamics and Control,* 33(5):1052–1072.

HOMMES, C., D. MASSARO, AND M. WEBER (2017): "Monetary Policy under Behavioral Expectations: Theory and Experiment." Bank of Lithuania Working Paper Series 42, Bank of Lithuania

HOMMES, C. H. (2014): "Behaviorally Rational Expectations and Almost Self-Fulfilling Equilibria." *Review of Behavioral Economics,* 1(1-2):75–97.

IMF (2013): *The dog that didnt bark: has inflation been muzzled or was it just sleeping?* In *World Economic Outlook,* chapter 3, pages 1–17..

KUTTNER, K. AND T. ROBINSON (2010): "Understanding the flattening Phillips curve." *North American Journal of Economics and Finance,* 21(2):110–125.

LOPEZ-PEREZ, V. (2017): "Do professional forecasters behave as if they believed in the New Keynesian Phillips Curve for the euro area?" *Empirica,* 44(1):147–174.

MACKOWIAK, B. AND M. WIEDERHOLT (2009): "Optimal Sticky Prices under Rational Inattention." *American Economic Review,* 99(3):769–803.

MANKIW, N. G. AND R. REIS (2007): "Sticky Information in General Equilibrium." *Journal of the European Economic Association,* 5(2-3):603–613.

MARCET, A. AND T. J. SARGENT (1989): "Convergence of least squares learning mechanisms in self-referential linear stochastic models." *Journal of Economic Theory,* 48(2):337–368.

MATEJKA, F. AND A. MCKAY (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review,* 105(1):272–98.

MOLNAR, K. (2007): "Learning with Expert Advice." *Journal of the European Economic Association,* 5(2-3):420–432.

ORMENO, A. AND K. MOLNAR (2015): "Using Survey Data of Inflation Expectations in the Estimation of Learning and Rational Expectations Models." *Journal of Money, Credit and Banking,* 47(4):673–699.

PFAJFAR, D. AND B. ŽAKELJ (2014): "Experimental Evidence on Inflation Expectation Formation." *Journal of Economic Dynamics and Control,* 44:147–168.

PFAJFAR, D. AND B. ŽAKELJ (2017): "Inflation Expectations and Monetary Policy Design: Evidence from the Laboratory." *Macroeconomic Dynamics,* forthcoming:1–41.

SIMS, C. A. (2003): "Implications of rational inattention." *Journal of Monetary Economics,* 50 (3):665–690.

SLOBODYAN, S. AND R. WOUTERS (2012): "Learning in an estimated medium-scale DSGE model." *Journal of Economic Dynamics and Control,* 36(1):26–46.

WATSON, M. W. (2014): "Inflation Persistence, the NAIRU, and the great recession." *American Economic Review,* 104(5):31–36.

Appendix A: A simple of model

Proof. Proposition 2.0.1.

Deriving RPE1. The agents' PLM, consistent with this RPE, is

$$y_t = a_1 + b_1 w_t^1, \tag{A1}$$

therefore, the ALM is given by

$$y_t \quad = \quad \alpha + \beta a_1 + \bar{b}_1 w_t^1 + \gamma_2 w_t^2 + \eta_t, \tag{A2}$$

with $\bar{b}_1 = \beta\rho_1 b_1 + \gamma_1$ In what follows, we will set $\alpha = 0$ and assume that the agents know this; therefore, $a_1 = 0$ as well.

In order for the agents to be using (A1) in equilibrium, it must be the case that b_1 is a coefficient in the regression of y_t on w_t^1, or

$$b_1 = \frac{Cov\left(y_t, w_t^1\right)}{Var\left(w_t^1\right)}. \tag{A3}$$

Computing the above expression, we get

$$\begin{aligned} Cov\left(y_t, w_t^1\right) &= E_t\left[\bar{b}_1 w_t^1 + \gamma_2 w_t^2 + \eta_t, w_t^1\right] = \bar{b}_1 \sigma_1^2 + \gamma_2 \cdot \rho\sigma_1\sigma_2, && \text{(A4)}\\ b_1 &= \frac{\bar{b}_1\sigma_1^2 + \gamma_2 \cdot \rho\sigma_1\sigma_2}{\sigma_1^2} = \bar{b}_1 + \gamma_2 \cdot \rho\frac{\sigma_2}{\sigma_1} = && \text{(A5)}\\ &= \beta\rho_1 b_1 + \gamma_1 + \gamma_2 \cdot \rho\frac{\sigma_2}{\sigma_1} \Rightarrow b_1 = \frac{\gamma_1 + \gamma_2 \cdot \rho\frac{\sigma_2}{\sigma_1}}{1 - \beta\rho_1}. && \text{(A6)} \end{aligned}$$

Deriving RPE2. Similarly to the RPE1 case, we have now the PLM

$$y_t = a_2 + b_2 w_t^2, \tag{A7}$$

which implies that b_2 must be equal to the regression coefficient:

$$\begin{aligned} b_2 &= \frac{Cov\left(y_t, w_t^2\right)}{Var\left(w_t^2\right)} = \frac{E_t\left[\bar{b}_1 w_t^1 + \gamma_2 w_t^2 + \eta_t, w_t^2\right]}{\sigma_2^2} = && \text{(A8)}\\ &= \frac{(\beta\rho_1 b_1 + \gamma_1)\cdot\rho\sigma_1\sigma_2 + \gamma_2\sigma_2^2}{\sigma_2^2} = (\beta\rho_1 b_1 + \gamma_1)\cdot\rho\frac{\sigma_1}{\sigma_2} + \gamma_2. && \text{(A9)} \end{aligned}$$

E-stability. For solution in (A6) and (A9) to be E-stable the following should hold: $\frac{\partial T_{b_1}}{\partial b_1} < 1$ and $\frac{\partial T_{b_2}}{\partial b_2} < 1$. T_{b_1} is given by (A6) and T_{b_2} by (A9) . That is, $\frac{\partial T_{b_1}}{\partial b_1} = \frac{\partial\left[\beta\rho_1 b_1 + \gamma_1 + \gamma_2\cdot\rho\frac{\sigma_2}{\sigma_1}\right]}{\partial b_1} = \beta\rho_1$ and $\frac{\partial T_{b_2}}{\partial b_2} = \frac{\partial\left[(\beta\rho_1 b_1 + \gamma_1)\cdot\rho\frac{\sigma_1}{\sigma_2} + \gamma_2\right]}{\partial b_2} = 0$. Thus, the only condition to be satisfied is:

$$\beta\rho_1 < 1. \tag{A10}$$

With both β<1 and $\rho_1 < 1$ both solutions are E-stable. Next, we want to ensure that the Mean Squared Forecast Error (MSFE) of the agent living in RPE1 and using (A1) is lower than the MSFE of the agent using (A7), otherwise a small proportion of the latter could outperform the majority and lead to increasing deviations from the RPE1.

The forecast error and MSFE of the agents using (A1) are:

$$\begin{aligned} e_t^1 &= \bar{b}_1 w_t^1 + \gamma_2 w_t^2 + \eta_t - b_1 w_t^1, && \text{(A11)} \\ MSFE_1 &= E\left[e_t^1\right] = E\left[\left(\left(\bar{b}_1 - b_1\right) w_t^1 + \gamma_2 w_t^2 + \eta_t\right)^2\right]. && \text{(A12)} \end{aligned}$$

Similarly, for $MSFE_2$ we have the following expression

$$\begin{aligned} e_t^2 &= \bar{b}_1 w_t^1 + \gamma_2 w_t^2 + \eta_t - b_2 w_t^2, && \text{(A13)} \\ MSFE_2 &= E\left[e_t^2\right] = E\left[\left(\bar{b}_1 w_t^1 + \left(\gamma_2 - b_2\right) w_t^2 + \eta_t\right)^2\right]. && \text{(A14)} \end{aligned}$$

We are looking for the conditions under which $MSFE_1 < MSFE_2$:

$$\begin{aligned}
&: E\left[\left(\left(\bar{b}_1 - b_1\right) w_t^1 + \gamma_2 w_t^2 + \eta_t\right)^2\right] < E\left[\left(\bar{b}_1 w_t^1 + \left(\gamma_2 - b_2\right) w_t^2 + \eta_t\right)^2\right], \\
&: \left\{ \begin{array}{c} \left(\beta\rho_1 b_1 + \gamma_1 - b_1\right)^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2 + \\ 2\gamma_2 \left(\beta\rho_1 b_1 + \gamma_1 - b_1\right) \cdot \rho\sigma_1\sigma_2 \end{array} \right\} < \left\{ \begin{array}{c} \left(\beta\rho_1 b_1 + \gamma_1\right)^2 \sigma_1^2 + \left(\gamma_2 - b_2\right)^2 \sigma_2^2 + \\ 2\left(\beta\rho_1 b_1 + \gamma_1\right)\left(\gamma_2 - b_2\right) \cdot \rho\sigma_1\sigma_2 \end{array} \right\}, \\
&: \left\{ \begin{array}{c} \left(\beta\rho_1 b_1 + \gamma_1\right)^2 \sigma_1^2 + b_1^2 \sigma_1^2 \\ -2b_1 \left(\beta\rho_1 b_1 + \gamma_1\right) \sigma_1^2 + \gamma_2^2 \sigma_2^2 \\ +2\gamma_2 \left(\beta\rho_1 b_1 + \gamma_1\right) \cdot \rho\sigma_1\sigma_2 - \\ -2\gamma_2 b_1 \cdot \rho\sigma_1\sigma_2 \end{array} \right\} < \left\{ \begin{array}{c} \left(\beta\rho_1 b_1 + \gamma_1\right)^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2 \\ -2\gamma_2 b_2 \sigma_2^2 + b_2^2 \sigma_2^2 + \\ +2\gamma_2 \left(\beta\rho_1 b_1 + \gamma_1\right) \cdot \rho\sigma_1\sigma_2 - \\ -2b_2 \left(\beta\rho_1 b_1 + \gamma_1\right) \cdot \rho\sigma_1\sigma_2 \end{array} \right\}, \\
&: \left\{ \begin{array}{c} b_1^2 \sigma_1^2 - 2b_1 \left(\beta\rho_1 b_1 + \gamma_1\right) \sigma_1^2 - \\ -2\gamma_2 b_1 \cdot \rho\sigma_1\sigma_2 \end{array} \right\} < \left\{ \begin{array}{c} -2\gamma_2 b_2 \sigma_2^2 + b_2^2 \sigma_2^2 - \\ -2b_2 \left(\beta\rho_1 b_1 + \gamma_1\right) \cdot \rho\sigma_1\sigma_2 \end{array} \right\}, \\
&: \left\{ \begin{array}{c} b_1^2 \sigma_1^2 - \\ -2b_1 \sigma_1^2 \underbrace{\left(\beta\rho_1 b_1 + \gamma_1 + \gamma_2 \rho \frac{\sigma_2}{\sigma_1}\right)}_{=b_1} \end{array} \right\} < \left\{ \begin{array}{c} b_2^2 \sigma_2^2 - \\ -2b_2 \sigma_2^2 \underbrace{\left(\gamma_2 + \left(\beta\rho_1 b_1 + \gamma_1\right) \cdot \rho \frac{\sigma_1}{\sigma_2}\right)}_{=b_2} \end{array} \right\}, \\
&: -b_1^2 \sigma_1^2 < -b_2^2 \sigma_2^2 \Rightarrow b_1^2 \sigma_1^2 > b_2^2 \sigma_2^2. \qquad \text{(A15)}
\end{aligned}$$

The condition $b_1^2 \sigma_1^2 > b_2^2 \sigma_2^2$ has a very simple interpretation: in order to have MSFE of PLM1 lower than of PLM2, the share of variance of y_t explained by the PLM1 must be higher than that of PLM2. Alternatively, the R^2 of the regression (A1) must be higher than the R^2 of regression (A7). □

A.1 Analyzing conditions for $MSFE_1 < MSFE_2$

$$
\begin{aligned}
b_1^2\sigma_1^2 &> b_2^2\sigma_2^2 \Leftrightarrow b_1^2 > \left[(\beta\rho_1 b_1+\gamma_1)\cdot\rho\frac{\sigma_1}{\sigma_2}+\gamma_2\right]^2\frac{\sigma_2^2}{\sigma_1^2},\\
b_1^2 &> \left[(\beta\rho_1 b_1+\gamma_1)\cdot\rho+\gamma_2\frac{\sigma_2}{\sigma_1}\right]^2 = \left[\bar{b}_1\cdot\rho+\gamma_2\frac{\sigma_2}{\sigma_1}\right]^2,\\
&: \left[\bar{b}_1+\gamma_2\cdot\rho\frac{\sigma_2}{\sigma_1}\right]^2 > \left[\bar{b}_1\cdot\rho+\gamma_2\frac{\sigma_2}{\sigma_1}\right]^2,\\
&: \bar{b}_1^2+2\bar{b}_1\gamma_2\rho\frac{\sigma_2}{\sigma_1}+\left(\gamma_2\frac{\sigma_2}{\sigma_1}\right)^2\rho^2 > \bar{b}_1^2\rho^2+2\bar{b}_1\rho\gamma_2\frac{\sigma_2}{\sigma_1}+\left(\gamma_2\frac{\sigma_2}{\sigma_1}\right)^2,\\
&: \bar{b}_1^2\left(1-\rho^2\right) > \left(\gamma_2\frac{\sigma_2}{\sigma_1}\right)^2\left(1-\rho^2\right) \Rightarrow \bar{b}_1^2 > \left(\gamma_2\frac{\sigma_2}{\sigma_1}\right)^2,\\
&: \left|\bar{b}_1\right| > \gamma_2\frac{\sigma_2}{\sigma_1}.
\end{aligned}
\tag{A16}
$$

Theoretically, there could be 2 separate cases:

$$
CASE\ I \quad : \quad \bar{b}_1 > \gamma_2\frac{\sigma_2}{\sigma_1}, \tag{A17}
$$

$$
CASE\ II \quad : \quad \bar{b}_1 < -\gamma_2\frac{\sigma_2}{\sigma_1}. \tag{A18}
$$

As the CASE I is the most obvious, and will happen the easiest (assuming $\gamma_{1,2} > 0$ which is what we impose; otherwise, just re-define variable w_t^i so that $\gamma_{1,2}$ become positive), we start with this case.

A.1.1 CASE I: $\bar{b}_1 > \gamma_2\frac{\sigma_2}{\sigma_1}$

Coming back to the condition of $MSFE_1 < MSFE_2$, we get

$$
\begin{aligned}
\bar{b}_1 &> \gamma_2\frac{\sigma_2}{\sigma_1},\\
\frac{\gamma_1+\gamma_2\cdot\rho\frac{\sigma_2}{\sigma_1}\cdot\beta\rho_1}{1-\beta\rho_1} &> \gamma_2\frac{\sigma_2}{\sigma_1},\\
\gamma_1+\gamma_2\cdot\rho\frac{\sigma_2}{\sigma_1}\cdot\beta\rho_1 &> \gamma_2\frac{\sigma_2}{\sigma_1}-\gamma_2\frac{\sigma_2}{\sigma_1}\cdot\beta\rho_1,\\
\gamma_1+\gamma_2\frac{\sigma_2}{\sigma_1}\beta\rho_1(1+\rho) &> \gamma_2\frac{\sigma_2}{\sigma_1}\\
\frac{\gamma_1}{\gamma_2}\frac{\sigma_1}{\sigma_2}+\beta\rho_1\rho+\beta\rho_1-1 &> 0.
\end{aligned}
\tag{A19}
$$

Denoting the ratio of coefficients at the observable shocks $\frac{\gamma_1}{\gamma_2}$ as Γ, and the ratio of standard deviations $\frac{\sigma_1}{\sigma_2}$ as Σ, we see that the condition for CASE I to be true is $\Gamma\Sigma+\rho\cdot\beta\rho_1 > 1-\beta\rho_1$, or

$$
\rho > \frac{1-\beta\rho_1-\Gamma\Sigma}{\beta\rho_1}. \tag{A20}
$$

This condition is satisfied when $\rho_1 \to 1$ and $\Gamma\Sigma$ is large. Alternatively, when $\rho_1 \sim 0$ and $\Gamma\Sigma$ is small so that the numerator is positive, this condition might amount to $\rho > 1$ and thus be impossible to satisfy.

A.1.2 CASE II: $\bar{b}_1 < -\gamma_2 \frac{\sigma_2}{\sigma_1}$

In this case, we have

$$\begin{aligned}
\bar{b}_1 &< -\gamma_2 \frac{\sigma_2}{\sigma_1}, \\
\frac{\gamma_1 + \gamma_2 \cdot \rho \frac{\sigma_2}{\sigma_1} \cdot \beta\rho_1}{1-\beta\rho_1} &< -\gamma_2 \frac{\sigma_2}{\sigma_1}, \\
\gamma_1 + \gamma_2 \cdot \rho \frac{\sigma_2}{\sigma_1} \cdot \beta\rho_1 &< -\gamma_2 \frac{\sigma_2}{\sigma_1} + \gamma_2 \frac{\sigma_2}{\sigma_1} \cdot \beta\rho_1, \\
\frac{\gamma_1}{\gamma_2}\frac{\sigma_1}{\sigma_2} + \beta\rho_1\rho &< -1 + \beta\rho_1. \qquad \text{(A21)}
\end{aligned}$$

Using the notation just introduced, the condition $\bar{b}_1 < -\gamma_2 \frac{\sigma_2}{\sigma_1}$ amounts to $\Gamma\Sigma + \rho \cdot \beta\rho_1 + (1-\beta\rho_1) < 0$. Consider an intersection of with horizontal axis where $\Gamma\Sigma = 0$. Then, $\rho < -\frac{(1-\beta\rho_1)}{\beta\rho_1}$. As $|\rho| < 1$, the area where the solution exists is $-\frac{(1-\beta\rho_1)}{\beta\rho_1} > -1$, meaning that $\beta\rho_1 > 1/2$. That is, for MSFE condition satisfied for $\bar{b}_1 < 0$, $\beta\rho_1$ must be larger than $1/2$.

To sum up, combining to cases together, we see that we need:

$$|\Gamma\Sigma + \rho \cdot \beta\rho_1| > 1 - \beta\rho_1. \qquad \text{(A22)}$$

A.2 Deriving the Sparse Weights

Sparse weights are derived by minimizing the following expression:

$$\min_{m\in[0,1]^n} \frac{1}{2} \sum_{i,j=1\ldots n} (1-m_i)\Lambda_{ij}\left(1-m_j\right) + \kappa \sum_{i,j=1\ldots n} m_i \qquad \text{(A23)}$$

with:

$$\begin{aligned}
\Lambda_{ij} &= -\sigma_{ij} a_{w_i} u_{aa} a_{w_j}, \\
a_{w_i} &= -u_{aa}^{-1} u_{aw_i}, \\
u_{aa} &= \frac{\partial^2 u_a}{\partial a^2} = \frac{\partial}{\partial a}\left(-(a - \bar{b}_1 w_t^1 - \bar{b}_2 w_t^2)\right) = -1, \\
u_{aw_i} &= \bar{b}_i.
\end{aligned}$$

Then the cost of inattention is

$$\Lambda_{ij} = \sigma_{ij}\bar{b}_i\bar{b}_j. \qquad \text{(A24)}$$

Plugging the cost of inattention as in (A24) into (A23) we get the following problem:

$$\begin{aligned}
&\min_{m\in[0,1]^n} \frac{1}{2}\left\{(1-m_1)^2 \sigma_1^2 \bar{b}_1^2 + 2(1-m_1)(1-m_2)\sigma_{12}\bar{b}_1\bar{b}_2 + (1-m_2)^2 \sigma_2^2 \bar{b}_2^2\right\} + \\
&+\kappa(|m_1| + |m_2|). \qquad \text{(A25)}
\end{aligned}$$

s.t.

$$\begin{aligned}
m_i &\leq 1, \qquad \text{(A26)} \\
m_i &\geq 0, \qquad \text{(A27)} \\
i &= 1,2. \qquad \text{(A28)}
\end{aligned}$$

There are nine cases depending on which restriction is binding. Let us as start with the simpliest case, with the inner solution for both weights: $0 < m_i < 1$.

1. First order conditions of (A25) with respect to m_1 and m_2:

$$[m_1] \quad : \quad \kappa + \frac{1}{2}(-2\bar{b}_1^2\sigma_1^2(1-m_1) - 2\bar{b}_1\bar{b}_2\,(1-m_2)\,\sigma_{12} = 0, \tag{A29}$$

$$[m_2] \quad : \quad \kappa + \frac{1}{2}(-2\bar{b}_2^2\sigma_2^2(1-m_2) - 2\bar{b}_1\bar{b}_2\,(1-m_1)\,\sigma_{12} = 0. \tag{A30}$$

Solving (A29) and (A30) for m_1 and m_2 gives the expressions (13) and (14) in the text.

2. Consider the second case where $0 < m_2 < 1$, but $m_1 = 0$. The first order conditions are then modified as:

$$[m_1] \quad : \quad \kappa + \frac{1}{2}(-2\bar{b}_1^2\sigma_1^2 - 2\bar{b}_1\bar{b}_2\,(1-m_2)\,\sigma_{12}) \geq 0, \tag{A31}$$

$$[m_2] \quad : \quad \kappa + \frac{1}{2}(-2\bar{b}_2^2(1-m_2)\sigma_2^2 - 2\bar{b}_1\bar{b}_2\sigma_{12}) = 0, \tag{A32}$$

Resulting in the following expressions:

$$[m_1] \quad : \quad m_2 \geq \frac{\bar{b}_1^2\sigma_1^2 - \kappa}{\bar{b}_1\bar{b}_2\sigma_{12}} + 1, \tag{A33}$$

$$[m_2] \quad : \quad m_2 = \frac{\bar{b}_1\bar{b}_2\sigma_{12} - \kappa}{\bar{b}_2^2\sigma_2^2} + 1. \tag{A34}$$

3. Consider the third case where $0 < m_1 < 1$, but $m_2 = 0$. The first order conditions are then modified as:

$$[m_1] \quad : \quad \kappa + \frac{1}{2}(-2\bar{b}_1^2\sigma_1^2 + 2m_1\bar{b}_1^2\sigma_1^2 - 2\bar{b}_1\bar{b}_2\sigma_{12}) = 0, \tag{A35}$$

$$[m_2] \quad : \quad \kappa + \frac{1}{2}(-2\bar{b}_2^2\sigma_2^2 - 2\bar{b}_1\bar{b}_2\sigma_{12}(1-m_1)) \geq 0, \tag{A36}$$

Resulting in the following expressions:

$$[m_1] \quad : \quad m_1 = \frac{\bar{b}_1\bar{b}_2\sigma_{12} - \kappa}{\bar{b}_1^2\sigma_1^2} + 1, \tag{A37}$$

$$[m_2] \quad : \quad m_1 \geq \frac{\bar{b}_2^2\sigma_2^2 - \kappa}{\bar{b}_1\bar{b}_2\sigma_{12}} + 1. \tag{A38}$$

Combining these two expressions yields:

$$\frac{\bar{b}_1\bar{b}_2\sigma_{12} - \kappa}{\bar{b}_1^2\sigma_1^2} \geq \frac{\bar{b}_2^2\sigma_2^2 - \kappa}{\bar{b}_1\bar{b}_2\sigma_{12}}, \tag{A39}$$

$$\begin{cases} \bar{b}_1^2\bar{b}_2^2\sigma_{12}^2 - \kappa\bar{b}_1\bar{b}_2\sigma_{12} \geq \bar{b}_2^2\sigma_2^2\bar{b}_1^2\sigma_1^2 - \kappa\bar{b}_1^2\sigma_1^2 & \text{if} \;\; \bar{b}_1\bar{b}_2\sigma_{12} > 0 \\ \bar{b}_1^2\bar{b}_2^2\sigma_{12}^2 - \kappa\bar{b}_1\bar{b}_2\sigma_{12} \leq \bar{b}_2^2\sigma_2^2\bar{b}_1^2\sigma_1^2 - \kappa\bar{b}_1^2\sigma_1^2 & \text{if} \;\; \bar{b}_1\bar{b}_2\sigma_{12} < 0 \end{cases} \tag{A40}$$

$$\begin{cases} \bar{b}_1^2\bar{b}_2^2\sigma_{12}^2 - \bar{b}_2^2\sigma_2^2\bar{b}_1^2\sigma_1^2 \geq \kappa(\bar{b}_1\bar{b}_2\sigma_{12} - \bar{b}_1^2\sigma_1^2) & \text{if} \;\; \bar{b}_1\bar{b}_2\sigma_{12} > 0 \\ \bar{b}_1^2\bar{b}_2^2\sigma_{12}^2 - \bar{b}_2^2\sigma_2^2\bar{b}_1^2\sigma_1^2 \leq \kappa(\bar{b}_1\bar{b}_2\sigma_{12} - \bar{b}_1^2\sigma_1^2) & \text{if} \;\; \bar{b}_1\bar{b}_2\sigma_{12} < 0, \end{cases} \tag{A41}$$

As $\sigma_{12}^2 < \sigma_2^2\sigma_1^2$, the first inequlity implies $\bar{b}_1\bar{b}_2\sigma_{12} < \bar{b}_1^2\sigma_1^2$, while the second: $\kappa \leq \frac{\bar{b}_1^2\bar{b}_2^2\sigma_{12}^2 - \bar{b}_2^2\sigma_2^2\bar{b}_1^2\sigma_1^2}{(\bar{b}_1\bar{b}_2\sigma_{12} - \bar{b}_1^2\sigma_1^2)}$
From the condition $0 < m_1 < 1$ follows that attention costs must satisfy:

$$: \quad \kappa < \bar{b}_1^2\sigma_1^2 + \bar{b}_1\bar{b}_2\sigma_{12}, \tag{A42}$$
$$: \quad \kappa > \bar{b}_2^2\sigma_2^2. \tag{A43}$$

4. Consider the third case where both weights are zero. The first order conditions are then modified as:

$$[m_1] \quad : \quad \kappa + \frac{1}{2}(-2\bar{b}_1^2\sigma_1^2 - 2\bar{b}_1\bar{b}_2\sigma_{12}) \geq 0, \tag{A44}$$
$$[m_2] \quad : \quad \kappa + \frac{1}{2}(-2\bar{b}_2^2\sigma_2^2 - 2\bar{b}_1\bar{b}_2\sigma_{12}) \geq 0, \tag{A45}$$

which give the condition on attention costs, κ:

$$\kappa \geq \bar{b}_2^2\sigma_2^2 + \bar{b}_1\bar{b}_2\sigma_{12}, \tag{A46}$$
$$\kappa \geq \bar{b}_1^2\sigma_1^2 + \bar{b}_1\bar{b}_2\sigma_{12}. \tag{A47}$$

5. $m_1 = 1$ and $m_2 = 0$, with λ_1 as a langrange multipliers associated with (A26) for m_1:

$$[m_1] \quad : \quad \kappa - \bar{b}_1\bar{b}_2\sigma_{12} + \lambda_1 = 0, \tag{A48}$$
$$[m_2] \quad : \quad \kappa - \bar{b}_2^2\sigma_2^2 \geq 0, \tag{A49}$$

resulting in condition for κ: $\kappa \geq \bar{b}_2^2\sigma_2^2$ and $\kappa \leq \bar{b}_1\bar{b}_2\sigma_{12}$ (otherwise $\lambda_1 < 0$ which would violate the optimality conditions)

6. Similarly for $m_1 = 0$ and $m_2 = 1$, with λ_2 as a langrange multipliers associated with (A26) for m_2:

$$[m_1] \quad : \quad \kappa - \bar{b}_1^2\sigma_1^2 \geq 0, \tag{A50}$$
$$[m_2] \quad : \quad \kappa - \bar{b}_1\bar{b}_2\sigma_{12} + \lambda_2 = 0, \tag{A51}$$

resulting in condition for κ: $\kappa \geq \bar{b}_1^2\sigma_1^2$ and $\kappa \leq \bar{b}_1\bar{b}_2\sigma_{12}$ (otherwise $\lambda_2 < 0$ which would violate the optimality conditions)

7. For both weights equal to unity, $m_1 = 1$ and $m_2 = 1$

$$[m_1] \quad : \quad \kappa + \lambda_1 = 0, \tag{A52}$$
$$[m_2] \quad : \quad \kappa + \lambda_2 = 0. \tag{A53}$$

Resulting in $\lambda_2 = \lambda_1 = \kappa = 0$.

8. For $m_1 = 1$ and $0 < m_2 < 1$:

$$[m_1] \quad : \quad \kappa - \bar{b}_1\bar{b}_2(1 - m_2)\sigma_{12} + \lambda_1 = 0, \tag{A54}$$
$$[m_2] \quad : \quad \kappa - \bar{b}_2^2(1 - m_2)\sigma_2^2 = 0, \tag{A55}$$

resulting in:

$$m_2 = 1 - \frac{\kappa}{\bar{b}_2^2 \sigma_2^2}, \tag{A56}$$

$$\kappa < \bar{b}_2^2 \sigma_2^2, \tag{A57}$$

$$\bar{b}_1 \bar{b}_2 \sigma_{12} \geq \bar{b}_2^2 \sigma_2^2, \tag{A58}$$

where the last inequility comes from expression for $\lambda_1 = \kappa(\frac{\bar{b}_1 \bar{b}_2 \sigma_{12}}{\bar{b}_2^2 \sigma_2^2} - 1)$ and the condition $\lambda_1 \geq 0$.

9. For $m_2 = 1$ and $0 < m_1 < 1$:

$$[m_1] \quad : \quad \kappa - \bar{b}_1^2 (1 - m_1) \sigma_1^2 = 0, \tag{A59}$$

$$[m_2] \quad : \quad \kappa - \bar{b}_1 \bar{b}_2 (1 - m_1) \sigma_{12} + \lambda_2 = 0, \tag{A60}$$

resulting in:

$$m_1 = 1 - \frac{\kappa}{\bar{b}_1^2 \sigma_1^2}, \tag{A61}$$

$$\kappa < \bar{b}_1^2 \sigma_1^2. \tag{A62}$$

Now define the value function as V_j with j being the above solution case and compare which of the solutions results in the minimum.

$$[0 < m_1 < 1, 0 < m_2 < 1] \qquad V_1 = 2\kappa - \frac{\kappa^2 (\sigma_1^2 \bar{b}_1^2 - 2\sigma_{12} \bar{b}_2 \bar{b}_1 + \sigma_2^2 \bar{b}_2^2)}{2(\sigma_2^2 \bar{b}_2^2 \sigma_1^2 \bar{b}_1^2 - \sigma_{12}^2 \bar{b}_1^2 \bar{b}_2^2)}, \tag{A63}$$

$$[m_1 = 0, 0 < m_2 < 1] \qquad V_2 = \frac{1}{2} \sigma_1^2 \bar{b}_1^2 + \kappa - \frac{(\sigma_{12} \bar{b}_1 \bar{b}_2 - \kappa)^2}{2\sigma_2^2 \bar{b}_2^2}, \tag{A64}$$

$$[0 < m_1 < 1, m_2 = 0] \qquad V_3 = \frac{1}{2} \sigma_2^2 \bar{b}_2^2 + \kappa - \frac{(\sigma_{12} \bar{b}_1 \bar{b}_2 - \kappa)^2}{2\sigma_1^2 \bar{b}_1^2}, \tag{A65}$$

$$[m_1 = 0, m_2 = 0] \qquad V_4 = \frac{1}{2} (\sigma_2^2 \bar{b}_2^2 + \sigma_1^2 \bar{b}_1^2 + 2\sigma_{12} \bar{b}_1 \bar{b}_2) \tag{A66}$$

$$[m_1 = 1, m_2 = 0] \qquad V_5 = \frac{1}{2} \sigma_2^2 \bar{b}_2^2 + \kappa, \tag{A67}$$

$$[m_1 = 0, m_2 = 1] \qquad V_6 = \frac{1}{2} \sigma_1^2 \bar{b}_1^2 + \kappa, \tag{A68}$$

$$[m_1 = 1, m_2 = 1] \qquad V_7 = 2\kappa, \tag{A69}$$

$$[m_1 = 1, 0 < m_2 < 1] \qquad V_8 = 2\kappa - \frac{\kappa^2}{2\sigma_2^2 \bar{b}_2^2}, \tag{A70}$$

$$[0 < m_1 < 1, m_2 = 1] \qquad V_9 = 2\kappa - \frac{\kappa^2}{2\sigma_1^2 \bar{b}_1^2} \tag{A71}$$

$$\tag{A72}$$

Now, compare the value functions.

$$V_8 < V_9 \quad : \quad 2\kappa - \frac{\kappa^2}{2\sigma_2^2 \bar{b}_2^2} < 2\kappa - \frac{\kappa^2}{2\sigma_1^2 \bar{b}_1^2}, \tag{A73}$$

$$\Rightarrow \quad \sigma_2^2 \bar{b}_2^2 < \sigma_1^2 \bar{b}_1^2. \tag{A74}$$

$$V_1 < V_3 \quad : \quad 2\kappa - \frac{\kappa^2(\sigma_1^2\bar{b}_1^2 - 2\rho\sigma_2\bar{b}_2\sigma_1\bar{b}_1 + \sigma_2^2\bar{b}_2^2)}{2\sigma_2^2\bar{b}_2^2\sigma_1^2\bar{b}_1^2(1-\rho^2)} < \frac{1}{2}\sigma_2^2\bar{b}_2^2 + \kappa - \frac{(\sigma_{12}\bar{b}_1\bar{b}_2 - \kappa)^2}{2\sigma_1^2\bar{b}_1^2}, \tag{A75}$$

$$: \quad \kappa(1 - \frac{\rho\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1}) - \frac{\kappa^2(1 - 2\frac{\rho\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1} + \rho^2\frac{\sigma_2^2\bar{b}_2^2}{\sigma_1^2\bar{b}_1^2})}{2\sigma_2^2\bar{b}_2^2(1-\rho^2)} < \frac{1}{2}\sigma_2^2\bar{b}_2^2(1-\rho^2), \tag{A76}$$

$$: \quad 2\kappa(1-\rho^2)(1 - \frac{\rho\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1})\sigma_2^2\bar{b}_2^2 - \kappa^2(1 - \frac{\rho\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1})^2 - \sigma_2^4\bar{b}_2^4, < 0, \tag{A77}$$

$$: \quad -(\kappa(1 - \frac{\rho\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1}) - \sigma_2\bar{b}_2)^2 < 0. \tag{A78}$$

$$V_1 < V_5 \quad : \quad 2\kappa - \frac{\kappa^2(\sigma_1^2\bar{b}_1^2 - 2\rho\sigma_2\bar{b}_2\sigma_1\bar{b}_1 + \sigma_2^2\bar{b}_2^2)}{2\sigma_2^2\bar{b}_2^2\sigma_1^2\bar{b}_1^2(1-\rho^2)} < \frac{1}{2}\sigma_2^2\bar{b}_2^2 + \kappa, \tag{A79}$$

$$: \quad 2\kappa\sigma_2^2\bar{b}_2^2 - \frac{\kappa^2(1 - 2\rho\sigma_2\bar{b}_2\sigma_1\bar{b}_1 + \frac{\sigma_2^2\bar{b}_2^2}{\sigma_1^2\bar{b}_1^2})}{(1-\rho^2)} < \sigma_4^2\bar{b}_4^2, \tag{A80}$$

$$: \quad -(\kappa - \sigma_2^2\bar{b}_2^2)^2 - \frac{\kappa^2(1 - 2\rho\frac{\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1} + \frac{\sigma_2^2\bar{b}_2^2}{\sigma_1^2\bar{b}_1^2})}{(1-\rho^2)} + \kappa^2 < 0, \tag{A81}$$

$$: \quad -(\kappa - \sigma_2^2\bar{b}_2^2)^2 - \frac{\kappa^2(-2\rho\frac{\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1} + \frac{\sigma_2^2\bar{b}_2^2}{\sigma_1^2\bar{b}_1^2} + \rho^2)}{(1-\rho^2)} < 0, \tag{A82}$$

$$: \quad -(\kappa - \sigma_2^2\bar{b}_2^2)^2 - \frac{\kappa^2(\rho - \frac{\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1})^2}{(1-\rho^2)} < 0, \tag{A83}$$

$$\tag{A84}$$

$$V_1 < V_8 \quad : \quad 2\kappa - \frac{\kappa^2(\sigma_1^2\bar{b}_1^2 - 2\rho\sigma_2\bar{b}_2\sigma_1\bar{b}_1 + \sigma_2^2\bar{b}_2^2)}{2\sigma_2^2\bar{b}_2^2\sigma_1^2\bar{b}_1^2(1-\rho^2)} < 2\kappa - \frac{\kappa^2}{2\sigma_2^2\bar{b}_2^2}, \tag{A85}$$

$$: \quad -\frac{\kappa^2(\rho^2 - 2\rho\frac{\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1} + \frac{\sigma_2^2\bar{b}_2^2}{\sigma_1^2\bar{b}_1^2})}{2\sigma_2^2\bar{b}_2^2(1-\rho^2)} < 0, \tag{A86}$$

$$: \quad -\frac{\kappa^2(\rho - \frac{\sigma_2\bar{b}_2}{\sigma_1\bar{b}_1})^2}{2\sigma_2^2\bar{b}_2^2(1-\rho^2)} < 0, \tag{A87}$$

$$V_3 < V_2 \quad : \quad \frac{1}{2}\sigma_2^2\bar{b}_2^2 + \kappa - \frac{(\sigma_{12}\bar{b}_1\bar{b}_2 - \kappa)^2}{2\sigma_1^2\bar{b}_1^2} < \frac{1}{2}\sigma_1^2\bar{b}_1^2 + \kappa - \frac{(\sigma_{12}\bar{b}_1\bar{b}_2 - \kappa)^2}{2\sigma_2^2\bar{b}_2^2}, \tag{A88}$$

$$\Rightarrow \quad \sigma_2^2\bar{b}_2^2 < \sigma_1^2\bar{b}_1^2. \tag{A89}$$

$$V_5 < V_6 \quad : \quad \frac{1}{2}\sigma_2^2\bar{b}_2^2 + \kappa < \frac{1}{2}\sigma_1^2\bar{b}_1^2 + \kappa, \tag{A90}$$

$$: \quad \Rightarrow \sigma_2^2\bar{b}_2^2 < \sigma_1^2\bar{b}_1^2. \tag{A91}$$

Thus, if inner solution exists for $M_{\pi y}$, it outperforms corner solution for M_π: $V_1 < V_5 < V6$, $V_1 < V_3 < V2$, and corner solution for $M_{\pi y}$ with $w_1 = 1$: $V_1 < V_8 < V_9$.

Representing inattention costs as an ellipse. Note, that (A23) can be written as:

$$\frac{1}{2}(\bar{x}^2 + 2\rho\, bs\, \bar{x}\bar{y} + \bar{y}^2 bs^2) = c, \tag{A92}$$

where c is constant, $\bar{x} = 1 - m_1$, $\bar{y} = 1 - m_2$ and $bs = \frac{\sigma_2 \bar{b}_2}{\sigma_1 \bar{b}_1}$. Before finding ellipse focal points, we have to rotate axes to eliminate $\bar{x}\bar{y}$ term. For this we introduce new variables x, y such that $\bar{x} = x\,cos(\theta) - y\,sin(\theta)$ and $\bar{y} = x\,sin(\theta) + y\,cos(\theta)$, with $cot(2\theta) = \frac{1-bs^2}{2\rho\, bs}$. Then, plugging these variables back into the ellipse equation, $(\bar{x}^2 + 2\rho\, bs\, \bar{x}\bar{y} + \bar{y}^2 bs^2)$ and opening the brackets we get:

$$\begin{aligned}
&x^2 cos^2(\theta) - 2yx\,cos(\theta)sin(\theta) + y^2 sin(\theta)^2 + 2\rho\, bs\, x^2 cos(\theta)sin(\theta) + 2\rho\, bs\, xy\,cos^2(\theta)\\
&-2\rho\, bs\, yx\,sin^2(\theta) - 2\rho\, bs\, y^2 cos(\theta)sin(\theta) + bs^2 x^2 sin^2(\theta)\\
&+2bs^2\, xy\,sin(\theta)\,cos(\theta) + bs^2 y^2 cos^2(\theta).
\end{aligned} \tag{A93}$$

Now, we group xy terms:

$$\begin{aligned}
&xy\,(-2cos(\theta)sin(\theta) + 2\rho\, bs\, cos^2(\theta) - 2\rho\, bs\, sin^2(\theta) + 2bs^2 sin(\theta)cos(\theta)) =\\
&xy\,(-sin(2\theta) + 2\rho\, bs\,(cos^2(\theta) - sin^2(\theta)) + bs^2 sin(2\theta)) =\\
&xy\,(-sin(2\theta)2\rho\, bs\,\frac{(1-bs^2)}{2\rho\, bs} + 2\rho\, bs\, cos(2\theta) =\\
&xy\,(-sin(2\theta)2\rho\, bs\, cot(2\theta) + 2\rho\, bs\, cos(2\theta)) =\\
&xy\,(-sin(2\theta)2\rho\, bs\,\frac{cos(2\theta)}{sin(2\theta)} + 2\rho\, bs\, cos(2\theta)) =\\
&xy\,(-2\rho\, bs\, cos(2\theta) + 2\rho\, bs\, cos(2\theta)) = 0.
\end{aligned} \tag{A94}$$

Then, proceed to the rest of the terms:

$$
\begin{aligned}
&x^2(cos^2(\theta)+2\rho\, bs\, cos(\theta)sin(\theta)+bs^2 sin^2(\theta))\\
&+y^2(sin^2(\theta)-2\rho\, bs\, cos(\theta)sin(\theta)+bs^2cos^2(\theta))=\\
&x^2(1-sin^2(\theta)+bs^2sin^2(\theta)+\rho\, bs\, sin(2\theta))\\
&+y^2(1-cos^2(\theta)+bs^2cos^2(\theta)-\rho\, bs\, sin(2\theta))=\\
&x^2(1-sin^2(\theta)2\rho\, bs\frac{(1-bs^2)}{2\rho\, bs}+\rho\, bs\, sin(2\theta))\\
&+y^2(1-cos^2(\theta)2\rho\, bs\frac{(1-bs^2)}{2\rho\, bs}-\rho\, bs\, sin(2\theta))=\\
&x^2(1-sin^2(\theta)2\rho\, bs\frac{cos(2\theta)}{sin(2\theta)}+\rho\, bs\, sin(2\theta))\\
&+y^2(1-\rho\, bs\frac{cos(2\theta)(cos(2\theta)+1)}{sin(2\theta)}-\rho\, bs\, sin(2\theta))=\\
&x^2(1-\rho\, bs\frac{cos(2\theta)(1-cos(2\theta))}{sin(2\theta)}+\rho\, bs\, sin(2\theta))\\
&+y^2(1-\rho\, bs\frac{cos^2(2\theta)+cos(2\theta))}{sin(2\theta)}-\rho\, bs\, sin(2\theta))=\\
&x^2(1-\rho\, bs\frac{cos(2\theta)-cos^2(2\theta)-sin^2(2\theta))}{sin(2\theta)})\\
&+y^2(1-\frac{\rho\, bs\,(cos^2(2\theta)+sin^2(2\theta)+cos(2\theta))}{sin(2\theta)})=\\
&x^2(1+\rho\, bs\frac{1-cos(2\theta)}{sin(2\theta)})+y^2(1-\frac{\rho\, bs\,(1+cos(2\theta))}{sin(2\theta)})=\\
&x^2(1+\rho\, bs\, tg(\theta))+y^2(1-\rho\, bs\, cot(\theta)).
\end{aligned}
\tag{A95}
$$

Finally, the equation for ellipse in the traditional form is:

$$
x^2\frac{1}{\sqrt{\frac{2c}{(1+\rho\, bs\, tg(\theta))}}}+y^2\frac{1}{\sqrt{\frac{2c}{(1-\rho\, bs\, cot(\theta))}}}=1. \tag{A96}
$$

Proof. . Proposition 2.0.2. Consider when $m_1 > m_2$ for the inner solution:

$$
\begin{aligned}
-\frac{\kappa}{\bar{b}_1\bar{b}_2\sigma_1\sigma_2\left(1-\rho^2\right)}\frac{\bar{b}_2\sigma_2-\bar{b}_1\rho\sigma_1}{\bar{b}_1\sigma_1} &> -\frac{\kappa}{\bar{b}_1\bar{b}_2\sigma_1\sigma_2\left(1-\rho^2\right)}\frac{\bar{b}_1\sigma_1-\rho\bar{b}_2\sigma_2}{\bar{b}_2\sigma_2}\\
-\frac{1}{\bar{b}_1}\frac{\bar{b}_2\sigma_2-\bar{b}_1\rho\sigma_1}{\bar{b}_1\sigma_1} &> -\frac{1}{\bar{b}_1}\frac{\bar{b}_1\sigma_1-\rho\bar{b}_2\sigma_2}{\bar{b}_2\sigma_2}\\
\left(\bar{b}_2\sigma_2-\bar{b}_1\rho\sigma_1\right)\bar{b}_2\sigma_2 &< \bar{b}_1\sigma_1\left(\bar{b}_1\sigma_1-\rho\bar{b}_2\sigma_2\right)\\
\bar{b}_2^2\sigma_2^2-\bar{b}_1\bar{b}_2\rho\sigma_1\sigma_2 &< \bar{b}_1^2\sigma_1^2-\bar{b}_1\bar{b}_2\rho\sigma_1\sigma_2\\
\bar{b}_2^2\sigma_2^2 &< \bar{b}_1^2\sigma_1^2
\end{aligned}
\tag{A97}
$$

In the range of κ, where inner solution does not exist, solutions with $m_1 > m_2$ are optimal for $\bar{b}_2^2\sigma_2^2 < \bar{b}_1^2\sigma_1^2$, see (A73)-(A90). □

Proof. Proposition 2.0.3. Assume that the (A97) is satisfied. Then, we need

$$1-\kappa\frac{1-\rho\frac{\bar{b}_2}{\bar{b}_1}\frac{\sigma_2}{\sigma_1}}{\bar{b}_2^2\sigma_2^2\left(1-\rho^2\right)}<0.$$

We distinguish between the following cases depending on the sign of $1-\rho\frac{\bar{b}_2}{\bar{b}_1}\frac{\sigma_2}{\sigma_1}$.

CASE I: $\bar{b}_1>\frac{\bar{b}_2\sigma_2}{\sigma_1}$. Then $\frac{\bar{b}_2}{\bar{b}_1}\frac{\sigma_2}{\sigma_1}<1$ and $1-\rho\frac{\bar{b}_2}{\bar{b}_1}\frac{\sigma_2}{\sigma_1}>0$. **CASE II:** $\bar{b}_1<-\frac{\bar{b}_2\sigma_2}{\sigma_1}<0$. Then $1-\rho\frac{\bar{b}_2}{\bar{b}_1}\frac{\sigma_2}{\sigma_1}>0$. Note, that for RPE1 consistent rule be an equilibrium for $\bar{b}_1<0$, it must hold that $\beta\rho_1>1/2$ (see CASE II of Appendix A.1). In terms of parameter of the model (15) could be written as

$$\kappa<\gamma^2\sigma_2^2\left(1-\rho^2\right)\frac{(\Gamma\Sigma+\beta\rho\rho_1)}{\Gamma\Sigma+\rho\left(2\beta\rho_1-1\right)}>0.$$

The condition also states that if RPE1 consistent rule is an equilibrium, that is (10) is, satisfied, there is no region in the parameter space such that for any κ this equilibrium survives. □

Appendix B: Three-equation New Keynesian Model

B.1 Rational Expectations Equilibrium

Proof. Proposition 3.0.1

Deriving REE with forward looking policy rule. The solution of the model under forward-looking rule is:

$$y_t = \frac{1-\phi_\pi}{\sigma}E_t\pi_{t+1}+\frac{\phi_\pi-1}{\sigma}\bar{\pi}+y_{t-1}+g_t, \tag{B1}$$

$$\pi_t = \omega\frac{\phi_\pi-1}{\sigma}\bar{\pi}+(\beta+\omega\frac{1-\phi_\pi}{\sigma})E_t\pi_{t+1}+\omega y_{t-1}+\omega g_t+u_t. \tag{B2}$$

We re-define in (B1) and (B2) as:

$$a_y = \frac{\phi_\pi-1}{\sigma}\bar{\pi}, \tag{B3}$$

$$b_y = \frac{1-\phi_\pi}{\sigma}, \tag{B4}$$

$$c_y = 1, \tag{B5}$$

$$a_\pi = \omega\frac{\phi_\pi-1}{\sigma}\bar{\pi}, \tag{B6}$$

$$b_\pi = \beta+\omega\frac{1-\phi_\pi}{\sigma}, \tag{B7}$$

$$c_\pi = \omega. \tag{B8}$$

There are two rational expectations equilibria. One solution is associated with the following forecasting rule:

$$\pi_t^{REE} = \alpha+cy_{t-1}+b\pi_{t-1}, \tag{B9}$$

$$E_t\pi_{t+1}^{REE} = \alpha+cE_ty_t+b\pi_t=\alpha+b\alpha+c(1+b)y_{t-1}+b^2\pi_{t-1}, \tag{B10}$$

where in the last equality we make use of the assumption that our agents have naive expectations about future output gap. Plugging the forecasting rule in (B2) and in (B1) yields:

$$\pi_t = a_\pi + b_\pi \alpha(1+b) + b^2 b_\pi \pi_{t-1} + (b_\pi c(1+b) + c_\pi) y_{t-1} + c_\pi g_t + u_t, \tag{B11}$$
$$y_t = a_y + b_y \alpha(1+b) + b_y b^2 \pi_{t-1} + (b_y c(1+b) + c_y) y_{t-1} + g_t. \tag{B12}$$

Using the method of undetermined coefficients, we solve for the forecasting rule coefficients.

$$\alpha = a_\pi + b_\pi \alpha(1+b), \tag{B13}$$
$$b = b^2 b_\pi, \tag{B14}$$
$$c = b_\pi c(1+b) + c_\pi. \tag{B15}$$

The rational expectation solution is:

$$\alpha = -\frac{a_\pi}{b_\pi}, \tag{B16}$$
$$b = \frac{1}{b_\pi}, \tag{B17}$$
$$c = -\frac{c_\pi}{b_\pi}, \tag{B18}$$

The ALM coefficients are then:

$$\bar{a}_y^{REE} = a_y - b_y a_\pi \frac{1+b_\pi}{b_\pi^2}, \tag{B19}$$
$$\bar{b}_y^{REE} = \frac{b_y}{b_\pi^2}, \tag{B20}$$
$$\bar{c}_y^{REE} = c_y - b_y c_\pi \frac{1+b_\pi}{b_\pi^2}, \tag{B21}$$
$$\begin{aligned}\bar{a}_\pi^{REE} &= a_\pi - b_\pi \frac{a_\pi}{b_\pi}\frac{(1+b_\pi)}{b_\pi} \\ &= -\frac{a_\pi}{b_\pi},\end{aligned} \tag{B22}$$
$$\bar{b}_\pi^{REE} = \frac{1}{b_\pi}, \tag{B23}$$
$$\begin{aligned}\bar{c}_\pi^{REE} &= c_\pi - b_\pi \frac{c_\pi}{b_\pi}\frac{1+b_\pi}{b_\pi} \\ &= -\frac{c_\pi}{b_\pi}.\end{aligned} \tag{B24}$$

E-stability. For the REE equilibrium to be E-stable the eigenvalues of the following matrix must be smaller than unity:

$$\begin{pmatrix} (1+b)b_\pi & b_\pi \alpha & 0 \\ 0 & 2bb_\pi & 0 \\ 0 & cb_\pi & b_\pi(1+b). \end{pmatrix}$$

With the eigenvalues equal to $(2,\ 1+b_\pi,\ 1+b_\pi)$ the solution is not E-stable.

Another REE solution is the MSV solution of the form:

$$\pi_t^{MSV} = \alpha + c y_{t-1}, \tag{B25}$$
$$E_t \pi_{t+1}^{MSV} = \alpha + c E_t y_t = \alpha + c y_{t-1}, \tag{B26}$$

where in the last equality we make use of the assumption that our agents have naive expectations about future output gap. Plugging the forecasting rule in (B2) and in (B1) yields:

$$\pi_t = a_\pi + b_\pi \alpha + (b_\pi c + c_\pi) y_{t-1} + c_\pi g_t + u_t, \tag{B27}$$
$$y_t = a_y + b_y \alpha + (b_y c + c_y) y_{t-1} + g_t. \tag{B28}$$

With the coefficients defined as in (B1) and (B2). Using the method of undetermined coefficients, we solve for the forecasting rule coefficients.

$$\alpha = a_\pi + b_\pi \alpha, \tag{B29}$$
$$c = b_\pi c + c_\pi. \tag{B30}$$

The MSV solution is:

$$\alpha = \frac{a_\pi}{1 - b_\pi}, \tag{B31}$$
$$c = \frac{c_\pi}{1 - b_\pi}, \tag{B32}$$

The ALM coefficients are then:

$$\begin{aligned} \bar{a}_y^{MSV} &= a_y + b_y \alpha \\ &= a_y + b_y \frac{a_\pi}{1 - b_\pi}, \end{aligned} \tag{B33}$$
$$\bar{b}_y^{MSV} = 0, \tag{B34}$$
$$\bar{c}_y^{MSV} = (b_y \frac{c_\pi}{1 - b_\pi} + c_y), \tag{B35}$$
$$\begin{aligned} \bar{a}_\pi^{MSV} &= a_\pi + b_\pi \alpha \\ &= \frac{a_\pi}{1 - b_\pi}, \end{aligned} \tag{B36}$$
$$\bar{b}_\pi^{MSV} = 0, \tag{B37}$$
$$\begin{aligned} \bar{c}_\pi^{MSV} &= b_\pi \frac{c_\pi}{1 - b_\pi} + c_\pi \\ &= \frac{c_\pi}{1 - b_\pi}. \end{aligned} \tag{B38}$$

Determinacy. Representing the ALM as $z_t = A + B z_{t-1} + U_t$, with $z_t = (y_t, \pi_t)'$, A and U_t are vectors of constants and shocks respectively, and

$$B = \begin{pmatrix} \bar{c}_y^{MSV} & \bar{b}_y^{MSV} \\ \bar{c}_\pi^{MSV} & \bar{b}_\pi^{MSV} \end{pmatrix}$$

For the solution to be determinate, both eigenvalues of B.1 must be inside the unit circle. The characteristic polynomial of B.1 is given by $p(\lambda) = \lambda^2 + a_1 \lambda + a_0$, where:

$$a_0 = 0, \tag{B39}$$
$$a_1 = \frac{(\beta - 1)\sigma}{-\beta\sigma + \omega(\phi_\pi - 1) + \sigma} = -\frac{(1 - \beta)\sigma}{\omega(\phi_\pi - 1) + \sigma(1 - \beta)}. \tag{B40}$$

For both eigenvalues of B.1 to be inside the unit circle, the following conditions must hold[10]:

$$\begin{aligned} |a_0| &< 1, \quad &\text{(B41)} \\ |a_1| &< 1 + a_0. \quad &\text{(B42)} \end{aligned}$$

With $a_0 = 0$ condition (B41) is always satisfied. Condition (B42) implies:

$$-\frac{(1-\beta)\sigma}{\omega(\phi_\pi - 1) + \sigma(1-\beta)} > -1, \implies \frac{(1-\beta)\sigma}{\omega(\phi_\pi - 1) + \sigma(1-\beta)} < 1$$
$$\implies \begin{cases} \phi_\pi > 1 \\ \omega(\phi_\pi - 1) < -\sigma(1-\beta) \end{cases} \implies \begin{cases} \phi_\pi > 1 \\ \phi_\pi < 1 - \frac{\sigma}{\omega}(1-\beta) \end{cases} \quad \text{(B43)}$$

E-stability. For the MSV solution to be E-stable the following should hold

$$eig \begin{pmatrix} \frac{\partial T_{\alpha,c}}{\alpha} & \frac{\partial T_{\alpha,c}}{\partial c} \\ \frac{\partial T_c}{\partial \alpha} & \frac{\partial T_c}{\partial c} \end{pmatrix} < 1.$$

With $T_{\alpha,c} = a_\pi + b_\pi \alpha$ and $T_c = b_\pi c + c_\pi$, the above condition converges to

$$eig \begin{pmatrix} b_\pi & 0 \\ 0 & b_\pi \end{pmatrix} < 1, \implies b_\pi < 1, \implies \beta - \omega\frac{\phi_\pi - 1}{\sigma} < 1, \implies \phi_\pi > 1 - (1-\beta)\frac{\sigma}{\omega}.$$

That is, for $\phi_\pi > 1$ the MSV solution is both determinate and E-stable. **Deriving REE when the rule with contemporaneous inflation expectations** is used. The model solution is:

$$\begin{aligned} y_t &= \frac{\phi_\pi - 1}{\sigma}\bar{\pi} - \frac{\phi_\pi}{\sigma}E_t\pi_t + \frac{1}{\sigma}E_t\pi_{t+1} + y_{t-1} + g_t, \quad &\text{(B44)} \\ \pi_t &= \omega(\frac{\phi_\pi - 1}{\sigma})\bar{\pi} + (\beta + \frac{\omega}{\sigma})E_t\pi_{t+1} - \frac{\omega\phi_\pi}{\sigma}E_t\pi_t + \omega y_{t-1} + \omega g_t + u_t. \quad &\text{(B45)} \end{aligned}$$

Definine the coefficients as

$$\begin{aligned} a_y &= \frac{\phi_\pi - 1}{\sigma}\bar{\pi}, \quad &\text{(B46)} \\ b_y^f &= \frac{1}{\sigma}, \quad &\text{(B47)} \\ b_y^c &= -\frac{\phi_\pi}{\sigma}, \quad &\text{(B48)} \\ c_y &= 1, \quad &\text{(B49)} \\ a_\pi &= \omega\frac{\phi_\pi - 1}{\sigma}\bar{\pi}, \quad &\text{(B50)} \\ b_\pi^f &= \beta + \frac{\omega}{\sigma}, \quad &\text{(B51)} \\ b_\pi^c &= -\frac{\omega\phi_\pi}{\sigma}, \quad &\text{(B52)} \\ c_\pi &= \omega. \quad &\text{(B53)} \end{aligned}$$

[10] See the references in Appendix A of Bullard and Mitra (2002).

With the REE forecasting rules as in (B9) and the rational expectation solution is:

$$\alpha = -\frac{a_\pi}{b_\pi^f}, \tag{B54}$$
$$b = \frac{1-b_\pi^c}{b_\pi^f}, \tag{B55}$$
$$c = -\frac{c_\pi}{b_\pi^f}. \tag{B56}$$

E-stability. To check for E-stability, we calculate the eigenvalues of the following matrix:

$$\begin{pmatrix} (1+b)b_\pi^f + b_\pi^c & b_\pi^f \alpha & 0 \\ 0 & 2bb_\pi^f b_\pi^c & 0 \\ 0 & cb_\pi^f & (1+b)b_\pi^f + b_\pi^c \end{pmatrix} \Longrightarrow \begin{pmatrix} b_\pi^f + 1 & -a_\pi & 0 \\ 0 & 2-b_\pi^c & 0 \\ 0 & -c_\pi & b_\pi^f + 1. \end{pmatrix}$$

The eigenvalues are $[b_\pi^f + 1, b_\pi^f + 1, 2-b_\pi^c]$. With $b_\pi^f > 0$, $b_\pi^f + 1 \geq 1$, what contradict E-stability condition.

Let us now consider MSV solution. With the MSV forecasting rules as in (B25), the MSV solution is:

$$\alpha = \frac{a_\pi}{1-b_\pi^f - b_\pi^c}, \tag{B57}$$
$$b = 0, \tag{B58}$$
$$c = \frac{c_\pi}{1-b_\pi^f - b_\pi^c}, \tag{B59}$$

The ALM coefficients are given by:

$$\bar{a}_y^{MSV} = a_y + (b_y^f + b_y^c)\frac{a_\pi}{1-b_\pi^f - b_\pi^c}, \tag{B60}$$
$$\bar{b}_y^{MSV} = 0, \tag{B61}$$
$$\bar{c}_y^{MSV} = c_y + (b_y^f + b_y^c)\frac{c_\pi}{1-b_\pi^f - b_\pi^c}, \tag{B62}$$
$$\bar{a}_\pi^{MSV} = a_\pi \frac{1+b_\pi^f + b_\pi^c}{1-b_\pi^f - b_\pi^c}, \tag{B63}$$
$$\bar{b}_\pi^{MSV} = 0, \tag{B64}$$
$$\bar{c}_\pi^{MSV} = c_\pi \frac{1+b_\pi^f + b_\pi^c}{1-b_\pi^f - b_\pi^c}. \tag{B65}$$

Determinacy. Similar to the forward-looking rule, we use the conditions (B41) and (B42) to get:

$$\alpha_0 = 0, \alpha_1 = -\bar{c}_y^{MSV} = -(c_y + (b_y^f + b_y^c)\frac{c_\pi}{1-b_\pi^f - b_\pi^c}) = -\frac{(1-\beta)\sigma}{\omega(\phi_\pi - 1) + \sigma(1-\beta)}, \tag{B66}$$

which is the same condition as for forward looking rule. That is why, for the contemporaneous rule, the condition on the policy feedback is the identical:

$$\begin{cases} \phi_\pi > 1 \\ \phi_\pi < 1 - \frac{\sigma}{\omega}(1-\beta) \end{cases} \tag{B67}$$

E-stability. We calculate the eigenvalues of the following matrix:

$$\begin{pmatrix} b^f_\pi + b^c_\pi & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & b^f_\pi + b^c_\pi. \end{pmatrix}$$

For the solution to be E-stable, $b^f_\pi + b^c_\pi < 1$. Plugging the values for the b^f_π and b^f_π, we arrive to the same conditions as under forward-looking rule: $\phi_\pi > 1 - \frac{\sigma}{\omega}(1-\beta)$. □

B.2 E-stability, determinacy and MSFE criterion

Proof. Proposition 3.0.2. **Forward-looking policy rule**. We first derive the coefficients in agents' forecast rules and in the actual law of motion for inflation and output conditional on agents using M_π rule:

$$\pi_t = \bar{a}_\pi + \bar{b}_\pi \pi_{t-1} + \bar{c}_\pi y_{t-1} + \omega g_t + u_t, \tag{B68}$$
$$y_t = \bar{a}_y + \bar{b}_y \pi_{t-1} + \bar{c}_y y_{t-1} + g_t, \tag{B69}$$

where ALM coefficients are the following:

$$\bar{a}_y = a_y + b_y \alpha_\pi (1 + \beta_\pi), \tag{B70}$$
$$\bar{b}_y = b_y (\beta_\pi)^2, \tag{B71}$$
$$\bar{c}_y = c_y, \tag{B72}$$
$$\bar{a}_\pi = a_\pi + b_\pi \alpha_\pi (1 + \beta_\pi), \tag{B73}$$
$$\bar{b}_\pi = b_\pi (\beta_\pi)^2, \tag{B74}$$
$$\bar{c}_\pi = c_\pi, \tag{B75}$$

with a_π, b_π, c_π, a_y, b_y and c_y defined in (B3) - (B8).

Deriving M_π. As in the case of the simple model we allow our agents to be econometricians who estimate the coefficients for their learning rule as regression coefficients. Then for M_π the coefficient is:

$$\beta_\pi = \frac{Cov(\pi_t, \pi_{t-1})}{Var(\pi_{t-1})} = \frac{Cov\left(\bar{a}_\pi + \bar{b}_\pi \pi_{t-1} + \bar{c}_\pi y_{t-1}, \pi_{t-1}\right)}{Var(\pi_{t-1})},$$

denoting $\sigma_{\pi y} \equiv Cov(y, \pi)$, $\sigma^2_\pi \equiv Var(\pi)$, $\sigma^2_y \equiv Var(y)$:

$$\beta_\pi = \frac{\bar{b}_\pi \sigma^2_\pi + \bar{c}_\pi \sigma_{\pi y}}{\sigma^2_\pi}$$
$$= \bar{b}_\pi + \bar{c}_\pi \frac{\sigma_{\pi y}}{\sigma^2_\pi}, \tag{B76}$$
$$\alpha_\pi = \bar{\pi} - \beta_\pi \bar{\pi}. \tag{B77}$$

Determinacy. Re-writing actual law of motion as $\bar{z}_t = \bar{A} + \bar{B}\bar{z}_{t-1} + U_t$, with $\bar{z}_t = (y_t, \pi_t)'$, $\bar{A}$ and U_t are vectors of constants and shocks respectively, and

$$\bar{B} = \begin{pmatrix} \bar{c}_y & \bar{b}_y \\ \bar{c}_\pi & \bar{b}_\pi \end{pmatrix},$$

the solution in (B76) is determinate only if real parts of eigenvalues of $\bar{B}$ are inside the unit circle. The coefficients of the characteristic polynomial $p(\lambda) = \lambda^2 + a_1\lambda + a_0$ are:

$$a_0 = \beta\beta_\pi^2, \tag{B78}$$
$$a_1 = -(1 + b_\pi\beta_\pi^2) \tag{B79}$$

Using the conditions in (B41) and (B42), this solution is determinate if and only if:

$$|-(1+b_\pi\beta_\pi^2)| < 1 + \beta\beta_\pi^2 \implies \begin{cases} 1+b_\pi\beta_\pi^2 > -1-\beta\beta_\pi^2 & \text{if } b_\pi\beta_\pi^2 > -1 \\ 1+b_\pi\beta_\pi^2 < 1+\beta\beta_\pi^2 & \text{if } b_\pi\beta_\pi^2 < -1 \end{cases}$$

$$\implies \begin{cases} 1+\beta_\pi^2(b_\pi+\beta) > 0 & \text{if } b_\pi\beta_\pi^2 > -1 \\ b_\pi < \beta & \text{if } b_\pi\beta_\pi^2 < -1 \end{cases} \tag{B80}$$

$$\beta < \frac{1}{\beta_\pi^2}. \tag{B81}$$

Conditions (B80) and (B81) are satisfied for any $\beta_\pi < 1$.

Deriving $\mathbf{M}_y$. Coefficient c_π^y is then determined as regression coefficient:

$$c_\pi^y = \frac{Cov(\pi_t, y_{t-1})}{Var(y_{t-1})} = \frac{Cov\left(\bar{a}_\pi + \bar{b}_\pi\pi_{t-1} + \bar{c}_\pi y_{t-1}, y_{t-1}\right)}{Var(y_{t-1})} =$$
$$= \frac{\bar{b}_\pi\sigma_{\pi y} + \bar{c}_\pi\sigma_y^2}{\sigma_y^2} = \bar{c}_\pi + \bar{b}_\pi\frac{\sigma_{\pi y}}{\sigma_y^2}, \tag{B82}$$
$$\alpha_\pi^y = \bar{\pi} - c_\pi^y\bar{y}. \tag{B83}$$

E-stability. With the coefficients of the actual law of motion defined as in (B70) -(B75). For M_π to be E-stable under least-squares learning the following must be satisfied

$$eig\begin{pmatrix} \frac{\partial T_\alpha}{\partial\alpha} & \frac{\partial T_\alpha}{\partial\beta} \\ \frac{\partial T_\beta}{\partial\alpha} & \frac{\partial T_\beta}{\partial\beta} \end{pmatrix} < 1,$$

where T_β and T_α are as in (B76) and (B77) accordingly. As $\frac{\partial T_\alpha}{\partial\alpha} = 0$, the conditions collapses for $\frac{\partial T_\beta}{\partial\beta} < 1$, see Figure B1.

For M_y to be E-stable under least-squares learning the following must be satisfied

$$eig\begin{pmatrix} \frac{\partial T_{\alpha_y}}{\partial\alpha_y} & \frac{\partial T_{\alpha_y}}{\partial c_y} \\ \frac{\partial T_{c_y}}{\partial\alpha_y} & \frac{\partial T_{c_y}}{\partial c_y} \end{pmatrix} < 1.$$

With T_{c_y} and T_{α_y} defined in (B82) and (B83) the above matrix has only one non-zero element - $\frac{\partial T_{\alpha_y}}{\partial\alpha_y} = -\bar{y}$. For our parameters calibration, steady state output gap is positive, and the condition is satisfied.

Moving to **contemporaneous inflation in the policy rule**, we solve for the coefficients of actual law of motion as:

$$\bar{a}_y = a_y + \alpha_\pi (b_y^c + b_y^f (1+\beta_\pi)), \tag{B84}$$
$$\bar{b}_y = \beta_\pi (b_y^c + b_y^f \beta_\pi), \tag{B85}$$
$$\bar{c}_y = 1, \tag{B86}$$
$$\bar{a}_\pi = a_\pi + \alpha_\pi (b_\pi^c + b_\pi^f (1+\beta_\pi)), \tag{B87}$$
$$\bar{b}_\pi = \beta_\pi (b_\pi^c + b_\pi^f \beta_\pi), \tag{B88}$$
$$\bar{c}_\pi = \omega. \tag{B89}$$

The solutions for M_π and M_y look the same as in (B76)-(B77) and in (B82)-(B83) respectively, where ALM coefficients are in (B84)-(B89).

Determinacy. Similarly to the above derivations, the corresponding conditions are:

$$|\beta_\pi^2 \beta| < 1, \tag{B90}$$
$$|-(1+\beta_\pi(b_\pi^c + \beta_\pi b_\pi^f)| < 1 + \beta_\pi^2 \beta, \tag{B91}$$

which can be re-arranged as:

$$\beta_\pi < \phi_\pi < \beta_\pi + \frac{2(1+\beta_\pi^2\beta)\sigma}{\beta_\pi \omega} \tag{B92}$$

E-stability. As $\frac{\partial T_\alpha}{\partial \alpha} = 0$, for M_π the conditions collapses for $\frac{\partial T_\beta}{\partial \beta} < 1$, see figure B1. For M_y the same logic as for the forward looking rule holds, and it is always E-stable under our parameter calibration.

MSFE. For the agents to use M_π as a forecasting rule in the equilibrium, the rule must produce better forecast as the alternative M_y. We compare the quality of the forecasts based on the mean squared forecast error criterion. In calculations below we denote the correlation between output and inflation as R. The correlation is then $R = \frac{\sigma_{\pi y}}{\sigma_y \sigma_\pi}$ with $\sigma_{\pi y} \equiv Cov(\pi, y)$, $\sigma_\pi \equiv \sqrt{Var(\pi)}$, $\sigma_y \equiv \sqrt{Var(y)}$. We start with mean forecast error of M_π. The forecast error of M_π is the difference between the forecast and the actual inflation:

$$\begin{aligned}
&: \quad e_t^\pi = (\alpha_\pi - \bar{a}_\pi) + (\beta_\pi - \bar{b}_\pi)\pi_{t-1} - \bar{c}_\pi y_{t-1} \\
&= \bar{\pi} - \bar{\pi}\bar{b}_\pi - \bar{\pi}\bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi} - a_\pi - \bar{\pi} b_\pi (1-\beta_\pi)(1+\beta_\pi) + (\bar{b}_\pi + \bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi} - \bar{b}_\pi)\pi_{t-1} - \bar{c}_\pi y_{t-1} \\
&= \bar{\pi} - \bar{\pi}\bar{b}_\pi - \bar{\pi}\bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi} - a_\pi - \bar{\pi} b_\pi + \bar{\pi} b_\pi \beta_\pi^2 + \bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi} \pi_{t-1} - \bar{c}_\pi y_{t-1}.
\end{aligned} \tag{B93}$$

Using that $\bar{b}_\pi = b_\pi \beta_\pi^2$, we group the expression to obtain

$$\begin{aligned}
&: \quad \bar{\pi} - a_\pi - \bar{\pi} b_\pi - \bar{\pi}\bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi} + \bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi}\pi_{t-1} - \bar{c}_\pi y_{t-1} \\
&= \bar{\pi} - a_\pi - \bar{\pi} b_\pi - \bar{c}_\pi \bar{y} + \bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi}(\pi_{t-1} - \bar{\pi}) - \bar{c}_\pi (y_{t-1} - \bar{y}),
\end{aligned} \tag{B94}$$

where in the last line we added and substracted $\bar{c}_\pi \bar{y}$. Now, with $\bar{c}_\pi = c_\pi = \omega$, $a_\pi = \bar{\pi}\omega\frac{\phi_p - 1}{\sigma}$, $b_\pi \bar{\pi} = -a_\pi + \beta\bar{\pi}$ and $c_\pi = \omega$, $\bar{y} = \frac{1-\beta}{\omega}\bar{\pi}$:

$$: \quad \bar{\pi} - a_\pi - \bar{\pi} b_\pi - \bar{c}_\pi \bar{y} = 0, \tag{B95}$$

$$: \quad e_t^\pi = \bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi}(\pi_{t-1} - \bar{\pi}) - \bar{c}_\pi (y_{t-1} - \bar{y}), \tag{B96}$$

$$\begin{aligned}
: \quad & MSFE_\pi = E_t(e_t^\pi)^2 = E_t[\bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi}(\pi_{t-1} - \bar{\pi}) - \bar{c}_\pi (y_{t-1} - \bar{y})]^2 \\
: \quad & = E_t[\bar{c}_\pi^2 R^2 \frac{\sigma_y^2}{\sigma_\pi^2}(\pi_{t-1} - \bar{\pi})^2 - 2\bar{c}_\pi R \frac{\sigma_y}{\sigma_\pi}\bar{c}_\pi (\pi_{t-1} - \bar{\pi})(y_{t-1} - \bar{y}) + \bar{c}_\pi^2 (y_{t-1} - \bar{y})^2] \\
: \quad & = \bar{c}_\pi^2 (R^2 \frac{\sigma_y^2}{\sigma_\pi^2}\sigma_\pi^2 + \sigma_y^2 - 2R\frac{\sigma_y}{\sigma_\pi}\sigma_{\pi y}) \\
: \quad & = \bar{c}_\pi^2 (R^2 \sigma_y^2 + \sigma_y^2 - 2R^2 \sigma_y^2) \\
: \quad & = \bar{c}_\pi^2 \sigma_y^2 (1 - R^2) + \sigma_\mu^2.
\end{aligned} \tag{B97}$$

Similarly, the forecast error of M$_y$:

$$\begin{aligned}
: \quad & e_t^y = \left(\alpha_\pi^y - \bar{a}_\pi\right) + \left(c_y^y - \bar{c}_\pi\right) y_{t-1} - \bar{b}_\pi \pi_{t-1} + \mu_t = \\
: \quad & = \left(\bar{\pi} - \bar{y}\bar{c}_\pi - \bar{y}\bar{b}_\pi \frac{\sigma_{\pi y}}{\sigma_y^2} - \bar{a}_\pi - \bar{\pi} b_\pi + \bar{\pi}\bar{b}_\pi\right) + \bar{b}_\pi \frac{\sigma_{\pi y}}{\sigma_y^2} y_{t-1} - \bar{b}_\pi \pi_{t-1} + \mu_t = \\
: \quad & = \bar{\pi} - a_\pi - b_\pi \bar{\pi} - \bar{y}\bar{c}_\pi + \bar{b}_\pi [\frac{\sigma_{\pi y}}{\sigma_y^2}(y_{t-1} - \bar{y}) - (\pi_{t-1} - \bar{\pi})] + \mu_t \\
: \quad & = \bar{b}_\pi [R\frac{\sigma_\pi}{\sigma_y}(y_{t-1} - \bar{y}) - (\pi_{t-1} - \bar{\pi})] + \mu_t
\end{aligned} \tag{B98}$$

$$\begin{aligned}
: \quad & MSFE_y = E[\bar{b}_\pi [R\frac{\sigma_\pi}{\sigma_y}(y_{t-1} - \bar{y}) - (\pi_{t-1} - \bar{\pi})]^2 \\
: \quad & = \bar{b}_\pi^2 \sigma_\pi^2 [1 - R^2] + \sigma_\mu^2.
\end{aligned} \tag{B99}$$

We are looking for the conditions under which $MSFE_\pi < MSFE_y$:

$$\bar{c}_\pi^2 \left[\sigma_y^2 - R^2 \sigma_y^2\right] < \bar{b}_\pi^2 \left[\sigma_\pi^2 - R^2 \sigma_\pi^2\right], \tag{B100}$$

denoting $\Sigma = \frac{\sigma_\pi}{\sigma_y}$ and $\bar{\Gamma} = \frac{\bar{b}_\pi}{\bar{c}_\pi}$.And the criterion is simply:

$$\bar{\Gamma}^2 \Sigma^2 > 1. \tag{B101}$$

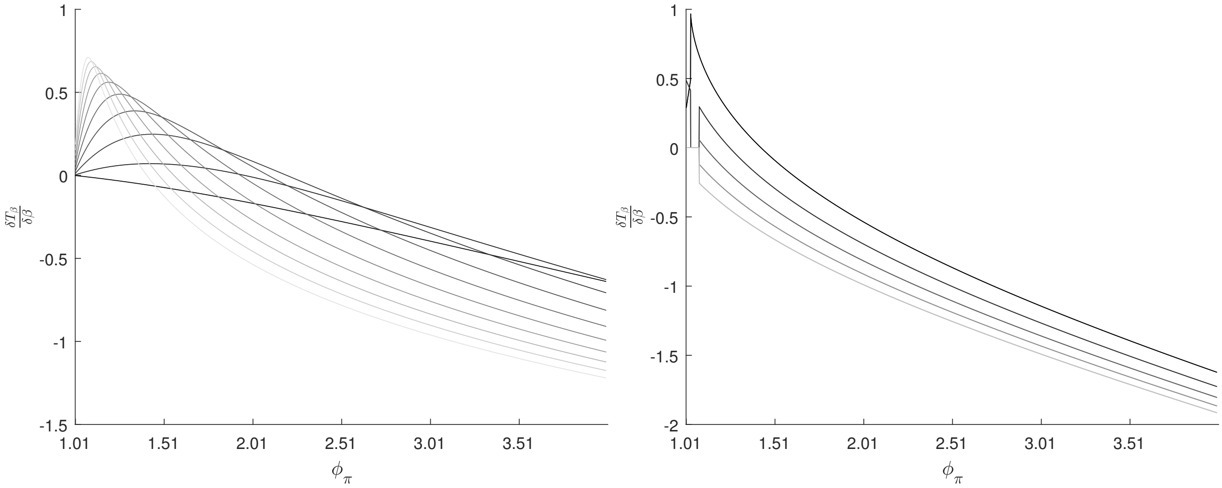
□

Variances and covariance

The procedure is similar to Adam (2005) With the actual law of motion as in (B68) and (B69) under different policy rules we derive variances and covariance of output and inflation. Denote $z_t = (\pi_t, y_t)'$ and $B = \begin{pmatrix} \bar{b}_\pi & \bar{c}_\pi \\ \bar{b}_y & \bar{c}_y \end{pmatrix}$, $U_t = (u_t, g_t)'$. Then the (B68) can be represented as:

$$z_t = Bz_{t-1} + U_t, \tag{B102}$$

Figure B1: E-stability



(a) Forward Looking Rule ***(b) Contemporaneous Rule***

Note: the figure for forward-looking rule is drawn for r in the range [0.1 2], where each line corresponds to different r - the darker is the line, the larger is r. The figure for contemporaneuos rule is drawn for r in the range [0.1 1]

where for the purpose of variance and covariance calculation we have dropped the unimportant constants. Denote $r^2 \equiv \frac{\sigma_u^2}{\sigma_g^2}$ and $\Omega \equiv E(ug)^2 = \sigma_g^2 \begin{pmatrix} r^2+\omega^2 & \omega \\ \omega & 1 \end{pmatrix}$ is variance covariance matrix of the shocks. We take variance of (B102) to get:

$$\mathscr{E} = B\mathscr{E}B' + \Omega, \tag{B103}$$

where $\mathscr{E} = E(zz')$. Vectorizing matrix $\mathscr{E}$:

$$\begin{aligned} vec(\mathscr{E}) &= (B \otimes B)vec(\mathscr{E}) + vec(\Omega) \\ &= (I - B \otimes B)^{-1} vec(\Omega). \end{aligned} \tag{B104}$$

To obtain the covariance with lagged values, we multiply both sides of (B102) by z_{t-1} and take expectations to get:

$$\Gamma = B\mathscr{E}, \tag{B105}$$

where $\Gamma = E(z_t z_{t-1}')$ is covariance matrix with the lagged values and

$$vec(\Gamma) = (I \otimes B)vec(\mathscr{E}). \tag{B106}$$

Then variances and covariances are as follows:

$$\begin{aligned} \sigma_{y\pi} &= vec(\mathscr{E})[2], && \text{(B107)} \\ \sigma_\pi^2 &= vec(\mathscr{E})[1], && \text{(B108)} \\ \sigma_y^2 &= vec(\mathscr{E})[4], && \text{(B109)} \\ E(\pi_t \pi_{t-1}) &= vec(\Gamma)[1], && \text{(B110)} \\ E(\pi_t y_{t-1}) &= vec(\Gamma)[3]. && \text{(B111)} \end{aligned}$$

Appendix C: Weights

Proof. Proposition 3.0.3. Consider (45) and (46) and using that $\bar{c}_\pi > 0$ and $\bar{b}_\pi \neq 0$ [11] :

$$
\begin{aligned}
m_\pi > m_y \quad &: \quad \frac{\kappa}{\bar{b}_\pi^2 \sigma_\pi^2 \sigma_y \left(1 - R^2\right)} \frac{\bar{c}_\pi \sigma_y - \bar{b}_\pi \sigma_\pi R}{\bar{c}_\pi} < \frac{\kappa}{\bar{c}_\pi^2 \sigma_y^2 \sigma_\pi \left(1 - R^2\right)} \frac{\bar{b}_\pi \sigma_\pi - \bar{c}_\pi \sigma_y R}{\bar{b}_\pi}, \\
&: \quad \begin{cases} \frac{\bar{c}_\pi \sigma_y - \bar{b}_\pi \sigma_\pi R}{\bar{b}_\pi \sigma_\pi} < \frac{\bar{b}_\pi \sigma_\pi - \bar{c}_\pi \sigma_y R}{\bar{c}_\pi \sigma_y}, & \text{if } \bar{b}_\pi > 0 \\ \frac{\bar{c}_\pi \sigma_y - \bar{b}_\pi \sigma_\pi R}{\bar{b}_\pi \sigma_\pi} > \frac{\bar{b}_\pi \sigma_\pi - \bar{c}_\pi \sigma_y R}{\bar{c}_\pi \sigma_y}, & \text{if } \bar{b}_\pi < 0 \end{cases} \\
&: \quad \left(\bar{c}_\pi \sigma_y - \bar{b}_\pi \sigma_\pi R\right) \bar{c}_\pi \sigma_y < \bar{b}_\pi \sigma_\pi \left(\bar{b}_\pi \sigma_\pi - \bar{c}_\pi \sigma_y R\right), \\
&: \quad \bar{c}_\pi^2 \sigma_y^2 - \bar{c}_\pi \sigma_y \bar{b}_\pi \sigma_\pi R < \bar{b}_\pi^2 \sigma_\pi^2 - \bar{b}_\pi \sigma_\pi \bar{c}_\pi \sigma_y R, \\
&: \quad \bar{c}_\pi^2 \sigma_y^2 < \bar{b}_\pi^2 \sigma_\pi^2, \\
&: \quad \bar{\Gamma}^2 \Sigma^2 > 1.
\end{aligned}
$$

□

Proof. Proposition 3.0.4 Rearranging (45)

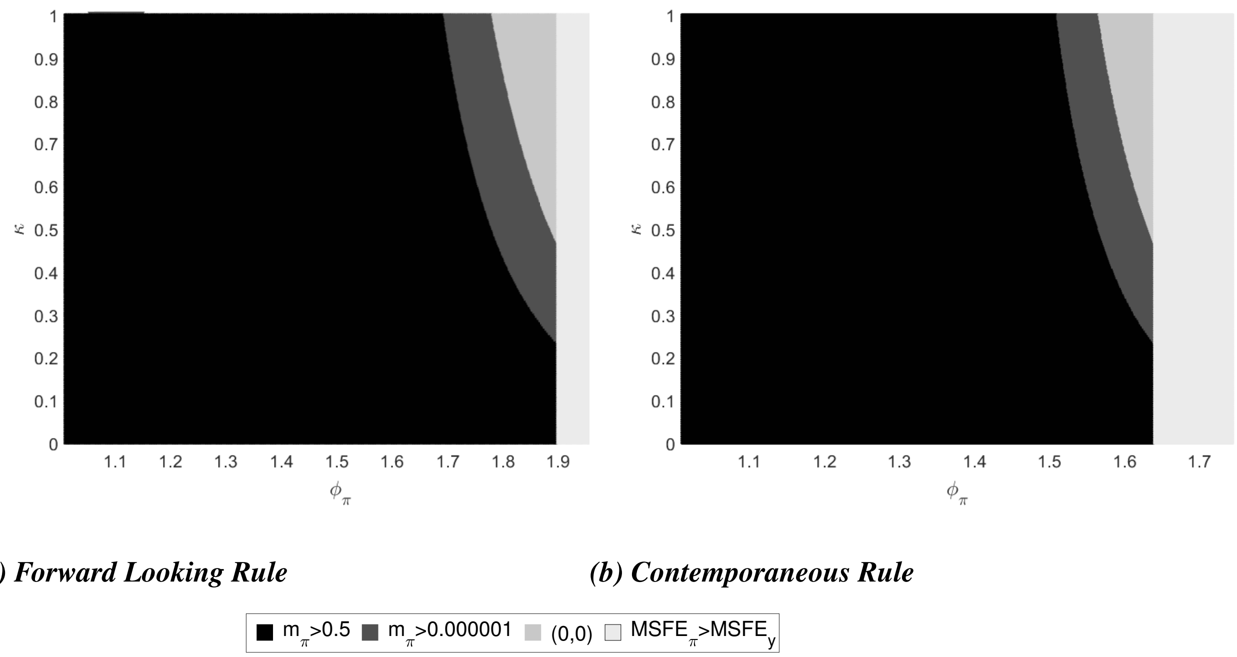
$$
\begin{aligned}
&: \quad \frac{\kappa}{\bar{c}_\pi^2 \sigma_y^2 \sigma_\pi \left(1 - R^2\right)} \frac{\bar{b}_\pi \sigma_\pi - \bar{c}_\pi \sigma_y R}{\bar{b}_\pi} < 1 \\
&: \quad \kappa < \frac{\bar{b}_\pi \bar{c}_\pi^2 \sigma_y^2 \sigma_\pi \left(1 - R^2\right)}{\bar{b}_\pi \sigma_\pi - \bar{c}_\pi \sigma_y R} = \frac{\left(1 - R^2\right) \bar{c}_\pi^2 \sigma_y^2}{1 - R \frac{\bar{c}_\pi \sigma_y}{\bar{b}_\pi \sigma_\pi}}
\end{aligned}
$$

where in the second step we again use that (42) implies $|\bar{b}_\pi| \sigma_\pi > \bar{c}_\pi \sigma_y$, which means that even for $\bar{b}_\pi < 0$, $\frac{\bar{b}_\pi \sigma_\pi - \bar{c}_\pi \sigma_y R}{\bar{b}_\pi} > 0$. $R \in [-1, 1]$, that is maximum of the expression $-\bar{c}_\pi \sigma_y R$, linear in R, is $\bar{c}_\pi \sigma_y < |\bar{b}_\pi| \sigma_\pi$. As was shown in Appendix A.2, the condition for inner solution is suficient for shock to have positive weight in the forecast. □

Appendix D: Weight on Inflation

[11] With $\bar{c}_\pi > 0$, $\bar{b}_\pi = 0$ means violation of (42), hence with $\bar{b}_\pi = 0$ M_π can not be equilibrium choice.

Figure D1: Model Selection under Sparse Weights, M_π.



(a) Forward Looking Rule ***(b) Contemporaneous Rule***