Forward Guidance without Common Knowledge*

George-Marios Angeletos[†] **Chen Lian**[‡]

December 29, 2017

Abstract

How does the economy respond to news about future policies or future fundamentals? Workhorse models answer this question by imposing common knowledge of the news and of their likely effects. Relaxing this assumption anchors the expectations of future outcomes, effectively leading to heavier discounting of the future. By the same token, general-equilibrium mechanisms that hinge on forward-looking expectations, such as the Keynesian feedback loop between inflation and spending, are attenuated. We establish these insights within a class of games ("dynamic beauty contests") which nest the New Keynesian model along with other applications. We next show how these insights help resolve the forward-guidance puzzle and offer a rationale for the front-loading of fiscal stimuli.

Keywords: Forward-guidance puzzle, fiscal stimuli, Keynesian multiplier, deflationary spiral, incomplete information, higher-order uncertainty, coordination, beauty contests.

*We are grateful to to Alisdair McKay, Alexandre Kohlhas, Kristoffer Nimark, Franck Portier, and Mirko Wiederholt for conference discussions. For helpful comments, we also thank Adrien Auclert, Harris Dellas, Xavier Gabaix, Alessandro Pavan, Karthik Sastry, and seminar participants at the 2016 and 2017 NBER Summer Institutes, the 2016 ESEM in Edinburgh, the 2017 NBER EFG and ME meetings, the 2017 SED meeting, the NYU Stern IUB workshop, the ECB NSM workshop, the 2017 Hydra Workshop in Sardinia, the 20th Central Bank Macroeconomic Modeling Workshop at the Bank of France, the University of Chicago, Duke University, Northwestern University, and Princeton University.

†MIT and NBER; angelet@mit.edu.

‡MIT; lianchen@mit.edu.

1 Introduction

How does the economy respond to news about future economic conditions or future policies? Consider the example of forward guidance, that is, news about future monetary policy. Such news are meant to influence current economic activity, not only by shifting the term structure, but also by "managing expectations" of future outcomes such as output and inflation.[1]

In macroeconomics and finance alike, this kind of forward-looking expectations is typically pinned down by assuming that the agents have common knowledge of the state of the economy and of one another's beliefs about the future. This imposes a certain "perfection" in the way agents adjust their expectations of future economic outcomes, and their decisions, to any news about the future. We instead perturb this perfection and proceed to show how this anchors the relevant expectations and attenuates the general-equilibrium effects of news, especially when the latter relates to events in the more distant future.

Our contribution proceeds in three steps. We start by revisiting the New Keynesian model and by embedding its building blocks, the modern analogues of the IS and Philips curves, into a class of "dynamic beauty contests." This class stylizes environments in which aggregate outcomes today—such as consumer spending, inflation, or asset prices—depend positively on expectations of the aggregate outcomes tomorrow.

Within this broader framework, we show that the absence of common knowledge anchors the expectations of future outcomes at a rate that increases with the horizon of the news. It is therefore *as if* the agents are discounting the influence of the distant future to current decisions more heavily than what seems rational. By the same token, general-equilibrium mechanisms that rest on the frictionless adjustment of forward-looking expectations are attenuated.

We finally return to the New Keynesian model and apply our insights to the context of a liquidity trap. In this context, the absence of common knowledge anchors expectations of future inflation and income, attenuates the general-equilibrium effects that operate both within and across the two blocks of the model, lessens the forward-guidance puzzle, and offers a rationale for the front-loading of fiscal stimuli.

The friction. By "lack of common knowledge" or "incomplete information" we refer to situations in which the agents have an imperfect and idiosyncratic understanding of the exogenous impulses and of their equilibrium implications (for instance, of the policy news and of their effects on output and inflation). This could be the product of dispersed private information, which gets only imperfectly aggregated through markets. But it could also be the product of rational inattention (Sims, 2003; Tirole, 2015). The key is that, similarly to the literatures on "global games" and "beauty contests" (Morris and Shin, 1998, 2002, 2003; Angeletos and Lian, 2016*c*), we accommodate an imperfection in the coordination of beliefs and decisions by allowing for higher-order uncertainty (i.e., uncertainty about the beliefs and the decisions of others).

Removing common knowledge from the New Keynesian model. Similarly to Mankiw and Reis (2002), Woodford (2003*a*) and Mackowiak and Wiederholt (2009), our setting allows the firms to have an imperfect and idiosyncratic understanding of the state of the economy. But it also adds two other features.

[1] For evidence of how such news matter and for the difficulty of workhorse models to capture this evidence, see Campbell et al. (2012) and Del Negro, Giannoni and Patterson (2015).

First, we let the informational friction coexist with the Calvo friction so as to include a forward-looking element in the firms' price-setting decisions. This links current inflation to expectations of future inflation, giving rise to a GE feedback mechanism within the supply block of the model.[2]

Second, we assume that the consumers, too, to have an imperfect and idiosyncratic understanding of the state of the economy. This allows us to pay closer attention to the GE mechanism that operates within the demand block of the model and that regards the feedback loop between individual and aggregate spending for given real interest rates. This is a modern version of the Keynesian multiplier.[3]

We show how each of the two blocks of the model can be represented as a "dynamic beauty contest." With this term we refer to a class of games in which the aggregate outcome in any given period depends on the average expectation of that outcome in the future; such games nest, not only the two blocks of the New Keynesian model, but also other applications, such as the dispersed-information asset-pricing models studied in Singleton (1987) and Allen, Morris and Shin (2006).

When agents share the same information and the same beliefs in all dates and all realizations of uncertainty (clearly, a very strong assumption), the obtained beauty contests reduce to the familiar equations of the textbook New Keynesian model, namely the representative consumer's Euler condition (a.k.a. dynamic IS curve) and the New Keynesian Philips curve. Away from that benchmark, our representation helps unearth the role of the rich higher-order beliefs that are hidden beneath these deceptively simple equations.

Attenuation and horizons in dynamic beauty contests. Motivated by the previous result, in Section 5 we study how lack of common knowledge influences the equilibrium outcomes of the aforementioned class of games. We focus, in particular, on the following question: how do outcomes in period t (say, $t = 0$) vary with the current average forecast of the fundamental in period $t + T$?

Let ϕ_T denote the response of the former to the latter.[4] We establish three properties. First, the absence of common knowledge reduces ϕ_T regardless of how precise the available information may be; we refer to this result as the "attenuation effect." Second, this attenuation is increasing in T; we refer to this result as the "horizon effect." Third, under a mild condition, ϕ_T becomes vanishing small relative to its common-knowledge counterpart (i.e., the attenuation effect increases without bound) as $T \to \infty$.[5]

Let us sketch the proof of these findings. In our setting, there is a unique rational-expectations equilibrium, regardless of the information structure. Along this equilibrium, the aggregate outcome today can be expressed as a function of the first- and higher-order beliefs of the fundamental T periods later. When common knowledge is imposed *in addition* to the rational-expectations hypothesis, higher-order beliefs collapse to first-order beliefs, yielding the standard predictions. When, instead, common knowledge is broken,

[2] Nimark (2008) also employs both frictions but pursues other objectives.

[3] In the textbook version of the New Keynesian model, the Keynesian multiplier reflects merely the dependance of consumption on permanent income. In more realistic settings, it contains additional mechanisms through which consumer spending today responds to expectations of economic conditions in the short to medium run, such precautionary motives, borrowing constraints, and "hand-to-mouth consumers" (Galí, López-Salido and Vallés, 2007; Kaplan, Moll and Violante, 2016; Auclert, 2017).

[4] More precisely, because both the average forecast of the future fundamental and the aggregate outcome are random variables, ϕ_T is defined as the coefficient of regressing the former on the latter.

[5] The sharpest version of these results is obtained under the assumption that the information structure remains constant between t and $t + T$, at which point the true fundamental becomes known. This also clarifies that our contribution is different from prior work that focuses on the effects of learning.

higher-order beliefs move less than one-to-one with first-order beliefs. This explains the attenuation effect. The horizon effect, on the other hand, follows from the combination of the following two properties: that the covariation between first- and h-order beliefs falls with h; and that longer horizons raise the relative importance of higher-order beliefs. The first property is a hallmark of higher-order uncertainty. The second follows from the fact that longer horizons involve more iterations of the best responses (the forward-looking, or Euler-like, equations of the model), which in turn map to beliefs of higher order. In short, iterating on the forward-looking equations of a model is akin to ascending the hierarchy of beliefs.

We complement these findings with an additional result, which helps recast ϕ_T as the solution to a representative-agent model in which the forward-looking expectations that regard aggregate (as opposed to own) outcomes are discounted by a factor that is directly related to the severity of the informational friction. It is therefore *as if* each agent is "myopic" vis-a-vis the future aggregate dynamics.

Policy lessons. In light of these broad insights, we revisit the predictions that the New Keynesian model makes about monetary and fiscal policy in a liquidity trap. To do this, we must deal with the higher-layer beauty contest that obtains once we take into account the *interaction* of the two blocks of the model, as opposed to looking at each block in isolation from the other. The essence, however, remains the same.

We first consider the so-called forward guidance puzzle. This puzzle refers to the prediction that a credible promise to keep monetary policy lax far in the future can have an incredibly large effect on current economic activity. Suppose, in particular, that the nominal interest rate is pegged to zero between $t = 0$ and $t = T - 1$, for some known $T \geq 2$, due to a binding zero lower bound (ZLB) constraint. Because of this constraint, the monetary authority is unable to stimulate the economy by reducing its current policy rate. It can nevertheless try to achieve the same goal by committing to low interest rates at $t = T$ (or later), that is, after the economy has exited the liquidity trap and the ZLB constraint has ceased to bind. The standard New Keynesian model predicts the effect of such increases with T and explodes as $T \to \infty$: the longer the horizon at which the interest rate is lowered, the larger the effect on current aggregate spending. What is more, for plausible parameterizations, the effect is quantitatively large even for modest T.

These predictions are at odds with the available evidence (Del Negro, Giannoni and Patterson, 2015). They are also considered to be a priori counter-intuitive. In our view, this is largely because these predictions are driven by general-equilibrium mechanisms, which hinge on large adjustments in expectations of future outcomes and which work at "full capacity" when common knowledge is imposed.

To appreciate what we mean, consider the following question: how does an individual respond to news about a future interest-rate change that applies *only to herself*, as opposed to the entire economy? Answering this question helps isolate the direct or partial-equilibrium (PE) effect of future interest rates from their general-equilibrium (GE) effects. Because agents discount the future, the PE effect diminishes with T and vanishes as $T \to \infty$. It follows that the aforementioned predictions are driven by GE mechanisms.

What are these GE mechanisms? Two of them have already been mentioned: one is the Keynesian multiplier that is hidden behind the IS curve, another is the feedback from future to current inflation that is hidden behind the NKPC. But there is an additional GE mechanism, which is operates across the two blocks of the model: the feedback loop between aggregate spending and inflation. Reducing the nominal interest rate at $t = T$ causes inflation at $t = T$. Because the nominal interest rate is pegged prior to T, this translates to a low *real* interest rate between $T-1$ and T. This stimulates aggregate spending at $T-1$, contributing to even higher inflation at $T-1$, which in turn feeds to even higher spending at $T-2$, and so on. Clearly, the cumulative effect at $t = 0$ increases with T, which explains why the power of forward guidance also increases with T—indeed without bound—according to the New Keynesian model.[6]

Removing common knowledge attenuates *all* the three GE mechanisms, and the more so the larger T is, thus also reducing the gap between macroeconomic predictions and partial-equilibrium intuitions. To illustrate, suppose that each agent (a consumer or firm) worries that any other agent may be unaware of the policy news with a 25% probability. Under a textbook parameterization of the New Keynesian model, this "doubt" about the awareness of others causes the power of forward guidance to be reduced by about 90% at the 5-year horizon. Importantly, this attenuation is relative to the movements in *expectations* of interest rates. Our theory therefore helps explain why forward guidance may a have a weak effect on expectations of inflation and income *relative* to its effect on expectations of future rates.[7]

Turning to fiscal policy, the standard New Keynesian model predicts that, in the presence of a binding ZLB constraint, a fiscal stimulus of a given size is more effective when it is back-loaded, i.e., when it is announced now but implemented later on.[8] This prediction hinges on the same GE mechanisms and the same kind of forward-looking expectations as those that govern the power of forward guidance. A variant of the aforementioned results therefore applies: removing common knowledge reduces the magnitude of fiscal multipliers, and the more so the longer the horizon of the fiscal stimulus. To put it differently, our paper offers a rationale for the front-loading of fiscal stimuli.

Layout. The remainder of the paper is organized as follows. Section 2 discusses the related literature. Section 3 introduces our incomplete-information version of the New Keynesian model. Section 4 nests the demand and supply blocks of the New Keynesian model within a class of dynamic beauty contests. Section 5 studies this class of games and develops our key insights regarding the effects of news about the future. Sections 6 and 7 work out the implications for, respectively, monetary and fiscal policy during a liquidity trap. Section 8 concludes the main text. The Appendices contain the proofs and a few additional results.

[6]This feedback loop is the flip-side of the "deflationary spiral" that helps the New Keynesian model generate a sizable recession during a liquidity trap. The initial trigger is different, and the sign is reversed, but the GE mechanism at work is the same.

[7]Let us also emphasize that the issue at stake is neither the credibility of central-bank communications (Bassetto, 2015) nor the size of their effect on the term structure of interest rates (can be thought of as a proxy for the expectations of future rates). For our purposes, the shifts in expectations of future interest rates are the exogenous impulses and the question of interest is how these impulses translate to movements in expectations of endogenous outcomes such as income and inflation. Finally, note that we rule out the confounding of news about one kind of fundamentals (say, about future interest rates) with news about another kind of fundamentals (say, about future growth potential). We are thus concerned with what Campbell et al. (2012) call "Odyssean" forward guidance. The alternative scenario, that of "Delphic" forward guidance, is emphasized in Andrade et al. (2015).

[8]See Christiano, Eichenbaum and Rebelo (2011), Woodford (2011) and Werning (2012).

2 Related Literature

Our paper builds on the macroeconomic literature on incomplete information and beauty contests; see Angeletos and Lian (2016*c*) for a survey. Morris and Shin (2002), Woodford (2003*a*), Angeletos and Pavan (2007), Angeletos and La'O (2010, 2013), and Bergemann and Morris (2013) focus on *static* beauty contests, namely, settings in which the agents are *not* forward-looking and only have to forecast the *concurrent* actions of others. By contrast, Allen, Morris and Shin (2006), Bacchetta and van Wincoop (2006), Morris and Shin (2006), and Nimark (2008, 2017) study *dynamic* beauty contests, namely, settings in which the agents must forecast the *future* actions of others.

Our paper shares with the latter set of works the emphasis on forward-looking expectations; the class of beauty contests studied in these works is indeed a special case of the broader class studied here. But whereas these works focus on how outcomes vary with the learning that takes place over time, our paper focuses on how outcomes vary with the horizon of the events that the agents have to forecast. This explains our main theoretical contribution, which is to show how longer horizons translate to larger attenuation, holding constant the information structure. An additional contribution, which is crucial for our applied purposes, is to map the building blocks of the New Keynesian model to the considered class of games.

By allowing for dispersed private information, we also connect to Lucas (1972). Note, however, that the latter abstracts from strategic complementarity and studies, in effect, a single-agent decision context featuring confusion between two kinds of fundamentals (namely, monetary shocks and relative demand shocks). By contrast, we do not allow the news of future monetary policy to be confounded with news of other fundamentals. And whereas higher-order uncertainty is irrelevant in Lucas (1972), it is of absolute essence in our context.[9]

On the applied side, our paper adds to the literature on the forward-guidance puzzle. Del Negro, Giannoni and Patterson (2015) and McKay, Nakamura and Steinsson (2016*b*) seek to resolve the puzzle by modifying the decision rules of the model: the one by introducing short horizons and adjustment costs ("internal habit"), the other by adding liquidity constraints. We make an orthogonal point: we anchor the adjustment of the expectations of future inflation and income, thus also dulling the GE effects of forward guidance for given decision rules. This kind of friction is consistent with the expectations evidence in Coibion and Gorodnichenko (2012, 2015) and Vellekoop and Wiederholt (2017).

Closely related in this regard is Wiederholt (2015). Similarly to ours, this paper stresses how higher-order uncertainty anchors inflation expectations and reduces the power of forward guidance. However, the modeling choices of that paper prevent one from detecting the higher-order beliefs that operate *within* each block of the New Keynesian model.[10] Most importantly, that paper does not study how the bite of higher-order uncertainty varies across horizons, which is the core theme of our paper.

[9] For a detailed explanation of why higher-order uncertainty is irrelevant in Lucas (1972), see Section 8.4 of Angeletos and Lian (2016*c*).

[10] On the supply side, the role of higher-order beliefs is shut down by assuming that the firms have common knowledge; and on the demand side, the attenuation of "income multiplier" is assumed away by letting the consumers can pool their income through insurance even though they cannot pool their information.

Chung, Herbst and Kiley (2015) and Kiley (2016) argue that some of the paradoxical predictions of the New Keynesian model are resolved once the nominal rigidity is attributed to sticky information as in Mankiw and Reis (2002). But this is largely because these works abstract entirely from the price stickiness seen at the micro data and from the forward-looking aspect in the firms' price-setting decisions. These works also rule out information frictions among the consumers. They therefore do not share our insights regarding the anchoring of forward-looking expectations, and the attenuation of the associated GE effects.

Last but not least, our paper is related to Garcıa-Schmidt and Woodford (2015), Farhi and Werning (2017), and Gabaix (2016). These works drop Rational Expectations from the New Keynesian model. The first two consider Level-k Thinking (or a close cousin of it, "Reflective Equilibrium"); the third assumes that the agents apply a "cognitive discount" on the influence of shocks on future outcomes. By design, these concepts arrest the responsiveness of expectations and attenuate the associated GE effects. We instead show that similar effects are accommodated without the methodological sacrifice of dropping Rational Expectations, once one "liberates" that concept from auxiliary common-knowledge assumptions.[11]

3 Framework

In this section, we introduce the framework used for the applied purposes of the paper. This is the same as the textbook New Keynesian model (Woodford, 2003*b*; Galí, 2008), except that we relax the common-knowledge assumption and let the agents have different beliefs about the future prospects of the economy.

Consumers. There is a measure-one continuum of ex-ante identical consumers in the economy, indexed by $i \in \mathcal{I}_c = [0,1]$. The preferences of consumer i are given by

$$\mathcal{U}_0 = \sum_{t=0}^{+\infty} \beta^t U(c_{i,t}, n_{i,t}), \tag{1}$$

where $c_{i,t}$ and $n_{i,t}$ denotes her consumption and labor supply at period t, $\beta = e^{-\rho} \in (0,1)$ is the discount factor, and U is the per-period utility function. The latter is specified as $U(c,n) = \frac{1}{1-1/\sigma} c^{1-1/\sigma} - \frac{1}{1+\epsilon} n^{1+\epsilon}$, where $\sigma > 0$ is the elasticity of intertemporal substitution and $\epsilon > 0$ is the inverse of the Frisch elasticity. The budget constraint in period t is given, in nominal terms, by the following:

$$p_t c_{i,t} + a_{i,t} = R_{t-1} a_{i,t-1} + p_t (w_{i,t} n_{i,t} + e_{i,t}), \tag{2}$$

where p_t is the price level in period t, $a_{i,t}$ is the household's nominal saving in that period, R_{t-1} is the nominal gross interest rate between between $t-1$ and t, and, finally, $w_{i,t}$ and $e_{i,t}$ are, respectively, the real wage and the real dividends received in period t. The wage and the dividend are allowed to be household-specific for reasons that will be explained shortly. We finally denote aggregate consumption by $c_t = \int_{\mathcal{I}_c} c_{i,t} di$, aggregate labor by $n_t = \int_{\mathcal{I}_c} n_{i,t} di$, and so on.

[11] For a sharper exposition of how the aforementioned departures from rational expectations compare to our approach, which maintains rational expectations but removes common knowledge, see Angeletos and Lian (2016*a*).

Firms. There is a measure-one continuum of ex-ante identical firms, indexed by $j \in \mathcal{I}_f = (1, 2]$. Each of these firms is a monopolist that produces a differentiated intermediate-good variety. The output of firm j is denoted by y_t^j, its nominal price is denoted by p_t^j, and its realized profit by e_t^j. The technology is assumed to be linear in labor and productivity is fixed to one, so that

$$y_t^j = l_t^j, \tag{3}$$

where l_t^j is the labor input. These intermediate goods are used by a competitive sector as inputs in the production of the final good. The technology is CES with elasticity ς. Aggregate output is thus given by

$$y_t = \left(\int_{\mathcal{I}_f} \left(y_t^j \right)^{\frac{\varsigma - 1}{\varsigma}} dj \right)^{\frac{\varsigma}{\varsigma - 1}}, \tag{4}$$

The corresponding price index—that is, the price level—is denoted by p_t.

Sticky Prices. We introduce nominal rigidity in the usual, Calvo-like, fashion: in each period, a randomly selected fraction $\theta \in (0, 1]$ of the firms must keep their prices unchanged, while the rest can reset them.

Monetary Policy. For the time being, we do not need to specify how monetary policy is conducted. To fix ideas, however, it is useful to think of a situation where the agents know what the policy will be in the short run (say, because the ZLB is binding) but face uncertainty about the policy in the more distant future. In this context, the question of interest is how outcomes today vary with the expectations of the future policy.

Rational Expectations and Common Knowledge. Throughout, we impose Rational Expectations Equilibrium (REE). We nevertheless depart from standard practice by letting the agents lack common knowledge of the underlying state of Nature and, thereby, of the future path of the economy.[12] In the aforementioned context, this translates to letting the agents lack common knowledge of the future policy and of its likely impact on inflation and income.

We will spell out the details in due course. For the time being, let us clarify the following point. In abstract games, such as those studied in Morris and Shin (2002, 2003) or in Section 5 of our paper, lack of common knowledge can be *directly* imposed by endowing each agent with exogenous private information (or different Harsanyi types). In macro and finance applications, however, there is the complication that market signals, such as prices, aggregate information. To preserve the absence of common knowledge, we either have to make sure there is "noise" in the available market signals (Grossman and Stiglitz, 1980), or assume that the observation of all the relevant variables—including the available market signals—is contaminated by idiosyncratic noise due to rational inattention (Sims, 2003, Woodford, 2003*a*).

We follow the first approach within the context of our New Keynesian model in order to make clear that our policy lessons are robust to allowing each consumer to observe perfectly her own income, each firm to observe perfectly her own demand and supply conditions, and everybody to have common knowledge of the

[12] By definition, the state of Nature contains the entire profile of the information sets in the population. In equilibrium, information sets pin down beliefs of future outcomes. It follows that the state of Nature can encode arbitrary "news" about the future, and that the agents can lack a common belief about the future outcomes only if they lack common knowledge of the state of Nature.

current monetary policy and the current price level. We nevertheless like the second approach, too, because it bypasses the need for the "auxiliary shocks" described below and because it allows the re-interpretation of the assumed friction as the product of cognitive limitations, or "costly contemplation" (Tirole, 2015). As part of the more abstract analysis in Section 5, we thus sketch how our main insights can be recast under the lenses of rational inattention.

The auxiliary shocks. We conclude this section by describing the auxiliary shocks, and the associated informational assumptions, that permit us to apply the first of the aforementioned two modeling approaches. Take any period t and let w_t, e_t, and μ_t denote, respectively, the average real wage, the average firm profit, and the average markup in the economy. The real wage and the dividend received by consumer i at t are given by, respectively, $w_{i,t} = w_t \xi_{i,t}$ and $e_{i,t} = e_t \zeta_{i,t}$, where $\xi_{i,t}$ and $\zeta_{i,t}$ are i.i.d. across i and t, independent of one another, and independent of any other random variable in the economy. On the other hand, the real wage paid by firm j is $w_t^j = w_t u_t^j$, and the markup charged by it is $\mu_t^j = \mu_t \nu_t^j$, where u_t^j and ν_t^j are i.i.d. across both j and t, independent of one another, and independent of any other random variable in the economy. One can interpret u_t^j and ν_t^j as idiosyncratic shocks to a firm's marginal cost and her optimal markup, and $\xi_{i,t}$ and $\zeta_{i,t}$ as idiosyncratic shocks to a household's labor and financial income.[13]

As anticipated, the modeling role of all these shocks is to "noise up" the information that each individual agent—be it a consumer or a firm—can extract from the available market signals about the future economic conditions. We can thus accommodate the desired friction while allowing, in every t, each consumer i to have private knowledge of $(w_{i,t}, e_{i,t})$, each monopolist j to have private knowledge of (w_t^j, μ_t^j), and everybody to have common knowledge of the current interest rate, R_t, of the current prices, $(p_t^j)_{j \in [0,1]}$, and therefore also for the current inflation rate.[14]

Log-linearization. To keep the analysis tractable, we work with the log-linearization of the model around a steady state in which inflation is zero and the nominal interest rate equals the natural rate (i.e., $\beta R = 1$). With abuse of notation, we henceforth reinterpret all the variables as the log-deviations from their steady state counterparts. We also let the means of the (logs of the) idiosyncratic shocks be zero.

4 Deconstructing the New Keynesian Model into Two Beauty Contests

In this section, we develop our beauty-contest representation of the New Keynesian model. Apart from facilitating the subsequent analysis, this representation clarifies the general-equilibrium mechanisms and the kind of forward-looking expectations that are hidden behind the familiar equations of the textbook version of that model.

[13] To simplify the exposition, we treat the aforementioned shocks as exogenous. Yet, it is not hard to see how the "missing micro-foundations" could be filled in. For instance, Angeletos, Iovino and La'O (2016) engineer the desired firm-specific markup shocks by allowing for good-specific shocks that shift the elasticity of the demand faced by each monopolist.

[14] We could have also allowed common knowledge of the current aggregate output by adding demand (discount-factor) shocks.

The two beauty contests. We let $\pi_t = p_t - p_{t-1}$ denote the period-t inflation rate and study the joint determination of π_t and c_t (or, equivalently, of π_t and y_t). The next result establishes that, regardless of the information structure, the equilibrium values of these variables can be understood as the joint solution to two fixed-point problems, one for the demand block (consumers) and another for the supply block (firms). After stating the result and explaining its derivation, we discuss how these fixed-point problems can be interpreted as "dynamic beauty contests" and elaborate on the usefulness of this interpretation.

Proposition 1 (Beauty Contests) *In any equilibrium, and regardless of the information structure, the following properties are true. First, aggregate spending satisfies*

$$c_t = -\sigma \sum_{k=1}^{+\infty} \beta^{k-1} \bar{E}_t^c[r_{t+k}] + \tfrac{1-\beta}{\beta} \left\{ \sum_{k=1}^{+\infty} \beta^k \bar{E}_t^c \left[c_{t+k}\right] \right\}, \tag{5}$$

where $\bar{E}_t^c[\cdot] \equiv \int_{\mathcal{I}_c} E_{i,t}[\cdot] di$ denotes the average expectation of the consumers and $r_t \equiv R_{t-1} - \pi_t$ denotes the realized real interest rate between $t-1$ and t. Second, inflation satisfies[15]

$$\pi_t = \sum_{k=0}^{+\infty} (\beta\theta)^k \, \bar{E}_t^f \left[mc_{t+k}\right] + \tfrac{1-\theta}{\theta} \sum_{k=1}^{+\infty} (\beta\theta)^k \, \bar{E}_t^f \left[\pi_{t+k}\right] + \mu_t, \tag{6}$$

where $\bar{E}_t^f[\cdot] \equiv \int_{\mathcal{I}_f} E_{j,t}[\cdot] dj$ denotes the average expectation of the firms, $mc_{t+k} = \kappa y_{t+k}$ is the real marginal cost in period $t+k$, and $\kappa \equiv \frac{(1-\theta)(1-\beta\theta)(\epsilon+1)}{\theta} \geq 0$. Finally, $y_t = c_t$.

Condition (5) is a general-equilibrium variant of the textbook version of the Permanent Income Hypothesis. It follows, essentially, from log-linearizing the individual consumption functions, aggregating across all consumers, and using market clearing in the product, labor, and debt markets. It gives current aggregate consumption as an increasing function of the consumers' average expectation of the future path of aggregate income, which in equilibrium coincides with future aggregate consumption, and as a decreasing function of the expected path of the real interest rates.

Condition (6), on the other hand, follows from aggregating the optimal reset prices of the firms. To see the logic, recall that the optimal reset price of any given firm $j \in \mathcal{I}_f$ is given by her expected discounted present value of her current and future *nominal* marginal cost:

$$p_t^{j*} = (1-\beta\theta) \left\{ (mc_t^j + p_t) + \sum_{k=1}^{+\infty} (\beta\theta)^k \, E_{j,t} \left[mc_{t+k}^j + p_{t+k} \right] \right\} + (1-\beta\theta) \, \mu_t^j, \tag{7}$$

where $E_{j,t}^f[\cdot]$ denotes the firm's expectation in period t, mc_t^j is its *real* marginal cost in period t, and μ_t^j is its (exogenous) markup shock. Aggregating this condition across all the firms that get to reset their prices in period t, and using the fact that the resulting inflation is $\pi_t = (1-\theta)(p_t^* - p_{t-1})$, where $(1-\theta)$ is the fraction of firms that reset their prices and $p_t^* \equiv \int_{\mathcal{I}_f} p_t^{j*} dj$ is the average reset price, we get condition (6).

[15] To economize notation, when we transition from (7) to (6), we rescale the markup shock: the μ_t that appears in (6) equals $\frac{\kappa}{1+\epsilon}$ times the original μ_t that appears in (7).

For our purposes, the key observation is that conditions (5) and (6) can be interpreted as a pair of dynamic beauty-contest games, one capturing the "demand block" of the New Keynesian model and another capturing its "supply block." Let us elaborate.

Consider first the aggregate behavior of the consumers and, momentarily, treat the process for the real interest rate as exogenous. Condition (5) then defines a game in which the players are the consumers, the actions are the consumers' spending levels at different periods, and the "fundamental" is the path of real interest rates. A dynamic strategic complementarity is present in this game because the optimal spending of each consumer depends on her expectation of her future income, which in turn depends on the spending decisions of other consumers in the future. Intuitively, when a consumer expects the others to spend more in the future, she expects her own income to be high, and finds it optimal to spend more now. Condition (5) therefore isolates the strategic complementarity that operates within the demand block of the model and that encapsulates the GE mechanism that we refer to as the "income multiplier."[16]

Consider next the aggregate behavior of the firms and, momentarily, treat the process of the aggregate real output and, hence, of the average real marginal cost as exogenous. Condition (6) then defines a game in which the players are the firms, the actions are the prices, and the "fundamental" is the path of the average real marginal cost. A dynamic strategic complementary is present in this game because the optimal reset price of a firm depends on her expectation of future real marginal costs and of future price levels, which in turn depend on the pricing decisions of other firms. Intuitively, when firms expect inflation to be high, they also expect their nominal marginal costs to be high; they therefore find it optimal to set higher nominal prices when given the option to adjust, contributing to higher inflation today. Condition (6) therefore isolates the strategic complementarity that operates within the supply block.

The preceding discussion looks at each block of the model in isolation from the other. The interaction of the two blocks can be understood as a "meta-game" among the consumers and the firms. This meta-game features strategic complementarity in the following sense: the consumers find it optimal to spend more when they expect the firms to raise their prices, because high inflation means low real returns to saving; and, symmetrically, the firms find it optimal to raise their prices when they expect the consumers to spend more, because high aggregate output means high marginal costs. It is this kind of strategic complementarity, or GE mechanism, that we refer as the "inflation-output feedback."[17]

From the beauty contests to the Euler condition and the NKPC. Before proceeding to the core of our paper, it is useful to relate our somewhat exotic representation of the New Keynesian model to its more familiar, textbook treatments.

[16]In the simple version of the New Keynesian model that we employ in this paper, this multiplier reflects *merely* the dependence of consumption on permanent income. In more realistic versions, it contains additional mechanisms via which aggregate demand responds to expectations of economic conditions in the short to medium run, such as precautionary motives or the feedback from expectations of economic activity to housing prices, and thereby to borrowing constraints and aggregate spending. We would like to think of our beauty-contest representation of the demand block as a proxy for this kind of forces as well.

[17]This particular from of strategic complementarity, and its attenuation in the presence of informational frictions, is emphasized also in Wiederholt (2015),

Since Proposition 1 holds true regardless of the information structure, it *has* to nest the standard version of the New Keynesian model as a special case. Indeed, suppose that information is complete, namely, that the agents share the same expectations about all the relevant future variables (interest rates, inflation, income) in all periods and all states of Nature. In this case, we have that $E_{i,t}[.] = E_t[\cdot]$ for every consumer $i \in \mathcal{I}_c$ and every firm $i \in \mathcal{I}_f$ and, therefore, $\bar{E}_t^c[\cdot] = \bar{E}_t^f[\cdot] = E_t[\cdot]$, where $E_t[\cdot]$ denotes the rational expectation conditional on all information in the economy available at period t (i.e., the information of the representative agent).[18] Using this fact along with condition (5), we can express the period-t spending as follows:

$$c_t = -\sigma E_t\left[r_{t+1}\right] + (1-\beta)E_t\left[c_{t+1}\right] - \sigma \sum_{k=2}^{+\infty} \beta^{k-1} E_t[r_{t+k}] + \tfrac{1-\beta}{\beta} \sum_{k=2}^{+\infty} \beta^k E_t\left[c_{t+k}\right].$$

Next, using Law of Iterated Expectations, we can rewrite the above as

$$c_t = -\sigma E_t[r_{t+1}] + (1-\beta)E_t\left[c_{t+1}\right] + \beta E_t\left[-\sigma \sum_{k=1}^{+\infty} \beta^{k-1} E_{t+1}[r_{t+1+k}] + \tfrac{1-\beta}{\beta} \sum_{k=1}^{+\infty} \beta^k E_t\left[c_{t+1+k}\right]\right].$$

But now note that, from condition (5), the term inside the big brackets equals c_{t+1}. It follows that, when information is complete, condition (5) can be restated in recursive form as

$$c_t = -\sigma E_t[r_{t+1}] + E_t[c_{t+1}], \tag{8}$$

which is the familiar Euler condition. Similarly, condition (6) can be reduced to

$$\pi_t = mc_t + \beta E_t\left[\pi_{t+1}\right] + \mu_t, \tag{9}$$

which is the familiar NKPC.

We conclude that the "missing link" between the familiar representation of the New Keynesian model and the one offered in Proposition 1 is the common-knowledge assumption. Without this assumption, the Law of Iterated Expectations cannot be applied to the "average" agent in the economy *despite* the imposition of rational expectations and, as a result, the beauty contests seen in conditions (5) and (6) cannot be reduced to, respectively, the representative consumer's Euler condition and the NKPC.

PE, GE, and Higher-Order Beliefs. Inspecting condition (5) reveals the following basic point: holding constant the expectations of inflation and aggregate spending, a one-unit reduction in $\bar{E}_t[r_{T+1}]$ raises y_t by $\beta^{T-t}\sigma$. This effect can be interpreted as the PE effect of forward guidance.[19] Note that this effect decays with the horizon T, due to discounting. By contrast, as we explain later on, the overall equilibrium effect of forward guidance increases with T, due to the aforementioned GE effects.

[18] Note that the argument made here restricts the information to be common but allows it to be otherwise arbitrary.

[19] Strictly speaking, this is the PE effect plus the "within-period" GE effect, namely, the feedback effect between aggregate spending and the *contemporaneous* level of income. This effect, however, vanishes as the length of the time interval shrinks to zero, which justifies the interpretation of $\beta^{T-t}\sigma$ as the PE effect.

These points are well understood in the literature; see, for example, Del Negro, Giannoni and Patterson (2015) and McKay, Nakamura and Steinsson (2016*a*,*b*). But whereas these papers proceed to modify the preferences and/or the market structures, we shift the focus to the formation of expectations. In the sequel, we first recast the expectations of future income and future inflation, and the associated GE mechanisms, in terms of higher-order beliefs. We then use this window to shed new light on how economy responds to forward guidance or, more generally, to news about the future.

5 Dynamic Beauty Contests: Attenuation and Horizon Effect

In the previous section, we showed how we can represent each block of the New Keynesian model as a dynamic beauty contest. In this section, we study a more abstract class of beauty contests that nests the two blocks of our New Keynesian model along with other applications, such as the asset-pricing model of Singleton (1987), Allen, Morris and Shin (2006) and Nimark (2017). This permits us to sidestep the underlying micro-foundations and to illustrate in a transparent manner the two broader insights of our paper: (i) how the absence of common knowledge anchors forward-looking expectations and attenuates GE mechanisms; and (ii) how this attenuation increases with the horizon at which these mechanisms operate.[20]

The beauty contest. Time is discrete, indexed by $t \in \{0, 1, ...\}$, and there is a continuum of players, indexed by $i \in [0, 1]$. In each period t, each agent i chooses an action $a_{i,t} \in \mathbb{R}$. We denote the corresponding average action by $a_t = \int_0^1 a_{i,t} di$ and let $\Theta_t \in \mathbb{R}$ denote an exogenous fundamental that becomes known in period t. We specify the best response of player i in period t as follows:

$$a_{i,t} = \Theta_t + \gamma E_{i,t}\left[a_{i,t+1}\right] + \alpha E_{i,t}\left[a_{t+1}\right], \tag{10}$$

where $\alpha, \gamma > 0$ are fixed parameters. Note that a player's optimal action in any given period depends on her expectation of both her own and the aggregate action in the next period. The former effect is parameterized by γ, the latter by α.[21]

By iterating the above best response, aggregating across i, and letting $\bar{E}_t[\cdot]$ denote the average expectation in period t, we reach the following representation of the considered beauty contest:

$$a_t = \Theta_t + \gamma \left\{ \sum_{k=1}^{+\infty} \gamma^{k-1} \bar{E}_t\left[\Theta_{t+k}\right] \right\} + \alpha \left\{ \sum_{k=1}^{+\infty} \gamma^{k-1} \bar{E}_t\left[a_{t+k}\right] \right\} \qquad \forall t \geq 0 \tag{11}$$

This is the key equation we work with in the sequel. It relates the aggregate outcome in any given period to

[20] By sidestepping the micro-foundations, we do not only speak to a larger class of environments, but also abstract from the kind of auxiliary shocks that were necessary before in order to limit the endogenous aggregation of information through markets,. This rules out the possibility that agents confuse one kind of fundamental for another (say, the news about future monetary policy with news about the idiosyncratic components of wage and markups), isolates the role of higher-order uncertainty about the fundamental of interest, and clarifies why our contribution is entirely orthogonal to that of Lucas (1972).

[21] Implicit behind condition (10) is, of course, a quadratic payoff structure. For example, player i's payoff can be $\mathcal{U}_i = \sum_t \beta^t U(a_{i,t}, a_{i,t+1}; \Theta_t, a_t)$, where U is reverse engineered so that the player's best-response condition is given by (10).

the concurrent fundamental, the average forecasts of the future fundamentals, and the average forecasts of the future aggregate outcomes.

When all agents share the same information, so that $\bar{E}_t[\cdot] = E_t[\cdot]$ is the rational expectation of a representative agent, we can use the Law of Iterated Expectations to restate condition (11) as follows:

$$a_t = \Theta_t + \delta E_t[a_{t+1}] \qquad \forall t \geq 0. \tag{12}$$

where $\delta \equiv \gamma + \alpha$. The above can be thought of as an aggregate-level Euler condition and δ as the "effective" discount factor that governs how much current outcomes depend on expectations of future outcomes. This dependence combines two distinct effects: a direct or PE effect that is parameterized by γ and that regards what each agent expects *herself* to do in the future, and an indirect or GE effect that is parameterized by α and that regards what each agent expects the *others* to do. Yet, by looking at equation (12) alone, it is impossible to tell these two effects apart. This explains why it is important to "open up" this equation to (11), just as we did with the representative consumer's Euler condition and the NKPC in the previous section.

Interpretation. Condition (11) directly nests the supply block of our New Keynesian model: just interpret Θ_t and a_t as, respectively, the real marginal cost and inflation, and set $\gamma = \beta\theta$ and $\alpha = \beta(1-\theta)$. Similarly, to nest the demand block, interpret Θ_t as the real interest rate scaled by the elasticity of inter-temporal substitution, interpret a_t as the aggregate level of spending, and set $\gamma = \beta$ and $\alpha = 1 - \beta$.[22] Moving to a completely different application, consider the class of incomplete-information asset-pricing models studied in Singleton (1987), Allen, Morris and Shin (2006), Bacchetta and van Wincoop (2006), and Nimark (2017). These models feature a single asset ("stock market") and overlapping generations of differentially informed traders. The equilibrium asset price is shown to satisfy the following condition:

$$p_t = d_t - s_t + \beta \bar{E}_t[p_{t+1}],$$

where p_t, d_t, and s_t denote the period-t price, dividend, and supply, respectively, and $\bar{E}_t[\cdot]$ is the average expectation of the period-t traders. The above can be nested in condition (11) by letting $\Theta_t = d_t - s_t$, $a_t = p_t$, $\alpha = \beta$ and $\gamma = 0$.[23] This indicates, not only the likely broader applicability of the insights we develop in the rest of this section, but also that the aggregate outcome a_t can be a market-clearing price in a Walrasian setting rather than the average action of a set of players.

[22] A minor qualification is needed here. In our version of the New Keynesian model, the consumers do not know the *real* interest rate between today and tomorrow because they have to forecast tomorrow's inflation. It follows that we cannot simply let $\Theta_t = -\sigma r_{t+1}$. That said, we can nest the demand block in condition (11) *as is* if we assume that prices are completely rigid ($\theta = 1$), fix inflation at zero, and let $\Theta_t = -\sigma R_t$. For the more general case in which inflation is variable, we can either modify the model so that r_{t+1} is observed, or nest the existing version of the demand block in the variant of (11) that replaces Θ_t with $\bar{E}_t[\Theta_t]$. The results developed in the sequel are robust to such modifications.

[23] The restriction $\gamma = 0$ means that the aggregate outcome today depends only on the average expectation of the aggregate outcome tomorrow, as opposed to the entire path of the aggregate outcome in the future. Relative to the more general case we study here, this restriction decreases dramatically the dimensionality of the higher-order beliefs that the aforementioned works had to deal with. The most essential difference, however, between these works and ours is that they focus on the effects of learning whereas we focus on the effects of different horizons.

The question of interest. How does aggregate spending respond to news about future real interest rates? How does inflation respond to news about future real marginal costs? How do asset prices respond to news about future dividends? In the rest of this section, we seek to answer this kind of questions without taking a specific stand on how precise or credible the available news is, or how exactly it has to be modeled. We achieve this by focusing on how the aggregate outcome covaries with the concurrent forecasts of future fundamentals, and by abstracting from the exact source and magnitude of the variation in these forecasts.

More specifically, we fix an arbitrary $T \geq 2$ and isolate the role of first- and higher- uncertainty about the period-T fundamental, Θ_T, by shutting down the uncertainty about the fundamentals in any other period. Without any loss, we then let $\Theta_t = 0$ for all $t \neq T$ and treat Θ_T as the *only* random fundamental. We also anchor the beliefs of the aggregate outcome "at infinity" by imposing that $\lim_{k\to\infty} \gamma^k E_{i,t}\left[a_{t+k}\right] = 0$ with probability one. This is akin to ruling out infinite-horizon bubbles and can be justified either by letting the game end at *any* finite period $T' > T$, or by imposing that $\gamma < 1$ and that a_t is bounded. We finally state the question of interest as follows: how does a_0, the current outcome, covaries with $\bar{E}_0[\Theta_T]$, the current forecasts of the future fundamental?[24]

Complete vs. Incomplete Information. In what follows, we study how the answer to the aforementioned question varies as we move from the complete-information benchmark to the more realistic scenario in which the agents face uncertainty about one another's beliefs and actions. To fix language, and to rule out trivial cases in which the incompleteness of information relates only to variables that are immaterial for our purposes (e.g. sunspots), we introduce the following definitions.

Definition 1 *We say that information is complete, or that the agents have common knowledge of the available news, if $E_{i,t}\left[\Theta_T\right] = E_{j,t}\left[\Theta_T\right]$ with probability one for all (i, j, t) such that $i \neq j$ and $t \leq T - 1$. (And when the converse is true, we say that information is incomplete.)*

Definition 2 *We say that the agents are able to reach a consensus about the future trajectory of the economy if $E_{i,t}\left[a_\tau\right] = E_{j,t}\left[a_\tau\right]$ with probability one for all (i, j, t, τ) such that $i \neq j$ and $t < \tau \leq T$.*

Note that the first notion regards the beliefs the agents hold about the *exogenous* fundamentals, whereas the second notion regards the beliefs the agents form about the *endogenous* outcomes. While conceptual distinct, the two notions become tied together under the rational-expectations hypothesis.

Proposition 2 *Along any rational-expectations equilibrium, lack of consensus about the future outcomes is possible only when information is incomplete.*

This clarifies the modeling role, and our preferred interpretation, of the introduced friction. What is at stake here is not how much the agents know about the fundamentals, but rather the ability to coordinate

[24]Setting $\Theta_t = 0$ for all $t \neq T$ is equivalent to isolating the variation in hierarchy of beliefs about the period-T fundamental that is orthogonal to belief hierarchy about the fundamental in any other period. Our result can thus be read as an "orthogonalization" of the effects of different horizons. Without any loss, we also normalize the unconditional mean of Θ_T to be zero.

their beliefs and their responses to the exogenous impulses (the news). In line with this point, the lessons obtained in the sequel are robust to replacing Θ_t in condition (10) with $\theta_{i,t} = \Theta_t + v_{i,t}$, where $v_{i,t}$ is i.i.d. across agents, and letting each agent i know at $t = 0$ the *entire* path of $\theta_{i,t}$; that is, we could have eliminated the uncertainty the typical agent faces about the fundamentals that matter for her *own* decisions (e.g., the uncertainty a consumer faces about her own interest rates), and yet preserve her uncertainty about the beliefs and the decisions of *others* (e.g., the consumer's uncertainty about aggregate spending and inflation).

The Frictionless Benchmark. As a reference point for our subsequent analysis, we first answer the question of interest in the absence of the considered friction.

Lemma 1 *Suppose that information is complete. For all states of Nature,*

$$a_t = \phi^*_{T-t} \cdot E_t[\Theta_T] \quad \forall t < T, \tag{13}$$

where $\phi^*_T = \delta^T \equiv (\gamma + \alpha)^T$.

The scalar ϕ^*_T measures how much the aggregate outcome covaries with the average forecast of the fundamental T periods later. This scalar is easily computed by iterating on condition (12), which is the Euler condition of the complete-information economy. Indeed, since $\Theta_t = 0$ for all $t \neq T$, we have $a_T = \Theta_T$ and $a_t = \phi^*_{T-t} E_t[\Theta_T]$ for every $t \leq T - 1$, where $\{\phi^*_{T-t}\}_{t=0}^{T-1}$ solves the following recursion:

$$\phi^*_{T-t} = \delta \phi^*_{T-t-1} \quad \forall t \leq T - 1, \tag{14}$$

with "terminal" condition $\phi^*_0 = 1$. It follows that $\phi^*_T = \delta^T$, where, recall, $\delta \equiv \gamma + \alpha$ and where γ and α parameterize, respectively, PE and GE effects.

The exact interpretation of ϕ^*_T and its magnitude depend, of course, on the application under consideration. Consider, for example, the demand block of the New Keynesian model, in isolation of the supply block. In this context, ϕ^*_T measures the response of aggregate spending in period 0 to the concurrent expectation of the real interest rate in period T, holding constant the expectations of the real interest rate in all other periods. In the representative-agent version of the model, this object is the same regardless of T; this can readily be seen by iterating the Euler condition of the representative consumer. In the heterogeneous-agent variant studied by McKay, Nakamura and Steinsson (2016*a*), on the other hand, the corresponding object is decaying with T, due to liquidity constraints. Alternatively, consider the supply block of the model, or the NKPC. In this context, ϕ^*_T is given by β^T and measures the response of inflation to news about marginal costs in period T; this can been seen by iterating the NKPC. Finally, in Singleton's asset-pricing model, ϕ^*_T measures the response of the period-0 asset price to news about dividends in period T.

What is of interest to us, however, is not how ϕ^*_T varies as we move from one complete-information application to another, but rather how it compares to its incomplete-information counterpart. We address this question in the rest of the section. But lets us first note two key facts about ϕ^*_T. First, ϕ^*_T is invariant to the precision of the representative agent's information that is available at $t = 0$. Second, ϕ_T^* is the same regardless of whether the representative agent expects to receive additional information between $t = 0$ and $t = T$ or not.[25]

The Frictional Case. We henceforth focus on the scenario in which the agents are unable to reach a perfect consensus about the response of the economy to any given news about the future fundamentals. This raises the delicate question of whether and how the level of higher-order uncertainty evolves over time. For the time being, we bypass this complication by making the following simplifying assumption.

Assumption 1 *There is no learning between $t = 0$ and $t = T$, at which point Θ_T becomes commonly known.*

As noted a moment ago, learning is completely inconsequential in the complete-information benchmark. This is no more true when information is incomplete, for reasons discussed later. Nevertheless, Assumption 1 is a useful starting point for three reasons. First, it affords an otherwise general specification of the information structure and yields the sharpest version of our results. Second, it separates our results from prior work that studies the implications of learning (Woodford, 2003*a*, Allen, Morris and Shin, 2006, Bacchetta and van Wincoop, 2006). Last but not least, it can be relaxed without compromising our main lessons.

Let us now fill in the remaining details. Having ruled out learning, we consider an otherwise arbitrary specification of the information structure. We first let a random variable s encapsulate the realization of the fundamental Θ_T along with *any* other aggregate shock that influences the information and the beliefs of the agents; we refer to s as the underlying state of Nature. We next represent the information of agent i by a random variable ω_i, which is itself correlated with s. Along with the fact that there is no learning, this means that, for all $t \leq T - 1$, $E_{i,t}[\cdot] = E_{i,0}[\cdot] = \mathbb{E}[\cdot|\omega_i]$, where $\mathbb{E}[\cdot|\omega_i]$ is the rational expectation conditional on ω_i. We refer to s as the aggregate state of Nature and to ω_i as the signal of agent i. Conditional on s, this signal is an i.i.d. draw from a fixed distribution, whose p.d.f. is given by $\phi(\omega|s)$. This signal therefore encodes the information that agent i has, not only about the fundamental Θ_T, but also about the entire state of Nature and thereby about the information and the beliefs of other agents; in short, it is the Harsanyi type of the agent. We finally assume that a law of large number applies in the sense that $\phi(\omega|s)$ is also the the realized cross-sectional distribution of the signals in the population when the realized aggregate state is s.

Because the state s can be an arbitrary random variable, the considered specification allows for rich first- and higher-order uncertainty and nests a variety of examples that can be found in the literature. For instance, we may let the aggregate state be $s = (\Theta_T, u)$, where $\Theta_T \sim N(0, \sigma_\theta^2)$ and $u \sim N(0, \sigma_u^2)$ are independent of one another, and specify the signal as $\omega_i = (z, x_i)$, where $z = \Theta_T + u$, $x_i = \Theta_T + v_i$, and $v_i \sim N(0, \sigma_v^2)$ is independent of s and i.i.d. across i. This case nests the information structure assumed in Morris and Shin (2002): each agent receives two signals about Θ_T, a private one given by x_i, and a public one given by z. Alternatively, we may modify the aforementioned case by letting $\omega_i = (z_i, x_i)$, with x_i as

[25] Clearly, these properties hinge on the linear structure of the best responses. Nonetheless, they are useful for our purposes because they guarantee that *only* higher-order uncertainty affects the elasticity of outcomes to news about future fundamentals.

before and $z_i = z + \eta_i$, $\eta_i \sim N(0, \sigma_\eta^2)$. In this case, there is a public signal z, whose observation is, however, contaminated by idiosyncratic noise, perhaps due to rational inattention a la Sims (2003). As yet another example, we could let the aggregate state be $s = (\Theta_T, \sigma)$, where $\Theta_T \sim N(0, \sigma_\theta^2)$ and $\sigma \sim U[\underline{\sigma}, \bar{\sigma}]$, and specify the signal as $\omega_i = (x_i, \sigma_i)$, where $x_i = \Theta_T + v_i$, $v_i \sim N(0, \sigma_i^2)$, $\sigma_i \sim U[\sigma - \Delta, \sigma + \Delta]$, and $\Delta, \underline{\sigma}, \bar{\sigma}$ being known scalars such that $0 < \Delta < \underline{\sigma} < \bar{\sigma}$. In this case, some agents are better informed than others, and each agent is uncertain about how informed or uninformed the other agents are.

By adopting this level of generality, we seek, not only to clarify the robustness of our insights, but also to bypass the need of taking a specific stand on what the available signals are and how the higher-order uncertainty is generated. This also explains why most of our results in this section are formulated in terms of how outcomes depend on the belief hierarchy, as opposed to how they depend on a specific set of signals. A concrete example, however, will also be considered.

Higher-order beliefs. An instrumental step towards addressing the question of interest is to understand the role of higher-order beliefs. Since Θ_t is fixed at zero for all $t \neq T$ and Θ_T becomes commonly known at $t = T$, we have $a_t = 0$ for all $t \geq T+1$, $a_T = \Theta_T$, and, for all $t \leq T-1$,

$$a_t = \gamma^{T-t} \bar{E}_t [\Theta_T] + \alpha \sum_{k=1}^{T-t} \gamma^{k-1} \bar{E}_t [a_{t+k}] . \tag{15}$$

By Assumption 1 (no learning), $\bar{E}_t[\cdot] = \bar{E}_0[\cdot]$ for all $t \leq T-1$. Using this fact, and iterating on the above condition, we reach the following lemma, which represents the period-0 outcome as a linear function of the concurrent average first- and higher-order expectations of Θ_T.

Lemma 2 *For every (α, γ, T), there exist positive scalars $\{\chi_{h,T}\}_{h=1}^T$ such that, regardless of the information structure,*

$$a_0 = \sum_{h=1}^{T} \left\{ \chi_{h,T} \bar{E}_0^h [\Theta_T] \right\} \tag{16}$$

where $\bar{E}_0^h[\cdot]$ is defined recursively by $\bar{E}_0^1[\cdot] = \bar{E}_0[\cdot]$ and $\bar{E}_0^h = \bar{E}_0[\bar{E}_0^{h-1}[\cdot]]$ for every $h \geq 2$.

The weights $\chi_{h,T}$ can be constructed in a recursive manner, as functions of α, γ, h, and T alone; see the Appendix for details. Here, we focus on the interpretation and the implications of Lemma 2.

Applied to the demand block of the New Keynesian model, this lemma means that today's aggregate spending is determined by the hierarchy of beliefs about future real interest rates. Applied to the supply block, it means that today's inflation is determined by the hierarchy of beliefs about future real marginal costs. In prior work, this two kinds of higher-order beliefs are "swept under the carpet" because complete information lets higher-order beliefs collapse to first-order beliefs.

To see this point in the abstract context under consideration, note that when higher-order beliefs coincide with first-order beliefs condition (16) reduces to $a_0 = \left(\sum_{h=1}^{T} \chi_{h,T} \right) \cdot \bar{E}_0 [\Theta_T]$. Along with Lemma 1, this also means that the complete-information outcome satisfies the following restriction:

$$\phi_T^* = \sum_{h=1}^{T} \chi_{h,T}. \tag{17}$$

When, instead, information is incomplete, higher-order beliefs no more coincide with first-order expectations. To understand how a_0 covaries with $\bar{E}_0[\Theta_T]$, we must therefore understand both how higher-order beliefs covary with first-order beliefs and how the beliefs of different order load into a_0. We complete these tasks in the sequel—but first we state the ultimate lesson.

Theorem 1 *Suppose that information is incomplete. There exists a positive scalar ϕ_T, which depends on both T and the information structure, such that the following properties hold:*

(i) For all states of Nature,

$$a_0 = \phi_T \cdot \bar{E}_0[\Theta_T] + \epsilon, \tag{18}$$

where ϵ is either identically zero or is random but orthogonal to $\bar{E}_0[\Theta_T]$.

(ii) The ratio ϕ_T/ϕ_T^* is strictly less than 1.

(iii) The ratio ϕ_T/ϕ_T^* is strictly decreasing in T.

Part (i) identifies ϕ_T as the incomplete-information counterpart of ϕ_T^*. Part (ii) establishes that the absence of common knowledge reduces the extent to which the aggregate outcome covaries with the average forecast of the fundamental, regardless of how precise the latter is. We refer to this finding as the "attenuation effect." Part (iii) establishes that this kind of attenuation increases with T: the longer the horizon, the larger the reduction in the responsiveness of the aggregate outcome to the forecasts of the fundamental relative to the frictionless benchmark. We refer to this finding as the "horizon effect."

Let us now sketch the proof of the result. Part (i) is trivial: it follows directly from projecting a_0 on $\bar{E}_0[\Theta_T]$ and letting ϕ_T be the coefficient of this projection and ϵ the residual. The latter captures any variation in higher-order beliefs that is orthogonal to the variation in first-order beliefs, such as the kind of "sentiment shocks" studied in Angeletos and La'O (2013).

To prove parts (ii) and (iii), note first that Lemma 2 implies that the following condition holds regardless of the information structure:

$$\phi_T = \sum_{h=1}^{T} \chi_{h,T}\beta_h, \tag{19}$$

where $\{\chi_{h,T}\}$ are the same scalars as those appearing in Lemma 2 and

$$\beta_h \equiv \frac{Cov\left(\bar{E}_0^h[\Theta_T], \bar{E}_0^1[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} \tag{20}$$

is the coefficient of the projection of $\bar{E}_0^h[\Theta_T]$ on $\bar{E}_0^1[\Theta_T]$, for any $h \geq 1$. When information is complete, $\bar{E}_0^h[\Theta_T]$ coincides with $\bar{E}_0^1[\Theta_T]$ for all h and all realizations of uncertainty, which in turn means that β^h is identically 1 for all h. By contrast:

Proposition 3 *With incomplete information, β_h is bounded in $(0,1)$ for all $h \geq 2$ and is decreasing in h.*[26]

In words, when information is incomplete, higher-order beliefs co-move less with first-order beliefs than lower-order beliefs. This result is easy to establish for the specific information structure studied in Morris and Shin (2002), but requires more work for the more general structure allowed here. It has a similar flavor as the result found in Samet (1998), which essentially states that higher-order beliefs are "anchored" by the common prior, but is distinct from it, because that result regards only the asymptotic properties of higher-order beliefs as $h \to \infty$ and contains no information on whether the covariance between first- and h-order forecasts varies monotonically with h.

Let us now go back to the proof of Theorem 1. Part (ii) now follows directly from the fact that $\beta^h \in (0,1)$, along with conditions (17) and (19). In a nutshell, ϕ_T is less than ϕ_T^* simply because higher-order beliefs move less than one-to-one with first-order beliefs when, and only when, information is incomplete. Importantly, this has nothing to do with how much the first-order beliefs themselves co-move with the fundamental.

Corollary 1 *Fix (α, γ, T). An incomplete-information economy features a lower response to expectations of future fundamentals (i.e., a lower ϕ_T) than a complete-information economy, even if the agents have more precise information in the former than in the latter. In this sense, the standard modeling practice "overstates" the aforementioned response.*

What remains to prove is part (iii), namely, the property that the attenuation of ϕ_T relative to ϕ_T^* increases with the horizon. This follows from combining our earlier result that the covariation between first- and higher-order beliefs decreases with the belief order h (Proposition 3) with another result, which we establish next and which sheds light on how the horizon T affects the relative importance of higher-order beliefs in the period-0 outcome.

Theorem 2 *Fix (α, γ). For any (h, T) such that $1 \leq h \leq T$, let $s_{h,T}$ be the total weight on beliefs of order up to, and including, h; that is, $s_{h,T} \equiv \sum_{r=1}^{h} \chi_{r,T}$, where the χ's are the same coefficients as those appearing in condition (16). The ratio $s_{h,T}/s_{T,T}$, which measures the relative contribution of the first h orders of beliefs to the aggregate outcome, strictly decreases with the horizon T and converges to 0 as $T \to \infty$.*

This result is crucial (which is why it qualifies, at least in our eyes, to be called a "theorem"). Longer horizons increase the number of loops from future aggregate actions to current actions. But when one increases the number of loops, one is effectively walking down the hierarchy of beliefs: forecasting outcomes further and further into the future maps to forecasting the forecasts of others at higher and higher orders. This in turn explains why longer horizons increase the relative importance of higher-order beliefs. Part (iii) of Theorem 1 then follows from combining this finding with the fact that high-order beliefs themselves covary with less with first-order beliefs than what lower-beliefs.

[26]Obviously, for $h = 1$, we have $\beta_1 = 1$ regardless of the information structure.

To sum up: The "attenuation effect" ($\phi_T/\phi_T^* < 1$) follows merely from the fact that incomplete information dampens the covariation of higher-order beliefs with first-order beliefs ($\beta_h < 1$ for all h). The "horizon effect" (ϕ_T/ϕ_T^* decreases with T), on the other hand, follows for the fact that this dampening is larger the higher the order of beliefs (β_h decreases with h) together with the fact that longer horizons raise the relative importance of higher-order beliefs (Theorem 2).

The limit as $T \to \infty$. We now complement the preceding results by establishing that, as long as higher-order beliefs are sufficiently anchored in the sense made precise below, ϕ_T becomes *vanishingly* small relative to its complete-information counterpart as the horizon gets larger and larger.

Proposition 4 (Limit) *If* $\lim_{h\to\infty} Var\left(\bar{E}^h[\Theta_T]\right) = 0$, *then*

$$\lim_{T\to\infty} \frac{\phi_T}{\phi_T^*} = 0,$$

To understand this result, note first that, by Theorem 1, the ratio $\frac{\phi_T}{\phi_T^*}$ is strictly decreasing in T and bounded in $(0, 1)$. It follows that this ratio necessarily converges to a constant $\varphi \in [0, 1)$ as $T \to \infty$. In the appendix we show that $\varphi = \lim_{h\to\infty} \beta_h$, where β_h is defined as before. This fact is, essentially, a corollary of Theorem 2: in the limit as $T \to \infty$, only the infinite-order beliefs matter. Establishing that $\frac{\phi_T}{\phi_T^*}$ converges to zero as $T \to \infty$ is therefore equivalent to establishing that β_h converges to zero as $h \to \infty$. Clearly, a sufficient (in fact, also a necessary) condition for this to be the case is that $\lim_{h\to\infty} Var\left(\bar{E}^h[\Theta_T]\right) = 0$.

This condition need not hold for every information structure, but can be said to be "generic" in the following sense. Take, as a reference point, the familiar case in which Θ_T is Normal and each agent observes a noisy private signal and a noisy public signal, as in Morris and Shin (2002). In this case, $\lim_{h\to\infty} \beta^h$ is strictly positive and is pinned down by the precision of the public signal. But now let us perturb this economy by allowing each agent's observation of the public signal to be contaminated by idiosyncratic noise, perhaps due to rational inattention or some other cognitive limitation. Then, $\lim_{h\to\infty} \beta^h$ is necessarily 0, even if the aforementioned noise is *arbitrarily* small.[27] In short, it takes a tiny perturbation to break the common-knowledge nature of a public signal and to guarantee the limit result in Proposition 4 holds.

Of course, this limit result is not useful of quantitative purposes. It nevertheless offers a sharp illustration of how the predictions made on the basis of common knowledge are particularly fragile at long horizons. It therefore reinforces our take-home message. It also builds a bridge between our results and the uniqueness result in global games (Morris and Shin, 1998, 2003): the common thread is the disproportionate effect of the (very) high orders of belief and the resulting discontinuity in small perturbations of common knowledge. Finally, the result holds true regardless of whether ϕ_T^* itself converges or explodes, which in turn depends on whether $\delta \equiv \gamma + \alpha$ is lower or higher than 1. As it will become clear in the next section, the latter case, where ϕ_T^* explodes, is directly relevant in the context of a liquidity trap.

[27] This case is formalized by letting agent i observe $z_i = z + \eta_i$, where $z = \Theta_T + u$ is the underlying public signal and η_i is the contaminating idiosyncratic noise. Our statement is that $\lim_{h\to\infty} \beta_h = 0$ as long as $Var(\eta_i)$ is not exactly zero.

An example and an equivalence result. The preceding analysis was heavy because we sought generality. We now present an example that simplifies the analysis by imposing an "exponential" structure on the hierarchy of beliefs. This example also serves two additional purposes: it offers a concrete interpretation of the information structure along the lines of the "news" literature; and it helps recast the lack of common knowledge as a certain form of myopia, or discounting.

For this example, we let $\Theta_T = z + \eta$, where $z \sim N(0, \sigma_z^2)$ and $\eta \sim N(0, \sigma_\eta^2)$ are independent of one another. We interpret z as the component of Θ_T that is realized at $t = 0$ and η as a residual that is realized at $t = T$. Finally, we let the signal of agent i received at $t = 0$ be $\omega_i = z + v_i$, where $v_i \sim N(0, \sigma_v^2)$ is i.i.d across i and orthogonal to both z and η. We can then think of z as "news" about Θ_T and v_i as an idiosyncratic noise in the observation, or interpretation, of this news by agent i.[28]

Regardless of its interpretation, the key property of this example is that it imposes an exponential structure in the belief hierarchy. By this we mean that, for every $h \geq 2$,

$$\bar{E}_0^h[\Theta_T] = \lambda \bar{E}_0^{h-1}[\Theta_T],$$

where $\lambda \equiv \frac{\sigma_v^{-2}}{\sigma_v^{-2}+\sigma_z^{-2}} \in (0,1)$. This scalar therefore controls the speed with which the covariation between first and h-order beliefs decay with h. A lower λ captures a lower degree of common knowledge; the frictionless, complete-information, benchmark is nested in the limit as $\lambda \to 1$.

The following result can then be shown.

Proposition 5 (Discounting) *The elasticity ϕ_T in the incomplete-information economy described above is the same as the elasticity ϕ_T^* in a representative-agent economy in which, for all t $\leq T-2$, the Euler condition (12) holds with $\delta \equiv \gamma + \alpha$ replaced by $\delta' \equiv \gamma + \alpha\lambda$.*

This result illustrates that, under appropriate conditions, the object of interest can be calculated with the same ease as in the complete-information benchmark. It also offers a sharp formalization of the idea that removing common knowledge causes the economy to act as if the representative agent is myopic and discounts the future more heavily that what in the frictionless benchmark (the effective δ is reduced). The extent of this kind of myopia is inversely related to λ, the degree of common knowledge.

The logic behind this "as if" result is simple. In the considered example, all the average higher-order forecast are linear transformations of the average first-order forecast. This guarantees that, for all $t \in \{1, ..., T-1\}$, the aggregate outcome can be expressed as a multiple of the average first-order forecast: $a_t = \phi_{T-t}\bar{E}_0[\Theta_T]$, for some known scalar ϕ_{T-t}. It follows that the prior-period average forecast of that outcome equals the same multiple of the average *second-order* forecast: $\bar{E}_{t-1}[a_t] = \phi_{T-t}\bar{E}_0^2[\Theta_T]$. And since $\bar{E}_0^2[\Theta_T] = \lambda\bar{E}_0[\Theta_T]$, we conclude that $\bar{E}_{t-1}[a_t] = \lambda a_t$. It is therefore *as if* the average agent systematically underestimates the variation in the future aggregate outcome, which in turn explains the result in Proposition 5.

[28]This noise can be justified in multiple ways; see the discussion of rational inattention in this section, as well as the discussion of Assumption 3 in Section 6.

Clearly, the exact result relies on the particular information structure assumed above. The logic, however, applies more generally. As already mentioned, the key feature of the considered example is that λ controls the speed with which β_h, the covariation between first and h-order beliefs, decays with h. The kind of myopia, or discounting, reported above is merely a manifestation of the more general property that β_h decays with h. Theorem 1 and Proposition 5 are therefore mirror images of each other: saying that there is more attenuation at longer horizons is the same as saying that the agents are, effectively, myopic.

Allowing for learning. We now return to the role played by Assumption 1. Relaxing this assumption does not appear to invalidate the insights we have developed, but complicates the analysis and precludes us from establishing Theorem 1 for arbitrary information structures. This is due to the increased complexity of the kind of higher-order beliefs that emerge once information is changing between 0 and T.

To appreciate what we mean by this, note that the absence of learning (namely, $\bar{E}_t[\cdot] = \bar{E}_0[\cdot]$ for all $t < T$) guarantees that the following properties hold for $h \in \{2, ..., T\}$ and any $\{t_2, t_3, ...t_h\}$ such that $0 < t_2 < ... < t_h < T$:

$$\bar{E}_0[\bar{E}_{t_2}[\cdots[\bar{E}_{t_h}[\Theta_T]\cdots]] = \bar{E}_0^h[\cdot]$$

That is, Assumption 1 helps collapse the "cross-period" higher-order beliefs to the "within-period" higher-order beliefs. This is the key step for obtaining the convenient representation of $\bar{a}_0$ in Lemma 2 and, thereby, for applying Proposition 3. Without Assumption 1, we must instead express $\bar{a}_0$ as a function of all the aforementioned kind of "cross-period" higher-order beliefs, which greatly complicates the analysis. Note, in particular, that there are $T-1$ types of second-order beliefs (namely, $\bar{E}_0\left[\bar{E}_t[\cdot]\right]$ for all t such that $1 \leq t \leq T-1$), plus $\frac{(T-1)\times(T-2)}{2}$ types of third-order beliefs, plus $\frac{(T-1)\times(T-2)\times(T-3)}{6}$ types of fourth-order beliefs, and so on. What is more, the correlation structure between first- and higher-order beliefs is more intricate, reflecting the anticipation of learning. Intuitively, the first-order belief $\bar{E}_0[\Theta_{10}]$ can be *less* correlated with the second-order belief $\bar{E}_0\left[\bar{E}_1[\Theta_{10}]\right]$ than with the third-order belief $\bar{E}_0\left[\bar{E}_8\left[\bar{E}_9[\Theta_{10}]\right]\right]$ if there is little or no learning between periods 0 and 1 but a lot of learning between periods 1 and 8. While this possibility does not invalidate our lessons,[29] it precludes us from extending Theorem 1 to arbitrary forms of learning.

It is possible, however, to overcome this caveat for two leading forms of learning studied in the literature. In the one, we let the agents become gradually aware of Θ_T, as in Mankiw and Reis (2002) and Wiederholt (2015). In the other, we allow the agents to receive a new private signal about Θ_T in each period prior to T, as in Woodford (2003*a*), Nimark (2008), and Mackowiak and Wiederholt (2009). In these two cases, we can not only recover Theorem 1, but also obtain a variant of Proposition 5. See Appendix B for details.

Also note that Theorem 2, which establishes that longer horizons raise the relative importance of higher-order beliefs, follows directly from the best-response structure and is therefore independent of the information structure. This is suggestive of why our horizon effect may extend well beyond the aforementioned cases. In line with this idea, we are able to show that the result of Proposition 4, namely the property that ϕ_T becomes vanishingly small relative to ϕ_T^* as $T \to \infty$, extends to *arbitrary* forms of learning as long as the

[29] This possibility is indeed allowed in the cases studied in Appendix B, for which that theorem continues to apply.

higher-order uncertainty remains bounded way from zero, in the sense made precise in Appendix B.

On the basis of these findings, we conclude that our lessons regarding the interaction of horizons and lack of common knowledge are robust to the introduction of learning. Two additional lessons, however, emerge once we take into account learning.

The first is that, holding the period-0 information constant, ϕ_T is closer to ϕ_T^* in the scenario in which subsequent learning is allowed relative to the scenario in which such learning is ruled out. Intuitively, the anticipation of a lower (higher) friction in the future eases (exacerbates) the friction in the present. Note that this effect hinges on the agents being forward-looking and is therefore absent in *static* beauty contests, such as those studied in Morris and Shin (2002), Woodford (2003*a*), and Angeletos and La'O (2010).

The second lesson is that, as time passes and agents accumulate more information, higher-order beliefs converge to first-order beliefs, causing the anchoring of the expectations and the attenuation of GE effects to decay with the lag of time since the news has arrived.[30] Although this prediction is not a core theme of our paper, it is worth noting that it is consistent with the available evidence on the impulse responses of expectations to identified shocks (Coibion and Gorodnichenko, 2012; Vellekoop and Wiederholt, 2017). By contrast, this evidence seems incompatible with the theories developed in Gabaix (2016) and Farhi and Werning (2017): by design, these theories attenuate the response of expectations to shocks but do not allow this attenuation to decay with the lag since the shock has hit the economy.

Rational Inattention. We now briefly sketch how the considered friction can be recast as the product of rational inattention. We let Θ_T be Normally distributed and, to sharpen the exposition, we assume that Θ_T is realized at $t = 0$ (think of Θ_T as being itself the "news" about the future). The typical agent is nevertheless unable to observe Θ_T perfectly. Instead, for every t, the action $a_{i,t}$ must be measurable in $\omega_i^t \equiv \{\omega_{i,\tau}\}_{\tau \leq t}$, where $\omega_{i,\tau}$ is a noisy signal obtained in period τ. We next assume that the noise is independent across the agents.[31] This guarantees that all aggregate outcomes are functions of Θ_T and, therefore, we can reduce the rational-inattention problem faced by each agent to the choice of a sequence of signals about Θ_T. We finally let these signals be chosen optimally, that is, so as to maximize the agent's ex ante payoff, subject to the following constraint:

$$\mathcal{I}\left(\, \omega_{i,t}, \Theta_T \mid \omega_i^{t-1} \,\right) \leq \kappa^{RI}, \tag{21}$$

where $\mathcal{I}\left(\omega_{i,t}, \Theta_T | \omega_i^{t-1}\right)$ is the (entropy-based) information flow between the period-t signal and the underlying fundamental, conditional on the agent's past information, and $\kappa^{RI} > 0$ is an exogenous scalar.

The usual interpretation of constraint (21) is that it captures the agent's limited cognitive capacity in

[30]Formally, consider the setting with learning studied in Appendix B, pick any $\tau \in \{1, ..., T-2\}$, and look at the response of a_t at $t = \tau$ rather than at $t = 0$. It is easy to check that this response increases with the precision of the information that arrives between $t = 0$ and $t = \tau$. A variant of this property is at the core of Woodford (2003*a*), Bacchetta and van Wincoop (2006), Nimark (2008, 2017), and Angeletos and La'O (2010): these papers show how, following persistent shocks to the underlying fundamentals, there is initially a large "wedge" between first- and higher-order beliefs, but this wedge decays as time passes and learning occurs.

[31]Although this assumption is separate the information-flow constraint (21), it is standard in the literature (e.g., Woodford, 2003*a*, Mackowiak and Wiederholt, 2009, Luo et al., 2017) and seems appealing if one interprets the noise as the product of cognitive limitations. It is also broadly consistent with experimental evidence (e.g., Khaw, Stevens and Woodford, 2016).

tracking Θ_T. But since beliefs about Θ_T map, in equilibrium, to beliefs of future outcomes, one can also think of (21) as a constraint on the agent's ability to figure out the likely effects of the underlying variation in Θ_T. This echoes Tirole (2015), who interprets rational inattention in games as a form of "costly contemplation".

As long as the prior about Θ_T is Gaussian and the objective function is quadratic, which is the case here by assumption, the optimal signal is also Gaussian. Furthermore, the noise in the signal has to be independent across periods, or else the agent could economize on cognitive costs, that is, relax the constraint in (21). These arguments are standard; see, e.g., Mackowiak, Matejka and Wiederholt (2017). The case studied here is actually far simpler than the one studied in the literature, because the relevant fundamental (Θ_T) does not vary as time passes. We infer that the optimal signal at every $t \leq T-1$ is given by

$$\omega_{i,t} = \Theta_T + v_{i,t},$$

where the noise $v_{i,t}$ is orthogonal to both Θ_T and $\{v_{i,\tau}\}_{\tau<t}$. Letting τ_t denotes the precision (i.e., the reciprocal of the variance) of this noise, we have that the period-t information flow is given by

$$\mathcal{I}\left(\, \omega_{i,t}, \Theta_T \mid \omega_i^{t-1} \, \right) = \tfrac{1}{2} \log_2 \left(1 + \frac{\tau_t}{\varsigma_t} \right),$$

where ς_t denotes the precision of the agent's prior in the beginning of period t; the latter is defined recursively by $\varsigma_0 = \sigma_\theta^{-2}$ and $\varsigma_{t+1} = \varsigma_t + \tau_t$ for $t \in \{0, ..., T-1\}$. It follows that the information constraint (21) pins down the sequence $\{\tau_t\}_{t=0}^{T-1}$ as a function of δ and σ_θ^2 alone.

All in all, the setting described above is nested in one of the cases studied in Appendix B, for which Theorem 1 applies. It follows that the documented attenuation and horizon effects can be readily recast as the consequences of rational inattention. In this sense, the considered friction can also be thought as a form of "bounded rationality."

Higher-order beliefs, rational expectations, and uncertainty about the responses of others. Throughout, we have invited the analyst (the theorist, the econometrician, or the policy maker) to pay close attention to the structure of higher-order beliefs in forward-looking models. This, however, does *not* mean that the agents inside the model (the consumers or the firms) must also engage in higher-order reasoning. Following the tradition of Muth and Lucas, we instead think of the consumers and the firms in the real world as non-strategic agents, who treat all the macroeconomic outcomes as exogenous stochastic processes and use a "statistical model" when trying to forecast these variables. By requiring that this model is consistent with actual behavior, the rational expectations hypothesis imposes a joint restriction on exogenous impulses, expectations, and outcomes. From this perspective, the exercise conducted in this section, and throughout our paper, is to revisit these restrictions when the auxiliary assumption that the exogenous impulses are common knowledge is relaxed.

6 Revisiting Forward Guidance

The preceding section dealt with an abstract dynamic beauty contest. This helped deliver the broader insights of the paper in a sharp and transparent manner. We now return to the application of interest, the New Keynesian model, and study the implications of our insights for the effectiveness of forward guidance.

A prelude. The following lesson follows *directly* the previous section. Suppose either that the monetary authority commits on implementing a specific path for the *real* interest rate, or that prices are infinitely rigid ($\theta = 1$) so that inflation is identically zero and the real interest rate coincides with the nominal one. In either case, we can understand the response of the economy to news about future rates by inspecting the demand block alone. Under complete information, we know, by iterating the familiar Euler condition, that the slope of y_0 with respect to $-E_0[r_T]$ is given by σ, the elasticity of intertemporal substitution, regardless of T. By contrast, the results of the previous section imply that, as long as information is incomplete, the relevant slope is bounded from above by σ and is decreasing in T, In this sense, the absence of common knowledge reduces the control that monetary policy has on aggregate demand even if we take for granted her control over the path of the real interest rate. What is more, this reduction is larger the longer the horizon at which the monetary authority commits to change the real interest rate. The reason is that the absence of common knowledge anchors the expectations that consumers form about aggregate spending, thus also attenuating the income multiplier that runs inside the demand block.

The obvious limitation of this lesson is that it treats the path of the *real* interest rate as the exogenous impulse or, equivalently, that it shuts down inflation. While this is useful for understanding how monetary policy controls aggregate demand holding constant inflation expectations, it is inappropriate for understanding how forward guidance matters during a liquidity trap. In that context, it is essential to treat the real interest rate as endogenous and to capture the feedback loop between aggregate spending and inflation.[32]

As noted before, this feedback loop is a GE mechanism that runs *between* the two blocks of the model. This means that the preceding analysis, which effectively dealt with each block in isolation, is not directly applicable. To understand what the complication is and how it can be addressed, let us momentarily assume that information is complete among the firms, even though it is incomplete among the consumers. In this special case, the standard NKPC remains valid, implying that inflation can be expressed as the present value of future real marginal costs and, thereby, of future consumer spending. Using this fact to substitute away inflation from condition (5), we arrive at the following representation of the equilibrium.

Lemma 3 *When the firms have complete information, equilibrium income (also, consumption) satisfies the following condition at every t:*

$$y_t = -\sigma R_t - \sigma \sum_{k=1}^{\infty} \beta^k \bar{E}_t^c[R_{t+k}] + \sum_{k=1}^{\infty} (1 - \beta + k\sigma\kappa) \beta^{k-1} \bar{E}_t^c[y_{t+k}]. \tag{22}$$

[32] Accordingly, for the rest of the analysis, we let $\theta < 1$ (equivalently, $\kappa > 0$).

This can, once again, be understood as a beauty contest among the consumers. But unlike the one seen earlier in condition (5), the one obtained here subsumes the feedback loop between inflation and spending. This in turn explains why the expectations of income show up with different weights that in condition (5), as well as why these weights now depend on κ, the slope of the NKPC.

The beauty contest seen in Lemma 3 is not directly nested in the abstract analysis of the previous section. Yet, it is possible to extend all the lessons of Section 5 to the new kind of beauty contest. In particular, Lemma 2 and Theorem 2 continue to hold, even though the exact structure of the χ weights that appear in these results is now different.[33] Note next that Proposition 3, which regards the structure in higher-order beliefs, is of course invariant to the best-response structure. It follows that Theorem 1 and all the other results of Section 5 can be directly applied to the question of interest; all we have to do is to interpret the (negative of the) nominal interest rate *T* periods ahead as the fundamental Θ_T in the context of Section 5.

A concrete exercise. To make things more concrete, and to accommodate the possibility that the firms, too, have incomplete information, we henceforth specify the monetary policy.

Assumption 2 (Monetary Policy) *There exists a known $T \geq 2$ such that:*

(i) At any $t < T$, the nominal interest rate is pegged at zero, namely $R_t = -\rho$. (Remember R_t is the log-deviation from steady state.)

(ii) At any $t > T$, monetary policy replicates flexible-price outcomes.

(iii) The period-T nominal rate is such that

$$R_T = z + \eta, \tag{23}$$

where z and η are random variables, independent of one another and of any other shock in the economy, with $z \sim N(0, \sigma_z^2)$ and $\eta \sim N(0, \sigma_\eta^2)$. The former is realized at $t = 0$; the latter is realized at $t = T$ and is unpredictable prior to that point.

This assumption permits us to interpret $T - 1$ as the length of the liquidity trap and T as the horizon over which forward guidance operates. In particular, part (i) is directly motivated by the idea that ZLB constraint is binding during a liquidity trap. Part (ii), on the other hand, permits us to concentrate on expectations of R_T, the nominal interest rate that the monetary authority will implement in the first period after the ZLB has ceased to bind.[34]

Finally, part (iii) splits the randomness of R_T into two orthogonal components, along the lines of the example considered in the end of the last section. The first component, z, can be interpreted as the "anticipated component" of R_T, or as the "news" about future monetary policy; such news could be the product of a policy announcement. The second component, η, captures the residual uncertainty, or the "unanticipated

[33] This is proved in Appnedix C.

[34] Needless to say, we let the interest rate be moved for just one period, as opposed to many periods, only to simplify the exposition.

component", which is revealed at $t = T$. The variance of z relative to that of η can be interpreted as the precision of the available news, or as the credibility of the policy announcements.

We next specify the information structure as follows.

Assumption 3 (Information) *(i) At $t = 0$, each agent i, be it a consumer or a firm, observes a private signal of z, given by*

$$\omega_i = z + v_i,$$

where ϵ_i is an idiosyncratic noise, Normally distributed, with mean zero and variance $\sigma_c^2 \geq 0$ or $\sigma_f^2 \geq 0$, depending on whether the agent is, respectively, a consumer or a firm.

(ii) No other exogenous information arrives till $t = T$, at which point R_T becomes commonly knowledge.

(iii) The volatilities of the markup and idiosyncratic shocks are arbitrarily large relative to that of z.

Part (i) lets each agent's observation of z to be contaminated by idiosyncratic noise. At first glance, this assumption seems at odds with the fact that central bank communications enjoy ample coverage in the financial press; one may instead be tempted to treat forward guidance as a public signals. But it is one thing to say that something is "public news" in the real world and it is a different thing to assume that something is a "public signal." Doing the latter requires, not only that every agent herself is aware of and attentive to the news, but also that she is fully confident that every other agent is also aware of and attentive to the news, and so on. Clearly, that kind of common knowledge does not have an obvious counterpart in the real world. What is more, even if one insists on modeling the central bank communications as a public signal, we can recast the idiosyncratic noise as the product of rational idiosyncratic variation in the *interpretation* of such communications.[35]

To sum up, part (i) captures the friction we are interested in. Part (ii), on the other hand, shuts down any *exogenous* learning. Finally, part (iii) shuts down the *endogenous* learning that obtains through the observation of the realized market outcomes (wages, prices, etc) in every period. As in the previous section, these assumptions are not strictly needed, but facilitate a sharp characterization of the belief hierarchy.[36] In particular, similarly to the example studied in Section 5, the higher-order beliefs of, respectively, the

[35]Formally, this can be done by modeling the central bank communication itself as a public signal of z (or, equivalently, of R_T) and by letting each agent have private information about the noise in that signal. To see this, let the central bank's announcement be $w = R_T + \epsilon$, where ϵ is independent of R_T. Next, let each agent receive a private signal $x_i = \epsilon + \xi_i$, where ξ_i is i.i.d. across agents and independent of both ϵ and R_T. Then, the posterior of an agent about R_T conditional on both w and x_i is given by $E[R_T|w, x_i] = \psi_w w + \psi_x x_i = \psi_w R_T + (\psi_w + \psi_x)\epsilon + \psi_x \xi_i$, where $\psi_w > 0$ and $\psi_x < 0$ are pinned down by the volatilities of R_T, ϵ, and ξ_i. Note then that the observation of (w, x_i) is informationally equivalent to the observation of $\tilde{x}_i \equiv R_T + \tilde{\epsilon} + \tilde{\xi}_i$, where $\tilde{\epsilon} \equiv \left(1 + \frac{\psi_x}{\psi_w}\right)\epsilon$ and $\tilde{\xi}_i \equiv \frac{\psi_x}{\psi_w}\xi_i$. It is therefore *as if* the agent has observed a private signal, which happens to contain both an idiosyncratic and a correlated noise component. This kind of signal, which herein represents the idiosyncratic interpretation of agent i about the information contained in the central bank's communication, is directly nested in the analysis of Section 5. It follows that the policy lessons delivered herein can go through even if one insists on letting the central bank communications themselves be common knowledge.

[36]The last assumption can also be motivated on empirical grounds. In the US data, the estimated contribution of interest-rate shocks to the business-cycle volatility of real output, the price level, and inflation is quite small, whether in the context of SVARs or in the context of estimated DSGE models. It follows that, even the consumers and firms are attentive to macroeconomic statistics, the latter are very imprecise signals of the relevant fundamentals.

consumers and the firms satisfy, for all $h \geq 2$,

$$\bar{E}_0^{c,h}[R_T] = \lambda_c^{h-1} \cdot \bar{E}_0^c[R_T] \qquad \text{and} \qquad \bar{E}_0^{f,h}[R_T] = \lambda_f^{h-1} \cdot \bar{E}_0^f[R_T],$$

where

$$\lambda_c \equiv \frac{\sigma_c^{-2}}{\sigma_c^{-2} + \sigma_z^{-2}} \in (0,1] \qquad \text{and} \qquad \lambda_f \equiv \frac{\sigma_f^{-2}}{\sigma_f^{-2} + \sigma_z^{-2}} \in (0,1],$$

Note that λ_c and λ_f control how much higher-order beliefs co-move with first-order beliefs; they therefore parameterize the friction of interest. The frictionless, complete-information, benchmark is nested with $\lambda_c = \lambda_f = 1$, and a larger friction corresponds to lower values for the λ's.

We next characterize how the variation in expectations of R_T (triggered, here, by the variation in z) translates into variation in the equilibrium level of output. We start with the common-knowledge benchmark.

Proposition 6 (Forward Guidance with CK) *Suppose $\lambda_c = \lambda_f = 1$, which means that z is common knowledge among the consumers and the firms. There exists a scalar $\phi_T^* > \sigma$ such that, for all realizations of uncertainty,*

$$y_0 = y_0^{trap} - \phi_T^* E_0[R_T]. \tag{24}$$

where y_0^{trap} is the "liquidity trap" level of output (i.e., the one obtained when the period-T nominal interest rate is fixed at the steady-state value). Furthermore, ϕ_T^* is strictly increasing in T and $\phi_T^* \to \infty$ as $T \to \infty$.

This result contains the predictions of the textbook New Keynesian model regarding the power of forward guidance during a liquidity trap: this power, as measured by ϕ_T^*, is predicted to increase without bound as the time of action is pushed further and further into the future. It is this prediction, along with its quantitative evaluation, that constitutes the so-called "forward guidance puzzle"(Carlstrom, Fuerst and Paustian, 2012; Del Negro, Giannoni and Patterson, 2015; McKay, Nakamura and Steinsson, 2016*b*).

Contrast this prediction with the PE effect of future interest rates: as noted earlier, the PE effect is decreasing in T, simply because of the discounting that is embedded in intertemporal preferences. The reason that ϕ_T^* exhibits the opposite pattern is because of the GE effects that run within and between the two blocks of the model. In particular, the income multiplier implies that ϕ_T^* would stay constant with T even if we were to shut down the inflation response.[37] The feedback loop between aggregate spending and inflation then explains why ϕ_T^* actually increases with T.

Let us explain. Reducing the interest rate at $t = T$ increases spending and causes inflation at $t = T$. Because the nominal interest rate is pegged prior to T, this translates to a low *real* interest rate between $T-1$ and T. This stimulates demand at $T-1$, contributing to even higher inflation at $T-1$, which feeds to even higher demand at $T-2$, and so on. A longer horizon therefore maps to a larger the number of iterations in this feedback look and, thereby, to a stronger cumulative effect at $t = 0$. Finally, as $T \to \infty$, this feedback loop explodes, which explains why ϕ_T^* increases without bound.

[37] As noted before, this is readily seen by iterating on the Euler condition of the representative consumer.

It is worth noting that the scalar ϕ_T^* is invariant to the ratio σ_z/σ_η, which parameterizes the precision of the available news: varying the ratio σ_z/σ_η affects how much $E_0[R_T]$ varies with R_T, but does not affect how much either y_0 or π_0 vary with $E_0[R_T]$. This property extends to the incomplete-information scenario studied next and explains the sense in which both the puzzle and the resolution we offer are orthogonal to the question of how accurate or credible the news about the future interest rates might be.

Proposition 7 (Forward Guidance without CK) *Suppose that $\lambda_c < 1$ and/or that $\lambda_f < 1$, which means that at least one group of agents lacks common knowledge of the news. There exists a scalar ϕ_T such that, for all realizations of uncertainty,*

$$y_0 = y_0^{trap} - \phi_T \cdot \bar{E}_0^c[R_T]. \tag{25}$$

Furthermore, the following properties hold:

(i) ϕ_T is bounded between the PE effect and the complete-information counterpart: $\sigma\beta^T < \phi_T < \phi_T^*$.

(ii) ϕ_T is strictly increasing in both λ_c and λ_f; the ratio ϕ_T/ϕ_T^* is strictly decreasing in T and converges to 0 as $T \to \infty$; finally, when λ_c is sufficiently low, ϕ_T also converges to 0 as $T \to \infty$.

Part (i) formalizes the sense in which the standard model "maximizes" or "overstates" the power of forward guidance: ϕ_T^* is an upper bound to the prediction that the analyst can make if she maintains the rational-expectations hypothesis but removes common knowledge of the news. By varying the degree of common knowledge, we effectively span the range between this upper bound and the underlying PE effect.

Part (ii) is a version of our earlier "horizon effect": by attenuating the feedback loop between inflation and spending, as well as the GE effects that run within each block of the model, lack of common knowledge reduces the power of forward guidance by a factor that increases with T. This attenuation increases without bound in the sense that ϕ_T becomes vanishingly small relative to ϕ_T^* as $T \to \infty$, even if the friction is small (i.e., if λ_c and λ_f are arbitrarily close to 1). Finally, if the friction is large enough, the documented effect can be strong enough that ϕ_T is *decreasing* in T, not only relative to ϕ_T^*, but also in absolute value.

A numerical illustration. We now use a numerical example to illustrate our findings. We interpret the period length as a quarter and adopt the calibration of the textbook New Keynesian model found in Galí (2008). That is, we set the discount factor to 0.99, the Frisch elasticity to 1, the elasticity of intertemporal substitution to 1, and the price revision rate to $1/3$.

What remains is to pick the values of λ_c and λ_f, that is, the departure from common knowledge. The existing literature offers little guidance on how to make this choice. In want of a better alternative, we let $\lambda_c = \lambda_f = 0.75$ and interpret this as a situation in which every agent who has heard the policy announcement believes that any other agent has failed to hear, or "trust", the announcement with a probability equal to 25%. This is arguably a modest "grain of doubt" in the minds of people about their ability to coordinate the

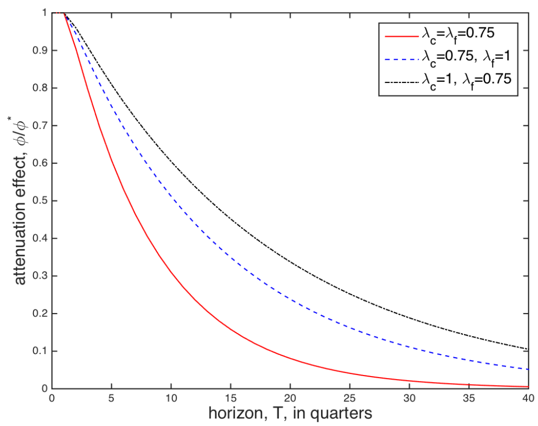

Figure 1: Attenuation effect, relative to common-knowledge benchmark, at different horizons.

adjustment in their beliefs and their behavior.[38,39]

The solid red line in Figure 1 plots the resulting attenuation effect, as measured by the ratio ϕ_T/ϕ_T^*, against the horizon length, T. By setting $\lambda_c = \lambda_f$, this line assumes that the consumers and the firms are subject to the same informational friction. The dashed blue line isolates the friction in the consumer side ($\lambda_c = .75$) by shutting it down in the firm side ($\lambda_f = 1$). The dotted black line does the converse.

The attenuation is strongest when the friction is present in both sides. Furthermore, the effect is quantitatively significant. For example, at a horizon of 5 years ($T = 20$), the power of forward guidance is only *one tenth* of its common-knowledge counterpart. This is on top of any mechanical effect that the noise may have on the size of the shift of the expectations of future interest rates: by construction, the documented attenuation effect is normalized by the size of the variation in the first-order beliefs of R_T.[40]

This illustrates the following broader point: by anchoring the movements in the expectations of economic outcomes *relative* to the movements in the expectations of the policy instrument, our paper helps operationalize the idea that policy makers may have harder time managing the former kind of expectations than the latter. This may help explain, for example, why forward guidance may trigger a large movement in the term structure without a commensurate movement in expectations of inflation and income.

In Appendix C, we elaborate on the mechanics behind Figure 1. We first show how the the effects of the considered friction can be understood through the lenses of a discounted Euler condition and a

[38]The proposed re-interpretation of the informational friction is exact: if we let the signal of an agent be binary, revealing the true z with probability λ and nothing with the residual probability, we obtain exactly the same characterization for ϕ_T as the one under the signal structure assumed above.

[39]Although Coibion and Gorodnichenko (2012) provide evidence in support of the kind of informational friction we have accommodated here, that evidence does not directly relate to forward guidance. Yet, taking that evidence at face value could justify an even larger departure from common knowledge: whenever a shock hits the economy, the average forecast error in the expectations of inflation appears to be half as large as the actual response in inflation, which translates to a value for λ close to 0.5.

[40]One can, of course, quibble about the appropriate parameterization. For instance, estimated DSGE models typically assume a higher degree of price stickiness than the one assumed here. This tends to reduce the attenuation effect. For instance, raising the value of θ from 2/3, the value in Galí (2008), to 0.85, the value in Christiano, Eichenbaum and Rebelo (2011), implies that the ratio ϕ_T/ϕ_T^* increases from about 0.1 to about 0.3 at $T = 20$. That said, note that menu-cost models calibrated to micro data indicate that the "right" value for θ is probably even smaller than 2/3. In any event, our goal here is only to illustrate; a more comprehensive quantitative evaluation is left for future work.

discounted NKPC; this is, in effect, an application of our earlier, more abstract, result in Proposition 5. This representation helps connect our paper to McKay, Nakamura and Steinsson (2016*b*,*a*) and Gabaix (2016). It also helps explains why almost all of the effect seen in Figure 1 comes from the attenuation of two GE mechanisms: the one that runs within the supply block of the model (the dependence of current inflation to future inflation); and the one that runs between the two blocks (the feedback loop between inflation and spending). By contrast, the attenuation of the income multiplier play a small role.

Intuitively, this is because, in the textbook version of the New Keynesian model, which we have employed here, consumers have infinite horizons and their spending depends on expectations of output *only* through the present value of permanent income. This means that varying the expectations of income in the next, say, 5 years has a minuscule effect on current spending. It follows that the attenuation of this particular GE mechanism (the income multiplier) is also quantitatively insignificant. Allowing for short horizons, precautionary motives, and feedback effects between housing prices and consumer spending is likely to increase the quantitative importance of this mechanism and, in so doing, also increase the relative importance of anchoring the consumers' expectations of income. This indicates the likely value of extending our analysis to more realistic versions of the demand block of the economy, such as those studied in McKay, Nakamura and Steinsson (2016*b*), Auclert (2017), and Kaplan, Moll and Violante (2016).

That said, note that the aforementioned papers are concerned *exclusively* with the demand block. By contrast, our paper revisits *both* blocks at once. This seems important given that (i) the NKPC is the key equation that distinguishes the New Keynesian model from its RBC counterparts; (ii) many challenges faced by the New Keynesian model are tied to the poor empirical performance of the NKPC; and (iii) our approach helps explain why inflation may move little in response to news about future demand and future marginal costs. Complementary in this regard are Nimark (2008) and Wiederholt (2015), which also emphasize how lack of common knowledge helps anchor inflation expectations.

7 Revisiting Fiscal Multipliers

We now introduce government spending in the model and study how aggregate income responds to "news" about government spending in the future. As usual, we treat the level of government spending as exogenous and assume that its is financed with lump-sum taxation. But unlike the standard practice, we remove common knowledge of the future level of government spending and of its macroeconomic effects.

The beauty-contest representation of the economy in conditions (5) and (6) still holds, with the following adjustment: the real marginal cost is now given by $mc_t = \Omega_c c_t + (1 - \Omega_c)\, g_t$, where g_t the level of government spending and $\Omega_c \in (0, 1)$ is a constant that increases with the steady-state fraction of GDP that is absorbed by private consumption. Note then that government spending enters the equilibrium dynamics of inflation and consumer spending only insofar as the real marginal cost increases with the level of output and, thereby, with the level of government spending.

Pick now any pair (T, T') such that $2 \leq T \leq T'$. We want T' to capture the length of the liquidity trap and T the period during which a fiscal stimulus is going to be enacted.[41] Accordingly, we let monetary policy satisfy Assumption 2, with T replaced by T'. We also fix $R_{T'} = 0$, so as to focus on fiscal policy. Finally, similar to Assumption 2, we assume that $g_t = 0$ for all $t \neq T$ and so as to focus on beliefs about the level of government spending at horizon T, g_T. Under an information structure similar to Assumption 3, we can then show the following.

Proposition 8 *Propositions 6 and 7 hold true with ϕ_T^* and ϕ_T reinterpreted as, respectively, the complete- and the incomplete-information slope of c_0 with respect to $\bar{E}_0[g_T]$.*

The part of this result that regards ϕ_T^* is the analogue of Proposition 6: under common knowledge, the current impact of a fiscal stimulus of a given size increases as the stimulus is pushed further into the future. This is a fiscal-policy variant of the forward guidance puzzle. The underlying logic is essentially the same: the PE effect of a fiscal stimulus, which is its direct effect on the concurrent real marginal costs, decays with the horizon T, but the GE effects that run within and between the two blocks of the model overturn this property. The other part of the result, which regards ϕ_T, is the analogue of Proposition 7: once we relax the assumption that there is common knowledge of the size and the consequences of the fiscal stimulus, the "fiscal multiplier" is reduced at every T, and the more so the larger T is.

In short, while the standard model predicts that fiscal stimuli must be back-loaded in order to "pile up" the feedback effects between aggregate spending and inflation, our approach offers a rationale for front-loading them. Such front-loading reduces the "coordination friction" in the economy: the sooner the fiscal stimulus is enacted, the less concerned the agents have to be about the beliefs and the responses of others at long horizons, and, thereby, the easier it is to coordinate on a large response today.[42,43]

8 Conclusion

Modern macroeconomics assigns a crucial role to forward-looking expectations, such as consumer expectations of future income and future real interest rates or firm expectations of future inflation and future real marginal costs. This property seems desirable and realistic. However, by assuming common knowledge along with rational expectations, the dominant modeling practice hardwires a certain kind of perfection in the ability of economic agents to understand what happens around them, to align their beliefs, and to coordinate their responses to any exogenous impulse. In so doing, it also maximizes the general-equilibrium

[41] Similarly to Christiano, Eichenbaum and Rebelo (2011), Woodford (2011), and Werning (2012), the fiscal multipliers we have been concerned with are those that obtain when the ZLB constraint binds. The investigation of how our insights extend when the economy is away from this constraint and monetary policy a Taylor rule or replicates flexible prices, is the subject of ongoing work.

[42] Of course, whether the provided rationale for front-loading is sufficiently strong to offset the standard back-loading property depends on the severity of the informational friction and all the parameters that govern the magnitude of the GE feedback loops, such as the value of κ. Letting κ be small enough guarantees that the incomplete-information fiscal multiplier decreases with T, not only relative to its complete-information counterpart, but also in absolute value.

[43] One can enrich the channels through which fiscal policy matters by adding financial frictions and "hand-to-mouth consumers" (Galí, López-Salido and Vallés, 2007; Kaplan, Moll and Violante, 2016). But even in these settings, fiscal multipliers are driven, in part, by the adjustment in the expectations of the unconstrained agents and the associated GE effects. We thus expect our insights to extend to these models as well, but leave for future work the exploration and quantitative evaluation of such an extension.

multipliers on fiscal and monetary policies.

Conversely, allowing for higher-order uncertainty helps accommodate a realistic friction in the ability of economic agents to forecast, or comprehend, the macroeconomic effects of policy news. This in turn arrests the underlying general-equilibrium feedback loops and reduces the ability of policy makers to steer the economy, especially when the policy news regard longer horizons.

We first formalized these ideas within a context of an abstract beauty contest game, which was flexible enough to nest either the demand or the supply block of the New Keynesian model, as well as other applications. We next showed how these ideas lessen the forward guidance puzzle and offer a rationale for front-loading fiscal stimuli.[44]

By anchoring the movements in the expectations of economic outcomes *relative* to the movements in the expectations of the policy instrument, our paper helps operationalize the idea that policy makers may have harder time managing the former kind of expectations than the latter. But it also hints to the following possibility: especially when it comes to longer horizons, it may be more important for the policy maker to communicate her intended path for the economy (e.g., for output and inflation) than her intended path for the policy instrument. What is more, these communications must be "loud and clear, not only in the ears of "Wall Street" (financial markets), but also in the ears of "Main Street" (firms and consumers). We leave the exploration of these ideas open for future research.

Another venue for future research is the investigation of the implications of our insights for the equilibrium-determinacy issues and the "neo-Fisherian" effects discussed in, inter alia, Cochrane (2016*a*,*b*) and Garcıa-Schmidt and Woodford (2015): these issues hinge on the same general-equilibrium feedback loops and higher-order beliefs as the one we have been concerned with in this paper.

A third possibility is the extension of our analysis to settings with financial frictions. Such frictions can give rise to feedback loops between asset prices and economic activity, such as those studied in Kiyotaki and Moore (1997) and Brunnermeier and Sannikov (2014). Under the lenses of our analysis, the extent to which the underlying shocks, or the news about future fundamentals, are common knowledge emerge as a crucial determinant of the equilibrium volatility in asset prices and economic activity.

Let us conclude by emphasizing once more the broader meaning of the friction we have sought to accommodate in this paper. Higher-order uncertainty can be the by-product either of "noise" in the aggregation of information through markets and other social institutions, or of "bounded rationality" in the form of rational inattention. Either way, higher-order uncertainty helps accommodate a friction in how well the agents can forecast, or comprehend, what *others* are doing and how aggregate outcomes may respond to exogenous impulses. We believe that this adds realism to macroeconomic theory, while allowing one to remain inside the "comfort zone" of rational expectations.

[44] As shown in Appendix C, the same logic helps reduce the paradox of flexibility.

Appendix A: Proofs

In this Appendix, we prove the results stated in the main text. For all the proofs that regard the New-Keynesian model (as opposed to the abstract analysis in Section 5), we use a tilde over a variable to denote the log-deviation of this variable from its steady-state counterpart, and reserve the non-tilde notation for the original variables.

Proof of Proposition 1. We proceed in four steps, starting with the behavior of the consumers, proceeding with the behavior of the firms, and concluding with market clearing and with the derivation of the two beauty contests shown in the main text.

Step 1: Consumers. Let $\pi_t = \frac{p_t}{p_{t-1}}$ denote the period-t gross inflation rate. Consider an arbitrary consumer $i \in \mathcal{I}_c$. As usual, the following intertemporal budget constraint holds in all periods and all states of Nature:[45]

$$\sum_{k=0}^{+\infty} \left\{ \Pi_{j=1}^{k} \left(\tfrac{R_{t+j-1}}{\pi_{t+j}} \right)^{-1} c_{i,t+k} \right\} \;=\; a_{i,t} + \sum_{k=0}^{+\infty} \left\{ \Pi_{j=1}^{k} \left(\tfrac{R_{t+j-1}}{\pi_{t+j}} \right)^{-1} \left(w_{i,t+k} n_{i,t+k} + e_{i,t+k} \right) \right\} \quad . \tag{26}$$

Taking the log-linear approximation of the above around the steady state,[46] we get the following:

$$\sum_{k=0}^{+\infty} \beta^k \tilde{c}_{i,t+k} = \tilde{a}_{i,t} + \sum_{k=0}^{+\infty} \beta^k \left\{ \Omega \left(\tilde{w}_{i,t+k} + \tilde{n}_{i,t+k} \right) + (1-\Omega)\, \tilde{e}_{i,t+k} \right\}, \tag{27}$$

where Ω is the ratio of labor income to total income in steady state. The consumer's optimality conditions, on the other hand, can be expressed as follows:

$$\tilde{n}_{i,t} = \frac{1}{\epsilon} \left(\tilde{w}_{i,t} - \tilde{c}_{i,t} \right), \tag{28}$$

$$\tilde{c}_{i,t} = E_{i,t} \left[\tilde{c}_{i,t+1} - \tilde{R}_t + \tilde{\pi}_{t+1} \right] = E_{i,t} \left[\tilde{c}_{i,t+1} - \tilde{r}_{t+1} \right], \tag{29}$$

where $E_{i,t}[\cdot]$ is the expectation of consumer i in period t. The first condition describes optimal labor supply; the second is the *individual-level* Euler condition, which describes optimal consumption and saving.

At this point, it is worth emphasizing that our analysis preserves the *standard* Euler condition at the individual level. This contrasts with McKay, Nakamura and Steinsson (2016*a*,*b*) and Werning (2015), where liquidity constraints cause this condition to be violated for *some* agents, as well as with Gabaix (2016), where a cognitive friction causes this condition to be violated for *every* agent. We revisit this point in Appendix C, when we show that our analysis rationalizes a *discounted* Euler condition at the aggregate level, *in spite of* the preservation of the standard condition at the individual level.

[45] One should think of the state of Nature as a realization of the exogenous payoff relevant shocks along with the cross-sectional distribution of the exogenous signals (information) received by the agents.

[46] In a symmetric steady state, $a_{i,t} = a^* = 0$. For this reason, we let $\tilde{a}_{i,t} \equiv \frac{a_{i,t}}{c^*}$, where c^* is steady-state consumption.

Combining conditions (27), (28) and (29), we obtain the optimal expenditure of consumer i in period t as a function of the current and the expected future values of wages, dividends, and real interest rates:

$$\begin{aligned}\tilde{c}_{i,t} = \tfrac{\epsilon(1-\beta)}{\epsilon+\Omega}\tilde{a}_{i,t} - \sum_{k=1}^{+\infty}\beta^k E_{i,t}\left[\tilde{r}_{t+k}\right] \qquad (30)\\ + (1-\beta)\left[\tfrac{(\epsilon+1)\Omega}{\epsilon+\Omega}\tilde{w}_{i,t} + \tfrac{\epsilon(1-\Omega)}{\epsilon+\Omega}\tilde{e}_{i,t}\right] + (1-\beta)\sum_{k=1}^{+\infty}\beta^k E_{i,t}\left[\tfrac{(\epsilon+1)\Omega}{\epsilon+\Omega}\tilde{w}_{i,t+k} + \tfrac{\epsilon(1-\Omega)}{\epsilon+\Omega}\tilde{e}_{i,t+k}\right].\end{aligned}$$

This condition, which is a variant of the consumption function seen in textbook treatments of the Permanent Income Hypothesis,[47] contains two elementary insights. First, all future variables—wages, dividends, and real interest rates—are discounted. Second, the *current* spending of a consumer depends on the present value of her income, which in turn depends, in equilibrium, on the *future* spending of other consumers.

The first property guarantees that the decision-theoretic, or partial-equilibrium, effect of forward guidance diminishes with the horizon at which interest rates are changed; the second represents a dynamic strategic complementarity, which is the modern reincarnation what was known as the "income multiplier" in the IS-LM framework. We elaborate on these two points more in the main text. For the time being, we aggregate condition (30), and use the facts that assets average to zero and that future idiosyncratic shocks are unpredictable, to obtain the following condition for aggregate spending:

$$\begin{aligned}\tilde{c}_t = -\sum_{k=1}^{+\infty}\beta^k \bar{E}^c_t\left[\tilde{r}_{t+k}\right] + (1-\beta)\left[\tfrac{(\epsilon+1)\Omega}{\epsilon+\Omega}\tilde{w}_t + \tfrac{\epsilon(1-\Omega)}{\epsilon+\Omega}\tilde{e}_t\right] \qquad (31)\\ + (1-\beta)\sum_{k=1}^{+\infty}\beta^k \bar{E}^c_t\left[\tfrac{(\epsilon+1)\Omega}{\epsilon+\Omega}\tilde{w}_{t+k} + \tfrac{\epsilon(1-\Omega)}{\epsilon+\Omega}\tilde{e}_{t+k}\right],\end{aligned}$$

where $\bar{E}^c_t\left[\cdot\right]$ henceforth denotes the *average* expectation of the consumers in period t.

Step 2: Firms. Consider a firm $j \in \mathcal{I}_f$ that gets the chance to reset its price during period t. The optimal reset price, denoted by p^{j*}_t, is given by the following:

$$\tilde{p}^{j*}_t = (1-\beta\theta)\left\{(\tilde{mc}^j_t + \tilde{p}_t) + \sum_{k=1}^{+\infty}(\beta\theta)^k E_{j,t}\left[\tilde{mc}^j_{t+k} + \tilde{p}_{t+k}\right]\right\} + (1-\beta\theta)\,\tilde{\mu}^j_t, \qquad (32)$$

where $E^f_{j,t}\left[\cdot\right]$ denotes the firm's expectations in period t, $\tilde{mc}^j_t = \tilde{w}^j_t$ is its *real* marginal cost in period t, and $\tilde{\mu}^j_t$ is the corresponding markup shock. The interpretation of this condition is familiar: the optimal "reset" price is given by the expected nominal marginal cost over the expected lifespan of the new price, plus the markup. Aggregating the above condition, using the fact that the past price level is known and that inflation

[47] To see this more clearly, suppose that initial assets are zero, that the real interest rate is expected to equal the discount rate at all periods, and that labor supply is fixed ($\epsilon \to \infty$). Condition (30) then reduces to $\tilde{c}_{i,t} = (1-\beta)\left[\Omega\tilde{w}_{i,t} + (1-\Omega)\,\tilde{e}_{i,t}\right] + (1-\beta)\sum_{k=1}^{+\infty}\beta^k E_{i,t}\left[\Omega\tilde{w}_{i,t+k} + (1-\Omega)\,\tilde{e}_{i,t+k}\right]$, which means that optimal consumption equals "permanent income" (the annuity value of current and future income). Relative to this benchmark, condition (30) adjusts for three factors: for the endogeneity of labor supply, which explains the different weights on wages and dividends; for initial assets, which explains the first term in condition (30); and for the potential gap between the real interest rate and the subjective discount rate, which explains the second term.

is given by $\tilde{\pi}_t = (1-\theta)(\tilde{p}_t^* - \tilde{p}_{t-1})$, where $\tilde{p}_t^* \equiv \int_{\mathcal{I}_f} \tilde{p}_t^{j*} dj$, we obtain the following condition for the level of inflation in period t:[48]

$$\tilde{\pi}_t = \tfrac{(1-\theta)(1-\beta\theta)}{\theta} \tilde{mc}_t + \tfrac{(1-\theta)(1-\beta\theta)}{\theta} \sum_{k=1}^{+\infty} (\beta\theta)^k \bar{E}_t^f [\tilde{mc}_{t+k}] + \tfrac{1-\theta}{\theta} \sum_{k=1}^{+\infty} (\beta\theta)^k \bar{E}_t^f [\tilde{\pi}_{t+k}] + \tilde{\mu}_t, \tag{33}$$

where $\bar{E}_t^f[\cdot]$ henceforth denotes the average expectation of the firms. The latter may or may not be the same as the average expectation of the consumers.

Step 3: Market Clearing, Wages, and Profits. Because the final-good sector is competitive and the technology satisfies (3) and (4), we have that $\tilde{p}_t = \int_{\mathcal{I}_f} \tilde{p}_t^j dj$ and $\tilde{y}_t = \int_{\mathcal{I}_f} \tilde{y}_t^j dj = \int_{\mathcal{I}_f} \tilde{l}_t^j dj$. The latter, together with market clearing in the labor market, gives $\tilde{y}_t = \tilde{n}_t \equiv \int_{\mathcal{I}_c} \tilde{n}_{i,t} di$. Market clearing in the market for the final good, on the other hand, gives

$$\tilde{y}_t = \tilde{c}_t \equiv \int_{\mathcal{I}_c} \tilde{c}_{i,t} di.$$

Finally, note that the real profit of monopolist j at period t is given by $e_t^j = \left(\frac{p_t^j}{p_t} - w_t^j\right) y_t^j$. Log-linearizing and aggregating it gives $\tilde{e}_t = -\frac{\Omega}{1-\Omega}\tilde{w}_t + \tilde{y}_t$. Combining all these facts with (28), the optimality condition for labor supply, we arrive at the following characterization of the aggregate wages and the profits:

$$\tilde{w}_t = \tilde{mc}_t = (\epsilon+1)\,\tilde{y}_t, \qquad \tilde{e}_t = \left[1 - \tfrac{\Omega(\epsilon+1)}{1-\Omega}\right]\tilde{y}_t, \quad \text{and} \quad \tfrac{(\epsilon+1)\Omega}{\epsilon+\Omega}\tilde{w}_t + \tfrac{\epsilon(1-\Omega)}{\epsilon+\Omega}\tilde{e}_t = \tilde{y}_t. \tag{34}$$

Step 4: Beauty Contests. Condition (31), which follows merely from consumer optimality, pins down aggregate spending as a function of the average beliefs of wages, profits, interest rates, and inflation. As we impose REE, a consumer can infer that (34) holds, aggregate spending can then be expressed as a function of the consumers' average beliefs of interest rates, of inflation, and of aggregate spending itself. This is condition (5), the consumption beauty contest. Similarly, combining (33) and (34), we can express aggregate inflation as a function of the firms' average beliefs of aggregate spending and of inflation itself. This is condition (6), the inflation beauty contest.

Proof of Proposition 2. Because Θ_t is zero for all $t > T$, a_t is also zero for all $t > T$.[49] Using this fact along with the fact that Θ_t is zero also for $t < T$, and iterating on condition (11), we can obtain a_t for all $t < T$ as a linear function of the average first- and higher-order beliefs about Θ_T; see, e.g., Lemma 2 below for an explicit characterization in the case without learning. When information is complete, all agents share the same first-order beliefs about Θ_T with probability one, and this fact is itself common knowledge. It follows that higher-order beliefs collapse to first-order beliefs and, therefore, a_t becomes a linear function of $E_t[\Theta_T]$, the commonly shared expectation of Θ_T. Now take any $t < \tau \leq T$ and any pair of agents i, j. Complete information guarantees that $E_{i,t}[E_\tau[\Theta_T]] = E_t[\Theta_T] = E_{j,t}[E_\tau[\Theta_T]]$ with probability one. And

[48] To economize notation, when we transit from (32) to (33), we rescale the markup shock: the $\tilde{\mu}_t$ that appears in (33) equals $\frac{\kappa}{1+\epsilon}$ times the original $\tilde{\mu}_t$.

[49] As mentioned in main text, we assume $\lim_{k\to\infty} \gamma^k E_{i,t}[a_{t+k}] = 0$ and rule out "extrinsic bubbles."

since we already argued that, in equilibrium, a_τ is a known linear function of $E_\tau[\Theta_T]$, it is also the case $E_{i,t}[a_\tau] = E_{j,t}[a_\tau]$. That is, complete information (in the sense of Definition 1) rules out imperfect consensus (in the sense of Definition 2).

Proof of Lemma 1. Lemma 1 directly follows from the argument in main text.

Proof of Lemma 2. We prove the following stronger result: there exists positively-valued coefficients $\{\chi_{h,k}\}_{k\geq 1, 1\leq h\leq k}$, such that, for any $t \leq T-1$,

$$a_t = \sum_{h=1}^{T-t} \left\{ \chi_{h,T-t} \bar{E}_t^h \left[\Theta_T\right] \right\}, \tag{35}$$

where each $\chi_{h,k}$ is a function of (α, γ, h, k) and $\bar{E}_t^h[\cdot]$ is defined recursively by $\bar{E}_t^1[\cdot] = \bar{E}_t[\cdot]$ and $\bar{E}_t^h[\cdot] = \bar{E}_t\left[\bar{E}_t^{h-1}[\cdot]\right]$ for every $h \geq 2$. We now prove this claim by induction. First, consider $t = T-1$. From $a_T = \Theta_T$[50] and condition (15), we have $a_{T-1} = (\gamma+\alpha)\bar{E}_{T-1}[\Theta_T]$. It follows that condition (35) holds for

$$\chi_{1,1} = \gamma + \alpha.$$

Now, pick an arbitrary $t \leq T-2$, assume that condition (35) holds for all $\tau \in \{t+1, ..., T-1\}$, and let us prove that it also holds for t. From condition (15), we have

$$\begin{aligned} a_t &= \gamma^{T-t-1}(\gamma+\alpha)\bar{E}_t[\Theta_T] + \alpha \sum_{k=1}^{T-t-1} \gamma^{k-1} \bar{E}_t\left[\sum_{h=1}^{T-t-k} \left\{\chi_{h,T-t-k}\bar{E}_{t+k}^h[\Theta_T]\right\}\right] \\ &= \gamma^{T-t-1}(\gamma+\alpha)\bar{E}_t[\Theta_T] + \sum_{h=1}^{T-t-1}\sum_{k=1}^{T-t-h}\left(\alpha\gamma^{k-1}\chi_{h,T-t-k}\right)\bar{E}_t^{h+1}[\Theta_T] \end{aligned} \tag{36}$$

where the second line uses Assumption 1 (no learning). As a result, condition (35) holds for

$$\chi_{1,T-t} = \gamma^{T-t-1}(\gamma+\alpha) \quad \text{and} \quad \chi_{h+1,T-t} = \sum_{k=1}^{T-t-h} \alpha\gamma^{k-1}\chi_{h,T-t-k} \;\; h \in \{1, \cdots T-t-1\}. \tag{37}$$

This finishes the proof.

Proof of Theorem 1. This theorem builds on Proposition 3 and Theorem 2, which are proved in the sequel. We invite the reader to read first the proofs of these two results. Here, we prove Theorem 1 taking for granted these results.

Part (i) follows directly from projecting a_0 on $\bar{E}_0[\Theta_T]$ and letting ϕ_T be the coefficient of this projection and ϵ the residual.

[50]As mentioned in main text, we assumelim$_{k\to\infty}\gamma^k E_{i,t}[a_{t+k}] = 0$ and rule out "extrinsic bubbles." Together with the fact Θ_t is zero for all $t > T$, a_t is also zero for all $t > T$. As a result, $a_T = \Theta_T$ from condition (11).

To prove part (ii), note that from Lemma 2, we have $\phi_T = \sum_{h=1}^{T} \chi_{h,T}\beta_h$, which is condition (19) in the main text. Together with the expression of ϕ_T^*, condition (17), and the fact that $\beta_h < 1$ for all $h \geq 2$ (from Lemma 2), we have $\phi_T/\phi_T^* < 1$ for all $T \geq 2$.

To prove part (iii), from condition (19), we have $\phi_T/\phi_T^* = \left[\sum_{h=1}^{T-1} s_{h,T}\left(\beta_h - \beta_{h+1}\right) + s_{T,T}\beta_T\right]/s_{T,T}$ and $\phi_{T+1}/\phi_{T+1}^* = \left[\sum_{h=1}^{T-1} s_{h,T+1}\left(\beta_h - \beta_{h+1}\right) + s_{T,T+1}\left(\beta_T - \beta_{T+1}\right) + s_{T+1,T+1}\beta_{T+1}\right]/s_{T+1,T+1}$. From Lemma 2 we know $\beta_h > \beta_{h+1}$ for all h. Together with Theorem 2, we have, for all $T \geq 1$,

$$
\begin{aligned}
\phi_{T+1}/\phi_{T+1}^* &< \left[\sum_{h=1}^{T-1} s_{h,T+1}\left(\beta_h - \beta_{h+1}\right) + s_{T+1,T+1}\left(\beta_T - \beta_{T+1}\right) + s_{T+1,T+1}\beta_{T+1}\right]/s_{T+1,T+1} \\
&= \left[\sum_{h=1}^{T-1} s_{h,T+1}\left(\beta_h - \beta_{h+1}\right) + s_{T+1,T+1}\beta_T\right]/s_{T+1,T+1} \\
&\leq \left[\sum_{h=1}^{T-1} s_{h,T}\left(\beta_h - \beta_{h+1}\right) + s_{T,T}\beta_T\right]/s_{T,T} = \phi_T/\phi_T^*.
\end{aligned}
$$

Proof of Proposition 3. Note that every agent i's information set at period 0 is drawn i.i.d. from the aggregate state of the Nature (for simplicity, we call this property "symmetry" in the rest of this proof), we have, for all $h \geq 2$,

$$
\begin{aligned}
Cov\left(\bar{E}_0^h[\Theta_T], \bar{E}_0^1[\Theta_T]\right) &= Cov\left(E_{i,0}\left[\bar{E}_0^{h-1}[\Theta_T]\right], \bar{E}_0^1[\Theta_T]\right) = Cov\left(E_{i,0}\left[\bar{E}_0^{h-1}[\Theta_T]\right], E_{i,0}\left[\bar{E}_0^1[\Theta_T]\right]\right), \\
&= Cov\left(\bar{E}_0^{h-1}[\Theta_T], E_{i,0}\left[\bar{E}_0^1[\Theta_T]\right]\right) = Cov\left(\bar{E}_0^{h-1}[\Theta_T], \bar{E}_0^2[\Theta_T]\right),
\end{aligned}
$$

where the second and the third equality come from the law of iterated expectations. By the same argument, we have, for all $h \geq 2$ and $j \in \{1, 2, \cdots h-1\}$,

$$
Cov\left(\bar{E}_0^h[\Theta_T], \bar{E}_0^1[\Theta_T]\right) = Cov\left(\bar{E}_0^{h-j}[\Theta_T], \bar{E}_0^{1+j}[\Theta_T]\right). \tag{38}
$$

From the previous condition, for $k \geq 1$, we have

$$
\beta_{2k} = \frac{Cov\left(\bar{E}_0^k[\Theta_T], \bar{E}_0^{k+1}[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} = \frac{Cov\left(\bar{E}_0^k[\Theta_T], E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} = \frac{Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} \geq 0 \;\; \forall i, \tag{39}
$$

where the second equation follows from symmetry and the last equation follows from the law of iterated expectations. Similarly, for $k \geq 1$, we have

$$
\beta_{2k-1} = \frac{Cov\left(\bar{E}_0^k[\Theta_T], \bar{E}_0^k[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} = \frac{Var\left(\bar{E}_0^k[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} \geq 0. \tag{40}
$$

Now, note that for any random variable X, and any information set I, according to the law of total variance, we have:

$$
Var\left(\mathbb{E}[X|I]\right) \leq Var\left(X\right).
$$

As a result, $Var\left(\bar{E}_0^{k+1}[\Theta_T]\right) = Var\left(E\left[E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]|s\right]\right) \leq Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right)$ and $Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right) = Var\left(E\left[\bar{E}_0^k[\Theta_T]|\omega_i\right]\right) \leq Var\left(\bar{E}_0^k[\Theta_T]\right)$, where, as a reminder, s is the aggregate state of the Nature and ω_i is the information set of agent i. Together with conditions (39) and (40), we know that, for all $k \geq 1$, $\beta_{2k+1} \leq \beta_{2k} \leq \beta_{2k-1}$. This proves that, for all $h \geq 2$, $\beta_h \in [0,1]$ and is weakly decreasing in h.

Now we try to prove β_h is strictly decreasing in h. Note that from condition (39), $\beta_2 = \frac{Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} \leq \beta_1 = 1$. If $\beta_2 = \beta_1 = 1$, we have $Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right]\right) = Var\left(\bar{E}_0[\Theta_T]\right)$ for all i. This means that

$$\begin{aligned} Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right] - \bar{E}_0[\Theta_T]\right) &= Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right]\right) + Var\left(\bar{E}_0[\Theta_T]\right) - 2Cov\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right], \bar{E}_0[\Theta_T]\right) \\ &= Var\left(\bar{E}_0[\Theta_T]\right) - Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right]\right) = 0, \end{aligned} \tag{41}$$

where the second equality follows from the law of iterated expectations. As a result, for all i, $E_{i,0}\left[\bar{E}_0[\Theta_T]\right] = \bar{E}_0[\Theta_T]$ almost surely. We henceforth have that

$$\begin{aligned} Var\left(\bar{E}_0\left[\Theta_T\right]\right) &= Cov\left(\Theta_T, \bar{E}_0^2\left[\Theta_T\right]\right) = Cov\left(\Theta_T, E_{i,0}\left(\bar{E}_0\left[\Theta_T\right]\right)\right) \\ &= Cov\left(\Theta_T, \bar{E}_0\left[\Theta_T\right]\right) = Cov\left(\Theta_T, E_{i,0}\left[\Theta_T\right]\right) = Var\left(E_{i,0}\left[\Theta_T\right]\right), \end{aligned}$$

where the first equality follows a similar argument as condition (38), the second and fourth equalities follow from symmetry, the third equality follows from $E_{i,0}\left[\bar{E}_0[\Theta_T]\right] = \bar{E}_0[\Theta_T]$ almost surely, and the last equality follows from the law of iterated expectations. This means that

$$\begin{aligned} Var\left(E_{i,0}\left[\Theta_T\right] - \bar{E}_0\left[\Theta_T\right]\right) &= Var\left(E_{i,0}\left[\Theta_T\right]\right) + Var\left(\bar{E}_0\left[\Theta_T\right]\right) - Cov\left(E_{i,0}\left[\Theta_T\right], \bar{E}_0\left[\Theta_T\right]\right) \\ &= Var\left(E_{i,0}\left[\Theta_T\right]\right) - Var\left(\bar{E}_0\left[\Theta_T\right]\right) = 0, \end{aligned} \tag{42}$$

where the second equality follows from symmetry. As a result, $E_{i,0}\left[\Theta_T\right] = \bar{E}_0\left[\Theta_T\right]$ almost surely for all i, and $E_{i,0}\left[\Theta_T\right] = E_{j,0}\left[\Theta_T\right]$ almost surely for all i, j. This is contradictory to the definition of incomplete information. As a result, $\beta_2 < \beta_1 = 1$.

Now, suppose it is not the case that β_h is strictly decreasing in h. The there exists a smallest $h^* > 1$ such that $\beta_{h^*+1} = \beta_{h^*}$.

If $h^* = 2k$ for some $k \geq 1$. From conditions (39) and (40), we have $Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right) = Var\left(\bar{E}_0^{k+1}[\Theta_T]\right)$. Following a similar argument as condition (42), we have $E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right] = \bar{E}_0^{k+1}[\Theta_T]$ almost surely. We henceforth have

$$\begin{aligned} Cov\left(\bar{E}_0^k[\Theta_T], \bar{E}_0^{k+1}[\Theta_T]\right) &= Cov\left(E_{i,0}\left[\bar{E}_0^{k-1}[\Theta_T]\right], \bar{E}_0^{k+1}[\Theta_T]\right) = Cov\left(E_{i,0}\left[\bar{E}_0^{k-1}[\Theta_T]\right], E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right) \\ &= Cov\left(E_{i,0}\left[\bar{E}_0^{k-1}[\Theta_T]\right], \bar{E}_0^k[\Theta_T]\right) = Cov\left(\bar{E}_0^k[\Theta_T], \bar{E}_0^k[\Theta_T]\right), \end{aligned}$$

where the first and the last equalities follow from symmetry, the second equality follows from the fact that $E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right] = \bar{E}_0^{k+1}[\Theta_T]$ almost surely, and the third equality follows from the law of iterated expectations. This expression means $\beta_{h^*-1} = \beta_{2k-1} = \beta_{2k} = \beta_{h^*}$, which contradicts the fact that h^* is the smallest h such that $\beta_{h^*+1} = \beta_{h^*}$.

If $h^* = 2k-1$ for some $k \geq 2$, from conditions (39) and (40), we have $Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right) = Var\left(\bar{E}_0^k[\Theta_T]\right)$. Following a similar argument as condition (41) for all i, $E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right] = \bar{E}_0^k[\Theta_T]$ almost surely. We henceforth have

$$\begin{aligned}Var\left(\bar{E}_0^k[\Theta_T]\right) &= Cov\left(\bar{E}_0^{k-1}\left[\Theta_T\right], \bar{E}_0^{k+1}\left[\Theta_T\right]\right) = Cov\left(\bar{E}_0^{k-1}\left[\Theta_T\right], E_{i,0}\left(\bar{E}_0^{k}\left[\Theta_T\right]\right)\right) \\ &= Cov\left(\bar{E}_0^{k-1}\left[\Theta_T\right], \bar{E}_0^k[\Theta_T]\right) = Cov\left(\bar{E}_0^{k-1}\left[\Theta_T\right], E_{i,0}\left[\bar{E}_0^{k-1}\left[\Theta_T\right]\right]\right) = Var\left(E_{i,0}\left[\bar{E}_0^{k-1}[\Theta_T]\right]\right),\end{aligned}$$

where the first equality follows a similar argument as condition (38), the second and forth equalities follow from symmetry, the third equality follows from $E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right] = \bar{E}_0^k[\Theta_T]$ almost surely, and the last equality follows from the law of iterated expectations. This expression means that $\beta_{h^*-1} = \beta_{2k-2} = \beta_{2k-1} = \beta_{h^*}$, which contradicts the fact that h^* is the smallest h such that $\beta_{h^*+1} = \beta_{h^*}$.

As a result, β_h is strictly decreasing in h. This implies that $\beta_h < \beta_1 = 1,\ \forall h \geq 2$. It also means that $\beta_h > 0\ \forall h$. If not, there exists a h^* such that $\beta_{h^*} = 0$. From strict monotonicity, we then have $\beta_{h^*+1} < \beta_{h^*} = 0$, which contradicts $\beta_{h^*+1} \geq 0$. This finishes the proof of Proposition 3.

Proof of Corollary 1. Corollary 1 follows directly from part (ii) of Theorem 1.

Proof of Theorem 2 To simplify notation, we extend the definition of $s_{h,\tau} = \sum_{r=1}^{h} \chi_{r,\tau}$ for all $h > \tau$. In the case that $h > \tau$, from Lemma 2, we have $\chi_{h,\tau} = 0$. As a result, $s_{h,\tau} = s_{\tau,\tau}$ for all $h > \tau$. We also define $s_{0,\tau} = 0$ for all $\tau \geq 1$.
From condition (36) and the expression of $\chi_{h,\tau}$ in condition (37), we have

$$s_{h,\tau} = \gamma^{\tau-1}\left(\gamma+\alpha\right) + \sum_{l=1}^{\tau-1} \alpha\gamma^{l-1} s_{h-1,\tau-l} \quad \forall h \geq 1 \; and \; \tau \geq 1. \tag{43}$$

Now, for all $\tau \geq 1$, as $\chi_{h,\tau} = 0$ for $h > \tau$, we can use $d_\tau = s_{\tau,\tau}$ denote the combined effect of beliefs of all different orders. From condition (43), we have

$$d_\tau = \gamma^{\tau-1}\left(\gamma+\alpha\right) + \sum_{l=1}^{\tau-1} \alpha\gamma^{l-1} d_{\tau-l} \quad \forall \tau \geq 1, \tag{44}$$

where we use the fact that $s_{h,\tau} = s_{\tau,\tau}$ for all $h > \tau$. From condition (44), we can easily verify, by induction, that

$$d_\tau = \left(\gamma+\alpha\right)^\tau \quad \forall \tau \geq 1. \tag{45}$$

For any $h \geq 1$, we now prove that $s_{h,\tau}/s_{\tau,\tau} = s_{h,\tau}/d_\tau$, strictly decreases with $\tau \geq h$. Notice that from condition (43), we have, for all $\tau \geq h \geq 1$

$$
\begin{aligned}
s_{h,\tau+1} &= \gamma^{\tau}\left(\gamma+\alpha\right)+\alpha s_{h-1,\tau}+\sum_{l=1}^{\tau-1}\alpha\gamma^{l}s_{h-1,\tau-l} \\
&= \gamma s_{h,\tau}+\alpha s_{h-1,\tau} < \left(\gamma+\alpha\right)s_{h,\tau}.
\end{aligned}
\tag{46}
$$

Also note that from condition (45), we have $s_{\tau+1,\tau+1} = d_{\tau+1} = \left(\gamma+\alpha\right)d_{\tau} = \left(\gamma+\alpha\right)s_{\tau,\tau}$. Together, we have $s_{h,\tau+1}/s_{\tau+1,\tau+1} < s_{h,\tau}/s_{\tau,\tau}$ for all $\tau \geq h \geq 1$.

Finally, we prove that, for any $h \geq 1$, $s_{h,\tau}/s_{\tau,\tau} \to 0$ as $\tau \to +\infty$. Because $s_{1,\tau} = \gamma^{\tau-1}\left(\gamma+\alpha\right)$ and $s_{\tau,\tau} = \left(\gamma+\alpha\right)^{\tau}$, $\lim_{\tau\to\infty} s_{h,\tau}/s_{\tau,\tau} \to 0$ holds for $h=1$. Suppose there is some h such that $\lim_{\tau\to\infty} s_{h,\tau}/s_{\tau,\tau} \to 0$ does not hold, let $h^* > 1$ be the smallest of such h. As $s_{h^*,\tau}/s_{\tau,\tau}$ is strictly decreasing in τ, there exists $\Gamma > 0$ such that $\lim_{\tau\to\infty} s_{h^*,\tau}/s_{\tau,\tau} \to \Gamma$. From conditions (45) and (46), we have $\frac{s_{h^*,\tau+1}}{s_{\tau+1,\tau+1}} = \frac{\gamma}{\gamma+\alpha}\frac{s_{h^*,\tau}}{s_{\tau,\tau}} + \frac{\alpha}{\gamma+\alpha}\frac{s_{h^*-1,\tau}}{s_{\tau,\tau}}$. Let $\tau \to +\infty$, we have $\Gamma = \frac{\gamma}{\gamma+\alpha}\Gamma$. This cannot be true as $\alpha, \gamma > 0$. As a result, $\lim_{\tau\to\infty} s_{h,\tau}/s_{\tau,\tau} \to 0$ for all $h \geq 1$.

Proof of Proposition 4. By Theorem 1, the ratio $\frac{\phi_T}{\phi_T^*}$ is strictly decreasing in T and bounded in $(0,1)$. It follows that $\frac{\phi_T}{\phi_T^*}$ necessarily converges to some $\varphi \in [0,1)$ as $T \to \infty$. Similarly, by Proposition 3, β_h is strictly decreasing in T and bounded in $(0,1)$. It follows that β_h necessarily converges to some $\underline{\beta} \in [0,1)$ as $T \to \infty$.

We first prove $\varphi = \underline{\beta} \equiv \lim_{h\to\infty}\beta_h$. We note that for, any $\vartheta > 0$, there exists a h^*, such that $\left|\beta_h - \underline{\beta}\right| < \frac{\vartheta}{2}$ for all $h \geq h^*$. From Theorem 2, we can then find $T^* \in \mathbb{N}_+$ such that, for all $T \geq T^*$, $\frac{s_{h^*-1,T}}{s_{T,T}} \leq \frac{\vartheta}{2}$. Together with conditions (17) and (19), we have, for all $T \geq T^*$,

$$
\begin{aligned}
\left|\frac{\phi_T}{\phi_T^*} - \underline{\beta}\right| &= \left|\frac{\sum_{h=1}^{T}\chi_{h,T}\left(\beta_h - \underline{\beta}\right)}{s_{T,T}}\right| = \left|\frac{\sum_{h=1}^{h^*-1}\chi_{h,T}\left(\beta_h - \underline{\beta}\right) + \sum_{h=h^*}^{T}\chi_{h,T}\left(\beta_h - \underline{\beta}\right)}{s_{T,T}}\right| \\
&\leq \frac{\sum_{h=1}^{h^*-1}\chi_{h,T}}{s_{T,T}} + \frac{\sum_{h=h^*}^{T}\chi_{h,T}\frac{\vartheta}{2}}{s_{T,T}} \\
&\leq \frac{\vartheta}{2} + \frac{\vartheta}{2} = \vartheta,
\end{aligned}
$$

where the first inequality we use the fact that $\left|\beta_h - \underline{\beta}\right| \leq 1$ and the second inequality uses the fact that $\frac{\sum_{h=h^*}^{T}\chi_{h,T}}{s_{T,T}} \leq \frac{s_{T,T}}{s_{T,T}} = 1$. As a result, $\varphi \equiv lim_{T\to\infty}\frac{\phi_T}{\phi_T^*} = \underline{\beta}$.

Finally, from condition (40), we know $\beta_{2h-1} = \frac{Var\left(\bar{E}_0^h[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)}$. If $\lim_{h\to\infty} Var\left(\bar{E}^h[\Theta_T]\right) = 0$, we have $\lim_{h\to\infty}\beta_{2h-1} = 0$. As β_h is decreasing in h, we also have

$$
\underline{\beta} = \lim_{h\to\infty}\beta_h = 0. \tag{47}
$$

As a result, $lim_{T\to\infty}\frac{\phi_T}{\phi_T^*} = 0$.

Proof of Proposition 5. Under the assumed information structure, we have for any $h \in \{1, ..., T\}$ and $0 \leq t_1 < t_2 < \cdots < t_h < T$,

$$\bar{E}_{t_1}[\bar{E}_{t_2}[...[\bar{E}_{t_h}[\Theta_T]...]] = \lambda^h z. \tag{48}$$

Now we prove by induction that, for all $t \leq T-1$,

$$a_t = (\gamma + \alpha)\, \Pi_{\tau=t+1}^{T-1} (\gamma + \lambda\alpha)\, \bar{E}_t[\Theta_T]. \tag{49}$$

Since $\Theta_t = 0$ for all $t \neq T$, together with condition (15), we have $a_T = \Theta_T$ and $a_{T-1} = (\gamma + \alpha)\, \bar{E}_{T-1}[\Theta_T]$. As a result, condition (49) holds for $t = T-1$. Now, pick a $t \leq T-2$, assume that the claim holds for all $\tau \in \{t+1, ..., T-1\}$, and let us prove that it also holds for t. Using the claim for all $\tau \in \{t+1, ..., T-1\}$, condition (15), and condition (48), we have, for $t \leq T-2$,

$$a_t = \gamma^{T-t} \bar{E}_t [\Theta_T] + \alpha \bar{E}_t [a_{t+1}] + \alpha \sum_{k=2}^{T-t} \gamma^{k-1} \bar{E}_t [a_{t+k}];$$

$$\gamma \bar{E}_t [a_{t+1}] = \lambda \gamma^{T-t} \bar{E}_t [\Theta_T] + \lambda\alpha \sum_{k=2}^{T-t} \gamma^{k-1} \bar{E}_t [a_{t+k}].$$

As a result, we have $a_t = \left(\frac{\gamma}{\lambda} + \alpha\right) \bar{E}_t [a_{t+1}]$. Together with condition (49) for $t+1$, we have

$$a_t = (\gamma + \alpha)\, \Pi_{\tau=t+1}^{T-1} (\gamma + \lambda\alpha)\, \bar{E}_t[\Theta_T].$$

This proves condition (49) for all $t \leq T-1$. As a result, $\phi_T = (\gamma + \alpha)\, \Pi_{t=1}^{T-1} (\gamma + \lambda\alpha)$. This proves Proposition 5.

Proof of Lemma 3. As firms have complete information, the canonical NKPC in condition (9) holds. Substituting it into the consumption beauty contest, condition (5), and using the fact that future markup shocks are unpredictable, we have

$$\tilde{y}_t = -\sigma \tilde{R}_t - \sigma \sum_{k=1}^{\infty} \beta^k \bar{E}_t^c[\tilde{R}_{t+k}] + \sum_{k=1}^{\infty} (1 - \beta + k\sigma\kappa)\, \beta^{k-1} \bar{E}_t^c[\tilde{y}_{t+k}].$$

Proof of Proposition 6. Let $\left\{\tilde{y}_t^{trap}, \tilde{\pi}_t^{trap}\right\}_{t=0}^{T}$ denote the path of output and inflation that obtains when the nominal interest rate at all periods after, and including, T is fixed at its steady state value (meaning that $\tilde{R}_t = 0$ for all $t \geq T$); think of this as the "liquidity trap" path. From conditions (8) and (9), we have, for all $t \leq T-1$,

$$\tilde{y}_t - \tilde{y}_t^{trap} = \sigma E_t[\tilde{\pi}_{t+1} - \tilde{\pi}_{t+1}^{trap}] + E_t[\tilde{y}_{t+1} - \tilde{y}_{t+1}^{trap}], \tag{50}$$

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = \kappa \left(\tilde{y}_t - \tilde{y}_t^{trap}\right) + \beta E_t[\tilde{\pi}_{t+1} - \tilde{\pi}_{t+1}^{trap}]. \tag{51}$$

Now we will prove the following stronger result, which nests the representation in condition (24): there exists positive scalars $\{\phi^*_\tau, \varpi^*_\tau\}_{\tau\geq 0}$ such that, whenever Assumptions 2 hold and z is commonly known, the equilibrium spending and inflation at any $t \leq T$ are given by

$$\tilde{y}_t - \tilde{y}^{trap}_t = -\phi^*_{T-t} \cdot E_t[\tilde{R}_T], \tag{52}$$

$$\tilde{\pi}_t - \tilde{\pi}^{trap}_t = \kappa \left(\tilde{y}_t - \tilde{y}^{trap}_t\right) - \varpi^*_{T-t} \cdot E_t[\tilde{R}_T]. \tag{53}$$

We prove this result by induction, starting with $t = T$ and proceeding backwards. When $t = T$, under Assumption 2, we have $\tilde{y}_T - \tilde{y}^{trap}_T = -\sigma \tilde{R}_T$ and $\tilde{\pi}_T - \tilde{\pi}^{trap}_T = \kappa \left(\tilde{y}_T - \tilde{y}^{trap}_T\right)$. This verifies (52) and (53) for $t = T$, with

$$\phi^*_0 = \sigma \text{ and } \varpi^*_0 = 0. \tag{54}$$

Now suppose that the result holds for arbitrary $t \in \{1, ..., T\}$ and let's prove that it also holds for $t-1$. By the assumption that (52) and (53) hold at t along with the Law of Iterated Expectations, we have

$$E_{t-1}[\tilde{y}_t - \tilde{y}^{trap}_t] = -\phi^*_{T-t} \cdot E_{t-1}[\tilde{R}_T],$$

$$E_{t-1}[\tilde{\pi}_t - \tilde{\pi}^{trap}_t] = -\left(\kappa\phi^*_{T-t} + \varpi^*_{T-t}\right) \cdot E_{t-1}[\tilde{R}_T].$$

Using the above together with conditions (50) and (51) verifies that (52) and (53) hold also for $t-1$, with

$$\phi^*_{T-t+1} = (1 + \sigma\kappa)\,\phi^*_{T-t} + \sigma\varpi^*_{T-t}, \tag{55}$$

$$\varpi^*_{T-t+1} = \beta\kappa\phi^*_{T-t} + \beta\varpi^*_{T-t}. \tag{56}$$

This completes the proof and gives a recursive formula that can be used to compute ϕ^*_T.

Now we prove the Proposition. From conditions (55) and (56), we have that, for all $\tau \geq 0$,

$$\phi^*_{\tau+1} = (1 + \sigma\kappa)\,\phi^*_\tau + \sigma\varpi^*_\tau, \tag{57}$$

$$\varpi^*_{\tau+1} = \beta\kappa\phi^*_\tau + \beta\varpi^*_\tau. \tag{58}$$

Together with condition (54), we know, when $\kappa > 0$, $\varpi^*_\tau > 0, \ \forall \tau \geq 1$. Then, from equation (57), we have $\phi^*_\tau > \sigma, \ \forall\tau \geq 1$, and ϕ^*_τ is strictly increasing in τ. Moreover, when $\kappa > 0$, $1 + \sigma\kappa > 1$. From (57), we know ϕ^*_τ explodes to infinity as $\tau \to \infty$.

Finally, we prove a few more results useful for the rest of the paper. First, we prove a recursive relationship about $\{\phi^*_\tau\}_{\tau\geq 0}$.

$$\frac{\phi^*_{\tau+1}}{\phi^*_\tau} + \beta\frac{\phi^*_{\tau-1}}{\phi^*_\tau} = 1 + \beta + \sigma\kappa \quad \forall\tau \geq 1. \tag{59}$$

From condition (57), we have, for all $\tau \geq 1$,

$$\beta\phi^*_\tau = \beta\,(1 + \sigma\kappa)\,\phi^*_{\tau-1} + \sigma\beta\varpi^*_{\tau-1}.$$

Together with conditions (57) and (58), we arrive at condition (59).

Second, we prove that, when $\kappa > 0$,

$$\frac{\phi^*_\tau}{\phi^*_{\tau-1}} \text{ is strictly increasing in } \tau \geq 1. \tag{60}$$

From conditions (57) and (58), we have $\phi^*_1 = \sigma\left(1+\sigma\kappa\right)$ and $\phi^*_2 = \sigma\left(\left(1+\sigma\kappa\right)^2 + \sigma\kappa\beta\right)$. As a result, when $\kappa > 0$,

$$\frac{\phi^*_2}{\phi^*_1} = 1 + \sigma\kappa + \frac{\sigma\kappa\beta}{1+\sigma\kappa} > \frac{\phi^*_1}{\phi^*_0}.$$

Now we proceed by induction. Suppose that, for $\tau \geq 1$, we have $\frac{\phi^*_{\tau+1}}{\phi^*_\tau} > \frac{\phi^*_\tau}{\phi^*_{\tau-1}}$. From condition (59), we know

$$\frac{\phi^*_{\tau+2}}{\phi^*_{\tau+1}} + \beta\frac{\phi^*_\tau}{\phi^*_{\tau+1}} = 1 + \beta + \sigma\kappa \quad \forall \tau \geq 1.$$

Together with $\frac{\phi^*_{\tau+1}}{\phi^*_\tau} > \frac{\phi^*_\tau}{\phi^*_{\tau-1}}$, we have $\frac{\phi^*_{\tau+2}}{\phi^*_{\tau+1}} > \frac{\phi^*_{\tau+1}}{\phi^*_\tau}$. This proves (60).

Finally, from condition (59), we know $\frac{\phi^*_\tau}{\phi^*_{\tau-1}}$ is bounded above. Together with (60), $\frac{\phi^*_\tau}{\phi^*_{\tau-1}}$ must converge to $\Gamma^* > 0$, as $\tau \to \infty$. From condition (59) again, we know Γ^* satisfy

$$\Gamma^* + \beta\frac{1}{\Gamma^*} = 1 + \beta + \sigma\kappa. \tag{61}$$

Proof of Proposition 7. With $\left\{\tilde{y}^{trap}_t, \tilde{\pi}^{trap}_t\right\}_{t=0}^T$ defined as in the proof of Proposition 6, we can rewrite the two beauty contests as follows:

$$\tilde{y}_t - \tilde{y}^{trap}_t = -\sigma\beta^{T-t}\bar{E}^c_t[\tilde{R}_T] + \sum_{k=1}^{T-t}\sigma\beta^{k-1}\bar{E}^c_t\left[\tilde{\pi}_{t+k} - \tilde{\pi}^{trap}_{t+k}\right] + (1-\beta)\sum_{k=1}^{T-t}\beta^{k-1}\bar{E}^c_t\left[\tilde{y}_{t+k} - \tilde{y}^{trap}_{t+k}\right], \tag{62}$$

$$\tilde{\pi}_t - \tilde{\pi}^{trap}_t = \kappa\left(\tilde{y}_t - \tilde{y}^{trap}_t\right) + \kappa\sum_{k=1}^{T-t}(\beta\theta)^k\,\bar{E}^f_t\left[\tilde{y}_{t+k} - \tilde{y}^{trap}_{t+k}\right] + \tfrac{1-\theta}{\theta}\sum_{k=1}^{T-t}(\beta\theta)^k\,\bar{E}^f_t\left[\tilde{\pi}_{t+k} - \tilde{\pi}^{trap}_{t+k}\right]. \tag{63}$$

Consider the following claim, which nests the representation in condition (25): under Assumption 3, there exists functions $\phi, \varpi : (0,1] \times (0,1] \times \mathbb{N} \to \mathbb{R}_+$ such that, for any $t \leq T-1$,

$$\tilde{y}_t - \tilde{y}^{trap}_t = -\phi\left(\lambda_c, \lambda_f, T-t\right)\bar{E}^c_t[\tilde{R}_T], \tag{64}$$

$$\tilde{\pi}_t - \tilde{\pi}^{trap}_t = \kappa\left(\tilde{y}_t - \tilde{y}^{trap}_t\right) - \varpi\left(\lambda_c, \lambda_f, T-t\right)\bar{E}^f_t[\tilde{R}_T] \tag{65}$$

We now establish this claim by induction.

First, consider $t = T-1$. From conditions (62) and (63) along with the fact that it is common knowledge monetary policy replicates flexible-price allocations from $T+1$ and on, we have

$$\tilde{y}_T - \tilde{y}^{trap}_T = -\sigma\tilde{R}_T \qquad \text{and} \qquad \tilde{\pi}_T - \tilde{\pi}^{trap}_T = \kappa\left(\tilde{y}_T - \tilde{y}^{trap}_T\right),$$

and therefore

$$\tilde{y}_{T-1} - \tilde{y}^{trap}_{T-1} = -\sigma(1+\sigma\kappa)\bar{E}^{c}_{T-1}[\tilde{R}_T],$$

$$\tilde{\pi}_{T-1} - \tilde{\pi}^{trap}_{T-1} = \kappa\left(\tilde{y}_{T-1} - \tilde{y}^{trap}_{T-1}\right) - \sigma\kappa\beta\bar{E}^{f}_{T-1}[\tilde{R}_T].$$

It follows that the claim holds for $t = T-1$ with

$$\phi(\lambda_c, \lambda_f, 1) = \sigma\left(1+\sigma\kappa\right) \qquad \text{and} \qquad \varpi(\lambda_c, \lambda_f, 1) = \sigma\kappa\beta. \tag{66}$$

Now, pick an arbitrary $t \leq T-2$, assume that conditions (64) and (65) hold for all $\tau \in \{t+1, ..., T-1\}$, and let us prove that it also holds for t. Since the claim holds for $\tau \in \{t+1, ...T-1\}$, and since $\tilde{y}_T - \tilde{y}^{trap}_T = -\sigma\tilde{R}_T$ and $\tilde{\pi}_T - \tilde{\pi}^{trap}_T = -\kappa\sigma\tilde{R}_T + \tilde{\mu}_T$, from condition (62), we have

$$\begin{aligned}
\tilde{y}_t - \tilde{y}^{trap}_t &= -\sigma\beta^{T-t-1}\left(1+\sigma\kappa\right)\bar{E}^c_t[\tilde{R}_T] - \left(1-\beta+\sigma\kappa\right)\sum_{k=1}^{T-t-1}\beta^{k-1}\phi\left(\lambda_c, \lambda_f, T-t-k\right)\bar{E}^c_t[\bar{E}^c_{t+k}[\tilde{R}_T]] \\
&\qquad - \sigma\sum_{k=1}^{T-t-1}\beta^{k-1}\varpi\left(\lambda_c, \lambda_f, T-t-k\right)\bar{E}^c_t\left[\bar{E}^f_{t+k}[\tilde{R}_T]\right] \\
&= -\sigma\beta^{T-t-1}\left(1+\sigma\kappa\right)\bar{E}^c_t[\tilde{R}_T] \\
&\qquad - \sum_{k=1}^{T-t-1}\beta^{k-1}\left[\left(1-\beta+\sigma\kappa\right)\lambda_c\phi\left(\lambda_c, \lambda_f, T-t-k\right) + \sigma\lambda_f\varpi\left(\lambda_c, \lambda_f, T-t-k\right)\right]\bar{E}^c_t[\tilde{R}_T],
\end{aligned}$$

where we have used the fact that, under Assumption 3, for $1 \leq k \leq T-t-1$,

$$\bar{E}^c_t[\bar{E}^c_{t+k}[\tilde{R}_T]] = \lambda_c\bar{E}^c_t[\tilde{R}_T] \quad \text{and} \quad \bar{E}^c_t[\bar{E}^f_{t+k}[\tilde{R}_T]] = \lambda_f\bar{E}^c_t[\tilde{R}_T]. \tag{67}$$

This proves the part of the claim that regards output, condition (64), with

$$\phi\left(\lambda_c, \lambda_f, T-t\right) = \sigma\beta^{T-t-1}\left(1+\sigma\kappa\right) + \sum_{k=1}^{T-t-1}\beta^{k-1}\left[\left(1-\beta+\sigma\kappa\right)\lambda_c\phi\left(\lambda_c, \lambda_f, T-t-k\right) + \sigma\lambda_f\varpi\left(\lambda_c, \lambda_f, T-t-k\right)\right]. \tag{68}$$

Similarly, the inflation beauty contest in condition (63) gives

$$\begin{aligned}
\tilde{\pi}_t - \tilde{\pi}^{trap}_t &= \kappa\left(\tilde{y}_t - \tilde{y}^{trap}_t\right) + \tilde{\mu}_t - \sigma\tfrac{\kappa}{\theta}\left(\beta\theta\right)^{T-t}\bar{E}^f_t[\tilde{R}_T] - \left(\kappa + \kappa\tfrac{1-\theta}{\theta}\right)\sum_{k=1}^{T-t-1}\left(\beta\theta\right)^k\phi\left(\lambda_c, \lambda_f, T-t-k\right)\bar{E}^f_t[\bar{E}^c_{t+k}[\tilde{R}_T]] \\
&\qquad - \tfrac{1-\theta}{\theta}\sum_{k=1}^{T-t-1}\left(\beta\theta\right)^k\varpi\left(\lambda_c, \lambda_f, T-t-k\right)\bar{E}^f_t[\bar{E}^f_{t+k}[\tilde{R}_T]] \\
&= \kappa\left(\tilde{y}_t - \tilde{y}^{trap}_t\right) + \tilde{\mu}_t \\
&\quad - \left\{\sigma\tfrac{\kappa}{\theta}\left(\beta\theta\right)^{T-t} + \sum_{k=1}^{T-t-1}\left(\beta\theta\right)^k\left[\tfrac{\kappa\lambda_c}{\theta}\phi\left(\lambda_c, \lambda_f, T-t-k\right) + \tfrac{(1-\theta)\lambda_f}{\theta}\varpi\left(\lambda_c, \lambda_f, T-t-k\right)\right]\right\}\bar{E}^f_t[\tilde{R}_T],
\end{aligned}$$

where we have used the fact that, similarly to the consumers' case, for $1 \leq k \leq T-t-1$,

$$\bar{E}_t^f[\bar{E}_{t+k}^c[\tilde{R}_T]] = \lambda_c \bar{E}_t^f[\tilde{R}_T] \quad \text{and} \quad \bar{E}_t^f[\bar{E}_{t+k}^f[\tilde{R}_T]] = \lambda_f \bar{E}_t^f[\tilde{R}_T]. \tag{69}$$

This proves the part of the claim that regards inflation, condition (65) with

$$\varpi\left(\lambda_c, \lambda_f, T-t\right) = \sigma \frac{\kappa}{\theta}\left(\beta\theta\right)^{T-t} + \sum_{k=1}^{T-t-1} \left(\beta\theta\right)^k \left(\tfrac{\kappa\lambda_c}{\theta}\phi\left(\lambda_c, \lambda_f, T-t-k\right) + \tfrac{(1-\theta)\lambda_f}{\theta}\varpi\left(\lambda_c, \lambda_f, T-t-k\right)\right). \tag{70}$$

We finally provide a recursive formula for computing $\phi(\lambda_c, \lambda_f, T-t)$ and $\varpi\left(\lambda_c, \lambda_f, T-t\right)$, which will be useful later. From condition (68), we have, for $t \leq T-2$,

$$\begin{aligned} \phi\left(\lambda_c, \lambda_f, T-t\right) &= \beta\phi\left(\lambda_c, \lambda_f, T-t-1\right) + \left(1-\beta+\sigma\kappa\right)\lambda_c\phi\left(\lambda_c, \lambda_f, T-t-1\right) + \sigma\lambda_f\varpi\left(\lambda_c, \lambda_f, T-t-1\right) \\ &= \left(\beta + \left(1-\beta+\sigma\kappa\right)\lambda_c\right)\phi\left(\lambda_c, \lambda_f, T-t-1\right) + \sigma\lambda_f\varpi\left(\lambda_c, \lambda_f, T-t-1\right). \end{aligned} \tag{71}$$

Similarly, from condition (70), we have, for $t \leq T-2$,

$$\begin{aligned} \varpi\left(\lambda_c, \lambda_f, T-t\right) &= \beta\theta\varpi\left(\lambda_c, \lambda_f, T-t-1\right) + \beta\theta\left(\tfrac{\kappa\lambda_c}{\theta}\phi\left(\lambda_c, \lambda_f, T-t-1\right) + \tfrac{(1-\theta)\lambda_f}{\theta}\varpi\left(\lambda_c, \lambda_f, T-t-1\right)\right) \\ &= \kappa\beta\lambda_c\phi\left(\lambda_c, \lambda_f, T-t-1\right) + \beta\left[\theta + \left(1-\theta\right)\lambda_f\right]\varpi\left(\lambda_c, \lambda_f, T-t-1\right). \end{aligned} \tag{72}$$

From now on, to simplify notation, we use ϕ_τ and ϖ_τ as shortcuts for, respectively, $\phi\left(\lambda_c, \lambda_f, \tau\right)$ and $\varpi\left(\lambda_c, \lambda_f, \tau\right)$.

We first prove part (i) of Proposition 7. From condition (68), we know $\phi_\tau > \sigma\beta^\tau$. The fact that $\phi_\tau < \phi_\tau^*$ is a direct corollary from the monotonicity of ϕ_τ with respect to λ_c and λ_f, which will be proved shorty.

We then prove part (ii) of Proposition 7. When $\kappa > 0$, from conditions (66), (71) and (72), we know that $\phi_\tau, \varpi_\tau > 0$ for all $\tau \geq 1$.

We will first prove, for $\tau \geq 2$, $\phi_\tau = \phi\left(\lambda_c, \lambda_f, \tau\right)$ is strictly increasing in both λ_c and λ_f. We will proceed by induction on τ. For $\tau = 2$, from (66), (71) and (72), we have ϕ_2 and ϖ_2 is strictly increasing in both λ_c and λ_f. Suppose for $\tau \geq 2$, ϕ_τ, ϖ_τ is strictly increasing in both λ_c and λ_f. From conditions (71) and (72), we know $\phi_{\tau+1}$ and $\varpi_{\tau+1}$ are strictly increasing in both λ_c and λ_f, where we use the fact that $\phi_\tau, \varpi_\tau > 0$. This proves that, for $\tau \geq 2$, $\phi_\tau = \phi\left(\lambda_c, \lambda_f, \tau\right)$ is strictly increasing in both λ_c and λ_f. Because of the strict monotonicity, we have, for $\tau \geq 2$, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$, $\frac{\phi_\tau}{\phi_\tau^*} = \frac{\phi(\lambda_c, \lambda_f, \tau)}{\phi(1,1,\tau)} < 1$.

We now prove that, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$, the ratio $\frac{\phi_\tau}{\phi_\tau^*} = \frac{\phi(\lambda_c, \lambda_f, \tau)}{\phi_\tau^*}$ is strictly decreasing in $\tau \geq 1$. We start by noticing, from the proof of Proposition 6, we have, for $\tau \geq 3$,

$$\frac{\phi_\tau^*}{\phi_{\tau-1}^*} + \beta\frac{\phi_{\tau-2}^*}{\phi_{\tau-1}^*} = 1 + \beta + \sigma\kappa \quad . \tag{73}$$

Now we prove that ϕ_τ satisfies an inequality with a similar form as (73):

$$\frac{\phi_\tau}{\phi_{\tau-1}} + \beta\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \leq 1 + \beta + \sigma\kappa\lambda_c < 1 + \beta + \sigma\kappa \qquad \forall \tau \geq 3. \tag{74}$$

From condition (71), we have, for $\tau \geq 3$,

$$\phi_\tau = (\beta + (1-\beta)\lambda_c)\phi_{\tau-1} + \sigma\kappa\lambda_c\phi_{\tau-1} + \sigma\lambda_f\varpi_{\tau-1},$$

$$\frac{\beta}{\beta + (1-\beta)\lambda_c}\phi_{\tau-1} = \beta\phi_{\tau-2} + \frac{\sigma\beta\kappa\lambda_c}{\beta + (1-\beta)\lambda_c}\phi_{\tau-2} + \frac{\sigma\beta\lambda_f}{\beta + (1-\beta)\lambda_c}\varpi_{\tau-2}.$$

From the previous two conditions, we have, for $\tau \geq 3$,

$$\begin{aligned}\phi_\tau + \beta\phi_{\tau-2} = &(\beta + (1-\beta)\lambda_c)\phi_{\tau-1} + \sigma\kappa\lambda_c\phi_{\tau-1} + \sigma\lambda_f\varpi_{\tau-1} \\ &+ \frac{\beta}{\beta + (1-\beta)\lambda_c}\phi_{\tau-1} - \frac{\sigma\beta\kappa\lambda_c}{\beta + (1-\beta)\lambda_c}\phi_{\tau-2} - \frac{\sigma\beta\lambda_f}{\beta + (1-\beta)\lambda_c}\varpi_{\tau-2}.\end{aligned} \tag{75}$$

Note that, for $\tau \geq 3$ and $\lambda_c, \lambda_f \in (0,1]$, from condition (72), we have

$$\left[(\beta + (1-\beta)\lambda_c) + \sigma\kappa\lambda_c + \frac{\beta}{\beta + (1-\beta)\lambda_c}\right]\phi_{\tau-1} \leq (1 + \beta + \sigma\kappa\lambda_c)\phi_{\tau-1},$$

and

$$\begin{aligned}&\sigma\lambda_f\varpi_{\tau-1} - \frac{\sigma\beta\kappa\lambda_c}{\beta + (1-\beta)\lambda_c}\phi_{\tau-2} - \frac{\sigma\beta\lambda_f}{\beta + (1-\beta)\lambda_c}\varpi_{\tau-2} \\ =&\sigma\lambda_f\left(\kappa\beta\lambda_c\phi_{\tau-2} + \beta\left[\theta + (1-\theta)\lambda_f\right]\varpi_{\tau-2}\right) - \frac{\sigma\beta\kappa\lambda_c}{\beta + (1-\beta)\lambda_c}\phi_{\tau-2} - \frac{\sigma\beta\lambda_f}{\beta + (1-\beta)\lambda_c}\varpi_{\tau-2} \\ =&\sigma\kappa\beta\lambda_c\left(\lambda_f - \frac{1}{\beta + (1-\beta)\lambda_c}\right)\phi_{\tau-2} + \sigma\beta\lambda_f\left[\theta + (1-\theta)\lambda_f - \frac{1}{\beta + (1-\beta)\lambda_c}\right]\varpi_{\tau-2} \\ \leq&0.\end{aligned}$$

Together with condition (75), we arrive at condition (74).

Now we can prove that, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$, $\frac{\phi_\tau}{\phi_\tau^*}$ is strictly decreasing in τ. We already prove $\frac{\phi_2}{\phi_2^*} < 1 = \frac{\phi_1}{\phi_1^*}$. We proceed by induction on τ. If $\frac{\phi_\tau}{\phi_\tau^*} < \frac{\phi_{\tau-1}}{\phi_{\tau-1}^*}$ for $\tau \geq 2$, we have $\frac{\phi_{\tau-1}}{\phi_\tau} > \frac{\phi_{\tau-1}^*}{\phi_\tau^*}$. From (73) and (74), we have $\frac{\phi_{\tau+1}}{\phi_\tau} < \frac{\phi_{\tau+1}^*}{\phi_\tau^*}$ and thus $\frac{\phi_{\tau+1}}{\phi_{\tau+1}^*} < \frac{\phi_\tau}{\phi_\tau^*}$. This finishes the proof that $\frac{\phi_\tau}{\phi_\tau^*}$ is strictly decreasing in $\tau \geq 1$, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$.

Now we prove that, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$, $\frac{\phi_\tau}{\phi_\tau^*}$ converges to 0 as $\tau \to \infty$. Because $\frac{\phi_\tau}{\phi_\tau^*} > 0$ is strictly decreasing in $\tau \geq 1$, there exists $\Gamma \in [0,1)$ such that $\frac{\phi_\tau}{\phi_\tau^*} \to \Gamma$ as $\tau \to \infty$. We next prove by contradiction that $\Gamma = 0$.

Suppose first that $\lambda_c < 1$. If $\Gamma > 0$, we have $\frac{\phi_\tau}{\phi_\tau^*}\frac{\phi_{\tau-1}^*}{\phi_{\tau-1}} \to 1$ as $\tau \to \infty$. Because $\frac{\phi_\tau^*}{\phi_{\tau-1}^*} \to \Gamma^*$, we have $\frac{\phi_\tau}{\phi_{\tau-1}} \to \Gamma^*$ and $\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \to \frac{1}{\Gamma^*}$ as $\tau \to \infty$. From condition (61), we have $\frac{\phi_\tau}{\phi_{\tau-1}} + \beta\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \to 1 + \beta + \sigma\kappa$ as $\tau \to \infty$. However, this is inconsistent with (74) when $\lambda_c < 1$ and $\kappa > 0$. As a result, $\Gamma = 0$.

Suppose next that $\lambda_c = 1$ but $\lambda_f < 1$. We prove a stronger version of (74):

$$\frac{\phi_\tau}{\phi_{\tau-1}} + (1 + \sigma\kappa(1-\lambda_f))\,\beta\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \leq 1 + \beta + \sigma\kappa \quad \forall \tau \geq 3. \tag{76}$$

From conditions (71) and (72), we have, for $\tau \geq 3$,

$$\phi_\tau = (1+\sigma\kappa)\,\phi_{\tau-1} + \sigma\lambda_f\varpi_{\tau-1},$$

$$\beta\phi_{\tau-1} = \beta\phi_{\tau-2} + \beta\sigma\kappa\phi_{\tau-2} + \beta\sigma\lambda_f\varpi_{\tau-2},$$

$$\varpi_{\tau-1} = \kappa\beta\phi_{\tau-2} + \beta\left[\theta + (1-\theta)\,\lambda_f\right]\varpi_{\tau-2}.$$

As a result, for $\tau \geq 3$,

$$\begin{aligned}\phi_\tau + \beta\phi_{\tau-2} &= (1+\sigma\kappa+\beta)\,\phi_{\tau-1} + \sigma\lambda_f\varpi_{\tau-1} - \beta\sigma\kappa\phi_{\tau-2} - \beta\sigma\lambda_f\varpi_{\tau-2} \\ &\leq (1+\sigma\kappa+\beta)\,\phi_{\tau-1} + \sigma\,(\lambda_f - 1)\,\kappa\beta\phi_{\tau-2}.\end{aligned}$$

This proves (76).

Now, if $\Gamma > 0$, similarly, we have $\frac{\phi_\tau}{\phi_\tau^*}\frac{\phi_{\tau-1}^*}{\phi_{\tau-1}} \to 1$ as $\tau \to \infty$. Because $\frac{\phi_\tau^*}{\phi_{\tau-1}^*} \to \Gamma^*$, we have $\frac{\phi_\tau}{\phi_{\tau-1}} \to \Gamma^*$ and $\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \to \frac{1}{\Gamma^*}$ as $\tau \to \infty$. From condition (61), we have $\frac{\phi_\tau}{\phi_{\tau-1}} + (1+\sigma\kappa(1-\lambda_f))\,\beta\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \to 1 + \beta + \sigma\kappa + \sigma\kappa(1-\lambda_f)\,\beta\frac{1}{\Gamma^*}$ as $\tau \to \infty$. However, this is inconsistent with equation (76) when $\lambda_f < 1$. As a result, $\Gamma = 0$ when $\lambda_c = 1$, but $\lambda_f < 1$.

Finally, we prove that, when λ_c is sufficiently low, $\phi(\lambda_c, \lambda_f, \tau)$ converges to zero as $\tau \to \infty$. The eigenvalues of the dynamical system (ϕ_τ, ϖ_τ) based on conditions (71) and (72) are

$$m_1 = \frac{\beta + (1-\beta+\sigma\kappa)\,\lambda_c + \beta\left[(1-\theta)\,\lambda_f + \theta\right] - \sqrt{\left(\beta + (1-\beta+\sigma\kappa)\,\lambda_c - \beta\left[(1-\theta)\,\lambda_f+\theta\right]\right)^2 + 4\sigma\beta\lambda_f\lambda_c\kappa}}{2} > 0$$

$$m_2 = \frac{\beta + (1-\beta+\sigma\kappa)\,\lambda_c + \beta\left[(1-\theta)\,\lambda_f + \theta\right] + \sqrt{\left(\beta + (1-\beta+\sigma\kappa)\,\lambda_c - \beta\left[(1-\theta)\,\lambda_f+\theta\right]\right)^2 + 4\sigma\beta\lambda_f\lambda_c\kappa}}{2} > m_1.$$

Note that $\lim_{\lambda_c \to 0} m_2 = \beta < 1$. As a result, when λ_c is sufficiently low, both eigenvalues are below 1, which means that $\phi(\lambda_c, \lambda_f, \tau)$ converges to zero as $\tau \to \infty$.

Proof of Proposition 8. We use $\tilde{g}_t$ to denote the amount of government spending at period t. As mentioned in main text, the government spending $\tilde{g}_t$ is financed by lump sum tax at period t, $\tilde{t}_t = \tilde{g}_t$. Similar to the analysis for monetary policy, we assume $\tilde{g}_t$ becomes commonly known at period t and only allow higher-order uncertainty about future $\tilde{g}$.

Similar to the main text, now we start to work with log-linearized representation. Because the introduction of lump-sum tax, the individual budget constraint becomes

$$\sum_{k=0}^{+\infty}\beta^k\tilde{c}_{i,t+k}=\tilde{a}_{i,t}+\sum_{k=0}^{+\infty}\beta^k\left\{\Omega_1\left(\tilde{w}_{i,t+k}+\tilde{n}_{i,t+k}\right)+\Omega_2\tilde{e}_{i,t+k}-(\Omega_1+\Omega_2-1)\tilde{t}_t\right\},$$

where Ω_1 is the ratio of labor income to total income (net of tax) in steady state, Ω_2 is the ratio of dividend income to total income (net of tax) in steady state, and $(\Omega_1+\Omega_2-1)$ is the ratio of lump sum tax to total income (net of tax) in steady state. On the other hand, the individual optimal labor supply and Euler equation, conditions (28) and (29), still hold here. Together, this gives rise to the optimal expenditure of consumer $i\in\mathcal{I}_c$ in period t,

$$\begin{aligned}\tilde{c}_{i,t}=&\tfrac{\epsilon(1-\beta)}{\epsilon+\Omega_1}\tilde{a}_{i,t}-\sum_{k=1}^{+\infty}\beta^kE_{i,t}\left[\tilde{r}_{t+k}\right]+(1-\beta)\left[\tfrac{(\epsilon+1)\Omega_1}{\epsilon+\Omega_1}\tilde{w}_{i,t}+\tfrac{\epsilon\Omega_2}{\epsilon+\Omega_1}\tilde{e}_{i,t}-\frac{\epsilon(\Omega_1+\Omega_2-1)}{\epsilon+\Omega_1}\tilde{t}_t\right]\\&+(1-\beta)\sum_{k=1}^{+\infty}\beta^kE_{i,t}\left[\tfrac{(\epsilon+1)\Omega_1}{\epsilon+\Omega_1}\tilde{w}_{i,t+k}+\tfrac{\epsilon\Omega_2}{\epsilon+\Omega_1}\tilde{e}_{i,t+k}-\frac{\epsilon(\Omega_1+\Omega_2-1)}{\epsilon+\Omega_1}\tilde{t}_{t+k}\right].\end{aligned}\tag{77}$$

Using the fact that assets average to zero and that future idiosyncratic shocks are unpredictable, we obtain the following condition for aggregate spending:

$$\begin{aligned}\tilde{c}_t=&-\sum_{k=1}^{+\infty}\beta^k\bar{E}_t^c\left[\tilde{r}_{t+k}\right]+(1-\beta)\left[\tfrac{(\epsilon+1)\Omega_1}{\epsilon+\Omega_1}\tilde{w}_t+\tfrac{\epsilon\Omega_2}{\epsilon+\Omega_1}\tilde{e}_t-\frac{\epsilon(\Omega_1+\Omega_2-1)}{\epsilon+\Omega_1}\tilde{t}_t\right]\\&+(1-\beta)\sum_{k=1}^{+\infty}\beta^k\bar{E}_t^c\left[\tfrac{(\epsilon+1)\Omega_1}{\epsilon+\Omega_1}\tilde{w}_{t+k}+\tfrac{\epsilon\Omega_2}{\epsilon+\Omega_1}\tilde{e}_{t+k}-\frac{\epsilon(\Omega_1+\Omega_2-1)}{\epsilon+\Omega_1}\tilde{t}_{t+k}\right].\end{aligned}\tag{78}$$

The firm side, on the other hand, is essentially same as the case without government spending, as a result, condition (6) still holds, but the formula for marginal cost are different. In particular, from the production function (3) and the optimal labor supply condition (28), we have

$$\tilde{mc}_t=\tilde{w}_t=\epsilon\int_{\mathcal{I}_c}\tilde{n}_{i,t}di+\tilde{c}_t=\epsilon\tilde{y}_t+\tilde{c}_t=\left(\epsilon\Omega_3+1\right)\tilde{c}_t+\epsilon\left(1-\Omega_3\right)\tilde{g}_t,\tag{79}$$

where $\tilde{y}_t=\Omega_3\tilde{c}_t+(1-\Omega_3)\,\tilde{g}_t$, $\Omega_3=\frac{1}{\Omega_1+\Omega_2}$ is the steady state consumption to output ratio, and $1-\Omega_3=\frac{\Omega_1+\Omega_2-1}{\Omega_1+\Omega_2}$ is the steady state government spending to output ratio.[51] As a result, the inflation beauty contest in condition (6) can be written as

$$\tilde{\pi}_t=\kappa\left(\Omega_c\tilde{c}_t+(1-\Omega_c)\,\tilde{g}_t\right)+\kappa\sum_{k=1}^{+\infty}(\beta\theta)^k\,\bar{E}_t^f[\Omega_c\tilde{c}_{t+k}+(1-\Omega_c)\,\tilde{g}_{t+k}]+\tfrac{1-\theta}{\theta}\sum_{k=1}^{+\infty}(\beta\theta)^k\,\bar{E}_t^f[\tilde{\pi}_{t+k}]+\tilde{\mu}_t,\tag{80}$$

where $\Omega_c=\frac{\epsilon\Omega_3+1}{\epsilon+1}$.

Finally, note that the real profit of monopolist j at period t is given by $e_t^j=\left(\frac{p_t^j}{p_t}-w_t^j\right)y_t^j$. After log-

[51] In steady state, the ratio of government spending to consumption will be equal to ratio of lump sum tax to total income (net of tax), $\Omega_1+\Omega_2-1$. This explains the formula for Ω_3.

linearization, we have $\tilde{e}_t = -\frac{\frac{\Omega_1}{\Omega_1+\Omega_2}}{1-\frac{\Omega_1}{\Omega_1+\Omega_2}}\tilde{w}_t + \tilde{y}_t = -\frac{\Omega_1}{\Omega_2}\tilde{w}_t + \tilde{y}_t$.[52] Together with condition (79), we have, for all t,

$$\begin{aligned}
\tfrac{(\epsilon+1)\Omega_1}{\epsilon+\Omega_1}\tilde{w}_t + \tfrac{\epsilon\Omega_2}{\epsilon+\Omega_1}\tilde{e}_t - \frac{\epsilon(\Omega_1+\Omega_2-1)}{\epsilon+\Omega_1}\tilde{t}_t &= \tfrac{(\epsilon+1)\Omega_1}{\epsilon+\Omega_1}\tilde{w}_t + \tfrac{\epsilon\Omega_2}{\epsilon+\Omega_1}\left(-\frac{\Omega_1}{\Omega_2}\tilde{w}_t + \tilde{y}_t\right) - \frac{\epsilon(\Omega_1+\Omega_2-1)}{\epsilon+\Omega_1}\tilde{g}_t \\
&= \tfrac{\Omega_1}{\epsilon+\Omega_1}\tilde{w}_t + \tfrac{\epsilon\Omega_2}{\epsilon+\Omega_1}\tilde{y}_t - \frac{\epsilon(\Omega_1+\Omega_2-1)}{\epsilon+\Omega_1}\tilde{g}_t \\
&= \tfrac{\epsilon(\Omega_1+\Omega_2)}{\epsilon+\Omega_1}\tilde{y}_t + \tfrac{\Omega_1}{\epsilon+\Omega_1}\tilde{c}_t - \frac{\epsilon(\Omega_1+\Omega_2-1)}{\epsilon+\Omega_1}\tilde{g}_t \\
&= \tfrac{\epsilon(\Omega_1+\Omega_2)}{\epsilon+\Omega_1}\left[\frac{1}{\Omega_1+\Omega_2}\tilde{c}_t + \frac{\Omega_1+\Omega_2-1}{\Omega_1+\Omega_2}\tilde{g}_t\right] + \tfrac{\Omega_1}{\epsilon+\Omega_1}\tilde{c}_t - \frac{\epsilon(\Omega_1+\Omega_2-1)}{\epsilon+\Omega_1}\tilde{g}_t \\
&= \tilde{c}_t
\end{aligned}$$

Substitute it into condition (78), we have

$$\tilde{c}_t = -\sigma\sum_{k=1}^{+\infty}\beta^{k-1}\bar{E}_t^c[\tilde{r}_{t+k}] + \tfrac{1-\beta}{\beta}\left\{\sum_{k=1}^{+\infty}\beta^k\bar{E}_t^c\left[\tilde{c}_{t+k}\right]\right\}. \tag{81}$$

This is exactly the same form of the consumption beauty contest, as condition (5).

Now let us state Proposition 8 formally here. Similar to Assumption 2, we assume $\tilde{g}_T = z + \eta$, where z and η are random variables, independent of one another and of any other shock in the economy, with $z \sim N(0,\sigma_z^2)$ and $\eta \sim N(0,\sigma_\eta^2)$. The former is realized at $t = 0$, and could be interpreted as news about government spending; the latter is realized at $t = T$ and is unpredictable prior to that point.

First consider the complete information outcome. Suppose z is commonly known starting at $t = 0$, we can find a scalar $\phi_{g,T}^*$ such that $\tilde{c}_0 - \tilde{c}_0^{trap} = \phi_{g,T}^* E_0[\tilde{g}_T]$, where $\tilde{c}_0^{trap}$ denotes the "liquidity trap" level of consumption (i.e., the one obtained when it is common knowledge that $\tilde{g}_T = 0$.) We have, when $\kappa > 0$,

$$\phi_{g,T}^* > 0, \text{ is strictly increasing in } T, \text{ and diverges to infinity as } T \to \infty. \tag{82}$$

Now consider the case in which z is not common knowledge. Similar to Section 6, we consider the information structure specified in Assumption 3, in which let each agent receives a private signal about z at period 0. We can then find a scalar $\phi_{g,T}$ such that $\tilde{c}_0 = \phi_{g,T}^*\bar{E}_0^c[\tilde{g}_T]$. We have, as long as $\kappa > 0$ and

[52] This expression is equivalent to $\frac{\Omega_2}{\Omega_1+\Omega_2}\tilde{e}_t + \frac{\Omega_1}{\Omega_1+\Omega_2}(\tilde{w}_t + \tilde{y}_t) = \frac{\Omega_2}{\Omega_1+\Omega_2}\tilde{e}_t + \frac{\Omega_1}{\Omega_1+\Omega_2}\left(\tilde{w}_t + \int_{\mathcal{I}_f}\tilde{l}_t^j dj\right) = \tilde{y}_t$. The last equation is true because $\frac{\Omega_1}{\Omega_1+\Omega_2}$ is steady state labor income to total income ratio (before deducting tax) and $\frac{\Omega_2}{\Omega_1+\Omega_2}$ is steady state dividend income to total income ratio (before deducting tax).

information is incomplete, that is $\lambda_c < 1$,[53]

$$\begin{aligned}\phi_{g,T} \in \left(0, \phi^*_{g,T}\right), \text{ is strictly increasing in } \lambda_c \text{ and } \lambda_f;\\ \text{the ratio } \phi_{g,T}/\phi^*_{g,T} \text{ is strictly decreasing in } T \text{ and converges to 0 as } T \to \infty;\\ \text{finally, when } \lambda_c \text{ is sufficiently low, } \phi_{g,T} \text{ also converges to 0 as } T \to \infty.\end{aligned} \tag{83}$$

We start from the proof of condition (82). Similar to the proof Proposition 6, we can establish that there exists non-negative scalars $\left\{\phi^*_{g,\tau}, \varpi^*_{g,\tau}\right\}_{\tau \geq 0}$ such that, when z is commonly known, the equilibrium spending and inflation at any $t \leq T$ are given by

$$\tilde{c}_t - \tilde{c}^{trap}_t = \phi^*_{g,T-T} \cdot E_t[\tilde{g}_T], \tag{84}$$

$$\tilde{\pi}_t - \tilde{\pi}^{trap}_t = \kappa \left(\Omega_c \left(\tilde{c}_t - \tilde{c}^{trap}_t\right) + (1-\Omega_c)\,\tilde{g}_t\right) + \varpi^*_{g,T-t} \cdot E_t[\tilde{g}_T], \tag{85}$$

where $\tilde{c}^{trap}_t$ and $\tilde{\pi}^{trap}_t$ denotes the "liquidity trap" level of consumption and inflation (i.e., the one obtained when it is common knowledge that $\tilde{g}_T = 0$.) We will use a version of Euler condition and NKPC as conditions (8) and (9):

$$\tilde{c}_t \;=\; -\sigma \left\{\tilde{R}_t - E_t\left[\tilde{\pi}_{t+1}\right]\right\} + E_t\left[\tilde{c}_{t+1}\right] \tag{86}$$

$$\tilde{\pi}_t \;=\; \kappa\left(\Omega_c \tilde{c}_t + (1-\Omega_c)\,\tilde{g}_t\right) + \beta E_t\left[\tilde{\pi}_{t+1}\right] + \tilde{\mu}_t. \tag{87}$$

Using the above expressions, similar to the proof of Proposition 6, we can establish that $\phi^*_{g,0} = 0$, $\varpi^*_{g,0} = 0$, $\phi^*_{g,1} = \sigma\kappa\,(1-\Omega_c)$, $\varpi^*_{g,1} = \beta\kappa\,(1-\Omega_c)$, and for all $\tau \geq 1$,

$$\phi^*_{g,\tau+1} = (1 + \sigma\kappa\Omega_c)\,\phi^*_{g,\tau} + \sigma\varpi^*_{g,\tau}, \tag{88}$$

$$\varpi^*_{g,\tau+1} = \beta\kappa\Omega_c\phi^*_{g,\tau} + \beta\varpi^*_{g,\tau}. \tag{89}$$

From condition (88), we can see when $\kappa > 0$, $\phi^*_{g,\tau}$ is positive, strictly increasing in τ and diverges to infinity. This proves condition (82). Similar to condition (59, one can also prove the following recursive relationship about ϕ^*_g:

$$\frac{\phi^*_{g,\tau+1}}{\phi^*_{g,\tau}} + \beta\frac{\phi^*_{g,\tau-1}}{\phi^*_{g,\tau}} = 1 + \beta + \sigma\kappa\Omega_c \quad \forall \tau \geq 1. \tag{90}$$

Moreover, as condition (61), we know $\frac{\phi^*_{g,\tau}}{\phi^*_{g,\tau-1}}$ must converge to $\Gamma^*_g > 0$, as $\tau \to \infty$:

$$\Gamma^*_g + \beta\frac{1}{\Gamma^*_g} = 1 + \beta + \sigma\kappa\Omega_c. \tag{91}$$

[53] For simplicity here, we always remove common knowledge about z among consumer here. We allow $\lambda_f \in (0,1]$. In other words, we nest the case in which firms have perfect knowledge about z.

We now turn to the case of incomplete information and establish the proof of condition (83). Similar to the proof of Proposition 7, we can find $\phi_g, \varpi_g : (0,1] \times (0,1] \times \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that, for any $t \leq T-1$,

$$\begin{aligned} \tilde{c}_t - \tilde{c}_t^{trap} &= \phi_g\left(\lambda_c, \lambda_f, T-t\right) \bar{E}_t^c[\tilde{g}_T], &(92)\\ \tilde{\pi}_t - \tilde{\pi}_t^{trap} &= \kappa\left(\Omega_c\left(\tilde{c}_t - \tilde{c}_t^{trap}\right) + (1-\Omega_c)\,\tilde{g}_t\right) + \varpi_g\left(\lambda_c, \lambda_f, T-t\right) \bar{E}_t^f[\tilde{R}_T]. &(93) \end{aligned}$$

Using conditions (80) and (81), we have $\phi_g(\lambda_c, \lambda_f, 1) = \sigma\kappa(1-\Omega_c)$, $\varpi_g(\lambda_c, \lambda_f, 1) = \beta\kappa(1-\Omega_c)$ and, for all $\tau \geq 2$,

$$\phi_g\left(\lambda_c, \lambda_f, \tau\right) = \sigma\beta^{\tau-1}\left(1-\Omega_c\right)\kappa + \sum_{k=1}^{\tau-1}\beta^{k-1}\left[\left(1-\beta+\sigma\kappa\Omega_c\right)\lambda_c\phi_g\left(\lambda_c, \lambda_f, \tau-k\right) + \sigma\lambda_f\varpi_g\left(\lambda_c, \lambda_f, \tau-k\right)\right]; \tag{94}$$

$$\varpi_g\left(\lambda_c, \lambda_f, \tau\right) = \left(1-\Omega_c\right)\frac{\kappa}{\theta}\left(\beta\theta\right)^{T-t} + \sum_{k=1}^{\tau-1}\left(\beta\theta\right)^k\left(\tfrac{\kappa\lambda_c\Omega_c}{\theta}\phi_g\left(\lambda_c, \lambda_f, \tau-k\right) + \tfrac{(1-\theta)\lambda_f}{\theta}\varpi_g\left(\lambda_c, \lambda_f, \tau-k\right)\right). \tag{95}$$

Together, we can establish, for all $\tau \geq 2$,

$$\phi_g\left(\lambda_c, \lambda_f, \tau\right) = \left(\beta + \left(1-\beta+\sigma\kappa\Omega_c\right)\lambda_c\right)\phi_g\left(\lambda_c, \lambda_f, \tau-1\right) + \sigma\lambda_f\varpi_g\left(\lambda_c, \lambda_f, \tau-1\right); \tag{96}$$

$$\varpi_g\left(\lambda_c, \lambda_f, \tau\right) = \kappa\beta\lambda_c\Omega_c\phi_g\left(\lambda_c, \lambda_f, \tau-1\right) + \beta\left[\theta + \left(1-\theta\right)\lambda_f\right]\varpi_g\left(\lambda_c, \lambda_f, \tau-1\right). \tag{97}$$

From the above conditions, we can see that for for all $\tau \geq 2$, $\phi_{g,\tau} = \phi_g\left(\lambda_c, \lambda_f, \tau\right)$ is strictly increasing in λ_c and λ_f. As $\phi^*_{g,\tau} = \phi_g(1,1,T)$, we also have $\phi_{g,\tau} \in \left(0, \phi^*_{g,\tau}\right)$.

Now, we now prove that, whenever $\lambda_c < 1$, the ratio $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}} = \frac{\phi_g(\lambda_c,\lambda_f,\tau)}{\phi^*_{g,\tau}}$ is strictly decreasing in $\tau \geq 1$ and converges to 0 as $\tau \to \infty$. To this goal, similar to condition (74), we try to establish that

$$\frac{\phi_{g,\tau}}{\phi_{g,\tau-1}} + \beta\frac{\phi_{g,\tau-2}}{\phi_{g,\tau-1}} \leq 1 + \beta + \sigma\kappa\Omega_c\lambda_c < 1 + \beta + \sigma\kappa\Omega_c \quad \forall\tau \geq 3. \tag{98}$$

From condition (96), we have, for $\tau \geq 3$,

$$\phi_{g,\tau} = \left(\beta + \left(1-\beta\right)\lambda_c\right)\phi_{g,\tau-1} + \sigma\kappa\Omega_c\lambda_c\phi_{g,\tau-1} + \sigma\lambda_f\varpi_{g,\tau-1},$$

$$\frac{\beta}{\beta + \left(1-\beta\right)\lambda_c}\phi_{g,\tau-1} = \beta\phi_{g,\tau-2} + \frac{\sigma\beta\kappa\Omega_c\lambda_c}{\beta + \left(1-\beta\right)\lambda_c}\phi_{g,\tau-2} + \frac{\sigma\beta\lambda_f}{\beta + \left(1-\beta\right)\lambda_c}\varpi_{g,\tau-2}.$$

From the previous two conditions, we have, for $\tau \geq 3$,

$$\begin{aligned} \phi_{g,\tau} + \beta\phi_{g,\tau-2} = {} & \left(\beta + \left(1-\beta\right)\lambda_c\right)\phi_{g,\tau-1} + \sigma\kappa\Omega_c\lambda_c\phi_{g,\tau-1} + \sigma\lambda_f\varpi_{g,\tau-1} \\ & + \frac{\beta}{\beta + \left(1-\beta\right)\lambda_c}\phi_{g,\tau-1} - \frac{\sigma\beta\kappa\Omega_c\lambda_c}{\beta + \left(1-\beta\right)\lambda_c}\phi_{g,\tau-2} - \frac{\sigma\beta\lambda_f}{\beta + \left(1-\beta\right)\lambda_c}\varpi_{g,\tau-2}. \end{aligned} \tag{99}$$

Note that, for $\tau \geq 3$ and $\lambda_c, \lambda_f \in (0,1]$, from condition (97), we have

$$\left[(\beta + (1-\beta)\lambda_c) + \sigma\kappa\Omega_c\lambda_c + \frac{\beta}{\beta + (1-\beta)\lambda_c}\right]\phi_{g,\tau-1} \leq (1 + \beta + \sigma\kappa\Omega_c\lambda_c)\,\phi_{g,\tau-1},$$

and

$$\begin{aligned}
&\sigma\lambda_f\varpi_{g,\tau-1} - \frac{\sigma\beta\kappa\Omega_c\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{g,\tau-2} - \frac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{g,\tau-2}\\
=&\sigma\lambda_f\left(\kappa\beta\lambda_c\Omega_c\phi_{g,\tau-2} + \beta\left[\theta + (1-\theta)\lambda_f\right]\varpi_{g,\tau-2}\right) - \frac{\sigma\beta\kappa\Omega_c\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{g,\tau-2} - \frac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{g,\tau-2}\\
=&\sigma\kappa\beta\Omega_c\lambda_c\left(\lambda_f - \frac{1}{\beta+(1-\beta)\lambda_c}\right)\phi_{g,\tau-2} + \sigma\beta\lambda_f\left[\theta + (1-\theta)\lambda_f - \frac{1}{\beta+(1-\beta)\lambda_c}\right]\varpi_{g,\tau-2}\\
\leq&0.
\end{aligned}$$

Together with condition (99), we reach at condition (98). To prove $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}$ is strictly decreasing in τ, note that we already prove that $\frac{\phi_{g,2}}{\phi^*_{g,2}} < 1 = \frac{\phi_{g,1}}{\phi^*_{g,1}}$. We proceed by induction on τ. If $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}} < \frac{\phi_{g,\tau-1}}{\phi^*_{g,\tau-1}}$ for $\tau \geq 2$, we have $\frac{\phi_{g,\tau-1}}{\phi_{g,\tau}} > \frac{\phi^*_{g,\tau-1}}{\phi^*_{g,\tau}}$. From (90) and (98), we have $\frac{\phi_{g,\tau+1}}{\phi_{g,\tau}} < \frac{\phi^*_{g,\tau+1}}{\phi^*_{g,\tau}}$ and thus $\frac{\phi_{g,\tau+1}}{\phi^*_{g,\tau+1}} < \frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}$. This finishes the proof that $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}$ is strictly decreasing in $\tau \geq 1$.

To prove that $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}$ converges to 0 as $\tau \to \infty$. Because $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}} > 0$ is strictly decreasing in $\tau \geq 1$, there exists $\Gamma_g \in [0,1)$ such that $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}} \to \Gamma_g$ as $\tau \to \infty$. If $\Gamma_g > 0$, we have $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}\frac{\phi^*_{g,\tau-1}}{\phi_{g,\tau-1}} \to 1$ as $\tau \to \infty$. Because $\frac{\phi^*_{g,\tau}}{\phi^*_{g,\tau-1}} \to \Gamma^*_g$, we have $\frac{\phi_{g,\tau}}{\phi_{g,\tau-1}} \to \Gamma^*_g$ and $\frac{\phi_{g,\tau-2}}{\phi_{g,\tau-1}} \to \frac{1}{\Gamma^*_g}$ as $\tau \to \infty$. From condition (91), we have $\frac{\phi_{g,\tau}}{\phi_{g,\tau-1}} + \beta\frac{\phi_{g,\tau-2}}{\phi_{g,\tau-1}} \to 1 + \beta + \sigma\kappa\Omega_c$ as $\tau \to \infty$. However, this is inconsistent with (98) when $\lambda_c < 1$ and $\kappa > 0$. As a result, $\Gamma_g = 0$.

Finally, we prove that, when λ_c is sufficiently low, $\phi_g(\lambda_c, \lambda_f, \tau)$ converges to zero as $\tau \to \infty$. The eigenvalues of the dynamical system $(\phi_{g,\tau}, \varpi_{g,\tau})$ based on conditions (96) and (97) are

$$\begin{aligned}
m_1 = &\frac{\beta + (1-\beta+\sigma\kappa\Omega_c)\lambda_c + \beta\left[(1-\theta)\lambda_f + \theta\right]}{2}\\
&- \frac{\sqrt{\left(\beta + (1-\beta+\sigma\kappa\Omega_c)\lambda_c - \beta\left[(1-\theta)\lambda_f + \theta\right]\right)^2 + 4\sigma\beta\lambda_f\lambda_c\Omega_c\kappa}}{2} > 0;
\end{aligned}$$

$$\begin{aligned}
m_2 = &\frac{\beta + (1-\beta+\sigma\kappa\Omega_c)\lambda_c + \beta\left[(1-\theta)\lambda_f + \theta\right]}{2}\\
&+ \frac{\sqrt{\left(\beta + (1-\beta+\sigma\kappa\Omega_c)\lambda_c - \beta\left[(1-\theta)\lambda_f + \theta\right]\right)^2 + 4\sigma\beta\lambda_f\lambda_c\Omega_c\kappa}}{2} > m_1;
\end{aligned}$$

Note that $\lim_{\lambda_c \to 0} m_2 = \beta < 1$. As a result, when λ_c is sufficiently low, both eigenvalues are below 1, which means that $\phi_g(\lambda_c, \lambda_f, \tau)$ converges to zero as $\tau \to \infty$.

Appendix B. Learning

In this appendix, we first prove that Theorem 1 holds for two leading forms of learning studied in the literature. We next prove that an asymptotic version of the result holds for arbitrary forms of learning, provided that higher-order uncertainty is bounded away from zero, in a sense to be made precise.

Horizon effects with sticky information or noisy private learning Consider the following two cases of learning.

Case 1. Agents become gradually aware of $\Theta_T \sim N\left(0, \sigma_\theta^2\right)$, as in Mankiw and Reis (2002) and Wiederholt (2015). Specially, at each period $t \in \{0, ..., T-1\}$, a fraction λ_{sticky} of agents who have not become aware about Θ_T become aware about Θ_T. Θ_T becomes commonly known at period T.

Case 2. Agents receive a new private signal about $\Theta_T \sim N\left(0, \sigma_\theta^2\right)$ in each period, as in Woodford (2003*a*), Nimark (2008), and Mackowiak and Wiederholt (2009). In each period $t \in \{0, ..., T-1\}$, each agent i's new information about Θ_T can be summarized by a private signal of Θ_T: $s_{i,t} = \Theta_T + v_{i,t}$, with $v_{i,t} \sim N\left(0, \sigma_{v,t}^2\right)$, i.i.d across i and t, and independent with respect to Θ_T.

In both cases, there exists $\{\lambda_t\}_{t=0}^{T-1}$ such that, for all t, $\lambda_t \in (0,1)$ and, for any $h \in \{1, ..., T\}$ and $0 \leq t_1 < t_2 < \cdots < t_h < T$,

$$\bar{E}_{t_1}[\bar{E}_{t_2}[...[\bar{E}_{t_h}[\Theta_T]...]] = \lambda_{t_1} \cdots \lambda_{t_h} \Theta_T, \tag{100}$$

with $\lambda_t = 1 - (1 - \lambda_{sticky})^{t+1}$ in case 1 and $\lambda_t = \frac{\sum_{\tau=0}^{t} \sigma_{v,t}^{-2}}{\sum_{\tau=0}^{t} \sigma_{v,t}^{-2} + \sigma_\theta^{-2}}$ in case 2.

Now we prove by induction that, for all $t \leq T-1$,

$$a_t = (\gamma + \alpha) \Pi_{\tau=t+1}^{T-1} (\gamma + \lambda_\tau \alpha) \bar{E}_t[\Theta_T]. \tag{101}$$

Since $\Theta_t = 0$ for all $t \neq T$, together with condition (15), we have $a_T = \Theta_T$ and $a_{T-1} = (\gamma + \alpha) \bar{E}_{T-1}[\Theta_T]$. As a result, condition (101) holds for $t = T-1$. Now, pick a $t \leq T-2$, assume that the claim holds for all $\tau \in \{t+1, ..., T-1\}$, and let us prove that it also holds for t. Using the claim for all $\tau \in \{t+1, ..., T-1\}$, condition (15), and condition (100), we have, for $t \leq T-2$,

$$a_t = \gamma^{T-t} \bar{E}_t [\Theta_T] + \alpha \bar{E}_t [a_{t+1}] + \alpha \sum_{k=2}^{T-t} \gamma^{k-1} \bar{E}_t [a_{t+k}];$$

$$\gamma \bar{E}_t [a_{t+1}] = \lambda_{t+1} \gamma^{T-t} \bar{E}_t [\Theta_T] + \alpha \lambda_{t+1} \sum_{k=2}^{T-t} \gamma^{k-1} \bar{E}_t [a_{t+k}].$$

As a result, we have $a_t = \left(\frac{\gamma}{\lambda_{t+1}} + \alpha\right) \bar{E}_t [a_{t+1}]$. Together with condition (101) for $t+1$, we have

$$a_t = (\gamma + \alpha) \Pi_{\tau=t+1}^{T-1} (\gamma + \lambda_\tau \alpha) \bar{E}_t[\Theta_T].$$

This proves condition (101) for all $t \leq T-1$. As a result, $\phi_T = (\gamma + \alpha) \Pi_{t=1}^{T-1} (\gamma + \lambda_t \alpha)$. Together with the fact that $\lambda_t \in (0,1)$ and $\phi_T^* = (\gamma+\alpha)^T$, we prove Theorem 1 for the case with learning.

The limit property with arbitrary learning. As noted in the main text, it is possible to prove, under quite general conditions, an *asymptotic* version of our horizon effect: as long as the higher-order uncertainty is bounded away from zero (in a sense we make precise now), the scalar ϕ_T becomes vanishingly small relative to ϕ_T^* as $T \to \infty$.

For any $t \leq T-1$ and any $k \in \{1,...,T-t\}$, we henceforth let B_t^k collect all the relevant k-order beliefs, as of period t :

$$B_t^k \equiv \left\{x \ : \ \exists (t_1, t_2, ..., t_k), \text{ with } t = t_1 < t_2 < ... < t_k \leq T-1, \text{ such that } x = \bar{E}_{t_1}[\bar{E}_{t_2}[\cdots \bar{E}_{t_k}[\Theta_T]\cdots]]\right\}.$$

We next introduce the following assumption.

Assumption 4 (Non-Vanishing Higher-Order Uncertainty) *There exists an $\epsilon > 0$ such that:*

(i) For all $t \in \{0,....,T-1\}$, there exists at least a mass ϵ of agents such that

$$Var\left(E_t[x]|\omega_i^t\right) \geq \epsilon Var\left(E_t\left[x\right]\right),$$

for all $x \in B_\tau^k \cup \{\Theta_T\}$, $\tau \in \{t+1,...,T-1\}$, and $k \in \{1,...,T-\tau\}$, where ω_i^t is agent i's information set at period t and $E_t\left[x\right]$ denotes the rational expectation of variable x conditional on the union of information sets of all agents in the economy available at period t.

(ii) $Var\left(\bar{E}_0[\Theta_T]\right) \geq \epsilon$.

To interpret this assumption, note that complete information imposes that $E_t\left[x\right]$ is known to every agent, and therefore that $Var\left(E_t\left[x\right]]|\omega_i^t\right) = 0$, regardless of how volatile $E_t\left[x\right]$ itself is. By contrast, letting $Var\left(E_t\left[x\right]|\omega_i^t\right) > 0$ whenever $Var\left(E_t\left[x\right]\right) > 0$ is essentially *tautological* to assuming that agents have incomplete information or, equivalently, that they face higher-order uncertainty. Relative to this tautology, part (i) introduces an arbitrarily small bound on the level of higher-order uncertainty. This bound guarantees that the higher-order uncertainty does not vanish as we let T go to infinity. Part (ii), on the other hand, means simply that there is non-trivial variation in first-order beliefs in the first place. The next result then follows from Theorem 1 and formalizes our point that our horizon effect, at least in its limit form, holds for arbitrary forms of learning.

Proposition 9 (Limit) *Under Assumption 4, the ratio $\frac{\phi_T}{\phi_T^*}$ converges to zero as $T \to \infty$.*

Proof of Proposition 9. We first prove that, under Assumption (4),

$$Var\left(y\right) \leq \left(1-\epsilon^2\right)^k Var\left(\Theta_T\right). \tag{102}$$

for any $t \leq T-1$ and $y = \bar{E}_t\bar{E}_{t_2}...\bar{E}_{t_k}[\Theta_T] \in B_t^k$.

To simplify notation, let $x = \bar{E}_{t_2}...\bar{E}_{t_k}[\Theta_T]$ for $k \geq 2$, and $x = \Theta_T$ for $k = 1$. From Assumption 4, we have, there is at least a mass ϵ of agents such that

$$Var\left(E_t\left[x\right]|\omega_i^t\right) \geq \epsilon Var\left(E_t\left[x\right]\right).$$

As a result,

$$E\left[Var\left(E_t\left[x\right]|\omega_i^t\right)|\Omega_t\right] \geq \epsilon^2 Var\left(E_t\left[x\right]\right),$$

where Ω_t is the cross-sectional distribution of agent's information set ω_i^t at period t. Using the law of total variance, we have

$$Var\left(E_t\left[x\right]\right) = E\left[Var\left(E_t\left[x\right]|\omega_i^t\right)|\Omega_t\right] + Var\left(E\left[E_t\left[x\right]|\omega_i^t\right]\right) = E\left[Var\left(E_t\left[x\right]|\omega_i^t\right)|\Omega_t\right] + Var\left(E\left[x|\omega_i^t\right]\right).$$

As a result, we have[54]

$$\begin{aligned} Var\left(y\right) = Var\left(\bar{E}_t\bar{E}_{t_2}...\bar{E}_{t_k}[\Theta_T]\right) = Var\left(\bar{E}_t\left[x\right]\right) \leq Var\left(E\left[x|\omega_i^t\right]\right) \leq \left(1-\epsilon^2\right)Var\left(E_t\left[x\right]\right) \\ \leq \left(1-\epsilon^2\right)Var\left(x\right) = \left(1-\epsilon^2\right)Var\left(\bar{E}_{t_2}...\bar{E}_{t_k}[\Theta_T]\right). \end{aligned}$$

Iterating the previous condition proves (102).

Condition (102) provides an upper bound for the variance of all k-th order belief. Together with the fact that $\phi_T^* = s_{T,T}$ and, for any random variables X, Y and scalars $a, b \geq 0$,

$$\begin{aligned} Var(aX+bY) &= a^2Var(X) + 2abCov\left(X,Y\right) + b^2Var\left(Y\right) \\ &\leq a^2Var(X) + 2ab\sqrt{Var\left(X\right)Var\left(Y\right)} + b^2Var\left(Y\right) \\ &= \left(a\sqrt{Var\left(X\right)} + b\sqrt{Var\left(Y\right)}\right)^2, \end{aligned} \tag{103}$$

we have

$$\begin{aligned} \left(\frac{\phi_T}{\phi_T^*}\right)^2 = \left(\frac{Cov\left(a_0,\bar{E}_0\left[\Theta_T\right]\right)}{\phi_T^*Var\left(\bar{E}_0\left[\Theta_T\right]\right)}\right)^2 \leq \frac{Var\left(a_0\right)}{\left[\phi_T^*\right]^2Var\left(\bar{E}_0[\Theta_T]\right)} \leq \frac{1}{Var\left(\bar{E}_0[\Theta_T]\right)}\left[\sum_{k=1}^{T}\left(\frac{\chi_{k,T}}{s_{T,T}}\left(1-\epsilon^2\right)^{\frac{k}{2}}\sqrt{Var\left(\Theta_T\right)}\right)\right]^2 \\ = \left[\sum_{k=1}^{T}\left(\frac{\chi_{k,T}}{s_{T,T}}\left(1-\epsilon^2\right)^{\frac{k}{2}}\right)\right]^2\frac{Var\left(\Theta_T\right)}{Var\left(\bar{E}_0[\Theta_T]\right)}. \end{aligned} \tag{104}$$

For any $\vartheta > 0$, there exists $h \in \mathbb{N}_+$ such that $\frac{\left(1-\epsilon^2\right)^{\frac{h}{2}}}{1-\left(1-\epsilon^2\right)^{\frac{1}{2}}} \leq \frac{\vartheta}{2}$. From Theorem 2, there exists $T^* \in \mathbb{N}_+$ such

[54] We use the fact that for any random variable X, and any information set I, $Var\left(E\left[X|I\right]\right) \leq Var\left(X\right)$. We also use the fact that $\bar{E}_t\left[\cdot\right] = E\left[E\left[\cdot\,|\omega_i^t\right]|\Omega_t\right]$, where Ω_t is the cross sectional distribution of ω_i^t at time t,

that, for all $T \geq T^*$, $\sum_{k=1}^{h-1} \frac{\chi_{k,T}}{s_{T,T}} \leq \frac{\vartheta}{2}$. As a result, for all $T \geq T^*$,

$$\sum_{k=1}^{T} \frac{\chi_{k,T}}{s_{T,T}} \left(1-\epsilon^2\right)^{\frac{k}{2}} \leq \sum_{k=1}^{h-1} \frac{\chi_{k,T}}{s_{T,T}} + \sum_{k=h}^{T} \left(1-\epsilon^2\right)^{\frac{k}{2}} \leq \frac{\vartheta}{2} + \frac{\left(1-\epsilon^2\right)^{\frac{h}{2}}}{1-\left(1-\epsilon^2\right)^{\frac{1}{2}}} \leq \vartheta.$$

This proves

$$\sum_{k=1}^{T} \left(\frac{\chi_{k,T}}{s_{T,T}} \left(1-\epsilon^2\right)^{\frac{k}{2}} \right) \to 0 \text{ as } T \to +\infty.$$

Together with (104) and the fact that $Var\left(\bar{E}_0[\Theta_T]\right) \geq \epsilon$, the proof of Proposition 9 is completed.

Appendix C. Additional Results for the New Keynesian Model

In this Appendix, we provide a few additional results regarding the application of our insights in the context of a liquidity trap. We first explain how our results regarding the forward-guidance puzzle can be understood under the lenses of a discounted Euler condition and a discounted NKPC, and draw certain connections to the literature. We next show how our insights help lessen the paradox of flexibility. We finally show that all the results of Section 5 extend to the new type of beauty contest seen in condition (22).

Discounted Euler Condition and Discounted NKPC. Proposition 5 has already indicated how the lack of common knowledge is akin to introducing additional discounting in the forward-looking equations of a macroeconomic model. We now illustrate how this helps recast our results regarding forward guidance and fiscal multipliers under the lenses of a discounted Euler condition and a discounted NKPC.

For the present purposes, we make a minor modification to the setting used in Sections 6 and 7: for $t \leq T$, we let the firms lack knowledge of the concurrent level of marginal cost. For simplicity, we also let the firms and the consumers face the same level of friction, that is, we set $\lambda_c = \lambda_f = \lambda$. These modifications are not strictly needed but sharpen the representation offered below.[55]

Proposition 10 *The power of forward guidance in the absence of common knowledge, ϕ_T, is the same as that in a representative-agent variant in which the Euler condition and the NKPC are modified as follows, for all $t \leq T-1$:*

$$\tilde{y}_t = -\sigma \left\{ \tilde{R}_t - \lambda E_t \left[\tilde{\pi}_{t+1} \right] \right\} + M_c E_t \left[\tilde{y}_{t+1} \right] \tag{105}$$

$$\tilde{\pi}_t = \kappa' \tilde{y}_t + \beta M_f E_t \left[\tilde{\pi}_{t+1} \right] + \tilde{\mu}_t, \tag{106}$$

where $M_c \equiv \beta + (1-\beta)\lambda \in (\beta, 1]$, $M_f \equiv \theta + (1-\theta)\lambda \in (\theta, 1]$, and $\kappa' \equiv \kappa\lambda$.[56]

This result, which is analogous to Proposition 5 in our abstract setting, maps the incomplete-information ϕ_T of the economy under consideration to the complete-information ϕ_T^* of a *variant* economy, in which the Euler condition and the NKPC have been "discounted" in the manner described above. When we remove common knowledge, it is *as if* the representative consumer discounts her expectations of next period's aggregate income and inflation by a factor equal to, respectively, M_c and λ; and it is *as if* the representative firm discounts the future inflation by a factor equal to M_f.[57]

Consider first the discounting that shows up in the Euler condition. When β is close to 1, M_c is close to 1, even if λ is close to zero. This underscores that the multiplier inside the demand block—which gets attenuated by the absence of common knowledge—is weak in the textbook version of the New Keynesian

[55] Without these modifications, the obtained representation is a bit less elegant, but the essence remains the same; see Proposition 10 in the earlier, NBER version of our paper (Angeletos and Lian, 2016*b*).

[56] To be precise, condition (105) holds with $M_c = 1$ for $t = T-1$.

[57] The change in the slope of the NKPC, from κ to κ', is of relative little interest to us, because the effect of the informational friction through this slope cannot be identified separately from that of a higher Frisch elasticity or less steep marginal costs.

model. As mentioned in the main text, short horizons, counter-cyclical precautionary savings, and feedback effects between housing prices and consumer spending tend to reinforce this multiplier, thereby also increasing the discounting caused by the absence of common knowledge.

Consider next the discounting that shows up in the NKPC. For the textbook parameterization of the degree of price stickiness (meaning a price revision rate, $1-\theta$, equal to $2/3$), the effective discount factor, M_f, falls from 1 to .83 as we move from common knowledge ($\lambda = 1$) to the level of imperfection assumed in our numerical example ($\lambda = .75$). The magnitude of this discount helps explain the sizable effects seen in Figure 1. Under the considered parameterization, the *actual* response of inflation to news about future demand is greatly reduced relative the common-knowledge benchmark. The fact that the average consumer underestimates the inflation response, as well as the spending of other consumers, reinforces this effect and helps further attenuate the feedback loop between inflation and spending.

Relation to McKay, Nakamura and Steinsson (2016*a*) and Gabaix (2016). Related forms of discounting appear in McKay, Nakamura and Steinsson (2016*a*), for the Euler condition, and in Gabaix (2016), for both the Euler condition and the NKPC. In this regards, these papers and ours are complementary to one another. However, the underlying theory and its empirical manifestations are different.

McKay, Nakamura and Steinsson (2016*a*) obtain a discounted Euler condition at the aggregate level by introducing a specific combination of heterogeneity and market incompleteness that forces some agents to hit their borrowing constraints and breaks the individual-level Euler condition. This theory therefore ties the resolution of forward guidance to microeconomic evidence about the response of individual consumption to idiosyncratic shocks. By contrast, our theory ties the resolution of forward guidance to survey evidence about the response of average forecast errors to the underlying policy news. The two theories can therefore be quantified independently from one another—and it's an open question which is one is more relevant in the context of forward guidance.

Gabaix (2016) on the other hand, assumes two kinds of friction. The first is that agents are less responsive to any variation in interest rates and incomes due to "sparsity" (a form of adjustment cost). The second is that agents underestimate the response of future *aggregate* outcomes to exogenous shocks. The first is of purely decision-theoretic nature and, as the one in McKay, Nakamura and Steinsson (2016*a*), amounts to a distortion of the individual-level Euler condition. The second is more closely related to the one we have obtained here: by anchoring expectations of *aggregate* outcomes, it gives rise to discounting only at the aggregate level. In this regard, Gabaix's theory and ours have a similar empirical implication: they both let the average forecast of future inflation and income respond less than the complete-information, rational-expectations, benchmark. Yet, the two theories make distinct empirical predictions in two other dimensions. First, our theory ties that inertia of the forecasts to their cross-sectional heterogeneity. Second, our theory predicts that the forecast errors, and the associated attenuation, ought to decrease as time passes, agents accumulate more information, and higher-order beliefs converge to first-order beliefs. This prediction, which is not shared by Gabaix's approach, is validated by the evidence in Coibion and Gorodnichenko (2012).

Proof of Proposition 10. Let us first focus on the incomplete-information ϕ_T. When firms lack common knowledge of the concurrent level of marginal cost, condition (62) continues to hold but condition (63) becomes, for any $t \leq T$,

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = \kappa \sum_{k=0}^{T-t} (\beta\theta)^k \bar{E}_t^f \left[\tilde{y}_{t+k} - \tilde{y}_{t+k}^{trap}\right] + \tfrac{1-\theta}{\theta} \sum_{k=1}^{T-t} (\beta\theta)^k \bar{E}_t^f \left[\tilde{\pi}_{t+k} - \tilde{\pi}_{t+k}^{trap}\right].$$

Slightly different from conditions (64) and (65),[58] we can find functions $\phi, \omega : (0,1] \times \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that, for any $t \leq T-1$,

$$\tilde{y}_t - \tilde{y}_t^{trap} = -\phi(\lambda, T-t)\, \bar{E}_t^c[\tilde{R}_T], \tag{107}$$

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = -\omega(\lambda, T-t)\, \bar{E}_t^f[\tilde{R}_T], \tag{108}$$

where $\phi(\lambda, 1) = \sigma(1 + \sigma\lambda\kappa)$, $\omega(\lambda, 1) = \kappa\lambda\sigma(1 + \sigma\lambda\kappa) + \kappa\beta(\theta + (1-\theta)\lambda)\sigma$, and for $t \leq T-2$,

$$\phi(\lambda, T-t) = (\beta + (1-\beta)\lambda)\, \phi(\lambda, T-t-1) + \sigma\lambda\omega(\lambda, T-t-1), \tag{109}$$

$$\omega(\lambda, T-t) = \beta(\theta + (1-\theta)\lambda)\, \omega(\lambda, T-t-1) + \kappa\lambda\phi(\lambda, T-t). \tag{110}$$

We now derive the complete-information ϕ_T^* and ω_T^* of a *variant* economy, where they denote how the output and inflation at $t = 0$ responds to shocks to the representative agent's belief about $\tilde{R}_T$ at $t = 0$. From conditions (105), (106) and footnote 56, we have $\phi_1^* = \sigma(1 + \sigma\lambda\kappa)$, $\omega_1^* = \kappa\lambda\sigma(1 + \sigma\lambda\kappa) + \kappa\beta(\theta + (1-\theta)\lambda)\sigma$, and, for $t \leq T-2$,

$$\phi_{T-t}^* = (\beta + (1-\beta)\lambda)\, \phi_{T-t-1}^* + \sigma\lambda\omega_{T-t-1}^*, \tag{111}$$

$$\omega_{T-t}^* = \beta(\theta + (1-\theta)\lambda)\, \omega_{T-t-1}^* + \kappa\lambda\phi_{T-t}^*. \tag{112}$$

The previous conditions coincide with conditions (109) and (110), and prove Proposition 10.

On the Paradox of Flexibility We now consider the implications of our insights for the paradox of flexibility. In the standard model, the power of forward guidance and the fiscal multiplier vis-a-vis future government spending increase with the degree of price flexibility: ϕ_T^* increases with κ.[59] This property is directly related to the "paradox of flexibility" (Eggertsson and Krugman, 2012). The next result proves, in effect, that the mechanism identified in our paper helps diminish this paradox as well.

Proposition 11 (Price Flexibility) *Let ϕ_T^* be the scalar characterized in either Proposition 7 or Proposition 8 and set $\lambda_f = 1$. We have $\frac{\partial \phi_T^*}{\partial \kappa} > 0$ and $\frac{\partial}{\partial \lambda_c}\left(\frac{\partial \phi_T^*}{\partial \kappa}\right) > 0$. That is, the power of forward guidance and the fiscal*

[58] There are two differences compared to conditions (64) and (65). First, as we impose $\lambda_c = \lambda_f = \lambda$ here, ϕ and ω are functions of λ, the common parameter characterizing the degree of information friction. Second, as firms lack common knowledge of the concurrent level of marginal cost, it is easier to let ω measure how inflation as a whole responds to the average belief about $\tilde{R}_T$.

[59] Whenever we vary κ, we vary θ while keeping the Frisch elasticity constant, which means that variation in κ maps one-to-one to variation in the degree of price flexibility.

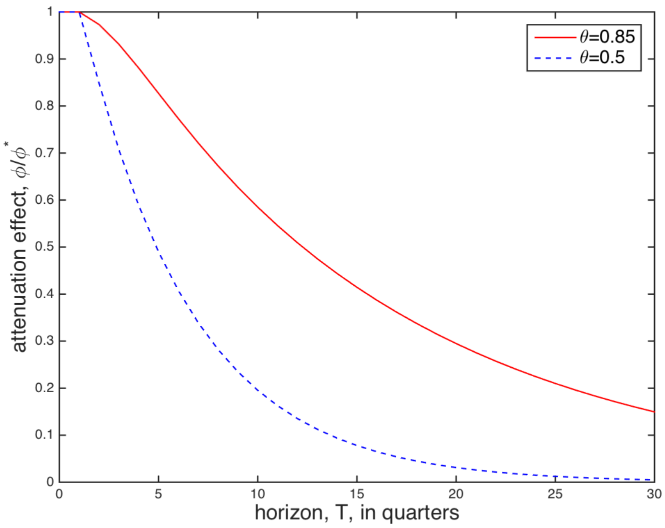

Figure 2: Varying the degree of price flexibility.

multiplier vis-a-vis future government spending increase with the degree of price flexibility, but at a rate that is slower the greater the departure from common knowledge.

This finding is an example of how lack of common knowledge reduces the paradox of flexibility more generally. In the standard model, a higher degree of price flexibility raises the GE effects of all kinds of demand shocks—whether these come in the form of forward guidance, discount rates, or borrowing constraints—because it intensifies the feedback loop between aggregate spending and inflation. By intensifying this kind of macroeconomic complementarity, however, a higher degree of price flexibility also raises the relative importance of higher-order beliefs, which in turn contributes to stronger attenuation effects of the type we have documented in this paper. In a nutshell, the very same mechanism that creates the paradox of flexibility within the New Keynesian framework also helps contain that paradox once we relax the common-knowledge assumptions of that framework.

Note that we have proved the above result only under the restriction $\lambda_f = 1$, which means that only the consumers lack common knowledge. Whenever $\lambda_f < 1$, there is a conflicting effect, which is that higher price flexibility reduces the strategic complementarity that operates within the supply block, thereby also reducing the role of λ_f itself. For the numerical example considered earlier, however, the overall effect of higher price flexibility is qualitatively the same whether $\lambda_f = 1$ or $\lambda_f = \lambda_c$.

We illustrate this in Figure 2. We let $\lambda_f = \lambda_c$, use the same parameter values as those used in Figure 1, and plot the relation between the ratio ϕ_T/ϕ_T^* and the horizon T under two values for θ. The solid red line corresponds to a higher value for θ, while the dashed blue line corresponds to a lower value for θ, that is, to more price flexibility. As evident in the figure, more price flexibility maps, not only to a lower ratio ϕ_T/ϕ_T^* (i.e., stronger attenuation) for any given T, but also to a more rapid decay in that ratio as we raise T.

Proof of Proposition 11 Consider first the environment studied in Section 6 and let us study the cross-partial derivative of the power of forward guidance with respect to κ and λ_c. To simplify notation, we use ϕ_τ and ϖ_τ as shortcuts for, respectively, $\phi(\lambda_c, \lambda_f, \tau)$ and $\varpi(\lambda_c, \lambda_f, \tau)$, where the functions ϕ and ϖ are defined as in the proof of Proposition 7. From conditions (66), we have

$$\frac{\partial \phi_1}{\partial \kappa} = \frac{\partial \phi(\lambda_c, 1, 1)}{\partial \kappa} = \sigma^2 > 0 \qquad \text{and} \qquad \frac{\partial \varpi_1}{\partial \kappa} = \frac{\partial \varpi(\lambda_c, 1, 1)}{\partial \kappa} = \sigma\beta > 0. \tag{113}$$

For any $\tau \geq 2$, when $\lambda_f = 1$, conditions (71) and (72) become

$$\phi_\tau = (\beta + (1 - \beta + \sigma\kappa)\lambda_c)\phi_{\tau-1} + \sigma\varpi_{\tau-1} \qquad \text{and} \qquad \varpi_\tau = \kappa\beta\lambda_c\phi_{\tau-1} + \beta\varpi_{\tau-1}.$$

As a result, for all $\tau \geq 2$, we have

$$\frac{\partial \phi_\tau}{\partial \kappa} = (\beta + (1 - \beta + \sigma\kappa)\lambda_c)\frac{\partial \phi_{\tau-1}}{\partial \kappa} + \sigma\lambda_c\phi_{\tau-1} + \sigma\frac{\partial \varpi_{\tau-1}}{\partial \kappa}, \tag{114}$$

$$\frac{\partial \varpi_\tau}{\partial \kappa} = \kappa\beta\lambda_c\frac{\partial \phi_{\tau-1}}{\partial \kappa} + \beta\lambda_c\phi_{\tau-1} + \beta\frac{\partial \varpi_{\tau-1}}{\partial \kappa}. \tag{115}$$

From conditions (113), (114) and (115), $\frac{\partial \phi_\tau}{\partial \kappa}$ and $\frac{\partial \varpi_\tau}{\partial \kappa}$ are strictly positive for any $\tau \geq 1$ by induction. Moreover, from conditions (66), (114) and (115), we have that $\frac{\partial \phi_2}{\partial \kappa}$ and $\frac{\partial \varpi_2}{\partial \kappa}$ are strictly increasing in λ_c. Then, from conditions (114), (115) and the fact that ϕ_τ itself is strictly increasing in λ_c for all $\tau \geq 2$, we have $\frac{\partial \phi_\tau}{\partial \kappa}$ and $\frac{\partial \varpi_\tau}{\partial \kappa}$ are strictly increasing in λ_c for all $\tau \geq 2$ by induction.

Consider now the environment studied in Section 7 and let us study the cross-partial derivative of the relevant fiscal multiplier with respect to κ and λ_c when $\lambda_f = 1$. Similarly to conditions (113), (114) and (115), we have

$$\frac{\partial \phi_{g,1}}{\partial \kappa} = \sigma(1 - \Omega_c) > 0 \qquad \text{and} \qquad \frac{\partial \varpi_{g,1}}{\partial \kappa} = \beta(1 - \Omega_c) > 0.$$

$$\frac{\partial \phi_{g,\tau}}{\partial \kappa} = (\beta + (1 - \beta + \sigma\kappa\Omega_c)\lambda_c)\frac{\partial \phi_{g,\tau-1}}{\partial \kappa} + \sigma\Omega_c\lambda_c\phi_{g,\tau-1} + \sigma\frac{\partial \varpi_{g,\tau-1}}{\partial \kappa},$$

$$\frac{\partial \varpi_{g,\tau}}{\partial \kappa} = \kappa\beta\lambda_c\Omega_c\frac{\partial \phi_{g,\tau-1}}{\partial \kappa} + \beta\lambda_c\Omega_c\phi_{g,\tau-1} + \beta\frac{\partial \varpi_{g,\tau-1}}{\partial \kappa}.$$

The result then follows from the same argument as before.

Extension of Lemma 2 and Theorems 1 and 2. Here we show that Lemma 2, Theorem 2 and, by implication, Theorem 1 extend to the kind of multi-layer beauty contest seen in condition (22) of Lemma 3.

Similar to the question of interest studied in Section 5, we impose Assumption 2 in condition (22). Similar to the proof of Lemma 2, we can find positively-valued coefficients $\{\chi_{h,\tau}\}_{\tau\geq 1, 1\leq h\leq \tau}$, such that, for any $t \leq T - 1$,

$$\tilde{y}_t - \tilde{y}_t^{trap} = \sum_{h=1}^{T-t}\left\{\chi_{h,T-t}\bar{E}_t^h\left[\tilde{R}_T\right]\right\}, \tag{116}$$

with $\tilde{y}_t^{trap}$ defined as in the proof of Proposition 6 and

$$\chi_{1,\tau} = \sigma\left(1+\tau\sigma\kappa\right)\beta^{\tau-1} \quad \forall \tau \geq 1, \tag{117}$$

$$\chi_{k,\tau} = \sum_{l=1}^{\tau-k+1}\left(1-\beta+l\sigma\kappa\right)\beta^{l-1}x_{k-1,\tau-l} \quad \forall k \geq 2 \text{ and } \tau \geq k. \tag{118}$$

We can then characterize the combined effect of beliefs of order up to k on spending, $s_{k,\tau}$, as[60]

$$s_{k,\tau} = \sigma\left(1+\tau\sigma\kappa\right)\beta^{\tau-1} + \sum_{l=1}^{\tau-1}\left(1-\beta+l\sigma\kappa\right)\beta^{l-1}s_{k-1,\tau-l} \quad \forall k \geq 1 \text{ and } \tau \geq 1. \tag{119}$$

Let $d_\tau = s_{\tau,\tau}$ denote the combined effect of beliefs of all different orders on spending. Similar to condition (17), $d_\tau = \phi_\tau^*$. Following the proof of Proposition 6, we have

$$d_0 = \sigma \qquad \text{and} \qquad d_1 = \sigma\left(1+\sigma\kappa\right),$$

$$\frac{d_\tau}{d_{\tau-1}} + \beta\frac{d_{\tau-2}}{d_{\tau-1}} = 1+\beta+\sigma\kappa \qquad \forall \tau \geq 2. \tag{120}$$

Now we prove $s_{k,\tau}$ satisfies an inequality with a similar form as condition (120):

$$\frac{s_{k,\tau}}{s_{k,\tau-1}} + \beta\frac{s_{k,\tau-2}}{s_{k,\tau-1}} \leq 1+\beta+\sigma\kappa \quad \forall \tau \geq 3 \text{ and } k \geq 1. \tag{121}$$

From condition (119), we have

$$\beta s_{k,\tau-1} = \sigma\left(1+\left(\tau-1\right)\sigma\kappa\right)\beta^{\tau-1} + \sum_{l=2}^{\tau-1}\left(1-\beta+\left(l-1\right)\sigma\kappa\right)\beta^{l-1}s_{k-1,\tau-l} \quad \forall k \geq 1 \text{ and } \tau \geq 2.$$

As a result, we can write $s_{k,\tau}$ in a recursive form:

$$s_{k,\tau} = \beta s_{k,\tau-1} + \left(1-\beta\right)s_{k-1,\tau-1} + \sigma^2\kappa\beta^{\tau-1} + \sigma\kappa\sum_{l=1}^{\tau-1}\beta^{l-1}s_{k-1,\tau-l} \quad \forall k \geq 1 \text{ and } \tau \geq 2,$$

$$\beta s_{k,\tau-1} = \beta^2 s_{k,\tau-2} + \beta\left(1-\beta\right)s_{k-1,\tau-2} + \sigma^2\kappa\beta^{\tau-1} + \sigma\kappa\sum_{l=2}^{\tau-1}\beta^{l-1}s_{k-1,\tau-l} \quad \forall k \geq 1 \text{ and } \tau \geq 3.$$

Using the previous two conditions, we have, for all $k \geq 1$ and $\tau \geq 3$,

$$s_{k,\tau} + \beta^2 s_{k,\tau-2} + \beta\left(1-\beta\right)s_{k-1,\tau-2} = 2\beta s_{k,\tau-1} + \left(1-\beta+\sigma\kappa\right)s_{k-1,\tau-1},$$

$$s_{k,\tau} + \beta s_{k,\tau-2} = \left(1+\beta+\sigma\kappa\right)s_{k,\tau-1} + \beta\left(1-\beta\right)\chi_{k,\tau-2} - \left(1-\beta+\sigma\kappa\right)\chi_{k,\tau-1}. \tag{122}$$

[60]Similar to the proof of Theorem 2, for notation simplicity, we extend the definition of $s_{h,\tau} = \sum_{r=1}^{h}\chi_{r,\tau}$ for all $h > \tau$. As for $h > \tau$, $\chi_{h,\tau} = 0$, we have $s_{h,\tau} = s_{\tau,\tau}$ for all $h > \tau$. We also define $s_{0,\tau} = 0$ for all $\tau \geq 1$.

To prove (121), we only need to prove:

$$\beta\left(1-\beta\right)\chi_{k,\tau-2} \leq \left(1-\beta+\sigma\kappa\right)x_{k,\tau-1} \quad \forall k \geq 1 \; and \; \tau \geq 3. \tag{123}$$

In fact, we prove the following stronger result:

$$\beta\chi_{k,\tau-2} \leq x_{k,\tau-1} \quad \forall k \geq 1 \; and \; \tau \geq 3. \tag{124}$$

From condition (117), we know that (124) is true for $k=1$ and $\tau \geq 3$. From condition (118), we know that

$$\chi_{k,\tau-1} = \sum_{l=1}^{\tau-k}\left(1-\beta+\sigma l\kappa\right)\beta^{l-1}x_{k-1,\tau-1-l} \quad \forall k \geq 2 \; and \; \tau \geq k+1, \tag{125}$$

$$\beta\chi_{k,\tau-2} = \sum_{l=2}^{\tau-k}\left(1-\beta+\sigma\left(l-1\right)\kappa\right)\beta^{l-1}x_{k-1,\tau-1-l} \quad \forall k \geq 2 \; and \; \tau \geq k+2.$$

This proves $\beta\chi_{k,\tau-2} \leq x_{k,\tau-1}$ for $k \geq 2$ and $\tau \geq k+2$. Together with the fact that, $\chi_{k,\tau-2} = 0 \; \forall k \geq \tau-1$, we prove (124) and thus (123). This finishes the proof of (121).

Based on (120) and (121), we can then establish a result akin to Theorem 2. That is, for any given $k \geq 1$ and $\tau \geq k$, the relative contribution of the first k orders, $\frac{s_{k,\tau}}{s_{\tau,\tau}} = \frac{s_{k,\tau}}{d_\tau}$, strictly decreases with τ.

First, note that, for any given $k \geq 1$,$1 = \frac{s_{k,k}}{d_k} > \frac{s_{k,k+1}}{d_{k+1}}$, because $x_{k+1,k+1} > 0$. Then, we can proceed by induction on $\tau \geq k$, for any fixed $k \geq 1$. If we have $\frac{s_{k,\tau}}{d_\tau} > \frac{s_{k,\tau+1}}{d_{\tau+1}}$ for some $\tau \geq k$, we have $\frac{s_{k,\tau}}{s_{k,\tau+1}} > \frac{d_\tau}{d_{\tau+1}}$. Using (120) and (121), we have $\frac{s_{k,\tau+2}}{s_{k,\tau+1}} < \frac{d_{\tau+2}}{d_{\tau+1}}$, and thus $\frac{s_{k,\tau+1}}{d_{\tau+1}} > \frac{s_{k,\tau+2}}{d_{\tau+2}}$. This completes the proof that, for any $k \geq 1$ and any $\tau \geq k$, the ratio $\frac{s_{k,\tau}}{s_{\tau,\tau}}$, strictly decreases with the horizon τ.

Finally, we prove that, for any $k \geq 1$,

$$\frac{s_{k,\tau}}{s_{\tau,\tau}} \to 0, \quad as \; \tau \to \infty. \tag{126}$$

In other words, we want to prove the relative contribution of the first k orders of beliefs to aggregate spending converges to zero when the horizon τ goes to infinity.

First note that, from condition (119), we have $s_{1,\tau} = \sigma\left(1+\sigma\tau\kappa\right)\beta^{\tau-1} \to 0,$ as $\tau \to +\infty$. From the proof of Proposition 6, we know $s_{\tau,\tau} = d_\tau = \phi^*_\tau \geq \sigma$. As a result, (126) is true for $k=1$.

If there exists $k \geq 2$ such that (126) does not hold, we let $k^* \geq 2$ denote the smallest of such k. Then, (126) holds for $1 \leq k \leq k^*-1$. Because we already prove that $\frac{s_{k^*,\tau}}{s_{\tau,\tau}} \geq 0$ is decreasing with the horizon τ, there exists $0 < \Gamma < 1$ such that $\frac{s_{k^*,\tau}}{s_{\tau,\tau}} = \frac{s_{k^*,\tau}}{\phi^*_\tau} \to \Gamma$ as $\tau \to \infty$. As a result, $\frac{s_{k^*,\tau}}{\phi^*_\tau}\frac{\phi^*_{\tau-1}}{s_{k^*,\tau-1}} \to 1$ as $\tau \to \infty$. Because we already prove that, in the proof of Proposition 6, $\frac{\phi^*_\tau}{\phi^*_{\tau-1}} \to \Gamma^*$, we have

$$\frac{s_{k^*,\tau}}{s_{k^*,\tau-1}} \to \Gamma^* \quad and \quad \frac{s_{k^*,\tau-2}}{s_{k^*,\tau-1}} \to \frac{1}{\Gamma^*} \quad as \; \tau \to \infty. \tag{127}$$

Note that since $s_{k^*,\tau} = s_{k^*-1,\tau} + \chi_{k^*,\tau}$ and $\frac{s_{k^*-1,\tau}}{s_{\tau,\tau}} \to 0 \; as \; \tau \to \infty$, we have $\frac{\chi_{k^*,\tau}}{s_{\tau,\tau}} = \frac{\chi_{k^*,\tau}}{\phi^*_\tau} \to \Gamma \; as \; \tau \to \infty$.

As a result,

$$\frac{\chi_{k^*,\tau}}{s_{k^*,\tau}} = \frac{\chi_{k^*,\tau}}{\phi^*_\tau}\frac{\phi^*_\tau}{s_{k^*,\tau}} \to 1 \quad as\ \tau \to \infty. \tag{128}$$

Now we prove a stronger version of (121)

$$\frac{s_{k,\tau}}{s_{k,\tau-1}} + \beta\frac{s_{k,\tau-2}}{s_{k,\tau-1}} + \sigma\kappa\frac{\chi_{k,\tau-1}}{s_{k,\tau-1}} \leq 1 + \beta + \sigma\kappa \quad \forall\tau \geq 3\ and\ k \geq 1. \tag{129}$$

This comes from the fact that (124) can be written as

$$\beta\left(1-\beta\right)\chi_{k,\tau-2} + \sigma\kappa\chi_{k,\tau-1} \leq \left(1-\beta+\sigma\kappa\right)x_{k,\tau-1} \quad \forall\tau \geq 3\ and\ k \geq 1. \tag{130}$$

Using (61), (127) and (128), we have

$$\frac{s_{k^*,\tau}}{s_{k^*,\tau-1}} + \beta\frac{s_{k^*,\tau-2}}{s_{k^*,\tau-1}} + \sigma\kappa\frac{\chi_{k^*,\tau-1}}{s_{k^*,\tau-1}} \to \Gamma^* + \beta\frac{1}{\Gamma^*} + \sigma\kappa = 1 + \beta + 2\sigma\kappa \quad as\ \tau \to \infty.$$

This contradicts (129) when $\kappa > 0$ and proves (126). This finishes the proof of the result akin to Theorem 2. Together with Proposition 3, we then establish the “horizon effects” akin to Theorem 1. Similarly, Theorem 4 also holds here.

References

Allen, Franklin, Stephen Morris, and Hyun Song Shin. (2006) "Beauty Contests and Iterated Expectations in Asset Markets." *Review of financial Studies*, 19(3): 719–752.

Andrade, Philippe, Gaballo Gaetano, Eric Mengus, and Benoit Mojon. (2015) "Forward Guidance and Heterogeneous Beliefs." *Banque De France mimeo*.

Angeletos, George-Marios, and Alessandro Pavan. (2007) "Efficient Use of Information and Social Value of Information." *Econometrica*, 75(4): 1103–1142.

Angeletos, George-Marios, and Chen Lian. (2016*a*) "Dampening General Equilibrium: From Micro to Macro." *Massachusetts Institute of Technology mimeo*.

Angeletos, George-Marios, and Chen Lian. (2016*b*) "Forward Guidance without Common Knowledge." *NBER Working Paper No. 22785*.

Angeletos, George-Marios, and Chen Lian. (2016*c*) "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination." *Handbook of Macroeconomics*, 2: forthcoming.

Angeletos, George-Marios, and Jennifer La'O. (2010) "Noisy business cycles." In *NBER Macroeconomics Annual 2009, Volume 24*. 319–378. University of Chicago Press.

Angeletos, George-Marios, and Jennifer La'O. (2013) "Sentiments." *Econometrica*, 81(2): 739–779.

Angeletos, George-Marios, Luigi Iovino, and Jennifer La'O. (2016) "Real Rigidity, Nominal Rigidity, and the Social Value of Information." *The American Economic Review*, 106(1): 200–227.

Auclert, Adrien. (2017) "Monetary Policy and the Redistribution Channel." *NBER Working Paper No. 23451*.

Bacchetta, Philippe, and Eric van Wincoop. (2006) "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?" *The American Economic Review*, 96(3): 552.

Bassetto, Marco. (2015) "Forward Guidance: Communication, Commitment, or Both?" *FRB of Chicago mimeo*.

Bergemann, Dirk, and Stephen Morris. (2013) "Robust Predictions in Games with Incomplete Information." *Econometrica*, 81(4): 1251–1308.

Brunnermeier, Markus K, and Yuliy Sannikov. (2014) "A Macroeconomic Model with a Financial sector." *The American Economic Review*, 104(2): 379–421.

Campbell, Jeffrey R, Charles L Evans, Jonas DM Fisher, and Alejandro Justiniano. (2012) "Macroeconomic Effects of Federal Reserve Forward Guidance." *Brookings Papers on Economic Activity*, 2012(1): 1–80.

Carlstrom, Charles T, Timothy S Fuerst, and Matthias Paustian. (2012) "How Inflationary is an Extended Period of Low Interest Rates?"

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo. (2011) "When Is the Government Spending Multiplier Large?" *Journal of Political Economy*, 119(1): 78–121.

Chung, Hess, Edward Herbst, and Michael T Kiley. (2015) "Effective Monetary Policy Strategies in New Keynesian Models: A Reexamination." *NBER Macroeconomics Annual*, 29(1): 289–344.

Cochrane, John. (2016*a*) "Do Higher Interest Rates Raise or Lower Inflation?" *Stanford University mimeo*.

Cochrane, John. (2016*b*) "The New-Keynesian Liquidity Trap." *Stanford University mimeo*.

Coibion, Olivier, and Yuriy Gorodnichenko. (2012) "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy*, 120(1): 116–159.

Coibion, Olivier, and Yuriy Gorodnichenko. (2015) "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review*, 105(8): 2644–78.

Del Negro, Marco, Marc P Giannoni, and Christina Patterson. (2015) "The Forward Guidance Puzzle." *FRB of New York mimeo*.

Eggertsson, Gauti B, and Paul Krugman. (2012) "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach." *The Quarterly Journal of Economics*, 127(3): 1469–1513.

Farhi, Emmanuel, and Iván Werning. (2017) "Monetary Policy, Bounded Rationality, and Incomplete Markets." *MIT mimeo*.

Gabaix, Xavier. (2016) "A Behavioral New Keynesian Model." *NBER Working Paper No. 22954*.

Galí, Jordi. (2008) *Monetary policy, Inflation, and the Business Cycle: an Introduction to the New Keynesian Framework and its Applications.* Princeton University Press.

Galí, Jordi, J David López-Salido, and Javier Vallés. (2007) "Understanding the Effects of Government Spending on Consumption." *Journal of the European Economic Association*, 5(1): 227–270.

Garcıa-Schmidt, Mariana, and Michael Woodford. (2015) "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis." *NBER Working Paper No. 21614*.

Grossman, Sanford J, and Joseph E Stiglitz. (1980) "On the Impossibility of Informationally Efficient Markets." *The American Economic Review*, 70(3): 393–408.

Kaplan, Greg, Benjamin Moll, and Giovanni L Violante. (2016) "Monetary Policy According to HANK." *NBER Working Paper No. 21897*.

Khaw, Mel Win, Luminita Stevens, and Michael Woodford. (2016) "Discrete Adjustment to a Changing Environment: Experimental Evidence." *NBER Working Paper No. 22978*.

Kiley, Michael T. (2016) "Policy Paradoxes in the New Keynesian Model." *Review of Economic Dynamics*, 21: 1–15.

Kiyotaki, Nobuhiro, and John Moore. (1997) "Credit Cycles." *Journal of Political Economy*, 105(2): 211–248.

Lucas, Robert E. (1972) "Expectations and the Neutrality of Money." *Journal of Economic Theory*, 4(2): 103–124.

Luo, Yulei, Jun Nie, Gaowang Wang, and Eric R. Young. (2017) "Rational inattention and the dynamics of consumption and wealth in general equilibrium." *Journal of Economic Theory*, 172(C): 55 – 87.

Mackowiak, Bartosz, and Mirko Wiederholt. (2009) "Optimal Sticky Prices under Rational Inattention." *American Economic Review*, 99(3): 769–803.

Mackowiak, Bartosz, Filip Matejka, and Mirko Wiederholt. (2017) "The Rational Inattention Filter." *Goethe University Frankfurt mimeo*.

Mankiw, N Gregory, and Ricardo Reis. (2002) "Sticky Information Versus Sticky Prices: a Proposal to Replace the New Keynesian Phillips Curve." *The Quarterly Journal of Economics*, 117(4): 1295–1328.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson. (2016*a*) "The Discounted Euler Equation: A Note." *NBER Working Paper No. 22129*.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson. (2016*b*) "The Power of Forward Guidance Revisited." *American Economic Review*, 106(10): 3133–3158.

Morris, Stephen, and Hyun Song Shin. (1998) "Unique equilibrium in a model of self-fulfilling currency attacks." *American Economic Review*, 587–597.

Morris, Stephen, and Hyun Song Shin. (2002) "Social Value of Public Information." *The American Economic Review*, 92(5): 1521–1534.

Morris, Stephen, and Hyun Song Shin. (2003) "Global games: theory and applications." In *Advances in Economics and Econometrics (Proceedings of the Eighth World Congress of the Econometric Society)*. Cambridge University Press.

Morris, Stephen, and Hyun Song Shin. (2006) "Inertia of Forward-looking Expectations." *The American Economic Review*, 152–157.

Nimark, Kristoffer. (2008) "Dynamic Pricing and Imperfect Common Knowledge." *Journal of Monetary Economics*, 55(2): 365–382.

Nimark, Kristoffer. (2017) "Dynamic Higher Order Expectations." *Cornell Univeristy mimeo*.

Samet, Dov. (1998) "Iterated Expectations and Common Priors." *Games and Economic Behavior*, 24(1): 131–141.

Sims, Christopher A. (2003) "Implications of Rational Inattention." *Journal of Monetary Economics*, 50(3): 665–690.

Singleton, Kenneth J. (1987) "Asset Prices in a Time-series Model with Disparately Informed, Competitive Traders." In *New Approaches to Monetary Economics*. , ed. William A. Barnett and Kenneth J. Singleton, 249–272. Cambridge University Press.

Tirole, Jean. (2015) "Cognitive Games and Cognitive Traps." *Toulouse School of Economics mimeo*.

Vellekoop, Nathanael, and Mirko Wiederholt. (2017) "Inflation Expectations and Choices of Households." *Goethe University Frankfurt mimeo*.

Werning, Iván. (2012) "Managing a Liquidity Trap: Monetary and Fiscal Policy." *NBER Working Paper No. 17344*.

Werning, Iván. (2015) "Incomplete Markets and Aggregate Demand." *NBER Working Paper No. 21448*.

Wiederholt, Mirko. (2015) "Empirical Properties of Inflation Expectations and the Zero Lower Bound." *Goethe University Frankfurt mimeo*.

Woodford, Michael. (2003*a*) "Imperfect Common Knowledge and the Effects of Monetary Policy." *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*.

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

Woodford, Michael. (2011) "Simple Analytics of the Government Expenditure Multiplier." *American Economic Journal: Macroeconomics*, 3(1): 1–35.