

# A Behavioral New Keynesian Model

Xavier Gabaix\*

March 31, 2018

## Abstract

This paper presents a framework for analyzing how bounded rationality affects monetary and fiscal policy. The model is a tractable and parsimonious enrichment of the widely-used New Keynesian model – with one main new parameter, which quantifies how poorly agents understand future economic disturbances. That myopia parameter, in turn, affects the power of monetary and fiscal policy in a microfounded general equilibrium.

A number of consequences emerge. (i) Fiscal stimulus or “helicopter drops of money” are powerful and, indeed, pull the economy out of the zero lower bound. More generally, the model allows for the joint analysis of optimal monetary and fiscal policy. (ii) The Taylor principle is strongly modified: even with passive monetary policy, equilibrium is determinate, whereas the traditional rational model yields multiple equilibria, which reduces its predictive power, and generates indeterminate economies at the zero lower bound (ZLB). (iii) The ZLB is much less costly than in the traditional model. (iv) The model helps solve the “forward guidance puzzle”: the fact that in the rational model, shocks to very distant rates have a very powerful impact on today’s consumption and inflation; because agents are partially myopic, this effect is muted. (v) Optimal policy changes qualitatively: the optimal commitment policy with rational agents demands “nominal GDP targeting”; this is not the case with behavioral firms, as the benefits of commitment are less strong with myopic firms. (vi) The model is “neo-Fisherian” in the long run, but Keynesian in the short run: a permanent rise in the interest rate decreases inflation in the short run but increases it in the long run. The non-standard behavioral features of the model seem warranted by extant and new empirical evidence.

---

\*[xgabaix@fas.harvard.edu](mailto:xgabaix@fas.harvard.edu). I thank Igor Cesarec, Vu Chau, Antonio Coppola, Wu Di, James Graham and Lingxuan Wu for excellent research assistance. For useful comments I thank the editor and referees, Marios Angeletos, Adrien Auclert, Larry Ball, Olivier Blanchard, Jeff Campbell, Larry Christiano, John Cochrane, Tim Cogley, Gauti Eggertsson, Emmanuel Farhi, Roger Farmer, Jordi Galí, Mark Gertler, Narayana Kocherlakota, Greg Mankiw, Ricardo Reis, Dongho Song, Jim Stock, Michael Woodford, and participants at various seminars and conferences. I am grateful to the CGEB, the Institute for New Economic Thinking, the NSF (SES-1325181) and the Sloan Foundation for financial support.

# 1 Introduction

This paper proposes a way to analyze monetary and fiscal policy when agents are not fully rational. To do so, it enriches the basic model of monetary policy, the New Keynesian (NK) model, by incorporating behavioral factors. In the baseline NK model the agent is fully rational (though prices are sticky). Here, in contrast, the agent is partially myopic to unusual events and does not anticipate the future perfectly. The formulation takes the form of a parsimonious generalization of the traditional model that allows for the analysis of monetary and fiscal policy. This has a number of strong consequences for aggregate outcomes.

1. Fiscal policy is much more powerful than in the traditional model.<sup>1</sup> In the traditional model, rational agents are Ricardian and do not react to tax cuts. In the present behavioral model, agents are partly myopic, and consume more when they receive tax cuts or “helicopter drops of money” from the central bank. As a result, we can study the interaction between monetary and fiscal policy.
2. The Taylor principle is strongly modified. Equilibrium selection issues vanish in many cases: for instance, even with a constant nominal interest rate there is just one (bounded) equilibrium.
3. Relatedly, the model can explain the stability in economies stuck at the zero lower bound (ZLB), something that is difficult to achieve in traditional models.
4. The ZLB is much less costly.
5. Forward guidance is much less powerful than in the traditional model, offering a natural behavioral resolution of the “forward guidance puzzle”.
6. Optimal policy changes qualitatively: for instance, the optimal commitment policy with rational agents demands “price level targeting”. This is not the case with behavioral firms.
7. A number of neo-Fisherian paradoxes are resolved. A permanent rise in the nominal interest rate causes inflation to fall in the short run (a Keynesian effect) and rise in the long run (so that the long-run Fisher neutrality holds with respect to inflation).

In addition, I will argue that there is reasonable empirical evidence for the main non-standard features of the model. The paper estimates behavioral factors, and finds that they indeed are warranted by the empirical evidence.

Let me expand on the above points.

*Fiscal policy and helicopter drops of money.* In the traditional NK model, agents are fully rational. So Ricardian equivalence holds, and fiscal policy (i.e. lump-sum tax changes, as opposed to government expenditure) has no impact. Here, in contrast, the agent is not Ricardian because he fails to perfectly anticipate future taxes. As a result, tax cuts and transfers are unusually stimulative, particularly if they happen in the present. As the agent is partially myopic, taxes are best enacted in the present.

At the ZLB, only forward guidance (or, in more general models, quantitative easing) is available, and in the rational model optimal policy only leads to a complicated second best. However, in this

---

<sup>1</sup>By “fiscal policy” I mean government transfers, i.e. changes in (lump-sum) taxes. In the traditional Ricardian model, they have no effect (Barro (1974)). This is in contrast to government consumption, which does have an effect even in the traditional model.

model, the central bank (and more generally the government) has a new instrument: it can restore the first best by doing “helicopter drops of money”, i.e. by sending checks to people – via fiscal policy.

*Zero lower bound (ZLB).* Depressions due to the ZLB are unboundedly large in the rational model, probably counterfactually so (e.g. Werning (2012)). This is because agents unflinchingly respect their Euler equations. In contrast, depressions are moderate and bounded in this behavioral model – closer to reality and common sense.

*The Taylor principle reconsidered and equilibrium determinacy.* When monetary policy is passive (e.g. via a constant interest rate rule, or when it violates the Taylor principle that monetary policy should strongly lean against economic conditions), the traditional model has a continuum of (bounded) equilibria, so that the response to a simple question like “What happens when interest rates are kept constant?” is ill-defined: it is mired in the morass of equilibrium selection. In contrast, in this behavioral model there is just one (bounded) equilibrium: things are clean and definite theoretically.

*Economic stability.* Determinacy is not just a purely theoretical question. In the rational model, if the economy is stuck at the ZLB forever the Taylor principle is violated (as the nominal interest rate is stuck at 0%). The equilibrium is therefore indeterminate: we could expect the economy to jump randomly from one period to the next (we shall see that a similar phenomenon happens if the ZLB lasts for a large but finite duration). However, we do not see that in Japan since the late 1980s or in the Western economies in the aftermath of the 2008 crisis (Cochrane (2017)). This can be explained with this behavioral model if agents are myopic enough and if firms rely enough on “inflation guidance” by the central bank.

*Forward guidance.* With rational agents, “forward guidance” by the central bank is predicted to work very powerfully, most likely too much so, as emphasized by Del Negro et al. (2015) and McKay et al. (2016). The reason is again that the traditional consumer rigidly respects his Euler equation and expects other agents to do the same, so that a movement of the interest rate far in the future has a strong impact today. However, in the behavioral model I put forth, this impact is muted by the agent’s myopia, which makes forward guidance less powerful. The model, in reduced form, takes the form of a “discounted Euler equation”, where the agent reacts in a discounted manner to future consumption growth.

*Optimal policy changes qualitatively.* With rational firms, the optimal commitment policy entails “price level targeting” (which gives, when GDP is trend-stationary, “nominal GDP targeting”): after a cost-push shock, monetary policy should partially let inflation rise, but then create deflation, so that eventually the price level and nominal GDP come back to their pre-shock trend. This is because with rational firms, there are strong benefits from commitment to being very tough in the future (Clarida et al. (1999)). With behavioral firms, in contrast, the benefits from commitment are lower, and after the cost-push shock the central bank does not find it useful to engineer a great deflation and come back to the initial price level. Hence, price level targeting and nominal GDP targeting are not desirable when firms are behavioral.

*A number of neo-Fisherian paradoxes vanish.* A number of authors, especially Cochrane (2017), highlight that in the rational New Keynesian model, a permanent rise in interest rates leads to an immediate rise in inflation, which is paradoxical.<sup>2</sup> This is called the “neo-Fisherian” property. In the present behavioral model, the property holds in the long run: the long-run real rate is independent of monetary policy (Fisher neutrality holds). However, in the short run, raising rates does lower inflation and output, as in the Keynesian model.

---

<sup>2</sup>This can depend on which equilibrium is selected, leading to some cacophony in the dialogue.

I build on the large New Keynesian literature, as distilled in Woodford (2003b) and Galí (2015). I am indebted to the number of authors who identified paradoxes in the New Keynesian model, e.g. Cochrane (2017), Del Negro et al. (2015), McKay et al. (2016). For the behavioral model, I rely on the general dynamic setup derived in Gabaix (2016), itself building on a general static “sparsity approach” to behavioral economics laid out in Gabaix (2014). The sparsity model is particularly tractable because it uses deterministic models (unlike models with noisy signals), and continuous parameters. As a result, it applies to microeconomic problems like basic consumer theory and Arrow-Debreu-style general equilibrium (Gabaix (2014) – something as of yet not done by other modelling techniques), dynamic macroeconomics (Gabaix (2016)), and public economics (Farhi and Gabaix (2017)).

At the same time, this research is sympathetic to many works studying departures from traditional rationality in macroeconomics. They are discussed in detail in Section 6.

Section 2 presents basic model assumptions and derives its main building blocks, summarized in Proposition 2.10. Section 3 derives the positive implications of the model. Section 4 studies optimal monetary and fiscal policy with behavioral agents. Section 5 econometrically evaluates the model. For that, it also supplies a simple extension that can handle changes to trend inflation. Section 6 discusses the literature, and Section 7 concludes. Section 8 presents detailed microfoundations for the behavioral model. Section 9 presents an elementary 2-period model with behavioral agents. I recommend it to entrants to this literature. The rest of the appendix contains additional proofs and details.

*Notations.* I distinguish between  $\mathbb{E}[X]$ , the objective expectation of  $X$ , and  $\mathbb{E}^{BR}[X]$ , the expectation under the agent’s boundedly rational (BR) model of the world.

Though the exposition is largely self-contained, this paper is in part a behavioral version of Chapters 2-5 of the Galí (2015) textbook, itself in part a summary of Woodford (2003b). My notations are typically those of Galí, except that  $\gamma$  is risk aversion, something that Galí denotes with  $\sigma$ . In concordance with the broader literature, I use  $\sigma$  for the (“effective”) intertemporal elasticity of substitution.

I call the economy “determinate” (in the sense of Blanchard and Kahn (1980)) if, given initial conditions, there is only one non-explosive equilibrium path.

## 2 A Behavioral Model

Let us first recall the notations of the rational NK model. I call  $x_t$  the output gap (i.e. the deviation of GDP from its efficient level). Hence, positive  $x_t$  corresponds to a boom, negative  $x_t$  to a recession. With rational agents, the traditional NK model gives microfoundations that lead to:

$$x_t = \mathbb{E}_t[x_{t+1}] - \sigma(i_t - \mathbb{E}_t\pi_{t+1} - r_t^n), \tag{1}$$

$$\pi_t = \beta\mathbb{E}_t[\pi_{t+1}] + \kappa x_t, \tag{2}$$

where  $i_t$  is the nominal short term interest rate,  $\pi_t$  is inflation, and  $r_t^n$  is the “natural real interest rate”, which is the interest rate that would prevail if all pricing frictions were removed.

I now present other foundations leading to a behavioral model that has the traditional rational outcome in (1)-(2) as a particular case.

## 2.1 Behavioral Agent: Basic model for the IS curve

**Setup: Objective reality.** I consider an agent with standard utility

$$U = \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t, N_t) \text{ with } u(c, N) = \frac{c^{1-\gamma} - 1}{1-\gamma} - \frac{N^{1+\phi}}{1+\phi}, \quad (3)$$

where  $c_t$  is consumption, and  $N_t$  is labor supply (as in Number of hours supplied). The real wage is  $\omega_t$ . The real interest rate is  $r_t$  and agent's real income is  $y_t = \omega_t N_t + y_t^f$ : the sum of labor income  $\omega_t N_t$  and profit income  $y_t^f$  (as in income coming from firms); later we will add taxes. His real financial wealth  $k_t$  evolves as:

$$k_{t+1} = (1 + r_t) (k_t - c_t + y_t). \quad (4)$$

The agent's problem is  $\max_{(c_t, N_t)_{t \geq 0}} U$  subject to (4), and the usual transversality condition ( $\lim_{t \rightarrow \infty} \beta^t c_t^{-\gamma} k_t = 0$ ), which I will omit mentioning from now on.

The aggregate production of the economy is  $c_t = e^{\zeta_t} N_t$ , where productivity  $\zeta_t$  follows an AR(1) process with mean 0. There is no capital, as in the baseline New Keynesian model.

Consider first the case where the economy is deterministic at the steady state ( $\zeta_t \equiv 0$ ), so that the interest rate, income, and real wage, consumption and labor supply are at their steady-state values  $\bar{r}$ ,  $\bar{y}$ ,  $\bar{\omega}$ ,  $\bar{c}$ ,  $\bar{N}$ . We have a simple deterministic problem. Defining  $R := 1 + \bar{r}$ , we have  $R = 1/\beta$ . To correct monopolistic distortions, I assume that the government has put in place the usual corrective production subsidies, financed by a lump-sum tax on firms (so that profits are 0 on average). Hence, at the steady state the economy operates efficiently and  $\bar{c} = \bar{N} = \bar{\omega} = \bar{y} = 1$ .<sup>3</sup>

Let us now go back to the general case, outside of the steady state. There is a state vector  $\mathbf{X}_t$  (comprising productivity  $\zeta_t$ , as well as announced actions in monetary and fiscal policy), that will evolve in equilibrium as:

$$\mathbf{X}_{t+1} = \mathbf{G}^{\mathbf{X}}(\mathbf{X}_t, \epsilon_{t+1}) \quad (5)$$

for some equilibrium transition function  $\mathbf{G}^{\mathbf{X}}$  and mean-0 innovations  $\epsilon_{t+1}$ .

I decompose the values as deviations from the above steady state, for example:

$$r_t = \bar{r} + \hat{r}_t, \quad y_t = \bar{y} + \hat{y}_t,$$

and those deviations are function of the state:

$$\hat{r}_t = \hat{r}(\mathbf{X}_t), \quad \hat{y}_t = \hat{y}(N_t, \mathbf{X}_t) := \omega(\mathbf{X}_t) N_t + y^f(\mathbf{X}_t) - \bar{y},$$

where the functions of  $\mathbf{X}_t$  are determined in equilibrium. The law of motion for private financial wealth  $k_t$  is<sup>4</sup>

$$k_{t+1} = G^k(c_t, N_t, k_t, \mathbf{X}_t) := (1 + \bar{r} + \hat{r}(\mathbf{X}_t)) (k_t + \bar{y} + \hat{y}(N_t, \mathbf{X}_t) - c_t), \quad (6)$$

<sup>3</sup>Indeed, when  $\zeta = 0$ ,  $\bar{\omega} = 1$ , and labor supply satisfies  $\bar{\omega} u_c + u_N = 0$ , i.e.  $\bar{N}^\phi = \bar{\omega} \bar{c}^{-\gamma}$ , with the resource constraint:  $\bar{c} = \bar{N}$ .

<sup>4</sup>As there is no aggregate capital, financial wealth is  $k_t = 0$  in equilibrium in the basic model without government debt. But we need to consider potential deviations from  $k_t = 0$  when studying the agent's consumption problem. When later we add government debt  $B_t$ , we will have  $k_t = B_t$  in equilibrium.

so the agent’s problem can be rewritten as  $\max_{(c_t, N_t)_{t \geq 0}} U$  subject to (5) and (6).

I assume that  $\mathbf{X}_t$  has mean 0, i.e. has been de-meanned. Linearizing, the law of motion becomes:

$$\mathbf{X}_{t+1} = \mathbf{\Gamma} \mathbf{X}_t + \boldsymbol{\epsilon}_{t+1} \quad (7)$$

for some matrix  $\mathbf{\Gamma}$ , after perhaps a renormalization of  $\boldsymbol{\epsilon}_{t+1}$ . Likewise, linearizing we will have  $\hat{r}(\mathbf{X}) = \mathbf{b}_{\mathbf{X}}^r \mathbf{X}$ , for some factor  $\mathbf{b}_{\mathbf{X}}^r$ .

**Setup: Reality perceived by the behavioral agent** I can now describe the behavioral agent. The main assumption is the following:<sup>5</sup>

**Assumption 2.1** (Cognitive discounting of the state vector) *The agent perceives that the state vector evolves as:*

$$\mathbf{X}_{t+1} = \bar{m} \mathbf{G}^{\mathbf{X}}(\mathbf{X}_t, \boldsymbol{\epsilon}_{t+1}), \quad (8)$$

where  $\bar{m} \in [0, 1]$  is a “cognitive discounting” parameter measuring attention to the future.

Then, given this perception, the agent solves  $\max_{(c_t, N_t)_{t \geq 0}} U$  subject to (6) and (8).

To better interpret  $\bar{m}$ , let us linearize (8):

$$\mathbf{X}_{t+1} = \bar{m} (\mathbf{\Gamma} \mathbf{X}_t + \boldsymbol{\epsilon}_{t+1}). \quad (9)$$

Hence the expectation of the behavioral agent is  $\mathbb{E}_t^{BR}[\mathbf{X}_{t+1}] = \bar{m} \mathbf{\Gamma} \mathbf{X}_t$  and, iterating,  $\mathbb{E}_t^{BR}[\mathbf{X}_{t+k}] = \bar{m}^k \mathbf{\Gamma}^k \mathbf{X}_t$ , while the rational expectation is  $\mathbb{E}_t[\mathbf{X}_{t+k}] = \mathbf{\Gamma}^k \mathbf{X}_t$  (the rational policy always obtains from setting the attention parameters to 1).<sup>67</sup> Hence:

$$\mathbb{E}_t^{BR}[\mathbf{X}_{t+k}] = \bar{m}^k \mathbb{E}_t[\mathbf{X}_{t+k}], \quad (10)$$

where  $\mathbb{E}_t^{BR}[\mathbf{X}_{t+k}]$  is the subjective expectation by the behavioral agent, and  $\mathbb{E}_t[\mathbf{X}_{t+k}]$  is the rational expectation. The more distant the events in the future, the more the behavioral agent “sees them dimly”, i.e. sees them with a dampened cognitive discount factor  $\bar{m}^k$  at horizon  $k$  (recall that  $\bar{m} \in [0, 1]$ ). The parameter  $\bar{m}$  models a form of “global cognitive discounting” – discounting future disturbances more as they are more distant in the future. Importantly, this implies that all perceived variables will embed some cognitive discounting:<sup>8</sup>

<sup>5</sup>I particularize the formalism in Gabaix (2016), which is a tractable way to model dynamic programming with limited attention. “Cognitive discounting” was laid out as a possibility in that paper (as a misperception of autocorrelations), but its concrete impact was not studied in any detail there.

<sup>6</sup>When the mean of  $\mathbf{X}_t$  is not 0, but rather  $\mathbf{X}_*$  such that  $\mathbf{X}_* = \mathbf{G}(\mathbf{X}_*, 0)$ , then the process perceived by the behavioral agent is:  $\mathbf{X}_{t+1} = (1 - \bar{m}) \mathbf{X}_* + \bar{m} \mathbf{G}(\mathbf{X}_t, \boldsymbol{\epsilon}_{t+1})$ . Then, we have, linearizing,  $\mathbb{E}_t^{BR}[\mathbf{X}_{t+k} - \mathbf{X}_*] = \bar{m}^k \mathbb{E}_t[\mathbf{X}_{t+k} - \mathbf{X}_*]$ .

<sup>7</sup>There is no long term growth in this model, as in the basic New Keynesian model. It is easy though not central to introduce it (see Section 11.8 of the online appendix). The behavioral agent would be rational with respect to the values around the balanced growth path, but myopic for the deviations from it.

<sup>8</sup>Linearizing, we have  $z(\mathbf{X}) = \mathbf{b}_{\mathbf{X}}^z \mathbf{X}$  for some row vector  $\mathbf{b}_{\mathbf{X}}^z$ , and:

$$\mathbb{E}_t^{BR}[z(\mathbf{X}_{t+k})] = \mathbb{E}_t^{BR}[\mathbf{b}_{\mathbf{X}}^z \mathbf{X}_{t+k}] = \mathbf{b}_{\mathbf{X}}^z \mathbb{E}_t^{BR}[\mathbf{X}_{t+k}] = \mathbf{b}_{\mathbf{X}}^z \bar{m}^k \mathbb{E}_t[\mathbf{X}_{t+k}] = \bar{m}^k \mathbb{E}_t[\mathbf{b}_{\mathbf{X}}^z \mathbf{X}_{t+k}] = \bar{m}^k \mathbb{E}_t[z(\mathbf{X}_{t+k})].$$

**Lemma 2.2** (Cognitive discounting of all variables) *For any variable  $z(\mathbf{X}_t)$  with  $z(0) = 0$ , the beliefs of the behavioral agent satisfy, for all  $k \geq 0$ , and linearizing:*

$$\mathbb{E}_t^{BR} [z(\mathbf{X}_{t+k})] = \bar{m}^k \mathbb{E}_t [z(\mathbf{X}_{t+k})], \quad (11)$$

where  $\mathbb{E}_t^{BR}$  is the subjective (behavioral) expectation operator, which uses the misperceived law of motion (8), and  $\mathbb{E}_t$  is the rational one, which uses the rational law of motion (5).

For instance, the interest rate perceived in  $k$  periods is

$$\mathbb{E}_t^{BR} [\bar{r} + \hat{r}(\mathbf{X}_{t+k})] = \bar{r} + \bar{m}^k \mathbb{E}_t [\hat{r}(\mathbf{X}_{t+k})].$$

The agent perceives correctly the average interest rate  $\bar{r}$  and is globally patient, like the rational agent, but he perceives myopically future deviations from the average interest rate (i.e.  $\mathbb{E}_t [\hat{r}(\mathbf{X}_{t+k})]$  is dampened by  $\bar{m}^k$ ).

**Behavioral IS curve** We can now derive the IS (investment-saving) curve. The Euler equation of a rational agent is:  $\mathbb{E}_t \left[ \beta R_t \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} \right] = 1$ . Linearizing, we get:<sup>9</sup>

$$\hat{c}_t = \mathbb{E}_t [\hat{c}_{t+1}] - \frac{1}{\gamma R} \hat{r}_t. \quad (12)$$

This is the traditional derivation of the IS curve, with rational agents.

Now call  $c(\mathbf{X}_t, k_t)$  the equilibrium consumption of the behavioral agent. Under the agent's subjective model, we have:<sup>10</sup>  $\mathbb{E}_t^{BR} \left[ \beta R_t \left( \frac{c(\mathbf{X}_{t+1}, k_{t+1})}{c(\mathbf{X}_t, k_t)} \right)^{-\gamma} \right] = 1$ . Now, in general equilibrium, there is zero financial wealth,  $k_t = 0$ , and income and consumption are the same (so  $c_t = \bar{y} + \hat{y}(N_t, \mathbf{X}_t)$ ) and private wealth is  $k_t = 0$ . Hence, given (6), the agent correctly anticipates that her beginning of period  $t + 1$  private wealth will be  $k_{t+1} = 0$ .<sup>11</sup> It follows that aggregate consumption  $c(\mathbf{X}_t) = c(\mathbf{X}_t, 0)$  satisfies

$$\mathbb{E}_t^{BR} \left[ \beta R_t \left( \frac{c(\mathbf{X}_{t+1})}{c(\mathbf{X}_t)} \right)^{-\gamma} \right] = 1.$$

Linearizing, this gives:

$$\hat{c}(\mathbf{X}_t) = \mathbb{E}_t^{BR} [\hat{c}(\mathbf{X}_{t+1})] - \frac{1}{\gamma R} \hat{r}_t.$$

---

<sup>9</sup>Indeed, using  $\beta \bar{R} = 1$  and  $R_t = \bar{R} + \hat{r}_t$ ,

$$1 = \mathbb{E}_t [\beta R_t \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma}] = \mathbb{E}_t [\beta \bar{R} \left( 1 + \frac{\hat{r}_t}{\bar{R}} \right) \left( \frac{1 + \hat{c}_{t+1}}{1 + \hat{c}_t} \right)^{-\gamma}] \simeq 1 + \frac{\hat{r}_t}{\bar{R}} - \gamma \mathbb{E}_t [\hat{c}_{t+1} - \hat{c}_t],$$

which gives (12). Galí (2015) does not have the  $\frac{1}{R}$  term as he defines the interest rate as  $r_t^{\text{Galí}} := \ln R_t$ , whereas in the present paper it is defined as  $r_t := R_t - 1$ , so that  $\hat{r}_t^{\text{Galí}} \equiv \frac{\hat{r}_t}{R}$ . The predictions are the same, adjusting for the slightly different convention.

<sup>10</sup>To be very formal,  $\mathbb{E}_t \left[ \beta R_t (c(\bar{m} \mathbf{G}^{\mathbf{X}}(\mathbf{X}_t, \epsilon_{t+1}), G^k(c_t, N_t, k_t, \mathbf{X}_t)) / c(\mathbf{X}_t, k_t))^{-\gamma} \right] = 1$ .

<sup>11</sup>When the agent has non-zero private wealth (which is the case with taxes) or when she misperceives her income, the derivation is more complex, as we shall see in Section 2.3.

Now, by Lemma 2.2,  $\mathbb{E}_t^{BR} [\hat{c}_t(\mathbf{X}_{t+1})] = \bar{m}\mathbb{E}_t[\hat{c}_t(\mathbf{X}_{t+1})]$ , so we obtain

$$\hat{c}_t = M\mathbb{E}_t[\hat{c}_{t+1}] - \sigma\hat{r}_t, \quad (13)$$

with  $M = \bar{m}$  and  $\sigma = \frac{1}{\gamma R}$ . Equation (13) is a “discounted aggregate Euler equation”. I call  $M$  the macro parameter of attention. Here  $M = \bar{m}$ , but in more general specifications coming later,  $M \neq \bar{m}$ , so, anticipating them, I keep the notation  $M$  for the macro attention.

Let us next link (13) to the output gap. First, the static first order condition for labor supply holds:<sup>12</sup>

$$N_t^\phi = \omega_t c_t^{-\gamma}. \quad (14)$$

Next, call  $c_t^n$  and  $r_t^n$  the natural rate of output and interest, defined as the quantity of output and interest that would prevail if we removed all pricing friction, and use hats to denote them as deviations from the steady state,  $\hat{c}_t^n := c_t^n - \bar{c}$  and  $\hat{r}_t^n := r_t^n - \bar{r}$ . The natural rate of output is easy to derive;<sup>13</sup> it is

$$\hat{c}_t^n = \frac{1 + \phi}{\gamma + \phi} \zeta_t. \quad (15)$$

Next, note that equation (13) also holds in that “natural” economy that would have no pricing frictions. So,

$$\hat{c}_t^n = M\mathbb{E}_t[\hat{c}_{t+1}^n] - \sigma\hat{r}_t^n, \quad (16)$$

which gives that the natural rate of interest  $r_t^n = r_t^{n0}$ , where

$$r_t^{n0} = \bar{r} + \frac{1 + \phi}{\sigma(\gamma + \phi)} (M\mathbb{E}_t[\zeta_{t+1}] - \zeta_t). \quad (17)$$

I call this interest rate  $r_t^{n0}$  the “pure” natural rate of interest—this is the interest rate that prevails in an economy without pricing frictions, and undisturbed by government policy (in particular, budget deficits). So when there are no budget deficits (as is the case here)  $r_t^n = r_t^{n0}$ , but in later specifications the two concepts will differ. Behavioral forces don’t change the natural rate of output, but they do change the pure natural rate of interest.<sup>14</sup>

The output gap is  $x_t := \hat{c}_t - \hat{c}_t^n$ . Then, taking (13) minus (16), we obtain:

$$x_t = M\mathbb{E}_t[x_{t+1}] - \sigma(\hat{r}_t - \hat{r}_t^n). \quad (18)$$

Rearranging,  $\hat{r}_t - \hat{r}_t^n = (r_t - \bar{r}) - (r_t^n - \bar{r}) = i_t - \mathbb{E}_t[\pi_{t+1}] - r_t^n$ , where  $i_t$  is the nominal interest rate. We obtain the following result:<sup>15</sup>

**Proposition 2.3** (Discounted Euler equation) *Consider the simplest model with only cognitive discounting ( $\bar{m}$ ). In equilibrium, the output gap  $x_t$  follows:*

$$x_t = M\mathbb{E}_t[x_{t+1}] - \sigma(i_t - \mathbb{E}_t[\pi_{t+1}] - r_t^n), \quad (19)$$

<sup>12</sup>It holds under the behavioral agent’s subjective model, and is identical to the rational one.

<sup>13</sup>The resource constraint is  $c_t = e^{\zeta_t} N_t$ , and with flexible prices,  $\omega_t = e^{\zeta_t}$ . Together with (14), we obtain natural rate of output,  $\ln c_t^n = \frac{1+\phi}{\gamma+\phi} \zeta_t$ ; linearizing around  $\bar{c} = 1$ , so that  $\ln c_t^n \simeq \hat{c}_t^n$ , we get the announced value.

<sup>14</sup>In a model with physical capital, behavioral forces would change the natural rate of output.

<sup>15</sup> Substantially, the agents anchors on the steady state. This implies that cognitive discounting is about the deviation of output from the steady state (and not just the output gap).

where  $M = \bar{m} \in [0, 1]$  is the macro attention parameter, and  $\sigma := \frac{1}{\gamma R}$ . In the rational model,  $M = 1$ .

The behavioral NK IS curve (18) implies:

$$x_t = -\sigma \sum_{k \geq 0} M^k \mathbb{E}_t [\hat{r}_{t+k} - \hat{r}_{t+k}^n], \quad (20)$$

In the rational case with  $M = 1$ , a one-period change in the real interest rate  $\hat{r}_{t+k}$  in 1000 periods has the same impact on the output gap as a change occurring today. This is intuitively very odd, and is an expression of the forward guidance puzzle. However, when  $M < 1$ , a change occurring in 1000 periods has a much smaller impact as a change occurring today.<sup>16</sup>

## 2.2 Phillips Curve with Behavioral Firms

Next, I explore what happens if firms do not fully pay attention to future macro variables either. The economy consists of a Dixit-Stiglitz continuum of firms. Firm  $i$  produces output  $Y_{it} = N_{it} e^{\zeta_t}$ , and sets a price  $P_{it}$ . The final good is produced competitively in quantity  $Y_t = \left( \int_0^1 Y_{it}^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}$ , so that its price is:

$$P_t = \left( \int_0^1 P_{it}^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}. \quad (21)$$

Firms have the usual Calvo pricing friction: at each period, they can reset their price with probability  $1 - \theta$ .

**Setup: Objective reality facing firms** Consider a firm  $i$ , and call  $q_{i\tau} := \ln \frac{P_{i\tau}}{P_\tau} = p_{i\tau} - p_\tau$  its real log price at time  $\tau$ . Its real profit is

$$v_\tau = \left( \frac{P_{i\tau}}{P_\tau} - MC_\tau \right) \left( \frac{P_{i\tau}}{P_\tau} \right)^{-\varepsilon} c_\tau.$$

Here  $\left( \frac{P_{i\tau}}{P_\tau} \right)^{-\varepsilon} c_\tau$  is the total demand for the firm's good, with  $c_\tau$  aggregate consumption;  $MC_\tau = (1 - \tau_f) \frac{\omega_\tau}{e^{\zeta_\tau}} = (1 - \tau_f) e^{-\mu_\tau}$  is the real marginal cost;  $-\mu_\tau := \ln \omega_\tau - \zeta_\tau$  is the social real marginal cost.<sup>17</sup> A corrective wage subsidy  $\tau_f = \frac{1}{\varepsilon}$  ensures that there are no price distortions on average. For simplicity I assume that this subsidy is financed by a lump-sum tax on firms, which affects  $v_\tau$  by an additive value, so that it does not change the pricing decision:  $v_\tau$  is the firm's profit before the lump-sum tax. It is equal to:

$$v(q_{i\tau}, \mu_\tau, c_\tau) := \left( e^{q_{i\tau}} - (1 - \tau_f) e^{-\mu_\tau} \right) e^{-\varepsilon q_{i\tau}} c_\tau. \quad (22)$$

I consider the worldview at time  $t$  of a firm simulating the future. Call  $\mathbf{X}_\tau$  the extended macro state vector  $\mathbf{X}_\tau = (\mathbf{X}_\tau^M, \Pi_\tau)$  where  $\Pi_\tau := p_\tau - p_t = \pi_{t+1} + \dots + \pi_\tau$  is inflation between times  $t$  and  $\tau$ , and  $\mathbf{X}_\tau^M$  is the vector of macro variables: TFP  $\zeta_t$ , as well as possible announcements about future policy. Then, if the firm hasn't changed its price between  $t$  and  $\tau$ , its real price is  $q_{i\tau} = q_{it} - \Pi_\tau$ , so the flow profit at  $\tau$  is:

$$v^{\text{rat}}(q_{it}, \mathbf{X}_\tau) := v(q_{it} - \Pi(\mathbf{X}_\tau), \mu(\mathbf{X}_\tau), c(\mathbf{X}_\tau)), \quad (23)$$

<sup>16</sup>I defer to future research the exploration of asset pricing implications of this sort of model. That would require adding non-trivial risks, e.g. disaster risk.

<sup>17</sup>Equivalently,  $\mu_t$  is a "social markup"; and  $\mu_t = 0$  at efficiency.

where  $\Pi_\tau := \Pi(\mathbf{X}_\tau)$  is aggregate future inflation, and similarly for  $\mu$  and  $c$ . A traditional Calvo firm which can reset its price at  $t$  wants to choose the optimal real price  $q_{it}$  to maximize total profits, as in:

$$\max_{q_{it}} \mathbb{E}_t \sum_{\tau=t}^{\infty} (\beta\theta)^{\tau-t} \frac{c(\mathbf{X}_\tau)^{-\gamma}}{c(\mathbf{X}_t)^{-\gamma}} v^{\text{rat}}(q_{it}, \mathbf{X}_\tau), \quad (24)$$

where  $\frac{c(\mathbf{X}_\tau)^{-\gamma}}{c(\mathbf{X}_t)^{-\gamma}}$  is the adjustment in the stochastic discount factor between  $t$  and  $\tau$ . Linearizing around the deterministic steady state,  $\frac{c(\mathbf{X}_\tau)^{-\gamma}}{c(\mathbf{X}_t)^{-\gamma}} \simeq 1$ , so that term will not matter in the linearizations.

**Setup: Reality perceived by a behavioral firm** The behavioral firm faces the same problem, with a less accurate view of reality. Most importantly, I posit that the behavioral firm also perceives the future via the cognitive discounting mechanism in (8). To be precise, I model that, at time  $t$ , the firm perceives the future profit at date  $\tau \geq t$  as:

$$v^{BR}(q_{it}, \mathbf{X}_\tau) := v(q_{it} - m_\pi^f \Pi(\mathbf{X}_\tau), m_x^f \mu(\mathbf{X}_\tau), c(\mathbf{X}_\tau)), \quad (25)$$

where  $v$  is as in (22). This means that the firm, when simulating the future, sees only a fraction  $m_\pi^f$  of future inflation  $\Pi(\mathbf{X}_\tau)$ , and a fraction  $m_x^f$  of the future marginal cost  $-\mu(\mathbf{X}_\tau)$  (recall that those two quantities have been normalized to have mean 0 at the steady state). When all the  $m$ 's are equal to 1, we recover the traditional rational firm from the New Keynesian model. The most important parameter is  $\bar{m}$ , while the other parameters  $m_\pi^f$ ,  $m_x^f$  should be considered optional enrichments. The behavioral firm wants to optimize its initial real price level  $q_{it}$ :

$$\max_{q_{it}} \mathbb{E}_t^{BR} \sum_{\tau=t}^{\infty} (\beta\theta)^{\tau-t} \frac{c(\mathbf{X}_\tau)^{-\gamma}}{c(\mathbf{X}_t)^{-\gamma}} v^{BR}(q_{it}, \mathbf{X}_\tau) \quad (26)$$

with the perceived law of motion given in (8), reflecting cognitive dampening. The nominal price that firm  $i$  will choose will be  $p_t^* = q_{it} + p_t$ , and its value is as in the following lemma (the derivation is in section 10.2).

**Lemma 2.4** (Optimal price for a behavioral firm resetting its price) *A behavioral firm resetting its price at time  $t$  will set it to a value  $p_t^*$  equal to:*

$$p_t^* = p_t + (1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta\bar{m})^k \mathbb{E}_t [m_\pi^f (\pi_{t+1} + \dots + \pi_{t+k}) - m_x^f \mu_{t+k}], \quad (27)$$

where  $m_\pi^f$  and  $m_x^f$  parameterize attention to inflation and macro disturbances, respectively, and  $\bar{m}$  is the overall cognitive discounting factor.

**The resulting aggregate behavior of inflation** Tracing out the implications of (27), the macro outcome is as follows (the derivation is in section 10.2).

**Proposition 2.5** (Phillips curve with behavioral firms) *When firms are partially inattentive to future macro conditions, the Phillips curve becomes:*

$$\pi_t = \beta M^f \mathbb{E}_t [\pi_{t+1}] + \kappa x_t, \quad (28)$$

with the attention coefficient  $M^f := \bar{m} \left( \theta + \frac{1-\beta\theta}{1-\beta\theta\bar{m}} m_\pi^f (1-\theta) \right)$  and  $\kappa = \bar{\kappa} m_x^f$ , where  $\bar{\kappa}$  (given in (117)) is the value of  $\kappa$  in the traditional model with full attention. Aggregate inflation is more forward-looking ( $M^f$  is higher) when prices are sticky for a longer period of time ( $\theta$  is higher) and when firms are more attentive to future macroeconomic outcomes ( $m_\pi^f, \bar{m}$  are higher). When  $m_\pi^f = m_x^f = \bar{m} = 1$  (traditional firms), we recover the usual model, and  $M^f = 1$ .

In the traditional model, the coefficient on future inflation in (28) is exactly  $\beta$  and, quite miraculously, does not depend on the adjustment rate of prices  $\theta$ . In the behavioral model (with  $m_\pi^f < 1$ ), in contrast, the coefficient ( $\beta M^f$ ) is higher when prices are stickier for longer (higher  $\theta$ ).<sup>18</sup>

Firms can be fully attentive to all idiosyncratic terms (something that would be easy to include), e.g. the idiosyncratic part of their productivity. For the purposes of this result, they simply have to pay limited attention to macro outcomes. If we include idiosyncratic terms, and firms are fully attentive to them, the aggregate NK curve does not change. Also, firms are still fully rational for steady state variables (e.g., in the steady state they discount future profits at rate  $R = \frac{1}{\beta}$ ).<sup>19</sup> It is only their sensitivity to deviations from the deterministic steady state that is partially myopic.

## 2.3 Extension: Term structure of attention and misperception of fiscal policy

In this section, I enrich the assumptions of the basic model, though most of the paper could be conducted with (8) only. At the first reading, I recommend to skip to Section 2.4.

For conceptual and empirical reasons, I wish to explore the possibility that agent may, for instance, perceive the interest rate less accurately than income. To capture that, I assume that the agent perceives the law of motion of wealth as:

$$k_{t+1} = \mathbf{G}^{k, BR}(c_t, N_t, k_t, \mathbf{X}_t) := (1 + \bar{r} + \hat{r}^{BR}(\mathbf{X}_t)) (k_t + \bar{y} + \hat{y}^{BR}(N_t, \mathbf{X}_t) - c_t), \quad (29)$$

where  $\hat{r}^{BR}(\mathbf{X}_t)$  and  $\hat{y}^{BR}(N_t, \mathbf{X}_t)$  are the perceived interest rate and income, given by:

$$\hat{r}^{BR}(\mathbf{X}_t) = m_r \hat{r}(\mathbf{X}_t), \quad \hat{y}^{BR}(N_t, \mathbf{X}_t) = m_y \hat{y}(\mathbf{X}_t) + \omega(\mathbf{X}_t) (N_t - N(\mathbf{X}_t)), \quad (30)$$

and where  $m_r, m_y$  are attention parameters in  $[0, 1]$ , and  $\hat{r}(\mathbf{X}_t), \hat{y}(N_t, \mathbf{X}_t)$  are the true values of interest rate and personal income, while  $\hat{y}(\mathbf{X}_t) = \hat{y}(N(\mathbf{X}_t), \mathbf{X}_t)$  is the true value aggregate income (given aggregate labor supply  $N(\mathbf{X}_t)$ ) – all expressed as deviations from the steady state. When  $m_r, m_y$  and  $\bar{m}$  are equal to 1, the agent is the traditional, rational agent. Here  $m_r, m_y$  capture the attention to the interest rate and income. For instance, if  $m_r = 0$ , the agent “doesn’t pay attention” to the interest rate – formally, he replaces it by  $\bar{r}$  in his perceived law of motion. When  $m_r \in (0, 1)$ , he partially takes into account the interest rate – really, the deviations of the interest rate from its mean.

Here  $\hat{y}^{BR}(N_t, \mathbf{X}_t)$  is his perceived income, and perceived aggregate income is  $\hat{y}^{BR}(\mathbf{X}_t) = \hat{y}^{BR}(N(\mathbf{X}_t), \mathbf{X}_t) = m_y \hat{y}(\mathbf{X}_t)$ : the agent perceives only a fraction of income. However, he correctly perceives that the marginal income is  $\frac{\partial}{\partial N} \hat{y}^{BR}(N_t, \mathbf{X}_t) = w_t$ . This captures that the agent is smart enough to appreciate fully today’s marginal impact of working more, though anticipating his total income is

<sup>18</sup>Here I use the same  $\bar{m}$  for consumer and firms. If firms had their own rate of cognitive discounting  $\bar{m}^f$ , then one would simply replace  $\bar{m}$  by  $\bar{m}^f$  in the expression for  $M^f$  and in (27).

<sup>19</sup>Note that  $\beta R = 1$  pins down the value of  $\beta$ . So, one could not accommodate an anomalous Phillips curve by just changing  $\beta$ : that would automatically change the interest rate.

harder, especially in the future. Given these perceptions, the agent solves  $\max_{(c_t, N_t)_{t \geq 0}} U$  subject to (8) and (29).

**Term structure of attention to interest rate and income** This formulation, together with Lemma 2.2, implies:<sup>20</sup>

**Lemma 2.6** (Term structure of attention) *We have:*

$$\mathbb{E}_t^{BR} [\hat{r}^{BR}(\mathbf{X}_{t+k})] = m_r \bar{m}^k \mathbb{E}_t [\hat{r}(\mathbf{X}_{t+k})], \quad \mathbb{E}_t^{BR} [\hat{y}^{BR}(\mathbf{X}_{t+k})] = m_y \bar{m}^k \mathbb{E}_t [\hat{y}(\mathbf{X}_{t+k})]. \quad (31)$$

In words, for the interest rate (the same holds for income):

$$\text{Perceived deviation in } k \text{ periods} = m_r \bar{m}^k \times (\text{True deviation in } k \text{ periods}).$$

Hence, we obtain a “term structure of attention”. The factor  $m_r$  is the “level” or “intercept” of attention, while the factor  $\bar{m}$  is the “slope” of attention as a function of the horizon. The same holds for aggregate income.

If the reader seeks a model with just one free parameter, I recommend setting  $m_r = m_y = 1$  (the rational values) and keeping  $\bar{m}$  as the main parameter governing inattention.

**Consumption and labor supply** I now detail the consequences of these enrichments, for a behavioral agent with small initial wealth (Section 10.2 gives the derivation).<sup>21</sup>

**Proposition 2.7** (Behavioral consumption function) *In this behavioral model, consumption is:  $c_t = c_t^d + \hat{c}_t$ , with  $c_t^d = \bar{y} + b_k k_t$ ,  $b_k := \frac{\bar{r}}{R} \frac{\phi}{\phi + \gamma}$ , and, up to second order terms:*

$$\hat{c}_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} \frac{\bar{m}^{\tau-t}}{R^{\tau-t}} \left( b_r m_r \hat{r}(\mathbf{X}_\tau) + m_y \frac{\bar{r}}{R} \hat{y}(\mathbf{X}_\tau) \right) \right], \quad (32)$$

with  $b_r := \frac{-1}{\gamma R^2}$ , and  $m_y := \frac{\phi m_y + \gamma}{\phi + \gamma}$ . Labor supply satisfies the usual condition  $N_t^\phi = \omega_t c_t^{-\gamma}$ , i.e., in deviations from the steady state,  $\hat{N}_t = \frac{1}{\phi} \hat{\omega}_t - \frac{\gamma}{\phi} \hat{c}_t$ . The policy of the rational agent is a particular case, setting  $\bar{m}, m_r, m_y$  to 1.

In (32), consumption reacts to future interest rates and income deviations, dampening future values by a factor  $\bar{m}^{\tau-t}$  at horizon  $\tau - t$ , as in (31). Note that this agent is “globally patient” for steady-state variables. For instance, her marginal propensity to consume wealth is  $\frac{\bar{r}}{R} \frac{\phi}{\phi + \gamma}$ , like for the rational agent.<sup>22</sup> However, she is myopic to small macroeconomic disturbances in the economy.

<sup>20</sup>We have:

$$\mathbb{E}_t^{BR} [\hat{r}^{BR}(\mathbf{X}_{t+k})] = \mathbb{E}_t^{BR} [m_r \hat{r}(\mathbf{X}_{t+k})] = m_r \mathbb{E}_t^{BR} [\hat{r}(\mathbf{X}_{t+k})] = m_r \bar{m}^k \mathbb{E}_t [\hat{r}(\mathbf{X}_{t+k})].$$

<sup>21</sup>I allow for  $k_t$  different from 0, as private wealth will be non-zero when there is an active fiscal policy.

<sup>22</sup>There is a subtlety here, which Section 11.2 details. The MPC out of wealth is only  $\frac{\bar{r}}{R} \frac{\phi}{\phi + \gamma}$ , because higher wealth translates into not just more consumption of goods, but also more leisure. However, future booms have an impact of consumption that is  $\frac{\bar{r}}{R}$  when  $m_y = 1$ . This is because they affect the agent’s decisions both through higher income, and through higher wages.

**The behavioral IS curve with imperfect attention to income and interest rate** I next solve for the general equilibrium consequences of policy (32). To simplify notations, I now call  $r = R - 1 = \bar{r}$  the steady-state real interest rate. The resulting IS curve is next (the derivation, in Section 10.2, is instructive, and quite simple).

**Proposition 2.8** (Discounted Euler equation) *In the enriched model with partial attention to income and interest rate, we obtain a variant of the behavioral IS curve (18) in which  $M = \frac{\bar{m}}{R - rm_Y} \in [0, 1]$  for the macro parameter of attention, and  $\sigma := \frac{m_r}{\gamma R(R - rm_Y)} \in \left[0, \frac{1}{\gamma R}\right]$ . In the rational model,  $M = 1$ .*

**Understanding discounting in rational and behavioral models** It is worth pondering where the discounting by  $M$  comes from in (20). What is the impact at time 0 of a one-period fall of the real interest rate  $\hat{r}_\tau$ , in partial and general equilibrium, in both the rational and the behavioral model (as in Angeletos and Lian (2017a) and Farhi and Werning (2017))? For simplicity, I take  $\hat{r}_\tau^n = 0$  here.

Let us start with the *rational model*. In partial equilibrium (i.e., taking future income as given), a change in the future real interest rate  $\hat{r}_\tau$  changes time-0 consumption by the direct (i.e. partial equilibrium) impact (see (32)):

$$\text{Rational agent: } \Delta^{\text{direct}} := \left. \frac{\partial \hat{c}_0}{\partial \hat{r}_\tau} \right|_{(y_t)_{t \geq 0} \text{ held constant}} = -\alpha \frac{1}{R^\tau},$$

where  $\alpha := \frac{1}{\gamma R^2}$ . Hence, there is discounting by  $\frac{1}{R^\tau}$ . However, in general equilibrium (i.e., when the impact of  $\hat{r}_\tau$  on income flows  $(y_t)_{t \geq 0}$  is taken into account), the impact is (see (20) with  $M = 1$ ),

$$\text{Rational agent: } \Delta^{\text{GE}} := \frac{d\hat{c}_0}{d\hat{r}_\tau} = -\alpha R,$$

so that there is no discounting by  $\frac{1}{R^{\tau+1}}$ . The reason is the following: the rational agent sees the “first round of impact”, that is  $-\alpha \frac{\hat{r}_\tau}{R^\tau}$ ; a future interest rate cut will raise consumption. But he also sees how this increase in consumption will increase other agents’ future consumptions, hence increase his future income, hence his own consumption: this is the second-round effect. Iterating all other rounds (as in the Keynesian cross), the initial impulse is greatly magnified via this aggregate demand channel: though the first round (direct) impact is  $-\alpha \frac{\hat{r}_\tau}{R^\tau}$ , the full impact (including indirect channels) is  $-\alpha R \hat{r}_\tau$ . This means that the total impact is larger than the direct effect by a factor

$$\frac{\Delta^{\text{GE}}}{\Delta^{\text{direct}}} = R^{\tau+1}.$$

At large horizons  $\tau$ , this is a large multiplier. Note that this large general equilibrium effect relies upon common knowledge of rationality: the agent needs to assume that other agents are fully rational. This is a very strong assumption, typically rejected in most experimental setups (see the literature on the  $p$ -beauty contest, e.g. Nagel (1995)).

In contrast, in the *behavioral model*, the agent is not fully attentive to future innovations. So first, the direct impact of a change in interest rates is smaller:

$$\text{Behavioral agent: } \Delta^{\text{direct}} := \left. \frac{\partial \hat{c}_0}{\partial \hat{r}_\tau} \right|_{(y_t)_{t \geq 0} \text{ held constant}} = -\alpha m_r \bar{m}^\tau \frac{1}{R^\tau},$$

which comes from (32). Next, the agent is not fully attentive to indirect effects (including general equilibrium) of future policies. This results in the total effect in (20):

$$\text{Behavioral agent: } \Delta^{\text{GE}} := \frac{d\hat{c}_0}{d\hat{r}_\tau} = -\alpha m_r M^\tau \frac{R}{R - rm_Y},$$

so the multiplier for the general equilibrium effect is (as  $M = \frac{\bar{m}}{R - rm_Y}$ )

$$\frac{\Delta^{\text{GE}}}{\Delta^{\text{direct}}} = \left( \frac{R}{R - rm_Y} \right)^{\tau+1} \in [1, R^{\tau+1}] \quad (33)$$

and is smaller than the multiplier  $R^{\tau+1}$  in economies with common knowledge of rationality. As  $m_Y$  becomes smaller, the multiplier weakens: distant changes in interest rates will be very ineffective if agents are extremely myopic.

**Extension: Behavioral IS curve with fiscal policy** Finally, I generalize the above IS curve to the case of an active fiscal policy. In this paper, fiscal policy means cash transfers from the government to agents and lump-sum taxes (government consumption is zero). Hence, it would be completely ineffective in the traditional model, which features rational, Ricardian consumers. I call  $B_t$  the real value of government debt in period  $t$ , before period- $t$  taxes. Linearizing, it evolves as  $B_{t+1} = R(B_t + \mathcal{T}_t)$ , where  $\mathcal{T}_t$  is the lump-sum transfer given by the government to the agent (so that  $-\mathcal{T}_t$  is a tax).<sup>23</sup> I also define  $d_t$ , the budget deficit (after the payment of the interest rate on debt) in period  $t$ ,  $d_t := \mathcal{T}_t + \frac{r}{R}B_t$ , so that public debt evolves as:<sup>24</sup>

$$B_{t+1} = B_t + Rd_t. \quad (34)$$

Section 10.1 details the specific assumption I use to capture the agent's worldview. Summarizing, the situation is thus. Suppose that the government runs a deficit and gives a rebate  $d_t$  to the agents. Agents see the increase in their income, but, because of cognitive discounting, they see only partially the associated future taxes. Hence, they spend some of that transfer, and increases their consumption. The macroeconomic impact of that is as follows.

**Proposition 2.9** (Discounted Euler equation with sensitivity to budget deficits) *Because agents are not Ricardian, budget deficits temporarily increase economic activity. The IS curve (18) becomes:*

$$x_t = M\mathbb{E}_t[x_{t+1}] + b_d d_t - \sigma(i_t - \mathbb{E}_t[\pi_{t+1}] - r_t^{n0}), \quad (35)$$

where  $r_t^{n0}$  is the "pure" natural rate with zero deficits (derived in (17)),  $d_t$  is the budget deficit and  $b_d = \frac{\phi m_y r R(1-\bar{m})}{(\phi+\gamma)(R-m_Y r)(R-\bar{m})}$  is the sensitivity to deficits. When agents are rational,  $b_d = 0$ , but with behavioral agents,  $b_d > 0$ . In the sequel, we will write this equation by saying that the behavioral IS curve (19) holds, but with the following modified natural rate, which captures the stimulative action of deficits:

$$r_t^n = r_t^{n0} + \frac{b_d}{\sigma} d_t. \quad (36)$$

<sup>23</sup>Without linearization,  $B_{t+1} = \frac{1+i_t}{1+\pi_{t+1}}(B_t + \mathcal{T}_t)$ , where  $\frac{1+i_t}{1+\pi_{t+1}}$  is the realized gross return on bonds. Linearizing,  $B_{t+1} = R(B_t + \mathcal{T}_t)$ . Formally I consider the case of small debts and deficits, which allows us to neglect the variations of the real rate (i.e. second-order terms  $O\left(\left|\frac{1+i_t}{1+\pi_{t+1}} - R\right|(|B_t| + |d_t|)\right)$ ).

<sup>24</sup>Indeed,  $B_{t+1} = R(B_t - \frac{r}{R}B_t + d_t) = B_t + Rd_t$ .

Hence, bounded rationality gives both a discounted IS curve and an impact of fiscal policy:  $b_d > 0$ .<sup>25</sup>

Here I assume a representative agent. This analysis complements analyses that assume heterogeneous agents to model non-Ricardian agents, in particular rule-of-thumb agents à la Campbell and Mankiw (1989), Galí et al. (2007), Mankiw (2000), Bilbiie (2008), Mankiw and Weinzierl (2011) and Woodford (2013).<sup>26,27</sup> When dealing with complex situations, a representative agent is often simpler. In particular, it allows us to evaluate welfare unambiguously.

## 2.4 Synthesis: Behavioral New Keynesian Model

I now gather the above results.

**Proposition 2.10** (Behavioral New Keynesian model – two equation version) *We have the following behavioral version of the New Keynesian model, for the behavior of the output gap  $x_t$  and of inflation  $\pi_t$ :*

$$x_t = M\mathbb{E}_t[x_{t+1}] - \sigma(i_t - \mathbb{E}_t\pi_{t+1} - r_t^n) \quad (IS \text{ curve}), \quad (37)$$

$$\pi_t = \beta M^f \mathbb{E}_t[\pi_{t+1}] + \kappa x_t \quad (Phillips \text{ curve}), \quad (38)$$

where  $M, M^f \in [0, 1]$  are the macro-attention of consumers and firms, respectively, to macroeconomic outcomes

$$M = \frac{\bar{m}}{R - rm_Y}, \quad \sigma = \frac{m_r}{\gamma R(R - rm_Y)}, \quad (39)$$

$$M^f = \bar{m} \left( \theta + \frac{1 - \beta\theta}{1 - \beta\theta\bar{m}} m_\pi^f (1 - \theta) \right), \quad \kappa = \bar{\kappa} m_x^f. \quad (40)$$

*Bounded rationality causes agents to be non-Ricardian, so that fiscal policy is stimulative: the natural rate follows*

$$r_t^n = r_t^{n0} + \frac{b_d}{\sigma} d_t, \quad (41)$$

where  $r_t^{n0}$  is the “pure” natural interest rate that prevails with zero deficits (derived in (17)), and  $b_d = \frac{\phi m_y r R(1 - \bar{m})}{(\phi + \gamma)(R - m_Y r)(R - \bar{m})} \geq 0$  is the impact of deficits.

In the traditional model,  $\bar{m} = m_Y = m_r = m_\pi^f = m_x^f = 1$ , so that  $M = M^f = 1$  and  $b_d = 0$ . In addition,  $\bar{\kappa} = \left(\frac{1}{\theta} - 1\right)(1 - \beta\theta)(\gamma + \phi)$  is the value of the Phillips curve slope with fully rational firms.

The reader may be bewildered by a model with five behavioral parameters. Fortunately, only one is very crucial – the cognitive discounting factor  $\bar{m}$ . The other four parameters ( $m_Y, m_r, m_x^f, m_\pi^f$ ) are

<sup>25</sup>The online appendix (Section 11.9) works out a slight variant, where debt mean-reverts to a fixed constant. The economics is quite similar.

<sup>26</sup>The model in Galí et al. (2007) is richer and more complex, as it features heterogeneous agents. Omitting the monetary policy terms, instead of  $x_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} M^{\tau-t} b_d d_\tau \right]$ , as obtained by integrating (35) forward, they generate  $\hat{c}_t = \Theta_n n_t - \Theta_t t_t^r$ . Here  $n_t$  is the deviation of employment from its steady state,  $t_t^r$  is the log-deviation from steady state of the taxes levied on a fraction of agents who are hand-to-mouth, and  $\Theta_n, \Theta_t$  are positive constants. Hence, one key difference is that in the present model, future deficits matter as well, whereas in their model, they do not.

<sup>27</sup>Mankiw and Weinzierl (2011) have a form of the representative agent with a partial rule of thumb behavior. They derive an instructive optimal policy in a 3-period model with capital (which is different from the standard New Keynesian model), but do not analyze an infinite horizon economy. Another way to have non-Ricardian agents is via rational credit constraints, as in Kaplan et al. (2018). The analysis is then rich and complex.

not essential, and could be set to 1 (the rational value) in most cases. Still, I keep them here for two reasons: conceptually, I found it instructive to see where the intercept, rather than the slope of attention, matters. Also given these various “intercepts of attention” are conceptually natural, they are likely to be empirically relevant as well when future studies measure attention.

**Macroeconomic evidence on the model’s deviations from pure rationality.** Before showing new evidence later, let us review the extant evidence for the effects described in the model. It appears to support the main deviations of the model from pure rationality.

*In the Phillips curve, firms do not appear to be fully forward looking:  $M^f < 1$ .* Empirically, the Phillips curve is not fully forward looking. For instance, Galí and Gertler (1999) find that we need  $\beta M^f \simeq 0.8$  at the quarterly frequency; given that  $\beta \simeq 1$ , that leads to an attention parameter of  $M^f \simeq 0.8$ . If  $m_\pi^f = 1$  and  $\theta = 0.7$ , this corresponds to  $\bar{m} \simeq 0.85$ . This is the value I will take in the numerical examples in section 2.6.

*In the Euler equation consumers do not appear to be fully forward looking:  $M < 1$ .* Fuhrer and Rudebusch (2004) estimate an IS curve and find  $M \simeq 0.6$ . The literature on the forward guidance puzzle concludes, qualitatively, that  $M < 1$ .

*Ricardian equivalence does not fully hold.* There is a lot of debate about Ricardian equivalence. The provisional median opinion is that it only partly holds. For instance, the literature on tax rebates (see Johnson et al. (2006)) appears to support  $b_d > 0$ .

Indeed, all three facts come out naturally from a model with cognitive discounting, i.e.  $\bar{m} < 1$ .

I next discuss the behavioral micro assumption of the model.

## 2.5 Discussion of the Behavioral Assumptions

Here I discuss the behavioral assumptions, especially the key Assumption 2.1.

**Microeconomic evidence** There is mounting microeconomic evidence for the existence of inattention to small dimensions of reality (Brown et al. (2010), Caplin et al. (2011), Gabaix and Laibson (2006), Gabaix (2017)) including taxes (Chetty et al. (2009), Taubinsky and Rees-Jones (2017)), and macroeconomic variables (Coibion and Gorodnichenko (2015a)). It is represented in a compact way by the inattention parameters – that is, the  $m$ ’s.<sup>28</sup>

This paper highlights another potential effect that has not specifically been investigated: a “slope” of inattention captured by  $\bar{m}$ , whereby agents perceive more dimly things that are further in the future. There is indirect evidence for it, though. For instance, evidence in Coibion and Gorodnichenko (2015a) might be explained by cognitive discounting. The present agents do not fully perceive variable realizations happening at a future date  $\tau > t$  due to cognitive discounting (they dampen them by  $\bar{m}^{\tau-t}$ ). As time goes by ( $t$  increases), the dampening is less strong, and they incompletely revise their forecast (as in Lemma 2.2). Thus, the revision is correlated to the ex-post forecast error – as in Coibion and Gorodnichenko (2015a).<sup>29</sup> Gabaix and Laibson (2017) argue that a large fraction of the vast literature on hyperbolic discounting reflects a closely related form of cognitive discounting. The empirical evidence

<sup>28</sup>In this paper, the theme is that of underreaction. It is possible to generate overreaction: if people overestimate the autocorrelation of productivity or income shocks (because it’s higher in their default model), they will overreact to them. See Gabaix (2017), section 2.3.13. Still, the evidence mentioned in footnote 29 points to under rather than over reaction.

<sup>29</sup>This is developed in Section 11.6. At the same time, the Coibion and Gorodnichenko (2015a) data comes from professional forecasters, who are likely to be more rational than the consumers in this model.

in Section 5.2 supports  $\bar{m} < 1$ . Here, and elsewhere, this paper gives functional forms and predictions that can be estimated in future research.

**Theoretical microfoundations** Section 8 discusses in detail the microfoundations of the inattention parameters, and proposes an endogenization for them. Here I give a summary of the situation. First, pragmatically, my preferred interpretation is that the formulation (8) can be taken as a useful idealization of the agent’s simulation process. This is in line with much behavioral economics, in which a plausible description of the thought process is posited, and its consequences analyzed – but the research on its deeper microfoundations is left for the future (for instance, loss aversion is observed and modeled, but there is still no agreement about its “deep” microfoundations, so that loss aversion is directly used, rather than its more remotely speculative microfoundations; and likewise for hyperbolic discounting, fairness etc.). Adjusting for the different stakes, this is similar for, say, equilibrium. One starts with a notion of equilibrium (supply equal demand; or Nash equilibrium), but the hypothetical nanofoundations for how the market will reach equilibrium (e.g. tâtonnement) are typically done in separate studies, and not actively used when thinking about the consequences of equilibrium for concrete economic analyses.

Section 8.2 proposes such a potential nanofoundation: formulation (8) (and the extension (29)) can be viewed as the “representative agent” version of a model in which the agent performs a mental simulation of the future, but receives only noisy signals about that simulation. As the agent receives noisy signals, his posteriors are the signals times a coefficient that is less than 1 – this creates inattention. Indeed, this generates a perception that on average is (8), where  $\bar{m}$  indexes the precision of the agent’s mental simulation. Hence, the average agent will behave according to the policy presented in Proposition 2.7.<sup>30</sup>

Importantly, and quite independently of the details of the nanofoundation leading to (8), one can allow agents to optimize their attention (Section 8.1). Then, the optimal level of attention reacts to the incentives to pay attention that the agent faces. The framework presented in Section 8.1 formalizes this (drawing from Gabaix (2014)), in a way that applies to general utility functions and distributional assumptions (as opposed to the usual Gaussian-quadratic setup).

To summarize, we have theoretical microfoundations for cognitive discounting and related inattention: these can be generated by noisy perceptions, and the optimal degree of attention can be made to vary with incentives.

**Lucas critique** In most of this paper the attention parameters are taken to be constant. But for completeness Section 8.1 discusses their endogenization. Attention will not change if, for instance, the volatility of the environment increases by a small or moderate amount (the “sparsity” feature of the theoretical microfoundation is useful for that, as it makes the agent locally non-reactive to things), but it will rise if the volatility of the environment increases a lot (see in particular the discussion after Proposition 8.1). Hence, the Lucas critique does not apply for small or moderate changes, but it does apply to large changes.

---

<sup>30</sup>For the welfare part of this paper, it is expedient to have a model in which the representative agent construct holds exactly (it would be interesting to study welfare when agents differ because of the different noisy signals that they receive, but I leave it to future research). Under the first interpretation, this is automatic. Under the second interpretation, we can posit that the agent is really a continuous family of such agents, each of whom takes a decision on consumption and labor supply, so that the representative agent perspective holds.

**Long run learning** Relatedly, the agent has forever a biased model of the world (biased by cognitive discounting)—in that sense, she does not learn in the long run. This makes sense, as attention is costly. We do sail through life without learning many things – for example, most people lead happy lives without learning quantum mechanics. Quantum mechanics is difficult, and not crucial to leading a good life. Likewise, in this model, fully understanding interest rates is difficult, and not crucial for life. Learning and attention are effortful, and typically we do not learn all things within a human lifespan.

**New degrees of freedom** This model is quite parsimonious: there is one main non-standard parameter, the cognitive discounting parameter  $\bar{m}$ . The other attention parameters are much less important, and can be disciplined via measurement.<sup>31</sup> In other contexts such as tax salience (Taubinsky and Rees-Jones (2017)), attention parameters are being progressively better measured (see Gabaix (2017) for a survey), and one can hope that the same thing will happen for macro parameters of attention.

**Reasonable variants** Like any model, the framework admits a large number of reasonable variants. I have explored a number of them, and the economics I present here reflects what is robust in those variants.<sup>32</sup> The model here presents one such set of assumptions – essentially, I chose them by aiming at a happy balance between tractability, parsimony and psychological and macroeconomic realism.

## 2.6 Values used in the Numerical Examples

Table 1 summarizes the main sufficient statistics for the output of the model, summarized in Figures 1–5. These values can in turn be rationalized in terms of “ancillary” parameters shown in Table 2. I call these parameters “ancillary” because they matter only via their impact on the aforementioned sufficient statistics listed in Table 1. For instance, the value of  $\kappa = \left(\frac{1}{\theta} - 1\right) (1 - \beta\theta) (\gamma + \phi) m_x^f$  can come from many combinations of  $\theta, \gamma, \phi$  etc. Table 2 shows one such combination.

The values are broadly consistent with those of the New Keynesian literature, and the empirical work of Section 5. The inattention parameters are drawn to be close to the myopia found in Galí and Gertler (1999). The inattention to the output gap,  $m_x^f$ , is there to match a low slope of the Phillips curve,  $\kappa$ .

---

<sup>31</sup>There is “meta” degree of freedom – where do the  $m$ ’s go? I note that this “meta” problem is present in all of economics. For example, in information economics, it’s normally assumed that the agent knows almost all the world perfectly, and has imperfect information about just one or a few variables. Likewise, we introduce adjustment costs in just one of a few variables, not to all. The modeler chooses which those are – guided by a sense of “what is relevant and interesting”. I try to do the same here.

<sup>32</sup>For instance, the agent might extrapolate too much from present income: this gives a high MPC out of current income, but otherwise the macro behavior does not change much (see Section 11.3). She might also suffer from nominal illusion in her perception of the interest rate (Section 11.7). Also, if we had growth, the agent would cognitively discount the deviations  $\mathbf{X}_t$  from the balanced growth path (see Section 11.8). One could also imagine a number of variants, e.g. (8) and (10) replace  $\bar{m}$  by a diagonal matrix  $diag(\bar{m}_i)$  of component-specific cognitive discounting factors.

Table 1:  
KEY PARAMETER INPUTS

Cognitive discounting by consumers and firms	$M = 0.85, M^f = 0.79$
Sensitivity to interest rates	$\sigma = 0.20$
Slope of the Phillips curve	$\kappa = 0.053$
Rate of time preference	$\beta = 0.99$
Deviation from Ricardian equivalence	$b_d = 0.0048$
Relative welfare weight on output	$\vartheta = 0.05$

*Notes.* This table reports the coefficients used in the model. Units are quarterly.

Table 2:  
ANCILLARY PARAMETERS

Coefficient of risk aversion	$\gamma = 1$
Inverse of Frisch elasticity	$\phi = 1$
Survival rates of prices	$\theta = 0.7$
Demand elasticity	$\varepsilon = 5.3$

*Attention parameters*

Cognitive discounting	$\bar{m} = 0.85$
Consumer's attention to income and interest rates	$m_y = 1, m_r = 0.2$
Firms' attention to inflation and future marginal costs	$m_\pi^f = 1, m_x^f = 0.2$

*Notes.* This table reports the coefficients used in the model to generate the parameters of Table 1. Units are quarterly. The parameters in turn imply  $m_Y = 1$ .

### 3 Consequences of this Behavioral Model

#### 3.1 The Taylor Principle Reconsidered: Equilibria are Determinate Even with a Fixed Interest Rate

The traditional model suffers from the existence of a continuum of multiple equilibria when monetary policy is passive. We will now see that if consumers are boundedly rational enough, there is just one unique (bounded) equilibrium. As monetary policy is passive at the ZLB, this topic will have strong impacts for the economy at the ZLB.

I assume that the central bank sets the nominal interest  $i_t$  in a Taylor rule fashion:

$$i_t = \phi_\pi \pi_t + \phi_x x_t + j_t, \quad (42)$$

where  $j_t$  is typically just a constant.<sup>33</sup> Calculations show that the system of Proposition 2.10 can be

<sup>33</sup>The reader will want to keep in mind the case of a constant  $j_t = \bar{j}$ . Generally,  $j_t$  is a function  $j_t = j(\mathbf{X}_t)$  where  $\mathbf{X}_t$  is

represented as:<sup>34</sup>

$$\mathbf{z}_t = \mathbf{A}\mathbb{E}_t[\mathbf{z}_{t+1}] + \mathbf{b}a_t, \quad (43)$$

where

$$\mathbf{z}_t := (x_t, \pi_t)'$$

will be called the state vector<sup>35</sup>,  $a_t := j_t - r_t^n$  (as in “action”) is the the baseline tightness of monetary policy (if the government pursues the first best,  $a_t = 0$ ),  $\mathbf{b} = \frac{-\sigma}{1+\sigma(\phi_x+\kappa\phi_\pi)}(1, \kappa)'$  and

$$\mathbf{A} = \frac{1}{1 + \sigma(\phi_x + \kappa\phi_\pi)} \begin{pmatrix} M & \sigma(1 - \beta^f\phi_\pi) \\ \kappa M & \beta^f(1 + \sigma\phi_x) + \kappa\sigma \end{pmatrix}, \quad (44)$$

where I use the notation

$$\beta^f := \beta M^f. \quad (45)$$

For simplicity, I assume an inactive fiscal policy,  $d_t = 0$ .<sup>36</sup>

The next proposition generalizes the well-known Taylor determinacy criterion to behavioral agents. I assume that  $\phi_\pi$  and  $\phi_x$  are weakly positive (the proof indicates a more general criterion).

**Proposition 3.1** (Equilibrium determinacy with behavioral agents) *There is a determinate equilibrium (i.e., all of  $\mathbf{A}$ 's eigenvalues are less than 1 in modulus) if and only if:*

$$\phi_\pi + \frac{(1 - \beta M^f)}{\kappa}\phi_x + \frac{(1 - \beta M^f)(1 - M)}{\kappa\sigma} > 1. \quad (46)$$

In particular, when monetary policy is passive (i.e., when  $\phi_\pi = \phi_x = 0$ ), we have a determinate equilibrium if and only if bounded rationality is strong enough, in the sense that

$$\text{Strong enough bounded rationality condition: } \frac{(1 - \beta M^f)(1 - M)}{\kappa\sigma} > 1. \quad (47)$$

Condition (47) does not hold in the traditional model, where  $M = 1$ . The condition means that agents are boundedly rational enough (i.e.,  $M$  is sufficiently less than 1) and the firm-level pricing or cognitive frictions are large enough (i.e, respectively,  $\bar{\kappa}$  is low (a pricing friction),  $m_x^f$  is low (a cognitive friction), so that  $\kappa = \bar{\kappa}m_x^f$  is low).<sup>37</sup> Quantitatively, it is quite easy to satisfy this criterion.<sup>38</sup>

a vector of primitives that are not affected by  $(x_t, \pi_t)$ , e.g. the natural rate of interest coming from stochastic preferences and technology.

<sup>34</sup>It is easier (especially for higher-dimensional variants) to proceed with the matrix  $\mathbf{B} := \mathbf{A}^{-1}$ , write the system as  $\mathbb{E}_t[\mathbf{z}_{t+1}] = \mathbf{B}\mathbf{z}_t + \tilde{\mathbf{b}}a_t$ , and to reason on the eigenvalues of  $\mathbf{B}$ :

$$\mathbf{B} = \frac{1}{M\beta^f} \begin{pmatrix} \beta^f(1 + \sigma\phi_x) + \kappa\sigma & -\sigma(1 - \beta^f\phi_\pi) \\ -\kappa M & M \end{pmatrix}.$$

<sup>35</sup>I call  $\mathbf{z}_t$  the “state vector” with some mild abuse of language. It is an outcome of the deeper state vector  $\mathbf{X}_t$ .

<sup>36</sup>Given a rule for fiscal policy, the sufficient statistic is the behavior of the “monetary and fiscal policy mix”  $i_t - \frac{b_d d_t}{\sigma}$ .

<sup>37</sup>Recall also that  $\kappa = \bar{\kappa}m_x^f$ . So, greater bounded rationality by firms (lower  $m_x^f$ ) helps achieving unicity. As the frequency of price changes becomes infinite,  $\kappa \rightarrow 0$  (see equation (117)). So to maintain determinacy (and more generally, insensitivity to the very long run), we need both enough bounded rationality and enough price stickiness, in concordance with Kocherlakota (2016)'s finding that we need enough price stickiness to have sensible predictions in long-horizon models.

<sup>38</sup>Call  $g(M) = \frac{(1-M)(1-\beta M)}{\kappa\sigma} - 1$  (to simplify this discussion, I take  $M = M^f$ ). The behavioral Taylor criterion (46) is

Why does bounded rationality eliminate multiple equilibria? This is because boundedly rational agents are less reactive to the future, hence less reactive to future agents' decisions. Therefore, bounded rationality lowers the complementarity between agents' actions (their consumptions). That force dampens the possibility of multiple equilibria.<sup>39,40</sup>

Condition (47) implies that the two eigenvalues of  $\mathbf{A}$  are less than 1. This implies that the equilibrium is determinate.<sup>41</sup> This is different from the traditional NK model, in which there is a continuum of non-explosive monetary equilibria, given that one root is greater than 1 (as condition (47) is violated in the traditional model).<sup>42</sup>

This absence of multiple equilibria is important, in particular when the central bank keeps an interest peg (e.g. at 0% because of the ZLB).

*Permanent interest rate peg.* First, take the (admittedly extreme) case of a permanent peg. Then, in the traditional model, there is always a continuum of bounded equilibria, technically, because matrix  $\mathbf{A}$  has a root greater than 1 (in modulus) when  $M = 1$ . As a result, there is no definite answer to the question “What happens if the central bank raises the interest rate?” – as one needs to select a particular equilibrium. In this paper's behavioral model, however, we do get a definite non-explosive equilibrium. In this behavioral model, we can simply write:

$$\mathbf{z}_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} \mathbf{A}^{\tau-t} \mathbf{b} a_\tau \right]. \quad (48)$$

Cochrane (2017) made the point that we'd expect an economy such as Japan's to be quite volatile, if the ZLB is expected to last forever: conceivably, the economy could jump from one equilibrium to the next at each period. This is a problem for the rational model, which is solved if agents are behavioral enough (i.e., if (47) holds).

*Long-lasting interest rate peg.* Second, the economy is still very volatile (in the rational model) in the less extreme case of a peg lasting for a long but finite duration. To see this, suppose that the ZLB is expected to last for  $T$  periods. Call  $\mathbf{A}_{ZLB}$  the value of matrix  $\mathbf{A}$  in (44) when  $\phi_\pi = \phi_x = j = 0$  in the Taylor rule. Then, the system (43) is, at the ZLB ( $t \leq T$ ):  $\mathbf{z}_t = \mathbb{E}_t \mathbf{A}_{ZLB} \mathbf{z}_{t+1} + \underline{\mathbf{b}}$  with  $\underline{\mathbf{b}} := (1, \kappa) \sigma \underline{r}$ , where  $\underline{r} < 0$  is the real interest rate that prevails during the ZLB. Iterating forward, we have:

$$\mathbf{z}_0(T) = (\mathbf{I} + \mathbf{A}_{ZLB} + \dots + \mathbf{A}_{ZLB}^{T-1}) \underline{\mathbf{b}} + \mathbf{A}_{ZLB}^T \mathbb{E}_0 [\mathbf{z}_T]. \quad (49)$$

Here I note  $\mathbf{z}_0(T)$ , the value of the state at time 0, given the ZLB will last for  $T$  periods. Let us focus on the last term,  $\mathbf{A}_{ZLB}^T \mathbb{E}_0 [\mathbf{z}_T]$ . In the traditional case, one of the eigenvalues of  $\mathbf{A}_{ZLB}$  is greater than 1 in modulus. This implies that very small changes to today's expectations of economic conditions after

---

that  $g(M) > 0$ , i.e.  $M < M^*$  where  $g(M^*) = 0$ . Using the calibration, this is the case if and only if  $M^* \simeq 0.90$ . If we divide  $\kappa\sigma$  by 10 (which is not difficult, given the small values of  $\kappa$  and  $\sigma$  often estimated) we get  $M^* \simeq 0.97$ .

<sup>39</sup>This theme that bounded rationality reduces the scope for multiple equilibria is general, and also holds in simple static models. I plan to develop it separately.

<sup>40</sup>One could also introduce nominal illusion as consumers perceiving the inflation to be  $\pi^{BR}(\mathbf{X}_t) = m_\pi^c \pi(\mathbf{X}_t)$ . In the IS curve (37), that will lead to replacing  $\mathbb{E}_t \pi_{t+1}$  by  $m_\pi^c \mathbb{E}_t \pi_{t+1}$ . More surprisingly, the Taylor criterion is modified by replacing, in the right-hand side of 46, the 1 by  $m_\pi^c$  (see Section 11.7). Again, bounded rationality makes the Taylor criterion easier to satisfy.

<sup>41</sup>The condition does not prevent unbounded or explosive equilibria, the kind that Cochrane (2011) analyzes. My take is that this issue is interesting (as are rational bubbles in general), but that the main practical problem is to eliminate bounded equilibria. The present behavioral model does that well.

<sup>42</sup>Of course, the selected equilibrium depends implicitly on the “default” model in the agents (which is a close cousin of the “prior” of Bayesian models). I tried to discipline them by adopting the long run value of variables,  $\bar{r}$ ,  $\bar{y}$ ,  $\bar{\mathbf{X}} = 0$ .

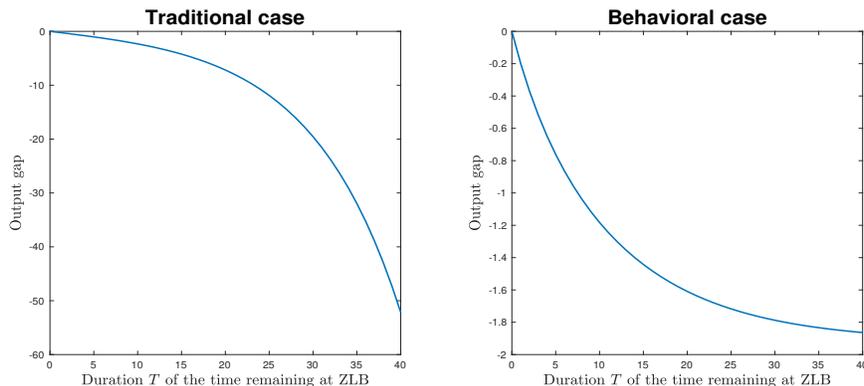


Figure 1: This figure shows the output gap  $x_0(T)$  at time 0, given that the economy will be at the ZLB for  $T$  more periods. The left panel is the traditional New Keynesian model, the right panel the present behavioral model. Parameters are the same in both models, except for the attention parameters  $M$ ,  $M^f$  which are equal to 1 in the rational model. The natural rate at the ZLB is  $-1\%$ . Output gap units are percentage point. Time units are quarters.

the ZLB (i.e., to  $\mathbb{E}_0[z_T]$ ), have an unboundedly large impact today ( $\lim_{T \rightarrow \infty} \|\mathbf{A}_{ZLB}^T\| = \infty$ ). Hence, we would expect the economy to be very volatile today, provided the ZLB period is long though finite, and a reasonable amount of fluctuating uncertainty about future policy.

### 3.2 The ZLB is Less Costly with Behavioral Agents

What happens when economies are at the ZLB? The rational model makes very stark predictions, which this behavioral model overturns. To see this, I follow the thought experiment in Werning (2012) (building on Eggertsson and Woodford (2003)), but with behavioral agents. I take  $r_t^n = \underline{r}$  for  $t \leq T$ , and  $r_t^n = \bar{r}$  for  $t > T$ , with  $\underline{r} < 0 \leq \bar{r}$ . I assume that for  $t > T$ , the central bank implements  $x_t = \pi_t = 0$  by setting  $i_t = \bar{r} + \phi_\pi \pi_t$  with  $\phi_\pi > 1$ , so that in equilibrium  $i_t = \bar{r}$ . At time  $t < T$ , I suppose that the CB is at the ZLB, so that  $i_t = 0$ .

**Proposition 3.2** *Call  $x_0(T)$  the output gap at time 0, given the ZLB will lasts for  $T$  periods. In the traditional rational case, we obtain an unboundedly intense recession as the length of the ZLB increases:  $\lim_{T \rightarrow \infty} x_0(T) = -\infty$ . This also holds when myopia is mild, i.e. (47) fails. However, suppose cognitive myopia is strong enough, i.e. (47) holds. Then, we obtain a boundedly intense recession:*

$$\lim_{T \rightarrow \infty} x_0(T) = \frac{\sigma(1 - \beta M^f)}{(1 - M)(1 - \beta M^f) - \kappa \sigma} \underline{r} < 0. \quad (50)$$

We see how impactful myopia can be. We see that myopia has to be stronger when agents are highly sensitive to the interest rate (high  $\sigma$ ) and price flexibility is high (high  $\kappa$ ). High price flexibility makes the system very reactive, and a high myopia is useful to counterbalance that.<sup>43</sup>

Figure 1 shows the result. The left panel shows the traditional model, the right one, the behavioral model. The parameters are the same in both models, except that attention is lower in the behavioral model. In the left panel, we see how costly the ZLB is: mathematically it is unboundedly costly as it

<sup>43</sup>The “paradox of flexibility” still holds though in a dampened way: if prices are more flexible,  $\kappa$  is higher (Proposition 2.5), and the higher disinflation worsens the recession at the ZLB (Eggertsson and Krugman (2012), Werning (2012)). However, bounded rationality moderates this, by lowering  $\kappa$  and  $M$  in (50).

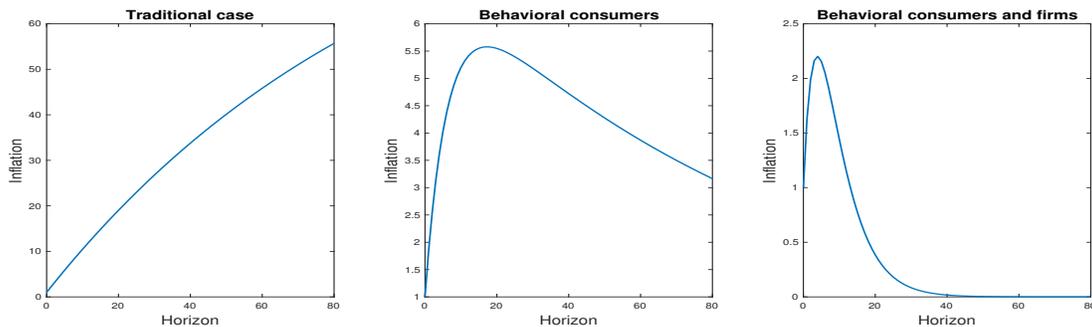


Figure 2: This Figure shows the response of current inflation to forward guidance about a one-period interest rate cut in  $T$  quarters, compared to an immediate rate change of the same magnitude. Left panel: traditional New Keynesian model. Middle panel: model with behavioral consumers and rational firms. Right panel: model with behavioral consumers and firms. Parameters are the same in both models, except for the attention parameters  $M$ ,  $M^f$  which are equal to 1 in the rational model.

becomes more long-lasting, displaying an exponentially bad recession as the ZLB is more long-lasting. In contrast, in this behavioral model, in the right panel we see a finite, though prolonged cost. Reality looks more like the mild slump of the behavioral model (right panel) – something like Japan since the 1990s – rather than the frightful abyss of the rational model (left panel), which is something like Japan in 1946. This sort of effect could be useful to empirically show that (47) likely holds.<sup>44</sup>

### 3.3 Forward Guidance Is Much Less Powerful

Suppose that the central bank announces at time 0 that in  $T$  periods it will perform a one-period, 1 percent real interest rate cut. What is the impact on today’s inflation? This is the thought experiment analyzed by McKay et al. (2016) with rational agents, which I pursue here with behavioral agents.

Figure 2 illustrates the effect. In the left panel, the whole economy is rational. We see that the further away the policy, the bigger the impact today – this is quite surprising, hence the term “forward guidance puzzle”. In the middle panel, consumers are behavioral but firms are rational, while in the right panel both consumers and firms are behavioral. We see that indeed, announcements about very distant policy changes have vanishingly small effects with behavioral agents – but they have the biggest effect with rational<sup>45,46</sup> agents.

<sup>44</sup>In this case, the economy is better off if agents are not too rational. This quite radical change of behavior is likely to hold in other contexts. For instance, in those studied by Kocherlakota (2016) where the very long run matters a great deal, it is likely that a modicum of bounded rationality would change the behavior of the economy considerably.

<sup>45</sup>Formally, we have  $x_t = Mx_{t+1} - \sigma\hat{r}_t$ , with  $\hat{r}_T = -\delta = -1\%$  and  $\hat{r}_t = 0$  if  $t \neq T$ . So  $x_t = \sigma M^{T-t}\delta$  for  $t \leq T$  and  $x_t = 0$  for  $t > T$ . This implies that time-0 response of inflation to a one-period interest-rate cut  $T$  periods into the future is:

$$\pi_0(T) = \kappa \sum_{t \geq 0} (\beta^f)^t x_t = \kappa \sigma \sum_{t=0}^T (\beta^f)^t M^{T-t} \delta = \kappa \sigma \frac{M^{T+1} - (\beta^f)^{T+1}}{M - \beta^f} \delta.$$

A rate cut in the very distant future has a powerful impact on today’s inflation ( $\lim_{T \rightarrow \infty} \pi_0(T) = \frac{\kappa \sigma}{1 - \beta^f} \delta$ ) in the rational model ( $M = 1$ ), and no impact at all in the behavioral model ( $\lim_{T \rightarrow \infty} \pi_0(T) = 0$  if  $M < 1$ )

<sup>46</sup>When attention is endogenous, the analysis could become more subtle. Indeed, if other agents are more attentive to the forward announcement by the Fed, their impact will be bigger, and a consumer will want to be more attentive to it. This positive complementarity in attention could create multiple equilibria in effective attention  $M, m_r$ . I do not pursue that here.

## 4 Optimal Monetary and Fiscal Policy

### 4.1 Welfare with Behavioral Agents and the Central Bank's Objective

Optimal policy needs a welfare criterion. Welfare here is the expected utility of the representative agent,  $\widetilde{W} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t, N_t)$ , under the objective expectation. This is as in much of behavioral economics, which views behavioral agents as using heuristics, but experience utility from consumption and leisure like rational agents.<sup>47</sup> I express  $\widetilde{W} = W^* + W$ , where  $W^*$  is first best welfare, and  $W$  is the deviation from the first best. The following lemma derives it.

**Lemma 4.1** (Welfare) *The welfare loss from inflation and output gap is*

$$W = -K \mathbb{E}_0 \sum_{t=0}^{\infty} \frac{1}{2} \beta^t (\pi_t^2 + \vartheta x_t^2) + W_-, \quad (51)$$

where  $\vartheta = \frac{\bar{\kappa}}{\varepsilon} = \frac{\kappa}{m_x^f \varepsilon}$ ,  $K = u_{cc}(\gamma + \phi) \frac{\varepsilon}{\bar{\kappa}}$ , and  $W_-$  is a constant (made explicit in (202)),  $\bar{\kappa}$  is independent of bounded rationality,  $\kappa = m_x^f \bar{\kappa}$  is the Phillips curve coefficient,  $\varepsilon$  is the elasticity of demand, and  $m_x^f \in (0, 1]$  is firms' attention to the output determinant of the markup.

Hence, the welfare losses are the same as in the rational model, when expressed in terms of deep parameters (including  $\bar{\kappa}$ ). However, when expressed in terms of the Phillips curve coefficient  $\kappa$ , the relative weight on the output gap ( $\vartheta$ ) is higher when firms are more behavioral (when  $m_x^f$  is lower). The traditional model gives a very small relative weight  $\vartheta$  on the output gap when it is calibrated from the Phillips curve – this is often considered a puzzle, which this lemma helps alleviate.

### 4.2 Optimal Policy: Response to Changes in the Natural Interest Rate

Suppose that there are productivity or discount factor shocks (the latter are not explicitly in the basic model, but can be introduced in a straightforward way). This changes the natural real interest rate,  $r_t^n$ . To find the policy ensuring the first best (i.e. 0 output gap and inflation), we inspect the two equations of this behavioral model (equations (37)-(38)). This reveals that the first best is achieved if and only if:  $i_t = r_t^n$ .

**Lemma 4.2** (First best) *When there are shocks to the natural rate of interest, the first best is achieved if and only if:*

$$i_t = r_t^n \equiv r_t^{n0} + \frac{b_d}{\sigma} d_t, \quad (52)$$

where  $r_t^{n0}$  is the “pure” natural rate of interest given in (17) and is independent of fiscal and monetary policy.

So, if the economy has a lower pure natural interest rate  $r_t^{n0}$  (hence “needs loosening”), the government can either decrease rates, or increase deficits. Monetary and fiscal policy are substitutes.<sup>48</sup>

<sup>47</sup>In particular I use the objective (not subjective) expectations. Also, I do not include thinking costs in the welfare metric. One reason is that thinking costs are very hard to measure (revealed preference arguments apply only if attention is exactly optimally set, something which is controversial). In the terminology of Farhi and Gabaix (2017), we are in the “no attention in welfare” case.

<sup>48</sup>If there are budget deficits, the central bank must “lean against behavioral biases interacting with fiscal policy”. For instance, suppose that (for some reason) the government is sending cash transfers to the agents,  $d_t > 0$ . That creates a boom. Then, the optimal policy is to still enforce zero inflation and output gap by raising interest rates.

**When the ZLB Doesn't Bind: Monetary Policy Attains the First Best** Suppose that the ZLB doesn't bind ( $r_t^{n0} \geq 0$ ). Then, we can turn off fiscal policy ( $d_t = 0$ ). With rational and behavioral agents, the optimal policy is still to set  $i_t = r_t^{n0}$ , i.e. to make the nominal rate track the natural real rate:<sup>49,50</sup>. This is the traditional, optimistic message in monetary policy.

$$\text{First best away from the ZLB: } i_t = r_t^{n0} \text{ and zero deficit: } d_t = 0. \quad (53)$$

### **When the ZLB Binds: “Helicopter Drops of Money” / Fiscal transfers as an Optimal Cure**

When the natural rate becomes negative (and with low inflation), the optimal nominal interest rate is negative, which is by and large not possible.<sup>51</sup> That is the ZLB. The first best is not achievable in the traditional model and the second best policy is quite complex.<sup>52</sup> However, with behavioral agents, there is an easy first best policy:<sup>53</sup>

$$\text{First best at the ZLB: } i_t = 0 \text{ and deficit: } d_t = \frac{-\sigma}{b_d} r_t^{n0}, \quad (54)$$

i.e. fiscal policy runs deficits to stimulate demand. By “fiscal policy” I mean transfers (from the government to the agents) or equivalently “helicopter drops of money”, i.e. checks that the central bank might send (this gives some fiscal authority to the central bank).<sup>54</sup> This is again possible because agents are not Ricardian. In conclusion, behavioral considerations considerably change policy at the ZLB, and allow the achievement of the first best.<sup>55,56</sup>

## **4.3 Optimal Policy with Complex Tradeoffs: Reaction to a Cost-Push Shock**

The previous shocks (productivity and discount rate shocks) allowed monetary policy to attain the first best. I now consider a shock that doesn't allow the monetary policy to reach the first best, so that trade-offs can be examined. Following the tradition, I consider a “cost-push shock”, i.e. a disturbance

---

<sup>49</sup>If the inflation target were  $\bar{\pi}$ , the nominal rate would be real rate plus inflation target  $i_t = r_t^n + \bar{\pi}$ . Throughout I assume  $\bar{\pi} = 0$  for simplicity.

<sup>50</sup>Sections 4.2-4.3 give optimal policy on the equilibrium path. To ensure determinacy, one simply adds a Taylor rule around it: if the equilibrium path predict values  $i_t^*, \pi_t^*, x_t^*$ , the policy function is:  $i_t = i_t^* + \phi_x (x_t - x_t^*) + \phi_\pi (\pi_t - \pi_t^*)$  with coefficients  $\phi$  that satisfy the modified Taylor criterion (46).

<sup>51</sup>It does not seem possible to obtain very negative nominal rates, say -5%, for long, because stockpiling cash in a vault is then a viable alternative.

<sup>52</sup>The first best is not achievable, and second best policies are complex, as has been analyzed by a large number of authors, e.g. Eggertsson and Woodford (2003), Werning (2012) and Galí (2015, section 5.4).

<sup>53</sup>A variant is:  $i_t = \varepsilon$  and  $d_t = \frac{-\sigma}{b_d} (r_t^n - \varepsilon)$ , for some small  $\varepsilon > 0$ , to ensure the determinacy of the Taylor rule around the policy (see footnote 50), which requires the possibility of lowering rates out of the equilibrium path.

<sup>54</sup>The central bank could also rebate the “seigniorage check” to the taxpayers rather than the government, and write bigger checks at the ZLB, and smaller checks outside the ZLB.

<sup>55</sup>Models with rational agents and credit constraints might work similarly (e.g. Kaplan et al. (2018)). However, they will be much harder to analyze, in part because future policy will have a complex, non-linear effect on savings etc.

<sup>56</sup>In Section 11.10 I analyze a richer situation, and show that the possibility of future fiscal policy can have ex-ante benefits – it makes agents confident about the future, as they know that the government will not run out of tools.

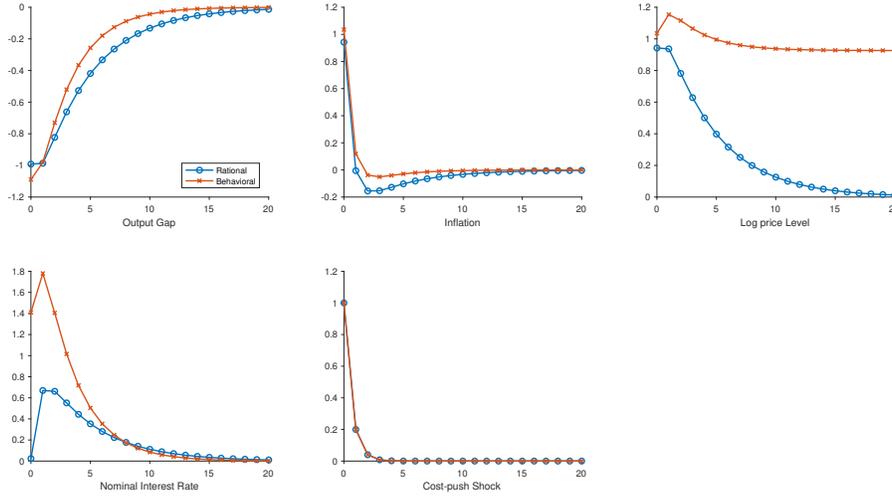


Figure 3: This figure shows the optimal interest rate policy in response to a cost-push shock ( $\nu_t$ ), when the central bank follows the optimal commitment strategy. When firms are rational, the optimal strategy entails “price level targeting”, i.e. the central bank will engineer a deflation later to come back to the initial price level. This is not the optimal policy with behavioral firms. This illustrates Proposition 4.3. Units are percentage points.

$\nu_t$  to the Phillips curve, which becomes:  $\pi_t = \beta M^f \mathbb{E}_t[\pi_{t+1}] + \kappa x_t + \nu_t$ , and  $\nu_t$  follows an AR(1):  $\nu_t = \rho_\nu \nu_{t-1} + \varepsilon_t^\nu$ .<sup>57,58</sup>

What is the optimal policy then? I examine the optimal policy first if the central bank can commit to actions in the future (the “commitment” policy), and then if it cannot commit (the “discretionary” policy).

**Proposition 4.3** (Optimal policy with commitment: suboptimality of price level targeting) *The optimal commitment policy entails*

$$\pi_t = \frac{-\vartheta}{\kappa} (x_t - M^f x_{t-1}), \quad (55)$$

so that the (log) price level ( $p_t = \sum_{\tau=0}^t \pi_\tau$ , normalizing the initial log price level to  $p_{-1} = 0$ ) satisfies

$$p_t = \frac{-\vartheta}{\kappa} \left( x_t + (1 - M^f) \sum_{\tau=0}^{t-1} x_\tau \right). \quad (56)$$

With rational firms ( $M^f = 1$ ), the optimal policy involves “price level targeting”: it ensures that the price level mean-reverts to a fixed target ( $p_t = \frac{-\vartheta}{\kappa} x_t \rightarrow 0$  in the long run). However, with behavioral firms, the price level goes up (even in the long run) after a positive cost-push shock: the optimal policy does not seek to bring the price level back to baseline.

“Price level targeting” and “nominal GDP targeting” are not optimal anymore when firms are behavioral. Price level targeting is optimal with rational firms, but not with behavioral firms. Qualitatively, the commitment to engineer a deflation later helps today, because firms are very forward looking (see Figure 3). That force is dampened in the present behavioral model. The recommendation of price level

<sup>57</sup> For instance, if firms’ optimal markup increases (perhaps because the elasticity of demand changes), they will want to increase prices and we obtain a positive  $\nu_t$  (see Clarida et al. (1999) and Galí (2015, Section 5.2)).

<sup>58</sup> Analyzing an early version of the present model, Bounader (2016) examined various constrained policies and derived independently some results in section 4.3, though not the key result on the non-optimality of price-level targeting.

targeting, one robust prediction of optimal policy model under the rational model, has been met with skepticism in the policy world – in part, perhaps, because its justification isn’t very intuitive.<sup>59</sup> This lack of intuitive justification may be caused by the fact that it is not robust to behavioral deviations, as Proposition 4.3 shows. Likewise, “nominal GDP targeting” is optimal in the traditional model, but is suboptimal with behavioral agents.<sup>60</sup> I next examine the optimal discretionary policy.

**Proposition 4.4** (Optimal discretionary policy) *The optimal discretionary policy entails:*

$$\pi_t = \frac{-\vartheta}{\kappa} x_t, \quad (57)$$

so that on the equilibrium path:  $i_t = K\nu_t + r_t^n$  with  $K = \frac{\kappa\sigma^{-1}(1-M\rho_\nu)+\vartheta\rho_\nu}{\kappa^2+\vartheta(1-\beta M^f\rho_\nu)}$ .

For persistent shocks ( $\rho_\nu > 0$ ), the optimal policy is less aggressive ( $K$  is lower) when firms are more behavioral (when  $M^f$  is lower, controlling the value of  $\kappa$ ,  $\sigma$ ,  $\vartheta$ ).<sup>61</sup> This is because with more myopic firms, future cost-push shocks do not affect much the firms’ pricing today, hence the central bank needs to respond less to them.

## 5 A Quantitative Exploration

Here I propose a quantitative version of the model. To do that, I present first an extension of the model that allows for non-zero trend inflation.

### 5.1 Enriching the Model with Changes to Trend Inflation

In the basic model so far, trend inflation is implicitly zero. Here I present a way to have a non-zero trend inflation. The reader eager to see empirical results may wish to jump to Proposition 5.2 and then to Section 5.2.

**Assumptions and basic model** The analytics will be very simple, but they require a bit of overhead.

*Default inflation.* When asked to forecast about inflation, firms may look at past inflation, as a number of papers have found, in particular Galí and Gertler (1999).<sup>62</sup> To capture this, I call “default inflation”,  $\pi_t^d$ , a signal about future inflation that firms form effortlessly. Conceptually, it might be an

---

<sup>59</sup>This is not an intuitive result even in the rational model: in the derivation, this is because the coefficient  $\beta$  in the Phillips curve and the rate of time preference for policy in (51) are the same—something that is not intuitive. That identity is broken in the behavioral model. This is analogous to the Slutsky symmetry in the rational model: there is no great intuition for its justification in the rational model; this is in part because it fails with behavioral agents (Gabaix 2014). Our intuitions are often (unwittingly) calibrated on our experience as living behavioral agents.

<sup>60</sup>Figure 3 gives some more intuition. Consider the behavior of the interest rate. The policy response is milder with rational firms than with behavioral firms. The reason is that monetary policy (especially forward guidance) is more potent with rational firms (they discount the future at  $\beta$ , not at the lower rate  $\beta M^f < \beta$ ), so the central bank can act more mildly to obtain the same effect. In addition, the gains from commitment are lower, as firms don’t react much to the future. At the same time, the optimal policy still features “history dependence” (in the terminology of Woodford (2003b)), even when the cost-push shock has no persistence: see equation (55).

<sup>61</sup>The analogue of Figure 3 for this no-commitment case is in Figure 7 of the online appendix.

<sup>62</sup>Galí and Gertler (1999) also present a model with partially backward looking firms, which this section extends. In the notations of this section, their model has  $\eta = 1$ ,  $M = 1$ , and  $\zeta = 0$ , which prevents the determinacy analysis below, where  $\zeta > 0$  is crucial.

optimal simple forecast based on past variables, e.g.<sup>63</sup>  $\pi_t^d = \tilde{\mathbb{E}}_t \left[ \pi_{t+1} | (\pi_\tau, \pi_\tau^{CB})_{\tau \leq t} \right]$  (as in Fuster et al. (2012)), where  $\tilde{\mathbb{E}}_t$  indicates that this need not be a completely optimal forecast. Here the variable  $\pi_t^{CB}$  is the target inflation announced by the central bank (e.g., 2%).

The specifics of default inflation will not matter much – all we need is *some* default inflation. For concreteness and the empirical implementation I will use the following functional form:<sup>64</sup>

$$\pi_t^d = (1 - \zeta) \bar{\pi}_t + \zeta \bar{\pi}_t^{CB}, \quad (58)$$

where  $\bar{\pi}_t$  and  $\bar{\pi}_t^{CB}$  are moving averages of past inflation and inflation guidance.<sup>65</sup> This means that default inflation puts a weight  $\zeta \in [0, 1]$  on the past central bank guidance,  $\pi_t^{CB}$ , and a weight  $1 - \zeta$  on past inflation.

*Indexation.* A number of authors have found that a form of automatic indexation is useful to fit the aggregate data, and not coincidentally, to hit conceptual targets such as long-run Fisher neutrality.<sup>66</sup> I will follow their lead, and assume here full indexation for firms not reoptimizing their prices (like Christiano et al. (2005), and Smets and Wouters (2007)). If a firm does not “actively” adjust its price in a forward looking manner like in the Calvo model, it just raises it by  $\pi_t^d$ .

Let us now call  $\hat{\pi}_t := \pi_t - \pi_t^d$  the deviation of inflation from the default. As the proof of the next Proposition spells out, then, we are in a world isomorphic to that of Section 2.2, except we replace  $\pi_t$  by  $\hat{\pi}_t$ .

**Proposition 5.1** (Behavioral New Keynesian model – augmented by a non-zero trend inflation) *In the extended model with non-zero trend inflation, we obtain the following behavioral version of the New Keynesian model. Decompose inflation as:  $\pi_t = \pi_t^d + \hat{\pi}_t$ , where  $\pi_t^d$  is default inflation, and  $\hat{\pi}_t$  is the deviation from default inflation. Then, we have*

$$x_t = M \mathbb{E}_t [x_{t+1}] - \sigma (i_t - \mathbb{E}_t \pi_{t+1} - r_t^n), \quad (59)$$

$$\hat{\pi}_t = \beta M^f \mathbb{E}_t [\hat{\pi}_{t+1}] + \kappa x_t. \quad (60)$$

We recover the same formulation as in the core model (Proposition 2.10); simply, in the Phillips curve (38), we replace inflation  $\pi_t$  by “deviation from default inflation”,  $\hat{\pi}_t$ , which gives (60).

This encompasses the basic behavioral model of Proposition 2.10, when default inflation is just 0, i.e.  $\pi_t^d \equiv 0$ , and  $\zeta = 1$ .

## 5.2 Quantitative Exploration

I now proceed to a quantitative exploration of the model. I perform a Bayesian estimation of the model’s parameters (Herbst and Schorfheide (2015)), which has some advantages over frequentist methods (Mavroeidis et al. (2014)): incorporating priors helps in identifying parameters that lie within

<sup>63</sup>The forecasted variable might be average future inflation,  $(1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^k \pi_{t+k}$ .

<sup>64</sup>Form (58) could come from optimal signal extraction – for instance, when inflation has low volatility,  $\zeta$  will naturally be bigger; but when inflation is variable and the central bank is not trusted, then  $\zeta$  will be low. That microfoundation would be easy to formalize, but I won’t pursue that here.

<sup>65</sup>This is,  $\bar{\pi}_t = (1 - \eta) \bar{\pi}_{t-1} + \eta \pi_{t-1}$  and  $\bar{\pi}_t^{CB} = (1 - \eta_{CB}) \bar{\pi}_{t-1}^{CB} + \eta_{CB} \pi_{t-1}^{CB}$ . This specific form does not really matter, e.g. it could have more lags.

<sup>66</sup>The traditional New Keynesian model is partially Fisher non-neutral: if inflation is permanently higher, then output is permanently higher: (2) gives  $x = \frac{1-\beta}{\kappa} \pi$ . Indexation restores full Fisher neutrality.

economically sensible domains, and the Bayesian framework provides a principled way of performing model comparisons.<sup>67,68</sup>

**Empirical models and data** The empirical model is the model of Proposition 2.10, with extra shocks  $\eta_t^k$  (with  $k \in \{d, s, m\}$ ) added:

$$x_t = M\mathbb{E}_t[x_{t+1}] - \sigma(i_t - \mathbb{E}_t\pi_{t+1}) + \eta_t^d, \quad (61)$$

$$\pi_t = \beta M^f \mathbb{E}_t[\pi_{t+1}] + \kappa x_t + \eta_t^s, \quad (62)$$

$$i_t = \rho_i i_{t-1} + (1 - \rho_i)(\phi_\pi \pi_t + \phi_x x_t) + \eta_t^m. \quad (63)$$

Equation (61) is the IS equation, with an additional AR(1) shock  $\eta_t^d$  that subsumes movements in the natural interest rate,<sup>69</sup> as well as potential model misspecification or measurement error. Equation (62) is the Phillips Curve, with  $\eta_t^s$  reflecting other disturbances, e.g. time-varying markups. Equation (63) is the Taylor rule, with an inertia coefficient  $\rho_i$ ,<sup>70</sup> and where  $\eta_t^m$  are shocks to monetary policy, e.g. coming from political pressure. Each of the shock processes follows an AR(1):  $\eta_t^k = \rho_k \eta_{t-1}^k + \varepsilon_t^k$  for  $k \in \{d, s, m\}$ , with  $\varepsilon_t^k$  i.i.d. Gaussians  $\mathcal{N}(0, \sigma_k^2)$ .<sup>71</sup>

I estimate the model using three observable time series: the unemployment rate, the CPI inflation rate, and the Federal Funds rate. Following e.g. Coibion and Gorodnichenko (2015b), I use the unemployment gap as a measure of the the output gap, i.e. set  $x_t = \bar{u} - u_t$ , where  $u_t$  is the unemployment rate, and  $\bar{u}$  its mean. In keeping with previous literature I detrend all three data series using the HP filter with smoothing parameter  $\lambda = 1600$ . The full sample period for estimation is 1960:Q1-2016:Q3.

The backwards looking New Keynesian model is the same as above, except (62) is replaced by

$$\pi_t = \beta M^f \mathbb{E}_t[\pi_{t+1}] + \pi_t^d - \beta M^f \pi_{t+1}^d + \kappa x_t + \eta_t^s \quad (64)$$

to implement (59), and the default inflation equation (58). Since all data are detrended, including the inflation rate, I set  $\pi_t^{CB} \equiv 0$ .

**Priors** Table 3 shows the priors. The prior for the attention parameters  $M$  and  $M^f$  follows a beta distribution with support in  $[0, 1]$ , consistent with their inattention interpretation. The prior also places a large amount of weight close to 1. For the model with a backwards looking Phillips curve, the parameter  $\eta$  reflects the notion that default inflation depends on past inflation, with a half-life of  $1/\eta$ . The low prior mean reflects a half life of around 2 years. The other priors are standard.<sup>72</sup>

<sup>67</sup>Bayesian models have drawbacks too, in particular assuming a Phillips curve via the prior that satisfies  $\kappa > 0$ .

<sup>68</sup>For a frequentist approach, Andrade et al. (2018) estimate the present behavioral model via GMM, finding evidence for  $M, M^f$  less than 1.

<sup>69</sup>This is why the equation has  $i_t - \mathbb{E}_t\pi_{t+1}$  rather than  $i_t - \mathbb{E}_t\pi_{t+1} - r_t^n$ :  $r_t^n$  is absorbed by  $\eta_t^d$ .

<sup>70</sup>This is standard, and captures the inertia or gradualism in empirical monetary policy.

<sup>71</sup>The shock  $\eta_t^m$  conveys that interest rates are also guided by other considerations than inflation and output gap, e.g. financial stability or political pressure.

<sup>72</sup>The priors for  $\sigma$  and  $\kappa$  are normal distributions truncated at zero to prevent negative values. The priors for  $\phi_x$  and  $\phi_\pi$  are normally distributed with mean values that satisfy the Taylor principle in a rational New Keynesian model, and which are typical of calibrated models in the New Keynesian literature. The small prior standard deviations reflect reasonably tight priors, which aids with model determinacy when estimating on the full sample. The priors for the standard deviations and persistence of the structural shocks are very standard. Note, however, that the persistence of the cost push shock,  $\rho_s$ , is set to zero (as in Galí and Gertler (1999)), for the following reason. In the model with the backwards looking Phillips curve there are lagged terms in inflation. Since there is very little persistence in inflation in the latter part of the sample,

Parameter	Distribution	Mean	Std. Dev.
$M$	Beta	0.85	0.10
$M^f$	Beta	0.85	0.10
$\sigma$	Normal	0.20	0.05
$\kappa$	Normal	0.15	0.10
$\phi_\pi$	Normal	1.50	0.15
$\phi_x$	Normal	0.50	0.05
$\sigma_d$	Inv. Gamma	0.30	1.00
$\sigma_s$	Inv. Gamma	0.30	1.00
$\sigma_m$	Inv. Gamma	0.30	1.00
$\rho_d$	Beta	0.50	0.20
$\rho_m$	Beta	0.50	0.20
$\rho_i$	Beta	0.50	0.20

Table 3: Prior specification

*Notes.* This table reports the prior distributions associated with Bayesian estimation of the model. The augmented model has two more parameters:  $\eta$  and  $\zeta$ . The priors on those are Beta distributions, with (mean, S.D.) equal to (0.2, 0.1) and (0.7, 0.1).

**Bayesian criterion** Table 4 reports the results of the estimation of several models. Column (1) reports the estimation for the rational model ( $M = M^f = 1$ ). Column (2) is the behavioral model where the attention parameters are restricted to be equal ( $M = M^f$ ) but may be less than one; column (3) reports the results for the model where  $M, M^f$  can be different. The remaining columns show the behavioral model with a backwards-looking Phillips Curve and attention parameters restricted to be equal (Column (4)) or not (Column (5)).

The first row reports the Bayes factor for each model  $i$  (plain behavioral, or behavioral augmented by backward looking terms) relative to the rational model. The exact Bayes factor is computed as the ratio of marginal data densities under each model, i.e.:

$$\mathcal{B}_i = \text{Bayes Factor for model } \mathcal{M}_i = \frac{\mathbb{P}(\text{Data}|\mathcal{M}_i)}{\mathbb{P}(\text{Data}|\mathcal{M}_{\text{rat}})}, \quad (65)$$

where the marginal data density for model  $i$  is just the integral of the likelihood with respect to the prior distribution for model  $i$ :  $\mathbb{P}(\text{Data}|\mathcal{M}_i) = \int_{\Theta_i} \mathbb{P}(\text{Data}|\Theta_i, \mathcal{M}_i) f(\Theta_i|\mathcal{M}_i) d\Theta_i$ . To get an intuitive feel for it, note that it has an interpretation in terms of the Bayesian Information Criterion (Kass and Raftery (1995)):

$$\ln \mathcal{B}_i \simeq \left( \hat{L}_i - \hat{L}_{\text{rat}} \right) - \frac{1}{2} n_i \ln T, \quad (66)$$

where  $\hat{L}_i$  is the log-likelihood of model  $i$ ,  $n_i$  is the number of parameters to estimate in model  $i$ , minus the number of parameters in the rational model, and  $T$  is the number of observations. Hence the log Bayes factor is essentially the difference in log likelihoods between behavioral model  $i$  and the rational model, minus a penalty for the extra number of free parameters  $n_i$  in the behavioral model  $i$ .

---

it is difficult to estimate persistent components in both the default inflation equation and the cost push shock. Thus, for comparability across models, cost push shocks are assumed to be i.i.d. in all model specifications.

Bayes Factor	Rational (1)			Behavioral with $M = M^f$ (2)			Behavioral (3)			Behavioral augmented with $M = M^f$ (4)			Behavioral augmented (5)		
	Mode	Mean	Std.	Mode	Mean	Std.	Mode	Mean	Std.	Mode	Mean	Std.	Mode	Mean	Std.
	1	1	0	0.732	0.642	0.105	0.736	0.669	0.110	0.733	0.637	0.113	0.740	0.668	0.105
$M$	1	1	0	–	–	–	0.916	0.808	0.109	–	–	–	0.916	0.803	0.124
$M^f$	0.050	0.065	0.041	0.011	0.050	0.033	0.013	0.050	0.036	0.006	0.048	0.036	0.005	0.047	0.033
$\sigma$	0.012	0.013	0.004	0.023	0.030	0.011	0.012	0.019	0.009	0.022	0.028	0.010	0.011	0.019	0.008
$\kappa$	1.022	1.010	0.141	1.085	1.057	0.145	1.083	1.056	0.136	1.088	1.061	0.141	1.089	1.059	0.135
$\phi_\pi$	0.458	0.463	0.043	0.451	0.456	0.042	0.452	0.455	0.041	0.451	0.456	0.042	0.451	0.456	0.042
$\eta$	–	–	–	–	–	–	–	–	–	–	–	–	0.609	0.607	0.065
$\zeta$	–	–	–	–	–	–	–	–	–	0.440	0.444	0.055	0.440	0.443	0.055
	0.058	0.061	0.008	0.102	0.132	0.029	0.101	0.125	0.030	0.101	0.132	0.031	0.099	0.125	0.029
$\sigma_d$	0.217	0.219	0.011	0.218	0.219	0.010	0.218	0.219	0.010	0.182	0.184	0.008	0.182	0.184	0.009
$\sigma_s$	0.182	0.184	0.009	0.182	0.184	0.009	0.182	0.184	0.009	0.182	0.184	0.009	0.182	0.184	0.008
$\sigma_m$	0.606	0.589	0.065	0.651	0.627	0.060	0.649	0.626	0.055	0.654	0.628	0.056	0.654	0.628	0.055
$\rho_i$	0.842	0.839	0.022	0.904	0.902	0.025	0.904	0.900	0.024	0.903	0.901	0.025	0.903	0.899	0.025
$\rho_d$	0.263	0.281	0.080	0.238	0.262	0.075	0.239	0.261	0.075	0.237	0.261	0.072	0.237	0.262	0.071
$\rho_m$															

Table 4: Estimation Results

*Notes.* This table reports the estimation results. The Bayes factor is in (65), and is approximately  $\ln \mathcal{B}_i = \hat{L}_i - \hat{L}_{\text{rat}} - \frac{1}{2}n_i \ln T$ , where  $\hat{L}_i$  is the log-likelihood of model  $i$ ,  $n_i$  is the number of parameters to estimate in model  $i$  minus the number of parameters in the rational model, and  $T$  is the number of observations. Hence the log Bayes factor is the difference in log likelihoods between the behavioral model “ $i$ ” and rational model “rat”, minus a penalty for the extra number of factors  $n_i$  in behavioral model  $i$ . The models are: the rational model (1); the basic behavioral NK model (forcing  $M = M^f$  in (2), or not, in (3)); and the augmented behavioral NK model with a backward looking component (with  $M = M^f$  in (4) or not, in (5)). *Sample period:* 1960:Q1-2016:Q3.

**Estimation results** We find that for each version of the behavioral model, the Bayes factor is considerably greater than one, suggesting the behavioral model is preferred to the rational model. For instance, it is over  $10^4$  in the main specification. Note that this is not simply a result of the additional free parameters estimated under the behavioral models (e.g.  $M, M^f$ ), as the Bayes factor in (66) contains a penalty for the number of free parameters. For instance, in the basic behavioral model with only 1 free parameter  $M = M^f$ ,  $n_i = 1$ . But when we allow for  $M, M^f$  to be estimated separately so that there are  $n_i = 2$  free extra parameters in the behavioral model (compared to the rational model), the Bayes factor falls (see the Bayes factor in Column (2) vs Column (3) of Table 4): the better in-sample fit is not enough to compensate for the loss in parsimony from having another parameter. The same thing happens between columns (4) and (5): the less parsimonious model has a lower Bayes factor.

The online appendix (Section 13) contains further results for the two subperiods, post-Volcker (1984:Q1-2016:Q3) and post-Volcker, pre-crisis (1984:Q1-2008:Q2). In this post-Volcker sample, the favored model is actually the basic model with  $M = M^f$  (Column (2)), over the augmented model and the rational model: in that post-Volcker sample, the backward looking feature does not seem useful – presumably because trend inflation was just stable. However, the backward looking extension is useful when studying the full sample (since 1960). This makes sense, as “trend inflation” (as captured by  $\pi_t^d$ ) was much more unstable then.

**Conclusion** The empirical analysis confirms the existence of partially myopic Euler equations and Phillips curves, and in order of magnitude it gives similar results to prior results by Galí and Gertler (1999) and Lindé (2005) for the Phillips curve, and Fuhrer and Rudebusch (2004) for the IS curve (still, it is useful to be able to estimate both equations together, in a coherent framework):<sup>73</sup>  $M, M^f \in [0.73, 0.92]$ , depending on the specifications, so roughly  $M, M^f \simeq 0.8$ .

For most purposes, I recommend the basic model of Proposition 2.10. However, when trend inflation is quite unstable, the extended one is useful.

### 5.3 Some consequences of the augmented model

I complete this study with some observations on the behavior of this model augmented with time-varying trend inflation.

**Fisher Neutrality holds** Fisher neutrality holds in the extended model of Proposition 5.1. Indeed, suppose indeed that long run inflation is  $\bar{\pi}$ , then the long run nominal rate is  $i = r^n + \bar{\pi}$ , and the economy is long-run Fisher neutral. Note that this is not the case in the basic model (which assumes that long run inflation is 0, see in (9)), and that the traditional NK model is only partially Fisher neutral (see footnote 66).

**Equilibrium determinacy revisited** Is the equilibrium determinate? The next Proposition generalizes the earlier criterion (46).<sup>74</sup>

---

<sup>73</sup>Hence we have evidence on the macro parameters of attention  $M, M^f$ . Gathering evidence on micro parameters  $\bar{m}$  would be much more costly. However, using micro data, Ganong and Noel (2017) find evidence for micro-level cognitive discounting, so that progress is being made in that direction too.

<sup>74</sup>This proposition states a necessary condition. The necessary and sufficient condition is (67) and a “Routh-Hurwitz auxiliary condition” stated in the online appendix (Proposition 12.1). This auxiliary condition is much more minor, and almost automatically valid in practice (see the discussion around Proposition 12.1).

**Proposition 5.2** (Equilibrium determinacy with behavioral agents – with backward looking terms) *In the extended model, the equilibrium is determinate only if:*

$$\phi_\pi + \zeta \frac{(1 - \beta M^f)}{\kappa} \phi_x + \zeta \frac{(1 - \beta M^f)(1 - M)}{\kappa \sigma} > 1. \quad (67)$$

Hence, we have a very similar criterion, except for the appearance of  $\zeta$ , the weight on the central bank guidance.

When monetary policy is passive ( $\phi_\pi = \phi_x = 0$ ), the economy can be determinate in this behavioral model if agents are behavioral enough (low  $M$ , low  $\kappa$  perhaps coming from low  $m_x^f$ ) and if their expectations are anchored enough, e.g. on central bank guidance (high  $\zeta$ ). However, when monetary policy is passive traditional models generate non-determinacy, as they violate the criterion (67). This is the case in the traditional New Keynesian model (which has  $M = M^f = 1$ ), in the indexation model of Galí and Gertler (1999) (which has  $\zeta = 0$ ,  $M = 1$ ,  $M^f \in [0, 1]$ ,  $\pi_t^d = \pi_{t-1}$ )<sup>75</sup>, and in the typical old Keynesian model (which has  $M^f = 0$ ,  $\zeta = 0$ ).<sup>76</sup>

Speculating somewhat more, this usefulness of “inflation guidance” may explain why central bankers in recent years did not wish to deviate from an inflation target of 2% (and go to a higher target, say 4%, which would leave more room to avoid the ZLB). They fear that “inflation expectations will become unanchored”, i.e. that  $\zeta$  will be lower: agents will believe the central bank less (as it “broke its word”), which in turn can make the economy equilibrium indeterminate (by (67)). This reasoning relies on agents’ bounded rationality.

**The 1970s** The stagflation of the 1970s has been attributed by Clarida et al. (2000) to a violation of the Taylor criterion – in essence,  $\phi_\pi < 1$ . But we have seen that Japan has arguably  $\phi_\pi = 0$ , and that this can still be consistent with a determinate equilibrium. How to reconcile these *prima facie* contradictory facts? In the present model, the 1970s can be interpreted as a moment where agents did not believe the central bank enough, i.e.  $\zeta$  is too low (in part because inflation was volatile, the central bank credibility was eroded) – while in Japan,  $\zeta$  is high enough. Together with the failure of the Taylor criterion documented in Clarida et al. (2000), this leads to criterion (67) to be violated.

The Bayesian analysis above offers some support for this conjecture. The weight on central bank’s guidance  $\zeta$  is estimated to be higher in the post-Volcker sample ( $\zeta = 0.78$ ) than in the sample since 1960 ( $\zeta = 0.44$ ) (see respectively Tables 5 and 4). This reflects quantitatively that agents put more weight on central banks announcements since 1984. It would be interesting to econometrically analyze this hypothesis more systematically (essentially, estimating a time-varying  $\zeta$ ), but this would take us too far afield here: methods such as the one in Section 5.2 do not easily allow for state-varying coefficients.

**Neo-Fisherian Experiment: A permanent shock to target inflation** I now consider a few impulse-responses for the system. I assume that the central bank announces at time 0 an immediate, permanent, unexpected rise of 1% in the nominal rate and of its corresponding target inflation ( $i_t = 1\%$  at all dates  $t \geq 0$ , and the central bank guidance is the corresponding long term target,  $\pi_t^{CB} = 1\%$ ).<sup>77</sup>

<sup>75</sup>With those values, the Phillips curve (59) is actually the limit of their Phillips curve when their  $\beta$  is 1.

<sup>76</sup>For instance, the old Keynesian model features a deflationary spiral, because it has  $\zeta = 0$ . However, we see that in the old Keynesian model, augmented with  $\zeta > 0$  (i.e., agents listen enough to the central bank when forming expectations), we can verify the criterion.

<sup>77</sup>To ensure determinacy, we can just add a Taylor rule around that the equilibrium path reported here, as in footnote 50.

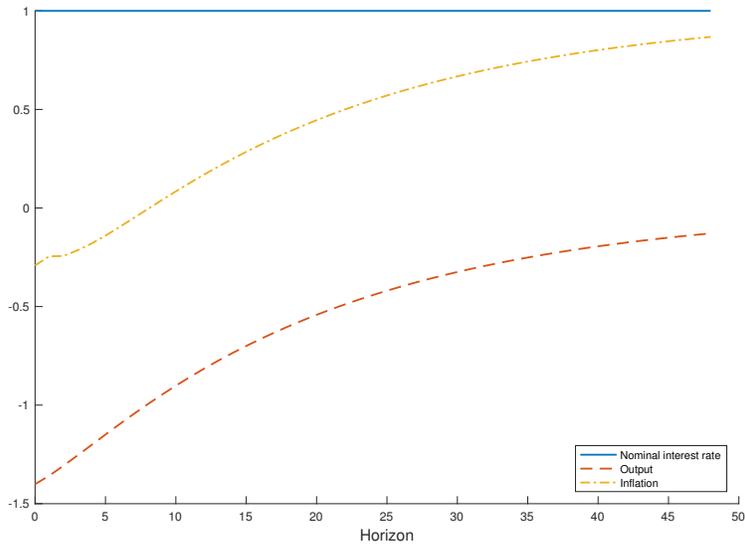


Figure 4: Impact of a permanent rise in the nominal interest rate. At time 0, the nominal interest rate is permanently increased by 1%. The Figure traces the impact on inflation and output. Units are percentage points.

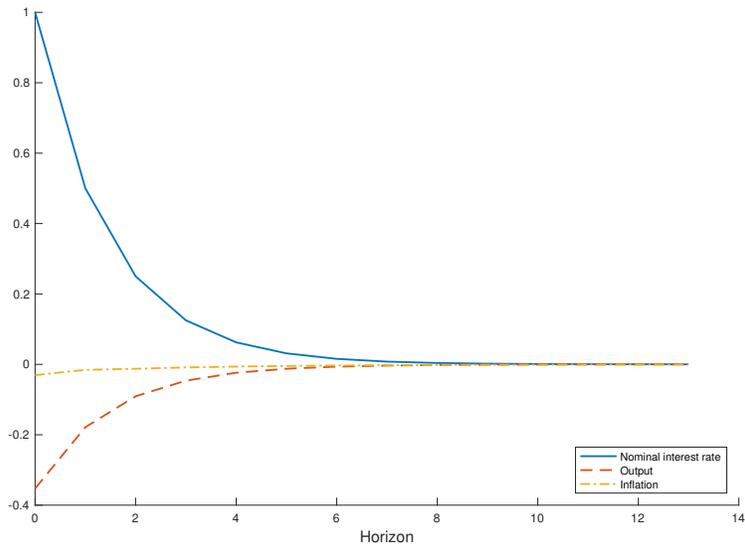


Figure 5: Impact of a temporary rise in the nominal interest rate. At time 0, the nominal interest rate is temporarily increased by 1%. The Figure traces the impact on inflation and output. Units are percentage points.

Figure 4 shows the result.<sup>78</sup> On impact, there is a recession: output and inflation are below trend. However, over time the default inflation increases: as the central bank gives “guidance”, inflation expectations are raised. In the long run, for this calibration, we obtain Fisher sign neutrality.

This effect is hard to obtain in a conventional New Keynesian model.<sup>79</sup> Cochrane (2017, p.3) summarizes the situation:<sup>80</sup>

“The natural starting place in this quest [for a negative short-run impact of interest rates on inflation] is the simple frictionless Fisherian model,  $i_t = r + \mathbb{E}_t \pi_{t+1}$ . A rise in interest rates  $i$  produces an immediate and permanent rise in expected inflation. In the search for a temporary negative sign [one can add] to this basic frictionless model: 1) new Keynesian pricing frictions, 2) backwards-looking Phillips curves, 3) monetary frictions. These ingredients robustly fail to produce the short-run negative sign.”

This paper gives a way to overturn this result, coming from agents’ bounded rationality. In this behavioral model, raising rates permanently first depresses output and inflation, then in the long run raises inflation (as Fisher neutrality approximately holds), via the credible inflation guidance.

This analysis, of course, is not an endorsement, as this policy is not first best, and leads to a prolonged recession.

**A temporary shock to the interest rate.** I now study a temporary increase of the nominal interest rate,  $i_t = i_0 e^{-\phi t} > 0$  for  $t \geq 0$ . As the long run is not modified, I assume an inflation guidance of 0,  $\pi_t^{CB} = 0$ . Figure 5 shows the result. On impact, inflation and output fall, and then mean-revert. The behavior is very close to what happens in the basic model of Section 2.

## 6 Related Work

The present paper benefits from many works that studied departures from full-knowledge and rational expectations. Before detailing them, let me summarize the comparative advantage of this paper. Given its tractability, the model offers a fairly unified and transparent way of reasoning about a host of issues that often require disconnected and less-tractable analyses in other models in the literature: for example, the cognitive discounting mechanism presented here applies to both the IS curve and the Phillips curve, and also extends to fiscal policy. Importantly, we can write the dynamics of the economy in the compact two-equation format familiar from the rational New Keynesian model – something that typically is not possible with other technologies (or only possible under very specific shock structures), which more commonly yield a discounted version of only one of the two New Keynesian equations (37)–(38), while leaving unchanged or much less tractable the other. As a result of this tractability, we can derive a number of results that characterize the interaction of bounded rationality with monetary and fiscal policy: these include a behavioral Taylor principle, deviations from Ricardian equivalence, the power of

<sup>78</sup>In addition to the basic calibration of Table 1, I use also parameters for default inflation:  $\eta = 0.5$ ,  $\eta_{CB} = 0.05$ ,  $\zeta = 0.7$ .

<sup>79</sup>For instance, in the traditional New Keynesian model a permanent change in the inflation target (i.e., of the intercept  $j_t$  of the Taylor rule) involves no transitional dynamics: it leads to an instantaneous jump of the whole economy to the new steady state, and this rise in interest rates leads to a one-for-one rise in inflation. Slow transition dynamics emerge only when departing from the basic model, e.g. by assuming imperfect information (Erceg and Levin (2003)) or sticky wages and/or indexation (Ascari and Ropele (2012)).

<sup>80</sup>Cochrane (2017) needs to select a particular equilibrium, which leads to some controversy.

forward guidance as a function of both consumers’ and firms’ bounded rationality, and the size of the recessions caused by adverse demand shocks at the ZLB. Moreover, the model’s tractability makes it easy to extend it in numerous directions, some of which are presented in the online appendix.

Let us now see the various strands to which this paper relates. First, it is a behavioral paper. There are many ways to model bounded rationality, of course, including limited information updating (Caballero (1995), Gabaix and Laibson (2002), Mankiw and Reis (2002), Reis (2006)), related differential salience (Bordalo et al. (2013)), and noisy signals (Sims (2003), Woodford (2012), Maćkowiak and Wiederholt (2015), Caplin et al. (2017)).<sup>81</sup> I use the “sparsity” approach developed progressively since Gabaix (2014), because it seems to capture a good deal of the psychology of attention (Gabaix (2017)), and it is particularly tractable. For instance, it allows us to give a behavioral version of consumer theory and of the Arrow-Debreu model.

Like much of behavioral economics, the model seeks to approximate human beings’ decisions, and looks also for general principles allowing to unify behavior. Inattention is such a candidate theme. It seeks to be constructively skeptical of the conventional candidate “first principles” (e.g. rational expectations, or Bayesian updating) which we inherited from the tradition, but are at best useful approximations of what people do, rather than inviolable true first principles of human nature.

On the macro theory front, this paper relates to the literature on “macroeconomics without the rational-expectations hypothesis” reviewed in Woodford (2013). My behavioral agent retains rational expectations dynamics, though dampened and discounted. In previous literature, a popular way to discipline beliefs under deviation from rational expectation is via learning, as reviewed in Eusepi and Preston (2018), and Evans and Honkapohja (2001). Agents in these models act as econometricians who update their forecast models as new data arrive. This literature focuses greatly on the local stability of the learning dynamics, which links tightly to whether the learning equilibrium converges to the rational expectations equilibrium. Behavioral agents in the present work, on the other hand, are behavioral in the sense that their mental model never converges to the rational expectations solution. This is also related to more recent work, such as that on level- $k$  thinking by García-Schmidt and Woodford (2015), Farhi and Werning (2017) and Woodford (2018). There, agents are rational with respect to partial equilibrium effects, but don’t quite understand general equilibrium effects. In both cases, the future is dampened. One difference in terms of predictions is that in the present framework agents’ cognitive myopia applies to both partial and general equilibrium effect.

The paper also relates to the literature on incomplete information and higher-order beliefs that has followed Morris and Shin (1998); see Angeletos and Lian (2016) for a review. This literature has long emphasized how lack of common knowledge and higher-order uncertainty can anchor expectations of endogenous outcomes, such as past inflation. Examples of such anchoring include Woodford (2003a), Nimark (2008), Angeletos and La’O (2010), and Morris et al. (2006). Building on this literature, Angeletos and Lian (2017b) have recently explored a version of the NK model that maintains the rational hypothesis but allows for higher-order uncertainty. This rationalizes a closely related form of discounting, and also illustrates the connection between the aforementioned literature and the approach taken in the present paper. This is because, in a general-equilibrium setting, higher-order uncertainty anchors forward-looking expectations in a manner that resembles cognitive discounting – though again, it is in general rather analytically complex with general shock structures. Angeletos and Lian (2017a) provides an abstract framework that embeds rational models (with incomplete information) and behavioral

---

<sup>81</sup>My notion of “behavioral” here is bounded rationality or cognitive myopia. I abstract from other interesting forces, like fairness (Eyster et al. (2017)) – they create an additional source of price stickiness.

models (without rational expectations).<sup>82</sup> One can hope that a healthy interplay between incomplete information models and behavioral models will continue.

Finally, it is linked to the literature around the “forward guidance puzzle” (Del Negro et al. (2015), McKay et al. (2016)). McKay et al. (2016) provide a microfoundation for an approximate version of the IS curve in (37) with  $M < 1$ , based on heterogeneous rational agents with limited risk sharing (see also Campbell et al. (2017)), without an analysis of deficits  $d_t$ . Their model has  $M^f = 1$ . The analysis of Werning (2015) yields a modified Euler equation with rational heterogeneous agents, which often yields  $M > 1$ . Del Negro et al. (2015) and Eggertsson and Mehrotra (2015) work out models with finitely-lived agents that give an  $M$  slightly below 1. Finite lives severely limit how myopic agents can be (e.g., the models predict an  $M$  very close to 1), given that life expectancies are quite high. Relatedly, Galí (2017) shows that the NK model with finite lives without any assets in positive net supply (government debt, capital, or bubbles) cannot generate discounting in the IS curve.<sup>83</sup> Caballero and Farhi (2017) offer a different explanation of the forward guidance puzzle in a model with a shortage of safe assets and endogenous risk premia. Relatedly, Fisher (2015) derives a discounted Euler equation with a safe asset premium: but the effect is very small; for example the coefficient  $1 - M$  is very close to 0 – close to the empirical “safety premium”, so at most  $M = 0.99$ . Models with hand-to-mouth consumers do generate an impact of deficits ( $b_d > 0$ ), but they also yield  $M = M^f = 1$ .<sup>84</sup>

## 7 Conclusion

This paper gives a simple way to think about the impact of bounded rationality on monetary and fiscal policy. Furthermore, we have seen that the model has good empirical support for its main non-standard elements. As shown in the prior literature and the empirical analysis of Section 5, the Phillips curve is partially myopic, so is the IS curve, and agents are partially non-Ricardian.

In conclusion, we have a theoretical model with empirical support for its non-standard features, that is also simple to use. This paper leads to a large number of natural questions.

*Theory.* I have studied only the most basic model. Doing a similar exploration for its richer variants would be very interesting and relevant both empirically and conceptually: e.g. capital accumulation, a more frictional labor market, distortionary taxes, agents that are heterogeneous in wealth or rationality. The tractable approach laid out in this paper makes the exploration of those questions quite accessible. Relatedly, it facilitates studying optimal central bank policy with behavioral agents under varied situations.<sup>85</sup>

*Empirics.* The present work suggests a host of questions for empirical work. One would like to estimate the intercept and slope of attention (i.e. attention to current variables, and how the under-

---

<sup>82</sup>In earlier work, Angeletos and La’O (2009) and Angeletos et al. (2017) explore how dropping the common-prior assumption can help mimic the dynamics of higher-order beliefs introduced by incomplete information, that generates a departure from traditional rational expectations.

<sup>83</sup>The Mankiw and Reis (2002) model changes the Phillips curve, which helps with some paradoxes (Kiley (2016)). But as it keeps the same IS curve as the traditional model, it still requires  $\phi_\pi > 1$  for determinacy – unlike the present behavioral model. A synthesis of Mankiw-Reis and the present model would be useful.

<sup>84</sup>Consider the case without fiscal policy. Suppose that a fraction  $f^h$  (resp.  $f^r = 1 - f^h$ ) consists of hand-to-mouth (resp. rational) agents who just consume their income  $c_t^h = y_t$ . Aggregate consumption is  $c_t = f^r c_t^r + f^h c_t^h$  and the resource constraint is  $y_t = c_t$ . But as  $c_t^h = y_t$ , this implies  $y_t = c_t^h = c_t^r$ . The hand-to-mouth consume exactly like rational agents. Hence, having hand-to-mouth agents changes *nothing* in the IS equation, and  $M = 1$ . With fiscal policy, however, those agents do make a difference, i.e. create something akin to  $b_d > 0$ , but still with  $M = 1$ .

<sup>85</sup>See Nakata et al. (2017) and Benchimol and Bounader (2018).

standing of future variables decreases with the horizon) using individual-level dynamics for consumers (equation (32)), for firms (equation (27)), and of the whole equilibrium economy (Proposition 2.10). One side-payoff of this work is to provide a parametrized model where these forces can be empirically assessed (i.e. measuring the various  $m$ 's in the economy).<sup>86</sup>

*Surveys.* It also suggests new questions for survey design. One would like to measure people's subjective model of the world – which, like that of this model's agents, may not be an accurate model of the world. For instance, one could design surveys about people's understanding of the impulse-response in the economy. They would ask questions such as: “Suppose that the central bank raises the interest rate now [or in a year, etc.], what do you think will happen in the economy? How will you change your consumption today?”. In contrast, most work assesses people's predictions of individual variables (e.g. Greenwood and Shleifer (2014)) rather than their whole causal model.<sup>87</sup> The parametrization in the present work allows for a way to explore potentially important deviations of the model from the rational benchmark, and suggests particular research designs that focus on the key differential predictions of a rational vs. a behavioral model.<sup>88</sup>

In conclusion, this paper offers a parsimonious way to think through the impact of bounded rationality on monetary and fiscal policy, both positively and normatively. It suggests a number of theoretical and empirical questions that would be fruitfully explored.

## 8 Appendix: Microfoundations for Cognitive discounting

There are three questions when handling a behavioral model of the type presented here.

1. How does the model generate in a coherent way the agent's consumption and labor supply policies (given attention  $m$ )? How does this affect economic outcomes?
2. How does the parameter  $m$  vary with incentives?
3. Is there a story for why we would achieve that formulation?

In my view, question 1 is the most important “practical” question – as it is crucial to handle demand functions. Accordingly, I detail it throughout the paper.

Question 2 is useful in some applications, and I detail it in Section 8.1. The upshot is this: in the “sparsity” framework, there is local rigidity and global flexibility. *Local rigidity*: when variances of primitive shocks remain within a certain bound<sup>89</sup>, then attention parameters remain exactly constant (there, “sparsity” is particularly useful). *Global flexibility*: However, if variances become very high, then attention does increase. The Lucas critique applies to large changes in the environment, but not to small ones. Hence, this paper will work with constant attention – that corresponds to variances that can vary moderately.

Question 3 is addressed in Section 8.2. For practical purposes, it is probably the least crucial. One perhaps useful example is the concept of “equilibrium prices”. The way over 99.9% of economics

---

<sup>86</sup>See Coibion and Gorodnichenko (2015a), Afrouzi (2017) and Fuhrer (2017) for progress on those issues.

<sup>87</sup>E.g. it asks questions like: “Are you optimistic about the economy today?” or “Where do you think the economy will be in a year?”. See Carvalho and Nechio (2014) for people's qualitative understanding of policy.

<sup>88</sup>E.g. one could ask “Suppose the central bank lowers interest rates by 1% [or the government gives \$1000 to all agents] for one period in eight quarters, what will happen to the rest of the economy, and to your decisions?”, plot the impulse response, vary the parameter “eight”, and compare that to the rational and behavioral models.

<sup>89</sup>For instance,  $\mathbb{E}[\zeta_t^2 + j_t^2] \leq K$ , for some bound  $K$ . Recall that  $j_t$  denotes innovations to monetary policy.

proceeds is to “assume that the market clears at price  $p$ ”, and solve for the price. There is small and worthy part of economics that thinks about microfoundations for the possibility of finding the equilibrium price (e.g. tâtonnement etc.) – that is a microfoundation that is useful to know, but you do not want to have to repeat it in each paper. On top of that, it is not clear that we have found the true microfoundations for market equilibrium.<sup>90</sup> Rather, those generate only something close, but not exactly equal to, the costless, instantaneous jump to market equilibrium. Hence, practicing economists are aware of those candidate microfoundations for “equilibrium prices”, but seldom actively use them. Still, it is healthy for concrete economics to have it as a benchmark in the background. Likewise, I develop a microfoundation in 8.2.

## 8.1 Endogenizing attention

**General formulation** The traditional New Keynesian model takes pricing frictions as given, and then studies their consequences. One can also endogenize the size of the pricing friction (of  $\theta$ , see Kiley (2000)), but most of the analysis is most cleanly done by taking the pricing friction as given. Likewise, in this paper I take the degree of inattention as given, and study its consequences. In this section I sketch how to endogenize it, drawing heavily on Gabaix (2014) and Gabaix (2016).

I show how to endogenize a parameter I call  $m$  – this could be  $\bar{m}$ , or the attention to the interest rate, or to inflation  $(m_r, m_\pi^f)$ , and so forth. Call  $a_t$  the action at time  $t$ , and  $\mathbf{S}_t$  the state vector. For instance, in the consumer’s problem,  $a_t = (c_t, N_t)$  (consumption and labor supply) and  $\mathbf{S}_t = (k_t, \mathbf{X}_t)$  ( $k_t$  is the agent’s personal wealth, which will be 0 in equilibrium in the model without taxes, and  $\mathbf{X}_t$  is vector of macro variables). The value at the default state is normalized to be  $\mathbf{S}^d = 0$ .

The agent has a subjective value function  $V(\mathbf{S}, m)$  that is the traditional, rational value function under the subjective model parameterized by  $m$ . So, at time  $t$ , the agent wishes to maximize

$$v(a_t, \mathbf{S}_t, m) = u(a) + \beta \mathbb{E}V(\mathbf{G}^{\mathbf{S}}(\mathbf{S}_t, a_t, m, \varepsilon_{t+1}), m), \quad (68)$$

where  $V(\mathbf{S}, m)$  is the subjective value function, i.e. the value function corresponding to the agent’s subjective model of the world – parameterized by  $m$ , i.e. the one with the transition function  $\mathbf{G}(\mathbf{S}_t, a_t, m, \varepsilon_{t+1})$ . She takes the action

$$a(m, \mathbf{S}_t) = \underset{a}{\operatorname{argmax}} v(a, \mathbf{S}_t, m). \quad (69)$$

Conceptually, the agent would like to maximize true utility, given the imperfect action  $a(m, \mathbf{S})$ , net of costs  $\mathcal{K}g(m - m^d)$  :

$$\max_m \mathbb{E}v(a(m, \mathbf{S}_t), \mathbf{S}_t, 1) - \mathcal{K}g(m - m^d), \quad (70)$$

i.e. the inclusive utility is evaluated under the true model (indexed by  $m = 1$ ), and the agent wants to avoid paying the thinking costs  $\mathcal{K}g(m - m^d)$ , with  $\mathcal{K} \geq 0$  (with  $\mathcal{K} = 0$  being the rational case). Here  $m^d$  is a “default” attention, processed for free by the agent. Typically in this paper,  $m^d > 0$ : we have “for free” some understanding of the future. The thinking cost  $g(\cdot)$  can be traced back in turn to the more primitive simulation technology used by the agent in Section 8.2. A typical functional form is the

---

<sup>90</sup>The same holds for related concepts, e.g. “Nash equilibrium” and “rational expectations equilibrium”. There are “infinitely iterated rounds of reasoning” stories to microfound those, and they generate something like those equilibrium concepts only under very strong and idealized conditions.

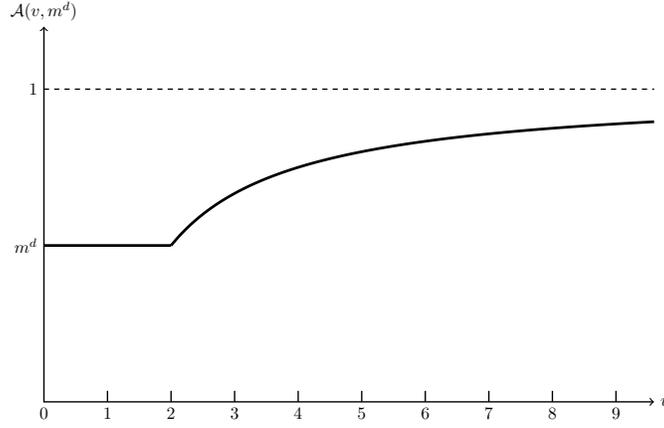


Figure 6: This Figure plots the attention function (73). When the volatility  $v$  of the environment is small or moderate, attention is at its default value  $m^d$ . However, when volatility increases a lot, attention increases, asymptotically toward 1.

parametrization  $g(m - m^d) = |m - m^d|$ , which gives an attention function that is always continuous, and constant in part of its domain (Gabaix (2014)).

The problem in (70) is typically intractable (both for the researcher and, presumably, for the agent), so some alternative formulation is needed. The sparse max model in Gabaix (2014) proposes a way to address that difficulty. There, the agent solves a linear-quadratic approximation of this problem (70), taking a Taylor expansion of the utility losses when evaluating optimal attention – but keeping the true nonlinear utility when taking her action, as in (69).<sup>91</sup> This means that (70) is replaced by:<sup>92</sup>

$$\max_m -\frac{1}{2}\Lambda(1-m)^2 - \mathcal{K}g(m - m^d), \quad (71)$$

with  $\Lambda = \lambda\sigma_{\mathbf{S}}^2$ ,

$$\lambda = -\mathbb{E}[\mathbf{S}_t^\circ a'_{m,\mathbf{S}}(m^d, 0) v_{aa}(a(m^d, 0), 0, m^d) a_{m,\mathbf{S}}(m^d, 0) \mathbf{S}_t^\circ], \quad (72)$$

where  $a_{m,\mathbf{S}}$  is the second-order cross partial derivative, and I scale  $\mathbf{S}_t - \mathbf{S}_t^d = \sigma_{\mathbf{S}}\mathbf{S}_t^\circ$ : a higher  $\sigma_{\mathbf{S}}$  parameterizes the volatility of  $\mathbf{S}_t$ . Here,  $\frac{1}{2}\Lambda(1-m)^2$  is the leading term in the Taylor expansion of utility losses. The solution is  $m = \mathcal{A}\left(\frac{\lambda\sigma_{\mathbf{S}}^2}{\mathcal{K}}, m^d\right)$  with:

$$\mathcal{A}(v, m^d) := \operatorname{argmin}_m \frac{1}{2}(1-m)^2 v + g(m, m^d) = \max\left(1 - \frac{1}{v}, m^d\right). \quad (73)$$

The following Proposition summarizes this.

**Proposition 8.1** (Endogenizing attention) *The agent's attention  $m$  is given by:*

$$m = \mathcal{A}\left(\frac{\lambda\sigma_{\mathbf{S}}^2}{\mathcal{K}}, m^d\right) = \max\left(1 - \frac{\mathcal{K}}{\lambda\sigma_{\mathbf{S}}^2}, m^d\right),$$

<sup>91</sup>The assumptions on  $\mathbf{G}^{\mathbf{S}}(\mathbf{S}, a, m)$  imply that  $a(m, 0)$  is independent of  $m$ : attention affects only the deviations from the default state.

<sup>92</sup>This generalizes to Gabaix (2014), Definition 1 and Lemma 2. The derivation is in Section 11.11.

where  $\lambda$  is in (72),  $\mathcal{K}$  is the cost of cognition, and the attention function  $\mathcal{A}$  is given by (73).

Proposition 8.1, illustrated by Figure 6, has a number of implications. First, *local rigidity from sparsity*: when  $\sigma_{\mathcal{S}}^2$  is low enough or  $\mathcal{K}$  is large enough (so that  $v = \frac{\lambda\sigma_{\mathcal{S}}^2}{\mathcal{K}}$  is low and we are on the left, constant part of Figure 6), we have  $m = m^d$ , without any reaction to incentives. Locally,  $m$  is constant, at its default value  $m^d$ . Second, *flexible reaction of attention to strong incentives*: for large enough  $\sigma_{\mathcal{S}}^2$  (so that  $v$  is high and we are on the right, increasing part of Figure 6), then attention  $m$  increases in the variance,  $\sigma_{\mathcal{S}}^2$  and in the stakes  $\lambda$ . That is, the Lucas critique applies to big changes in parameters, but not to small ones.

This predicts for instance that when economic volatility (coming from TFP or monetary policy shocks) is increased, then: for a while,  $m$  does not move, and then,  $m$  increases as a function of volatility. For instance, people react more to interest rates in a highly volatile interest rate environment. This kind of comparative statics is sensible, and could be tested more systematically.

**Concrete values for attention** I now apply Proposition 8.1 to the consumer's attention, then to the firm's attention. I consider the case where all fluctuations are driven by productivity  $\zeta_t$ , with the Taylor rule followed by monetary policy (I take  $i_t = \phi_\pi \pi_t + \phi_x x_t + \bar{r}$ , it is consistent with 0 inflation on average, and this economy is worked out in Section 11.12). The proofs suggest how other sources of shocks could be handled.

**Proposition 8.2** (Endogenizing consumer's attention) *In the consumer problem, the attention is*

$$\bar{m} = \bar{m}^c = \mathcal{A}\left(\frac{\lambda^{\bar{m}}\sigma_\zeta^2}{\mathcal{K}^c}, \bar{m}^d\right), \quad m_r = \mathcal{A}\left(\frac{\lambda^{m_r}\sigma_\zeta^2}{\mathcal{K}^c}, m_r^d\right), \quad m_y = \mathcal{A}\left(\frac{\lambda^{m_y}\sigma_\zeta^2}{\mathcal{K}^c}, m_r^d\right), \quad (74)$$

with

$$(\lambda^{\bar{m}}, \lambda^{m_r}, \lambda^{m_y}) = \frac{\gamma(\gamma + \phi)}{\phi} \left( c_{\bar{m},\zeta}^2, c_{m_r,\zeta}^2, c_{m_y,\zeta}^2 \right),$$

where  $\mathcal{K}^c$  is the consumer's cost of cognition, the attention function  $\mathcal{A}$  is in (73), and the coefficients  $c_{\bar{m},\zeta}$ ,  $c_{m_r,\zeta}$ ,  $c_{m_y,\zeta}$  on the right-hand side are given in equations (212), (214) and (215) in the online appendix.

The intuition for those expressions is as follows. In this expression,  $c_{m_r,\zeta}\zeta_t$  represents how much consumption changes (for a given  $\zeta_t$ ) if the consumer pays more attention to the interest rate. Hence, attention to the interest rate is higher if the interest rate matters more, i.e.  $c_{m_r,\zeta}^2$  is high. So, when interest rates have moderate volatility, attention doesn't move (it stays at  $m_r^d$ ), but when volatility increases much, then attention increases.

**Proposition 8.3** (Endogenizing firms' attention) *The firm's attention is:*

$$\bar{m} = \bar{m}^f = \mathcal{A}\left(\frac{\lambda^{\bar{m}^f}\sigma_\zeta^2}{\mathcal{K}^f}, \bar{m}^d\right), \quad m_x^f = \mathcal{A}\left(\frac{\lambda^{m_x^f}\sigma_\zeta^2}{\mathcal{K}^f}, m_x^{f,d}\right), \quad m_\pi^f = \mathcal{A}\left(\frac{\lambda^{m_\pi^f}\sigma_\zeta^2}{\mathcal{K}^f}, m_\pi^{f,d}\right), \quad (75)$$

with

$$\left(\lambda^f, \lambda^{m_x^f}, \lambda^{m_\pi^f}\right) = \frac{\varepsilon - 1}{1 - \beta\theta} \left( q_{\bar{m},\zeta}^2, q_{m_x^f,\zeta}^2, q_{m_\pi^f,\zeta}^2 \right), \quad (76)$$

where  $\mathcal{K}^f$  is the firms' cost of cognition, and the coefficients  $q_{\bar{m},\zeta}$ ,  $q_{m_x^f,\zeta}$ ,  $q_{m_\pi^f,\zeta}$  on the right-hand side are given in equations (220), (223) and (224) in the online appendix.

Note that this might enrich the policy analysis – e.g. attention to inflation depends on the Fed’s aggressiveness in controlling inflation and vice-versa. I do not pursue this here.

## 8.2 A possible “noisy simulations” foundation for cognitive discounting

Here is a possible “noisy simulations” microfoundation for cognitive discounting (8), i.e. the fact that the agent perceives:<sup>93</sup>

$$\mathbf{X}_{t+1} = \bar{m}\mathbf{G}^{\mathbf{X}}(\mathbf{X}_t, \boldsymbol{\epsilon}_{t+1}). \quad (77)$$

**One period: basic idea** To clarify ideas, let us consider a simpler scenario where at time  $t = 0$  the agent simply simulates  $\mathbf{X}_1$ . The true next-period value is  $\mathbf{X}_1 = \mathbf{G}^{\mathbf{X}}(\mathbf{X}_0, \boldsymbol{\epsilon}_1)$ . However, the agent receives a noisy signal  $\mathbf{Y}_1$  about this:

$$\mathbf{Y}_1 = \begin{cases} \mathbf{X}_1 & \text{with probability } q, \\ \mathbf{X}'_1 & \text{with probability } 1 - q. \end{cases} \quad (78)$$

That is, with a probability  $q$ , he receives the correct value  $\mathbf{X}_1$ , while with probability  $1 - q$ , there is a “random reset” in the agent’s simulation process: then the agent receives a random i.i.d. draw  $\mathbf{X}'_1$  from the distribution of  $\mathbf{X}_1$ .<sup>94</sup> This reset captures a form of disruption in the reasoning process.

Normalizing throughout  $\bar{\mathbf{X}} = 0$ , this implies that the conditional mean given the signal  $\mathbf{y}_1$  is:<sup>95</sup>

$$\mathbf{X}_1^e(\mathbf{y}_1) := \mathbb{E}[\mathbf{X}_1 | \mathbf{Y}_1 = \mathbf{y}_1] = q\mathbf{y}_1, \quad (79)$$

and the average perceived value given the truth is:

$$\bar{\mathbf{X}}_1^e(\mathbf{X}_1) := \mathbb{E}[\mathbf{X}_1^e(\mathbf{Y}_1) | \mathbf{X}_1] = q^2 \mathbf{X}_1.$$

Hence, defining

$$\bar{m} := q^2 \quad (80)$$

we have

$$\bar{\mathbf{X}}_1^e(\mathbf{X}_1) = \bar{m}\mathbf{X}_1 = \bar{m}\mathbf{G}^{\mathbf{X}}(\mathbf{X}_0, \boldsymbol{\epsilon}_1). \quad (81)$$

This way, the agent will perceive  $\bar{m}\mathbf{X}_1$  on average. Then, the representative agent (who averages over all agents) will behave according to (77).

<sup>93</sup>Here I give a microfoundation for cognitive discounting, which involves iterated simulations of the future. There is also a more static “noisy perceptions” microfoundation for the intercept parameters  $m_r, m_y$ : it is detailed in Gabaix (2014), Appendix B.

<sup>94</sup>One could use other formulations, e.g.  $\mathbf{Y}_1 = q\mathbf{X}_1 + \sqrt{1 - q^2}\mathbf{X}'_1$ . It works the same way (i.e.,  $\mathbb{E}[\mathbf{X}_1 | \mathbf{Y}_1] = q\mathbf{Y}_1$ , so  $\bar{m} = q^2$ ), but then  $\mathbf{X}_1$  needs to be Gaussian distributed.

<sup>95</sup>Indeed, calling  $g(\mathbf{X}_1)$  the distribution of  $\mathbf{X}_1$ , the joined density of  $(\mathbf{X}_1, \mathbf{Y}_1)$  is  $f(\mathbf{x}_1, \mathbf{y}_1) = qg(\mathbf{y}_1)\delta_{\mathbf{y}_1}(\mathbf{x}_1) + (1 - q)g(\mathbf{y}_1)g(\mathbf{x}_1)$ , where  $\delta$  is the Dirac function. So, as  $\int \mathbf{x}_1 g(\mathbf{x}_1) d\mathbf{x}_1 = 0$ ,

$$\mathbb{E}[\mathbf{X}_1 | \mathbf{Y}_1 = \mathbf{y}_1] = \frac{\int \mathbf{x}_1 f(\mathbf{x}_1, \mathbf{y}_1) d\mathbf{x}_1}{\int f(\mathbf{x}_1, \mathbf{y}_1) d\mathbf{x}_1} = \frac{qg(\mathbf{y}_1)\mathbf{y}_1 + (1 - q)g(\mathbf{y}_1)\int \mathbf{x}_1 g(\mathbf{x}_1) d\mathbf{x}_1}{g(\mathbf{y}_1)} = q\mathbf{y}_1.$$

**How to get all agents to do exactly like the representative agent?** For welfare, it is useful for exact representative-agent aggregation to hold. I show two ways to do that: via the family metaphor, and via the integration within the mind metaphor. The first yields the linearized version of cognitive discounting, (9), while the second story yields the full non-linear version of cognitive discounting, (8). In laying out foundations, it may be useful to have two potential ways to think about things, so I present both.

*The “family metaphor”.* Suppose each agent is really a “family”, made of a continuum of agents  $j \in [0, 1]$ . Agent  $j$  takes the optimal action<sup>96</sup> (expressed as a deviation  $\hat{a}$  from the steady state optimal action,  $\bar{a}$ ) given his perception, so  $\hat{a}_j = a_{\mathbf{X}} \mathbb{E}[\mathbf{X}_1 | \mathbf{Y}_{1j}] = a_{\mathbf{X}} q \mathbf{Y}_{1j}$ . The total action of the aggregate agent is then  $\hat{a} = \int_0^1 \hat{a}_j dj = a_{\mathbf{X}} q \int_0^1 \mathbf{Y}_{1j} dj = a_{\mathbf{X}} q^2 \mathbf{X}_1$ :

$$\hat{a}(\mathbf{X}_1) = a_{\mathbf{X}} \bar{m} \mathbf{X}_1. \quad (82)$$

This way, the aggregate agent is the representative agent, at least to the first order. His action is  $a(\mathbf{X}_1) = \bar{a} + a_{\mathbf{X}} \bar{m} \mathbf{X}_1$ .

*The “integration within the mind” metaphor.* The agent runs a continuum of simulations  $j \in [0, 1]$ , i.e. obtains draws  $\mathbf{Y}_{1j} = \mathbf{Y}_1(\mathbf{X}_1, s_{1j})$ , where  $s_{1j}$  indexes the simulation  $j$ . Each simulation  $j$  leads to a posterior mean  $\mathbf{X}_1^e(\mathbf{Y}_{1j}) = q \mathbf{Y}_{1j}$ . Then, the mind uses an average of those posteriors:

$$\mathbf{X}_1^{BR} = \int_0^1 \mathbf{X}_1^e(\mathbf{Y}_{1j}) dj = q^2 \mathbf{X}_1 = \bar{m} \mathbf{X}_1 = \bar{m} \mathbf{G}^{\mathbf{X}}(\mathbf{X}_0, \boldsymbol{\epsilon}_1).$$

This way, the perceived law of motion is exactly (77).

This law of motion is perceived for a given  $\boldsymbol{\epsilon}_1$ . If the agent just cares about the mean value of a linearized system, she does this for one value,  $\boldsymbol{\epsilon}_1 = 0$ . If the agent cares about the whole non-linear system, the procedure is done for all  $\boldsymbol{\epsilon}_1$ .<sup>97,98</sup>

This completes the microfoundation for a case where the agent simulates the next period.

**Several periods** Now that the one-period simulation is in place, it is easy to generalize to several periods. We have seen how a value  $\mathbf{X}_0$  leads to a value of  $\mathbf{X}_1$  which follows (8). Now, the agent does the same at all periods. She does this going from  $\mathbf{X}_1$  to  $\mathbf{X}_2$ , etc. By induction, the agent perceives (8) for all dates  $t$ .

<sup>96</sup>This action can be any action, e.g. consumption or labor supply.

<sup>97</sup>Implicitly, we assume that the mind can only integrate by taking the sample mean of the signal. It could conceivably take a more sophisticated procedure. When we model bounded rationality (as opposed to optimal information processing), there is a point at which the sophistication of the algorithm must stop. For instance, take level- $k$  models: suppose a level- $k$  agent with  $k = 1$ . Then, given her signal for the reaction at  $k = 1$ , she could optimize some more and find some better estimate of the optimal action. Level  $k$  models assume that the agent just stops there.

<sup>98</sup>I do not claim that agent exactly do this. I just delineate a stylized scenario that would generate (77). Understanding exactly how people calculate expected values (e.g. approximate  $\int V(\varepsilon) f(\varepsilon) d\varepsilon$  for some value  $V$  and distribution  $f$ ) would be very interesting, but completely outside the scope of this study.

## 9 Appendix: Behavioral New Keynesian Macro in a Two-Period Economy

Here I present a two-period model that captures some of the basic features of the behavioral New Keynesian model. I recommend it for entrants to this literature, as everything is very clear with two periods.

It is similar to the model taught in undergraduate textbooks, but with rigorous microfoundations: it makes explicit the behavioral economics foundations of that undergraduate model. It highlights the complementarity between cognitive frictions and pricing frictions.

It is a useful model in its own right: to consider extensions and variants, I found it easiest to start with this two-period model.

**Basic setup.** Utility is:

$$\sum_{t=0}^1 \beta^t u(c_t, N_t) \quad \text{with } u(c, N) = \frac{c^{1-\gamma} - 1}{1-\gamma} - \frac{N^{1+\phi}}{1+\phi}.$$

As in Section 2.2, there is an economy consisting of a Dixit-Stiglitz continuum of firms with Calvo pricing frictions. Calling GDP  $Y_t$ , the aggregate production function is  $Y_t = N_t$  and the aggregate resource constraint is:

$$\text{Resource constraint: } Y_t = c_t + G_t = N_t, \quad (83)$$

where  $G_t$  is real government consumption. The real wage is  $\omega_t$ . Labor supply is frictionless, so the agent respects his first order condition:  $\omega_t u_c + u_N = 0$ , i.e.

$$\text{Labor supply: } N_t^\phi = \omega_t c_t^{-\gamma}. \quad (84)$$

*The economy at time 1.* Let us assume that the time-1 economy has flexible prices and no government consumption, but for simplicity labor supply is rigid at  $N = 1$  (this is a technological constraint). The real wage must equal productivity,  $\omega_t = 1$ , and output is  $y_1 = c_1 = 1$ .

*The economy at time 0.* Now, consider the consumption demand at time 0, for the rational consumer. Taking for now personal income  $y_t$  as given, he solves  $\max_{(c_t)_{t=0,1}} \sum_{t=0}^1 \beta^t \frac{c_t^{1-\gamma}}{1-\gamma}$  subject to  $\sum_{t=0}^1 \frac{c_t}{R_0^t} = y_0 + \frac{y_1}{R_0}$ . That gives

$$c_0 = b \left( y_0 + \frac{y_1}{R_0} \right), \quad (85)$$

$$b := \frac{1}{1+\beta},$$

with log utility.<sup>99</sup> Here  $b$  is the marginal propensity to consume (given the labor supply).<sup>100</sup>

Let us assume for now that the government does not issue any debt nor consumes. Then, aggregate

<sup>99</sup>In the general case,  $b := \frac{1}{1+\beta^\psi R_0^{\psi-1}}$ , calling  $\psi = \frac{1}{\gamma}$  the intertemporal elasticity of substitution (IES). In this section I just use  $\psi = 1$ .

<sup>100</sup>This is different from the more subtle MPC inclusive of labor supply movements, which is  $\frac{\phi}{\gamma+\phi} \frac{1}{1+\beta}$  when evaluated at  $c = N = 1$ .

income equals aggregate consumption:  $y_t = c_t$ . Hence,<sup>101</sup>

$$c_0 = b \left( c_0 + \frac{c_1}{R_0} \right), \quad (86)$$

which yields the Euler equation  $\beta R_0 \frac{c_0}{c_1} = 1$ . I use the consumption function formulation (86) rather than this Euler equation. Indeed, the consumption function is the formulation that generalizes well to behavioral agents.

**Monetary policy is effective with sticky prices.** At time  $t = 0$ , a fraction  $\theta$  of firms have sticky prices – their prices are pre-determined at a value we will call  $P_0^d$  (if prices are sticky, then  $P_0^d = P_{-1}$ , but we could have  $P_0^d = P_{-1}e^{\pi_0^d}$ , where  $\pi_0^d$  is an “automatic” price increase pre-programmed at time  $-1$ , not reactive to time-0 economic conditions, as in Mankiw and Reis (2002) or Section 5.1).<sup>102</sup> As in section 2.2, a corrective wage subsidy is assumed to be in place, so that there are no price distortions on average. Other firms freely optimize their price, and hence optimally choose a price

$$P_0^* = \omega_0 P_0, \quad (87)$$

where  $\omega_0$  is the real wage. Indeed, prices will be flexible at  $t = 1$ , so only current conditions matter for the optimal price. By (21), the aggregate price level is:

$$P_0 = \left( \theta (P_0^d)^{1-\varepsilon} + (1-\theta) (P_0^*)^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}}, \quad (88)$$

as a fraction  $\theta$  of firms set the price  $P_0^d$  and a fraction  $1 - \theta$  set the price  $P_0^*$ .

To solve the problem, there are 6 unknowns ( $c_0, N_0, \omega_0, P_0, P_0^*, R_0$ ) and 5 equations ((83)–(84) and (86)–(88)). What to do?

In the model with flexible prices ( $\theta = 0$ ), this means that the price level  $P_0$  is indeterminate (as in the basic Arrow-Debreu model). However, real variables are determinate: for instance, any solution yields  $c_0 = N_0 = 1$ .

In the model with sticky prices ( $\theta > 0$ ), there is a one-dimensional continuum of real equilibria. It is the central bank that chooses the real equilibrium, by selecting the nominal interest rate, or equivalently here, by choosing the real interest rate  $R_0$ .<sup>103</sup> This is the great power of the central bank.

**The behavioral consumer and fiscal policy.** We can now consider the case where the consumer is behavioral. If his true income at time 1 is  $y_1 = y_1^d + \hat{y}_1$ , he sees only  $y_1^s = y_1^d + \bar{m}\hat{y}_1$  for some  $\bar{m} \in [0, 1]$ , which is the attention to future income shocks ( $\bar{m} = 1$  if the consumer is rational). Here the default is the frictionless case,  $y_1^d = c_1 = Y_1 = 1$ .

But now suppose that (85) becomes:

$$c_0 = b \left( y_0 + \frac{y_1^d + \bar{m}\hat{y}_1}{R_0} \right). \quad (89)$$

<sup>101</sup>The production subsidy by the government, designed to eliminate markup distortions, is paid for by lump-sum taxes. The consumer receives it in profits, then pays it in taxes, so that his total income is just labor income.

<sup>102</sup>This feature is not essential. The reader can imagine the case  $\pi_0^d = 0$ .

<sup>103</sup>The central bank chooses the nominal rate. Given equilibrium inflation, that allows it to choose the real rate (when there are pricing frictions).

Suppose that the government consumes  $G_0$  at 0, nothing at time 1, and makes a transfer  $\mathcal{T}_t$  to the agents at times  $t = 0, 1$ . Call  $d_0 = G_0 + \mathcal{T}_0$  the deficit at time 0. The government must pay its debt at the end of time 1, which yields the fiscal balance equation:

$$R_0 d_0 + \mathcal{T}_1 = 0. \quad (90)$$

The real income of a consumer at time 0 is

$$y_0 = c_0 + G_0 + \mathcal{T}_0 = c_0 + d_0.$$

Indeed, labor and profit income equal the sales of the firms,  $c_0 + G_0$ , plus the transfer from the government,  $\mathcal{T}_0$ . Income at time 1 is  $y_1 = Y_1 + \mathcal{T}_1$ : GDP, plus the transfer from the government.<sup>104</sup> Hence, (89) gives:

$$c_0 = b \left( c_0 + d_0 + \frac{Y_1 + \bar{m}\mathcal{T}_1}{R_0} \right).$$

Using the fiscal balance equation (90) we have:

$$c_0 = b \left( c_0 + (1 - \bar{m}) d_0 + \frac{Y_1}{R_0} \right),$$

and solving for  $c_0$ :

$$c_0 = \frac{b}{1 - b} \left( (1 - \bar{m}) d_0 + \frac{Y_1}{R_0} \right). \quad (91)$$

We see how the ‘‘Keynesian multiplier’’  $\frac{b}{1-b}$  arises.

When consumers are fully attentive,  $\bar{m} = 1$ , and deficits do not matter in (91). However, take the case of behavioral consumers,  $\bar{m} \in [0, 1)$ . Consider a transfer by the government  $\mathcal{T}_0$ , with no government consumption,  $G_0 = 0$ . Equation (91) means that a positive transfer  $d_0 = \mathcal{T}_0$  stimulates activity. If the government gives the agent  $\mathcal{T}_0 > 0$  dollars at time 0, he does not fully see that they will be taken back (with interest) at time 1, so that this is awash. Hence, given  $\frac{Y_1}{R_0}$ , the consumer is tempted to consume more.

To see the full effect, when prices are not frictionless, we need to take a stance on monetary policy to determine  $R_0$ . Here, assume that the central bank does not change the interest rate  $R_0$ .<sup>105</sup> Then, (91) implies that GDP ( $Y_0 = c_0 + G_0$ ) changes as:

$$\frac{dY_0}{d\mathcal{T}_0} = \frac{b}{1 - b} (1 - \bar{m}). \quad (92)$$

With rational agents,  $\bar{m} = 1$ , and fiscal policy has no impact. With behavioral agents,  $\bar{m} < 1$  and fiscal policy has an impact: the Keynesian multiplier  $\frac{b}{1-b}$ , times  $(1 - \bar{m})$ , a measure of deviation from full rationality.

I record these results in the next proposition.

<sup>104</sup>As we assumed that period 1 has frictionless pricing and no government consumption, we have  $c_1 = Y_1 = 1$ . If  $d_0 > 0$ , then the transfer  $\mathcal{T}_1$  is negative. Agents use the proceeds of the time-0 government bonds to pay their taxes at time 1.

<sup>105</sup>With flexible prices ( $\theta = 0$ ), we still have  $\omega_0 = 1$ , hence we still have  $c_0 = N_0 = 1$ . Hence, the interest rate  $R_0$  has to increase. Therefore, to obtain an effect of a government transfer, we need both monetary frictions (partially sticky prices) and cognitive frictions (partial failure of Ricardian equivalence).

**Proposition 9.1** *Suppose that we have (partially) sticky prices, and the central bank keeps the real interest rate constant. Then, a lump-sum transfer  $\mathcal{T}_0$  from the government at time 0 creates an increase in GDP:*

$$\frac{dY_0}{d\mathcal{T}_0} = \frac{b}{1-b} (1 - \bar{m}),$$

where  $b = \frac{1}{1+\beta}$  is the marginal propensity to consume. Likewise, government spending  $G_0$  has the multiplier:

$$\frac{dY_0}{dG_0} = 1 + \frac{b}{1-b} (1 - \bar{m}).$$

We see that  $\frac{dY_0}{d\mathcal{T}_0} > 0$  and  $\frac{dY_0}{dG_0} > 1$  if and only if consumers are non-Ricardian,  $\bar{m} < 1$ .

This proposition also announces a result on government spending, that I now derive. Consider an increase in  $G_0$ , assuming a constant monetary policy (i.e., a constant real interest rate  $R_0$  – alternatively, the central bank might choose to change rates).<sup>106</sup> Equation (91) gives  $\frac{dc_0}{dG_0} = \frac{b}{1-b} (1 - \bar{m})$ , so that GDP,  $Y_0 = c_0 + G_0$ , has a multiplier

$$\frac{dY_0}{dG_0} = 1 + \frac{b}{1-b} (1 - \bar{m}).$$

When  $\bar{m} = 1$  (Ricardian equivalence), a change in  $G_0$  creates no change in  $c_0$ . Only labor demand  $N_0$  increases, hence, via (84), the real wage increases, and inflation increases. GDP is  $Y_0 = c_0 + G_0$ , so that the multiplier  $\frac{dY_0}{dG_0}$  is equal to 1.

However, when  $\bar{m} < 1$  (so that Ricardian equivalence fails), the multiplier  $\frac{dY_0}{dG_0}$  is greater than 1. This is due to the reason invoked in undergraduate textbooks: people feel richer, so they spend more, which creates more demand. Here, we can assert that with good conscience – provided we allow for behavioral consumers.

Without Ricardian equivalence, the government consumption multiplier is greater than 1.<sup>107</sup> Again, this relies on monetary policy being passive, in the sense of keeping a constant real rate  $R_0$ . If the real interest rate rises (as it would with frictionless pricing), then the multiplier would fall to a value less than 1.

**Old vs. New Keynesian Model: a Mixture via Bounded Rationality.** The above derivations show that the model is a mix of Old and New Keynesian models. Here, we do obtain a micro-foundation for the Old Keynesian story (somewhat modified). We see what is needed: some form of non-Ricardian behavior (here via bounded rationality), and of sticky prices. This behavioral model allows for a simple (and I think realistic) mixture of the two ideas.

For completeness, I describe the behavior of realized inflation – the Phillips curve. I describe other features in Section 11.13.

**The Phillips Curve.** Taking a log-linear approximation around  $P_t = 1$ , with  $p_t = \ln P_t$ , (88) becomes:  $p_0 = \theta p_0^d + (1 - \theta) p_0^*$ . Subtracting  $p_0$  on both sides gives  $0 = \theta (p_0^d - p_0) + (1 - \theta) (p_0^* - p_0)$ , i.e.

$$p_0 - p_0^d = \frac{1 - \theta}{\theta} (p_0^* - p_0).$$

<sup>106</sup>See Woodford (2011) for an analysis with rational agents.

<sup>107</sup>This idea is known in the Old Keynesian literature. Mankiw and Weinzierl (2011) consider late in their paper non-Ricardian agents, and find indeed a multiplier greater than 1. But to do that they use two types of agents, which makes the analytics quite complicated when generalizing to a large number of periods. The methodology here generalizes well to static and dynamic contexts.

Recall that  $\overline{P_0^d} = P_{-1} e^{\pi_0^d}$ , so inflation is  $\pi_0 = p_0 - p_{-1} = (p_0 - p_0^d) + (p_0^d - p_{-1})$ , i.e.

$$\pi_0 = \frac{1 - \theta}{\theta} (p_0^* - p_0) + \pi_0^d. \quad (93)$$

Via (87),

$$p_0^* - p_0 = \hat{\omega}_0, \quad (94)$$

where  $\hat{\omega}_0 = \frac{\omega_0 - \omega_0^*}{\omega_0^*}$  is the percentage deviation of the real wage from the frictionless real wage,  $\omega_0^* = 1$ . Because of the labor supply condition (84), and  $c_0 = N_0$ , we have  $\omega_0 = c_0^{\phi + \gamma}$ . Therefore,  $\hat{\omega}_0 = (\phi + \gamma) \hat{c}_0$ . Hence (94) becomes  $p_0^* - p_0 = (\phi + \gamma) \hat{c}_0$ , and (93) yields:

$$\text{Phillips curve: } \pi_0 = \kappa \hat{c}_0 + \pi_0^d, \quad (95)$$

with  $\kappa := \frac{1 - \theta}{\theta} (\phi + \gamma)$ . Hence, we obtain an elementary Phillips curve: increases in economic activity  $\hat{c}_0$  lead to inflation. Inflation comes also from the automatic adjustment  $\pi_0^d$ .

To synthesize, we gather the results. Here  $x_0 = (Y_0 - Y_0^d) / Y_0^d$  is the deviation of GDP from its frictionless value,  $Y_0^d = 1$ , while  $\pi_0$  is the inflation between time -1 (the pre-time 0 price level) and time 0.<sup>108</sup> The deviations of  $(c_0, G_0)$  from trend are from the baseline of  $(1, 0)$ .

**Proposition 9.2** (Two-period behavioral Keynesian model) *In this 2-period model, we have for time-0 consumption and inflation:*

$$x_0 = \hat{G}_0 + b_d \hat{d}_0 - \sigma \hat{r}_0 \text{ (IS curve)}, \quad (96)$$

$$\pi_0 = \kappa \hat{c}_0 + \pi_0^d \text{ (Phillips curve)}, \quad (97)$$

where  $\hat{G}_0$  is government consumption,  $\hat{d}_0$  the budget deficit,  $b_d = \frac{b}{1-b} (1 - \bar{m})$  is the sensitivity to deficits,  $b = \frac{1}{1+\beta}$  is the marginal propensity to consume (given labor income) and  $\hat{r}_0 = i_0 - \mathbb{E}\pi_1$  is the real interest rate between periods 0 and 1, and  $\sigma = \frac{1}{R} = \beta$  with log utility.

This completes the derivation of the 2-period Keynesian model. The online appendix (Section 11.13) contains complements, including a discounted Euler equation.

## 10 Appendix: Complements

### 10.1 Details of the perception of future taxes

Here we flesh out the assumptions and results useful for the fiscal part of Section 2.3. First, we observe that iterating (34) gives  $B_\tau = B_t + R \sum_{u=t}^{\tau-1} d_u$ , so that the transfer at time  $\tau$ ,  $\mathcal{T}_\tau = -\frac{r}{R} B_\tau + d_\tau$  is:

$$\mathcal{T}_\tau = -\frac{r}{R} B_t + \left( d_\tau - r \sum_{u=t}^{\tau-1} d_u \right). \quad (98)$$

Here I detail the formalism useful for the perception of future taxes. This requires some mathematical overhead. Call  $\mathbf{Z}_\tau = (B_\tau, d_\tau, d_{\tau+1}, d_{\tau+2}, \dots)'$  the state vector (more properly, the part of it that concerns

<sup>108</sup>If the agent perceived only part of the change in the real rate, replacing  $R_0$  with  $(1 - m_r) R_0^d + m_r R_0$  in (91), then the expression in (96) would be the same, replacing  $\sigma = \frac{1}{R}$  with  $\sigma = \frac{m_r}{R}$ .

deficits). Under the rational model,  $\mathbf{Z}_{\tau+1} = \mathbf{H}\mathbf{Z}_\tau$  for a matrix  $\mathbf{H}$  characterized by:  $(\mathbf{H}\mathbf{Z})(1) = \mathbf{Z}(1) + R\mathbf{Z}(2)$  and  $(\mathbf{H}\mathbf{Z})(i) = \mathbf{Z}(i+1)$  for  $i > 1$ , where  $\mathbf{Z}(i)$  is the  $i$ -th component of vector  $\mathbf{Z}$ .

The true transfer at time  $\tau$  is  $\mathcal{T}_\tau = -\frac{r}{R}B_\tau + d_\tau = \mathcal{T}(\mathbf{Z}_\tau)$ , where

$$\mathcal{T}(Z) := \mathbf{e}^\mathcal{T} \cdot \mathbf{Z}, \quad \mathbf{e}^\mathcal{T} := \left(-\frac{r}{R}, 1, 0, 0, \dots\right)'. \quad (99)$$

We take a behavioral agent at time  $t$ . He forms a mental model of events and values at future dates  $\tau \geq t$ . Under his subjective model, the law of motion of vector  $\mathbf{Z}_\tau$  is:

$$\mathbf{Z}_{\tau+1} - \mathbf{Z}_t^d = \bar{m}\mathbf{H}(\mathbf{Z}_\tau - \mathbf{Z}_t^d). \quad (100)$$

This is, the agent ‘anchors’ future debt on the current debt captured by  $\mathbf{Z}_t^d = (B_t, 0, 0, \dots)$ , he does only a partial adjustment for the future innovations – as captured by  $\bar{m}$ . Likewise, the perceived law of motion for wealth (29) is extended to

$$k_{\tau+1} = (1 + \bar{r} + \hat{r}^{BR}(\mathbf{Z}_\tau)) (k_\tau + \bar{y} + \hat{y}_\tau^{BR}(\mathbf{Z}_\tau, N_\tau) + \mathcal{T}^{BR}(\mathbf{Z}_\tau, \mathbf{Z}_t^d) - c_\tau), \quad (101)$$

where

$$\mathcal{T}^{BR}(\mathbf{Z}_\tau, \mathbf{Z}_t^d) = (1 - m_y)\mathcal{T}(\mathbf{Z}_t^d) + m_y\mathcal{T}(\mathbf{Z}_\tau) = \mathcal{T}(\mathbf{Z}_t^d) + m_y\mathcal{T}(\mathbf{Z}_\tau - \mathbf{Z}_t^d) \quad (102)$$

is the perceived transfer. Note that  $\mathcal{T}(\mathbf{Z}_t^d) = -\frac{r}{R}B_t$  is the default transfer. Under the rational model, equation (98) gives  $\mathcal{T}(\mathbf{Z}_\tau - \mathbf{Z}_t^d) = d_\tau - r \sum_{u=t}^{\tau-1} d_u$ .

Equations (101)-(102) mean that the ‘baseline’ transfer  $\mathcal{T}(\mathbf{Z}_t^d)$  is seen clearly, but the additional future modifications  $\hat{\mathcal{T}}_\tau$  are seen with a discount,  $m_y$ . Because of cognitive discounting as in (10), we have:

$$\mathbb{E}_t^{BR}[\mathcal{T}(\mathbf{Z}_\tau - \mathbf{Z}_t^d)] = \bar{m}^{\tau-t}\mathbb{E}_t[\mathcal{T}(\mathbf{Z}_\tau - \mathbf{Z}_t^d)] = \bar{m}^{\tau-t}\mathbb{E}_t\left[d_\tau - r \sum_{u=t}^{\tau-1} d_u\right]. \quad (103)$$

This gives the future taxes, as perceived by the agent at time  $t$ :

$$\mathbb{E}_t^{BR}[\mathcal{T}^{BR}(\mathbf{X}_\tau)] = -\frac{r}{R}B_t + m_y\bar{m}^{\tau-t}\mathbb{E}_t\left[d_\tau - r \sum_{u=t}^{\tau-1} d_u\right]. \quad (104)$$

This reflects a partially rational consumer. Suppose that there are no future deficits. Given initial debt  $B_t$ , the consumer will see that it will have to be repaid: he accurately foresees the part  $-\frac{r}{R}B_t$  in the perception of future deficits (104). However, he sees only dimly future deficits and their impact on future taxes. This is captured by the term  $m_y\bar{m}^{\tau-t}$ .

## 10.2 Additional Proofs

**Proof of Lemma 2.4** *Notations.* The proof of this lemma and that of Proposition 2.5 follows the steps and notations of Galí (2015, Sections 3.2–3.3). I simplify matters by assuming constant returns to scale ( $\alpha = 0$  in Galí’s notations). So, the nominal marginal cost at  $t+k$  is simply  $\psi_{t+k}$ , not  $\psi_{t+k|t}$ . When referring to, say, equation (11) of Chapter 3 in Galí (2015), I write ‘equation (G11)’. I replace the coefficient of relative risk aversion ( $\sigma$  in his notations) by  $\gamma$  (as in  $u'(c) = c^{-\gamma}$ ).

If the firm was free to choose its real (log) price  $q_{it}$  freely, it would choose price  $q_{it}^*$  maximizing (22), i.e.  $e^{q_{it}^*} = \frac{1-\tau_f}{1-\frac{1}{\varepsilon}}MC_t$ . The subsidy  $\tau_f = \frac{1}{\varepsilon}$  was chosen to eliminate the monopoly distortion on average.

The FOC for the (subjectively) optimal flexible price is  $q_i^{*,BR}(\mathbf{X}_\tau) := \operatorname{argmax}_{q_i} v^{BR}(q_i, \mathbf{X}_\tau)$ . For firms facing the Calvo pricing friction, we have, much as in the traditional model, that the price is the weighted average of future optimal prices:<sup>109</sup>

$$q_{it} = (1 - \beta\theta) \sum_{\tau \geq t} (\beta\theta)^{\tau-t} \mathbb{E}_t^{BR} \left[ q_i^{*,BR}(\mathbf{X}_\tau) \right], \quad (105)$$

which is a behavioral counterpart to Galí's (G11).

Given the behavioral perceptions in (25), we have, linearizing:

$$q_i^{*,BR}(\mathbf{X}_\tau) = m_\pi^f \Pi(\mathbf{X}_\tau) - m_x^f \mu(\mathbf{X}_\tau). \quad (106)$$

Now, by the now usual cognitive discounting (11), we have:

$$\mathbb{E}_t^{BR} [\Pi(\mathbf{X}_\tau)] = \bar{m}^{\tau-t} \mathbb{E}_t [\Pi(\mathbf{X}_\tau)], \quad \mathbb{E}_t^{BR} [\mu(\mathbf{X}_\tau)] = \bar{m}^{\tau-t} \mathbb{E}_t [\mu(\mathbf{X}_\tau)].$$

So, we have the following counterpart to the equation right before (G16):

$$\begin{aligned} q_{it} &= (1 - \beta\theta) \sum_{\tau \geq t} (\beta\theta)^{\tau-t} \mathbb{E}_t^{BR} [m_\pi^f \Pi(\mathbf{X}_\tau) - m_x^f \mu(\mathbf{X}_\tau)] \\ &= (1 - \beta\theta) \sum_{\tau \geq t} (\beta\theta)^{\tau-t} \bar{m}^{\tau-t} \mathbb{E}_t [m_\pi^f \Pi_\tau - m_x^f \mu_\tau]. \end{aligned}$$

**Proof of Proposition 2.5** Let us define

$$\rho := \beta\theta\bar{m}, \quad (107)$$

and calculate

$$H_t := \sum_{k \geq 1} \rho^k (\pi_{t+1} + \dots + \pi_{t+k}) = \sum_{i \geq 1} \pi_{t+i} \sum_{k \geq i} \rho^k = \sum_{i \geq 1} \pi_{t+i} \frac{\rho^i}{1 - \rho} = \frac{1}{1 - \rho} \sum_{i \geq 0} \pi_{t+i} \rho^i 1_{i > 0}.$$

Firms who can reset their price choose a price  $p_t^*$  given in (27):

$$\begin{aligned} p_t^* - p_t &= (1 - \beta\theta) \sum_{k=0}^{\infty} \rho^k \mathbb{E}_t [m_\pi^f (\pi_{t+1} + \dots + \pi_{t+k}) - m_x^f \mu_{t+k}] = (1 - \beta\theta) \mathbb{E}_t \left[ m_\pi^f H_t - \sum_{k=0}^{\infty} \rho^k m_x^f \mu_{t+k} \right] \\ &= (1 - \beta\theta) \sum_{k \geq 0} \rho^k \mathbb{E}_t \left[ \frac{m_\pi^f}{1 - \rho} \pi_{t+k} 1_{k > 0} - m_x^f \mu_{t+k} \right] \\ &= \sum_{k \geq 0} \rho^k \mathbb{E}_t [m'_{f\pi} \pi_{t+k} 1_{k > 0} - \mu'_{t+k}], \quad (108) \end{aligned}$$

<sup>109</sup>The proof is as in the traditional model: the FOC of problem (26) is  $\mathbb{E}^{BR} \sum_{\tau \geq t} (\beta\theta)^{\tau-t} v_{q_i}^{BR}(q_{it}, \mathbf{X}_\tau) = 0$  and linearizing around  $q_i^{*,BR}(\mathbf{X}_\tau)$ , the FOC is  $\mathbb{E}^{BR} \sum_{\tau \geq t} (\beta\theta)^{\tau-t} v_{q_i q_i}^{BR}(q_i^{*,BR}(\mathbf{X}_\tau), \mathbf{X}_\tau) \cdot (q_{it} - q_i^{*,BR}(\mathbf{X}_\tau)) = 0$ . Taking the Taylor expansion around 0 disturbances so  $q_i^{*,BR}(\mathbf{X}_\tau)$  close to 0, the terms  $v_{q_i q_i}^{BR}(q_i^{*,BR}(\mathbf{X}_\tau), \mathbf{X}_\tau)$  are approximately constant and equal to  $v_{q_i q_i}^{BR}(0, 0)$  up to first order terms, and the FOC is (up to second order terms)  $\mathbb{E}^{BR} \sum_{\tau \geq t} (\beta\theta)^{\tau-t} (q_{it} - q_i^{*,BR}(\mathbf{X}_\tau)) = 0$ , which gives (105).

where  $m'_{f\pi} := \frac{1-\beta\theta}{1-\rho} m^f_\pi$  and  $\mu'_t := m^f_x (1 - \beta\theta) \mu_t$ .

*Inflation.* As a fraction  $1 - \theta$  of firms reset their price, starting from  $p_{t-1}$  on average:

$$\begin{aligned}\pi_t &= p_t - p_{t-1} = (1 - \theta) (p_t^* - p_{t-1}) \\ &= (1 - \theta) (p_t^* - p_t + p_t - p_{t-1}) = (1 - \theta) (p_t^* - p_t + \pi_t). \\ \pi_t &= \frac{1 - \theta}{\theta} (p_t^* - p_t).\end{aligned}\tag{109}$$

Plugging this in (108) gives:

$$\pi_t = \frac{1 - \theta}{\theta} \sum_{k \geq 0} \rho^k \mathbb{E}_t [m'_{f\pi} \pi_{t+k} \mathbf{1}_{k > 0} - \mu'_{t+k}].\tag{110}$$

Next, I use the forward operator  $F$  ( $Fy_t := y_{t+1}$ ), which allows me to evaluate infinite sums compactly, as in:

$$\sum_{k=0}^{\infty} \rho^k y_{t+k} = \sum_{k=0}^{\infty} \rho^k (F^k y_t) = \left( \sum_{k=0}^{\infty} \rho^k F^k \right) y_t = (1 - \rho F)^{-1} y_t.\tag{111}$$

Rewriting (110) using  $F$  gives

$$\pi_t = \frac{1 - \theta}{\theta} \mathbb{E}_t (1 - \rho F)^{-1} (m'_{f\pi} \rho F \pi_t - \mu'_t).$$

Hence, multiplying by  $1 - \rho F$ , we obtain the key equation (which is a behavioral version of (G17)):

$$\pi_t = \beta M^f \mathbb{E}_t [\pi_{t+1}] - \lambda \mu_t,\tag{112}$$

with

$$\beta^f := \rho \left( 1 + \frac{1 - \theta}{\theta} m'_{f\pi} \right) = \beta \bar{m} \left( \theta + \frac{1 - \beta\theta}{1 - \beta\theta \bar{m}} m^f_\pi (1 - \theta) \right) = \beta M^f,\tag{113}$$

$$M^f = \bar{m} \left( \theta + \frac{1 - \beta\theta}{1 - \beta\theta \bar{m}} m^f_\pi (1 - \theta) \right),\tag{114}$$

$$\lambda = m^f_x \frac{1 - \theta}{\theta} (1 - \beta\theta).$$

The rest of the proof is as in Galí. The labor supply is still (14),  $N_t^\phi = \omega_t c_t^{-\gamma}$ , and as the resource constraint is  $c_t = e^{\zeta_t} N_t$ ,  $\omega_t = e^{-\phi \zeta_t} c_t^{\gamma + \phi}$ , i.e.  $\hat{\omega}_t = -\phi \zeta_t + (\gamma + \phi) \hat{c}_t$ , so that with  $\mu_t := \zeta_t - \hat{\omega}_t$ ,

$$\mu_t = (1 + \phi) \zeta_t - (\gamma + \phi) \hat{c}_t.$$

Next, as in the “natural” economy without pricing frictions,  $\mu_t^n = 0$ :

$$0 = (1 + \phi) \zeta_t - (\gamma + \phi) \hat{c}_t^n.$$

Hence, subtracting the two equations, and using  $x_t = \hat{c}_t - \hat{c}_t^n$ ,

$$\mu_t = -(\gamma + \phi) x_t.\tag{115}$$

Plugging this into (112), we obtain the behavioral version of (G22):

$$\pi_t = \beta M^f \mathbb{E}_t [\pi_{t+1}] + \kappa x_t,$$

with  $\kappa = \lambda(\gamma + \phi)$ , i.e.

$$\kappa = \bar{\kappa} m_x^f, \tag{116}$$

$$\bar{\kappa} = \left( \frac{1}{\theta} - 1 \right) (1 - \beta\theta)(\gamma + \phi). \tag{117}$$

**Proof of Proposition 2.7** First, we state a simple result.

**Proposition 10.1** (Consumption given beliefs) *Consider an agent maximizing over  $(c_\tau, N_\tau)$  utility  $U_t = \mathbb{E}_t^{BR} \sum_{\tau=t}^{\infty} \beta^{\tau-t} u(c_\tau, N_\tau)$  subject to the law of motion for wealth (29). Up to second order terms (and for small wealth  $k_0$ ), consumption is:*

$$c_t = \frac{r}{R} k_t + \bar{y} + \mathbb{E}_t^{BR} \left[ \sum_{\tau \geq t} \frac{1}{R^{\tau-t}} (b_r \hat{r}^{BR}(\mathbf{X}_\tau) + b_y \hat{y}^{BR}(N_\tau, \mathbf{X}_\tau)) \right], \tag{118}$$

where expectations are taken under the agent's subjective model of the world,  $b_r = \frac{-1}{\gamma R^2}$ ,  $b_y = \frac{\bar{r}}{R}$ , and  $N_\tau$  labor supply at the perceived optimum. The chosen labor supply is given by  $N_t^\phi = \hat{y}_N^{BR}(N_t, \mathbf{X}_t) c_t^{-\gamma}$ .

It is stated as a function of an endogenous labor supply, because this is the form that is most useful in some derivations (Section 11.2 develops the case that explicitly solves for labor supply). Versions of this proposition were proven a number of times with minor variants (e.g. Eusepi and Preston (2011); Woodford (2013); Gabaix (2016); Auclert (2017)), but for completeness let us derive it.<sup>110</sup>

*Proof of Proposition 10.1.* For simplicity, we take the deterministic case (as we consider first order Taylor expansions, the general case is the deterministic case where the path of variables are their expected values). The agent wants to maximize, over  $c_t$  and  $N_t$ :

$$\mathcal{L} = \sum_{t \geq 0} \beta^t u(c_t, N_t) + \lambda \left( k_0 + \sum_{t \geq 0} q_t (\bar{y} + \hat{y}^{BR}(N_t, \mathbf{X}_t) - c_t) \right), \tag{119}$$

where  $q_t := 1 / \prod_{\tau=0}^{t-1} (1 + \bar{r} + \hat{r}^{BR}(\mathbf{X}_\tau))$ . Here we consider the decision at time 0, which is just a normalization. Consider first the problem of optimizing  $\mathcal{L}$  over  $c_t$  (taking the value of  $y_t := \bar{y} + \hat{y}^{BR}(N_t, \mathbf{X}_t)$  as given), the FOC is  $\beta^t c_t^{-\gamma} = \lambda q_t$ , so  $c_t = c_0 \left( \frac{\beta^t}{q_t} \right)^\psi$  (with  $\psi := \frac{1}{\gamma}$ ), for some  $c_0 = \lambda^{-\psi}$ . With  $\Omega := k_0 + \sum_{t \geq 0} q_t y_t$  the total perceived wealth, the perceived budget constraint is  $\Omega = \sum_{t \geq 0} q_t c_t = c_0 \sum_{t \geq 0} \beta^{t\psi} q_t^{1-\psi}$ . This gives consumption at decision time:

$$c_0 = \frac{k_0 + \sum_{t \geq 0} q_t y_t}{\sum_{t \geq 0} \beta^{t\psi} q_t^{1-\psi}}. \tag{120}$$

From there, let us see the impact of a small change in income. We have, at the default value of interest rates,  $q_t = \beta^t$ , so  $\sum_{t \geq 0} \beta^{t\psi} q_t^{1-\psi} = \frac{1}{1-\beta}$ , so  $c_0 = (1 - \beta) (k_0 + \sum_{t \geq 0} \beta^t y_t)$ . This yields  $b_y = 1 - \beta = \frac{r}{R}$  in

<sup>110</sup>See also the proof of Lemma 4.2 in Gabaix (2016) for another style of derivation, using dynamic programming.

(118). The impact of the interest rate on current consumptions is similar, though a little tedious (it is done in Section 12).

Next, for labor supply at time 0, the behavioral agent optimizes (119) given his perceived model of the world. This gives  $\mathcal{L}_{N_0} = 0$ , i.e.  $-N_0^\phi + \lambda \hat{y}_N^{BR}(N_0, \mathbf{X}_0) = 0$ . As we saw that  $c_0^{-\gamma} = \lambda$ , we obtain  $N_0^\phi = c_0^{-\gamma} \hat{y}_N^{BR}(N_0, \mathbf{X}_0)$ .  $\square$

**Application to this paper's behavioral agent** *When  $m_y = 1$ , and no initial wealth.* This is the simplest case. The behavioral agent perceives his dynamic budget constraint is (29) and (30). Hence, we apply Proposition 10.1 to (30). At the optimum policy, we have  $N_t = N(\mathbf{X}_t)$ , (under the perceived motion for  $\mathbf{X}_t$ ), so the planned labor supply also verifies  $N_\tau = N(\mathbf{X}_\tau)$ , so  $\hat{y}^{BR}(N_t, \mathbf{X}_t) = \hat{y}(\mathbf{X}_t)$ . Now, using cognitive discounting (10), we have (31). This gives the consumption:

$$\hat{c}_t = \mathbb{E}_t^{BR} \left[ \sum_{\tau \geq t} \frac{1}{R^{\tau-t}} \left( b_r \hat{r}^{BR}(\mathbf{X}_t) + \frac{r}{R} \hat{y}(\mathbf{X}_t) \right) \right] = \mathbb{E}_t \left[ \sum_{\tau \geq t} \frac{\bar{m}^{\tau-t}}{R^{\tau-t}} \left( b_r \hat{r}(\mathbf{X}_t) + \frac{r}{R} \hat{y}(\mathbf{X}_t) \right) \right].$$

*With a general  $m_y$  or non-zero initial wealth.* Here things are more complex, because the aggregate wealth is 0, and the agent plans to have non-zero wealth next period (as he misperceives income), so the agent doesn't plan that her future labor supply will be equal to the aggregate labor supply,  $N_\tau = N(\mathbf{X}_\tau)$ . Section 11.2 gives the proof.

*Another method: Dynamic programming with Taylor expansion.* The following method is a bit less intuitive, but may be handy to automatize when considering medium-scale extensions of this model. The subjective value function of the agent satisfies:

$$V(k, \mathbf{X}) = \max_{c, N} \{ u(c, N) + \beta \mathbb{E}V \left( (1 + r + m_r \hat{r}(\mathbf{X})) (k + \bar{y} + m_y \hat{y}(\mathbf{X}) + w(\mathbf{X})(N - N(\mathbf{X})) - c), \bar{m} \mathbf{G}^{\mathbf{X}}(\mathbf{X}, \epsilon) \right) \}$$

and optimal consumption satisfies  $u_c(c(k, \mathbf{X}), N) = V_k(k, \mathbf{X})$  (independently of  $N$  because utility is separable), so that  $c_{\mathbf{X}} = \frac{V_{k\mathbf{X}}}{u_{cc}} = -\frac{V_{k\mathbf{X}}}{\gamma}$ . In turn,  $\hat{c}_t = c_{\mathbf{X}} \mathbf{X}_t$ . Hence, to derive consumption, we simply need to calculate  $V_{k\mathbf{X}}$ . This is done in Section 11.5.

**Proof of Proposition 2.8** Proposition 2.7 gives:

$$\hat{c}_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} \frac{\bar{m}^{\tau-t}}{R^{\tau-t}} \left( \frac{r}{R} m_Y \hat{y}_\tau + b_r m_r \hat{r}_\tau \right) \right]. \quad (121)$$

Now, since there is no capital in the NK model, we have  $\hat{y}_\tau = \hat{c}_\tau$ : income is equal to aggregate demand. Hence, using  $\tilde{b}_y := \frac{r}{R} m_Y$  and  $\tilde{b}_r := b_r m_r = -\frac{m_r}{\gamma R^2}$ , (121) becomes:

$$\hat{c}_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} \frac{\bar{m}^{\tau-t}}{R^{\tau-t}} \left( \tilde{b}_y \hat{c}_\tau + \tilde{b}_r \hat{r}_\tau \right) \right]. \quad (122)$$

Taking out the first term yields:

$$\hat{c}_t = \tilde{b}_y \hat{c}_t + \tilde{b}_r \hat{r}_t + \mathbb{E}_t \left[ \sum_{\tau \geq t+1} \frac{\bar{m}^{\tau-t}}{R^{\tau-t}} \left( \tilde{b}_y \hat{c}_\tau + \tilde{b}_r \hat{r}_\tau \right) \right].$$

Given that (122), applied to  $t + 1$ , yields  $\hat{c}_{t+1} = \mathbb{E}_{t+1} \left[ \sum_{\tau \geq t+1} \frac{\bar{m}^{\tau-t-1}}{R^{\tau-t-1}} \left( \tilde{b}_y \hat{c}_\tau + \tilde{b}_r \hat{r}_\tau \right) \right]$ , we have:

$$\hat{c}_t = \tilde{b}_y \hat{c}_t + \tilde{b}_r \hat{r}_t + \frac{\bar{m}}{R} \mathbb{E}_t [\hat{c}_{t+1}] = \frac{r}{R} m_Y \hat{c}_t + \tilde{b}_r \hat{r}_t + \frac{\bar{m}}{R} \mathbb{E}_t [\hat{c}_{t+1}].$$

Multiplying by  $R$  and gathering the  $\hat{c}_t$  terms, we have  $\hat{c}_t = \frac{\bar{m} \mathbb{E}_t [\hat{c}_{t+1}] + R \tilde{b}_r \hat{r}_t}{R - r m_Y}$ . This suggests defining  $M := \frac{\bar{m}}{R - r m_Y}$  and  $\sigma := \frac{-R \tilde{b}_r}{R - r m_Y} = \frac{m_r}{\gamma R (R - r m_Y)}$ , and we get (13). This then translates into (19).<sup>111</sup>

**Proof of Proposition 3.1** We will use the following well-known fact (Bullard and Mitra (2002)):

**Proposition 10.2** (Roots in unit circle) *Consider the polynomial  $p(x) = x^2 + ax + b$ . Its two roots satisfy  $|x| < 1$  if and only if:  $|a| - 1 < b < 1$ .*

We calculate  $p(x) := \det(x\mathbf{I} - \mathbf{A}) = x^2 + ax + b$  with

$$a = -\frac{M + \beta^f + \kappa\sigma + \beta^f \sigma \phi_x}{D}, \quad b = \frac{M\beta^f}{D},$$

with  $D = 1 + \sigma(\phi_x + \kappa\phi_\pi)$ . Proposition 10.2 indicates that the equilibrium is determinate iff:  $|a| - 1 < b < 1$ . Given that we assume nonnegative coefficients  $\phi$ ,  $b < 1$  and  $a < 0$ . Hence the criterion is:  $1 + b + a > 0$ , i.e.  $p(1) > 0$ . Calculations show that this is (46).

**Proof of Proposition 3.2** Go back to (49), assuming the first best after the ZLB, so  $\mathbf{z}_T = 0$ . Then,

$$\mathbf{z}_0(T) = (\mathbf{A}_{ZLB} - \mathbf{I})^{-1} (\mathbf{A}_{ZLB}^T - \mathbf{I}) \underline{\mathbf{b}}.$$

When condition (47) fails, one of the eigenvalues of  $\mathbf{A}_{ZLB}$  is greater than 1 in modulus. Then,  $\lim_{T \rightarrow \infty} \|\mathbf{A}_{ZLB}^T \underline{\mathbf{b}}\| = \infty$  (it is easy to verify that  $\underline{\mathbf{b}}$  is not exactly the eigenvector corresponding to the root less than 1 in modulus). Hence,  $\lim_{T \rightarrow \infty} \|\mathbf{z}_0(T)\| = \infty$ . Furthermore, this explosion is a recession: given that the entries of  $\mathbf{A}_{ZLB}$  are positive, and those of  $\underline{\mathbf{b}}$  are negative, each of the terms in  $(\mathbf{I} + \mathbf{A}_{ZLB} + \dots + \mathbf{A}_{ZLB}^{T-1}) \underline{\mathbf{b}}$  is negative, hence  $\mathbf{z}_0(T)$  has unboundedly negative inflation and output gap.

When condition (47) holds, all roots of  $\mathbf{A}_{ZLB}$  are less than 1 in modulus. Hence,  $\lim_{T \rightarrow \infty} \mathbf{z}_0(T) = -(\mathbf{A}_{ZLB} - \mathbf{I})^{-1} \underline{\mathbf{b}}$ , a finite value.

**Proof of Proposition 4.3** The Lagrangian is

$$L = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[ -\frac{1}{2} (\pi_t^2 + \vartheta x_t^2) + \Xi_t (\beta M^f \pi_{t+1} + \kappa x_t - \pi_t + \nu_t) \right],$$

<sup>111</sup>Here, bounded rationality lowers  $\sigma$ , the effective sensitivity to the interest rate, in addition to lowering  $M$ . With heterogeneous agents (along the lines of Auclert (2017)), one can imagine that bounded rationality might increase  $\sigma$ : some high-MPC (marginal propensity to consume) agents will have to pay adjustable-rate mortgages, which will increase the stimulative effects of a fall in the rate (increase  $\sigma$ ).

where  $\Xi_t$  are Lagrange multipliers. The first order conditions are:  $L_{x_t} = 0$  and  $L_{\pi_t} = 0$ , which give respectively  $-\vartheta x_t + \kappa \Xi_t = 0$  and  $-\pi_t - \Xi_t + M^f \Xi_{t-1} = 0$ , i.e.  $\Xi_t = \frac{\vartheta}{\kappa} x_t$  and  $\pi_t = \frac{-\vartheta}{\kappa} (x_t - M^f x_{t-1})$ .

**Proof of Proposition 4.4** The central bank today takes its future actions as given, and chooses  $x_t, \pi_t, i_t$  to minimize today's loss  $-\frac{1}{2}(\pi_t^2 + \vartheta x_t^2)$  subject to the behavioral IS equation and behavioral NK Phillips curve. This is equivalent to

$$\max_{\pi_t, x_t} -\frac{1}{2} (\pi_t^2 + \vartheta x_t^2) \quad \text{subject to} \quad \pi_t = \beta M^f \mathbb{E} \pi_{t+1} + \kappa x_t + \nu_t,$$

and  $i_t$  can be read off the IS equation. Hence, the Lagrangian is simply:

$$L = -\frac{1}{2} (\pi_t^2 + \vartheta x_t^2) + \Xi (\beta M^f \mathbb{E} \pi_{t+1} + \kappa x_t + \nu_t - \pi_t).$$

The first order conditions are:  $L_{x_t} = 0$  and  $L_{\pi_t} = 0$ , i.e.  $-\vartheta x_t + \kappa \Xi = 0$  and  $-\pi_t - \Xi = 0$ , which together yields  $\pi_t = -\frac{\vartheta}{\kappa} x_t$ . The explicit value of  $i_t$  is in Section 12.

**Proof of Proposition 5.1** Call  $q_{it} := p_{it} - p_t$  the real log price of firm  $i$  at a date  $t$ . Consider a firm that has not done a Calvo reset between  $t$  and  $\tau > t$ , and instead has simply passively indexed on default inflation. Then (using  $\Delta z_\tau = z_\tau - z_{\tau-1}$ ),  $\Delta p_{i,\tau} = \pi_\tau^d$ , and  $\Delta p_\tau = \pi_\tau = \pi_\tau^d + \hat{\pi}_\tau$ . Hence,  $\Delta q_{i,\tau} = -\hat{\pi}_\tau$ , i.e.  $q_{i\tau} = q_{it} - \hat{\Pi}_{i\tau}$  where  $\hat{\Pi}_\tau := \hat{\pi}_{t+1} + \dots + \hat{\pi}_\tau$  is the cumulative inflation between  $t$  and  $\tau$ , but only in ‘‘hat space’’, i.e. considering the deviation of inflation from default inflation. Intuitively, the firm knows that it will indexed on default inflation, so it's important for it to forecast the deviation from default inflation,  $\hat{\pi}_\tau$ , not inflation itself.

Then, we are in a world isomorphic to that of Section 2.2, expect that we put hats on  $\pi$  and  $\Pi$  – this is, replace  $\pi$  and  $\Pi$  by  $\hat{\pi}$  and  $\hat{\Pi}$ . For instance, the state space is now  $\mathbf{X}_\tau = (\mathbf{X}_\tau^M, \hat{\Pi}_\tau)$ , where  $\mathbf{X}_\tau^M$  is the basic macro state vector. The firm's profit is (23) with hats on  $\Pi$ , and so the natural generalization of (25) is that the behavioral firm perceives the flow profit

$$v^{BR}(q_{it}, \mathbf{X}_\tau) := v \left( q_{it} - m_\pi^f \hat{\Pi}(\mathbf{X}_\tau), m_x^f \mu(\mathbf{X}_\tau), c(\mathbf{X}_\tau) \right), \quad (123)$$

and its objective is still (26), with that notation. This leads to the economy in Proposition 5.1.<sup>112</sup>

## References

- Afrouzi, H. (2017). Strategic inattention, inflation dynamics and the non-neutrality of money. *Working Paper*.
- Andrade, J., Cordeiro, P., and Lambais, G. (2018). Estimating a Behavioral New Keynesian Model. *In Preparation*.
- Angeletos, G.-M., Collard, F., and Dellas, H. (2017). Quantifying confidence. *NBER Working Paper No. 20807*.
- Angeletos, G.-M. and La'O, J. (2009). Incomplete information, higher-order beliefs and price inertia. *Journal of Monetary Economics*, 56:S19–S37.

<sup>112</sup>The derivation is very simple. Lemma 2.4 holds, putting hats on  $\pi$ , and as a result, the Phillips curve holds (28) also putting hats on  $\pi$ , i.e. (59) holds.

- Angeletos, G.-M. and La'O, J. (2010). Noisy business cycles. *NBER Macroeconomics Annual*, 24(1):319–378.
- Angeletos, G.-M. and Lian, C. (2016). Incomplete information in macroeconomics: Accommodating frictions in coordination. *Handbook of Macroeconomics*, 2:1065–1240.
- Angeletos, G.-M. and Lian, C. (2017a). Dampening general equilibrium: From micro to macro. *NBER Working Paper No. 22785*.
- Angeletos, G.-M. and Lian, C. (2017b). Forward Guidance without Common Knowledge. *Forthcoming at the American Economic Review*.
- Ascari, G. and Ropele, T. (2012). Sacrifice Ratio in a Medium-Scale New Keynesian Model. *Journal of Money, Credit and Banking*, 44(2-3):457–467.
- Auclert, A. (2017). Monetary Policy and the Redistribution Channel. *NBER Working Paper No. 23451*.
- Barro, R. J. (1974). Are Government Bonds Net Wealth? *Journal of political economy*, 82(6):1095–1117.
- Benchimol, J. and Bounader, L. (2018). Optimal monetary policy under bounded rationality. *Working Paper*.
- Bilbiie, F. O. (2008). Limited Asset Markets Participation, Monetary Policy and (Inverted) Aggregate Demand Logic. *Journal of Economic Theory*, 140(1):162–196.
- Blanchard, O. J. and Kahn, C. M. (1980). The Solution of Linear Difference Models Under Rational Expectations. *Econometrica*, 48(5):1305–1311.
- Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Saliency and consumer choice. *Journal of Political Economy*, 121(5):803–843.
- Bounader, L. (2016). Optimal Monetary Policy in Behavioral New Keynesian Model. *Working Paper*.
- Brown, J., Hossain, T., and Morgan, J. (2010). Shrouded attributes and information suppression: Evidence from the field. *Quarterly Journal of Economics*, 125(2):859–876.
- Bullard, J. and Mitra, K. (2002). Learning about monetary policy rules. *Journal of Monetary Economics*, 49(6):1105–1129.
- Caballero, R. J. (1995). Near-rationality, heterogeneity, and aggregate consumption. *Journal of Money, Credit and Banking*, 27(1):29–48.
- Caballero, R. J. and Farhi, E. (2017). The Safety Trap. *Review of Economic Studies*, 85(1):223–274.
- Campbell, J. R., Fisher, J. D., Justiniano, A., and Melosi, L. (2017). Forward Guidance and Macroeconomic Outcomes since the Financial Crisis. *NBER Macroeconomics Annual*, 31(1):283–357.
- Campbell, J. Y. and Mankiw, N. G. (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER Macroeconomics Annual*, 4:185–216.
- Caplin, A., Dean, M., and Leahy, J. (2017). Rationally inattentive behavior: Characterizing and generalizing Shannon entropy. *NBER Working Paper No. 23652*.
- Caplin, A., Dean, M., and Martin, D. (2011). Search and satisficing. *The American Economic Review*, 101(7):2899–2922.
- Carvalho, C. and Nechio, F. (2014). Do People Understand Monetary Policy? *Journal of Monetary Economics*, 66:108–123.
- Chetty, R., Looney, A., and Kroft, K. (2009). Saliency and taxation: Theory and evidence. *The American Economic Review*, 99(4):1145–1177.
- Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.
- Clarida, R., Gali, J., and Gertler, M. (1999). The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature*, 37(4):1661–1707.

- Clarida, R., Gali, J., and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly Journal of Economics*, 115(1):147–180.
- Cochrane, J. H. (2011). Determinacy and Identification with Taylor Rules. *Journal of Political Economy*, 119(3):565–615.
- Cochrane, J. H. (2017). Michelson-Morley, Occam and Fisher: The Radical Implications of Stable Inflation at Near-Zero Interest Rates. *Forthcoming at NBER Macroeconomics Annual*.
- Coibion, O. and Gorodnichenko, Y. (2015a). Information rigidity and the expectations formation process: A simple framework and new facts. *The American Economic Review*, 105(8):2644–2678.
- Coibion, O. and Gorodnichenko, Y. (2015b). Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation. *American Economic Journal: Macroeconomics*, 7(1):197–232.
- Del Negro, M., Giannoni, M. P., and Patterson, C. (2015). The Forward Guidance Puzzle. *Federal Reserve Bank of New York Staff Reports* 574.
- Eggertsson, G. B. and Krugman, P. (2012). Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach. *Quarterly Journal of Economics*, 127(3):1469–1513.
- Eggertsson, G. B. and Mehrotra, N. R. (2015). A Model of Secular Stagnation. *NBER Working Paper No. 20574*.
- Eggertsson, G. B. and Woodford, M. (2003). The Zero Bound on Interest Rates and Optimal Monetary Policy. *Brookings Papers on Economic Activity*, 34(1):139–235.
- Erceg, C. J. and Levin, A. T. (2003). Imperfect credibility and inflation persistence. *Journal of Monetary Economics*, 50(4):915–944.
- Eusepi, S. and Preston, B. (2011). Expectations, learning, and business cycle fluctuations. *The American Economic Review*, 101(6):2844–2872.
- Eusepi, S. and Preston, B. (2018). The science of monetary policy: An imperfect knowledge perspective. *Journal of Economic Literature*, 56(1):3–59.
- Evans, G. W. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press.
- Eyster, E., Madarasz, K., and Michailat, P. (2017). Pricing when Customers Care about Fairness but Misinfer Markups. *NBER Working Paper No. 23778*.
- Farhi, E. and Gabaix, X. (2017). Optimal Taxation with Behavioral Agents. *NBER Working Paper No. 21524*.
- Farhi, E. and Werning, I. (2017). Monetary Policy, Bounded Rationality, and Incomplete Markets. *NBER Working Paper No. 23281*.
- Fisher, J. D. (2015). On the Structural Interpretation of the Smets-Wouters ‘Risk Premium’ Shock. *Journal of Money, Credit and Banking*, 47(2-3):511–516.
- Fuhrer, J. (2017). Expectations as a source of macroeconomic persistence: Evidence from survey expectations in a dynamic macro model. *Journal of Monetary Economics*, 86:22–35.
- Fuhrer, J. C. and Rudebusch, G. D. (2004). Estimating the Euler equation for output. *Journal of Monetary Economics*, 51(6):1133–1153.
- Fuster, A., Hebert, B., and Laibson, D. (2012). Natural Expectations, Macroeconomic Dynamics, and Asset Pricing. *NBER Macroeconomics Annual*, 26(1):1–48.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*, 129(4):1661–1710.
- Gabaix, X. (2016). Behavioral macroeconomics via sparse dynamic programming. *NBER Working*

- Paper No. 21848.*
- Gabaix, X. (2017). Behavioral inattention. *NBER Working Paper No. 24096.*
- Gabaix, X. and Laibson, D. (2002). The 6D bias and the equity-premium puzzle. *NBER Macroeconomics Annual*, 16:257–312.
- Gabaix, X. and Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, 121(2):505–540.
- Gabaix, X. and Laibson, D. (2017). Myopia and discounting. *NBER Working Paper No. 23254.*
- Galí, J. (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and its Applications.* Princeton University Press.
- Galí, J. (2017). Monetary Policy and Bubbles in a New Keynesian Model with Overlapping Generations. *Working Paper.*
- Galí, J. and Gertler, M. (1999). Inflation Dynamics: A Structural Econometric Analysis. *Journal of Monetary Economics*, 44(2):195–222.
- Galí, J., López-Salido, J. D., and Vallés, J. (2007). Understanding the Effects of Government Spending on Consumption. *Journal of the European Economic Association*, 5(1):227–270.
- Ganong, P. and Noel, P. (2017). Consumer spending during unemployment: Positive and normative implications. *Unpublished manuscript, U. Chicago.*
- García-Schmidt, M. and Woodford, M. (2015). Are low interest rates deflationary? A paradox of perfect-foresight analysis. *NBER Working Paper 21614.*
- Greenwood, R. and Shleifer, A. (2014). Expectations of returns and expected returns. *The Review of Financial Studies*, 27(3):714–746.
- Herbst, E. P. and Schorfheide, F. (2015). *Bayesian Estimation of DSGE Models.* Princeton University Press.
- Johnson, D. S., Parker, J. A., and Souleles, N. S. (2006). Household Expenditure and the Income Tax Rebates of 2001. *The American Economic Review*, 96(5):1589–1610.
- Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108(3):697–743.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.
- Kiley, M. T. (2000). Endogenous price stickiness and business cycle persistence. *Journal of Money, Credit and Banking*, pages 28–53.
- Kiley, M. T. (2016). Policy Paradoxes in the New Keynesian Model. *Review of Economic Dynamics*, 21:1–15.
- Kocherlakota, N. (2016). Fragility of Purely Real Macroeconomic Models. *NBER Working Paper No. 21866.*
- Lindé, J. (2005). Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach. *Journal of Monetary Economics*, 52(6):1135–1149.
- Lubik, T. and Schorfheide, F. (2004). Testing for indeterminacy: an application to US monetary policy. *The American Economic Review*, 94(1):190–217.
- Maćkowiak, B. and Wiederholt, M. (2015). Business cycle dynamics under rational inattention. *Review of Economic Studies*, 82(4):1502–1532.
- Mankiw, N. G. (2000). The Savers-Spenders Theory of Fiscal Policy. *The American Economic Review Papers and Proceedings*, 90(2):120–125.
- Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the

- New Keynesian Phillips curve. *Quarterly Journal of Economics*, 117(4):1295–1328.
- Mankiw, N. G. and Weinzierl, M. (2011). An Exploration of Optimal Stabilization Policy. *Brookings Papers on Economic Activity*, 42(1 (Spring)):209–272.
- Mavroudis, S., Plagborg-Møller, M., and Stock, J. H. (2014). Empirical evidence on inflation expectations in the new keynesian phillips curve. *Journal of Economic Literature*, 52(1):124–188.
- McKay, A., Nakamura, E., and Steinsson, J. (2016). The Power of Forward Guidance Revisited. *The American Economic Review*, 106(10):3133–3158.
- Morris, S., Allen, F., and Shin, H. S. (2006). Beauty contests, bubbles and iterated expectations in asset markets. *The Review of Financial Studies*, 19(4):719–752.
- Morris, S. and Shin, H. S. (1998). Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks. *The American Economic Review*, pages 587–597.
- Nagel, R. (1995). Unraveling in Guessing Games: An Experimental Study. *The American Economic Review*, 85(5):1313–1326.
- Nakata, T., Schmidt, S., and Yoo, P. (2017). Attenuating the forward guidance puzzle: Implications for optimal monetary policy. *Working Paper*.
- Nimark, K. (2008). Dynamic pricing and imperfect common knowledge. *Journal of Monetary Economics*, 55(2):365–382.
- Reis, R. (2006). Inattentive consumers. *Journal of Monetary Economics*, 53(8):1761–1800.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.
- Smets, F. and Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *The American Economic Review*, 97(3):586–606.
- Taubinsky, D. and Rees-Jones, A. (2017). Attention variation and welfare: Theory and evidence from a tax salience experiment. *Forthcoming at the Review of Economic Studies*.
- Taylor, J. B. (1999). The Robustness and Efficiency of Monetary Policy Rules as Guidelines for Interest Rate Setting by the European Central Bank. *Journal of Monetary Economics*, 43(3):655–679.
- Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–30.
- Werning, I. (2012). Managing a Liquidity Trap: Monetary and Fiscal Policy. *NBER Working Paper 17344*.
- Werning, I. (2015). Incomplete Markets and Aggregate Demand. *NBER Working Paper No. 21448*.
- Woodford, M. (2003a). Imperfect common knowledge and the effects of monetary policy. In Aghion, P., Frydman, R., Stiglitz, J., and Woodford, M., editors, *Knowledge, information, and expectations in modern macroeconomics: In honor of Edmund S. Phelps*. Princeton University Press.
- Woodford, M. (2003b). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.
- Woodford, M. (2011). Simple Analytics of the Government Expenditure Multiplier. *American Economic Journal: Macroeconomics*, 3(1):1–35.
- Woodford, M. (2012). Inattentive valuation and reference-dependent choice. *Working Paper*.
- Woodford, M. (2013). Macroeconomic analysis without the rational expectations hypothesis. *Annual Reviews of Economics*, 5:303–346.
- Woodford, M. (2018). Monetary policy analysis when planning horizons are finite. *Forthcoming at NBER Macroeconomics Annual*.