

Gradualism in Monetary Policy: A Time-Consistency Problem?*

Jeremy C. Stein
Harvard University and NBER

Adi Sunderam
Harvard University and NBER

First draft: June 2015
This draft: September 2015

Abstract

We develop a model of monetary policy with two key features: (i) the central bank has private information about its long-run target for the policy rate; and (ii) the central bank is averse to bond-market volatility. In this setting, discretionary monetary policy is gradualist, or inertial, in the sense that the central bank only adjusts the policy rate slowly in response to changes in its privately-observed target. Such gradualism reflects an attempt to not spook the bond market. However, this effort ends up being thwarted in equilibrium, as long-term rates rationally react more to a given move in short rates when the central bank moves more gradually. The same desire to mitigate bond-market volatility can lead the central bank to lower short rates sharply when publicly-observed term premiums rise. In both cases, there is a time-consistency problem, and society would be better off appointing a central banker who cares less about the bond market. We also discuss the implications of our model for forward guidance once the economy is away from the zero lower bound.

* We are grateful to Adam Wang-Levine for research assistance and to seminar participants at Stanford and the NBER Summer Institute MEFM meeting for their comments. A special thanks to Mike Woodford for an extremely helpful discussion of the paper at the NBER Summer Institute.

I. Introduction

In this paper, we attempt to shed some light on the well-known phenomenon of gradualism in monetary policy. As described by Bernanke (2004), gradualism is the idea that “the Federal Reserve tends to adjust interest rates incrementally, in a series of small or moderate steps in the same direction.” This behavior can be represented empirically by an inertial Taylor rule, with the current level of the federal funds rate modeled as a weighted average of a target rate—which is itself a function of inflation and the output gap as in, e.g., Taylor (1993)—and the lagged value of the funds rate. In this specification, the coefficient on the lagged funds rate captures the degree of inertia in policy. In recent U.S. samples, estimates of the degree of inertia are strikingly high, on the order of 0.85 in quarterly data.¹

Several authors have proposed theories in which this kind of gradualism can be thought of as optimal behavior on the part of the central bank. One influential line of thinking, due originally to Brainard (1967) and refined by Sack (1998), is that moving gradually makes sense when there is uncertainty about how the economy will respond to a change in the stance of policy. An alternative rationale comes from Woodford (2003), who argues that committing to move gradually gives the central bank more leverage over long-term rates for a given change in short rates, a property which is desirable in the context of his model.

In what follows, we offer a different take on gradualism. In our model, the observed degree of policy inertia is not optimal from an ex ante perspective, but rather reflects a fundamental time-consistency problem. The distinguishing feature of our approach is that the Fed is assumed to have private information about its preferred value of the target rate. In other words, the Fed knows something about its reaction function that the market does not. Although this assumption is not standard in the literature on monetary policy, it should be emphasized that it is necessary if one wants to address the basic fact that financial markets respond to monetary policy announcements.²

¹ Coibion and Gorodnichenko (2012) provide a comprehensive recent empirical treatment; see also Rudebusch (2002, 2006). Campbell, Pflueger, and Viceira (2015) argue that the degree of inertia in Federal Reserve rate-setting became more pronounced after about 2000.

² Said differently, if the Fed mechanically followed a policy rule that was a function only of publicly observable variables (e.g., the current values of the inflation rate and the unemployment rate), then the market would react to news releases about movements in these variables but not to Fed policy statements.

We also assume that the Fed behaves as if it is averse to bond-market volatility. We model this concern in reduced form, by simply putting the volatility of long-term rates into the Fed's objective function, but note that a preference of this sort can ultimately be rooted in an effort to deliver on its traditional dual mandate. For example, a bout of bond-market volatility may be undesirable not just in its own right, but rather because it is damaging to the financial system and hence to real economic activity and employment.

Nevertheless, in a world of private information and discretionary meeting-by-meeting decision-making, an attempt by the Fed to moderate bond-market volatility can be welfare-reducing. The logic here is similar to that in signal-jamming models (Holmstrom, 1999; Stein, 1989). Suppose the Fed observes a private signal that its long-run target value for the funds rate has permanently increased by 100 basis points. If it adjusts policy fully in response to this signal, raising the funds rate by 100 basis points, long-term rates will move by a similar amount. If it is averse to such a movement in long-term rates, the Fed will be tempted to announce a smaller change in the funds rate, thereby trying to fool the market into thinking that its private signal was less dramatic. Hence, it will under-adjust to its signal, perhaps raising the funds rate by only 25 basis points.

However, if bond-market investors come to understand this dynamic, the Fed's efforts to reduce volatility will be frustrated in equilibrium. The market will see the 25 basis-point increase in the funds rate and understand that it is likely to be just the first in a series of similar moves, so long-term rates will react more than one-for-one to the change in short rates. Indeed, in a rational-expectations equilibrium, the Fed's private signal will always be fully inferred by the market, regardless of the degree of gradualism. Still, if it acts on a discretionary basis, the Fed will always keep trying to fool the market. This is because when it decides how much to adjust the policy rate, it *takes as given the market's conjecture* about the degree of inertia in its rate-setting behavior. As a result, the Fed's behavior is inefficient from an ex ante perspective: by moving gradually, it does not succeed in its attempts to reduce bond-market volatility, but gradualism means that the policy rate is further from its long-run target than it otherwise would be.

This inefficiency reflects a commitment problem. In particular, the Fed cannot commit to not trying to smooth the private information that it communicates to the market via its changes in the

policy rate.³ One institutional solution to this problem, in the spirit of Rogoff (1985), would be to appoint a central banker who cares less about bond-market volatility than the representative member of society. More broadly, appointing such a market-insensitive central banker can be thought of as a metaphor for building a certain kind of institutional culture and set of norms inside the central bank such that bond-market movements are not given as much weight in policy deliberations.

We begin with a simple static model that is designed to capture the above intuition in as parsimonious a way as possible. The main result here is that in any rational-expectations equilibrium, there is always under-adjustment of the policy rate, as compared to the first-best outcome in which the Fed adjusts the policy rate fully in response to changes in its privately-observed target. Moreover, for some parameter values, there can be Pareto-ranked multiple equilibria with different degrees of under-adjustment. The intuition for these multiple equilibria is that there is two-way feedback between the market's expectations about the degree of gradualism on the one hand and the Fed's optimal choice of gradualism on the other.⁴ Specifically, if the market conjectures that the Fed is behaving in a highly inertial fashion, it will react more strongly to an observed change in the policy rate: in an inertial world, the market knows that there are further changes to come. But this strong sensitivity of long-term rates to changes in the policy rate makes the Fed all the more reluctant to move the policy rate, hence validating the initial conjecture of extreme inertia.

Next, we ask whether the commitment problem that we have identified can be mitigated with forward guidance, whereby the Fed announces in advance a preferred path for future short rates. We show that it cannot. Because its future private information is by definition not forecastable, forward guidance is a one-sided commitment device. It can only commit the Fed to incorporating new private information more slowly, not more quickly. Moreover, we argue that once the economy is away from the zero lower bound, forward guidance can actually be harmful if it is not implemented carefully. If we are in a region of the parameter space where there are multiple equilibria, forward guidance can increase the risk of getting stuck in the Pareto-inferior, more-inertial equilibrium.

³ The literature on monetary policy has long recognized a different commitment problem, namely that, under discretion, the central bank will be tempted to create surprise inflation so as to lower the unemployment rate. See, e.g., Kydland and Prescott (1977) and Barro and Gordon (1983). More recently, Farhi and Tirole (2012) have pointed to the time-consistency problem that arises from the central bank's ex post desire to ease monetary policy when the financial sector is in distress; their focus on the central bank's concern with financial stability is somewhat closer in spirit to ours.

⁴ For an informal description of this two-way feedback dynamic, see Stein (2014).

We then enrich the model by adding term premium shocks as a source of publicly observable variation in long-term interest rates. We show that in its efforts to reduce bond-market volatility, the Fed is tempted to try to offset term premium shocks, cutting the policy rate when term premiums rise. Effectively, the Fed tries to pretend that it has received dovish private information about the optimal target rate whenever there is a positive term premium shock even though the term premium shock is public. Again, this tactic is unsuccessful in reducing volatility in equilibrium. Thus, the presence of term premium shocks exacerbates the Fed's time-consistency problem.

Finally, we extend the model to an explicitly dynamic setting, which allows us to more fully characterize how a given innovation to the Fed's target works its way into the funds rate over time. These dynamic results enable us to draw a closer link between the mechanism in our model and the empirical evidence on the degree of inertia in the funds rate.

The remainder of the paper is organized as follows. Section II discusses some motivating evidence, based on readings of Federal Open Market Committee (FOMC) transcripts. Section III presents the static version of the model and summarizes our basic results on under-adjustment of the policy rate, multiple equilibria, forward guidance, and term premiums. Section IV develops a dynamic extension of the model, and Section V concludes.

II. Motivating Evidence from FOMC Transcripts

In their study of monetary-policy inertia, Coibion and Gorodnichenko (2012) use FOMC transcripts to document two key points. First, FOMC members sometimes speak in a way that is suggestive of a gradual-adjustment model—that is, they articulate a target for the policy rate, and then put forward reasons why it is desirable to adjust only slowly in the direction of that target. Second, one of the stated rationales for such gradualism appears to be a desire not to create financial-market instability. Coibion and Gorodnichenko highlight the following quote from Chairman Alan Greenspan at the March 1994 FOMC meeting:

“My own view is that eventually we have to be at 4 to 4½ percent. The question is not whether but when. If we are to move 50 basis points, I think we would create far more instability than we realize, largely because a half-point is not enough to remove the question of where we are ultimately going. I think there is a certain advantage in doing 25 basis points....”

In a similar spirit, at the August 2004 meeting, shortly after the Fed had begun to raise the funds rate from the low value of one percent that had prevailed since mid-2003, Chairman Greenspan remarked:

“Consequently, the sooner we can get back to neutral, the better positioned we will be. We were required to ease very aggressively to offset the events of 2000 and 2001, and we took the funds rate down to extraordinarily low levels with the thought in the back of our minds, and often said around this table, that we could always reverse our actions. Well, reversing is not all that easy.... We’ve often discussed that ideally we’d like to be in a position where, when we move as we did on June 30 and I hope today, the markets respond with a shrug. What that means is that the adjustment process is gradual and does not create discontinuous problems with respect to balance sheets and asset values.”

These sorts of quotes help motivate our basic modeling approach, in which gradualism in monetary policy reflects the Fed’s desire to keep bond-market volatility in check—in Greenspan’s words, to “not create discontinuous problems with respect to balance sheets and asset values.” This same approach may also be helpful in thinking about changes in gradualism over time. Campbell, Pflueger, and Viceira (2015) show that the degree of inertia in Fed rate-setting behavior became significantly more pronounced after about 2000; given the logic of our model, one might wonder whether this heightened inertia was associated with an increase over time in the Fed’s concern with financial markets. In a crude attempt to speak to this question, we examine all 200 FOMC transcripts for the 25-year period 1985-2009, and for each meeting, simply measure the frequency of words related to financial markets.⁵ Specifically, we count the number of times the terms “financial market,” “equity market,” “bond market,” “credit market,” “fixed income,” and “volatility” are mentioned.⁶ For each year, we aggregate this count and divide by the total number of words in that year’s transcripts.

Figure 1 displays the results from this exercise. As can be seen, there is a strong upward trend in the data. While there is also a good deal of volatility around the trend— mentions of financial-market terms decline in the years preceding the financial crisis and then spike up—a simple linear time trend captures almost 59 percent of the variation in the time series, with a t-statistic of 5.0.⁷ And

⁵ Given the five-year lag in making transcripts public, 2009 is the last available year.

⁶ We obtain similar results if we use different subsets of these terms. For instance, the results are similar if we only count the frequency of the term “financial market”.

⁷ Restricting the sample to the pre-crisis period 1985-2006, we obtain an R^2 of 46 percent and a t-statistic of 4.3. So the

the fitted value of the series based on this time trend goes up about four-fold over the 25-year sample period. While this analysis is extremely simplistic, it does suggest an increasing emphasis over time by the FOMC on financial-market considerations. If there was in fact such a change in FOMC thinking, the model we develop below is well-suited to drawing out its implications, both for the dynamics of the policy rate, as well as for longer-term yields.

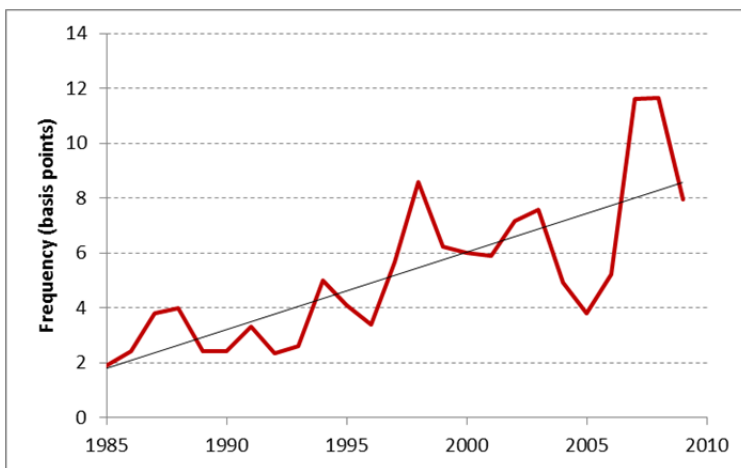


Figure 1. Frequency of terms related to financial markets in FOMC transcripts.

III. Static Model

We begin by presenting what is effectively a static version of the model, in which the Fed adjusts the funds rate only partially in response to a one-time innovation in its desired target rate. Later, in Section IV, we extend the model to a fully dynamic setting, in which we can be more explicit about how an innovation to the Fed’s target rate gradually makes its way into the funds rate over time. Though it is stylized, the static model highlights the main intuition for why a desire to limit bond-market volatility creates a time-consistency problem for the Fed.

A. Model Setup

We begin by assuming that at any time t , the Fed has a target rate based on its traditional dual-mandate objectives. This target rate is the Fed’s best estimate of the value of the federal funds rate

trend in the data is not primarily driven by post-2006 period.

that keeps inflation and unemployment as close as possible to their desired levels. For tractability, we assume that the target rate, denoted by i_t^* , follows a random walk, so that:

$$i_t^* = i_{t-1}^* + \varepsilon_t, \quad (1)$$

where $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2 \equiv \frac{1}{\tau_\varepsilon}\right)$ is normally distributed. Our key assumption is that i_t^* is private information of the Fed and is unknown to market participants before the Fed acts at time t . One can think of the private information embodied in i_t^* as arising from the Fed's attempts to follow something akin to a Taylor rule, where it has private information about either the appropriate coefficients to use in the rule (i.e., its reaction function), or about its own forecasts of future inflation or unemployment. As noted above, an assumption of private information along these lines is necessary if one wants to understand why asset prices respond to Fed policy announcements.

Once it knows the value of i_t^* , the Fed acts to incorporate some of its new private information ε_t into the federal funds rate i_t , which is observable to the market. We assume that the Fed picks i_t on a discretionary period-by-period basis to minimize the loss function L_t , given by:

$$L_t = (i_t^* - i_t)^2 + \theta (\Delta i_t^\infty)^2, \quad (2)$$

where i_t^∞ is the infinite-horizon forward rate. Thus, the Fed has the usual concerns about inflation and unemployment, as captured in reduced form by a desire to keep i_t^* close to i_t . However, when $\theta > 0$, the Fed also cares about the volatility of long-term bond yields, as measured by the squared change in the infinite-horizon forward rate.

For simplicity, we start by assuming that the expectations hypothesis holds, so there is no time-variation in the term premium. This implies that the infinite-horizon forward rate is equal to the expected value of the funds rate that will prevail in the distant future. Because the target rate i_t^* follows a random walk, the infinite-horizon forward rate is then given by the market's best estimate of the *current target* i_t^* at time t . Specifically, we have $i_t^\infty = E_t[i_t^*]$. In Section III.E below, we relax the assumption that the expectations hypothesis holds so that we can also consider the Fed's reaction to term-premium shocks.

Several features of the Fed's loss function are worth discussing. First, in our simple formulation, θ reflects the degree to which the Fed is concerned about bond-market volatility, over

and above its desire to keep i_t^* close to i_t . To be clear, this loss function need not imply that the Fed cares about asset prices for their own sake. An alternative interpretation is that volatility in financial-market conditions can affect the real economy and hence the Fed's ability to satisfy its traditional dual mandate. This linkage is not modeled explicitly here, but as one example of what we have in mind, the Fed might believe that a spike in bond-market volatility could damage highly-levered financial intermediaries and interfere with the credit-supply process. With this stipulation in mind, we take as given that θ reflects the socially "correct" objective function—in other words, it is exactly the value that a well-intentioned social planner would choose. We then ask whether there is a time-consistency problem when the Fed tries to optimize this objective function period-by-period, in the absence of a commitment technology.

Second, note that the Fed's target in the first term of (2) is the *short-term* policy rate, which it directly controls, whereas the bond-market volatility that it worries about in the second term of (2) refers to the variance of *long-term* market-determined rates. This short-versus-long divergence is crucial for our results on time consistency. To see why, consider two alternative loss functions:

$$L_t = (i_t^* - i_t)^2 + \theta(\Delta i_t)^2, \quad (2')$$

$$L_t = (i_t^{\infty*} - i_t^{\infty})^2 + \theta(\Delta i_t^{\infty})^2, \quad (2'')$$

In (2'), the Fed cares about the volatility of the funds rate, rather than the volatility of the long-term rate. This objective function mechanically produces gradual adjustment of the funds rate, but because there is no forward-looking long rate, there is no issue of the Fed trying to manipulate market expectations and hence no time-consistency problem. In other words, a model of this sort delivers a strong form of gradualism, but the resulting gradualism is entirely optimal from the Fed's perspective. However, by emphasizing only short-rate volatility, this formulation arguably fails to capture the financial-stability goal articulated by Greenspan, namely to "not create discontinuous problems with respect to balance sheets and asset values." We believe that putting the volatility of the long-term rate directly in the objective function, as in (2), does a better job in this regard.

In (2''), the Fed pursues its dual-mandate objectives not by having a target i_t^* for the funds rate, but instead by explicitly targeting a long rate of $i_t^{\infty*}$. It can be shown that in this case, too, given

that the same rate i_t^∞ appears in both parts of the objective function, there is no time-consistency problem.⁸ Moreover, one might argue that the first term in (2'') is a reasonable reduced-form way to model the Fed's efforts to achieve its dual mandate. In particular, in standard New-Keynesian models, it is long-term real rates, not short rates, that matter for inflation and output stabilization. Nevertheless, our preferred formulation of the Fed's objective function in (2) can be motivated in a couple of ways. First, (2) would appear to be more realistic than (2'') as a simple description of how the Fed actually behaves, and—importantly for our purposes—communicates about its future intentions.⁹ Second, in many models of the monetary transmission mechanism, including those based on a bank lending channel or a reaching-for-yield effect, the short rate has an independent effect on economic activity. This could help explain why it might make sense for the Fed to target the short rate per se.¹⁰

Finally, the assumption that the Fed cares about the volatility of the infinite-horizon forward rate i_t^∞ , along with the fact that the target rate i_t^* follows a random walk, makes for a convenient simplification. Because i_t^* follows a random walk, all of the new private information ε_t will, in expectation, eventually be incorporated into the future short rate. The reaction of the infinite-horizon forward rate reflects this revision in expectations and allows us to abstract from the exact dynamic path that the Fed follows in ultimately incorporating its new information into the funds rate. In contrast, revisions in finite-horizon forward rates depend on the exact path of the short rate. For example, suppose i_t^* were public information and changed from 1% to 2%. The infinite-horizon forward rate would then immediately jump to 2%, while the two-year forward rate might move by less if the market expected the Fed to tighten slowly. However, for our key conclusions, it is not

⁸ We are grateful to Mike Woodford for emphasizing this point to us, and for providing a simple proof.

⁹ For example, in its public communications, the Fed typically talks about its view of the appropriate level of short-term policy rates as a function of different economic scenarios. This can be seen in the Summary of Economic Projections, where FOMC members are asked to provide a preferred path for the funds rate (the so-called “dot plot”), and where the accompanying text notes that: “These charts are based on policymakers’ assessments of appropriate monetary policy, which, by definition, is the future path of [the funds rate] that each participant deems most likely to foster outcomes for economic activity and inflation that best satisfy his or her interpretation of the Federal Reserve’s dual objectives of maximum employment and stable prices.”

¹⁰ The large literature on the bank lending channel includes Bernanke and Blinder (1992), Kashyap and Stein (2000), and Drechsler, Savov, and Schnabl (2015). A few noteworthy recent papers on reaching for yield and its implications for the pricing of credit risk are Gertler and Karadi (2015), Jimenez, Ongena, Peydro, and Saurina (2014), and Dell’Ariccia, Laeven and Suarez (2013).

necessary that the Fed care about the volatility of the infinite-horizon forward rate; as we demonstrate below, a time-consistency problem still emerges (though in less extreme form) if the Fed cares instead about the volatility of a finite-horizon rate that is effectively a weighted average of short-term and infinite-horizon rates.¹¹

Once it observes i_t^* , we assume that the Fed sets the federal funds rate i_t by following a partial adjustment rule of the form:

$$i_t = i_{t-1} + k(i_t^* - i_{t-1}) + u_t, \quad (3)$$

where $u_t \sim N\left(0, \sigma_u^2 \equiv \frac{1}{\tau_u}\right)$ is normally distributed noise that is overlaid onto the rate-setting process.

The noise u_t is simply a technical modeling device; its usefulness will become clear shortly. Loosely speaking, this noise—which can be thought of as a “tremble” in the Fed’s otherwise optimally-chosen value of i_t —ensures that the Fed’s actions cannot be perfectly inverted to fully recover its private information i_t^* . As will be seen, this imperfect-inversion feature helps to avoid some degenerate equilibrium outcomes. In any event, in all the cases we consider, σ_u^2 is set to be very small relative to σ_ε^2 , so that one should not try to read too much economic content into the construct. One possibility would be to think of u_t as coming from the fact that the Fed uses round numbers (typically in 25 basis-point increments) for the funds rate settings that it communicates to the market, while its underlying private information about i_t^* is presumably continuous.

The market tries to infer the Fed’s private information i_t^* based on its observation of the funds rate i_t . To do so, it conjectures that the Fed follows a rule given by:

$$i_t = i_{t-1} + \kappa(i_t^* - i_{t-1}) + u_t. \quad (4)$$

Thus, the market correctly conjectures the form of the Fed’s smoothing rule but, crucially, it does not directly observe the Fed’s smoothing parameter k ; rather, it has to make a guess κ as to the value of

¹¹ The time-consistency problem arises because of the forward-looking nature of longer-term rates, and the resulting incentives that the Fed has to manipulate market expectations. As long as the Fed seeks to moderate the volatility of a finite-horizon rate that has some forward-looking component, an element of the time-consistency problem will remain.

this parameter. In a rational-expectations equilibrium, this guess will turn out to be correct, and we will have $\kappa = k$. However—and this is a key piece of intuition—when the Fed decides how much smoothing to do, it takes the market’s conjecture κ as a fixed parameter and does not impose that $\kappa = k$. The equilibrium concept here is thus of the “signal-jamming” type introduced by Holmstrom (1999): the Fed, taking the bond market’s estimate of κ as fixed, tries to fool the market into thinking that i_t^* has moved by less than it actually has, in an effort to reduce the volatility of long-term rates. In a Nash equilibrium, these attempts to fool the market wind up being fruitless, but the Fed cannot resist the temptation to try.¹²

B. Equilibrium

We are now ready to solve for the equilibrium in the static model. Suppose that the economy was previously in steady state at time $t-1$, with $i_{t-1} = i_{t-1}^*$. Given the Fed’s adjustment rule, the funds rate at time t satisfies:

$$i_t = i_{t-1} + k(i_t^* - i_{t-1}) + u_t = i_{t-1} + k\varepsilon_t + u_t. \quad (5)$$

Based on its conjecture about the Fed’s adjustment rule in Eq. (4), the market tries to back out i_t^* from its observation of i_t . Since both shocks ε_t and u_t are normally distributed, the market’s expectation is given by:

$$E[i_t^* | i_t] = i_{t-1} + \frac{\kappa\tau_u \times (i_t - i_{t-1})}{\tau_\varepsilon + \kappa^2\tau_u} = i_{t-1} + \chi(i_t - i_{t-1}), \quad (6)$$

where $\chi = \frac{\kappa\tau_u}{\tau_\varepsilon + \kappa^2\tau_u}$.

The less noise there is in the Fed’s adjustment rule, the higher is τ_u and the more the market reacts to the change in the rate i_t .

In light of Eqs. (5) and (6) and the random-walk property that $i_t^\infty = E_t[i_t^*]$, the Fed’s loss function taking expectations over the realization of the noise can be written as:

¹² There is a close analogy to models where corporate managers with private information pump up their reported earnings in an effort to impress the stock market. See, e.g., Stein (1989).

$$L_t = E_{u_t} \left[(i_t^* - i_t)^2 + \theta (\Delta i_t^\infty)^2 \right] = (1-k)^2 \varepsilon_t^2 + \sigma_u^2 + \theta \chi^2 (k^2 \varepsilon_t^2 + \sigma_u^2). \quad (7)$$

The Fed then minimizes this loss function by picking the optimal value of k . Again, we emphasize that in doing so, it takes the market's conjectures about its behavior, and hence the parameter χ , as fixed. The first-order condition with respect to k then yields:

$$k = \frac{1}{1 + \theta \chi^2}. \quad (8)$$

In rational-expectations equilibrium, the market's conjecture turns out to be correct, so we have $\kappa = k$. Imposing this condition, we have that, in equilibrium, the Fed's adjustment rule satisfies:

$$k = \frac{(\tau_\varepsilon + k^2 \tau_u)^2}{(\tau_\varepsilon + k^2 \tau_u)^2 + \theta (k \tau_u)^2}. \quad (9)$$

It follows immediately from (9) that the Fed follows a partial-adjustment rule in any rational expectations equilibrium:

Proposition 1: In any rational expectations equilibrium, the Fed's adjustment to a change in its target rate is partial: we have $k < 1$ so long as $\theta > 0$.

B.1 Equilibrium with No Noise in Rate-Setting Process

To build some intuition for the Fed's behavior, let us begin by considering the simple limiting case in which there is no noise in the rate setting process: $\sigma_u^2 = 1/\tau_u = 0$. In this case, the market's inference of i_t^* is simply:

$$E[i_t^* | i_t] = i_{t-1} + \frac{(i_t - i_{t-1})}{\kappa} = i_{t-1} + \frac{k}{\kappa} \varepsilon_t, \quad (10)$$

and the Fed's loss function is:

$$L_t = (i_t^* - i_t)^2 + \theta (\Delta i_t^\infty)^2 = ((1-k) \varepsilon_t)^2 + \theta \left(\frac{k}{\kappa} \varepsilon_t \right)^2. \quad (11)$$

When the Fed considers lowering k , it trades off being further away from the optimal i_t^* against the fact that it believes it can reduce bond market volatility by moving more slowly. In rational-expectations equilibrium, Eq. (9) reduces to:

$$k = \frac{k^2}{k^2 + \theta}. \quad (12)$$

First note that when $\theta = 0$, the only solution to (12) is given by $k = 1$. When the Fed does not care about bond market volatility, it fully adjusts to its new private information ε_t . By contrast, when $\theta > 0$, there may be more than one solution to (12), but in any equilibrium it must be the case that $k < 1$, and the Fed only partially adjusts.

Moreover, when $\theta > 0$, it is always the case that $k = 0$ satisfies Eq. (12), and there is therefore one equilibrium in which *the Fed does not adjust the funds rate at all*. This somewhat unnatural outcome is a function of the extreme feedback that arises between the Fed's adjustment rule and the market's conjecture when there is no noise in the rate-setting process. Specifically, when the market conjectures that the Fed never moves the funds rate at all (i.e., when the market conjectures that $\kappa = 0$), then even a tiny out-of-equilibrium move by the Fed would lead to an infinitely large change in the infinite-horizon forward rate. Anticipating this possibility, the Fed validates the market's conjecture by not moving at all, i.e. by choosing $k = 0$. However, as soon as there is even an infinitesimal amount of noise u_t in the rate-setting process, this extreme $k = 0$ equilibrium is ruled out, as small changes in the funds rate now lead to bounded market reactions. This explains our motivation for keeping a tiny amount of rate-setting noise in the more general model.

In addition to $k = 0$, for $0 < \theta < 0.25$ there are two other solutions to Eq. (12), given by:

$$k = \frac{1 \pm \sqrt{(1 - 4\theta)}}{2} \quad (13)$$

Of these, the larger of the two values also represents a stable equilibrium outcome. Thus, as long as θ is not too big, the model also admits a non-degenerate equilibrium, with $\frac{1}{2} < k < 1$, even with $\sigma_u^2 = 0$. Within this region, higher values of θ lead to lower values of k . In other words, the more intense the Fed's concern with bond-market volatility, the more gradually it adjusts the funds rate.

B.2 Equilibrium Outcomes across the Parameter Space

Now we return to the more general case where $\sigma_u^2 > 0$ and explore more fully the range of outcomes produced by the model for different parameter values. In each panel of Fig. 2 below, we

plot the Fed's best response k as a function of the market's conjecture κ . Any rational expectations equilibrium must lie on the 45-degree line where $k = \kappa$.

In Panel (a) of the figure, we begin with a relatively low value of θ , namely 0.12, and set $\sigma_\varepsilon^2 / \sigma_u^2 = 250$. As can be seen, this leads to a unique equilibrium with a high value of k , given by 0.86. When the Fed cares only a little bit about bond-market volatility, it engages in a modest amount of rate smoothing, because it does not want to deviate too far from its target rate i_t^* in its efforts to reduce volatility.

In Panel (b), we keep $\sigma_\varepsilon^2 / \sigma_u^2 = 250$ and increase θ to 0.40. In this high- θ case, there is again a unique equilibrium, but it now involves a very low value of k of just 0.04. Thus, when the Fed cares a lot about bond market volatility, the funds rate dramatically under-adjusts to new information.

In Panel (c), we set θ to an intermediate value of 0.22. Here, we have multiple equilibria: the Fed's best response crosses the 45-degree line in three places. Of these three crossing points, the two outer ones (at $k = 0.07$ and 0.69 respectively) correspond to stable equilibria, in the sense that if the market's initial conjecture κ takes on an out-of-equilibrium value, the Fed's best response to that conjecture will tend to drive the outcome towards one of these two extreme crossing points.

The existence of multiple equilibria highlights an essential feature of the model: the potential for market beliefs about Fed behavior to become self-validating. If the market conjectures that the Fed will adjust rates only very gradually, then even small changes are heavily freighted with informational content about the Fed's reaction function. Given this strong market sensitivity and its desire not to create too much volatility, the Fed may then choose to move very carefully, even at the cost of accepting a funds rate that is quite distant from its target. Conversely, if the market conjectures that the Fed is more aggressive in its adjustment of rates, it reacts less to any given movement, which frees the Fed up to track the target rate more closely.

When there are multiple equilibria, it is typically the case that the one with the higher value of k leads to a better outcome from the perspective of the Fed's objective function: with a higher k , i_t is closer to i_t^* , and yet in equilibrium, bond-market volatility is no greater. This suggests that if we are in a range of the parameter space where multiple equilibria are a possibility, it is important for the Fed to avoid doing anything in its communications that tends to lead the market to conjecture a low value of κ . Even keeping its own preferences fixed, fostering an impression of strong gradualism

among market participants can lead to an undesirable outcome in which the Fed gets stuck in the low- k equilibrium. We will return to this point in more detail shortly, when we discuss the implications of forward guidance in our model.

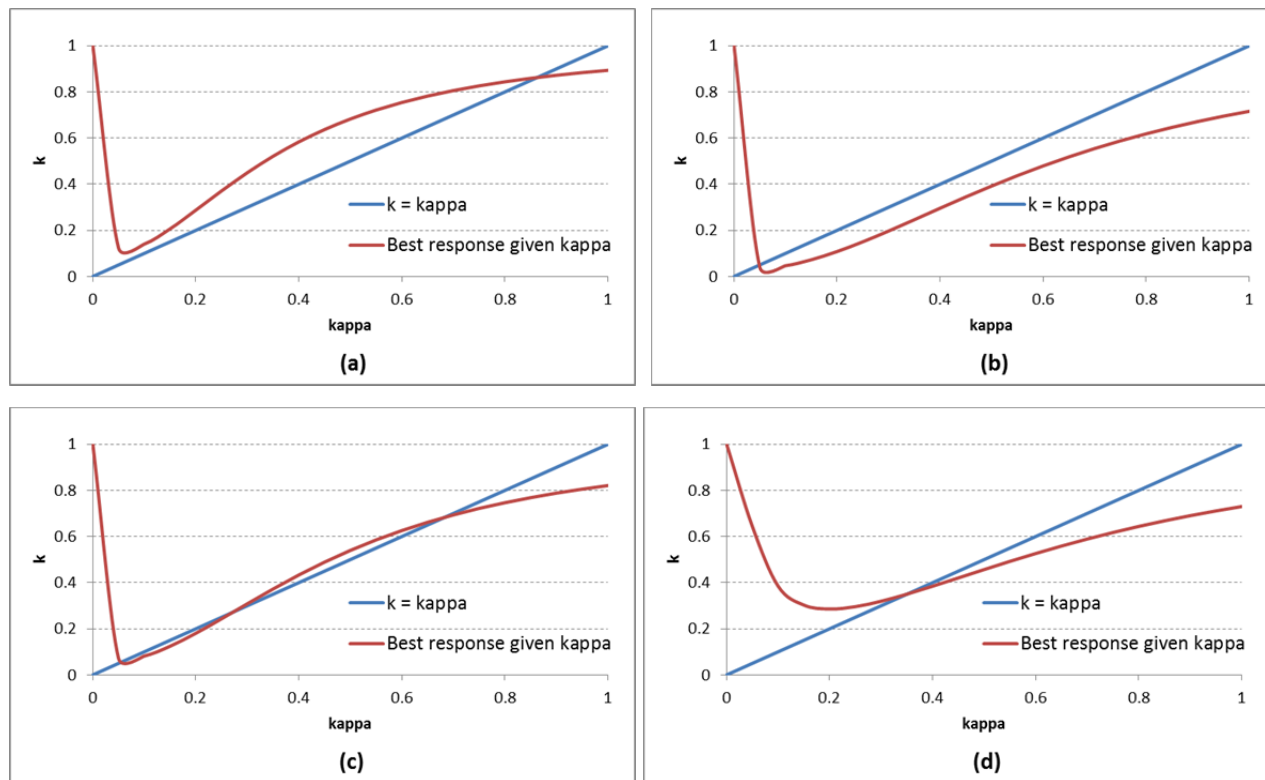


Figure 2. Behavior of the equilibrium for different parameter values. Panel (a): $\theta = 0.12$, $\sigma_\varepsilon^2 / \sigma_u^2 = 250$. Panel (b): $\theta = 0.4$, $\sigma_\varepsilon^2 / \sigma_u^2 = 250$. Panel (c): $\theta = 0.22$, $\sigma_\varepsilon^2 / \sigma_u^2 = 250$. Panel (d): $\theta = 0.4$, $\sigma_\varepsilon^2 / \sigma_u^2 = 25$.

As noted earlier, the very small but non-zero value of σ_u^2 plays an important role in pinning down the “low- k ” equilibria illustrated in Panels (b) and (c)—those where we have $k < 0.5$. If instead we were to set $\sigma_u^2 = 0$, these equilibria would degenerate to $k = 0$, for the reasons described above. By contrast, the “high- k ” equilibrium depicted in Panel (a) as well as the upper equilibrium in Panel (c) are much less sensitive to the choice of σ_u^2 . Thus, loosely speaking, the value of σ_u^2 really matters only when θ is sufficiently high that a low- k equilibrium exists.

Panel (d) of Fig. 2 illustrates this point. We maintain θ at 0.40 as in Panel (b), but now set the ratio $\sigma_\varepsilon^2 / \sigma_u^2$ to 25 instead of 250. This leads to an increase in the (unique) equilibrium value of k from 0.04 to 0.35. Intuitively, with more noise in the rate-setting process, the market reacts less

sensitively to a given change in rates, which allows the Fed to adjust rates more elastically in response to changes in its target.

C. The Time-Consistency Problem

A central property of our model is that there is a time-consistency problem: the Fed would do better in terms of optimizing its own objective function if it were able to commit to behaving as if it had a lower value of θ than it actually does. This is because while the Fed is always tempted to move the funds rate gradually so as to reduce bond-market volatility, this desire is frustrated in equilibrium: the more gradually the Fed acts, the more the market responds to any given change in rates. Thus, all that is left is a funds rate that is further from target on average. This point comes through most transparently if we consider the limiting case in which there is no noise in the Fed's rate-setting process. In this case, once we impose the rational-expectations assumption that $\kappa = k$, the Fed's attempts at smoothing have no effect at all on bond-market volatility in equilibrium:

$$\left(\Delta i_t^\infty\right)^2 = \left(\frac{k}{\kappa} \varepsilon_t\right)^2 = \varepsilon_t^2. \quad (14)$$

Thus, the value of the Fed's loss function is

$$L_t = \left(i_t^* - i_t\right)^2 + \theta \left(\Delta i_t^\infty\right)^2 = \left((1-k) \varepsilon_t\right)^2 + \theta \varepsilon_t^2, \quad (15)$$

which is decreasing in k for $k < 1$.

To the extent that the target rate i_t^* is non-verifiable private information, it is hard to think of a contracting technology that can readily implement the first-best outcome under commitment: how does one write an enforceable rule that says that the Fed must always react fully to its private information? Thus, discretionary monetary policy will inevitably be associated with some degree of inefficiency. However, even in the absence of a binding commitment technology, there may still be scope for improving on the fully discretionary outcome. One possible approach follows in the spirit of Rogoff (1985), who argues that society should appoint a central banker who is more hawkish on inflation than is society itself. The analogy in the current context is that society should aim to appoint a central banker who cares less about financial-market volatility (i.e., has a lower value of θ) than society as a whole. Or put differently, society—and the central bank itself—should seek to foster an institutional culture and set of norms that discourages the monetary policy committee from being overly attentive to market-volatility considerations.

To see this, consider the problem of a social planner choosing a central banker whose concern about financial market volatility is given by θ_c . This central banker will implement the rational expectations adjustment rule $k(\theta_c)$, where k is given by Eq. (12), replacing θ with θ_c . The planner's ex ante problem is then to pick θ_c to minimize its ex ante loss, recognizing that its own concern about financial-market volatility is given by θ :

$$E_{\varepsilon_t} \left[(i_t^* - i_t)^2 + \theta (\Delta i_t^\infty)^2 \right] = E \left[\left((1 - k(\theta_c)) \varepsilon_t \right)^2 + \theta \varepsilon_t^2 \right] = \left((1 - k(\theta_c))^2 + \theta \right) \sigma_\varepsilon^2. \quad (16)$$

Since $k(\theta_c)$ is decreasing in θ_c , the ex ante loss is minimized by picking $\theta_c = 0$, as the following proposition states.¹³

Proposition 2: In the absence of rate-setting noise, it is ex ante optimal to appoint a central banker with $\theta_c = 0$ so that $k(\theta_c) = 1$.

The time-consistency problem is especially stark in our setting because the Fed is assumed to care about the volatility of the infinite-horizon rate. As can be seen in (14), this volatility is a fixed constant in any rational-expectations equilibrium, independent of the degree of gradualism. Thus, the Fed's efforts to control the volatility of the infinite-horizon rate are completely unsuccessful in equilibrium, which is why it is ex ante optimal to appoint a central banker who does not care at all about this infinite-horizon volatility. By contrast, if we instead assume that the Fed cares about the volatility of a finite-horizon rate (e.g., the 10-year rate), there is still a time-consistency problem, but it is attenuated. If the expectations hypothesis holds, then the finite-horizon rate is given by the expected path of the funds rate over that finite horizon. By moving gradually, the Fed can actually reduce the volatility of the realized path. However, it still cannot fool the market about its ultimate destination.

¹³ In the alternative case where rate-setting noise is small but non-zero, a mild generalization of Proposition 2 obtains. While it is no longer true that the Fed would like to commit to behaving as if θ was exactly equal to zero, it would still like to commit to behaving as if θ was very close to zero and much smaller than its actual value. For example, for the parameter values in Panel D of Fig. 2, where the Fed's actual $\theta = 0.4$, it would minimize its loss function if it were able to commit to $\theta = 0.04$. In the appendix, we provide a fuller treatment of the optimal solution under commitment in the case of non-zero noise.

One way to develop this intuition in our setting is to note that any finite-horizon rate can be approximated by a weighted average of the current funds rate i_t and the infinite-horizon forward rate i_t^∞ , as long as we pick the weights correctly. This follows from the fact that there is only one factor, namely ε_t , in our model of the term structure of interest rates. Using this observation, it is straightforward to establish the following proposition, which is proven in the appendix.

Proposition 3: Suppose that $\theta > 0$ and that the Fed cares about the volatility of a finite-horizon rate, which can be represented as a weighted average of the short rate and the infinite-horizon rate. Then, it is ex ante optimal to appoint a central banker with a strictly positive value of θ_c , but to have $\theta_c < \theta$.

Under discretion, the Fed has two motives when it moves gradually. First, it aims to reduce the volatility of the component of the finite-horizon rate that is related to the current short rate i_t . Second, it hopes to reduce the volatility of the component of the finite-horizon rate that is related to the infinite-horizon forward rate i_t^∞ by fooling the market about its private information ε_t . The first goal can be successfully attained in rational-expectations equilibrium, but the second cannot—as in the baseline model, the Fed cannot fool the market about its private information in equilibrium.

Under commitment, only one of these motives for moving gradually remains, and thus it is less appealing to move gradually than it would be under discretion. Thus, if society is appointing a central banker whose concern about financial market volatility is given by θ_c , it would like to appoint one with $\theta_c < \theta$. In other words, there is still a time-consistency problem. However, unlike the case where the social planner cares about the volatility of the infinite-horizon forward rate, here the planner no longer wants to go to the extreme where $\theta_c = 0$.

D. Does Forward Guidance Help or Hurt?

Because the Fed's target rate i_t^* is non-verifiable private information, it is impossible to write a contract that effectively commits the Fed to rapid adjustment of the funds rate in the direction of i_t^* . So to the extent that the problem has not already been solved by appointing a central banker with the appropriate disregard for bond-market volatility, the first-best is generally unattainable. But this raises the question of whether there are other more easily enforced contracts that might be of some

help in addressing the time-consistency problem. In this section, we ask whether something along the lines of forward guidance could fill this role. By “forward guidance”, we mean an arrangement whereby the Fed announces at time t a value of the funds rate that it expects to prevail at some point in the future. Since both the announcement itself and the future realization of the funds rate are both publicly observable, market participants can easily tell ex post whether the Fed has honored its guidance, and one might imagine that it will suffer a reputational penalty if it fails to do so. In this sense, a “guidance contract” is more enforceable than a “don’t-smooth-with-respect-to-private-information contract.”

It is well understood that a semi-binding commitment to adhere to a pre-announced path for the short rate can be of value when the short rate is at the zero lower bound (ZLB); see e.g., Krugman (1998), Eggertsson and Woodford (2003), and Woodford (2012). We do not take issue with this observation here. Rather, we ask a different question, namely whether guidance can also be useful away from the ZLB, as a device for addressing the central bank’s tendency to adjust rates too gradually. It turns out the answer is no. Forward guidance can never help with the time-consistency problem we have identified, and attempts to use guidance away from the ZLB can potentially make matters strictly worse.

We develop the argument in two steps, working backwards. First, we assume that it is time t , and the Fed comes into the period having already announced a specific future rate path at time $t-1$. We then ask how the existence of this guidance influences the smoothing incentives analyzed above. Having done so, we then fold back to time $t-1$ and ask what the optimal guidance announcement looks like and whether the existence of the guidance technology increases or decreases the Fed’s expected utility from an ex ante perspective.

To be more precise, suppose we are in steady state at time $t-1$, and suppose again that the Fed cares about the volatility of the infinite horizon forward rate. We allow the Fed to publicly announce a value i_{t-1}^f as its guidance regarding i_t , the funds rate that will prevail one period later, at time t . To make the guidance partially credible, we assume that once it has been announced, the Fed bears a reputational cost of deviating from the guidance so that its augmented loss function at time t is now given by:

$$L_t = (i_t^* - i_t)^2 + \theta(\Delta i_t^\infty)^2 + \gamma(i_{t-1}^f - i_t)^2. \quad (17)$$

When it arrives at time t , the Fed takes the reputational penalty γ as exogenously fixed, but when we fold back to analyze guidance from an ex-ante perspective, it is more natural to think of γ as a choice variable for the Fed at time $t-1$. That is, the more emphatically it talks about its guidance ex ante or the more reputational chips it puts on the table, the greater will be the penalty if it deviates from the guidance ex post.

We start at time t and take the forward guidance i_{t-1}^f as given. To keep things simple, we assume that there is no noise in the Fed's adjustment rule ($\sigma_u^2 = 0$). Recall from Eq. (13) above that, in this case, there only exists a stable non-degenerate (i.e. positive k) equilibrium when $\theta < 1/4$, a condition that we assume to be satisfied in what follows. Moreover, in such an equilibrium, we have $1/2 < k < 1$.

Given the more complicated nature of the Fed's objective function, its optimal choice as to how much to adjust the short rate no longer depends on just its private information ε_t ; it also depends on the previously-established forward guidance i_{t-1}^f . To allow for a general treatment, we posit that this adjustment can be described by some potentially non-linear function of the two variables:

$$i_t = i_{t-1} + f(\varepsilon_t; i_{t-1}^f). \quad (18)$$

Moreover, the market conjectures that the Fed is following a potentially non-linear rule:

$$i_t = i_{t-1} + \phi(\varepsilon_t; i_{t-1}^f). \quad (19)$$

As before, given its conjecture about the Fed's adjustment rule, the market tries to back out i_t^* given i_t . Thus, the market's estimate of i_t^* satisfies

$$E[i_t^* | i_t] = i_{t-1} + \phi^{-1}(i_t - i_{t-1}; i_{t-1}^f) = i_{t-1} + \phi^{-1}(f(\varepsilon_t, i_{t-1}^f); i_{t-1}^f). \quad (20)$$

Since $i_t^\infty = E_t[i_t^*]$, the Fed's loss function can be written as:

$$\begin{aligned} & (i_t^* - i_t)^2 + \theta(\Delta i_t^\infty)^2 + \gamma(i_{t-1}^f - i_t)^2 \\ & = (\varepsilon_t - f(\varepsilon_t; i_{t-1}^f))^2 + \theta(\phi^{-1}(f(\varepsilon_t; i_{t-1}^f); i_{t-1}^f))^2 + \gamma(i_{t-1}^f - i_{t-1} - f(\varepsilon_t; i_{t-1}^f))^2. \end{aligned} \quad (21)$$

The Fed minimizes this loss function by choosing over the set of possible functions $f(\cdot; \cdot)$, taking as fixed the market's conjecture about its behavior $\phi(\cdot; \cdot)$. In the appendix, we use a calculus-of-

variations type of argument to establish that, in a rational-expectations equilibrium in which $f(\cdot; \cdot) = \phi(\cdot; \cdot)$, the Fed's adjustment behavior is given by:

$$i_t = i_{t-1} + k\varepsilon_t + \frac{\gamma}{1+\gamma}(i_{t-1}^f - i_{t-1}), \quad (22)$$

where the partial-adjustment coefficient k now satisfies:

$$k = (1+\gamma)k^2 + \theta. \quad (23)$$

Two features of the Fed's modified adjustment rule are worth noting. First, as can be seen in (22), the Fed moves i_t more strongly in the direction of its previously-announced guidance i_{t-1}^f as γ increases, which is intuitive. Second, as (23) indicates, the presence of guidance also alters the way that the Fed incorporates its new time- t private information into the funds rate i_t . Given that any equilibrium necessarily involves $k > 1/2$, we can show that $\partial k / \partial \gamma < 0$. In other words, the more the Fed cares about deviating from its forward guidance, the less it responds to its new private information ε_t . In this sense, the presence of semi-binding forward guidance has an effect similar to that of increasing θ : both promote gradualism with respect to the incorporation of new private information. The reason is straightforward: by definition, when guidance is set at time $t-1$, the Fed does not yet know the realization of ε_t . So ε_t cannot be impounded in the $t-1$ guidance, and anything that encourages the time- t rate to hew closely to the $t-1$ guidance must also discourage the incorporation of ε_t into the time- t rate. This is the key insight for why guidance is not helpful away from the ZLB.

To make this point more formally, let us now fold back to time $t-1$ and ask two questions. First, assuming that $\gamma > 0$, what is the optimal guidance announcement i_{t-1}^f for the Fed to make? And second, supposing that the Fed can choose the intensity of its guidance γ —i.e., it can choose how many reputational chips to put on the table when it makes the announcement—what is the optimal value of γ ? In the appendix, we demonstrate the following:

Proposition 4: *Suppose we are in steady state at time $t-1$, with $i_{t-1} = i_{t-1}^*$. If the Fed takes $\gamma > 0$ as fixed, its optimal choice of forward guidance at time $t-1$ is to announce $i_{t-1}^f = i_{t-1}$. Moreover, given*

this announcement policy, the value of the Fed's objective function is decreasing in γ . Thus, if it can choose, it is better off foregoing guidance altogether, i.e. setting $\gamma = 0$.

The first part of the proposition reflects the fact that because the Fed is already at its target at time $t-1$ and has no information about how that target will further evolve at time t , the best it can do is to set i_{t-1}^f at its current value of i_{t-1} . This guidance carries no incremental information to market participants at $t-1$. Nevertheless, even though it is uninformative, the guidance has an effect on rate-setting at time t . As Eq. (23) highlights, a desire to not renege on its guidance leads the Fed to underreact by more to its new time- t private information. Since this is strictly undesirable, the Fed is better off not putting guidance in place to begin with.

This negative result about the value of forward guidance in Proposition 4 should be qualified in two ways. First, as we have already emphasized, this result only speaks to the desirability of guidance away from the ZLB; there is nothing here that contradicts Krugman (1998), Eggertsson and Woodford (2003), and Woodford (2012), who make the case for using a relatively strong form of guidance (i.e., with a significantly positive value of γ) when the economy is stuck at the ZLB.

Second, and more subtly, an advocate of using forward guidance away from the ZLB might argue that if it is done with a light touch, it can help to communicate the Fed's private information to the market, without reducing the Fed's future room to maneuver. In particular, if the Fed is careful to make it clear that its guidance embodies no attempt at a binding commitment whatsoever (i.e., that $\gamma = 0$), the guidance may serve as a purely informational device and cannot hurt.

The idea that guidance can play a useful informational role strikes us as perfectly reasonable, even though it does not emerge in our model. The Fed's private information here is unidimensional, and because it is already fully revealed in equilibrium, there is no further information for guidance to transmit. However, it is quite plausible that, in a richer model, things would be different, and guidance could be incrementally informative. On the other hand, our model also suggests that care should be taken when claiming that light-touch ($\gamma = 0$) guidance has no negative effects in terms of increasing the equilibrium degree of gradualism in the funds rate.

This point is most easily seen by considering the region of the parameter space where there are multiple equilibria. In this region, anything that influences market beliefs can have self-fulfilling effects. So, for example, suppose that the Fed puts in place a policy of purely informational forward

guidance, and FOMC members unanimously agree that the guidance does not in any way represent a commitment on their parts—that is, they all plan to behave as if $\gamma = 0$ and indeed manage to follow through on this plan *ex post*. Nevertheless, if some market participants think that guidance comes with a degree of commitment ($\gamma > 0$), there can still be a problem. These market participants expect smaller deviations from the announced path of rates, and if we are in the multiple-equilibrium region, this expectation can be self-validating. Thus, the mistaken belief that guidance entails commitment can lead to an undesirable outcome where the Fed winds up moving more gradually in equilibrium, thereby lowering its utility. Again, this effect arises even if all FOMC members actually behave as if $\gamma = 0$ and attach no weight to keeping rates in line with previously-issued guidance. Thus, at least in the narrow context of our model, there would appear to be some potential downside associated with even the mildest forms of forward guidance once the economy is away from the ZLB.

E. Term Premium Shocks

We next enrich the model in another direction, to consider how the Fed behaves when financial-market conditions are not purely a function of the expected path of interest rates. Specifically, we relax the assumption that the expectations hypothesis holds and assume that the infinite-horizon forward rate consists of both the expected future short rate and an exogenous term premium component r_t^* :

$$i_t^\infty = E_t [i_t^*] + r_t^* \quad (24)$$

The term premium is assumed to be common information, observed simultaneously by market participants and the Fed. We allow the term premium to follow an arbitrary process and let η_t denote the innovation in the term premium:

$$\eta_t = r_t^* - E_{t-1} [r_t^*]. \quad (25)$$

The solution method in this case is similar to that used for forward guidance in Section III.D. Specifically, we again assume that there is no noise in the Fed's rate-setting rule ($\sigma_u^2 = 0$) but that the rule can be an arbitrary non-linear function of both the new private information ε_t that the Fed learns at time t as well as the term premium shock η_t , which is publicly observable:

$$i_t = i_{t-1} + f(\varepsilon_t; \eta_t). \quad (26)$$

The market again conjectures that the Fed is following a non-linear rule:

$$i_t = i_{t-1} + \phi(\varepsilon_t; \eta_t). \quad (27)$$

As before, given its conjecture about the Fed's adjustment rule, the market tries to back out i_t^* . Thus, the market's assessment of i_t^* given i_t satisfies:

$$E[i_t^* | i_t] = i_{t-1} + \phi^{-1}(f(\varepsilon_t; \eta_t); \eta_t). \quad (28)$$

Thus, the Fed picks the adjustment function $f(\cdot; \cdot)$ to minimize:

$$(i_t^* - i_t)^2 + \theta(\Delta i_t^\infty)^2 = (\varepsilon_t - f(\varepsilon_t; \eta_t))^2 + \theta(\phi^{-1}(f(\varepsilon_t; \eta_t); \eta_t) + \eta_t)^2. \quad (29)$$

In the appendix, we show that the solution to (29), combined with the rational-expectations condition that $f(\cdot; \cdot) = \phi(\cdot; \cdot)$, leads to a Fed adjustment rule given by:

$$i_t = i_{t-1} + k_\varepsilon \varepsilon_t + k_\eta \eta_t, \quad (30)$$

where

$$\begin{aligned} k_\varepsilon &= k_\varepsilon^2 + \theta & \text{and} \\ k_\eta &= -\frac{\theta}{k_\varepsilon}. \end{aligned} \quad (31)$$

The following proposition summarizes the key properties of the equilibrium.

Proposition 5: The Fed acts to offset term premium shocks, lowering the funds rate when the term premium shock is positive and raising it when the term premium shock is negative. It does so even though term premium shocks are publically observable and in spite of the fact that its efforts to reduce the volatility associated with term premium shocks are fruitless in equilibrium. As the Fed's concern with bond-market volatility θ increases, k_ε falls and k_η increases in absolute magnitude. Thus, when it cares more about the bond market, the Fed reacts more gradually to changes in its private information about its target rate, but more aggressively to changes in term premiums.

A first observation about the equilibrium is that the Fed's response to its own private information, k_ε , is exactly the same as it was in the absence of term premium shocks (assuming no rate-setting noise) as derived in Section III.B.1. One might then be tempted to think that nothing

changes when we introduce term premium shocks, because they are publicly observable. Proposition 5 shows that, strikingly, this is not the case. Short rates do in fact respond to movements in the term premium.

What explains this result? Essentially, when the term premium spikes up, the Fed is unhappy about the prospective increase in the volatility of long rates. So even if its private information about i_t^* has in fact not changed, it would like to make the market think it has become more dovish so as to offset the rise in the term premium. Therefore, it cuts the short rate in an effort to create this impression. Again, in equilibrium, this attempt to fool the market is not successful, but taking the market's conjectures at any point in time as fixed, the Fed is always tempted to try.

One way to see why the equilibrium must involve the Fed reacting to term premium shocks is to think about what happens if we try to sustain an equilibrium where it does not—that is, if we try to sustain an equilibrium in which $k_\eta = 0$. In such a hypothetical equilibrium, Eq. (30) tells us that when the market sees any movement in the funds rate, it attributes that movement entirely to changes in the Fed's private information ε_t about its target rate. But if this is the case, then the Fed can indeed offset movements in term premiums by changing the short rate, thereby contradicting the assumption that $k_\eta = 0$. Hence, $k_\eta = 0$ cannot be an equilibrium.

A second key feature of the equilibrium is that the absolute magnitude of k_η becomes larger as θ rises and as k_ε becomes smaller: when it cares more about bond-market volatility, the Fed's responsiveness to term premium shocks becomes more aggressive even as its adjustment to new private information becomes more gradual. In particular, because we are restricting ourselves to the region of the parameter space where the simple no-noise model yields a non-degenerate equilibrium for k_ε , this means that we must have $0 < \theta < 1/4$ from Eq. (13) above. As θ moves from the lower to the upper end of this range, k_ε declines monotonically from 1 to $1/2$, and k_η increases in absolute magnitude from 0 to $-1/2$.

This latter part of the proposition is especially useful, as it yields a sharp testable empirical implication. As noted earlier, Campbell, Pflueger, and Viceira (2015) have shown that the Fed's behavior has become significantly more inertial in recent years. If we think of this change as reflecting a lower value of k_ε , we might be tempted to use the logic of the model to claim that the

lower k_ε is the result of the Fed placing increasing weight over time on the bond market, i.e. having a higher value of θ than it used to. While the evidence on financial-market mentions in the FOMC transcripts that we plotted in Figure 1 is loosely consistent with this hypothesis, it is obviously far from being a decisive test. However with Proposition 4 in hand, if we want to attribute a decline in k_ε to an increase in θ , then we also have to confront the additional prediction that we ought to observe the Fed responding more forcefully over time to term premium shocks. That is, the absolute value of k_η must have gone up. If this is not the case, it would represent a rejection of the hypothesis.

Our earlier results about the value of commitment generalize straightforwardly to the case where there are term premium shocks. Consider, as before, the problem of a social planner appointing a central banker whose concern about financial market volatility is given by θ_c . This central banker will implement the adjustment rules characterized by $k_\varepsilon(\theta_c)$ and $k_\eta(\theta_c)$, which are given by Eq. (31), replacing θ with θ_c . Thus, the planner's ex ante problem is to pick θ_c to minimize:

$$E\left[\left(i_t^* - i_t\right)^2 + \theta\left(\Delta i_t^\infty\right)^2\right] = E\left[\left(\left(1 - k_\varepsilon(\theta_c)\right)\varepsilon_t - k_\eta(\theta_c)\eta_t\right)^2 + \theta\left(\varepsilon_t + \eta_t\right)^2\right]. \quad (32)$$

As can be seen from (32), the planner's optimum is attained when $k_\varepsilon(\theta_c) = 1$ and $k_\eta(\theta_c) = 0$, which, according to (31), can be implemented by setting $\theta_c = 0$. Thus we have:

Proposition 6: In the extended version of the model with term premium shocks, it is again optimal to appoint a central banker with $\theta_c = 0$.

As before, in the rational-expectations equilibrium with term premium shocks, the Fed cannot influence the volatility of the infinite horizon-forward rate. If it attempts to do so, it realizes no benefit and just ends up with a funds rate that is further from target on average. Thus it is again optimal to commit to not trying to manage volatility by setting $\theta_c = 0$.

IV. Dynamic Model

In the static model considered to this point, the phenomenon of “gradualism” is really nothing more than under-adjustment of the policy rate to a one-time shock to the Fed’s target. We now introduce a dynamic version of the model, which allows us to fully trace out how a shock to the Fed’s target works its way into the policy rate over time. This dynamic analysis generates predictions that can be more directly compared to the empirical evidence on gradualism.

A. The Fed’s Adjustment Rule

To study dynamics, we need to consider the Fed’s behavior when the economy is not in steady state at time $t-1$. Suppose we enter period t with a pre-existing gap between the time $t-1$ target rate and the time $t-1$ federal funds rate of $X_{t-1} \equiv i_{t-1}^* - i_{t-1}$. We allow the Fed to have a more general adjustment rule of the following form:

$$i_t = i_{t-1} + k_X X_{t-1} + k_\varepsilon \varepsilon_t. \quad (33)$$

In other words, we allow the Fed to adjust differentially to the pre-existing gap X_{t-1} and to the innovation to its target ε_t .

B. The Case of Fully Rational Investors

The first case to consider is one in which—as in the static model above—all investors are fully rational. To begin, note that if there is no rate-setting noise ($\sigma_u^2 = 0$), in a rational-expectations equilibrium, the gap X_{t-1} is known by investors as of time t . The gap is a function of past realizations of the ε ’s, and as we have seen, these are fully revealed in a no-noise equilibrium. A rational investor conjectures that the Fed is following a rule of the form:

$$i_t = i_{t-1} + \kappa_X X_{t-1} + \kappa_\varepsilon \varepsilon_t. \quad (34)$$

Given this conjecture, the rational investor’s estimate of the innovation to the Fed’s target ε_t , and hence his estimate of the innovation to the short rate that will prevail in the distant future, is given by:

$$E^R[\varepsilon_t] = \frac{(\Delta i_t - \kappa_X X_{t-1})}{\kappa_\varepsilon}. \quad (35)$$

This is similar to what we had in Eq. (10) in the static model, except that now the rational investor adjusts for the expected rate change of $\kappa_X X_{t-1}$ when making inferences about ε_t . In other words, when the funds rate goes up at time t , the rational investor understands that it is not necessarily due

entirely to a contemporaneous upwards revision to the Fed's target at time t . Rather, part of the change in the funds rate may reflect the fact that the Fed is playing catch-up: it could be that the Fed has entered period t with a positive pre-existing gap between its target and the prior value of the funds rate and is working to close that gap.

It is straightforward to show that in this case, the Fed's equilibrium behavior can be characterized as follows.¹⁴

Proposition 7: *In the dynamic model with fully rational investors, we have that $k_X = 1$, and $k_\varepsilon < 1$, with k_ε (as before) being given by: $k_\varepsilon = \frac{k_\varepsilon^2}{k_\varepsilon^2 + \theta}$.*

Thus with fully rational investors, the Fed still under-adjusts to a contemporaneous shock to its target ε_t , in exactly the same way as before. However, the under-adjustment is short-lived: in the next period, it fully impounds what was left of the time- t innovation into the rate. In other words, $k_X = 1$ implies that all the Fed's time- t private information has made its way into the funds rate by time $t+1$. Intuitively, by time $t+1$, rational investors have already figured out all of the Fed's time- t private information. Given that the Fed can no longer fool investors about this old information, there is no motive to continue to incorporate it into rates slowly.

Proposition 7 makes clear that there can be a considerable distance between our earlier theoretical results on under-adjustment and the empirical evidence on gradualism. In the empirical literature, a typical finding (in quarterly data) would be that $i_t = 0.85 \cdot i_{t-1} + 0.15 \cdot i_t^*$, which implies that changes in the short rate are significantly positively correlated even when they are several quarters apart. In other words, there is relatively long-lasting momentum in short-rate changes. By contrast, the fact that that $k_X = 1$ in Proposition 7 implies that $\text{cov}(\Delta i_t, \Delta i_{t-j}) = 0$ for all $j > 1$: changes in the short rate more than one period apart are uncorrelated with one another. Thus, Proposition 7 would appear to be strongly counterfactual.

¹⁴ As we discuss further in the appendix, to simplify the problem, we also assume that at any time t , the Fed picks k_ε and k_X taking X_{t-1} as given, but before knowing the realization of ε_t . This timing convention is purely a technical trick that makes the problem more tractable without changing anything of economic substance.

C. Allowing for Some Naïve Investors

One way to bring the model more closely into alignment with the empirical evidence is to allow for a class of investors with a modest form of bounded rationality. To motivate the presence of these “naïve” investors, note that the updating performed by the fully rational investors, as described by Eq. (35), is highly sophisticated: they base their estimate of the innovation to the Fed’s target ε_t not only on the observed change in the funds rate Δi_t , but also on the gap X_{t-1} . Thus they need to keep track of X_{t-1} separately, which in turn requires making a series of inferences based on past changes in the funds rate.

By contrast, we assume that while naïve investors also recognize that Δi_t is informative about the innovation to the Fed’s target ε_t , they are not sophisticated enough to simultaneously condition on the pre-existing gap X_{t-1} . Instead, naïve investors conjecture that the Fed follows the simple adjustment rule

$$i_t = i_{t-1} + \kappa_N \varepsilon_t. \quad (36)$$

Thus, naïve investors act as though the change in the Fed’s target, and hence the change in the short rate that will prevail in the distant future, has a simple univariate relationship with the change in the short rate Δi_t :

$$E^N[\varepsilon_t] = \frac{i_t - i_{t-1}}{\kappa_N} = \beta \cdot \Delta i_t, \quad (37)$$

where β is a fixed constant. The beliefs of the naïve investors are similar to those of the rational investors in the static version of the model, as described in Eq. (10): they treat the current change in the short rate as a sufficient statistic for the expected change in the Fed’s target. However, unlike the rational investors in the dynamic version, whose beliefs are given by (35), the naïve investors fail to adjust for X_{t-1} .

We assume that a fraction λ of investors are rational, and the remaining fraction $(1 - \lambda)$ are naïve, so that the change in the market-determined infinite-horizon forward rate is the weighted average of the change in their respective expectations for the short rate that will prevail in the distant future:

$$\Delta i_t^\infty = \lambda \cdot E^R[\varepsilon_t] + (1 - \lambda) \cdot E^N[\varepsilon_t] \quad (38)$$

As before, the equilibrium involves the Fed picking k_X and k_ε to minimize its loss function holding fixed the behavior of naïve investors (β) and the conjectures κ_X and κ_ε of rational investors. In the rational expectations equilibrium, the conjectures of the rational investors will be validated so we have $\kappa_X = k_X$ and $\kappa_\varepsilon = k_\varepsilon$. We further assume that naïve investors' estimate of β is optimal, given the Fed's equilibrium behavior and given their own failure to condition on X_{t-1} . In particular, we assume that β is pinned down as the regression coefficient of ε_t on the change in the short rate Δi_t :

$$\beta = \frac{\text{cov}(\varepsilon_t, \Delta i_t)}{\text{var}(\Delta i_t)}. \quad (39)$$

Hence our naïve investors are actually quite smart, apart from their mistake of ignoring X_{t-1} . However, it should be noted that our qualitative results below do not depend on the way the regression coefficient β is chosen—they emerge even if it is not chosen optimally.

With these assumptions in place, we are able to establish the following proposition:

Proposition 8: *In the dynamic model with both fully rational and naïve investors, we have $k_\varepsilon < k_X < 1$. As the fraction of naïve investors $(1 - \lambda)$ increases, k_X declines and approaches k_ε .*

The basic intuition behind the proposition is simple. Naïve investors indiscriminately treat *any change* in the funds rate—whether it is driven by ε_t or by X_{t-1} —as informative about changes in future expected short rates, and hence as relevant for pricing long-maturity bonds. Thus, the Fed is now motivated to move gradually in incorporating both ε_t and X_{t-1} into the current funds rate. In other words, in its efforts to manage market expectations, the Fed is driven to partially cater to the beliefs of the naïve investors. Since these investors do not discriminate between ε_t and X_{t-1} , there is a tendency for the Fed to treat these two variables similarly as well, that is, to push k_X closer to k_ε . This effect gets us away from the empirically unrealistic implication that $k_X = 1$.

Example 1: Suppose we have $\theta = 0.9$, and $\lambda = 0.1$. This leads to $k_\varepsilon = 0.260$ and to $k_X = 0.309$.

Example 2: Suppose we have $\theta = 0.3$, and $\lambda = 0.5$. This leads to $k_\varepsilon = 0.604$ and to $k_X = 0.776$.

The two examples illustrate the sorts of outcomes that obtain in the dynamic model with naïve investors. The first example is one in which the market is largely dominated by naïve investors, and the Fed has a relatively high value of θ , so it cares strongly about bond-market volatility. In this case, adjustment of the funds rate to both components ε_t and X_{t-1} is quite slow and very similar in magnitude. Thus, this example comes close to replicating the dynamic behavior of the funds rate that has been documented in the data. In the second example, the market is evenly balanced between rational and naïve investors, and θ is much lower. As might be expected, there is less gradualism overall, and there is a more noticeable difference between the speed with which the Fed impounds new (k_ε) versus old (k_X) innovations to its target into the funds rate. Nevertheless, both examples yield values of k_X that are well below unity, i.e., that differ substantially from the all-rational case.

Note that the time-consistency problem persists here. As we have set things up, both the rational and the naïve investors contribute to the time-consistency problem, because both have equilibrium beliefs that are endogenous to Fed behavior—that is, both react more strongly to a given change in the funds rate when the Fed moves more gradually. For the naïve investors, this feature is captured in the expression for β in Eq. (39). If we were instead to make the cruder assumption that β is a fixed constant independent of equilibrium Fed behavior, then the time-consistency problem would be driven entirely by the presence of the rational investors.

In sum, a dynamic version of our model seems capable of generating empirically realistic inertial behavior for the funds rate. However, doing so requires augmenting our basic private-information framework to include some boundedly rational investors. Specifically, we need some investors who recognize that movements in today’s funds rate are informative about future values of the funds rate, but who are not so sophisticated that their mapping between the two incorporates a time-varying estimate of the current gap between the Fed’s target and the prevailing funds rate.

V. Conclusion

Rather than restating our findings, we conclude with a brief caveat. We have argued that in our setting, it can be valuable for a central bank to develop an institutional culture and set of norms such that a concern with bond-market volatility does not play an outsized role in policy deliberations.

In other words, it can be useful for monetary policymakers to build a reputation for not caring too much about the bond market. This argument rests on a comparison of equilibrium outcomes with different values of the parameter θ_c , which measures the appointed central banker's concern about financial-market volatility. But crucially, an implicit assumption in making this comparison is that in any given equilibrium, the market has come to fully know the true value of θ_c . We have not addressed the more difficult question of out-of-equilibrium dynamics: how is it that the market learns about θ_c , from the central bank's observed behavior and other forms of communication?

For this reason, our model offers no guidance on the best way to make the transition to a less bond-market-sensitive equilibrium, one in which there is less inertia in the policy rate, and the market eventually comes to understand that θ_c has declined. At any point in time, taking market conjectures as fixed, a “cold turkey” approach that involves an unexpectedly sharp adjustment in the path of the policy rate (relative to prevailing expectations of gradualism) is likely to cause significant market volatility and perhaps some collateral costs for the real economy. To be clear, nothing in our analysis should be taken as advocating such an approach. In particular, we do not believe our model has much advice to give with respect to the near-term question of how rapidly the Fed should lift rates in the months following its liftoff from the zero lower bound. At this horizon, market conjectures regarding the equilibrium degree of gradualism are likely to be relatively fixed, and the costs of changing behavior suddenly could potentially be quite high. To the extent that an institution-building effort of the sort we have in mind is at all useful, it may be something that is better undertaken over a longer horizon or at a time when the economy is in a less fragile state.

References

- Barro, Robert J., and David B. Gordon, 1983. "Rules, Discretion and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics* 12, 101-121.
- Bernanke, Ben, 2004. "Gradualism," Remarks at an economics luncheon co-sponsored by the Federal Reserve Bank of San Francisco and the University of Washington, Seattle, Washington.
- Bernanke, Ben, and Alan Blinder, 1992. "The Federal Funds Rate and the Channels of Monetary Transmission," *American Economic Review* 82, 901-921.
- Brainard, William, 1967. "Uncertainty and the Effectiveness of Policy," *American Economic Review* 57, 411-425.
- Coibion, Olivier, and Yuriy Gorodnichenko, 2012. "Why Are Target Interest Rate Changes So Persistent?" *American Economic Journal: Macroeconomics* 4, 126-62.
- Campbell, John Y., Carolin Pflueger, and Luis Viceira, 2015. "Monetary Policy Drivers of Bond and Equity Risks," Harvard University working paper.
- Dell'Araccia, Giovanni, Luc Laeven, and Gustavo Suarez, 2013. "Bank Leverage and Monetary Policy's Risk-Taking Channel: Evidence from the United States," IMF working paper.
- Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, 2015. "The Deposits Channel of Monetary Policy," NYU working paper.
- Eggertsson, Gauti B. and Michael Woodford, 2003. "The Zero Bound on Interest Rates and Optimal Monetary Policy," *Brookings Papers on Economic Activity*, 139-211.
- Farhi, Emmanuel, and Jean Tirole, 2012. "Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts," *American Economic Review* 102, 60-93.
- Gertler, Mark, and Peter Karadi, 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics* 7, 44-76.
- Holmstrom, Bengt, 1999. "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies* 66, 169-182.
- Jimenez, Gabriel, Steven Ongena, Jose-Luis Peydro, and Jesus Saurina, 2014. "Hazardous Times for Monetary Policy: What do 23 Million Loans Say About the Impact of Monetary Policy on Credit Risk-Taking?," *Econometrica* 82, 463-505.
- Kashyap, Anil K, and Jeremy C. Stein, 2000. "What Do a Million Observations on Banks Say About the Transmission of Monetary Policy?," *American Economic Review* 90, 407-428.

- Krugman, Paul, 1998. "It's Baaack: Japan's Slump and the Return of the Liquidity Trap," *Brookings Papers on Economic Activity*, 137-2006.
- Kydland, Finn, and Edward Prescott, 1977. "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy* 85, 473-491.
- Rogoff, Kenneth, 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target," *Quarterly Journal of Economics* 100, 1169-1189.
- Rudebusch, Glenn D., 2002. "Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia," *Journal of Monetary Economics* 49, 1161-1187.
- Rudebusch, Glenn D., 2006. "Monetary Policy Inertia: Fact or Fiction?" *International Journal of Central Banking* 2, 85-135.
- Sack, Brian, 1998. "Uncertainty, Learning, and Gradual Monetary Policy," FEDS Working Paper Series, number 98-34.
- Stein, Jeremy C., 1989. "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior," *Quarterly Journal of Economics* 104, 655-669.
- Stein, Jeremy C., 2014. "Challenges for Monetary Policy Communication," Remarks at the Money Marketeers of New York University, New York, May 6.
- Taylor, John, 1993. "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* 39: 195-214.
- Woodford, Michael, 2003. "Optimal Interest-Rate Smoothing," *Review of Economic Studies* 70, 861-886.
- Woodford, Michael, 2012. "Methods of Policy Accommodation at the Interest-Rate Lower Bound," *The Changing Policy Landscape: 2012 Jackson Hole Symposium*, Federal Reserve Bank of Kansas City.

Appendix Proofs of Propositions

A.1 Generalization of Proposition 2 to case of non-zero rate-setting noise

Society's ex ante problem is to choose a central banker with concern about market volatility θ_c . In the absence of noise, this central banker will implement the rational-expectations equilibrium $k(\theta_c)$ given by Equation (12) in the main text replacing θ with θ_c . In the absence of noise, society's ex ante loss function is given by

$$\left((1 - k(\theta_c))^2 + \theta \right) \sigma_\varepsilon^2,$$

which is minimized by setting $\theta_c = 0$ so that $k(\theta_c) = 1$.

In the presence of noise, the appointed central banker will implement the rational-expectations equilibrium $k(\theta_c)$ given by Equation (9) in the main text replacing θ with θ_c . Differentiating $k(\theta_c)$ with respect to θ_c yields

$$\frac{\partial k}{\partial \theta_c} = - \left(1 - \frac{2\theta_c k \tau_u^2 (\tau_\varepsilon + k^2 \tau_u) (k^2 \tau_u - \tau_\varepsilon)}{\left((\tau_\varepsilon + k^2 \tau_u)^2 + \theta_c (k \tau_u)^2 \right)^2} \right)^{-1} \frac{(\tau_\varepsilon + k^2 \tau_u)^2 (k \tau_u)^2}{\left((\tau_\varepsilon + k^2 \tau_u)^2 + \theta_c (k \tau_u)^2 \right)^2} < 0.$$

In the presence of noise, society's ex ante loss function is given by

$$L = E \left[((1 - k) \varepsilon_t + u_t)^2 + \theta (\chi (k \varepsilon + u_t))^2 \right] = (1 - k)^2 \sigma_\varepsilon^2 + \sigma_u^2 + \theta \chi^2 (k^2 \sigma_\varepsilon^2 + \sigma_u^2).$$

Substituting in for the rational-expectations definition of χ and differentiating with respect to θ_c gives the first order condition

$$\frac{\partial L}{\partial \theta_c} = \left[-2(1 - k) \sigma_\varepsilon^2 + \frac{2k\theta\tau_u}{(\tau_\varepsilon + k^2\tau_u)^3} (k^2\tau_u + \tau_\varepsilon) \right] \frac{\partial k}{\partial \theta_c}.$$

At the optimum θ_c in the presence of noise, the term in square brackets will be zero. When we discuss the optimal θ_c in the presence of noise in footnote 13 of the main text, we numerically solve for the value of θ_c and therefore k that sets the term in square brackets to zero.

A.2 Proof of Proposition 3

We begin by characterizing the behavior of an appointed central banker who care about a finite-horizon yield, which can be represented as a weighted average of the short rate and the infinite-horizon rate:

$$i_t^{finite} = \alpha i_t + (1 - \alpha) i_t^\infty.$$

The degree of concern about the volatility of the finite-horizon yield is parameterized by θ_c . The central banker follows a rule of the form

$$i_t = i_{t-1} + k(\theta_c) \cdot \varepsilon_t,$$

while the market conjectures that the central banker is following a rule of the form

$$i_t = i_{t-1} + \kappa(\theta_c) \cdot \varepsilon_t.$$

Given this, the market's conjecture about ε_t is

$$\tilde{\varepsilon}_t = \frac{k}{\kappa} \varepsilon_t$$

and it sets the infinite-horizon forward rate to

$$i_t^\infty = i_{t-1} + \tilde{\varepsilon}_t.$$

Thus, the change in the finite-horizon yield is

$$\Delta i_t^{finite} = \alpha (i_t - i_{t-1}) + (1 - \alpha) \tilde{\varepsilon}_t.$$

The central banker picks $k(\theta_c)$ to minimize the loss function

$$\begin{aligned} E_t \left[(i_t^* - i_t)^2 + \theta_c \left(\Delta i_t^{finite} \right)^2 \right] &= (1 - k)^2 \sigma_\varepsilon^2 + \theta_c \alpha^2 k^2 \sigma_\varepsilon^2 \\ &\quad + \theta_c \frac{(1 - \alpha)^2}{\kappa^2} k^2 \sigma_\varepsilon^2 + 2\theta_c \alpha (1 - \alpha) \frac{k^2}{\kappa} \sigma_\varepsilon^2. \end{aligned}$$

Differentiating with respect to k yields the first order condition

$$0 = -(1 - k) \sigma_\varepsilon^2 + \theta_c \alpha^2 k \sigma_\varepsilon^2 + \theta_c \frac{(1 - \alpha)^2}{\kappa^2} k \sigma_\varepsilon^2 + 2\theta_c \alpha (1 - \alpha) \frac{k}{\kappa} \sigma_\varepsilon^2.$$

In rational expectations we have $\kappa = k$, so this reduces to

$$0 = -(1 - k) \sigma_\varepsilon^2 + \underbrace{\theta_c \alpha^2 k \sigma_\varepsilon^2}_A + \underbrace{\theta_c \frac{(1 - \alpha)^2}{k} \sigma_\varepsilon^2}_B + \underbrace{2\theta_c \alpha (1 - \alpha) \sigma_\varepsilon^2}_C. \quad (\text{A.1})$$

Term A reflects the direct effect of changing k on the short-rate component of our representation of the finite-horizon yield. Term B reflects the effect of changing k on the inferred shock $\tilde{\varepsilon}_t$, and term C reflects the interaction of the two effects. Note that (A.1) implies that

$$\frac{\partial k}{\partial \theta_c} = \left[\begin{array}{c} - \left(\sigma_\varepsilon^2 (1 + \theta_c \alpha^2 k) - \theta_c \frac{(1 - \alpha)^2}{k^2} \sigma_\varepsilon^2 \right)^{-1} \\ \times \left(\alpha^2 k \sigma_\varepsilon^2 + \frac{(1 - \alpha)^2}{k} \sigma_\varepsilon^2 + 2\theta_c \alpha (1 - \alpha) \sigma_\varepsilon^2 \right) \end{array} \right] < 0.$$

Society's ex ante loss is given by

$$E_t \left[(i_t^* - i_t)^2 + \theta \left(\Delta i_t^{finite} \right)^2 \right] = (1 - k(\theta_c))^2 \sigma_\varepsilon^2 + \theta \left(\begin{array}{c} \alpha^2 k(\theta_c)^2 \sigma_\varepsilon^2 + (1 - \alpha)^2 \sigma_\varepsilon^2 \\ + 2\alpha (1 - \alpha) k(\theta_c) \sigma_\varepsilon^2 \end{array} \right).$$

Differentiating with respect to θ_c yields the first order condition

$$\left[-(1 - k) \sigma_\varepsilon^2 + \theta \alpha k \sigma_\varepsilon^2 + \theta \alpha (1 - \alpha) \sigma_\varepsilon^2 \right] \frac{\partial k}{\partial \theta_c} = 0.$$

Evaluating at $\theta_c = \theta$ and using the appointed central banker's first order condition (A.1), we have that the term in the square brackets is

$$-(1 - k) \sigma_\varepsilon^2 + \underbrace{\theta \alpha k \sigma_\varepsilon^2}_A + \underbrace{\theta \alpha (1 - \alpha) \sigma_\varepsilon^2}_{C/2} = -\underbrace{\theta \frac{(1 - \alpha)^2}{k} \sigma_\varepsilon^2}_{-B} - \underbrace{\theta \alpha (1 - \alpha) \sigma_\varepsilon^2}_{-(C/2)} < 0.$$

Since $\partial k / \partial \theta_c < 0$, this implies that society's first order condition is positive at $\theta_c = \theta$ and its ex ante loss can be reduced by setting $\theta_c < \theta$.

A.3 Proof of Proposition 4

We begin by solving the time t problem, taking the forward guidance i_{t-1}^f as given. Given its conjecture about the rule the Fed is following, $\phi(\varepsilon_t; i_{t-1}^f)$, the market's conjecture about ε_t is

$$\tilde{\varepsilon}_t = \phi^{-1}\left(f\left(\varepsilon_t; i_{t-1}^f\right); i_{t-1}^f\right)$$

and the Fed's loss function is

$$\left(\varepsilon_t - f\left(\varepsilon_t; i_{t-1}^f\right)\right)^2 + \theta \left(\phi^{-1}\left(f\left(\varepsilon_t; i_{t-1}^f\right); i_{t-1}^f\right)\right)^2 + \gamma \left(i_{t-1}^f - i_{t-1} - f\left(\varepsilon_t; i_{t-1}^f\right)\right)^2.$$

Consider the effect on the value of the loss function of a small perturbation in the value of $f(\varepsilon_t; i_{t-1}^f)$, df . The effect of this perturbation is zero at the optimal $f(\cdot; \cdot)$ so we have

$$-\left(\varepsilon_t - f\left(\varepsilon_t; i_{t-1}^f\right)\right) + \theta \left(\phi^{-1}\left(f\left(\varepsilon_t; i_{t-1}^f\right); i_{t-1}^f\right)\right) \frac{\partial \phi^{-1}}{\partial i} \Big|_{i=f(\varepsilon_t; i_{t-1}^f)} - \gamma \left(i_{t-1}^f - i_{t-1} - f\left(\varepsilon_t; i_{t-1}^f\right)\right) = 0.$$

Since $\phi(\phi^{-1}(x)) = x$, we have

$$\frac{\partial \phi^{-1}}{\partial i} \Big|_{i=f(\varepsilon_t; i_{t-1}^f)} = \frac{1}{\frac{\partial \phi}{\partial i} \Big|_{\varepsilon=\phi^{-1}(f(\varepsilon_t; i_{t-1}^f); i_{t-1}^f)}}.$$

So the optimal $f(\cdot; \cdot)$ satisfies:

$$-\left(\varepsilon_t - f\left(\varepsilon_t; i_{t-1}^f\right)\right) + \theta \left(\phi^{-1}\left(f\left(\varepsilon_t; i_{t-1}^f\right); i_{t-1}^f\right)\right) \frac{1}{\frac{\partial \phi}{\partial i} \Big|_{\varepsilon=\phi^{-1}(f(\varepsilon_t; i_{t-1}^f); i_{t-1}^f)}} - \gamma \left(i_{t-1}^f - i_{t-1} - f\left(\varepsilon_t; i_{t-1}^f\right)\right) = 0.$$

Imposing rational expectations, we have $\phi = f$ so that this reduces to the differential equation

$$\frac{\partial f}{\partial \varepsilon} \Big|_{\varepsilon=\varepsilon_t} \left(\varepsilon_t - f\left(\varepsilon_t; i_{t-1}^f\right) + \gamma \left(i_{t-1}^f - i_{t-1} - f\left(\varepsilon_t; i_{t-1}^f\right)\right)\right) = \theta \varepsilon_t,$$

which the optimal $f(\cdot; \cdot)$ must satisfy.

Now conjecture that $f = k\varepsilon_t + c$. In this case the differential equation reduces to

$$k \left(\varepsilon_t - k\varepsilon_t - c + \gamma \left(i_{t-1}^f - i_{t-1} - k\varepsilon_t - c\right)\right) = \theta \varepsilon_t$$

or

$$k(1 - k - \gamma k) \varepsilon_t - kc + k\gamma \left(i_{t-1}^f - i_{t-1} - c\right) = \theta \varepsilon_t$$

Matching coefficients implies that

$$\begin{aligned} k(1 - k - \gamma k) &= \theta & \text{and} & & \text{(A.2)} \\ -kc + k\gamma \left(i_{t-1}^f - i_{t-1} - c\right) &= 0. \end{aligned}$$

Given the rule $f(\varepsilon_t; i_{t-1}^f)$ the Fed will follow at time t , we can now solve for the optimal forward guidance at $t-1$. The loss function at time t is given by

$$\begin{aligned} & \left((1 - k) \varepsilon_t - c\right)^2 + \theta \varepsilon_t^2 + \gamma \left(i_{t-1}^f - i_{t-1} - k\varepsilon_t - c\right)^2 \\ &= \left((1 - k) \varepsilon_t - c\right)^2 + \theta \varepsilon_t^2 + \gamma \left(k\varepsilon_t + \frac{1}{\gamma} c\right)^2 \end{aligned}$$

which in expectation at $t - 1$ is equal to

$$\left((1 - k)^2 + \gamma k^2 + \theta \right) \sigma_\varepsilon^2 + \frac{\gamma}{1 + \gamma} \left(i_{t-1}^f - i_{t-1} \right)^2. \quad (\text{A.3})$$

This is minimized by setting $i_{t-1}^f = i_{t-1}$.

Finally we solve for the optimal γ . Differentiating (A.2) with respect to γ , we have

$$\frac{\partial k}{\partial \gamma} = \frac{-k^2}{2k(1 + \gamma) - 1} < 0$$

so long as $k > 1/2$, which it will be in the parameter space we study. Setting $i_{t-1}^f = i_{t-1}$ and differentiating the ex ante loss function (A.3) with respect to γ yields

$$\left(-2(1 - k) \frac{\partial k}{\partial \gamma} + k^2 \left(1 - \frac{2k\gamma}{2k(1 + \gamma) - 1} \right) \right) \sigma_\varepsilon^2 > 0$$

since $\partial k / \partial \gamma < 0$ and $k > 1/2$. Since the ex ante loss is increasing in γ , it is optimal to set $\gamma = 0$.

A.4 Proof of Proposition 5

Given its conjecture about the rule the Fed is following, $\phi(\varepsilon_t; \eta_t)$, the market's conjecture about ε_t is

$$\tilde{\varepsilon}_t = \phi^{-1}(f(\varepsilon_t; \eta_t); \eta_t)$$

and the Fed's loss function is

$$(\varepsilon_t - f(\varepsilon_t; \eta_t))^2 + \theta \left(\phi^{-1}(f(\varepsilon_t; \eta_t); \eta_t) + \eta_t \right)^2.$$

Consider the effect on the value of the loss function of a small perturbation in the value of $f(\varepsilon_t; \eta_t)$, df . The effect of this perturbation is zero at the optimal $f(\cdot; \cdot)$ so we have

$$-(\varepsilon_t - f(\varepsilon_t; \eta_t)) + \theta \left(\phi^{-1}(f(\varepsilon_t; \eta_t); \eta_t) + \eta_t \right) \frac{\partial \phi^{-1}}{\partial i} \Big|_{i=f(\varepsilon_t; \eta_t)} = 0. \quad (\text{A.4})$$

Since $\phi(\phi^{-1}(x)) = x$, we have

$$\frac{\partial \phi^{-1}}{\partial i} \Big|_{i=f(\varepsilon_t; \eta_t)} = \frac{1}{\frac{\partial \phi}{\partial i} \Big|_{\varepsilon=\phi^{-1}(f(\varepsilon_t; \eta_t); \eta_t)}}.$$

Substituting into (A.4) gives

$$-(\varepsilon_t - f(\varepsilon_t; \eta_t)) + \theta \left(\phi^{-1}(f(\varepsilon_t; \eta_t); \eta_t) + \eta_t \right) \frac{1}{\frac{\partial \phi}{\partial i} \Big|_{\varepsilon=\phi^{-1}(f(\varepsilon_t; \eta_t); \eta_t)}} = 0.$$

Imposing rational expectations, we have $\phi = f$ so that this reduces to the differential equation

$$\frac{\partial f}{\partial \varepsilon} \Big|_{\varepsilon=\varepsilon_t} (\varepsilon_t - f(\varepsilon_t; \eta_t)) = \theta (\varepsilon_t + \eta_t),$$

which the optimal $f(\cdot; \cdot)$ must satisfy.

Now conjecture that $f = k_\varepsilon \varepsilon_t + c$. In this case the differential equation reduces to

$$k_\varepsilon (\varepsilon_t - k_\varepsilon \varepsilon_t - c) = \theta (\varepsilon_t + \eta_t)$$

or

$$k_\varepsilon (1 - k_\varepsilon) \varepsilon_t - k_\varepsilon c = \theta \varepsilon_t + \theta \eta_t$$

Matching coefficients yields

$$\begin{aligned} k_\varepsilon (1 - k_\varepsilon) &= \theta & \text{and} \\ c &= -\frac{\theta}{k_\varepsilon} \eta_t. \end{aligned} \tag{A.5}$$

Thus, we can write the optimal $f(\cdot; \cdot)$ as

$$f = k_\varepsilon \varepsilon_t + k_\eta \eta_t$$

where $k_\varepsilon (1 - k_\varepsilon) = \theta$ and $k_\eta = -\theta/k_\varepsilon$.

A.5 Proof of Proposition 7

Given ε_t (new private information about the target i_t^*) and X_{t-1} (the existing gap between i_{t-1}^* and i_{t-1} , which is public information), the central banker follows a rule of the form

$$i_t = i_{t-1} + k_X \cdot X_{t-1} + k_\varepsilon \cdot \varepsilon_t.$$

Rational investors understand that the Fed treats X_{t-1} and ε_t differently and conjecture it is following a rule of the form

$$i_t = i_{t-1} + \kappa_X \cdot X_{t-1} + \kappa_\varepsilon \cdot \varepsilon_t.$$

Given the conjecture and i_t , the investors back out

$$E^R[\varepsilon_t] = \frac{i_t - i_{t-1} - \kappa_X X_{t-1}}{\kappa_\varepsilon}.$$

To further simplify the problem, we also assume the following timing convention. First, based on its knowledge of X_{t-1} , the Fed decides on the value of the adjustment parameters k_X and k_ε it will use for the time t FOMC meeting. After making this decision, it deliberates further, and in so doing discovers the committee's consensus value of ε_t . Thus, the Fed picks k_X and k_ε , taking as given X_{t-1} , but before knowing the realization of ε_t . This timing convention is purely a technical trick that makes the problem more tractable without really changing anything of economic substance. Without it, the Fed's adjustment rule would turn out to depend on the realization of the product $X_{t-1} \varepsilon_t$. With the timing trick, what matters instead is the expectation of the product, which is zero. This simplification maintains the linearity of the Fed's optimal adjustment rule. We should emphasize that even with this somewhat strained intra-meeting timing, the Fed still behaves on a discretionary basis from one meeting to the next. Thus, while it agrees to values of k_X and k_ε in the first part of the time- t meeting, it has no ability to bind itself to those values across meetings. Hence the basic commitment problem remains.

Thus, the central banker picks k_X and k_ε to minimize the loss function

$$\begin{aligned} E_t \left[(i_t^* - i_t)^2 + \theta (\Delta i_t^\infty)^2 \right] &= E_t \left[((1 - k_X) X_{t-1} + (1 - k_\varepsilon) \varepsilon_t)^2 + \theta (E^R[\varepsilon_t])^2 \right] \\ &= (1 - k_X)^2 X_{t-1}^2 + (1 - k_\varepsilon)^2 \sigma_\varepsilon^2 + \theta \left(\left(\frac{1}{\kappa_\varepsilon} (k_X - \kappa_X) \right)^2 X_{t-1}^2 + \left(\frac{k_\varepsilon}{\kappa_\varepsilon} \right)^2 \sigma_\varepsilon^2 \right). \end{aligned}$$

Differentiating with respect to k_X yields

$$-(1 - k_X) + \theta \left(\frac{1}{\kappa_\varepsilon} (k_X - \kappa_X) \right) = 0.$$

Imposing rational expectations we have $k_X = \kappa_X$ which implies that $k_X = 1$.

Differentiating with respect to k_ε yields

$$0 = -(1 - k_\varepsilon) + \theta \left(\frac{k_\varepsilon}{\kappa_\varepsilon^2} \right).$$

Imposing rational expectations we have $k_\varepsilon = \kappa_\varepsilon$ which implies that

$$k_\varepsilon = \frac{k_\varepsilon^2}{k_\varepsilon^2 + \theta}$$

as before.

A.6 Proof of Proposition 8

Now the market is composed of two types investors. Fraction λ of investors is rational and makes the same inference about ε_t as before

$$E^R[\varepsilon_t] = \frac{i_t - i_{t-1} - \kappa_X X_{t-1}}{\kappa_\varepsilon}.$$

Fraction $1 - \lambda$ is naive. They ignore X_{t-1} and assume the Fed is following a rule of the form

$$i_t = i_{t-1} + \kappa^N \cdot \varepsilon_t$$

Given this conjecture and i_t , they back out

$$E^N[\varepsilon_t] = \frac{i_t - i_{t-1}}{\kappa^N} = \frac{k_X X_{t-1} + k_\varepsilon \varepsilon_t}{\kappa^N} = \beta \Delta i_t$$

The overall change in the infinite-horizon forward rate is

$$\Delta i_t^\infty = \lambda E^R[\varepsilon_t] + (1 - \lambda) E^N[\varepsilon_t].$$

The central banker picks k_X and k_ε to minimize the loss function

$$\begin{aligned} E_t \left[(i_t^* - i_t)^2 + \theta (\Delta i_t^\infty)^2 \right] &= (1 - k_X)^2 X_{t-1}^2 + (1 - k_\varepsilon)^2 \sigma_\varepsilon^2 \\ &+ \theta \left(\begin{aligned} &\left(\frac{\lambda}{\kappa_\varepsilon} (k_X - \kappa_X) + (1 - \lambda) \beta k_X \right)^2 X_{t-1}^2 \\ &+ \left(\frac{\lambda}{\kappa_\varepsilon} + (1 - \lambda) \beta \right)^2 k_\varepsilon^2 \sigma_\varepsilon^2 \end{aligned} \right). \end{aligned}$$

Differentiating with respect to k_X and imposing rational expectations for the rational investors ($k_X = \kappa_X$), we have

$$k_X = \frac{1}{1 + \theta (1 - \lambda) \beta \left(\frac{\lambda}{\kappa_\varepsilon} + (1 - \lambda) \beta \right)}. \quad (\text{A.6})$$

Note that when $\lambda = 1$, this implies $k_X = 1$ as before. For $\lambda < 1$, we have $k_X < 1$ if $\beta > 0$. When $\lambda = 0$, we have

$$k_X = \frac{1}{1 + \theta \beta^2}.$$

Differentiating with respect to k_ε and imposing rational expectations for the rational investors ($k_\varepsilon = \kappa_\varepsilon$), we have

$$k_\varepsilon = \frac{1}{1 + \theta \left(\frac{\lambda}{k_\varepsilon} + (1 - \lambda) \beta \right)^2}. \quad (\text{A.7})$$

If $\lambda = 1$, this reduces to

$$k_\varepsilon = \frac{k_\varepsilon^2}{k_\varepsilon^2 + \theta}$$

as before. If $\lambda = 0$ it reduces to

$$k_\varepsilon = \frac{1}{1 + \theta \beta^2} = k_X.$$

Note that we have $k_X \geq k_\varepsilon$ if

$$1 + \theta (1 - \lambda) \beta \left(\frac{\lambda}{k_\varepsilon} + (1 - \lambda) \beta \right) \leq 1 + \theta \left(\frac{\lambda}{k_\varepsilon} + (1 - \lambda) \beta \right)^2.$$

This is reduces to

$$(1 - \lambda) \beta \leq \frac{\lambda}{k_\varepsilon} + (1 - \lambda) \beta,$$

which is always true, so it is the case that $k_X \geq k_\varepsilon$.

The equilibrium relationship between ε_t and change in the short rate is given by

$$\begin{aligned} \hat{\beta} &= \frac{\text{Cov}(\varepsilon_t, \Delta i_t)}{\text{Var}(\Delta i_t)} \\ &= \frac{k_\varepsilon \sigma_\varepsilon^2}{k_X^2 \text{Var}[X_{t-1}] + k_\varepsilon^2 \sigma_\varepsilon^2}. \end{aligned}$$

Since $X_t = (1 - k_X) X_{t-1} + (1 - k_\varepsilon) \varepsilon_t$, the unconditional variance of X is given by

$$\text{Var}[X_t] = \text{Var}[X_{t-1}] = \frac{(1 - k_\varepsilon)^2}{1 - (1 - k_X)^2} \sigma_\varepsilon^2$$

so we can write

$$\hat{\beta} = \frac{k_\varepsilon \left(1 - (1 - k_X)^2 \right)}{k_X^2 (1 - k_\varepsilon)^2 + k_\varepsilon^2 \left(1 - (1 - k_X)^2 \right)}. \quad (\text{A.8})$$

Given θ and λ , the equilibrium we study satisfies (A.6), (A.7), and $\beta = \hat{\beta}$.

Finally, we show that the time consistency problem persists in this setting. As before, suppose that society's true concern with bond market volatility is given by θ , but that it appoints a central banker whose concern is given by θ_c . Recognizing rational investors' correct expectations in equilibrium, the Fed's utility is

$$E_t \left[(i_t^* - i_t)^2 + \theta (\Delta i_t^\infty)^2 \right] = (1 - k_X (\theta_c))^2 X_{t-1}^2 + (1 - k_\varepsilon (\theta_c))^2 \sigma_\varepsilon^2 + \theta \left(\frac{((1 - \lambda) \beta k_X (\theta_c))^2 X_{t-1}^2}{\left(\frac{\lambda}{k_\varepsilon (\theta_c)} + (1 - \lambda) \beta \right)^2 k_\varepsilon^2 (\theta_c) \sigma_\varepsilon^2} \right).$$

Differentiating with respect to θ_c yields

$$\begin{aligned} &\left[-(1 - k_X) + \theta (1 - \lambda)^2 \beta^2 k_X \right] X_{t-1}^2 \frac{\partial k_X}{\partial \theta_c} \\ &+ \left[-(1 - k_\varepsilon) + \theta \left(\frac{\lambda}{k_\varepsilon} + (1 - \lambda) \beta \right)^2 k_\varepsilon + \theta \left(\frac{\lambda}{k_\varepsilon} + (1 - \lambda) \beta \right)^2 (-\lambda) \right] \sigma_\varepsilon^2 \frac{\partial k_\varepsilon}{\partial \theta_c}. \end{aligned}$$

The first order conditions under discretion imply that

$$-\theta \left((1 - \lambda) \beta k_X \frac{\lambda}{k_\varepsilon} \right) = -(1 - k_X) + \theta (1 - \lambda)^2 \beta^2 k_X$$

and

$$0 = -(1 - k_\varepsilon) + \theta_c \left(\left(\frac{\lambda}{k_\varepsilon} + (1 - \lambda) \beta \right)^2 k_\varepsilon \right).$$

So we can rewrite the derivative with respect to θ_c as

$$\left[-\theta (1 - \lambda) \beta k_X \frac{\lambda}{k_\varepsilon} \right] X_{t-1}^2 \frac{\partial k_X}{\partial \theta_c} + \left[-\theta \left(\frac{\lambda}{k_\varepsilon} + (1 - \lambda) \beta \right)^2 \lambda \right] \sigma_\varepsilon^2 \frac{\partial k_\varepsilon}{\partial \theta_c}.$$

Since $\partial k_X / \partial \theta_c < 0$ and $\partial k_\varepsilon / \partial \theta_c < 0$, this is unambiguously positive so the Fed is better off with a lower θ_c .