# A Behavioral New Keynesian Model

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:34708481">http://nrs.harvard.edu/urn-3:HUL.InstRepos:34708481</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:34708481">http://nrs.harvard.edu/urn-3:HUL.InstRepos:34708481</a></td>
</tr>
</tbody>
</table>
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 policy and its impact. That myopia parameter in turn affects the power of monetary and fiscal policy in a microfounded general equilibrium.

A number of consequences emerge. First, 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. Second, 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 the agent is de facto myopic, this effect is muted. Third, the zero lower bound is much less costly than in the traditional model. Fourth, even with passive monetary policy, equilibrium is determinate, whereas the traditional rational model generates multiple equilibria, which reduce its predictive power. Fifth, 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. Sixth, the model is “neo-Fisherian” in the long run, but Keynesian in the short run – something that has proven difficult for other models to achieve: 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 the empirical evidence.
1 Introduction

This paper proposes a way to analyze what happens to 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, to incorporate 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: in the traditional model, rational agents are Ricardian and do not react to tax cuts. In the 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 zero lower bound (ZLB) is much less costly.

3. The model can explain the stability in economies stuck at the ZLB, something that is difficult to achieve in traditional models.

4. Equilibrium selection issues vanish in many cases: for instance, even with a constant nominal interest rate there is just one (bounded) equilibrium.

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 “nominal GDP 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.

---

1By “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.
Let me now expand on the above points.

**Fiscal policy.** In the traditional NK model, agents are fully rational. So Ricardian equivalence holds, and tax cuts have no impact. Hence, fiscal policy (i.e. lump-sum taxes, as opposed to government expenditure) has no impact. Here, in contrast, the agent is not Ricardian because he does not anticipate future taxes well. 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.

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

**Equilibrium determinacy.** When monetary policy is passive (e.g. constant interest rate, 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 not mired in the morass of equilibrium selection. In contrast, in the behavioral model there is just one (bounded) equilibrium: things are clean and definite theoretically.

**Economic stability.** Determinacy is not just purely a theoretical question. In the rational model, if the economy is stuck at the ZLB forever (or, I will argue, for a long period of time), the Taylor principle is violated (as the interest rate is pegged at 0%). So, the equilibrium is indeterminate: we could expect the economy to jump randomly from one period to the next. However, we do not see that in Japan since the late 1980s or in the Western world in the aftermath of the 2008 crisis (Cochrane 2015). This can be explained with the 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, Giannoni, Patterson (2015) and McKay, Nakamura and Steinsson (forth.). 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 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.

---

2However, I do need to rule out explosive equilibria.
Optimal policy changes qualitatively. With rational firm, the optimal commitment policy with rational agents demands “price level targeting” (which gives when GDP is trend-stationary, “nominal GDP targeting”): i.e. after a cost-push shock, the monetary policy should partially let inflation rise, but then create a deflation, so that eventually the price level (and nominal GDP) come back to their pre-shock trend. The reason is that the benefits of that commitment to being very tough in the future are high with rational firms. 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 (2015), highlight that in the strict (rational) NK model, a rise in interest rates (even temporary) leads to a rise in inflation (though this depends on which equilibrium is selected, leading to some cacophony in the dialogue). This is called the “neo-Fisherian” property. The property holds, in the behavioral model, in the long run: the long-run real rate is independent of monetary policy (Fisher neutrality holds in that sense). However, in the short run, raising rates does lower inflation and output, as in the Keynesian model. Cochrane (2015, p.1) summarizes the situation:

“If the Fed raises nominal interest rates, the [New Keynesian] model predicts that inflation will smoothly rise, both in the short run and long run. This paper presents a series of failed attempts to escape this prediction. Sticky prices, money, backward-looking Phillips curves, alternative equilibrium selection rules, and active Taylor rules do not convincingly overturn the result.”

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

Literature review. I build on the large New Keynesian literature, as summarized in Woodford (2003) and Galí (2015). I am also indebted to the identification of the “forward guidance puzzle” in Del Negro, Giannoni, Patterson (2015) and McKay, Nakamura and Steinsson (forth.). I discuss their proposed resolution (based on rational agents) below. I was also motivated by the paradoxes in the New Keynesian model outlined in Cochrane (2015).

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).

There are other ways to model bounded rationality, including related sorts of differential salience (Bordalo, Gennaioli, Shleifer 2016), rules of thumb (Campbell and Mankiw 1989), limited information updating (Caballero 1995, Gabaix and Laibson 2002, Mankiw and Reis 2002, Reis 2006), noisy signals (Sims 2003, Maćkowiak and Wiederholt forth., Woodford 2012). The sparsity approach aims at being tractable and fairly unified, as it applies to both microeconomic problems like basic consumer theory and Arrow-Debreu-style general equilibrium (Gabaix 2014), dynamic macroeconomics (Gabaix 2016) and public economics (Farhi and Gabaix 2015).

At any rate, this is the first paper to study how behavioral considerations affect forward guidance in the New Keynesian model, along Garcia-Schmidt and Woodford (2015). That paper was circulated simultaneously and offers very different modelling. Woodford (2013) explores non-rational expectations in the NK model, particularly of the learning type. However, he does not distill his rich analysis into something compact like the 2-equation NK model of Proposition 2.6 in this paper.

Section 2 presents basic model assumptions and derives its main building blocks, summarized in Proposition 2.6. Section 3 derives the positive implications of this model. Section 4 studies optimal monetary and fiscal policy with behavioral agents. Section 5 considers a slightly more complex version that handles backward looking terms (as in the old Keynesian models) and performs policy experiments with long run changes. Section 6 concludes.

Section 7 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 precisions.

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

2 A Behavioral Model

Let us recall the traditional NK model: there is no capital or government spending, so output equals consumption. Call $x_t = (C_t - C^n_t)/C^n_t = \hat{\epsilon}_t$ the output gap, where $C_t$ is actual consumption, and

---

4My notion of “behavioral” here is bounded rationality or cognitive myopia. I abstract from other interesting forces, like fairness (Eyster, Madarasz and Michaillat 2016) – they create an additional source of price stickiness. Also, I build the analysis on the new Keynesian framework. It would be interesting to build on a model on very different paradigms, e.g. old Keynesian paradigms (Barro and Grossman 1971, updated in Michaillat and Saez 2015) and for instance agent-based paradigms.

5The core of the forward guidance part of this paper was originally contained in what became Gabaix (2016), but was split off later.

6Two works circulated a year later pursue related themes. Farhi and Werning (2016) explore the interaction between bounded rationality and incomplete markets. Angeletos and Lian (2016) explore difference between direct and GE context, with an application to forward guidance. Their notion of “incomplete information” is a close cousin of the limited attention I study here, with different but related microfoundation.
$C_t^n$ (respectively $r_t^n$) is consumption (respectively the real interest rate) in an “underlying RBC economy” where all pricing frictions are removed (so that $r_t^n$ is the “natural interest rate”). The traditional NK model gives microfoundations that lead to:

$$x_t = E_t [x_{t+1}] - \sigma \left( i_t - E_t \pi_{t+1} - r_t^n \right)$$  \hspace{1cm} (1)$$
$$\pi_t = \beta E_t [\pi_{t+1}] + \kappa x_t.$$  \hspace{1cm} (2)

I now present other foundations leading to the behavioral model that has the traditional rational model as a particular case.

### 2.1 A Behavioral Agent: Microeconomic Behavior

In this section I present a way of modeling boundedly rational agents, drawing results from Gabaix (2016), which discusses more general sparse behavioral dynamic programming.

**Setup.** I consider an agent with standard utility

$$U = E \sum_{t=0}^{\infty} \beta^t u (C_t, N_t) \text{ with } u (C, N) = \frac{C^{1-\gamma} - 1}{1 - \gamma} - \frac{N^{1+\phi}}{1+\phi}$$

where $C_t$ is consumption, and $N_t$ is labor supply. The price level is $p_t$, the nominal wage is $w_t$, so that the real wage is $\omega_t = \frac{w_t}{p_t}$. The real interest rate is $r_t$ and agent’s real income is $y_t$ (it is labor income $\omega_t N_t$, profit income $\Pi_t$, and potential government transfers $T_t$, normally set to 0). His real financial wealth $k_t$ evolves as:

$$k_{t+1} = (1 + r_t) (k_t - c_t + y_t)$$  \hspace{1cm} (3)$$
$$y_t = \omega_t N_t + \Pi_t + T_t.$$  \hspace{1cm} (4)

The agent’s problem is $\max_{\{C_t, N_t\}_{t \geq 0}} U$ s.t. (3)-(4), and the usual transversality condition.

Consider first the case where the economy is deterministic, so that the interest rate, income, and real wage are at $\bar{r}$ (which I will soon call $r$), $\bar{y}$, and $\bar{\omega}$. Define $R := 1 + \bar{r}$ and assume that $R$ satisfies the macro equilibrium condition $\beta R = 1$. We have a simple deterministic problem, whose solution is: $c_t^d = \frac{rk_t}{R} + \bar{y}$, and labor supply is $g' (N) = \bar{\omega}$. If there are no taxes and profits, as in

### Notes

I change a bit the timing convention compared to Gabaix (2016): income innovation $\hat{y}_t$ is received at $t$, not $t+1$. The wealth $k_t$ is the wealth at the beginning of the period, before receiving any income.
the baseline model, $y = \bar{\omega} N$. In the New Keynesian model there is no aggregate capital, so that in most expressions we have $k_t = 0$ in equilibrium. Still, to track the consumption policy, it is useful to consider potential deviations from $k_t = 0$.

In general, there will be deviations from the steady state. I decompose the values as:

$$r_t = \bar{r} + \hat{r}_t, \ y_t = \bar{y} + \hat{y}_t, \ \omega_t = \bar{\omega} + \hat{\omega}_t, \ N_t = \bar{N} + \hat{N}_t.$$  

**Rational agent.** I first present a simple lemma describing the rational policy using Taylor expansions.\(^8\) To signify “up to the second order term,” I use the notation $O(\|x\|^2)$, where $\|x\|^2 := \mathbb{E}[\hat{y}_t^2]/\bar{y}^2 + \mathbb{E}[\hat{r}_t^2]/\bar{r}^2$ (the constants $\bar{y}, \bar{r}$ are just here to ensure units are valid).

**Lemma 2.1** (Traditional rational consumption function) *In the rational policy, optimal consumption is: $c_t = c^d_t + \hat{c}_t$, with $c^d_t = \frac{\bar{r}}{R} k_t + \bar{y}$ and

$$\hat{c}_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} \frac{1}{R^{\tau-t}} (b_r(k_t) \hat{r}_\tau + b_y \hat{y}_\tau) \right] + O(\|x\|^2),$$

$$b_r(k_t) := \frac{r}{R} k_t - \frac{1}{\gamma} c^d_t, \quad b_y := \frac{r}{R},$$

and labor supply is $\hat{N}_t / N = \frac{1}{\phi} \frac{\hat{\omega}_t}{\bar{\omega}} - \frac{\gamma}{\phi} \frac{\hat{c}_t}{c^d_t}$.\(^6\)

Consumption reacts to future interest rates and income changes according to the usual income and substitution effects (multiplied by $\frac{1}{\gamma}$). Note that here $\hat{y}_\tau$ is the change in total income, which includes the changes from endogenous (current and future) labor supply. To keep things tractable, it is best not to expand the endogenous $\hat{y}_\tau$ at this stage.

**Behavioral agent.** In the behavioral model, the agent is partially inattentive to the variable part of the interest rate, $\hat{r}_\tau$, and of income, $\hat{y}_\tau$. I detail the assumptions in Section 8.1. The behavioral policy is then as follows – see Gabaix (2016) for the derivation.

**Proposition 2.2** (Behavioral consumption function) *In the behavioral model, consumption is: $c_t = c^d_t + \hat{c}_t$, with $c^d_t = \frac{\bar{r}}{R} k_t + \bar{y}$ and

$$\hat{c}_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} \frac{\bar{m}}{R^{\tau-t+1}} (b_r(k_t) m_r \hat{r}_\tau + b_y m_y \hat{y}_\tau) \right] + O(\|x\|^2).$$

Labor supply is:
\[
\frac{\dot{N}_t}{N} = \frac{1}{\phi} \hat{\omega}_t - \gamma \frac{\hat{c}_t}{c_t^d}.
\]  
(8)

Parameters \(m_r, m_y, \bar{m}\) are attention parameters in \([0, 1]\). When they are all equal to 1, the agent is the traditional, rational agent. Here, \(m_r\) and \(m_y\) capture the attention to the interest rate and income, respectively. Parameter \(\bar{m}\) is a form of “cognitive discounting” – discounting future innovations more as they are more distant in the future.\(^9\)\(^10\)

There is mounting microeconomic evidence for the existence of inattention to macroeconomic variables (Coibon and Gorodnichenko 2015), taxes (Chetty, Looney and Kroft 2009, Taubinsky and Rees-Jones 2015), and, more generally, small dimensions of reality (Brown, Hossain and Morgan 2010, Caplin, Dean and Martin 2011). Those are represented in a compact way by the inattention parameters \(m_y, m_r\) and \(\bar{m}\). If the reader seeks maximum parsimony, I recommend setting \(m_r = m_y = 1\), and keeping \(\bar{m}\) as the main parameter governing inattention.

2.2 Behavioral IS curve: First Without Fiscal Policy

I start with a behavioral New Keynesian IS curve, in the case without fiscal policy. The derivation is instructive, and very simple. Proposition 2.2 gives:

\[
x_t = \frac{\hat{c}_t}{c_t^d} = \frac{1}{c_t^d} \mathbb{E}_t \left[ \sum_{\tau \geq t} \bar{m}^{\tau-t} \left( b_y m_y \hat{y}_\tau + b_r (k_t) m_r \hat{r}_\tau \right) \right].
\]  
(9)

The agent sees only partially the income innovations, with a dampening \(m_y\) for income and \(m_r\) for the interest rate. The cognitive dampening feature \(\bar{m}\) will mostly be useful for calibration purposes.

Now, since there is no capital in the NK model, we have \(\hat{y}_\tau = \hat{c}_\tau\): income is equal to aggregate demand. Conceptually, \(\hat{c}_\tau\) is the consumption of the other agents in the economy. Hence, using \(x_\tau = \frac{b_y}{R}\), we have with \(b_y = \frac{r}{R}\) and \(\bar{b}_r := \frac{b_r (k_t)_{k_t=0} m_r}{c_t^d} = -\frac{1}{2} \frac{m_r}{R^2}\), (9) becomes:

\[
x_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} \bar{m}^{\tau-t} \left( b_y m_y x_\tau + \bar{b}_r \hat{r}_\tau \right) \right].
\]  
(10)

\(^9\)See Gabaix (2016), Section 12.10 in the online appendix.

\(^{10}\)Here the labor supply comes from the first order condition. One could develop a more general version \(\frac{\dot{N}_t}{N} = \frac{\hat{\omega}_t - \gamma m_c N c_t^d}{\phi}\), where \(m_c^N \in [0, 1]\) is attention to consumption when choosing labor supply. When \(m_c^N = 0\), wealth effects are eliminated. Then, we have a behavioral microfoundation for the labor supply coming traditionally from the Greenwood, Hercowitz and Huffman (1988) preferences, \(u \left( C - \frac{N^{1+\phi}}{1+\phi} \right)\).
Taking out the first term yields:

\[ x_t = \frac{r}{R} m_y x_t + \tilde{b}_r \hat{r}_\tau + \mathbb{E}_t \left[ \sum_{\tau \geq t+1} \frac{\tilde{m}^{\tau-t}}{R^{\tau-t}} \left( b_y m_y x_{\tau} + \tilde{b}_r \hat{r}_\tau \right) \right]. \]

Given that (10), applied to \( t+1 \), yields \( x_{t+1} = \mathbb{E}_t \left[ \sum_{\tau \geq t+1} \frac{\tilde{m}^{\tau-t-1}}{R^{\tau-t-1}} \left( b_y m_y x_{\tau} + \tilde{b}_r \hat{r}_\tau \right) \right] \), we have:

\[ x_t = \frac{r}{R} m_y x_t + \tilde{b}_r \hat{r}_\tau + \frac{\tilde{m}}{R} \mathbb{E}_t \left[ x_{t+1} \right]. \]

Multiplying by \( R \) and gathering the \( x_t \) terms, we have:

\[ x_t = \tilde{m} \mathbb{E}_t \left[ x_{t+1} \right] + \frac{R \tilde{b}_r \hat{r}_\tau}{R - r m_y}. \]

Using \( M := \frac{\tilde{m}}{R - r m_y} \) and \( \sigma := -\frac{R \tilde{b}_r \hat{r}_\tau}{R - r m_y} = \frac{m_r / \gamma R}{R (R - r m_y)} \), we obtain the “discounted IS curve,” with discount \( M \):

\[ x_t = M \mathbb{E}_t \left[ x_{t+1} \right] - \sigma \hat{r}_t. \tag{11} \]

The next proposition records the result. The innovation in the interest rate is written in terms of the nominal rate,

\[ \hat{r}_t := i_t - \mathbb{E}_t [\pi_{t+1}] - r^n_t. \tag{12} \]

Here the natural interest rate \( r^n_t = r \) is the interest rate that would prevail in an economy with flexible price (but keeping cognitive frictions).\(^{11}\)

**Proposition 2.3 (Discounted Euler equation)** In equilibrium, the output gap \( x_t \) follows:

\[ x_t = M \mathbb{E}_t \left[ x_{t+1} \right] - \sigma \left( i_t - \mathbb{E}_t [\pi_{t+1}] - r^n_t \right), \tag{13} \]

where \( M := \frac{m}{R - r m_y} \in [0, 1] \) is a modified attention parameter, and \( \sigma := \frac{m_r}{\gamma R (R - r m_y)} \). In the rational model, \( M = 1 \).

Any kind of inattention (to aggregate variables via \( m_y \), cognitive discounting via \( \tilde{m} \)) creates \( M < 1 \). When the inattention to macro variables is the only force (\( \tilde{m} = 1 \)), then \( M \in \left[ \frac{1}{R}, 1 \right] \).

\(^{11}\)More precisely, this is the interest rate in a surrogate economy with no pricing friction, first best level of consumption and labor supply, but with the same cognitive frictions as in the original economy, and before government transfers to the agents. Section 8.2 details this.
Hence, the cognitive discounting gives a potentially powerful quantitative boost.\footnote{12}{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 (2015)), 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 $\sigma$).}

The behavioral NK IS curve (23) implies:

$$x_t = -\sigma \sum_{\tau \geq t} M^{t - \tau} \mathbb{E}_t [\hat{r}_\tau]$$

(14)

i.e. it is the discounted value of future interest rates that matters, rather than the undiscounted sum. This will be important when we study forward guidance below.

McKay, Nakamura and Steinsson (forth.) find that this equation fits better. They provide a microfoundation based on heterogeneous rational agents with limited risk sharing. In their model, wealthy, unconstrained agents with no unemployment risk would still satisfy the usual Euler equation. Werning (2015)’s analysis yields a modified Euler equation with rational heterogeneous agents, which often yields $M > 1$. Piergallini (2006), Nistico (2012), and Del Negro, Giannoni and Patterson (2015) offer microfoundations with heterogeneous mortality shocks, as in perpetual-youth models (this severely limits how myopic agents can be, given that life expectancies are quite high).\footnote{13}{Relatedly, Fisher (2015) derives a discounted Euler equation with a safe asset premium: but the effect is very small, e.g. the coefficient $1 - M$ is very close to 0 – it is the empirically very small “safety premium”.}

Caballero and Farhi (2015) offer a different explanation of the forward guidance puzzle in a model with endogenous risk premia and a shortage of safe assets (see also Caballero, Farhi and Gourinchas 2015).

My take, in contrast, is behavioral: the reason that forward guidance does not work well is that it is in some sense “too subtle” for the agents. In independent work, Garcia-Schmidt and Woodford (2015) offer another, distinct, behavioral take on the NK IS curve.

**Understanding discounting in rational and behavioral models.** It is worth pondering where the discounting comes from in (14). What is the impact at time 0 of a one-time fall of the real interest rate $\hat{r}_\tau$, in partial and general equilibrium, in both the rational and the behavioral model?

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\footnote{14}{See equation (5). I use continuous time notation, so replace $\frac{1}{\gamma R^2}$ by $\frac{1}{\gamma}$ to unclutter the analysis. I take the case without capital.}

\[ \text{Rational agent: } \hat{c}_0^{\text{direct}} = -\frac{\hat{r}_\tau}{\gamma R}. \]
Hence, there is discounting by $\frac{1}{R^\tau}$. However, in general equilibrium, the impact is (see (14) with $M = 1$)

$$\text{Rational agent: } \hat{c}_0^{\text{GE}} = -\frac{\hat{r}_\tau}{\gamma}$$

so that there is no discounting by $\frac{1}{R^\tau}$. The reason is the following: the rational agent sees the “first round of impact”, $-\frac{\hat{r}_\tau}{\gamma R^\tau}$: 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 other all rounds (as in the Keynesian cross, e.g. equation (56)), the initial impulse is greatly magnified: though the first round (direct) impact is $-\frac{\hat{r}_\tau}{\gamma R^\tau}$, the full impact (including indirect channels) is $-\frac{\hat{r}_\tau}{\gamma}$. This means that the total impact is larger than the direct effect by a factor

$$\frac{\hat{c}_0^{\text{GE}}}{\hat{c}_0^{\text{direct}}} = R^\tau.$$ 

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:

$$\text{Behavioral agent: } \hat{c}_0^{\text{direct}} = -m_r \bar{m} \frac{\hat{r}_\tau}{\gamma R^\tau}$$

which comes from (7). Next, the agent is not fully attentive to indirect effects (including general equilibrium) of future policies. This results in the total effect in (14):

$$\text{Behavioral agent: } \hat{c}_0^{\text{GE}} = -m_r M \frac{\hat{r}_\tau}{\gamma}$$

with $M = \frac{\bar{m}}{R - rm_y}$. So the multiplier for general equilibrium effect is:

$$\frac{\hat{c}_0^{\text{GE}}}{\hat{c}_0^{\text{direct}}} = \left( \frac{R}{R - rm_y} \right)^\tau \in [1, R^\tau].$$

and is smaller than the multiplier $R^\tau$ in economies with common knowledge of rationality.
2.3 Behavioral IS Curve with Fiscal Policy

In this subsection I generalize the above IS curve to the case of an active fiscal policy. The reader is encouraged to skip this section at first reading. I call $B_t$ the real value of government debt in period $t$, before period-$t$ taxes. It evolves as $B_{t+1} = \frac{1+i_{t+1}}{1+\pi_{t+1}} (B_t + \mathcal{T}_t)$ where $\mathcal{T}_t$ is the lump-sum transfer given by the government to the agent (so that $-\mathcal{T}_t$ is a tax), and $\frac{1+i_{t+1}}{1+\pi_{t+1}}$ is the realized gross real interest rate.\(^{15}\) Here, I take the Taylor expansion, neglecting the variations of the real rate (i.e. second-order terms $O \left( \left| \frac{1+i_{t+1}}{1+\pi_{t+1}} - R \right| \left( |B_t| + |d_t| \right) \right)$ around $R$.\(^{16}\) Hence, debt evolves as:

$$B_{t+1} = R (B_t + \mathcal{T}_t).$$

I also define $d_t$, the budget deficit (after the payment of the interest rate on debt) in period $t$:

$$d_t := \mathcal{T}_t + r B_t$$

so that public debt evolves as:\(^{17}\)

$$B_{t+1} = B_t + Rd_t.$$  

Iterating gives $B_u = B_t + R \sum_{u=t}^{\tau-1} d_u$, so that the transfer at time $\tau$, $\mathcal{T}_\tau = -\frac{r}{R} B_\tau + d_\tau$ is:

$$\mathcal{T}_\tau = -\frac{r}{R} B_t + \left( d_\tau - r \sum_{u=t}^{\tau-1} d_u \right). \quad (15)$$

This equation (15) is the objective law of motion of the transfer. The general formalism gives the following behavioral version of that equation, as perceived by the agent:\(^{18}\)

$$\mathbb{E}_t^{BR} [\mathcal{T}_\tau] = -\frac{r}{R} B_t + m_y \bar{m}^{\tau-t} \left( d_\tau - r \sum_{u=t}^{\tau-1} d_u \right). \quad (16)$$

This reflects a partially rational consumer. Given initial debt $B_t$, the consumer will see that it will have to be repaid: he accurately foresees the part $\mathbb{E}_t^{BR} [\mathcal{T}_\tau] = -\frac{r}{R} B_t$. If there were no future deficits, he would be rational. However, he sees future deficits dimly, those that come beyond the service of public debt. This is captured by the term $m_y \bar{m}^{\tau-t}$.

\(^{15}\)The debt is short-term. Debt maturity choice is interesting, but well beyond the scope of this paper.

\(^{16}\)That is, I formally consider the case of “small” debt.

\(^{17}\)Indeed, $B_{t+1} = R (B_t - \frac{r}{R} B_t + d_t) = B_t + Rd_t$.

\(^{18}\)See the derivation of Proposition 2.4 for details.
Consumption satisfies, again from Proposition 2.2:

\[ x_t = \mathbb{E}_t^{BR} \left[ \sum_{\tau \geq t} \frac{1}{R^{\tau-t}} \left( b_y(x_\tau + \tau) + b_r(k_\tau) \hat{r}_\tau \right) \right] + \frac{r}{R} k_t \]

where \( \mathbb{E}_t^{BR} \) is the expectation under the subjective model and in equilibrium \( k_t = B_t \).

Calculations in the Appendix give the following modifications of the IS curve. Note that here we only have deficits, not government consumption.

**Proposition 2.4 (Discounted Euler equation with sensitivity to budget deficits)** We have the following IS curve reflecting the impact of both fiscal and monetary policy:

\[ x_t = M \mathbb{E}_t [x_{t+1}] + b_d d_t - \sigma (i_t - \mathbb{E}_t \pi_{t+1} - r^*_{t}) \]  \hfill (17)

where \( d_t \) is the budget deficit and

\[ b_d = \frac{r m_y}{R - m_y r} \frac{R (1 - \bar{m})}{R - \bar{m}} \]  \hfill (18)

is the sensitivity to deficits. When agents are rational, \( b_d = 0 \), but with behavioral agents, \( b_d > 0 \). The values of \( M \) and \( \sigma \) are as in Proposition 2.3.

Hence, bounded rationality gives both a discounted IS curve and an impact of fiscal policy.

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-Mankiw (1989), Gali, López-Salido and Vallés (2007), Mankiw (2000), Bilbiie (2008), Mankiw and Weinzierl (2011) and Woodford (2013).\(^ {19,20} \) When dealing with complex situations, a representative agent is often simpler. In particular, it allows us to value assets unambiguously.\(^ {21} \)

\(^ {19} \text{Gali, López-Salido and Vallés (2007) is richer and more complex, as it features heterogeneous agents. Omitting the monetary policy terms, instead of } x_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} \frac{M^{\tau-t}}{R^{\tau-t}} b_d \right] \text{ (see (17)), they generate } x_t = \Theta_n n_t - \Theta_r t^r_t \text{, where } t^r_t \text{ are the rational agents. Hence, one key difference is that in the present model, the future deficits matter as well, whereas in their model, they do not.}\)

\(^ {20} \text{Mankiw and Weinzierl (2011) have a form of the representative agent with a partial rule of thumb behavior. They derive an instructive optimal policy in a 3-period model with capital (which is different from the standard New Keynesian model), but do not analyze an infinite horizon economy. Another way to have non-Ricardian agents is via rational credit constraints, as in Kaplan, Moll and Violante (2016). The analysis is then rich and complex.}\)

\(^ {21} \text{With heterogeneous agents and incomplete markets, there is no agreed-upon way to price assets: it is unclear “whose pricing kernel” one must take.}\)
2.4 Phillips Curve with Behavioral Firms

Next, I explore what happens if firms do not fully pay attention to future macro variables either. The reader may wish to skip this section upon the first reading, as this is less important (though it will be important for policy).

I assume that firms are partially myopic to the value of future markup. To do so, I adapt the classic derivation (e.g. chapter 3 of Galí (2015)) for behavioral agents. Firms can reset their prices with probability $1 - \theta$ each period. The general price level is $p_t$, and a firm that resets its price at $t$ sets a price $p_t^*$ according to:

$$p_t^* - p_t = (1 - \beta\theta) \sum_{\tau \geq t} (\beta\theta\bar{m})^{\tau-t} m^f t \mathbb{E}_t [\psi_{\tau} - p_t]$$

i.e. the price is equal to the present value of future marginal costs $\psi_{\tau} - p_t$. The log marginal cost is simply the log nominal wage (if productivity is constant), $\psi_{\tau} = \ln (\omega_{\tau} P_{\tau})$. Here $m^f \in [0, 1]$ indicates the imperfect attention to future markup innovations. Parameters $m^f$ and $\bar{m}$ as in some sense the “level” and “slope” of the cognitive discounting by firms (which is in $m^f \bar{m}^{\tau-t}$ at horizon $\tau - t$).\(^{22}\) Otherwise, the setup is as in Galí. When $\bar{m} = m^f = 1$, we have the traditional NK framework.

Tracing out the implications of (19), the macro outcome is as follows.\(^{23}\)

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

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

with the attention coefficient $M^f$:

$$M^f := \bar{m} \left[ \theta + (1 - \theta) \frac{1 - \beta\theta}{1 - \beta\theta \bar{m}} m^f \right]$$

and

$$\kappa = \bar{\kappa} m^f$$

where $\bar{\kappa}$ (given in (71)) is independent of attention. Firms are more forward-looking in their pricing

\(^{22}\)I could write $\bar{m}^f$ in (19) rather than $\bar{m}$, at it is the cognitive discounting of firms rather than workers. I didn’t do that do avoid multiplying notations, but the meaning should be clear. In particular, in (21), it is the $\bar{m}$ of firms that matters, not that of consumers.

\(^{23}\)The proof is in the appendix. Here I state the Proposition for the case where firms are constant return to scale ($\alpha = 0$). The expressions are similar (and available upon request) in the case where $\alpha \in [0, 1)$.
(\text{MF is higher}) when prices are sticky for a longer time (\theta is higher) and when firms are more attentive to future macroeconomic outcomes (m^f,\bar{m} are higher). When m^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 (20) is exactly \beta and, miraculously, does not depend on the adjustment rate of prices \theta. In the behavioral model, in contrast, the coefficient (\beta M^f) is higher when prices are stickier for longer (higher \theta).

Firms can be fully attentive to all idiosyncratic terms (something that will be easy to include in a future version of the paper), e.g. the idiosyncratic part of the productivity of demand. They simply have to pay attention \text{MF} to macro outcomes. If we include idiosyncratic terms, and firms are fully attentive to them, the aggregate NK curve does not change.

The behavioral elements simply change \beta into \beta M^f, where \text{MF} \leq 1 is an attention parameter. Empirically, this “extra discounting” (replacing \beta by \beta M^f in (20)) seems warranted, as we shall see in the next section.

Let me reiterate that firms are still forward-looking (with discount parameter \beta rather than \beta M^f) in the deterministic steady state. It is only their sensitivity to deviations around the deterministic steady state that is partially myopic.

2.5 Synthesis: Behavioral New Keynesian Model

I now gather the above results.

**Proposition 2.6** (Behavioral New Keynesian model – two equation version) We obtain the following behavioral version of the New Keynesian model, for the behavior of output gap \text{x}_t and inflation \text{\pi}_t:

\begin{align*}
\text{x}_t &= M \mathbb{E}_t [\text{x}_{t+1}] + b_d d_t - \sigma (i_t - \mathbb{E}_t \pi_{t+1} - \pi^n_t) \quad (IS \text{ curve}) \\
\text{\pi}_t &= \beta M^f \mathbb{E}_t [\text{\pi}_{t+1}] + \kappa \text{x}_t \quad (Phillips \text{ curve})
\end{align*}

(23)

(24)

where \text{M, M}^f \in [0, 1] are the attention of consumers and firms, respectively, to macroeconomic
outcomes, and \( b_d \geq 0 \) is the impact of deficits:

\[
M := \frac{\bar{m}}{R - rm_y} \quad M^f := \bar{m} \left[ \theta + (1 - \theta) \frac{1 - \beta \theta}{1 - \beta \bar{m}} m^f \right] \\
\]

\[ b_d = \frac{rm_y}{R - m_y r} \frac{R(1 - \bar{m})}{R - \bar{m}}. \]

In the traditional model, \( \bar{m} = m_y = m_r = 1 \), so that \( M = M^f = 1 \) and \( b_d = 0 \). In addition, \( \sigma := \frac{m_r \psi}{R(R - rm_y)} \), and \( \kappa = \bar{\kappa} m^f \), where \( \bar{\kappa} \) (given in 71) is independent of attention.

**Empirical Evidence on the Model’s Deviations from Pure Rationality.** The empirical evidence, we will now see, 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 very forward looking. For instance, Galí and Gertler (1999) find that we need \( \beta M^f \simeq 0.75 \) at the annual frequency; given that \( \beta \simeq 0.95 \), that leads to an attention parameter of \( M^f \simeq 0.8 \). If we have \( \theta = 0.2 \) (so that 80% of prices are reset after a year) and \( \bar{m} = 1 \), then this corresponds to \( m^f \simeq 0.75 \).

In the Euler equations consumers do not appear to be fully forward looking: \( M < 1 \). The literature on the forward guidance puzzle concludes, plausibly I think, that \( M < 1 \).

Ricardian equivalence does not fully hold. There is much debate about Ricardian equivalence. The provisional median opinion is that it only partly holds. For instance, the literature on tax rebates (Johnson, Parker and Souleles 2006) appears to support \( b^d > 0 \).

All three facts come out naturally from a model with cognitive discounting \( \bar{m} < 1 \), even without the auxiliary parameters \( m_y, m_r, m^f \). Those could be set to 1 (the rational value) in most cases.

**Continuous time version.** In continuous time, we write \( M = 1 - \xi \Delta t \) and \( \beta M^f = 1 - \rho \Delta t \).

In the small time limit \( (\Delta t \to 0) \), \( \xi \geq 0 \) is the cognitive discounting parameter due to myopia in the continuous time model, while \( \rho \) is the discount rate inclusive of firm’s myopia. The model (23) becomes, in the continuous time version:

\[
\dot{x}_t = \xi x_t - b_d d_t + \sigma (i_t - r_t - \pi_t) \\
\dot{\pi}_t = \rho \pi_t - \kappa x_t.
\]
When $\xi = b_d = 0$, we recover Werning (2012)’s formulation, in which agents are all rational.\footnote{Note that the way to read (25) is as a forward equation. Call $\hat{r}_t := i_t - r_t - \pi_t$ the interest rate gap. We take into account the fact that the model is stationary ($\lim_{t \to \infty} x_t = 0$). We have: $\dot{x}_t = -\sigma \int_{t}^{\infty} e^{-\xi(s-t)} \hat{r}_s ds$. Hence, the effect of future rate changes is dampened by myopia.}

We now study several consequences of these modifications for the forward guidance puzzle.

# 3 Consequences of this Behavioral Model

## 3.1 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.\footnote{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.}

I assume that the central bank follows a Taylor rule of the type:

$$i_t = \phi_\pi \pi_t + \phi_x x_t + j_t$$

(27)

where $j_t$ is typically just a constant.\footnote{The reader will want to keep in mind the case of a constant $j_t = \bar{j}$. More generally, $j_t$ is a function $j_t = j(Z_t)$ where $Z_t$ is 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 (captured by $Z_t$).}

I next express Proposition 2.6 with the notations: $z_t = (x_t, \pi_t)'$, forcing variables $a_t := j_t - r^n_t$, $m = (M, M^f)$, $\beta^f := \beta M^f$. For simplicity, I assume an inactive fiscal policy, $d_t = 0$.\footnote{Given a rule for fiscal policy, the sufficient statistic is the behavior of the “monetary and fiscal policy mix” $i_t - b_d d_t$. For instance, suppose that: $i_t - b_d d_t := \phi_\pi \pi_t + \phi_x x_t + j_t$, with some (unimportant) decomposition between $i_t$ and $d_t$. The analysis is then the same. More general analyses might add the total debt $D_t$ as a state variable in the rule for $d_t$.}

Calculations yield:\footnote{It is actually easier (especially when considering higher-dimensional variants) to proceed with the matrix $(A(m))^{-1}$, and a system $E_t [z_{t+1}] = A(m)^{-1} z_t + \tilde{b}(m) a_t$, and to reason on the roots of $(A(m))^{-1}$:}

$$z_t = A(m) \mathbb{E}_t [z_{t+1}] + b(m) a_t$$

(28)

\[
(A(m))^{-1} = \frac{1}{M \beta^f} \begin{pmatrix}
\beta^f (1 + \sigma \phi_\pi) + \kappa \sigma & -\sigma (1 - \beta^f \phi_\pi)
\end{pmatrix} \begin{pmatrix}
\beta^f (1 + \sigma \phi_\pi) + \kappa \sigma \\
-\kappa M
\end{pmatrix}.
\]
with

\[ A(m) = \frac{1}{1 + \sigma (\phi_x + \kappa \phi_\pi)} \left( \begin{array}{cc} M & \sigma (1 - \beta f \phi_\pi) \\ \kappa M & \beta f (1 + \sigma \phi_x) + \kappa \sigma \end{array} \right) , \]  
\[ b(m) = \frac{-\sigma}{1 + \sigma (\phi_x + \kappa \phi_\pi)} (1, \kappa)' , \]

\[ a_t = j_t - \rho_t. \]

The next proposition generalizes the well-known Taylor stability criterion to behavioral agents:

**Proposition 3.1** (Equilibrium determinacy with behavioral agents) There is a unique equilibrium (all of \( A \)‘s eigenvalues are less than 1 in modulus) if and only if:

\[ \phi_\pi + \frac{(1 - \beta f)}{\kappa} \phi_x + \frac{(1 - \beta f)}{\kappa \sigma} (1 - M) > 1. \]  

*In continuous time, the criterion (30) becomes: \( \phi_\pi + \frac{\rho}{\kappa} \phi_x + \frac{\rho \kappa}{\kappa \sigma} > 1. \)*

In particular, when monetary policy is passive (i.e., when \( \phi_\pi = \phi_x = 0 \)), we have a stable economy\(^{29}\) if and only if bounded rationality is strong enough, in the sense that

\[ \frac{(1 - \beta f)}{\kappa \sigma} (1 - M) > 1 \]  

in the discrete time version.\(^{30}\)

Condition (31) does not hold in the traditional model, where \( M = 1 \). The condition basically means that agents are boundedly rational enough, that is \( M \) is sufficiently less than 1 – and the pricing frictions are large enough.\(^{31}\)

Condition (31) implies that the two eigenvalues of \( A \) are less than 1.\(^{32}\) This implies that the equilibrium is determinate.\(^{33}\) This is different from the traditional NK model, in which there is a

\(^{29}\)So, \( \rho(A(m)) < 1 \), where \( \rho(V) \) is the the “spectral radius” of a matrix \( V \), i.e. the maximum of the modulus of its eigenvalues. If \( \rho(V) < 1 \), then \( \sum_{k=1}^{\infty} V^k \) converges.

\(^{30}\)In the continuous time version this condition is:

\[ \frac{\rho \kappa}{\kappa \sigma} > 1. \]  

\(^{31}\)As the frequency of price changes becomes infinite, \( \kappa \to 0 \) (see equation (71)). 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.

\(^{32}\)Indeed, (31) is equivalent to \( \phi(1) > 0 \), so both eigenvalues are either below or above 1. Given the sum of the two eigenvalues is \( \lambda_1 + \lambda_2 = \beta + M < 2 \), this implies that both eigenvalues are below 1.

\(^{33}\)The condition does not prevent unbounded or explosive equilibria, the kind that Cochrane (2011) wrestles with.
continuum of non-explosive monetary equilibria, given that one root is greater than 1 (as condition (31) is violated in the traditional model).

This absence of multiple equilibria is important. Indeed, take a central bank following a deterministic (e.g. constant) interest path – for instance in a period of a prolonged ZLB. Then, in the traditional model, there is always a continuum of (bounded) equilibria, technically, because matrix $A(m)$ 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 equilibrium.

Then, we can simply write, with $A(m)$:

$$z_t = \mathbb{E}_t \left[ \sum_{\tau \geq t} A(m)^{\tau-t} b(m) a_\tau \right]. \quad (33)$$

The detailed solution is in Section 10.3 of the Appendix.\(^{34}\)

### 3.2 Forward Guidance Is Much Less Powerful

Suppose that the central bank announces at time 0 that it will cut the rate at time $T$, following a policy $\delta_t = 0$ for $t \neq T$, $\delta_T < 0$, where $\delta$ is the interest rate gap. What is the impact? This is the thought experiment analyzed by McKay, Nakamura and Steinsson (forth.) with rational agents, which I pursue here with behavioral agents.

Figure 1 illustrates the effect. In the left panel, the whole economy is rational. 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 agents. We also see how the bounded rationality of both firms and consumers is useful for the effect.

Formally, we have $x_t = M x_{t+1} - \sigma \delta_t$, so $x_t = -\sigma M^{T-t} \delta_T$ for $t \leq T$ and $x_t = 0$ for $t > T$. This implies that inflation is:

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

\(^{34}\)My take is that this issue is interesting (as are rational bubbles in general), but that the largest practical problem is to eliminate bounded equilibria. The behavioral model does that well.

\(^{34}\)Here, I rule out any extra “explosive bubble” term.
Figure 1: This Figure shows the response of current inflation to forward guidance about interest rate in $T$ periods, compared to an immediate rate change of the same magnitude. Units are yearly. 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 that (annualized) attention is $M = M^f = e^{-\xi} = 0.7$ in the behavioral model, and $M = 1$ in the traditional model.

where $\beta^f := \beta M^f$ is the discount factor adjusted for firms’ inattention. A rate cut in the very distant future has a powerful impact on today’s inflation ($\lim_{T \to \infty} \pi_0(T) = -\kappa \sigma$) in the rational model ($M = 1$), and no impact at all in the behavioral model ($\lim_{T \to \infty} \pi_0(T) = 0$ if $M < 1$).

When attention is endogenous, the analysis could become more subtle. Indeed, if other agents are more attentive to the forward Fed announcement, 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.

3.3 The ZLB is Less Costly with Behavioral Agents

What happens when economies are at the ZLB? The rational model makes very stark predictions, which the 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 $r^a_t = \underline{r}$ for $t \leq T$, and $r^a_t = \bar{r}$ for $t > T$, with $\underline{r} < 0 < \bar{r}$. I assume that for $t > T$, the central bank implements $x_t = \pi_t = 0$ by setting $i_t = \bar{r}$. At time $t < T$, I suppose that the CB is at the ZLB, so that $i_t = 0$.

**Proposition 3.2** *In the traditional rational case ($\xi = 0$), we obtain an unboundedly intense recession as the length of the ZLB increases: \( \lim_{t \to -\infty} x_t = -\infty \). This also holds when myopia is mild, \( \frac{\rho \kappa}{\sigma \kappa} \leq 1 \). However, suppose cognitive myopia is strong enough (\( \frac{\rho \kappa}{\sigma \kappa} > 1 \)), which is the continuous-time version of condition (31). Then, we obtain a boundedly intense recession: \( \lim_{t \to -\infty} x_t = \frac{\rho \sigma r}{\rho \kappa - \sigma \kappa} < 0 \).*
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.

Figure 2 shows the dynamics. 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 (set to an annualized rate of $M = e^{-\xi} = 0.7$) in the behavioral model (against its value $M = 1$ in the traditional model). In the left panel, we see how costly the ZLB is (mathematically it is unboundedly costly as it becomes more long-lasting), while in the right panel we see a finite, though prolonged cost. Reality looks more like the prediction of the behavioral model (right panel) – something like Japan since the 1990s – rather than the prediction of the rational model (left panel) – which is something like Japan in 1945-46 or Rwanda.

I note that 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. Indeed, consider criterion (31).

The other parameters are: $\rho = 3\%$, $\kappa = 0.1$, $\sigma = 0.2$, $r = -5\%$. 

\footnote{35}{The other parameters are: $\rho = 3\%$, $\kappa = 0.1$, $\sigma = 0.2$, $r = -5\%$.}
4 Optimal Monetary and Fiscal Policy

4.1 Welfare with Behavioral Agents and the Central Bank’s Objective

Welfare is the expected utility of the representative agent, \( \tilde{W} = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, N_t) \). This is the traditional welfare measure taken over the objective expectations. This is in accordance to the typical practice in behavioral economics, which views behavioral agents as using heuristics, but have experience utility from consumption and leisure like rational agents. Following again the literature, I do a Taylor expansion, so that \( \tilde{W} = W^* + W \), where \( W^* \) is first best welfare, and \( W \) is the deviation from the first best. The next lemma calculates it.

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

\[
W = -KE_0 \sum_{t=0}^{\infty} \frac{1}{2} \beta^t (\pi_t^2 + \vartheta x_t^2) + W_-
\]

where

\[
\vartheta = \frac{\kappa}{m^f \varepsilon}
\]

\( K = u_c c(\gamma + \phi) \frac{\kappa}{m^f} \), and \( W_- \) is a constant (explicitied in (76)), and \( m^f \in (0, 1] \) is firms’ attention to the macro determinants of the markup, \( \kappa \) is Phillips curve coefficient, and \( \varepsilon \) is the elasticity of demand. In particular, controlling for the value of \( \kappa \), the relative weight on the output gap (\( \vartheta \)) is higher when firms are more behavioral (when \( m^f \) is lower).

The traditional model gives a very small relative weight \( \vartheta \) on the output gap when it is calibrated – this is often considered a puzzle. Here we obtain a larger weight – this is, a weight that is larger, conditional on a measure of \( \vartheta \). When firms react less to changes in inflation when setting their prices (when \( m^f \) is lower), transitory inflation is less important for price setting, hence it is less distortive for allocation and welfare: so, inflation receives a lower weight in the objective function (so the relative weight on output (\( \vartheta \)) is higher).

\[^{36}]In particular I use the objective (not subjective) expectations. Also, I do not count thinking costs in the welfare. 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 (2015), we are in the “no attention in welfare” case.
4.2 Optimal policy: Response to Changes in the Natural Interest Rate

4.2.1 When the ZLB doesn’t bind: Monetary policy attains the first best

Suppose that there is a productivity or discount factor shocks (they are not explicitly in the basic model, but can be introduced straightforwardly). This changes the natural real interest rate, $r^n_t$.\(^{37}\)

With rational and behavioral agents, the optimal policy is still to set $i_t = r^n_t$, i.e. to make the nominal rate track the natural real rate.\(^{38}\) This is consistent with $x_t = \pi_t = 0$ at all dates. We obtain the first best.\(^{39, 40}\)

This is the traditional, optimistic message in monetary policy. However, when the natural rate becomes negative (and with low inflation), the optimal nominal interest rate is negative, which is by and large not possible.\(^{41}\) That is the ZLB. Then, much research has shown that the policy is quite complex then. However, I now show how the policy becomes (in theory) easy and simple with rational agents.

4.2.2 When the ZLB binds: “Helicopter drops of money” as an optimal cure in the Optimal Mix of Fiscal and Monetary Policy

Now I explore how behavioral agent change a lot policy at the ZLB, and indeed use the model’s ability to have non-trivial monetary and fiscal policy. By “fiscal policy” I mean transfers (from the government to the agents), and “helicopter drops of money”, i.e. checks that the central bank might send (this gives some fiscal authority to the central bank).\(^{42}\)

To make the point, I suppose that we have a “crisis period” $I = (T_1, T_2)$, with $r^n_t < 0$ during that period, so that the ZLB binds. But $r^n_t > 0$ outside that period. With monetary policy only, the situation is dire, and we lose the first best.\(^{43}\) However, with fiscal policy and behavioral agents,

\(^{37}\)Behavioral biases modulate the way TFP shocks change the natural interest rate, as for instance they affect the effective intertemporal elasticity of substitution (see Section 8.2).

\(^{38}\)If the inflation target was $\bar{\pi}$, the nominal rate would be real rate plus inflation target $i_t = r^n_t + \bar{\pi}$. Throughout I assume $\bar{\pi} = 0$ for simplicity.

\(^{39}\)As is well understood, to ensure equilibrium determinacy, the central bank imbeds this in-sample policy into a more general rule, e.g. sets $i_t = r^n_t + \phi_\pi \pi_t + \phi_x x_t$ with coefficients $\phi_\pi, \phi_x$ sufficiently large (following (30). On the equilibrium path, $\pi_t = x_t = 0$, so that $i_t = r^n_t$.

\(^{40}\)If there are budget deficits, the central bank must “lean against behavioral biases”. 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 (in the IS curve (23)) by setting: $i_t = r^n_t + \frac{b_L}{\sigma} d_t$.

\(^{41}\)Recent events have seen nominal rates slightly below 0%, but it does not seem possible to obtain very low nominal rates, say -5%, for long, because stockpiling cash in a vault is then a viable alternative.

\(^{42}\)The central bank could also rebate the “seignorage check” to the taxpayers rather than the government, and write bigger checks at the ZLB, and smaller checks outside the ZLB.

\(^{43}\)The first best is not achievable, as been analyzed by a large number of authors, e.g. Eggertson and Woodford (2003), Eggertson and Krugman (2012), Werning (2012) and the survey in Gali (2015, section 5.4).
the first best can be restored.

**Proposition 4.2** (Optimal mix of fiscal and monetary policy in a ZLB environment). *The following monetary and fiscal policies yield the first best (x_t = π_t = 0) at all dates.* During the crisis (t ∈ (T_1, T_2)), use fiscal policy  

\[ d_t = -\frac{σr^m_t}{b_d}, \]

i.e. run a deficit with low interest rates, i_t = 0. After the crisis (t ≥ T_2), pay back the accumulated debt by running a government fiscal surplus and keeping the economy afloat with low rates, e.g.  

\[ d_t = R^{-1}(B_{T_2} - B_0)(1 - ρ_d)ρ_d^{t-T_2} < 0 \]

for some ρ_d ∈ (0, 1), and adjust i_t = \frac{b_d d_t}{σ} < 0 to ensure full macro stabilization, x_t = π_t = 0. Before the crisis (t < T_1), there is no preventive action to do, so set i_t = d_t = 0.

**Proof.** The proof is simply by examination of the basic equations of the NK model, (23)-(24). We adjust the instruments so that x_t = π_t = 0 at all dates. Note that there are multiple ways to soak up the debt after the crisis, so that  

\[ d_t = R^{-1}(B_{T_2} - B_0)(1 - ρ_d)ρ_d^{t-T_2} \]

is simply indicative. □

The ex-ante preventive benefits of potential ex-post fiscal policy. Proposition 4.2 shows that “the possibility of fiscal policy as ex-post cure produces ex-ante benefits”. Imagine that fiscal policy is not available. Then, the economy is depressed at the ZLB during (T_1, T_2). However, it is also depressed before: because the IS curve is forward looking, output threatens to be depressed before T_1, and that can put the economy to the ZLB at a time T_0 before T_1. Hence, the threat of a ZLB-depression in (T_1, T_2) creates an earlier recession at (T_0, T_2) with T_0 < T_1. Intuitively, agents feel “if something happens, monetary policy will be impotent, so large dangers loom”. However, if the government has fiscal policy in its arsenal, the agents feel “worse case, the government will use fiscal policy, so there is no real threat”, and there is no recession in (T_0, T_1). Hence, there is a possibility of fiscal policy as an ex-post cure to produce ex-ante benefits.

In general, monetary and fiscal policies are substitutes (d_t and i_t enter symmetrically in (23)), so a great number of policies achieve the first best. However, fiscal policy d_t helps monetary policy if there is a constraint (e.g. at the ZLB), so the possibility of future fiscal policy is a complement to the monetary policy (as it relieves the ZLB).\(^{44}\)

\(^{44}\)This “second instrument” could be very useful even in normal times, in a richer model with capital. Suppose that consumers get too optimistic about the future: the central bank should raise the interest rate. But then, that depresses investment. We do not get the first best any more, without a second instrument.
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 (without ZLB). I next 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 cost-push shock, i.e. a disturbance $\nu_t$ to the Phillips curve, which becomes:

$$\pi_t = \beta M^f E_t [\pi_{t+1}] + \kappa x_t + \nu_t$$  (36)

and the disturbance $\nu_t$ follows an AR(1), $\nu_t = \rho \nu_{t-1} + \epsilon_t'$. For instance, if firms’s optimal markup increases (perhaps because the elasticity of demand changes), they will want to increase prices and we obtain a positive $\nu_t$ (see Gali (2015, Section 5.2) for microfoundations).

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

4.3.1 Optimal commitment policy

The next proposition states the optimal policy with commitment. I normalize the initial (log) price level to be 0 ($p_{-1} = 0$).

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

$$\pi_t = -\frac{\vartheta}{\kappa} (x_t - M^f x_{t-1})$$  (37)

so that the (log) price level ($p_t = \sum_{\tau=0}^{t} \pi_{\tau}$) satisfies

$$p_t = -\frac{\vartheta}{\kappa} \left( x_t + (1 - M^f) \sum_{\tau=0}^{t-1} x_{\tau} \right)$$  (38)

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 ($p_t = \frac{\vartheta}{\kappa} x_t \to 0$ 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 lead the price level back to baseline.
Figure 3: This figure shows 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 optimum policy with behavioral firms. This illustrates Proposition 4.3.

“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 behavioral model. The recommendation of price level 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.\footnote{This is not a particularly intuitive fact, even in the rational model: technically, this is because the coefficient $\beta$ in the Phillips curve and the rate of time preference for policy are the same. That identity is broken in the behavior model. This is analogous to Slutsky symmetry in the rational model: there is no great intuition for its justification in rational model; this is in part because it fails with behavioral agents. Our intuitions are often (unwittingly) calibrated on our experience as living behavioral agents.} This lack of intuitive justification may be caused by that fact that it’s not robust to behavioral deviations, as Proposition 4.3 shows.

Likewise, “nominal GDP targeting” is optimal in the traditional model, but it is suboptimal with behavioral agents.
Figure 4: This figure shows optimal interest rate policy in response to a cost-push shock ($\nu_t$), when the central bank follows the optimal discretionary strategy. The behavior is very close, as the central bank does not rely on future commitments for its optimal policy. This illustrates Proposition 4.4.

**Other considerations**  Figure 3 gives some more intuition. Look at the behavioral 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.

The gains from commitment are lower, as agents don’t react much to the future.

### 4.3.2 Optimal discretionary policy

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

$$\pi_t = \frac{-\vartheta}{\kappa} x_t$$

and

$$i_t = \frac{\vartheta}{\kappa^2 + \vartheta(1 - \beta M^{f} \rho_\nu)} \nu_t$$  \hspace{1cm} (39)

which is the same expression as in the traditional model, up to the factor $M^{f}$. 

27
Let us first examine the comparative statics controlling for $\kappa$. For transitory shocks ($\rho_\nu = 0$), the optimal policy is independent of the firms’ bounded rationality. Future considerations don’t matter. However, for persistent shocks, the optimal policy is less aggressive ($\frac{dm}{d\nu}$ is lower) when firms are more behavioral. This is because, with more myopic firms, future cost push shocks do not affect firms’ pricing today much, hence the central bank needs to respond less to them.\footnote{Things are more complicated when we don’t control for $\kappa$. Plugging the endogenous values of $\kappa$ and $M^f$ (21-22) into (39), we see obtain:}

$$i_t = \frac{\vartheta}{(\bar{\kappa}m^f)^2 + \vartheta \left(1 - \beta \bar{m} \left[\theta + (1 - \theta) \frac{1 - \beta \bar{m}}{1 - \beta \bar{m} m^f} \right] \rho_\nu\right)^{\nu_t}}$$

Hence the optimal policy is again less aggressive when firms are more behavioral by decreasing $\bar{m}$, but the effect of $m^f$ is ambiguous.\footnote{Another take on this is in Section 11 of the Appendix.}

5 Enriching the Model with Long-Run Changes to Inflation

5.1 Enriched Model

So far all variables came back to a steady state value normalized to 0. This is sufficient for most of the analysis. Here, I extend the analysis to allow for the possibility that long run inflation might change.\footnote{Things are more complicated when we don’t control for $\kappa$. Plugging the endogenous values of $\kappa$ and $M^f$ (21-22) into (39), we see obtain:}

For good measure, I also extend the model to have backward looking terms, which has proven useful in empirical analyses, as this creates inertia in inflation (e.g. Galí and Gertler 1999). The interaction of backward looking terms for firms and permanent changes will prove fruitful.

To handle this case, it is good to have a slight generalization of the model. A microfoundation presented in the appendix (Section 9.2) leads to the following model.

**Proposition 5.1** (Behavioral New Keynesian model – three equation version) We obtain the following behavioral version of the New Keynesian model, for the behavior of output gap $x_t$ and inflation $\pi_t$:

$$x_t = M\mathbb{E}_t [x_{t+1}] + b_d d_t - \sigma (i_t - \mathbb{E}_t \pi_{t+1} - r^n) \tag{40}$$

$$\pi_t = \beta^f \mathbb{E}_t [\pi_{t+1}] + \alpha \pi^d_t + \kappa x_t \tag{41}$$

$$\pi^d_{t+1} = \pi^d_t + \gamma (\zeta \pi^CB_t + (1 - \zeta) \pi_t - \pi^d_t) \tag{42}$$

with $\alpha, \beta^f, \gamma, \zeta$ are all in $[0, 1]$ and $\alpha + \beta^f \leq 1$. The new term are $\pi^d_t$, “default inflation” coming from indexation, and $\pi^CB_t$, the “inflation guidance” by the central bank.
In the microfoundation, each firm has two ways of predicting future inflation: one is via “purely rational expectations”, with \( \pi_t \), another is via default inflation, \( \pi^d_t \). I view this default inflation as a simpler, more available source of signal about future inflation.

Leading old and new Keynesian models are embedded in the structure (40)-(42), as we shall see.

### 5.2 Long Run Behavior and Determinacy

Given our system (40)-(42), we ask two questions: does Fisher neutrality (or something close to it) hold? Is the economy stable? The analysis will reveal a connection between those properties.

We first make a few observations. Consider the long run value of inflation \( \pi_\infty \) and the nominal rate \( i_\infty \). Their link is as follows.

**Proposition 5.2** (Long run Fisher neutrality) If long run inflation is higher by \( d\pi_\infty \), then the long run nominal rate is higher by \( di_\infty \), where:

\[
\frac{di_\infty}{d\pi_\infty} = 1 - \frac{(1 - \alpha - \beta f)(1 - M)}{\kappa \sigma} \quad (43)
\]

**Proof.** Simply plug constant values of \( \pi_t = \pi^d_t = \pi^CB_t = \pi_\infty \), and \( x_t = x_\infty \) into (40)-(42). Then, we find \( i_\infty/\pi_\infty = 1 - \frac{(1-\alpha-\beta f)(1-M)}{\kappa \sigma} \). In addition, \( x_\infty = \frac{-i_\infty(1-\alpha-\beta f)\sigma}{\kappa \sigma(1-\alpha-\beta f)(1-M)} \). □

Next, I ask: is the equilibrium determinate? Is it stable?49

The next Proposition generalizes the earlier criterion (Proposition 3.1) to the case with backward looking terms.

**Proposition 5.3** (Equilibrium determinacy with behavioral agents – with backward looking terms)

The system is stable only if:

\[
\phi_\pi + \frac{(\alpha \zeta + 1 - \alpha - \beta f)(1 - M + \sigma \phi_x)}{\kappa \sigma} > 1. \quad (44)
\]

---

48Gali and Gertler (1999) present a model with partially backward looking firms: their model has \( \gamma = 1, \zeta = 0, M = 1 \). However, they have \( \zeta = 0 \), which prevents the stability analysis below, where \( \zeta > 0 \) is crucial.

49Technically, this is the following. Define \( z_t := (x_t, \pi_t, \pi^CB_t) \), taking \( \pi^CB_t \) as given, and write the system as \( E_t z_{t+1} = B z_t + a \pi^CB_t \), for a matrix \( B \). Given that \( \pi^d_t \) is a predetermined variable, we need \( B \) to have 2 eigenvalues with modulus greater than 1, and 1 with modulus less than 1. Also, in this draft I simply state the “central” necessary condition, that \( \det (I - B) > 0 \). The other, more minor, “Ruth-Hurwitz” conditions (see also Woodford 2003, pp.670-676) will be added in a subsequent iteration of this paper.
Now, which properties of real economies should a model reflect? First, in the long run, a steady state rise of in the nominal rate is associated with a rise in inflation: \( \frac{d\pi}{d\pi} > 0 \), something we might call “long run Fisher sign neutrality” (pure Fisher neutrality would be \( \frac{d\pi}{d\pi} = 1 \)). Most studies (e.g. Kandel, Ofer and Sarig 1996, Evans 1998) find \( \frac{d\pi}{d\pi} > 0 \) – though typically also they reject pure Fisher neutrality (\( \frac{d\pi}{d\pi} = 1 \)), and instead find \( \frac{d\pi}{d\pi} < 1 \), qualitatively as in Proposition 5.2.\(^{50}\) This means, given (43), that the data wants:

\[
\frac{(1 - \alpha - \beta f)(1 - M)}{\kappa \sigma} < 1.
\] (45)

Second, in the recent experience in Japan (since the late 1980s) and in Europe and US (since 2010), the interest rate has been stuck at the ZLB, but without strong vagaries of inflation or output. Hence, following Cochrane (2015), I hypothesize that another desirable empirical “target” for the model is that the economy is stable even if the monetary policy is stuck at the ZLB forever.\(^{51}\) In the model, that means that (use (44) in the case \( \phi_\pi = \phi_x = 0 \)):

\[
\frac{(\alpha \zeta + 1 - \alpha - \beta f)(1 - M)}{\kappa \sigma} > 1.
\] (46)

The next proposition records the tension between those two desirable properties, and a resolution.

**Proposition 5.4** (Long run links between inflation, nominal rates and stability) We have a positive long run link between inflation and nominal rates (45) and economic stability under passive monetary policy (46) if and only if \( \alpha \zeta \) is large enough and agents are boundedly rational (\( M < 1 \)), and prices are sticky enough (and as before, if prices are sticky enough). If \( \alpha \zeta = 0 \) (and “central bank guidance” has no impact) or \( M = 1 \), the two criteria cannot be simultaneously fulfilled.

This proposition means that the system is determinate if enough agents follow the central bank’s “inflation guidance”, \( \pi CB \) (i.e. if \( \alpha \zeta \) is large enough). Intuitively, then, agents are “anchored” enough and the system has fewer multiple equilibria. Bounded rationality makes people’s decision less responsive to the future (and in the old Keynesian model, to the past). As a result, it reduces the degree of complementarities, and we can more easily have only one equilibrium (this is quite a

\(^{50}\)The identification problems are very difficult, in part because we deal with fairly long run outcomes on which there are few observations.

\(^{51}\)This is a controversial issue, as other authors have argued that the instability of the 1970s in the US was due to the Taylor criterion being validated. Within the model, the 1970s can be interpreted as a moment where agents do not believe the central bank enough, i.e. \( \alpha \zeta \) is too low. This leads criterion (46) to be violated.
Leading old and new Keynesian models violate criterion (46), and allow for an unstable economy at the ZLB.

The Old Keynesian model of Taylor (1999) has: $\zeta = \beta f = 0, \alpha = 1$. As a result, criterion (46) is violated if monetary policy is passive (as it is at the ZLB). However, criterion (46) tells us how to get stability in an Old Keynesian model: have $\zeta > 0$. If the economy is at the ZLB, we avoid the deflationary spiral because of bounded rationality. We need $M < 1$, and also $\alpha \zeta > 0$, i.e. current inflation is not very responsive to its past and future values.

The traditional New Keynesian model has $M = 1$, so there is no stability (criterion (46) is violated). With $M < 1$, we can get stability. But to get “Fisher sign neutrality”, $\pi_-^\infty / i_-^\infty > 0$, we need $\alpha \zeta > 0$.

Hence, I conclude that the enrichment of this model is useful for both Old Keynesian and New Keynesian models.

Speculating somewhat more, this usefulness of “inflation guidance” may explain why central bankers these days do 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, which in turn can destabilize the economy.

### 5.3 Impulse Responses

A permanent shock to inflation. To study this system, I assume a permanent rise of 1% in the nominal rate, with $\pi_t^{CB} = 1\%$. Figure 5 shows the result. On impact, there is a recession: output and inflation are below trend. However, over time the default rate increases: as the central bank gives “guidance” $\pi_t^{CB} > 0$, inflation expectations are raised. In the long run, for this calibration, we obtain Fisher neutrality.

This effect is very hard to obtain in a conventional New Keynesian model. Cochrane (2015) documents this, and explores many variants: they all give that a rise in the interest rate creates a rise in inflation (though Cochrane needs to select one particular equilibrium, as the traditional model generates a continuum of bounded equilibria). However, here the bounded rationality of

---

52It also replaces $i_t - E_t \pi_{t+1}$ by $i_t - \pi_t$, and there is a $x_{t-1}$ term in the IS equation, but that is a fairly immaterial difference.

53The Taylor model does feature a deflationary spiral, because it has $\zeta = 0$.

54Parameters are: $M = 0.5$, $\beta f = 0.3$, $\alpha = 0.7$, $\sigma = 0.2$, $\kappa = 1.25$, $\gamma = 0.15$, $\zeta = 0.9$. They are provisional, and have not been particularly optimized.
agents overturns this result, with just one bounded equilibrium.

**A temporary shock to the interest rate.** I now study a temporary shock to the interest rate, \( i_t = i_0 e^{-\phi t} \) for \( t \geq 0 \). As the long run is not modified, I assume an inflation guidance of 0, \( \pi^{CB}_t = 0 \).

Figure 6 shows the result. On impact, inflation and output fall, and then mean-revert. The behavior is very close to what happens without the backward looking term, i.e. setting \( \alpha = 0 \).

For most purposes, I recommend the basic model of Proposition 2.6. However, when the long run changes, the extension proposed in this section is useful. Substantively, it yields the insight that the economy is stable if agents are boundedly rational (\( M < 1 \)) and they follow enough the “inflation guidance” by the government. Also, for empirical purposes, the extra backward term is helpful (Gali and Gerler 1999).

**6 Conclusion**

This paper gives a simple way to think about the impact of bounded rationality on monetary and fiscal policy.

It is grounded on fairly general microfoundations that model a behavioral agent in basic microe-
Figure 6: 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 percents.

Furthermore, we have seen that the model has good empirical support for its main non-standard elements. For instance, when Galí and Gertler (1999) estimate a Phillips curve, they estimate a coefficient on inflation of $\beta M^f \approx 0.75$ at the annual frequency, which leads to an attention parameter of $M^f \approx 0.8$ (at the annual frequency). In the IS curve, the literature on the forward guidance puzzle, using a mix of market data and thought experiments, gives good evidence that we need $M < 1$, a main contention of the model. Finally, the notion that a higher interest rate lowers inflation in the short run (Keynesian effect), then raises it in the long run (classical Fisherian effect) is generally well accepted, using again a mix of historical episodes and empirical evidence. This is generated by the model, and is hard to generate by other models (Cochrane 2015).

In conclusion, we have a model with quite systematic microfoundations and empirical support for its non-standard features, that is also simple to use.

This paper leads to a large number of natural questions.

I have studied only the most basic model. Doing a similar exploration for its richer variants would be very interesting, as relevant both empirically and interesting conceptually: e.g. with capital accumulation, a more frictional labor market, agents that are heterogeneous in wealth or
rationality. The tractable approach laid out in this paper makes the exploration of those questions quite accessible.

The present work has shown how important behavioral forces can be. More empirical work can assess them, e.g. using individual level dynamics for consumers (equation (7)), for firms (equation (19)), of the whole equilibrium economy (Proposition 2.6). 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).

In this model, agents do not have the “right” model of the world. This suggests empirical work trying to measure people’s subjective 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 understanding of individual dynamics (Greenwood and Shleifer (2014)) rather than not on their whole causal model.\(^{55}\) The tight parameterization in the present work allows 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 behavioral model.\(^{56}\)

\(^{55}\)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?”.

\(^{56}\)E.g. one could ask “suppose the central bank lowers interest rates by 1% [or the government gives $1000 to all agents)] for 1 period in 4 quarters, what will happen to the rest of the economy? to your decisions”, plot the impulse response, vary the “4” parameter, and compare that to rational and behavioral models.
7 Appendix: Behavioral 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 in 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.

7.1 Main model

Basic setup. Utility is:

\[ \sum_{t=0}^{1} \beta^t u(C_t, N_t) \quad \text{with} \quad u(C, N) = \frac{C^{1-\gamma} - 1}{1 - \gamma} - \frac{N^{1+\phi}}{1 + \phi}. \]

The economy consists of a Dixit-Stiglitz continuum of firms. Firm \( i \) produces \( q_{it} = N_{it} \) with unit productivity (there is no capital), and sets a price \( P_{it} \). The final good is produced competitively in quantity \( q_t = \left( \int_{0}^{1} q_{it}^{\epsilon} \right)^{\frac{1}{\epsilon-1}} \), and so that its price is:

\[ P_t = \left( \int_{0}^{1} P_{it}^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}. \] (47)

A corrective wage subsidy \( \tau = \tau \) (financed by lump-sum transfers) ensures that there are no price distortions on average, so that the optimal price set by a firm (when it can reset its price) is \( P_{it} = w_t \), so that price equals to marginal cost.\(^{57}\)

Calling GDP \( Y_t \), the aggregate resource constraint is:

\[ \text{Resource constraint: } Y_t = C_t + G_t = N_t. \] (48)

The real wage is \( \omega_t \). Labor supply is frictionless, so the agent respects his first order condition:

\(^{57}\)This is well-known: \( P_{it} = (1 - \tau) \mu w_t = w_t \) with \( \mu = \frac{\epsilon}{\epsilon-1} \).
\[
\omega_t u_c + u_N = 0, \text{ i.e.}
\]

Labor supply: \( N_t^\phi = \omega_t C_t^{-\gamma}. \) \hfill (49)

**The economy at time 1.** Let us assume that the time-1 economy has flexible prices and no government consumption. Then, the real wage must equal productivity, \( \omega_t = Z_t = 1. \) The labor supply equation (49) and \( C_t = N_t \) give: \( N_t^\phi = N_t^{-\gamma}, \) so

\[
N_1 = C_1 = 1.
\]

**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 maximizes

\[
\max_{(C_t)} \sum_{t=0}^{1} \beta^t C_t^{1-\gamma} \text{ s.t. } \sum C_t R_t = y_0 + \frac{y_1}{R_0}.
\]

That gives

\[
C_0 = b \left( y_0 + \frac{y_1}{R} \right)
\]

\[
b := \frac{1}{1+\beta}
\]

with log utility. \(^{58}\) But in this section I just use \( \psi = 1. \) Here \( b \) is the marginal propensity to consume (given the labor supply). \(^{59}\)

Without taxes, we also have the (real) income: \( y_t = C_t. \) Hence,

\[
C_0 = b \left( C_0 + \frac{C_1}{R_0} \right)
\]

which yields the Euler equation \( \beta R_0 C_0^{-\gamma} = 1. \) I use the consumption function formulation (51) rather than this Euler equation. Indeed, the consumption function is the formulation that generalizes well to behavioral agents (Gabaix 2016).

**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 \( P^d_0 \) (if prices are sticky, then \( P^d_0 = P_{-1}, \) but we could have \( P^d_0 = P_{-1} e^{\pi_0^d}, \) where \( \pi_0^d \) is an “automatic” price increase pre-programmed at time 1, not reactive to time-0 economic conditions, e.g. as in Mankiw and Reis

\(^{58}\)In the general case, \( b := \frac{1}{1+\beta \psi R_0^{-\gamma}}, \) calling \( \psi = \frac{1}{\gamma} \) the intertemporal elasticity of substitution (IES).

\(^{59}\)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 \( C = N = 1. \)
Other firms optimize freely their price, hence optimally choose a price

$$P^*_0 = \frac{\omega_0}{Z_0} P_0$$

(52)

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 (47), the aggregate price level is:

$$P_0 = \left( \theta \left( P^d_0 \right)^{1-\varepsilon} + (1-\theta) \left( P^*_0 \right)^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}}$$

(53)

as a fraction $\theta$ of firms set the price $P^d_0$ 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 ((48)-(49) and (51)-(53)). 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: $C_0 = N_0 = 1$.

In the model with sticky prices ($\theta > 0$), there is a one-dimensional continuum of real equilibria. It is the central bank who chooses the real equilibrium, by selecting the real interest rate $R_0$.\(^{60}\) 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 is $y_1 = y^d_1 + \hat{y}_1$, he sees only $y^*_1 = y^d_1 + \bar{m}\hat{y}_1$ 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 default, $y^d_1 = C_1 = 1$.

But now suppose that (50) becomes:\(^{61}\)

$$C_0 = b \left( y_0 + \frac{y^d_1 + \bar{m}\hat{y}_1}{R_0} \right).$$

(54)

Suppose that the government consumes $G_0$ at 0, nothing at time 1, and makes a transfer $T_t$ to the agents at times $t = 0, 1$. Call $d_0 = G_0 + T_0$ the deficit at time 0. The government must pay its

---

\(^{60}\)The central bank chooses the nominal rate. Given equilibrium inflation, that allows it to choose the real rate (when there are pricing frictions).

\(^{61}\)Formally in terms of behavioral dynamic programming as in Gabaix (2016), this comes from the consumers maximizing:

$$\mathbf{(C_0, N_0) = \arg \max_{C_0, N_0} u(C_0, N_0) + V \left( y^d_1 + m\hat{y}_1 + R_0 (T_0 + \omega_0 N_0 - C_0) \right)}$$

where $V$ is the continuation value function. To make things very straightforward, consider that $N_1$ is fixed at 1. Then, $V(x) = u(x, 1)$.
debt at the end of time 1, which yields the fiscal balance equation:

\[ R_0 d_0 + T_1 = 0. \]  

(55)

The real income of a consumer at time 0 is

\[ y_0 = C_0 + G_0 + T_0 = C_0 + d_0. \]

Indeed, labor and profit income equal the sales of the firms, \( C_0 + G_0 \), plus the transfer from the government, \( T_0 \). Income at time 1 is \( y_1 = Y_1 + T_1 \): GDP, plus the transfer from the government.\(^\text{62}\) Hence, (54) gives:

\[ C_0 = b \left( C_0 + d_0 + \frac{Y_1 + \bar{m} T_1}{R_0} \right). \]

Using the fiscal balance equation (55) we have:

\[ C_0 = b \left( C_0 + (1 - \bar{m}) d_0 + \frac{Y_1}{R} \right) \]

and solving for \( C_0 \):

\[ C_0 = \frac{b}{1 - b} \left( (1 - \bar{m}) d_0 + \frac{Y_1}{R_0} \right). \]  

(56)

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 (56). However, take the case of behavioral consumers, \( \bar{m} \in [0, 1) \). Consider a transfer by the government \( T_0 \), with no government consumption, \( G_0 = 0 \). Equation (56) means that a positive transfer \( d_0 = T_0 \) stimulates activity. If the government gives him \( T_0 > 0 \) 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 stand on monetary policy to determine \( R_0 \). Here assume that the central bank does not change the interest rate \( R_0 \).\(^\text{63}\)

\(^{62}\)Still, in equilibrium \( C_1 = Y_1 \). If \( d_0 > 0 \), then the transfer \( T_1 \) is negative. Agents use the proceed of the time-0 government bonds to pay their taxes at time 1.

\(^{63}\)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).
Then, (56) implies that GDP \( Y_0 = C_0 + G_0 \) changes as:

\[
\frac{dY_0}{dT_0} = \frac{b}{1 - b} (1 - \bar{m}).
\]  

(57)

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.

**Proposition 7.1** Suppose that we have (partially) sticky prices, and the central bank keeps the real interest rate constant. Then, a lump-sum transfer \( T_0 \) from the government at time 0 creates an increase in GDP:

\[
\frac{dY_0}{dT_0} = \frac{b}{1 - b} (1 - \bar{m})
\]

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 \( \frac{dY_0}{dT_0} > 0 \) 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 \( R_0 \)). Equation (56) 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 (49), 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 Ricardian equivalence fails), the multiplier \( \frac{dY_0}{dG_0} \) is greater than 1. This is for the reason evoked in undergraduate textbooks: people feel richer, so 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.\(^{64}\)

\(^{64}\)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,
Again, this relies on monetary policy here 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 microfoundation 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. The 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 7.2.

**The Phillips curve.** Taking a log-linear approximation around $P_t = 1$, with $p_t = \ln P_t$, (53) becomes: $p_0 = \theta p_0^d + (1 - \theta) p_0^*$, i.e.

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

Recall that $P_0^d = P_{-1} e^\pi_0^d$, so inflation is

$$\pi_0 = p_0 - p_{-1} = (p_0 - p_0^d) + (p_0^d - p_{-1}) = \frac{1 - \theta}{\theta} (p_0^* - p_0) + \pi_0^d. \quad (58)$$

Via (52),

$$p_0^* - p_0 = \hat{\omega}_0 \quad (59)$$

where $\hat{\omega}_0 = \frac{\omega_0 - \omega_0^*}{\omega_0^*}$ is the percentage deviation of the real wage from the frictionless real wage, $\omega_0^*$. Because of the labor supply condition (49), and $C_0 = N_0$, we have $\omega_0 = C_0^{\phi + \gamma}$. Therefore, $\hat{\omega}_0 = (\phi + \gamma) \hat{C}_0$. Hence (59) becomes $p_0^* - p_0 = (\phi + \gamma) \hat{C}_0$, and (58) yields:

Phillips curve: $\pi_0 = \kappa \hat{C}_0 + \pi_0^d \quad (60)$

with $\kappa := \frac{1 - \theta}{\theta} (\phi + \gamma)$. Hence, we obtain the elementary “New Keynesian 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 $x_0 = (C_0 + G_0 - Y_0^d) / Y_0^d$ is the deviation of GDP which makes the analytics quite complicated when generalizing to a large number of periods. The methodology here generalizes well to static and dynamic contexts.
Proposition 7.2 (Two-period behavioral Keynesian model) In this 2-period model, we have for time-0 consumption and inflation between periods 0 and 1:

\[ \dot{x}_0 = \dot{G}_0 + b_d \dot{d}_0 - \sigma \dot{r}_0 \] (IS curve)
\[ \pi_0 = \kappa \dot{C}_0 + \pi_0^d \] (Phillips curve)

where \( \dot{G}_0 \) is government consumption, \( \dot{d}_0 \) the budget deficit, \( b_d = \frac{b}{1-b} (1 - \hat{m}) \) is the sensitivity to future deficits, \( b = \frac{1}{1+\beta} \) is the marginal propensity to consume (given labor income) and \( \dot{r}_0 = r_0 - E\pi_1 \) is the real interest rate between periods 0 and 1, and \( \sigma = \frac{1}{R} = \beta \) with log utility.

7.2 Complements to the 2-period Model

This section gives complements to the 2-period model.

Discounted Euler equation in the 2-period model Also, this consumer satisfies a discounted Euler equation. Rewrite (51) as

\[ C_0 = b \left( C_0 + \frac{C_1^d + m \hat{C}_1}{R_0^d + m_r \hat{R}_0} \right) \]

where \( C_0^d = b \left( C_0^d + \frac{C_1^d}{R_0^d} \right) \). Then, we have

\[ \hat{C}_0 = b \hat{C}_0 + \frac{m \hat{C}_1 - m_r \hat{R}_0}{R_0^d} \]
i.e.

\[ \hat{C}_0 = \frac{b}{1-b} \hat{C}_1 \frac{1}{R_0^d} \left( m \hat{C}_1 - m_r \hat{R}_0 \right) \].

In the rational model, we have \( C_0 = \frac{b}{1-b} \hat{C}_1 \) and \( C_0 = C_1 = 1 \). Hence, \( \frac{b}{1-b} \hat{C}_0 = 1 \). We obtain:

\[ \hat{C}_0 = m \hat{C}_1 - m_r \hat{R}_0. \]

\[ \text{If the agent perceived only part of the change in the real rates, replacing } R_0 \text{ by } (1 - m_r) R_0^d + m_r R_0 \text{ in (56), then the expression in (61) would be the same, replacing } \sigma = \frac{1}{R} \text{ by } \sigma = \frac{m_r}{R}. \]
This is a “discounted Euler equation” (with discount factor $m$), i.e. instead of the rational Euler equation, $\dot{C}_0 = \mathbb{E}\left[\dot{C}_1\right] - \dot{R}_0$. The same factor $m$ gives power to fiscal policy, and yields a discounted Euler equation.

**Derivation of (54).** Call $k_1$ the wealth at the beginning of period 1 (before receiving labor income and profit), and $T_1$ the transfer received from the government, and $I_1$ the profit income from the oligopolistic firms (so that $\omega_1 N_1 + I_1 = C_1$ when aggregating). The rational value function at time 1 is:

$$V^r(k_1, T_1) = \max_{c_1, N_1} u(c_1, N_1) \text{ s.t. } c_1 \leq \omega_1 N_1 + I_1 + k_1 + T_1.$$ 

The decision at time 0 is

$$\text{smax}_{c_0, N_0, \bar{m}} u(c_0, N_0) + \beta V^r(R_0 (\omega_0 N_0 + I_0 + T_0), \bar{m} T_1)$$

where $\bar{m}$ is optimized upon in the sparse max. Taking here provisionally the $\bar{m}$ as given, then the decision is simply:

$$\max_{c_0, N_0} u(c_0, N_0) + \beta V^r(R_0 (\omega_0 N_0 + I_0 + T_0 - c_0), \bar{m} T_1).$$

The first order conditions are:

$$u_{c_0} = \beta R_0 V_{k_1},$$

$$u_{N_0} = -\omega_0 R_0 V_{k_1},$$

so that the intra-period labor supply condition $\omega_0 u_{c_0} + u_{N_0} = 0$ holds. Given that $V_{k_1} = u_{c_1}$, we obtain

$$u_{c_0}(c_0, N_0) = \beta R_0 u_{c_1}(c_0, N_0).$$

Now, we have $V^r_{k_1} = u'(c_1) = u'(k_1 + y_1)$ with $y_1 = \omega_1 N_1 + I_1 + m T_1$, so

$$\frac{1}{c_0} = \frac{\beta R_0}{c_1}$$

with $c_1 = y_1 + R (y_0 - c_0)$ i.e. $c_0 + \frac{y_1}{R} = y_0 + \frac{y_1}{R}$, and with the Euler equation $c_1 = \beta R_0 c_0$:

$$C_0 = \frac{1}{1 + \beta} \left( y_0 + \frac{y_1}{R} \right) = b \left( y_0 + \frac{y_1 + m y_1}{R} \right).$$

42
Appendix: Complements and Precisions

8.1 Details on the behavioral consumer

Here I state the assumptions regarding the behavioral consumer. I use the notions laid out in Gabaix (2016). The agent’s action is $a_t = (c_t, N_t)$, consumption and labor supply. The macroeconomy (excluding the agent’s personal wealth) is parametrized by a state vector $X_t$ that contains productivity shocks, announcements etc. For instance (linearizing), $( \hat{r}_t, \tilde{\omega}_t) = b X_t$ for some equilibrium vector $b$. The state vector is $z_t = (k_t, X_t)$, where $k_t$ is the agent’s wealth. The perceived laws of motions are:

\[
\begin{align*}
k_{t+1} &= F^k(a_t, z_t, m) = (1 + r + m_r \hat{r}_t)(k_t + \tilde{\omega} + m_y \tilde{\omega}_t - c_t) \\
X_{t+1} &= F^X(a_t, z_t, m) = \bar{m}AX_t
\end{align*}
\]

where $y_t = \omega_t N_t + \Pi_t + T_t$. Here $m = (m_r, m_y, \bar{m})$ is the vector of attention parameter. When all the components are 1 (so $m = \iota := (1, ..., 1)$), the agent is rational.

This agent pays full attention to his wealth, which ensures his dynamic budget constraint. He also has a rational understanding and rational reaction to the steady state: he understand the state value $\bar{r}$ of the interest rate, for instance.

But he’s a bit myopic to the deviation from it, e.g. he sees only a part $m_r \hat{r}_t$ of the deviation of the interest rate from trend.

The true law of motion of the macro state vector is $X_{t+1} = F^X(X_t, \iota) = AX_t$ (linearizing), but the agent has a subjective perception of it that it different: it is $X_{t+1} = F^X(\bar{m}X_t, \iota) = \bar{m}AX_t$. This captures that the future is hard to forecast so at each round in the future, the agent “cognitively discounts” it – sees those deviations from the steady state less and less as they are more remove.

The agent chooses his action, at each period, according to:

\[
a_t = \arg \max_{a,m} v(a, z_t, m)
\]

where the value function maximized is:

\[
v(a, z_t, m) := u(a) + \beta V^r(F^z(a, z_t, m), m)
\]

and $V^r$ is the rational value function in the economy parametrized by vector $m$, i.e. the value

\footnote{In his consumption function (Proposition 2.2, the term $c^d_t = \frac{\bar{r}k_t}{R} + \bar{y}$ reflects that attention to his wealth.}
function that the agent would attain if the world was indeed the world he perceives. As discussed in Gabaix (2016), other proxy value functions would only affect second order terms (in $O(\|X\|^2)$) in the agent’s decision.

In this draft of the paper I take the attention policy $m$ as fixed – a later draft will endogenize it. So basically the agent maximizes at each round under a slightly imperfect model of the world.

### 8.2 The “natural interest rate” in a behavioral economy

The natural interest rate is defined here as the interest rate that would prevail “if pricing frictions were removed”, but keeping cognitive frictions (and before any deficits). Let us examine this in detail. Take the IS curve (23), coming back to the more basic notion of $\hat{c}_t := \ln c_t - \ln \bar{c}$:

$$\hat{c}_t = M\mathbb{E}_t [\hat{c}_{t+1}] + b_d d_t - \sigma (r_t - \bar{r}) \quad (\text{IS curve})$$

where $r_t = i_t - \mathbb{E}_t \pi_{t+1}$ is the real rate.

Consider also the case with productivity shocks, so that $C_t = e^{\zeta_t} N_t$, so that the optimum frictionless consumption (see the derivation of (69)) is

$$\hat{c}_t^0 = \frac{1 + \phi}{\gamma + \phi} \hat{c}_t$$

So, if we removed all pricing frictions, we’d have $\hat{c}_t = \hat{c}_t^0$, and if we were in an environment with no deficit, we’d have:

$$\hat{c}_t^0 = M\mathbb{E}_t [\hat{c}_{t+1}^0] - \sigma (r_t^0 - \bar{r}),$$

which gives us the value of the natural rate, $r_t^n = \bar{r} + \frac{M\mathbb{E}_t [\hat{c}_{t+1}^n] - \hat{c}_t^0}{\sigma}$. Given we define the output gap as $x_t := \hat{c}_t - \hat{c}_t^n$, we have:

$$x_t = M\mathbb{E}_t [x_{t+1}] + b_d d_t - \sigma (r_t - r_t^n)$$

which is the formulation in the paper.

Note that we could have defined the “natural” rate as the rate that would prevail in an economy without pricing frictions, and given the actual deficits, i.e. defined it as the solution $\tilde{r}_t^n$ of:

$$\hat{c}_t^n = M\mathbb{E}_t [\hat{c}_{t+1}^n] + b_d d_t - \sigma (\tilde{r}_t^n - \bar{r})$$
i.e. \( r^n_t = r + \frac{b_d d_t + M \eta [\hat{c}^n_{t+1} - \hat{c}^n_t]}{\sigma} \). And then the IS curve would become:

\[
x_t = M \mathbb{E}_t [x_{t+1}] - \sigma (r_t - r^n_t)
\]

This would be mathematically equivalent, but the language would become more complicated. Then a policy change (via deficits) would change the natural rates. For instance, a temporary rise of the deficit would decrease the natural rate (as it makes people want to spend more). With that definition, the natural rate is not very “natural”.

9 Appendix: Additional Proofs and Some Closed Forms

9.1 Additional Proofs

Proof of Proposition 2.1 For consumption, this is simply a re-expression of Lemma 4.2 in Gabaix (2016), adapting the notations and the timing.

Labor supply is \( g' (N_t) = \omega_t u' (c_t) \), i.e. \( N_t^\phi = \omega_t c_t - \gamma_t \). Taking logs and deviations from the constant values, \( N^\phi_t = \hat{\omega}_t - \gamma^\phi_t \).

\[
\]

Proof of Proposition 2.4 Derivation of (16). This comes naturally for the general formalism. Call \( z_s = (B_s, d_s, d_{s+1}, d_{s+2}, ...) \) the state vector (more properly, the part of it that concerns deficits). Under the rational model, \( z_{s+1} = Hz_s \) for a matrix \( H \): \((Hz) (1) = z (1) + Rz (2) \) and \((Hz) (i) = z (i + 1) \) for \( i > 0 \), where \( z (i) \) is the \( i \)-th component of vector \( z \). Under the cognitive discounting model simulated by the agent at time \( t \), set \( z_t^d = (B_t, 0, 0, ...) \) and the subjective model \( z_s = z_t^d + \frac{m^s}{R} (z_s - z_t^d) \). This captures that the agent “sees” the debt \( B_t \), but more dimly the deficits \( d_t \). We also have

\[
T_s = -\frac{r}{R} B_s + d_s = e^T z_s \text{ with } e^T := \left( -\frac{r}{R}, 1, 0, 0, ... \right).
\]

So,

\[
\mathbb{E}_t^{BR} [T_s] = \mathbb{E}_t^{BR} [e^T \cdot z_s] = e^T \cdot \mathbb{E}_t^{BR} [z_s] = e^T \cdot (z_t^d + (\bar{m} A)^{s-t} (z_s - z_t^d))
\]

\[
= e^T \cdot (z_t^d + \bar{m}^{s-t} \mathbb{E}_t^{rat} [z_s - z_t^d]) = -\frac{r}{R} B_t + \bar{m}^{s-t} \left( T_s + \frac{r}{R} B_t \right).
\]

\[
= -\frac{r}{R} B_t + \bar{m}^{s-t} \left( d_s - r \sum_{u=t}^{s-1} d_u \right).
\]
Derivation of (17). We have:

\[ x_t = \frac{r}{R} k_t + \mathbb{E}^{BR}_t \left[ \sum_{s \geq t} \frac{1}{R^{s-t}} b_y (x_s + \tau_s) \right] \]

\[ = \frac{r}{R} B_t + b_y \sum_{s \geq t} \mathbb{E}^{BR} \left[ \tau_s \right] \frac{1}{R^{s-t}} + \mathbb{E}^{BR}_t \left[ \sum_{s \geq t} \bar{m}_{s-t} m_y b_y x_s \right] \]

\[ = \frac{r}{R} B_t + \frac{r}{R} \sum_{s \geq t} \left( m_y m_{s-t} (d_t - r \sum_{\tau=t}^{s-1} d_\tau) \right) \frac{1}{R^{s-t}} + \mathbb{E}^{BR}_t \left[ \sum_{s \geq t} \bar{m}_{s-t} m_y b_y x_s \right] \]

\[ = \mathbb{E}^{BR}_t \left[ \sum_{s \geq t} \bar{m}_{s-t} m_y b_y \left( x_s + d_s - r \sum_{\tau=t}^{s-1} d_\tau \right) \right]. \]

We see that the impact of \( B_t \) cancels out, a form of Ricardian equivalence. Old debt \( B_t \) does not make the agent feel richer. But a new deficit today \( (d_t) \) does.

This implies:

\[ x_t = m_y \frac{r}{R} (x_t + A d_t) + \bar{m} \frac{R}{R} x_{t+1} \]

with

\[ A = 1 - r \sum_{s \geq t+1} \bar{m}_{s-t} R^{s-t} = 1 - r \frac{\bar{m}}{R - \bar{m}} = 1 - \frac{r \bar{m}}{R - \bar{m}} = \frac{R (1 - \bar{m})}{R - \bar{m}}. \]

So, rearranging as in the derivation leading up to Proposition 2.3,

\[ x_t = \frac{1}{R - m_y r} (A m_y d_t + \bar{m} x_{t+1}) = b_d d_t + \frac{\bar{m}}{R - m_y r} x_{t+1} \]

with \( b_d = \frac{r m_y}{R - m_y r} R (1 - \bar{m}) / R - \bar{m} \). \( \square \)

**Proof of Proposition 2.5** The proof follows the steps and notations of Galí (2015, Chapter 3). I simplify matters by assuming constant return to scale \( (\alpha = 0 \text{ in Galí’s notations}) \). So, the marginal cost at \( t + k \) is simply \( \psi_{t+k} \), not \( \psi_{t+k|k} \).

**Notations.** When referring to equation 10 of Chapter 3 in Galí (2015), I write “equation (G10)” and do the same for (G11) and other equations. Lower-case letters denote logs. I replace the coefficient of relative risk aversion \( (\sigma \text{ in his notations}) \) by \( \gamma \) (as in \( u'(C) = C^{-\gamma} \)).

The law of motion is correctly perceived, but in the profit function, firms see only a fraction \( m^f \) of the (variable component) of the markup: they replace the real markup \( \psi_{t+k} - p_t \) by \( m^f (\psi_{t+k} - p_t) + \left( 1 - m^f \right) (\psi_{t+k}^d - p_t^d) \).

Firms can reset their price with probability \( 1 - \theta \). Mimicking Galí’s calculations, with behavioral
firms, firm \( i \)'s optimal price \( p^*_i \) satisfies:

\[
p^*_i - p_t = (1 - \beta \theta) \sum_{k \geq 0} (\beta \theta)^k E_t^{BR} [\psi_{t+k} - p_t].
\]

Given our assumption that firms underperceive the departure of the markup from the baseline:

\[
p^*_i - p_t = (1 - \beta \theta) \sum_{k \geq 0} (\beta \theta)^k \bar{m}_f E_t [\psi_{t+k} - p_t].
\]

Equation (G15) still holds, with \( \mu_t := p_t - \psi_t \), and becomes simply \( \psi_{t+k} = p_{t+k} - \mu_{t+k} \), so

\[
p^*_i - p_t = (1 - \beta \theta) \sum_{k \geq 0} (\beta \bar{m})^k m_f E_t [p_{t+k} - \mu_{t+k} - p_t]
\]

i.e., using \( \pi_t = (1 - \theta) (p^*_i - p_{t-1}) = \frac{1-\theta}{\theta} (p^*_i - p_t) \),

\[
\frac{\theta}{1-\theta} \pi_t = p^*_i - p_t = (1 - \beta \theta) \sum_{k \geq 0} (\beta \bar{m})^k m_f E_t [p_{t+k} - p_t - \mu_{t+k}] \tag{64}
\]

which is a close cousin of the equation right before (G16).

I now define: \( \lambda := \frac{1}{\delta} (1 - \theta) (1 - \beta \theta) m_f \), so that

\[
\frac{1}{\lambda} \pi_t = \frac{\theta}{(1 - \theta) (1 - \beta \theta) m_f} \pi_t = E_t \sum_{k \geq 0} (\beta \bar{m})^k (p_{t+k} - p_t) - E_t \sum_{k \geq 0} (\beta \bar{m})^k \mu_{t+k}. \tag{65}
\]

I use the forward operator \( F \) (\( Fy_t := y_{t+1} \)), which allows me to evaluate infinite sums compactly, as in (with \( \delta = \beta \bar{m} \)):

\[
B := \sum_{k=0}^{\infty} \delta F^k \mu_{t+k} = \sum_{k=0}^{\infty} \delta F^k \mu_t = \left( \sum_{k=0}^{\infty} \delta F^k \right) \mu_t = (1 - \delta F)^{-1} \mu_t. \tag{66}
\]

I also drop the expectation operator to simplify the notation. Hence (65) becomes:

\[
\frac{1}{\lambda} \pi_t = A - (1 - \beta \bar{m} F)^{-1} \mu_t
\]
where

\[ A := \sum_{i \geq 0} (\beta \bar{m})^i (p_{t+i} - p_t) = \sum_{i \geq 0} (\beta \bar{m})^i (\pi_{t+i} + \ldots + \pi_{t+1}) = \sum_{k \geq 1} \left[ \sum_{i \geq k} (\beta \bar{m})^i \right] \]

\[ = \sum_{k \geq 1} \pi_{t+k} \frac{(\beta \bar{m})^k}{1 - \beta \bar{m}} = (1 - \beta \bar{m}F)^{-1} \frac{\beta \bar{m}F \pi_t}{1 - \beta \bar{m}} \]

where I again used (66). Hence,

\[ \frac{1}{\lambda} \pi_t = A - (1 - \beta \bar{m}F)^{-1} \mu_t = (1 - \beta \bar{m}F)^{-1} \left( \frac{\beta \bar{m}F}{1 - \beta \bar{m}} \pi_t - \mu_t \right). \]

Multiplying both sides by \( \lambda (1 - \beta \bar{m}F) \) gives, defining \( C := \frac{\lambda}{1 - \beta \bar{m}} \):

\[ (1 - \beta \bar{m}F) \pi_t = C \beta \bar{m}F \pi_t - \lambda \mu_t \]

i.e. \( \pi_t = (1 + C) \theta \beta \bar{m}F \pi_t - \lambda \mu_t \), and reintroducing the expectation operator:

\[ \pi_t = \beta \bar{m} \theta (1 + C) E_t [\pi_{t+1}] - \lambda \mu_t. \]  \hfill (67)

Thus we obtain the key equation (which is a behavioral version of (G17)):

\[ \pi_t = \beta M^f E_t [\pi_{t+1}] - \lambda \mu_t \]  \hfill (68)

with

\[ M^f := \bar{m} \theta (1 + C) = \bar{m} \left[ \theta + (1 - \theta) \frac{1 - \beta \theta}{1 - \beta \bar{m}} m^f \right]. \]

We finally need to express the desired markup \( \mu_t \) as a function of primitives. Recall that \( \zeta_t \) is log productivity. The labor supply is still (49), \( N_t^\phi = \omega_t C_t^{-\gamma} \), and as the resource constraint is \( C_t = e^{\zeta_t} N_t, \omega_t = e^{-\phi \zeta_t} C_t^{(\gamma+\phi)} \). The real marginal cost is then \( \Psi_t/P_t = \frac{\omega_t}{C_t} = e^{-(1+\phi)\zeta_t} C_t^{(\gamma+\phi)} \). Then, recall the definition \( \mu_t = p_t - \psi_t \), we obtain

\[ \mu_t = (1 + \phi) \zeta_t - (\gamma + \phi) c_t \]

Next, if the pricing frictions disappeared, the markup would be 0 (recall that the government has
a subsidy yie, i.e. consumption would be at $c^n_t$ s.t.

$$0 = (1 + \phi) \zeta_t - (\gamma + \phi) c^n_t$$

which gives the efficient level of consumption:

$$c^n_t = \frac{1 + \phi}{\gamma + \phi} \zeta_t \quad \text{(69)}$$

So, the output gap is $x_t := c_t - c^n_t$ satisfies:

$$\mu_t = - (\gamma + \phi) x_t$$

Plugging this in (68), we obtain the behavioral version of (G22):

$$\pi_t = \beta M^F E_t [\pi_{t+1}] + \kappa x_t$$

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

$$\kappa = \bar{\kappa} m^f$$

$$\bar{\kappa} = \left( \frac{1}{\theta} - 1 \right) (1 - \beta \theta) (\gamma + \phi) \quad \text{(71)}$$

□

**Proof of Proposition 3.2** We have: $\dot{x}_t = \xi x_t - \sigma (\underline{r} + \pi_t)$. To solve for the system, note:

$$\ddot{x}_t = \xi \dot{x}_t - \sigma \ddot{x}_t = \xi \dot{x}_t - \sigma (\rho \pi_t - \kappa x_t) = \xi \dot{x}_t + \sigma \kappa x_t - \rho \sigma \pi_t$$

$$= \xi \dot{x}_t + \sigma \kappa x_t + \rho (\dot{x}_t - \xi x_t + \sigma \underline{r}) = (\rho + \xi) \dot{x}_t + (\sigma \kappa - \rho \xi) x_t + \rho \sigma \underline{r}$$

so that:

$$\ddot{x}_t - (\rho + \xi) \dot{x}_t + (\rho \xi - \sigma \kappa) x_t = \rho \sigma \underline{r} \quad \text{(72)}$$

and the boundary conditions are: $x_T = \pi_T = 0$, hence (taking the left derivative):

$$x_T = 0, \dot{x}_T = -\sigma \underline{r} \quad \text{(73)}$$
To analyze (72), we look for solutions of the type \( x_t = e^{\lambda t} \). Call \( \lambda \leq \lambda' \) the two roots of:

\[
\lambda^2 - (\rho + \xi) \lambda + \rho \xi - \sigma \kappa = 0. \tag{74}
\]

Then, with \( D = \frac{\rho \sigma r}{\rho \xi - \sigma \kappa} \) the solution is:

\[
x_t = D + \frac{(D \lambda - \sigma x) e^{\lambda (t-T)} - (D \lambda' - \sigma x) e^{\lambda' (t-T)}}{\lambda' - \lambda}. \tag{75}
\]

In the traditional case, \( \xi = 0 \), so that \( \lambda < 0 < \lambda' \). As \( D > 0 \), this implies that, as \( t \to -\infty \), \( x_t \to -\infty \). We obtain an unboundedly large recession. This is the logic that Werning (2012) analyzes.

However, take the case where cognitive myopia is strong enough, \( \xi > \frac{\sigma \kappa}{\rho} \). Then, both roots of (74) are positive. Hence, we have a bounded recession. Indeed, as \( D < 0 \) in that case, \( x_t \) is increasing in \( t \). \( \square \)

**Proof of Lemma 4.1** The proof mimics the ones in Woodford (2003) and Gali (2015). We have

\[
W = -\frac{1}{2} u_c E_0 \sum_{t=0}^{\infty} \beta^t \left[ (\gamma + \phi) x_t^2 + \epsilon \text{var}_i (p_t (i)) \right]
\]

where \( \text{var}_i (p_t (i)) \) is the dispersion of prices at time \( t \). As in Woodford (2003, Chapt. 6),

\[
\sum_{t=0}^{\infty} \beta^t \text{var}_i (p_t (i)) = \frac{\theta}{(1-\theta)(1-\beta \theta)} \sum_{t=0}^{\infty} \beta^t \pi_t^2 + \frac{1}{1-\beta \theta} v_{-1}
\]

\[
= \frac{\gamma + \phi}{\kappa} \sum_{t=0}^{\infty} \beta^t \pi_t^2 + \frac{1}{1-\beta \theta} v_{-1}
\]

using (71), and calling \( v_{-1} := \text{var}_i (p_{-1} (i)) \).

Hence,

\[
W = -\frac{1}{2} u_c E_0 \sum_{t=0}^{\infty} \beta^t \left[ (\gamma + \phi) x_t^2 + \frac{\gamma + \phi}{\kappa} \pi_t^2 \right] - \frac{1}{2} u_c c\epsilon \frac{1}{1-\beta \theta} v_{-1}
\]

\[
= -\frac{1}{2} u_c c (\gamma + \phi) E_0 \sum_{t=0}^{\infty} \beta^t \left( \frac{\pi_t^2}{\kappa} + \frac{\kappa}{\epsilon} x_t^2 \right) + W_-
\]

\[
= -\frac{1}{2} K E_0 \sum_{t=0}^{\infty} \beta^t \left[ \pi_t^2 + \varphi x_t^2 \right] + W_-
\]

50
with

\[
\begin{align*}
K & := u_c c (\gamma + \phi) \frac{\epsilon}{\kappa} = u_c c (\gamma + \phi) \frac{\epsilon}{\bar{m}_f} \\
\vartheta & := \frac{\bar{K}}{\epsilon} = \frac{\kappa}{m_f \epsilon} \\
W_- & := -\frac{1}{2} u_c c \epsilon \frac{1}{1 - \beta \theta} \text{var}_i (p_{-1} (i))
\end{align*}
\]

(76)

I used \( \kappa = \bar{m}_f \) from equation (70). Note that \( K \) and \( \vartheta \) are independent of behavioral factors, when expressed in terms of primitives including the components of \( \bar{K}, \epsilon \). However, when they’re expressed in terms of \( \kappa \), the behavioral term \( m_f \) intervenes.

**Proof of Proposition 4.3** The Lagrangian is

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

where \( \Xi_t \) are Lagrange multipliers. The first order conditions are: \( L_{x_t} = 0 \) and \( L_{\pi_t} = 0 \) which give respectively

\[
\begin{align*}
-\vartheta x_t + \kappa \Xi_t &= 0 \\
-\pi_t - \Xi_t + m_f \Xi_{t-1} &= 0
\end{align*}
\]

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 5.3** The state vector is \( z_t = (x_t, \pi_t, \pi_{t-1}) \). Write the system as \( \mathbb{E}_t z_{t+1} = B z_t + a \pi_t^{CB} \), for a matrix \( B \). To study stability, we dispense with the forcing term \( a \pi_t^{CB} \). We can write \( \mathbb{E}_t z_{t+1} = B z_t \), with

\[
B = \begin{pmatrix}
\frac{\kappa \sigma + \beta (1 + \sigma \phi)}{M \beta |} & \frac{\sigma (\beta \phi - 1)}{M \beta |} & \frac{\sigma \alpha}{M \beta |} \\
\frac{\beta}{\kappa} & \frac{1}{\beta \gamma} & \frac{\alpha}{\beta \gamma} \\
0 & \gamma (1 - \zeta) & 1 - \gamma
\end{pmatrix} .
\]

(77)

Consider also the characteristic polynomial of \( B \), \( \Phi (\Lambda) = \det (\Lambda I - B) \) (with \( I \) the identity matrix), which factorizes as \( \Phi (\Lambda) = \prod_{i=1}^{3} \Lambda_i \), where the \( \Lambda_i \)’s are the eigenvalues of \( B \).

When \( \alpha \neq 0 \), inflation \( \pi_t^d \) is a predetermined variable, not a jump variable. Hence, for determinacy, \( B \) needs to have 1 eigenvalue less than 1 in modulus (corresponding to the predetermined
variable \( \pi^d_t \), and 2 greater than 1 (corresponding to the free variables \( x_t, \pi_t \)). This implies that \( \Phi(1) > 0 \), which is equivalent to (44), given that direct calculation of \( \Phi(1) = \det(I - B) \) gives:

\[
\frac{\beta^f M}{\gamma \kappa \sigma} \Phi(1) = \phi_\pi - 1 + \frac{1}{\kappa \sigma} \left( \alpha \zeta + 1 - \alpha - \beta^f \right) (1 - M + \sigma \phi_x).
\]

Also, in this draft I simply state the “central” necessary condition, that \( \Phi(1) > 0 \). The other, more minor, “Ruth-Hurwitz” conditions (see also Woodford (2003, pp. 670-676)) will be added in a subsequent iteration of this paper. □

9.2 Microfoundations for the model with backward looking expectations

Here I provide microfoundations and derivations for the model in Section 5.

The consumer is as in the main model, so the IS equation (40) is as in the main model. The firms, however, have access to one more “signal” about future inflation, the default inflation \( \pi^d_t \). I assume that it follows (42), i.e. that it is a mix of past actual inflation and inflation guidance by the central bank. The key is to derive (41).

I will use the notation:

\[
\rho := \beta \theta \bar{m}
\]

(78)

I call \( F \) is the forward operator, \( Fy_t := y_{t+1}, \) which gives:

\[
\sum_{k \geq 0} \rho^k y_{t+k} = \sum_{k \geq 0} \rho^k y_{t+k} = \sum_{k \geq 0} \rho^k F^k y_t = (1 - \rho F)^{-1} y_t
\]

where \( F \) is the forward operator, \( Fy_t := y_{t+1}, \) and I used (66). I also drop the expectation operator to simplify the notation.

Let us start by calculating:

\[
A^0 := (1 - \beta \theta) \sum_{i \geq 1} (\beta \theta)^i (p_{t+i} - p_t) = (1 - \beta \theta) \sum_{i \geq 1} (\beta \theta)^i (\pi_{t+i} + \ldots + \pi_{t+1})
\]

\[
= (1 - \beta \theta) \sum_{k \geq 1} \pi_{t+k} \sum_{i \geq k} (\beta \theta)^i = \sum_{k \geq 1} \pi_{t+k} (\beta \theta)^k
\]

\[
= \sum_{k \geq 0} \pi_{t+k} (\beta \theta)^k 1_{k>0}
\]

---

\(^{67}\)Section 10.2 gives a continuous-time derivation, which is more compact and useful when studying variants of the model.
We call $\mu_t := p_t - \psi_t$ the desired markup. The firm $i$’s optimal price $p_t^*$ satisfies:

$$p_t^* - p_t = (1 - \beta \theta) E_t^{BR} \sum_{k \geq 0} (\beta \theta)^k (\psi_{t+k} - p_t)$$

$$= (1 - \beta \theta) E_t^{BR} \sum_{k \geq 0} (\beta \theta)^k (p_{t+k} - \mu_{t+k} - p_t)$$

$$= A^0 - (1 - \beta \theta) E_t^{BR} \sum_{k \geq 0} (\beta \theta)^k \mu_{t+k}$$

$$p_t^* - p_t = E_t^{BR} \sum_{k \geq 1} (\beta \theta)^k (\pi_{t+k+1} 1_{k > 0} - (1 - \beta \theta) \mu_{t+k})$$  \hspace{1cm} (79)

I assume\textsuperscript{68}

$$E_t^{BR}[\pi_{t+k}] = \bar{m}^k m_f E_t[\pi_{t+k}]$$  \hspace{1cm} (80)

$$\bar{\pi}_t := f \pi_t + (1 - f) \pi_{t-1}.$$  \hspace{1cm} (81)

and

$$E_t^{BR}[\mu_{t+k}] = \bar{m}^k m_f E_t[\mu_{t+k}]$$

Plugging this into (79) gives:

$$p_t^* - p_t = E_t \sum_{k \geq 0} m_f (\beta \theta \bar{m})^k \left( f \pi_{t+k+1} 1_{k > 0} + (1 - f) \pi_{t+k-1} 1_{k > 0} - (1 - \beta \theta) \mu_{t+k} \right)$$

$$= E_t \sum_{k \geq 0} m_f \rho^k \left( f \pi_{t+k+1} 1_{k > 0} + A_t \right)$$

$$= m_f E_t (1 - \rho F)^{-1} (f \rho F \pi_t + A_t)$$

where $\rho := \beta \theta \bar{m}$ and

$$A_t := (1 - f) \rho \pi_t^d - (1 - \beta \theta) \mu_t.$$  \hspace{1cm} (82)

Using

$$\pi_t = p_t - p_{t-1} = (1 - \theta) (p_t^* - p_{t-1})$$ as a fraction $1 - \theta$ of firms reset their price

$$= (1 - \theta) (p_t^* - p_t + p_t - p_{t-1}) = (1 - \theta) (p_t^* - p_t + \pi_t)$$

$$\pi_t = \frac{1 - \theta}{\theta} (p_t^* - p_t)$$

\textsuperscript{68}The assumption is slightly different than the one for Proposition 2.5, in part to probe the robustness of this mechanism.
so that with \( C := \frac{1-\theta}{\theta} \)

\[
\pi_t = C (p_t^* - p_t) \quad (83)
\]

\[
= C m_f (1 - \rho F)^{-1} (f \rho F \pi_t + A_t) \quad (84)
\]

Multiplying both sides by \((1 - \rho F)\) gives:

\[
(1 - \rho F) \pi_t = C m_f (f \rho F \pi_t + A_t)
\]

i.e.

\[
\pi_t = (1 + C m_f ) \rho F \pi_t + C m_f A_t
\]

Using (82) and reintroducing the expectation operator,

\[
\pi_t = (1 + C m_f ) \rho \mathbb{E}_t [\pi_{t+1}] + C m_f ((1 - f) \rho \pi_t^d - (1 - \beta \theta) \mu_t)
\]

\[
= \beta f \mathbb{E}_t [\pi_{t+1}] + \alpha \pi_t^d - \lambda m_f \mu_t \quad (85)
\]

with

\[
\beta^f = (1 + C m_f ) \rho = \left(1 + \frac{1-\theta}{\theta} f m_f \right) \beta \theta \bar{m} = \beta \bar{m} \left[ \theta + f m_f (1 - \theta) \right],
\]

\[
\beta^f = \beta M^f \quad (87)
\]

\[
M^f = \bar{m} \left( \theta + f m_f (1 - \theta) \right) \quad (88)
\]

\[
\alpha = \beta \bar{m} m_f (1 - f) (1 - \theta) \quad (89)
\]

Thus we obtain the key equation (which is a behavioral version of (G17)):

\[
\pi_t = \beta^f \mathbb{E}_t [\pi_{t+1}] + \alpha \pi_t^d - \lambda m_f \mu_t \quad (90)
\]

with \( \lambda = m_f C (1 - \beta \theta) \).

The rest of the proof is as in Galí. The labor supply is still (49), \( N_t^\phi = \omega_t C_t^{-\gamma} \), and as the resource constraint is \( C_t = N_t, \omega_t = C_t^{-(\gamma+\phi)} \), i.e. \( \dot{\omega}_t = - (\gamma + \phi) x_t \). This gives \( \mu_t = \dot{\omega}_t = - (\gamma + \phi) x_t \), and we obtain the behavioral version of (G22):

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

54
with $\kappa = m^f \lambda (\gamma + \phi)$, i.e.

$$\kappa = \bar{\kappa} m^f,$$

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

References


10 Further complements

10.1 Optimal policy in the enriched model with partially backward looking inflation

I now derive optimal policy in the enriched model of Section 5.1. As I consider shocks that do not affect the long run, I the case where the central bank chooses to keep stable long-run expectations, so $\pi_t^{CB} = 0$.

10.1.1 Optimal policy: Response to Changes in the Natural Interest Rate

The reasoning is exactly like for the baseline model. If the ZLB doesn’t bind, the first best ($\pi_t = x_t = 0$) is achieved by setting $i_t = r^n_t$. If the ZLB binds, the first best is achieved, by ensuring that $x_t = \pi_t = 0$ is a solution of (40)-(42), i.e.:

$$b_d d_t - \sigma (i_t - r^n_t) \quad 0$$

10.1.2 Optimal Policy with tradeoffs: Reaction to a Cost-Push Shock

The analytic solution is a bit messy, so here I just state the first order conditions, which allow for an easy numerical solution.

Optimal commitment policy

The Lagrangian is

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

$$+ \Psi_t \left( \pi_t^{d} + \gamma (\zeta \pi_t^{CB} + (1 - \zeta) \pi_t - \pi_t^{d}) - \pi_t^{d+1} \right) \right]$$

(91)

60
Figure 7: This figure shows 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 optimum policy with behavioral firms. This Figure is for the “enriched model” of Section 5.1.

where $\Xi_t$ and $\Psi_t$ are Lagrange multipliers associated with (41) and (42). The first order conditions are: $L_{x_t} = 0$, $L_{\pi_t} = 0$ and $L_{\pi_d} = 0$, which give respectively:

\[
-\vartheta x_t + \kappa \Xi_t = 0 \quad (92)
\]

\[
-\pi_t - \Xi_t + M^f \Xi_{t-1} + \gamma (1 - \zeta) \Psi_t = 0 \quad (93)
\]

\[
\alpha \Xi_t + (1 - \gamma) \Psi_t - \frac{1}{\beta} \Psi_{t-1} = 0 \quad (94)
\]

**Optimal discretionary policy** The reasoning is the same, so the optimality condition is $\pi_t = \frac{-\vartheta}{\kappa} x_t$.

The resulting policies are shown in Figures 7 and 8. Those are quite similar to the basic model, except that inflation is a more inertial (XX comment on magnitude).
Figure 8: This figure shows optimal interest rate policy in response to a cost-push shock ($\nu_t$), when the central bank follows the optimal discretionary strategy. The behavior is very close, as the central bank does not rely on future commitments for its optimal policy. This Figure is for the “enriched model” of Section 5.1.
10.2 Derivation of the Phillips curve (41) in continuous time

Here I show the derivation of the Phillips curve in continuous time. In exploring variants of the NK model, I found it much quicker to use this continuous-time derivation than the discrete time version (the 2-period model is also useful for basic conceptual issues).

If a firm can reset its price at time 0, it sets it to (using $\delta := \lambda + r$):

$$p^*_0 - p_0 = \mathbb{E}^{BR} \left[ \int_0^\infty \delta e^{-\delta t} (\mu x_t + p_t - p_0) \, dt \right] = \mathbb{E}^{BR} \left[ \int_0^\infty \delta e^{-\delta t} (\mu x_t + \int_0^t \pi_s \, ds) \, dt \right]$$

and using

$$\int_{t=0}^\infty \delta e^{-\delta t} \left( \int_{s=0}^t \pi_s \, ds \right) \, dt = \int_{s=0}^\infty \pi_s \, ds \left( \int_{t=s}^\infty \delta e^{-\delta t} \, dt \right) = \int_{s=0}^\infty e^{-\delta s} \pi_s \, ds$$

we have

$$p^*_0 - p_0 = \mathbb{E}^{BR} \left[ \int_0^\infty e^{-\delta t} (\delta \mu x_t + \pi_t) \, dt \right]. \quad (95)$$

We assume the following for the perceived inflation process:

$$\mathbb{E}^{BR}_0 [\pi_t] = e^{-\xi t} \mathbb{E}_0 [\bar{\pi}_t]$$

$$\bar{\pi}_s := f \pi_s + (1-f) \pi_s^d$$

so:

$$p^*_0 - p_0 = \mathbb{E} \left[ \int_0^\infty e^{-\delta (\delta + \xi) t} (\delta \mu x_t + \bar{\pi}_t) \, dt \right].$$

Now, we aggregate over all firms. Inflation at time 0 is $\pi_0 = \hat{p}_0 = \lambda (p^*_0 - p_0)$. Hence, we have:

$$\pi_0 = \lambda \mathbb{E} \left[ \int_0^\infty e^{-\delta (\delta + \xi) t} (\delta \mu x_t + \bar{\pi}_t) \, dt \right]. \quad (96)$$

To solve this, it is useful to use the differentiation operator, $D = \frac{d}{dt}$. With this notation, for a function $f$ (sufficiently regular), the Taylor expansion formula can be written as:

$$f(t + \tau) = \sum_{k=0}^\infty f^{(k)}(t) \frac{\tau^k}{k!} = \sum_{k=0}^\infty \left( D^k \frac{\tau^k}{k!} \right) f = e^{\tau D} f$$

i.e.

$$f(t + \tau) = e^{\tau D} f(t). \quad (97)$$
Hence, we have (formally at least):

\[ \int_0^\infty e^{-\rho \tau} f(\tau) \, d\tau = \int_0^\infty e^{-\rho \tau} e^{\tau D} f(0) \, d\tau = \frac{1}{\rho - D} f(0). \]  

(98)

Hence (96) becomes (dropping the expectations for ease of notation):

\[ \pi_t = \frac{\lambda}{\delta + \xi - D} \left( \delta \mu x_t + f \pi_t + (1 - f) \pi_t^d \right) \]  

(99)

and multiplying by \( \delta + \xi - D \),

\[ (\delta + \xi - D) \pi_t = \lambda \left( \delta \mu x_t + f \pi_t + (1 - f) \pi_t^d \right) \]

i.e.

\[ (\delta + \xi - \lambda f - D) \pi = \lambda \delta \left( 1 - f \right) \pi_t^d + \lambda \delta \mu x_t \]

(100)

This gives the continuous-time version of (41):

\[ (\delta + \xi - \lambda f) \pi = \frac{d}{dt} \pi_t + \lambda \delta \left( 1 - f \right) \pi_t^d + \lambda \delta \mu x_t. \]

10.3 Explicit Solutions for Impulse Responses

The correspondence between the discrete and continuous time version is that \( A = I - \ddot{A} \Delta t + o(\Delta t) \) and \( \Lambda_k = 1 - \lambda_k \Delta t + o(\Delta t) \) in the limit of small time intervals \( \Delta t \). I drew the graphs in continuous time.

**Continuous time.** I use the notations \( a_t = i_t - r_{nt} \) and \( b = (-\sigma, 0) \). In continuous time, the system (28) is

\[ 0 = \dot{z}_t - \ddot{A} z_t + b a_t \]

where \( \ddot{A}(m) = \begin{pmatrix} \xi & -\sigma \\ -\kappa & \rho \end{pmatrix} \), which has eigenvalues \( \lambda_1 \) and \( \lambda_2 \), solutions of \( \lambda^2 - (\xi + \rho) \lambda + \xi \rho - \kappa \sigma = 0 \). Using the stability criterion (32) they are both positive. Hence, the solution is

\[ z_t = \mathbb{E}_t \left[ \int_t^\infty e^{-\ddot{A}(s-t)} b a_s \, ds \right]. \]

(101)

To make the expression \( e^{-\ddot{A}(s-t)} \) a bit more explicit, we proceed as follows. We call \( w_k \) an
eigenvector corresponding to $\lambda_k$: $\bar{A}w_k = \lambda_k w_k$. We also find vectors $c_k$ so that $\sum_k w_k c_k' = I$, where $I$ is the identity matrix.

**Lemma 10.1** The solution can be written

$$z_t = \mathbb{E}_t \left[ \int_t^\infty \sum_{k=1}^2 e^{-\lambda_k(s-t)} v_k a_s ds \right]$$

where $v_k := w_k c_k' b$. Explicitly, $v_1 = \frac{\lambda_1 - \rho}{\lambda_1 - \lambda_2} \left( 1, \frac{\lambda_2 - \rho}{\sigma} \right)$ and symmetrically for $v_2$.

**Proof.** We diagonalize $\bar{A}$ and write $\bar{A} = V^{-1} D \lambda V$, where $D \lambda = \text{diag}(\lambda_k)$, and $w_k = V^{-1} e_k$, with $e_1 = (1, 0)$ and $e_2 = (0, 1)$. We also find vectors $c_k$ so that $\sum_k w_k c_k' = I$ (a “partition of unity”): explicitly, $c_k := V'e_k$. This gives

$$Ve^{-\bar{A} \tau} V^{-1} = \left( \text{diag}(e^{-\lambda_k \tau}) \right)_{k=1,2} = \sum_k e^{-\lambda_k \tau} e_k e_k'$$

Hence, we have

$$z_t = \mathbb{E}_t \left[ \int_t^\infty e^{-\bar{A}(s-t)} ba_s ds \right] = \mathbb{E}_t \left[ \int_t^\infty V^{-1} \left( Ve^{-\bar{A}(s-t)} V^{-1} \right) V ba_s ds \right]$$

$$= \mathbb{E}_t \left[ \int_t^\infty V^{-1} \left( \sum_k e^{-\lambda_k \tau} e_k e_k' \right) V ba_s ds \right]$$

$$= \mathbb{E}_t \left[ \int_t^\infty \left( \sum_k e^{-\lambda_k \tau} V^{-1} e_k e_k' V \right) ba_s ds \right] = \mathbb{E}_t \left[ \int_t^\infty \sum_{k=1}^2 e^{-\lambda_k(s-t)} w_k c_k' ba_s ds \right]$$

$$= \mathbb{E}_t \left[ \int_t^\infty \sum_{k=1}^2 e^{-\lambda_k(s-t)} v_k a_s ds \right]$$

as $v_k := w_k c_k' b$. □

**Discrete time.** In discrete time, the method is similar. The system is

$$z_t = \mathbb{E}_t [A(m) z_{t+1}] + ba_t$$

which yields, as $A(m)$ has eigenvalues less than 1 in modulus

$$z_t = \sum_{s \geq t} \mathbb{E}_t \left[ A(m)^{s-t} b(m) a_s \right].$$
After diagonalization:
\[ z_t = \mathbb{E}_t \left[ \sum_{s=t}^{\infty} \sum_{k=1}^{2} \Lambda_k^{s-t} v_k a_s ds \right] \]
where \( \Lambda_k \leq 1 \) are the eigenvalues of \( A(m) \).

## 11 Another Approach to Long-Run Changes

Here, I record another way of handling long run changes, alternative to the one discussed in Section 5. 

In behavioral models, agent’s actions and thoughts are anchored at a “default”. The “default” corresponds to: if the agent does not think, what kind of inflation does he expect? So far, I have assumed a constant default at 0 – this streamlines the analysis, at little cost in most situations. However, let us explore how to have a richer default, and what the consequences are.

### 11.1 Modelling the impact of long run policy expectation

For clarity, it is useful to be somewhat general and abstract. Suppose that we have a system

\[ z_t = A(m) \mathbb{E}_t [z_{t+1}] + b(m) a_t \quad (102) \]

where \( a_t \) is some exogenous “action” by the external world (e.g. the central bank), and \( z_t \) by endogenous variables. We assume, for the subjective model \( m \) of the agents, \( A(m) \) has only stable roots, less than 1 in modulus. Also, \( A^r = A(\iota) \) and \( b^r = b(\iota) \) are the responses that would happen where agents are rational (with \( \iota \) a vector of ones, representing full rationality); but \( A^r \) could have unstable roots (with eigenvalues greater than 1 in modulus). For instance, in our NK setup in (29) with passive policy \((\phi_x = \phi_\pi = 0)\), the behavioral response is \( A(m) = \begin{pmatrix} M & \sigma \\ \kappa M & \beta M^f + \kappa \sigma \end{pmatrix} \), and the rational response is \( A^r = A(\iota) = \begin{pmatrix} 1 \\ \kappa \end{pmatrix} \begin{pmatrix} \psi \\ \beta + \kappa \psi \end{pmatrix} \), where (from 19.3), I define

\[ \psi := \frac{1}{\gamma R} \quad (103) \]

---

69. Here, there is common knowledge that in the long run, frictions are unimportant – something that may be behaviorally a bit bold. Hence, I downgrade this section to the appendix.

70. In Bayesian models, the “default” is basically called the “prior” – which is a complex probability distribution, whereas the default is typically a point estimate.
which is basically the rational IES in the continuous time limit.

Suppose that we have a constant long run action: \( a_t = a \) for all \( t \). Then, (28) gives that the rational response \( z \) should satisfy:

\[
z = A^r z + b^r a,
\]

hence \( z = H^r a \), with

\[
H^r := (1 - A^r)^{-1} b^r.
\]  

(104)

Now consider an agent who forms, at time \( t \), some view of the long run action, e.g.

\[
a_{t}^{LR} = \lim_{\tau \to \infty} \mathbb{E}_t [a_{t+\tau}],
\]

(105)

but we will shortly consider a smoother version of this concept. Given the long run action \( a_{t}^{LR} \), the rational action is \( z_{t}^{LR} = b^r a_{t}^{LR} \).

Next, I posit that agents reason about the economy in the “deviation from the long run”, e.g. they think about a world:

\[
\hat{z}_{\tau|t} = A(m) \hat{z}_{\tau+1|t} + b(m) \hat{a}_{\tau|t}
\]

(106)

where \( \hat{z}_{\tau|t} \) and \( \hat{a}_{\tau|t} \) are the deviations from the time-\( t \) default:

\[
\hat{Z}_{\tau|t} := Z_{\tau} - m_{LR} Z_{t}^{LR} \text{ for } Z = z, a.
\]

Here, again \( m_{LR} \in [0, 1] \) is the weight on the LR as an anchor. In the formulation so far, we had \( m_{LR} = 0 \). If \( m_{LR} = 1 \), they think of economic outcomes as a deviation from the long run. I assume here that in their simulation at time \( t \), they set an anchor for the whole future path \( \hat{z}_{\tau|t} \) at times \( \tau \) after \( t \). I assume that agents do have access to this notion of “normatively correct long run response”, \( H(\nu) a_{t}^{LR} \). It is a bit of a strong assumption, though less strong than that of the traditional model. In future drafts, I plan to reexamine this assumption, and perhaps change it.

Calling \( H(m) = \sum_{i \geq 0} A(m)^i b(m) = (1 - A(m))^{-1} b(m) \), so that \( H(\nu) = H^r \).

---

\(^{71}\)Here I make an (arguably mild) equilibrium selection: I assume that a constant impulse \( a \) generates a constant response \( z \).

\(^{72}\)This is in the tradition of cognitive modelling, where thinking is anchored at a “default” and the agent considers partial adjustments from it (cf. Tversky and Kahneman (1974), Gabaix (2014, 2016)). Here, the default is the long run, which itself is influenced by the past, as we shall soon see.
Proposition 11.1 In the model with a non-zero long-run $a_t^{LR}$, we have

$$z_t = \sum_{\tau \geq t} \mathbb{E}_t [A(m)^{\tau-t} b(m) a_\tau] + m_{LR} (H(\iota) - H(m)) a_t^{LR}.$$ 

Proof. We have

$$z_t = \hat{z}_{t|t} + m_{LR} z_t^{LR} = \sum_{\tau \geq t} \mathbb{E}_t [A(m)^{\tau-t} b(m) (a_\tau - m_{LR} a_t^{LR})] + m_{LR} z_t^{LR}$$

$$= \sum_{\tau \geq t} \mathbb{E}_t [A(m)^{\tau-t} b(m) a_\tau] - m_{LR} \sum_{\tau \geq t} A(m)^{\tau-t} b(m) a_t^{LR} + m_{LR} H(\iota) a_t^{LR}$$

$$= \sum_{\tau \geq t} \mathbb{E}_t [A(m)^{\tau-t} b(m) a_\tau] + m_{LR} (H(\iota) - H(m)) a_t^{LR}.$$ 

□

This is what we had before, $z_t = \sum_{\tau \geq t} \mathbb{E}_t [A(m)^{\tau-t} b(m) a_\tau]$ (equation (33)) with a new term, $(H(\iota) - H(m)) a_t^{LR}$ for the adjustment to long run impact. When agents are rational, $m = \iota$ and this term is 0: there is no need for an extra adjustment term. When agents are less than rational, the anchor on the long term helps them be more rational.

To get clean expressions, I use the notations:

$$M = 1 - \xi, \quad M_f = \frac{1-\rho}{\beta}, \quad \beta = 1 - \rho + \chi$$

so that the rational case corresponds to $\xi = 0$ for consumers and $\chi = 0$ for firms; the expressions are similar in discrete and continuous time. Indeed, $(1 - \beta M_f) (1 - M) = \rho \xi$ then. Simple calculations show that we have:

$$H(m) = \frac{1}{\rho \xi - \kappa \sigma} (-\rho, \kappa) \sigma, \quad H(\iota) = \left( \frac{\rho - \chi}{\kappa}, 1 \right)$$

and $b^{LR}(m) = H(\iota) - H(m)$ is equal to

$$b^{LR}(m) = \frac{\rho \xi}{\rho \xi - \kappa \sigma} \left( \frac{\rho}{\kappa}, 1 \right)' - \left( \frac{\chi}{\kappa}, 0 \right).$$

(108)

Here, I gather the model.

\[73]\text{Actually, the value of } \kappa \text{ is a bit different in the rational model. This little bug will be fixed in the next iteration of the paper.}\]
Proposition 11.2 (Behavioral New Keynesian model, with adjustment for the long run) The generalization of the model with adjustment for the long run is as follows: Define \( \hat{x}_\tau, \hat{\pi}_\tau \) to be the solutions of the model of Proposition 2.6:

\[
\begin{align*}
\hat{x}_\tau &= M\mathbb{E}_\tau [\hat{x}_{\tau+1}] + b_d d_t - \sigma (i_{\tau} - \mathbb{E}_{\tau+1}\hat{\pi}_{\tau} - r^n_{\tau}) \quad (109) \\
\hat{\pi}_\tau &= \beta M f \mathbb{E}_\tau [\hat{\pi}_{\tau+1}] + \kappa \hat{x}_{\tau}. \quad (110)
\end{align*}
\]

The actual values of output and inflation are:

\[(x_t, \pi_t) = (\hat{x}_t, \hat{\pi}_t) + m_{LR} b_{LR} i_{LR} \]

where \( b_{LR} \) is given in (108), and \( i_{LR} \), the perception of long run policy, is given by \( i_{LR} = \lim_{\tau \to \infty} \mathbb{E}_t [i_{t+\tau}] \) in the “strict long run” case; or (112) applied to \( a_t = i_t \) in the “smoothed long run” case.

If indeed the action is constant, then \( a_{LR} = \bar{a} \), and the true long run is

\[\bar{z}_{LR} = [(1 - m_{LR}) H (m) + m_{LR} H (\bar{\nu})] \bar{a}.\]

Lemma 11.3 (In the long run, does inflation increase or decrease with interest rates?) Suppose that the nominal interest rate (minus the RBC normative interest rate) is constant at \( \bar{i} \) in the long run. Then, the steady state inflation is is:

\[\bar{\pi} = \left[ - (1 - m_{LR}) \frac{k\sigma}{\rho \xi - k\sigma} + m_{LR} \right] \bar{i}. \quad (111)\]

Hence, if \( m_{LR} = 1 \), long-run Fisher neutrality holds. More generally, if \( m_{LR} \) is close enough to 1, inflation increases with the interest rate in the long run.

Let me now detail the “smoothed long run” case.

11.2 A “Smoothed Long Run”

The simplest way to model the long run is \( a_{LR} = \lim_{\tau \to \infty} \mathbb{E}_t [a_{t+\tau}] \) (equation (105)). But, this notion captures only “mathematical infinity” and will not capture policies that last for 80 years, rather than forever. In addition, expectations of BR agents may be slow to adjust. Hence, I use a smooth
generalization of (105):

\[ a_t^{LR} = \sum_{D, \tau \geq 0} \mathbb{E}_{t-D} [a_{t-D+\tau}] g(D) f(\tau). \]  

(112)

Here, \( D \) represents a delay in the adjustment of the information set (as in Gabaix and Laibson 2002, Mankiw and Reis 2002), distributed according to \( g(D) = \phi e^{-\phi D} \) (i.e. \( g(D) = \phi (1 - \phi)^D \) in discrete time). Also, \( f(\tau) \) is the weight put on the future \( \tau \) periods ahead. In practice, I take \( f(\tau) = \zeta^2 e^{-\zeta \tau} \), which puts more weight on the future than on the immediate present.

When \( \zeta \to 0 \) and \( \phi \to \infty \), \( a_t^{LR} \) converges to \( \lim_{\tau \to \infty} \mathbb{E}_t [a_{t+\tau}] \). Let us evaluate its value for a typical case.

**Lemma 11.4** Suppose that a policy change \( a_t = e^{-\alpha t} a_0 \) is announced at time 0. Then, agents’ perception of its long-run value is (in continuous time):

\[ a_t^{LR} = \frac{\zeta^2 \phi}{(\zeta + \alpha)^2 (\phi - \alpha)} (e^{-\alpha t} - e^{-\phi t}) a_0. \]

For instance, if this is a permanent change, \( \alpha = 0 \), then \( a_t^{LR} = (1 - e^{-\phi t}) a_0 \). There is a delayed adjustment captured by \( \phi \). When \( \phi = \infty \) (no delay in expectations), then \( a_t^{LR} = a_0 \), expectations adjust immediately.

When the policy change will mean-revert (\( \alpha > 0 \)), then the “strict long run” is just 0: \( \lim_{\tau \to \infty} \mathbb{E}_t [a_{t+\tau}] = 0 \). However, the “smoothed long run” \( a_t^{LR} \) is not 0. Hence, a shock lasting, say, 50 years but not an infinite number of years is captured by the smoothed long run.

### 11.3 Impact of a Permanent Rise in the Nominal Interest Rate

We can now study the impact of a permanent rise in the nominal interest rate: \( i_t \) increase by \( J = 1\% \).

I take the following parameters, in continuous time with yearly units: \( \xi_0 = 0.45, \rho^f = -\ln (\beta Mf) = 0.2, m_{LR} = 1, \kappa_0 = 0.6, \sigma = 0.1, \phi = 0.25, \zeta = 0.05 \). They have not been particularly optimized.

The effect is shown in Figure 9. On impact, the rise in the rate lowers inflation and output – this is the conventional Keynesian effect. In the long run, however, the Fisherian prediction holds: the 1\% rise in the interest rate is coupled with a 1\% rise in inflation, so that the long term interest rate is unchanged.\(^74\)

As in Section 5.3, this effect is very hard to obtain in a conventional New Keynesian model\(^74\) However, this neutrality is partial – as in the basic NK model, output does increase permanently if inflation is permanently higher. This effect, however, is quite small.

\(^74\)
Figure 9: 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 percents.

(Cochrane (2015)). However, here again, the bounded rationality of the agents overturns this result, with just one bounded equilibrium.

To see analytically why, it is worth examining two polar cases. First, take the full rationality case in Proposition 2.6, we have

\[ x = x - \sigma (J - \pi), \]
\[ \pi = \beta \pi + \kappa x_t. \]  

Then, the solution with a constant coefficient is: \( \pi_t = J, x = \frac{(1-\beta)}{\kappa} J. \) Hence, a higher interest rate leads to higher inflation, as in the Fisher neutrality.

However, take a completely myopic model, so that \( M = M^f = 0. \) Then, Proposition 2.6 reduces to:

\[ x_t = -\sigma (i_t - \mathbb{E}_t \pi_{t+1} - r^n_t), \]
\[ \pi_t = \kappa x_t, \]

hence:

\[ \pi_t = \kappa \sigma \mathbb{E}_t \pi_{t+1} - \sigma J \]

so that a higher interest rate leads to lower inflation.

The enriched model with long run expectations gives something in between those two polar models. Hence, it generates the first Keynesian dynamics, with a high interest rate lowering inflation,
Figure 10: Impact of a permanent 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 percents.

and the long run Fisher effect of a higher inflation to restore constant real rates.

11.4 Impact of a Temporary Rise in Interest Rates

Let us now consider a short term rise in the interest rate, Figure 10. We find indeed that a temporary rise in interest rates decreases inflation and output. This is a result that was hard to get in NK models (though again this depends on issues of equilibrium selection), see Cochrane (2015). Here we get it easily, with a determinate equilibrium.