Optimal irreversible monetary policy

Hasui, Kohei

Kobayashi, Teruyoshi

Sugo, Tomohiro

Optimal irreversible monetary policy

Kohei Hasui
Teruyoshi Kobayashi
Tomohiro Sugo

GRADUATE SCHOOL OF ECONOMICS

KOBE UNIVERSITY

ROKKO, KOBE, JAPAN

Optimal irreversible monetary policy[*]

Kohei Hasui[†]   Teruyoshi Kobayashi[‡]   Tomohiro Sugo[§]

February 26, 2021

Abstract

Real-world central banks have a strong aversion to policy reversals. Nevertheless, theoretical models of monetary policy within the dynamic general equilibrium framework normally ignore the irreversibility of interest rate control. In this paper, we develop a formal model that incorporates a central bank's discretionary optimization problem with an aversion to policy reversals. We show that, even under a discretionary regime, the optimal timing of liftoff from the zero lower bound is characterized by its history dependence, which arises from the option value to waiting, and there exists an optimal degree of policy irreversibility at which the social loss is minimized.

JEL Classification: E31, E52, E58, E61
Keywords: Monetary policy, policy irreversibility, reversal aversion, liquidity trap

1 Introduction

Many studies argue that central banks prefer "gradualism" in which the targets for the official interest rates are smoothed. The hypothesis of gradualism in monetary policy, or "interest rate smoothing," has been supported by a number of empirical studies (Sack and Wieland, 2000; English et al., 2003). On the other hand, there are also studies that do not support the interest-rate-smoothing hypothesis. They show that the statistical significance of policy inertia may come from an inappropriate estimation procedure, such as omitting serially correlated variables (Rudebusch, 2002, 2006), ignoring the weak instrument problem (Consolo and Favero, 2009) or missing time-varying equilibrium interest rates (Trehan and Wu, 2007).

An important yet often ignored aspect of the interest rate policy in practice is that the decisions of central banks are rarely reversed within a short period of time (Blinder, 2006; Mendes et al., 2017). For instance, if the target interest rate is raised by 0.25% today, it is quite unlikely that the central bank would decide to cut it by 0.25% at the next policy meeting. Rather, the only option left for the central bank would be either to increase the interest rate or to keep it unchanged. It seems that monetary policy decisions are practically irreversible, or more generally, central banks have an aversion to policy reversals (Lowe and Ellis, 1997; Blinder, 2006; Mendes et al., 2017). In the US, the Fed changed the Federal funds target 95 times between 1990 and 2008, and there were only two cases in which the Fed reversed their policy directions within two quarters. Nevertheless, the discussion of interest rate smoothing has normally been based on the standard linear-quadratic optimization problem of a central bank or the empirical results obtained within a class of the partial adjustment models, in both of which immediate interest-rate reversals are allowed at no cost. Alan Blinder, former vice chairman of the Board of Governors of the Federal Reserve System, asserts:

> "Although the basic logic of optimization suggests that such policy reversals should not be uncommon, central bankers seem to avoid them like the plague." (Blinder, 2006)

Blinder (2006) points out three factors that would lead central banks to have reversal aversion; central bankers tend to i) be concerned about losing their credibility, ii) avoid creating unnecessary volatility in the financial market and iii) be unwilling to be seen as admitting errors.

Given that real-world monetary policies are rarely reversed within a short period of time, a feasible optimal policy would have to be described as a solution to an optimization problem with an irreversibility constraint or an aversion to policy reversals. In this paper, we show that the central bank's reversal aversion, if present, would become a source of policy inertia even under a discretionary regime. The intuition behind the emergence of policy inertia is that there arises an option value to waiting, in the same way that is discussed in the classical studies on the

irreversibility of investment (Bernanke, 1983; Dixit and Pindyck, 1994). Once an investment has been made, the firm holds a risk of not being able to resell the installed capital at the desired price. The firm thus may find it optimal to hold off on making an investment and wait to see what will happen to the capital price in the future. In monetary policy, a central bank needs to absorb shocks to the natural rate by controlling the interest rate (Woodford, 2003; Walsh, 2017). In the presence of an irreversibility constraint, however, it may be optimal not to immediately react to the current shocks, because a current contractionary (expansionary) policy will prohibit the central bank from implementing an expansionary (contractionary) policy in the future. If there is a chance that the natural rate would revert to the previous level in the near future, then the value of option to wait can be larger than the cost of not fully absorbing the current shock. In this way, the central bank needs to manage the risk of the current policy change itself constraining future policy decisions.

Within this realistic optimization framework, we ask one of the most important questions in recent monetary policy: how should a central bank determine the timing of liftoff from the zero lower bound (ZLB)? (Evans et al., 2015; Carlstrom et al., 2015; Nakata and Schmidt, 2019b). Under discretionary policy, a central bank always has to absorb the natural rate shock, so the interest rate will be lifted from zero as soon as the natural rate takes a positive value. However, this is suboptimal once the interest rate hits the ZLB, because the zero-interest-rate policy (ZIRP) would be contractionary as long as the natural rate takes a negative value at which there is a gap between the actual and the desirable interest rates. The fully optimal policy is to keep the interest rate at zero for an extended period of time even after the natural rate turns positive (Eggertsson and Woodford, 2003; Jung et al., 2005). However, the well-known difficulty is that a commitment to the ZIRP is time inconsistent and therefore it is not practical for real-world central banks to implement the optimal commitment policy (Bodenstein et al., 2012; Nakata, 2015).

In this study, we find that delegating monetary policy to a central banker having reversal aversion would improve social welfare, because an aversion to policy reversals effectively functions as a commitment device at the ZLB. Since there is a value to waiting, a reversal-averse central bank would keep the interest rate at zero even after the natural rate turns positive. In fact, we reveal that there exists an optimal degree of reversal aversion at which the social loss, arising from the volatility of inflation and output gap, will be minimized. The *optimal irreversible policy* thus provides an alternative and practically feasible delegation scheme for circumventing the well-known time-inconsistency problem (Rogoff, 1985; Walsh, 2003; Bilbiie, 2014). In particular, in contrast to the previous delegation schemes in which a central bank is assigned additional target variables other than inflation and output, the proposed delegation scheme based on policy irreversibility is *target-free* in the sense that no additional target

variables are required.

The analysis of policy reversal aversion can be relevant not only for interest-rate policies, but also for balance sheet policies. In particular, when a central bank conducts a flow-based balance sheet policy, it would be natural for the central bank to avoid reversing the pace of asset purchases, just as it would avoid reversing the direction of interest rate changes. For example, when the Fed was reducing the pace of asset purchases towards the end of the QE3 program, the change in pace was being carefully smoothed; in the early stage of the program, the Fed purchased $40 billion of mortgage-backed securities (MBS) per month and $45 billion of Treasuries per month (Swanson, 2020; Federal Reserve Bank of New York). In December 2013, the Fed decided to slow the pace to $35 billion per month for MBS and to $40 billion per month for Treasuries. Such a gradual reduction in pace continued until they ceased the QE3 program in October 2014. It is likely that this would have reflected the Fed's aversion to reversing the direction of ongoing policies. While we focus on interest-rate policies in the current study, the discussion about policy reversal aversion could also be extended to balance sheet policies.

Aside from the theoretical contribution to the literature on interest rate policy at the ZLB, our study also contributes to the general DSGE framework by developing a novel technique for solving a dynamic optimization problem with an irreversibility constraint, or more generally with an aversion to policy reversals. While Kobayashi (2010) solved an optimization problem with a strict irreversibility constraint using the Bellman formulation, the current method employs a more general approach that allows us to treat a strict irreversible control problem as a special case. Here, we formulate the central bank's problem as a dynamic loss-minimization problem with a penalty for immediate reversals, in which a strictly irreversible policy corresponds to the solution under a prohibitively high penalty. The constrained monetary policy is analyzed in an otherwise standard new Keynesian model, so there could be various possible applications of this method for the study of recent monetary policy tools such as quantitative easing (Chen et al., 2012; Gertler and Karadi, 2013).

2 Related literature

The current study analyzes the effect of a central bank's reversal aversion within the standard new Keynesian framework developed by Woodford (2003). Along with many other studies, our paper focuses on discretionary policy because the optimal commitment policy is in principle time-inconsistent and notoriously difficult to implement in practice (Bodenstein et al., 2012; Nakata, 2015). Many researchers have thus proposed "commitment devices," which would lead a discretionary policy to be closer to, or under certain circumstances even identical to, the optimal commitment policy. This is also known as the problem of optimal delegation of monetary policy

and has been an active research area over the past three decades (Rogoff, 1985; Walsh, 2003, 2017; Bilbiie, 2014). For a recent example, Bilbiie (2014) proposes a specific loss function for the central bank with which the discretionary policy becomes identical to the optimal commitment policy. Our finding that there is an optimal degree of reversal aversion suggests that delegating monetary policy to an independent central banker having an aversion to policy reversals would improve social welfare. Our study thus contributes to the literature on the optimal delegation problem by proposing a highly feasible delegation scheme.

As mentioned in Introduction, a disadvantage of discretionary policy at the ZLB is that the interest rate is controlled to move in sync with the natural rate, if possible, although the optimal commitment policy requires the interest rate to be kept at the ZLB for an extended period of time even after the natural rate turns positive (Eggertsson and Woodford, 2003; Jung et al., 2005; Adam and Billi, 2006, 2007; Nakov, 2008). Accordingly, as a solution to the delegation problem, various commitment devices have been proposed to allow central banks to delay the timing of liftoff from the ZLB. As argued by Woodford (2003), the point is how to effectively incorporate history dependence, or policy inertia, into practical monetary policy. Nakov (2008) argues that social loss will be reduced by employing a Taylor rule that depends on the lagged interest rate. Nakata and Schmidt (2019b) analyze optimal discretionary policy in a liquidity trap, showing that a central bank with an interest-rate-smoothing objective can lower social loss by mimicking the optimal commitment policy. Jeanne and Svensson (2007) assert that a central bank's concern about its balance sheet and the level of capital would provide a mechanism to commit to a low interest rate. Boneva et al. (2018) consider a "threshold-based forward guidance" in which a central bank makes a state-contingent commitment to the ZIRP until the values of particular macroeconomic variables exceed prespecified thresholds.

The current paper investigates the role of a central bank's intrinsic aversion to policy reversals as a commitment device at the ZLB. Although the prevalence of policy reversal aversion and the need to study it have been repeatedly pointed out by several central bankers (Lowe and Ellis, 1997; Blinder, 2006; Mendes et al., 2017), formal analysis of this phenomenon is still scarce, with the exception of the work by Kobayashi (2010). While Kobayashi (2010) analyzes strict policy irreversibility using a backward-looking model developed by Ball (1999) and Svensson (1997), the current study employs a more general approach within the forward-looking new Keynesian model.

An important property that follows from a central bank's reversal aversion is that there exists an option value to waiting. A similar mechanism has been well studied in the literature on the irreversibility of investment (Bernanke, 1983; Dixit and Pindyck, 1994), in which a firm has to determine the optimal amount of investment, assuming that the installed capital could not be resold (i.e., "reversed") at the desired price. Recently, Lei and Tseng (2019) considered

a "wait-and-see" monetary policy in which a central bank faced with fixed adjustment costs for policy shifts values the option to wait until new information arrives in the future. A crucial difference from our study is that in their model, the emergence of the value of waiting results from the exogenous fixed costs of policy changes, while a central bank's intrinsic reversal aversion is the only factor that generates the value of waiting in our model. Our model thus makes it possible to quantify the welfare effect of a central bank's intrinsic aversion to policy reversals.

3 Policy reversal aversion in practice

As explained in Introduction, central banks tend to avoid reversing their policy decisions once their new policy actions have been recognized publicly. In the literature, on the other hand, the observed gradual policy shifts have been interpreted as reflecting the central banks' objective of interest rate smoothing (Sack and Wieland, 2000; English et al., 2003; Coibion and Gorodnichenko, 2012). In this section, we first describe the differences and similarities between the well-examined interest rate smoothing and policy reversal aversion. Then, we discuss some empirical evidence that the Fed and the Bank of England really have an aversion to reversals.

3.1 Relationship between interest rate smoothing and reversal aversion

The nominal interest rate under interest rate smoothing is generally specified as a policy function of the form

$$i_t = f\left(i_{t-1}, \mathbf{Z}_t, \mathbf{e}_t\right), \tag{1}$$

where i_t is the nominal interest rate (or the policy rate), and $\mathbf{Z}_t$ and $\mathbf{e}_t$ denote the vectors of endogenous variables, typically inflation and output, and exogenous variables such as economic shocks, respectively. Eq. (1) can be regarded as a general functional form for interest rate smoothing in which the central bank tries to suppress the fluctuation of the change in the *level* of the interest rate, $\Delta i_t \equiv i_t - i_{t-1}$. The central bank takes i_{t-1} into account in setting i_t, but the direction of the current policy change, given by $\text{sgn}(\Delta i_t)$, is determined independently of $\text{sgn}(\Delta i_{t-1})$. For instance, in a widely used partial adjustment model, the current interest rate is expressed as a weighted average of the past interest rates and the desired value (English et al., 2003; Rudebusch, 2006).

On the other hand, the policy reversal aversion we will consider in this paper constrains the *direction* of policy changes. Specifically, a policy function of a central bank having policy reversal aversion takes the form:

$$i_t = g\left(i_{t-1}, \Delta i_{t-1}, \mathbf{Z}_t, \mathbf{e}_t\right). \tag{2}$$

The central bank now needs to take into account the previous policy shift Δi_{t-1}, as well as the previous level of the interest rate i_{t-1}, to avoid reversing the previous direction of the policy change in setting i_t (or equivalently, Δi_t). In particular, *policy irreversibility* is regarded as a special case in which policy reversals are strictly prohibited and thereby $\mathrm{sgn}(\Delta i_t) \cdot \mathrm{sgn}(\Delta i_{t-1}) \geq 0$.

Kobayashi (2010) argues that due to the functional similarity between Eqs. (1) and (2), one could not statistically identify the difference between the two regimes, interest rate smoothing and policy irreversibility, if the regression equation for i_t is based on a partial adjustment model. If the central bank really has an aversion to policy reversals, however, the sign of Δi_{t-1} would affect the sign of Δi_t, and thereby the level of the current interest rate i_t depends on i_{t-1} in a nonlinear manner.

3.2 Example: The Fed and the Bank of England

Now let us look at the empirical data on the policy changes of central banks to see if there is a tendency toward reversal aversion in practice. We consider the Federal Funds (FF) target rate and the Bank rate as the policy instruments of the Fed and the Bank of England, respectively. The data on the FF target rate ranges from August 1987, when the chairman Greenspan was appointed, to December 2008. For the UK, the data ranges between June 1997, when the BoE was given its independence, and December 2016.

Let the sign of a policy change be $\delta \equiv \mathrm{sgn}(\Delta i)$. If a central bank has an aversion to policy reversals, then the unconditional sign distribution $P(\delta)$ would be different from the conditional distribution $P(\delta|\delta_-)$, where δ_- denotes the sign of the last policy shift. By contrast, if the central bank's decision is independent of the direction of the last policy change, as in the case of interest rate smoothing, then it would be that $P(\delta|\delta_-) = P(\delta)$. For simplicity, in this section we only consider positive and negative policy shifts (i.e., $\delta \in \{-1, 1\}$), since it is hard to identify the timing of "no policy change" (i.e., $\delta = 0$). Because there was always a possibility of having a non-regular policy meeting especially in the 80's, looking only at the regularly held meetings does not fully capture the decision of the central banks to do nothing (Thornton, 2006; Kobayashi, 2009). It should be noted that the number of reversal policy changes would be lower if we included the events with $\delta = 0$. We will consider the possibility of $\delta = 0$ in the theoretical model.

The FF target rate and the Bank rate are plotted in Fig. 1a and b, respectively, and the left panels of Fig. 1c and d show that for both the Fed and the BoE, the unconditional sign distribution $P(\delta)$ is significantly different from $P(\delta|\delta_-)$. The observed common patterns regarding the relationship between $P(\delta)$ and $P(\delta|\delta_-)$ are summarized in Table 1. The current direction of policy change is largely affected by the sign of the last policy shift, and there are few
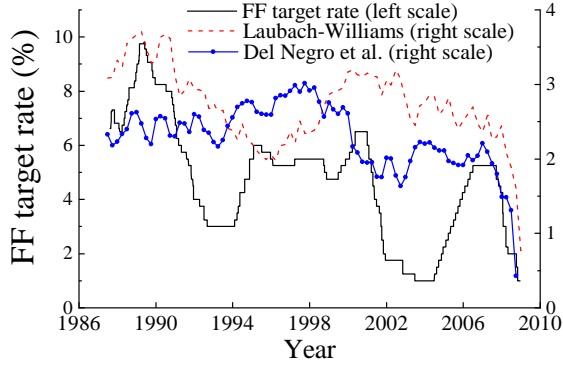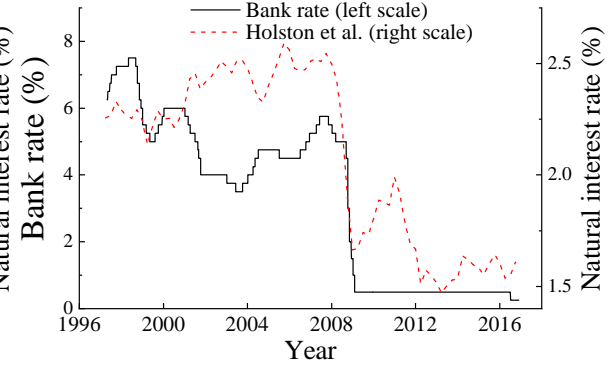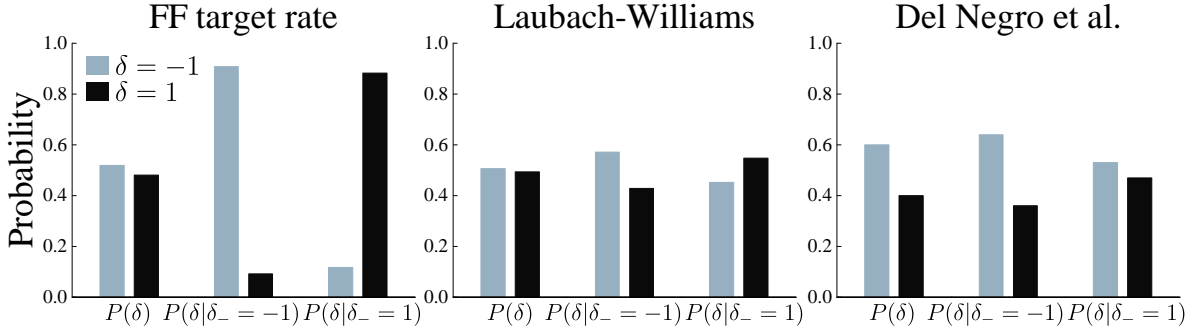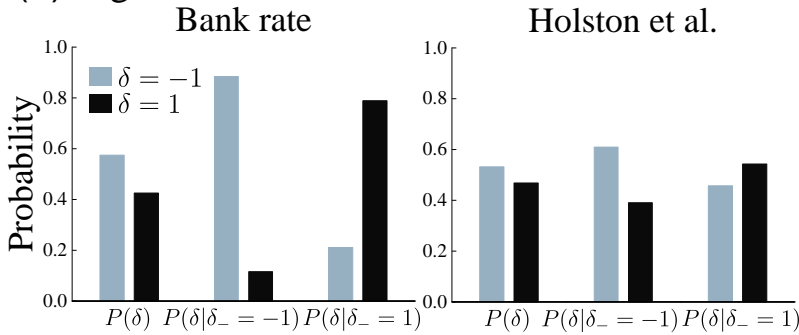
Figure 1: Reversal aversion of the Fed and the BoE. (a) The Federal Funds target rate (black solid) and the natural interest rates estimated by Laubach and Williams (2003) (red dashed) and Del Negro et al. (2017) (blue with circle). (b) The UK Bank rate (black solid) and the natural rates estimated by Holston et al. (2017) (red dashed). (c) Sign distributions of policy changes $\delta \equiv \mathrm{sgn}(\Delta i)$ (left panel) and the changes in the estimated natural rates, $\delta \equiv \mathrm{sgn}(\Delta r^n)$ (middle and right). In each panel, $P(\delta)$ and $P(\delta|\delta_-)$ denote the unconditional and conditional distributions, respectively, where δ_- denotes the sign of the last shift in the corresponding interest rate. (d) Same as (c) for the UK data.

Table 1: Observed relationship between $P(\delta)$ and $P(\delta|\delta_-)$ for policy changes. See the caption of Fig. 1 for the definitions of P, δ and δ_-.

	$\delta = -1$	$\delta = 1$
$\delta_- = -1$	$P(-1) < P(-1|-1),$	$P(1) > P(1|-1)$
$\delta_- = 1$	$P(-1) > P(-1|1),$	$P(1) < P(1|1)$

cases in which policy directions are reversed within a short period of time. Note that $P(1|-1)$ and $P(-1|1)$ are not zero because we exclude the possibility of $\delta = 0$ as explained above. Thus, we necessarily observe some policy reversals, but these changes were mostly made after a long period of "doing nothing." This property is not captured by the conventional models of interest rate smoothing, because they allow the policy rate to move in any direction independently of the past policy changes.

It should be noted that such a policy reversal aversion would naturally be observed if the central banks follow the natural rates exhibiting non-reversal behaviors. If this is the case, it is not appropriate to regard the central bank as having a reversal aversion. To explore this possibility, we also show the sign distributions of the changes in the estimated natural interest rates (Fig. 1c and d). We use the estimated natural rate data provided by Laubach and Williams (2003) and Del Negro et al. (2017) for the US and by Holston et al. (2017) for the UK.

We find that although the conditional sign distributions of policy shifts are totally different from the corresponding unconditional distribution (Fig. 1c and d, left), this is not true for the natural rate changes. In all the examined estimation methods, the conditional distributions of the natural rates are relatively similar to the unconditional distribution (Fig. 1c middle and right, Fig. 1d, right). Of course, we have to carefully interpret the result because there may be non-negligible errors and noises in the estimation of the natural rates. Nevertheless, given the striking difference between the sign distributions, it would be safe to say that the central banks have a reversal aversion at least to some extent.

4 Simplified framework: A finite-period model

To see how the introduction of reversal aversion would affect the optimal conduct of monetary policy, we first consider a simple four-period model with the ZLB. In this section, we assume for simplicity that the central bank is not allowed to reverse the direction of a policy shift made in the previous period. We will relax this assumption and consider an arbitrary degree of reversal aversion in Section 5.

4.1 Structural equations

We consider the following standard new Keynesian model developed by Woodford (2003):

$$x_t = \mathbb{E}_t x_{t+1} - \sigma(i_t - \mathbb{E}_t \pi_{t+1} - r_t^n), \tag{3}$$

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa x_t, \tag{4}$$

for $t = 0, 1, 2, 3$, where x_t, π_t, and i_t denote the output gap, inflation rate, and nominal interest rate, respectively. $\mathbb{E}_t$ is the expectations operator conditional on information available at time t. Equation (3) is the forward looking IS curve, which is derived from the standard Euler equation for consumption (Woodford, 2003; Walsh, 2017). The IS equation indicates that the current output gap is determined by the expected output gap and the deviation of the real interest rate from the natural rate, denoted by r_t^n. Equation (4) is the standard new Keynesian Phillips curve; the current inflation rate is expressed as the sum of the current output gap and inflation expectations when firms' opportunity to change prices is given by a constant probability (Calvo, 1983; Yun, 1996).

4.2 Monetary policy with an irreversibility constraint

Suppose that a central bank faces a strict irreversibility constraint in which reversing the direction of the policy shift made in period $t-1$ is not allowed in period t (Kobayashi, 2010). In the presence of a strict irreversibility constraint, the control space Ω_t for the current policy rate i_t is given by

$$\Omega_t = \begin{cases} \{i_t \,|\, i_t \leq i_{t-1}, i_t \in \Omega\} & \text{if } \delta_{t-1} = -1, \\ \Omega & \text{if } \delta_{t-1} = 0, \\ \{i_t \,|\, i_t \geq i_{t-1}, i_t \in \Omega\} & \text{if } \delta_{t-1} = 1, \end{cases} \tag{5}$$

where δ_t denotes the sign of the current policy shift $\Delta i_t \equiv i_t - i_{t-1}$, defined as $\delta_t \equiv \text{sgn}(\Delta i_t)$. Ω denotes the set of all possible values that the policy rate can take. To keep the analysis simple, we consider a small set of discrete values as the control space: $\Omega \equiv [0, 0.5, 1, 1.5] \times 1/4$. It should be noted that the set Ω does not include negative values, which effectively introduces the ZLB constraint: $i_t \geq 0$, $t = 0, 1, 2, 3$ (Eggertsson and Woodford, 2003; Jung et al., 2005; Adam and Billi, 2006, 2007; Billi, 2011; Coibion et al., 2012; Nakata and Schmidt, 2019b).

The social loss function to be minimized by the central bank is given by the sum of the squared deviations of inflation and of the output gap from their steady states: $\frac{1}{2}\mathbb{E}_0 \sum_{t=0}^{3} \beta^t \left(\pi_t^2 + \lambda x_t^2\right)$. Thus, the central bank's minimization problem constrained by the irreversibility condition is formulated as

$$\min_{i_t \in \Omega_t} \frac{1}{2}\mathbb{E}_0 \sum_{t=0}^{3} \beta^t \left(\pi_t^2 + \lambda x_t^2\right), \tag{6}$$

10

subject to Eqs. (3) and (4). Unlike the standard models of monetary policy, the central bank is faced with a *time-varying* control space, which suggests that the choice of the current policy rate is not independent of the change and the level of the previous policy rate. This adds a non-trivial nonlinearity to the model in addition to the ZLB constraint. We solve the minimization problem by formulating the value function.

4.3 Solving the model

We solve the four-period model in a backward manner. The central bank's problem in the last period (i.e., $t = 3$) can be specified by the value function:

$$V_3(r_3^n, i_2, \delta_2) = \min_{i_3 \in \Omega_3} \pi_3^2 + \lambda x_3^2, \tag{7}$$

$$\text{s.t.} \quad x_3 = -\sigma(i_3 - r_3^n), \tag{8}$$

$$\pi_3 = \kappa x_3, \tag{9}$$

where V_t denotes the value at time t. Because the economy terminates at $t = 3$, the expectations terms in the IS and the Phillips curves are dropped (Nakata et al., 2019). For $t = 0, 1, 2$, on the other hand, we need to include the expectations terms obtained in the previous step. The central bank's problem thus leads to

$$V_t(r_t^n, i_{t-1}, \delta_{t-1}) = \min_{i_t \in \Omega_t} \pi_t^2 + \lambda x_t^2 + \beta \mathbb{E}_t V_{t+1}(r_{t+1}^n, i_t, \delta_t), \tag{10}$$

$$\text{s.t.} \quad x_t = \mathbb{E}_t x_{t+1} - \sigma(i_t - \mathbb{E}_t \pi_{t+1} - r_t^n), \tag{11}$$

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa x_t, \tag{12}$$

for $t = 0, 1, 2$. We assume $i_{-1} = \delta_{-1} = 0$ to consider a situation in which the economy is already in a liquidity trap at the beginning of $t = 0$.

4.3.1 Parameters

Consider a simple scenario in which the natural rate takes either of the following two values: r_L or r_H, where $r_L < r_H$. The behavior of the natural rate is ruled by a Markov process whose conditional probabilities are given by

$$\text{Prob}(r_t^n = r_L | r_{t-1}^n = r_L) = q, \tag{13}$$

$$\text{Prob}(r_t^n = r_L | r_{t-1}^n = r_H) = p. \tag{14}$$

In this section, we set $r_L = -1/4$ and $r_H = 1.5/4$. Note that to investigate the optimal policy decision at the ZLB, r_L should be sufficiently small so that the policy rate could hit the ZLB. Following Adam and Billi (2007), the other model parameters are specified as follows: $\beta = 0.99$,

Table 2: Realized path of the natural rate in the four-period model

t	0	1	2	3
r_t^n	r_{L}	r_{H}	r_{H}	r_{H}

$\sigma = 1$, $\kappa = 0.024$, $\lambda = 0.003$ and $q = 0.875$. To see the uncertainty effect stemming from the natural rate, we will examine different values of p in the following.

To obtain an intuition behind the optimal (discretionary) irreversible policy at the ZLB, we specify the realized path of the natural rate as in Table 2. This scenario allows us to examine an illustrative situation in which the central bank is trapped at the ZLB at $t = 0$ but not necessarily for the periods from $t = 1$ onward.

4.3.2 Irreversible policy as risk management

As mentioned above, we consider that the economy has already been in a liquidity trap at the beginning of $t = 0$ so that the irreversibility constraint does not bind at $t = 0$. Since a positive natural rate shock hits the economy in period 1 (i.e., $r_1^n = r_{\mathrm{H}}$), the problem for the central bank at $t = 1$ is whether to raise the interest rate (i.e., $i_1 > 0$) or to keep it at the ZLB (i.e., $i_1 = 0$). Let $\ell_t \equiv \pi_t^2 + \lambda x_t^2$ be the temporal loss for the central bank. In period 1, the net *benefit* of raising the interest rate (i.e., $i_1 > i_0 = 0$) relative to keeping i_1 at 0 is given by

$$\mathcal{N}_1 = \ell_1|_{i_1=0} - \ell_1|_{i_1>0} + \beta \mathbb{E}_1 \left[V_2(r_2^n, 0, 0) - V_2(r_2^n, i_1, 1) \right]$$
$$= \ell_1|_{i_1=0} - \ell_1|_{i_1>0} + \beta \left\{ p \left[V_2(r_{\mathrm{L}}, 0, 0) - V_2(r_{\mathrm{L}}, i_1, 1) \right] \right.$$
$$\left. + (1-p) \left[V_2(r_{\mathrm{H}}, 0, 0) - V_2(r_{\mathrm{H}}, i_1, 1) \right] \right\}. \tag{15}$$

The central bank raises the interest rate in period 1 if $\mathcal{N}_1 > 0$, or

$$\underbrace{\ell_1|_{i_1=0} - \ell_1|_{i_1>0}}_{\text{Temporal benefit of raising } i_1} > \beta \underbrace{\mathbb{E}_1 \left[V_2(r_2^n, i_1, 1) - V_2(r_2^n, 0, 0) \right]}_{\text{Expected cost of raising } i_1}. \tag{16}$$

The LHS of Eq. (16) represents the temporal benefit of lifting off from the ZLB, which takes a positive value since the central bank could absorb the current natural rate shock by raising i_1, leading to a lower temporal loss. The RHS of Eq. (16) is interpreted as the (discounted) expected cost of raising i_1. Increasing the current interest rate might cause losses in the future because there is a possibility that a negative shock would occur again (i.e., $r_2^n = r_{\mathrm{L}}$). In that case, the central bank would need to lower the interest rate in period 2 to absorb the negative shock, reversing the direction of the policy change decided at $t = 1$. However, such a reversal policy shift is not allowed due to the irreversibility constraint. To maintain the controllability

of the interest rate in period 2, the current interest rate needs to be kept at the ZLB in period 1 (i.e., $i_1 = \delta_1 = 0$). Thus, the central bank faces a trade-off between the current benefit of absorbing the existing shock and the expected cost of abandoning flexible policies in the future.

An alternative way to interpret the optimal policy decision in period 1 is to regard the cost of not raising i_1 as the price of an option, in analogy with the irreversibility of investment (Bernanke, 1983; Dixit and Pindyck, 1994). If the central bank raises the interest rate in period 1, it would lose the option to cut the interest rate in period 2, due to irreversibility, which would cause losses if a negative shock hits the economy in period 2. The temporal loss that the central bank has to incur when keeping i_1 at the ZLB is therefore valuable as it gives the central bank an option to set $i_2 = 0$ when $r_2^n = r_L$. Thus, the choice of whether to raise i_1 can be considered as a problem of whether to purchase the option to cover the risk of failing to absorb a negative shock.

The condition that the central bank purchases the "option" (i.e., setting $i_1 = 0$) leads to

$$\underbrace{\ell_1|_{i_1=0} - \ell_1|_{i_1>0}}_{\text{option price}} < \beta \underbrace{\mathbb{E}_1 \left[V_2(r_2^n, i_1, 1) - V_2(r_2^n, 0, 0) \right]}_{\text{Expected payoff from option}}. \tag{17}$$

The LHS of Eq. (17) can be interpreted as the price of the option that guarantees a right to conduct flexible monetary policy in period 2 whatever happens to the natural rate. On the other hand, the RHS corresponds to the expected payoff from the option. The expected payoff from keeping i_1 at the ZLB, or the option value, plays a role as a price threshold below which the central bank is willing to buy the option to cover the irreversibility risk. This risk-management aspect of the central bank's problem adds one more dimension to the conventional formulation of monetary policy in which it is implicitly assumed that the central bank can freely control interest rates. In our model, the central bank needs to manage the risk of the current policy change itself constraining future policy decisions.

4.3.3 Optimal irreversible monetary policy

Because a change in policy itself creates a risk, the optimal level of the interest rate depends on the probability that the natural rate will return to the current value. If the central bank knows in period 1 that the natural rate will return to r_L in period 2 with high probability, then the central bank would not take a risk of losing the option to lower the interest rate in period 2. Thus, there is a negative relationship between the desired interest rate i_1 and $\text{Prob}(r_2^n = r_L | r_1^n = r_H) = p$ (Fig. 2). The figure indicates that the central bank should keep i_1 at zero if $p \geq 0.4$.

Now let us look at the path of the interest rates. For comparison, we also consider the following alternative policy regimes: i) optimal commitment, ii) pure discretionary policy and
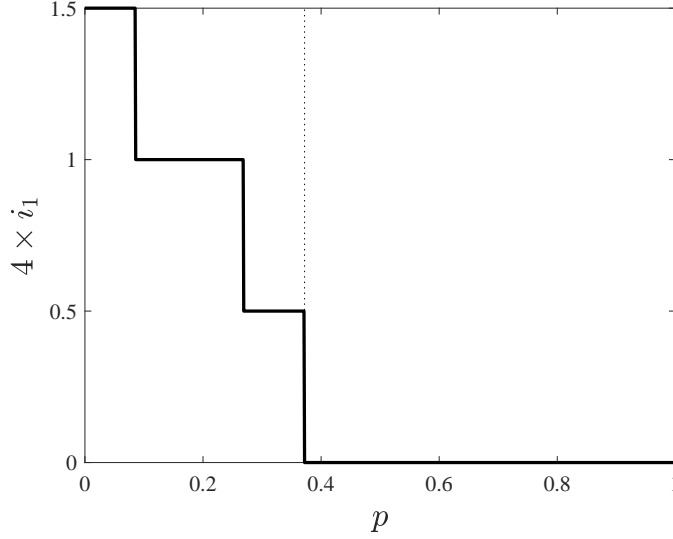
Figure 2: Optimal interest rate in period 1 under irreversible policy vs. $p = \text{Prob}(r_2^n = r_{\text{L}} | r_1^n = r_{\text{H}})$. Vertical dotted line denotes the threshold of p below which $\mathcal{N}_1 > 0$ (See, Eq. (15)).

iii) interest rate smoothing. Each of these three policies is assumed to be conducted without an irreversibility constraint. To obtain the path under interest rate smoothing, we consider the following standard temporal loss function (Nakata and Schmidt, 2019b; Debortoli et al., 2019):

$$\ell_t^{\Delta} = \pi_t^2 + \lambda x_t^2 + \lambda_{\Delta} \Delta i_t^2, \tag{18}$$

where λ_{Δ} is set at 0.01 for the moment. To illustrate the uncertainty effect that the probability p would have on the optimal policy rate, we consider two cases: $p = 0.01$ (low uncertainty) and $p = 0.4$ (high uncertainty). We note that setting $p = 0.01$ would be virtually equivalent to assuming that r_{H} is an "absorbing state" in which it is highly unlikely that the natural rate will return to r_{L} (Eggertsson and Woodford, 2003; Christiano et al., 2011).

As shown in Fig. 3, imposing a higher value of p makes the irreversible policy more history dependent, as is already indicated in Fig. 2. In other words, the risk of reversing the current policy decision in the future deters the central bank from controlling the interest rate in a flexible manner, delaying the timing of liftoff from the ZLB compared to the case of pure discretion (Fig. 3b). On the other hand, under interest rate smoothing, the timing of liftoff is the same as that under pure discretion, while changes in the interest rate are more smoothed. This illustrates the essential difference between the irreversible policy and the (conventional) interest-rate smoothing policy. The timing of liftoff is delayed in the irreversible policy because the interest rate cannot return to the ZLB right after liftoff, which makes the central bank more conservative. Interest-rate smoothing, on the other hand, does not constrain the direction
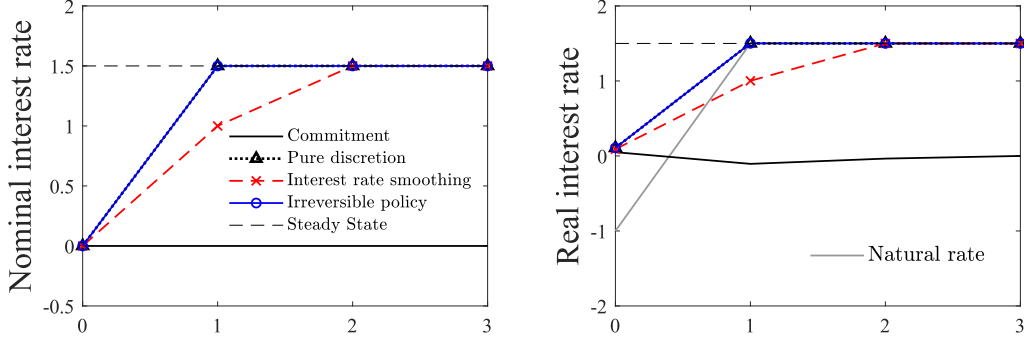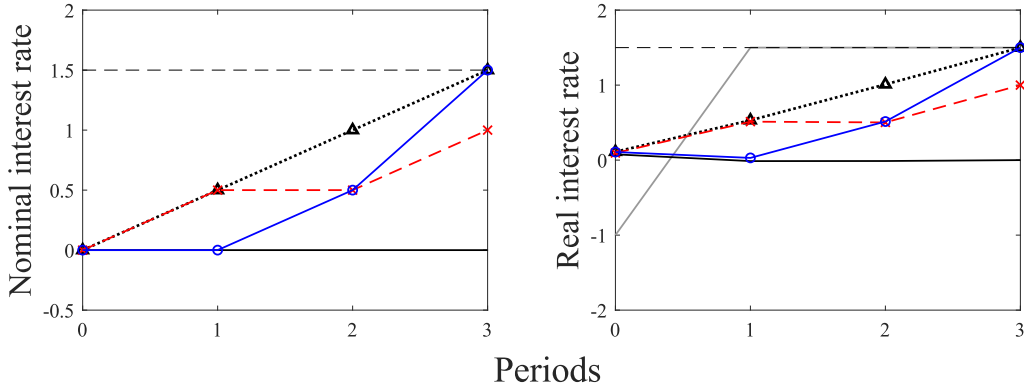
14

Figure 3: Nominal and real interest rates under different policy regimes. (a) r_H is nearly an absorbing state: $p = \text{Prob}(r_t^n = r_L | r_{t-1}^n = r_H) = 0.01$. (b) There is a reasonable probability of the natural rate returning to r_L from r_H: $p = 0.4$.

of policy shifts, so the timing of liftoff naturally coincides with that of the pure discretionary policy. However, since the (absolute) size of increment is penalized, the rate at which interest rate increases is slower for the interest-rate smoothing policy than the irreversible policy after liftoff. In passing, we note that interest rate is unchanged between periods 1 and 2 under interest rate smoothing, but this is simply because we use discrete values for interest rates with grid size 0.5 (see, description below Eq. (5)) to illuminate the characteristic of each policy regime.

This exercise illustrates an important property that irreversibility under uncertainty acts as a commitment device at the ZLB. Because the irreversible policy is more inertial than the pure discretionary policy, the behavior of the interest rate mimics that of optimal commitment at least to some extent (see, Fig. S1 in Supplementary Information (SI) for the responses of inflation and output). The introduction of an irreversibility constraint could thus reduce social loss, but the strict irreversibility constraint considered in this section may not be a desirable

15

choice. Rather, a more flexible policy constrained by an aversion to policy reversals would further reduce social loss if we could tune the degree of reversal aversion. We will address this issue in the next section.

5 General case: An infinite-horizon model with an arbitrary degree of reversal aversion

Now let us consider a general model in which the time horizon is infinite. In this section, we introduce a penalty term in the central bank's objective function to allow for an arbitrary degree of reversal aversion. This specification incorporates policies under a strict policy irreversibility constraint and pure discretion as two polar cases.

5.1 Model

5.1.1 Structure of the model

As in Section 4, we borrow the standard new Keynesian framework with the ZLB:

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

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa x_t, \tag{20}$$

$$r_t^n = \rho_r r_{t-1}^n + \epsilon_t, \tag{21}$$

$$i_t \geq 0. \tag{22}$$

Here, the natural rate is assumed to follow an AR(1) process whose disturbance ϵ_t is given by an *i.i.d.* random variable with mean zero and variance σ_ϵ. β, σ, κ and ρ_r are parameters satisfying $0 < \beta < 1$, $\sigma > 0$, $\kappa > 0$ and $|\rho_r| < 1$, respectively.

The social loss function is given by

$$\mathcal{L} = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j (\pi_{t+j}^2 + \lambda_x x_{t+j}^2), \quad \lambda_x \geq 0, \tag{23}$$

This loss function (23) is obtained in a micro-founded general equilibrium model with sticky prices (Calvo, 1983) within which the forward-looking IS curve (19) and the new Keynesian Phillips curve (20) are also derived (Woodford, 2003; Yun, 2005; Walsh, 2017).

5.1.2 Central bank with policy reversal aversion

To capture an arbitrary level of policy reversal aversion, here we introduce a penalty term in the periodic loss function:

$$L_t = \pi_t^2 + \lambda_x x_t^2 + \lambda_{\text{ir}} F(\delta_{t-1}, \Delta i_t), \tag{24}$$

where $F(\delta_{t-1}, \Delta i_t)$ is given by

$$
\begin{aligned}
&F(\delta_{t-1}, \Delta i_t) \\
&= \frac{\delta_{t-1}(1+\delta_{t-1})}{2} \times [\min(\Delta i_t, 0)]^2 - \frac{\delta_{t-1}(1-\delta_{t-1})}{2} \times [\max(\Delta i_t, 0)]^2.
\end{aligned}
\tag{25}
$$

Eq. (25) simply states that the penalty term F can take one of the following three forms depending on the sign of the policy shift in period $t-1$, δ_{t-1}:

$$
F(\delta_{t-1}, \Delta i_t) = \begin{cases} [\max(\Delta i_t, 0)]^2 & \text{if } \delta_{t-1} = -1, \\ 0 & \text{if } \delta_{t-1} = 0, \\ [\min(\Delta i_t, 0)]^2 & \text{if } \delta_{t-1} = 1. \end{cases}
\tag{26}
$$

Note that the central bank is not penalized if there was no policy shift in the last period (i.e., $\delta_{t-1} = 0$). If the central bank cut (raised) the policy rate in the previous period, then it would be penalized if it raises (cuts) the policy rate in the current period. The penalty term thus effectively introduces a "soft" irreversibility constraint, allowing us to capture an arbitrary degree of reversal aversion by tuning the weight parameter λ_{ir}. Clearly, the solution to the problem under a strict irreversibility constraint will be recovered if λ_{ir} is large enough, in which case our formulation is equivalent to the well-known penalty function method for solving a constrained optimization problem (cf., Luenberger and Ye, 2016, Ch. 13). By contrast, setting $\lambda_{\text{ir}} = 0$ recovers the standard new Keynesian model with the ZLB (Eggertsson and Woodford, 2003; Jung et al., 2005; Adam and Billi, 2007; Nakov, 2008). We also consider an absolute-value penalty term in section Appendix B.

We formulate the central bank's problem as the Bellman equation of the form

$$
\begin{aligned}
V(\mathbf{S}_t) &= \min_{i_t \geq 0} \pi(\mathbf{S}_t)^2 + \lambda_x x(\mathbf{S}_t)^2 + \lambda_{\text{ir}} F(\delta_{t-1}, \Delta i_t) + \beta \mathbb{E}_t V(\mathbf{S}_{t+1}), \\
\text{s.t.} \quad x(\mathbf{S}_t) &= \mathbb{E}_t x(\mathbf{S}_{t+1}) - \sigma(i_t - \mathbb{E}_t \pi(\mathbf{S}_{t+1}) - r_t^n), \\
\pi(\mathbf{S}_t) &= \kappa x(\mathbf{S}_t) + \beta \mathbb{E}_t \pi(\mathbf{S}_{t+1}),
\end{aligned}
\tag{27}
$$

where $\mathbf{S}_t$ denotes the state of the economy given by $\mathbf{S}_t \equiv [i_{t-1}, \delta_{t-1}, r_t^n]^\top$. In the following, we solve the problem by using a value function iteration method (cf., Miranda and Fackler, 2004). Details about the numerical algorithm are provided in Appendix.

5.1.3 Model parameters

The baseline parameter values are listed in Table 3. The parameter for the intertemporal elasticity of substitution for consumption σ is set at 1 following Adam and Billi (2007). The coefficient on the output gap in the Phillips curve and the weight on the output gap in the loss function are specified as $\kappa = 0.024$ and $\lambda_x = 0.003$, respectively, following Rotemberg and

Table 3: Baseline parameter values

Parameter	Value	Description
β	0.996	Subjective discount factor
i^*	1.5/4	Nominal interest rate at the steady state
σ	1	Intertemporal elasticity of substitution for consumption
κ	0.024	Slope of the Phillips curve
ρ_r	0.6	Persistence of natural rate shocks
σ_ϵ	0.233	Standard deviation of natural rate shocks
λ_x	0.003	Weight on output gap
λ_{ir}	0.0016	Degree of policy-reversal aversion (optimized)
λ_Δ	0.0008	Degree of interest rate smoothing (optimized)

Woodford (1998). As for the parameters related to the natural rate of interest, we set $\rho_r = 0.6$ and $\sigma_\epsilon = 0.233$ so that we can avoid using extrapolation in approximating the value function. Note that the calibrated value of $\rho_r(= 0.6)$ lies in between those of Woodford (2003) (i.e., 0.35) and Nakov (2008) (i.e., 0.65). The discount factor is set at $\beta = 0.996$ to ensure that the steady-state real interest rate is 1.5% annually, given that the targeted rate of inflation is zero. The key parameter of the model, namely the weight on the penalty term λ_{ir}, is set at 0.0016 since this turns out to minimize the social loss as we will show below. For discretionary policy under these parameter configurations, the average fraction of time periods during which the policy rate hits the ZLB is 8.87%.

5.2 Results

5.2.1 Policy function

In the presence of an aversion to policy reversals, the optimal policy rate under discretion is expressed as a nonlinear function of the previous interest rate i_{t-1}, the sign of previous policy shift δ_{t-1} and the current natural rate r_t^n:

$$i_t = i(i_{t-1}, \delta_{t-1}, r_t^n). \tag{28}$$

An important point is that, even under discretionary policy with no intrinsic economic inertia, the current policy rate is largely affected by the previous policy decision, leading to the emergence of an endogenous policy inertia. This suggests the possibility that the determination of the interest rate under discretionary policy would exhibit a history dependence, a key characteristic of optimal commitment policy.
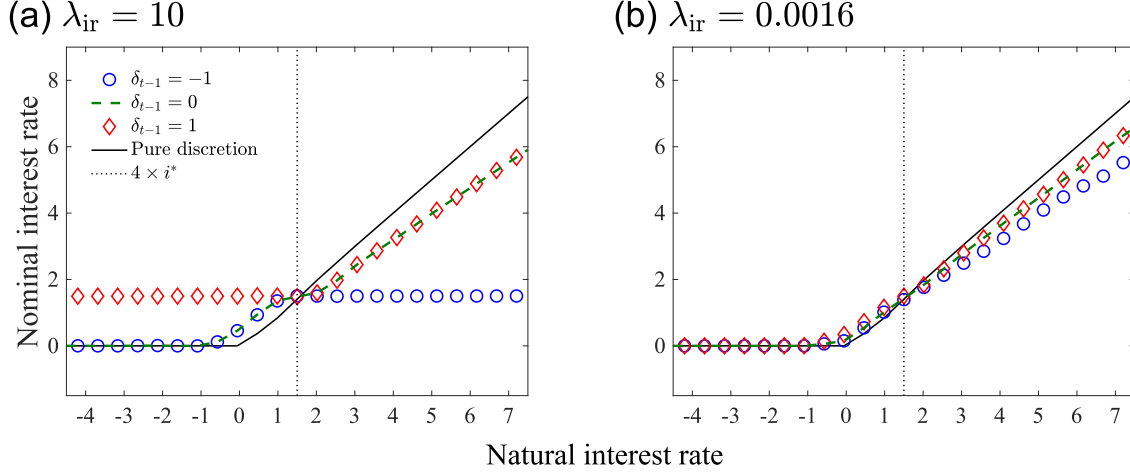
(a) $\lambda_{\mathrm{ir}} = 10$ (b) $\lambda_{\mathrm{ir}} = 0.0016$

Figure 4: Policy function for interest rate. Black solid line denotes a pure discretionary policy without reversal aversion (i.e., $\lambda_{\mathrm{ir}} = 0$). Blue circle, green dashed and red diamond respectively denote irreversible policies when the sign of policy shift in the previous period is -1, 0 and 1. We set $i_{t-1} = i^* = 1.5/4$ (vertical dotted line).

For comparison, let us first consider a situation in which the penalty for policy reversals is prohibitively high so that the central bank virtually faces an irreversibility constraint as in the previous section. With our parameter configurations, we find that the central bank will never reverse the direction of policy shifts implemented in the previous period if $\lambda_{\mathrm{ir}} \geq 10$. In Fig. 4a, the policy functions for irreversible policies and a pure discretionary policy are shown against the natural rate of interest. If $\delta_{t-1} = -1$ (denoted by blue circles), i.e., the policy rate was reduced in the previous period, then the central bank facing an irreversibility constraint cannot increase the current policy rate even if a large positive shock (in terms of the departure from the steady state level i^*) hits the economy. The converse is true for $\delta_{t-1} = 1$ (red diamonds), in which the policy rate was increased in the previous period and therefore the central bank cannot cut the current policy rate in the face of a large negative shock.

An interesting property of the policy function with reversal aversion is that the interest rates with $\delta_{t-1} = 1$ (red diamonds) and 0 (green dashed) for a given positive shock are lower than that under pure discretion. Note that if $\delta_{t-1} = 0$ or 1, the central bank would not be penalized by increasing the interest rate at t, and it would therefore be able to fully absorb a positive shock to the natural rate. Nevertheless, it can be optimal not to fully respond to a positive shock, because the higher the current interest rate, the more likely that the interest rate would need to be cut in the future. This property illustrates the emergence of a "conservative" or precautionary behavior of the central bank stemming solely from reversal aversion.

Fig. 4b shows the policy functions when reversals are optimally penalized (i.e., $\lambda_{\mathrm{ir}} = 0.0016$).

19

While the policy rate is not strictly irreversible, the optimal response to a positive natural rate shock is weaker when there was an interest rate cut in the previous period (i.e., $\delta_{t-1} = -1$, blue circles) than when the interest rate was kept unchanged (i.e., $\delta_{t-1} = 0$, green dashed) or increased (i.e., $\delta_{t-1} = 1$, red diamonds). This illustrates how such a "soft" irreversibility constraint works while keeping the flexibility of pure discretion to some extent. In fact, since the policy response is weaker when $\delta_{t-1} = 1$, the current inflation and output gap accordingly exhibit slightly larger values than in the case of pure discretion (Fig. S2b).

5.2.2 Impulse response

Here, we consider the following two different scenarios for impulse response analysis. In the first scenario, the natural rate drops significantly in period 1 so that the nominal interest rate hits the ZLB, and then the natural rate reverts to its steady state value in period 2. We call this scenario a *temporal liquidity trap*. In the second scenario, the natural rate drops in period 1 as in the first scenario, and it stays at the low level in period 2 and then reverts to its steady state in period 3. We call this situation a *prolonged liquidity trap*. In both scenarios, we assume $\delta_0 = 0$ and that shocks are not persistent (i.e., $\rho_r = 0$) to extract the endogenous persistence coming from the policy regime itself. As before, we also examine an interest-rate-smoothing policy, but the weight on the smoothing term, denoted by λ_Δ, is now determined such that the social loss is minimized.

Figure 5 plots impulse responses for the scenario of a temporal liquidity trap. Three points are worth mentioning. First, although the interest rate is raised in period 2 for all the examined regimes, the pace of increasing the interest rate is slower under the optimal discretionary policy with reversal aversion than under pure discretion (Fig. 5c). This is intuitive because raising the nominal interest rate in period 2 is penalized due to the reversal aversion term in the loss function. Second, the presence of reversal aversion slightly mitigates the sharp drops in the output gap and inflation that we see under pure discretion at $t = 1$ (Fig. 5a, b). This is because, with reversal aversion, private agents know that the central bank's incentive to raise the interest rate in period 2 will be partially suppressed, which leads them to expect a higher inflation rate and output gap. We obtain similar results in the second scenario, namely a prolonged liquidity trap (Fig. S3). Third, the discretionary policy with reversal aversion and the interest-rate-smoothing policy reduce the deflationary bias at the risky steady state. As is discussed by Hills, Nakata and Schmidt (2019) and Nakata and Schmidt (2019a), the risk of hitting the ZLB in the future would keep current inflation lower than the deterministic steady-state level even when the nominal interest rate is well above the ZLB (Figs. 5 and 6). The discretionary policy with reversal aversion and the interest-rate-smoothing policy would reduce the deflationary bias and thus improve welfare by lowering the probability of returning to the
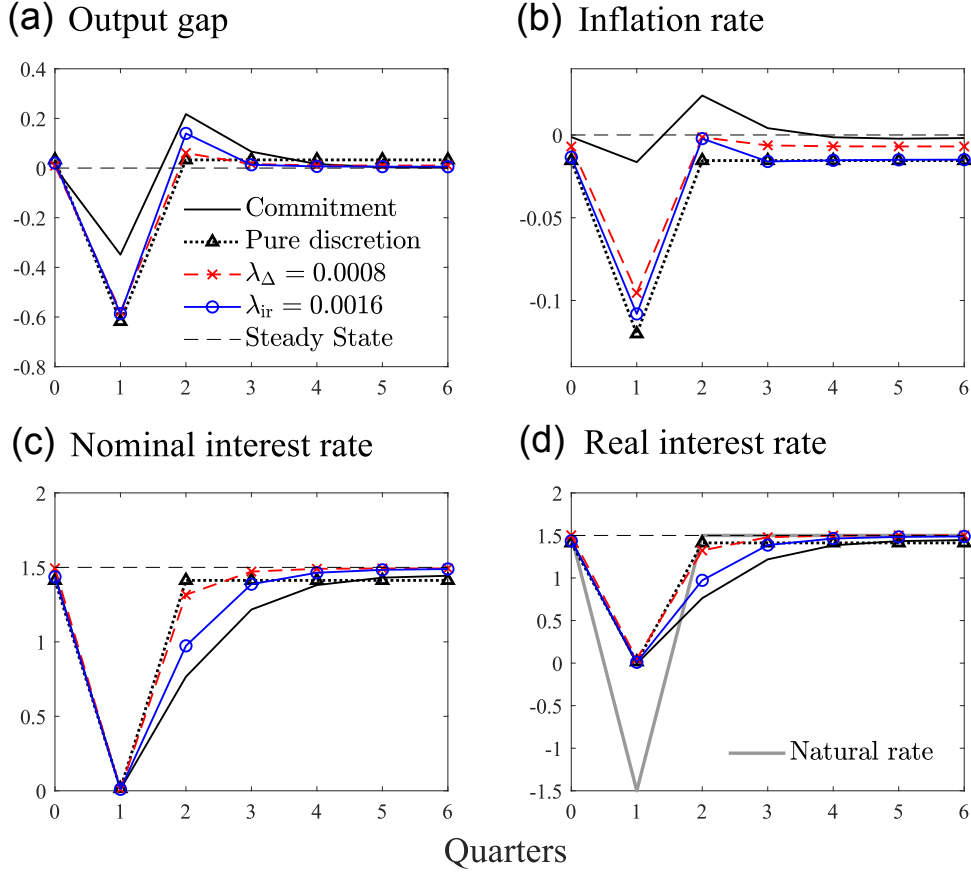
(a) Output gap

(b) Inflation rate

(c) Nominal interest rate

(d) Real interest rate

Legend:
— Commitment
···▲··· Pure discretion
– ✳ – $\lambda_\Delta = 0.0008$
–○– $\lambda_{ir} = 0.0016$
– – Steady State
— Natural rate

Quarters

Figure 5: Impulse responses to a temporal -3% natural rate shock. We set $\rho_r = 0$ to remove exogenous policy inertia.

ZLB after liftoff, as we will see in Section 5.3.

Although policy irreversibility can be socially beneficial when the interest rate is at the ZLB by creating an endogenous policy inertia, there is also a cost of not being able to promptly respond to shocks. For instance, it is optimal under the commitment policy to fully absorb a positive natural rate shock as long as the interest rate is not constrained by the ZLB (Fig. 6). However, even for positive shocks, an aversion to policy reversals would prohibit the central bank from raising the interest rate enough to completely offset the shocks, because doing so increases the likelihood that the interest rate would need to be cut in the future. This conservative policy creates an additional volatility of inflation and output and increases the social loss. Consequently, reversal aversion can be socially beneficial only in an environment in which the ZLB is relevant and thereby some extent of policy inertia is required. This gives us a trade-off between the benefit of creating an endogenous policy inertia at the ZLB and the cost of excess volatility caused by reversal aversion. In fact, we show in section 5.3.1 that there is an optimal

(a) Output gap

(b) Inflation rate

(c) Nominal interest rate
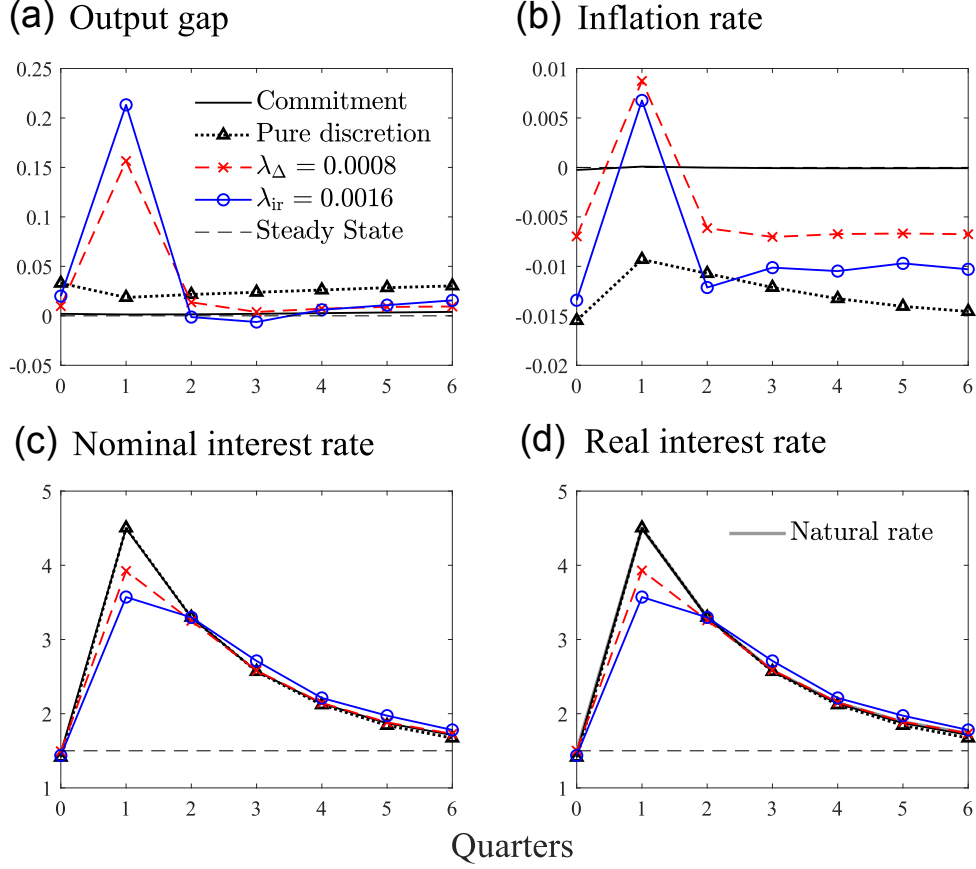
(d) Real interest rate

Quarters

Figure 6: Impulse responses to a temporal 3% natural rate shock. We set $\rho_r = 0.6$.

degree of reversal aversion at which the social loss is minimized.

5.2.3 Simulated path

Figure 7 illustrates sample simulated paths of the interest rates for $\lambda_{ir} = 0$ (pure discretion), $\lambda_{ir} = 10$ (strict irreversible policy) and $\lambda_{ir} = 0.0016$ (optimal irreversible policy). We note that the bandwidth within which the interest rate fluctuates becomes narrower as the penalty weight λ_{ir} increases. It turns out that the average variance of the interest rate is 0.945 for $\lambda_{ir} = 0.0016$ and 1.271 for $\lambda_{ir} = 0$. This is intuitive because even in a "normal" circumstance in which the central bank can shift the interest rate freely in the desired direction, it exhibits a more conservative behavior in that the size of the policy shift is smaller than that under pure discretion. This reflects the fact that the central bank with reversal aversion internalizes the influence that the current policy shift would have on the chance that the direction of a policy shift would need to be reversed in the future. Simulated paths of the interest rates under alternative regimes will be discussed in section Appendix B.
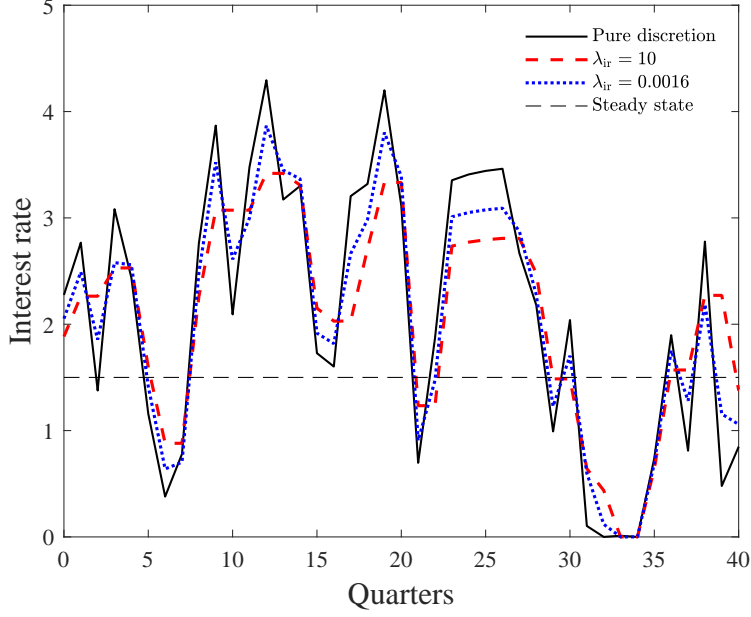
Figure 7: Simulated path of nominal interest rate.

5.3   Welfare consequence of policy reversal aversion

The current general equilibrium framework allows us to ask to what extent a central bank's policy reversals should be penalized. If policy reversals should not be penalized to flexibly react to economic shocks, then the desired value of λ_{ir} will be zero. By contrast, if penalizing policy reversals would increase social welfare, then the optimal value of λ_{ir} will take a positive value. As discussed above, a central bank with reversal aversion internalizes the possible influence of a current policy change on the future reversal probability. Searching for an optimal value of λ_{ir} is therefore essentially equivalent to finding the optimal extent of internalization or forward-lookingness of the central bank.

To find the optimal weight, we define the relative welfare gain from reversal aversion as

$$\mathcal{W} \equiv \left(\frac{\mathcal{L}^{\mathrm{disc}}}{\mathcal{L}^{\mathrm{ir}}} - 1 \right) \times 100, \tag{29}$$

where $\mathcal{L}^{\mathrm{ir}}$ and $\mathcal{L}^{\mathrm{disc}}$ denote the unconditional welfare losses under discretion with and without reversal aversion, respectively. Our computation of the unconditional losses is based on Eq. (23), in which we generate 1,000 paths for the inflation rate and the output gap with length 1,200 periods, and the simulated unconditional losses are averaged over the 1,000 runs. The initial 200 periods are discarded in each run to eliminate the influence of the initial conditions, where we set $\delta_0 = 0$.
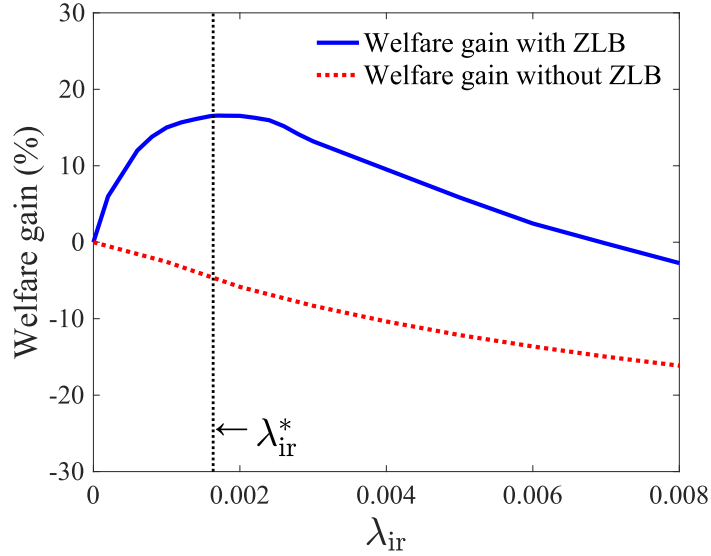
Figure 8: Welfare gain from reversal aversion. The welfare gain, defined by Eq. (29), is maximized at $\lambda_{ir}^* = 0.0016$.

5.3.1 Optimal degree of policy irreversibility

The welfare gain from reversal aversion against λ_{ir} is presented in Fig. 8. It is important to note that there is no welfare gain from reversal aversion in an environment in which there is no ZLB, i.e., the ZLB constraint is ignored (red dotted line in Fig. 8). In our model the natural rate is the only source of economic fluctuations, which could be fully offset if there were no ZLB (a phenomenon called "divine coincidence" (Blanchard and Galí, 2007)). Penalizing policy reversals in the absence of the ZLB therefore just worsens welfare by prohibiting the central bank from absorbing the natural rate shocks.

Fig. 8 confirms that the benefit of penalizing policy reversals could arise only when there is a chance that the policy rate would hit the ZLB, in which case the natural rate shock would affect the real economy under any kind of policy regime. As shown in Fig. 8, an increase in the weight λ_{ir} has a non-monotonic impact on the relative welfare gain. For a small value of λ_{ir}, the benefit of penalizing the central bank exceeds the cost of inhibiting flexible policy making, but the relative balance between the benefit and the cost is reversed when λ_{ir} is large enough. It turns out that there is a unique maximizer λ_{ir}^* at which the social loss is minimized. We now define *optimal irreversible policy* as follows:

Definition. *Optimal irreversible policy is a discretionary monetary policy with the relative weight on reversal aversion λ_{ir} in Eq. (24) equal to λ_{ir}^*.*

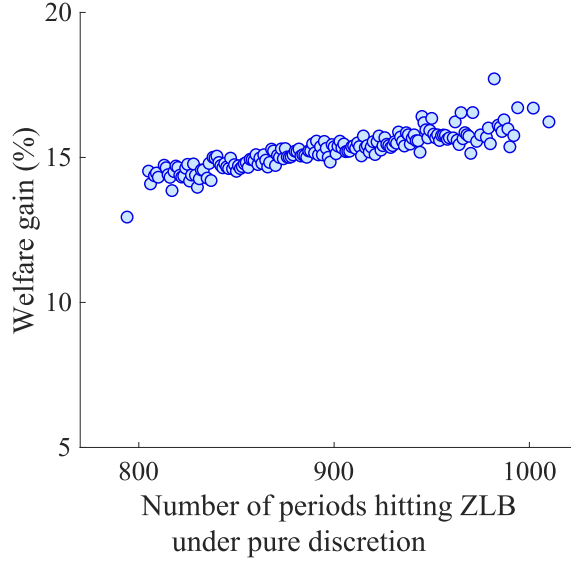We also compare the welfare gains under various policy regimes, such as optimal commitment

24

Figure 9: Relationship between welfare gain from reversal aversion and the frequency of hitting ZLB under pure discretion. Each dot represents a simulation of 10,000 periods for $\lambda_{ir} = 0.0016$.

policy, simple interest rate rules and interest rate smoothing, in section Appendix B. In the following sections we will examine the detailed mechanisms behind the optimality of reversal aversion.

5.3.2   Frequency of ZIRPs

Because the natural rate is the only source of economic fluctuations in our model, the optimal value of λ_{ir} should be 0 if the ZLB were absent. The benefit of assigning a positive weight on the penalty term thus depends on the extent to which the ZLB constrains monetary policy. In fact, our numerical simulations suggest that there is a strong positive correlation between the welfare gain and the frequency of hitting the ZLB under pure discretion for a given value of λ_{ir} (Fig. 9). This implies that the more frequently the central bank is likely to be trapped by the ZLB, the higher the welfare gain from penalizing policy reversals. It should be noted that the frequency of hitting the ZLB in Fig. 9 is independent of the degree of reversal aversion.

While the welfare benefit of policy reversal aversion is attributed to the distortions stemming from the ZLB, the frequency of interest rate hitting the ZLB would also be affected by the degree of reversal aversion. This suggests that there is an endogeneity in the relationship between the frequency of hitting the ZLB and the desirable degree of reversal aversion. Basically, with other things being equal, imposing a higher weight on the penalty term makes the central bank more conservative, and thereby the chance that the policy rate reaches the ZLB will be lowered. This
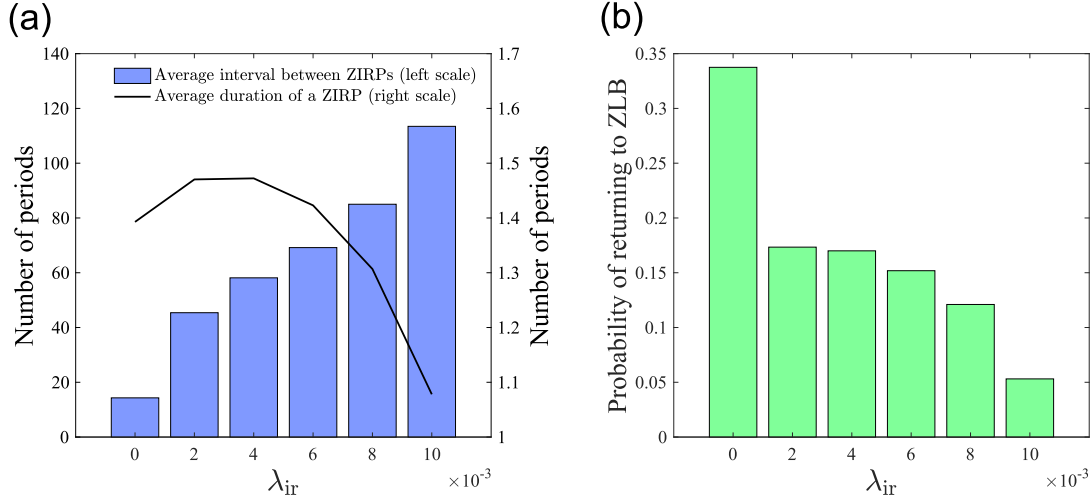
25

Figure 10: Frequency of zero interest rate policies (ZIRPs). (a) Interval time between two consecutive ZIRPs (left scale) and duration of a ZIRP (right scale). For a given λ_{ir}, the average is taken over 1,000 simulations with length 10,000 periods. (b) Probability that the interest rate returns to the ZLB within four periods of liftoff.

is reflected by a positive relationship between λ_{ir} and the average *interval time* between ZIRPs (see, bars in Fig. 10a). The average interval between ZIRPs is given by the average number of periods between the end of a ZIRP and the beginning of the following ZIRP.

In contrast, the relationship between λ_{ir} and the *duration time* of a ZIRP is not monotonic (see, solid line in Fig. 10a). The duration of a ZIRP is defined as the number of periods during which the interest rate is kept at the ZLB. On one hand, a higher value of λ_{ir} makes the central bank less aggressive, as mentioned above, which would make it more likely to keep the policy rate at zero once it reaches the ZLB. This effect would extend the average duration time of a ZIRP. On the other hand, the frequency of hitting the ZLB decreases with λ_{ir} because the policy rate hits the zero floor only when there are infrequent yet large negative shocks. If this happens, such a large negative shock is expected to quickly revert to the steady state in the succeeding periods, as long as it is stationary, which would lead the central bank to exit from the ZLB shortly. Due to these two opposite effects, the average duration time of a ZIRP exhibits an inverse U-shaped curve.

English et al. (2015) argue that there might be a substantial probability that the Federal funds rate will return to the effective lower bound (ELB) within one year of liftoff. They show, based on a small-scale macro model of the US economy, that the probability of returning to the ELB within one year of liftoff can be up to 40%, depending on the policy rule that the Fed would employ. Their result is roughly consistent with the case of $\lambda_{ir} = 0$ in our model,

where the probability of returning to the ZLB within four quarters of liftoff is about 35% (Fig. 10b). Indeed, our result suggests that the reversal-aversion policy with $\lambda_{\mathrm{ir}} > 0$ will reduce the probability of returning to the ZLB to a large extent compared with a simple discretionary policy.

5.4 Alternative policy regimes

Next, we compare the welfare performance of optimal irreversible policy with those of different policy regimes, namely optimal commitment policy, the Taylor rule, interest rate smoothing and a simple interest-rate rule with price-level targeting. We investigate price-level targeting based on the argument put forth by Eggertsson and Woodford (2003) and Fujiwara et al. (2013) that introducing a price-level target into a policy rule can be effective when the nominal interest rate is constrained by the ZLB.

We specify the Taylor rule and a simple rule with price-level targeting as follows:

- *Taylor rule with interest rate smoothing*:

$$i_t = \max\{0, \ \rho_i i_{t-1} + (1 - \rho_i)(\phi_\pi \pi_t + \phi_x x_t)\}, \tag{30}$$

  where we set $\phi_\pi = 1.5$ and $\phi_x = 0.5$. For ρ_i, we examine two values: $\rho_i = 0$ and 0.5.

- *Simple rule with price-level targeting*:

$$i_t = \max\{0, \ \phi_p(p_t - p^*) + \phi_x x_t\}, \tag{31}$$

  where $\phi_p = 1.5$ and $\phi_x = 0.5$. p_t and p^* denote the logarithms of the price level and its target value, respectively.

Within a discretionary regime, we also examine a different specification of the penalty term F in the loss function (24). Instead of penalizing the squared value of Δi_t, we consider the following absolute-value penalty F^{abs}.

- *Absolute value penalty*:

$$F^{\mathrm{abs}}(\delta_{t-1}, \Delta i_t) = \begin{cases} |\max(\Delta i_t, 0)| & \text{if } \delta_{t-1} = -1, \\ 0 & \text{if } \delta_{t-1} = 0, \\ |\min(\Delta i_t, 0)| & \text{if } \delta_{t-1} = 1. \end{cases} \tag{32}$$

Then, monetary policy is delegated to a central bank whose loss function is given by

$$L_t^{\mathrm{abs}} = \pi_t^2 + \lambda_x x_t^2 + \lambda_{\mathrm{ir}}^{\mathrm{abs}} F^{\mathrm{abs}}(\delta_{t-1}, \Delta i_t), \tag{33}$$

where the optimal weight on the penalty term turns out to be $\lambda_{\mathrm{ir}}^{\mathrm{abs}} = 0.0008$.

Table 4: Welfare losses under alternative policy regimes

Policy regime	Welfare loss
Optimal commitment	0.0033
Pure discretion	0.0197
Optimal irreversible policy, $\lambda_{\mathrm{ir}} = 0.0016$	0.0169
Optimal irreversible policy, $\lambda_{\mathrm{ir}}^{\mathrm{abs}} = 0.0008$	0.0168
Interest rate smoothing, $\lambda_\Delta = 0.0008$	0.0122
Taylor rule, $\rho_i = 0$	0.1712
Taylor rule with smoothing, $\rho_i = 0.5$	0.1365
Simple rule with price-level targeting	0.0851

Note: The values of λ_{ir}, $\lambda_{\mathrm{ir}}^{\mathrm{abs}}$ and λ_Δ are optimized.

Table 4 summarizes the social losses under the examined regimes. It turns out that the optimal irreversible policy yields a lower loss compared to the Taylor rules and a simple rule with price-level targeting, but the optimized interest rate smoothing outperforms all the examined policies other than the optimal commitment policy. As we saw above, interest rate smoothing can incorporate policy inertia by making the previous interest rate a current state variable through the penalization of $\Delta i_t{}^2$. Our result suggests that minimizing the squared deviation of Δi_t independently of $\mathrm{sgn}(\Delta i_{t-1})$ may be more efficient than penalizing policy reversals for the purpose of reducing the social loss.

We find that the interest rate volatility under the Taylor rule with interest rate smoothing, given by Eq.(30), is much smaller than that under alternative regimes (Fig. S4), but as shown in Table 4, the Taylor rules generally lead to high welfare losses. In fact, the inflation rate under the Taylor rule with interest smoothing turns out to be more volatile than that under optimal interest rate smoothing (Fig. S5). The examined Taylor-type rules are less efficient than discretionary policy even if a price-level targeting objective is included (see also Nakov (2008)).

It is important to note that the welfare performance computed here does not take into account the potential central bank's loss of reputation, which could be a serious concern for real-world central banks. As is pointed out by Blinder (2006), in reality central banks tend to avoid policy reversals "like the plague" based on the notion that immediate policy reversals would necessarily undermine their credibility. Given this fact, the degree of feasibility would have to be evaluated in quantifying the desirability of a policy scheme, a task we leave for future research.

6   Concluding remarks

Avoiding policy reversals is a common practice of central banks, but the consequent economic outcome and welfare effects have not been examined so far despite the recent development of theoretical models of monetary policy. Our study contributes to the literature by revealing that reversal aversion can be desirable for the society as long as the ZLB is relevant. This implies that a central bank *should* have an aversion to policy reversals from the welfare point of view, which provides a rationale for the widely observed phenomenon that policy reversals are rare.

There are some unsettled issues that should be addressed in future research. First, we argue that a central bank's policy reversal aversion would improve the social welfare measured by the standard loss function that incorporates the volatility of inflation and output gap. Our results thus favor delegating monetary policy to a central banker having an aversion to policy reversals even if policy reversals per se are not relevant to welfare. On the other hand, there might be a situation in which a benevolent central bank needs to avoid policy reversals simply because policy reversals directly affects social welfare. For example, as the volatility of interest rates itself would become a welfare loss when money is included in the utility function (Woodford, 2003) or there is stickiness in retail interest rate setting (Kobayashi, 2008; Teranishi, 2015), interest rate reversals themselves could directly worsen welfare if there are some frictions in financial markets. It would be interesting to explore in what circumstances a microfounded model would give us a social loss function with a penalty term on interest-rate reversals.

Second, because the introduction of an irreversibility constraint would generally make it harder to guarantee the determinacy of equilibrium, a detailed analysis on the determinacy condition will be needed. An irreversibility condition, or more generally a penalty for policy reversals, would occasionally constrain the direction of future policy shifts, so a necessary condition for local stability would become harder to satisfy compared to the standard models. Third, the presence of reversal aversion may provide a microfoundation for "interest-rate smoothing." As mentioned in Introduction, the standard specifications of interest-rate smoothing in empirical and theoretical models allow for policy reversals and are therefore not necessarily consistent with the widely observed reversal aversion. Instead, the prevailing gradual behavior of interest rates may be explained, at least to some extent, by the presence of reversal aversion. More detailed analysis is needed on these issues, and we hope our model will stimulate further research.

Appendix A   Numerical algorithm

We solve the central bank's optimization problem Eq. (27) by value function iteration. Since the ZLB introduces nonlinearity into the model, unlike in a simple linear-quadratic framework, we

need to employ an approximation technique to obtain the functional form of the value function.

We first specify the grids for the three state variables i_{t-1}, δ_{t-1}, and r_t^n. Let $\mathbf{s}^1$, $\mathbf{s}^2$, and $\mathbf{s}^3$ denote the vectors of grids for i_{t-1}, δ_{t-1}, and r_t^n, respectively, where the size of $\mathbf{s}^i$ is $n_i \times 1$. The whole state space $\overline{\mathbf{S}}$ is given by a tensor product of the three grid vectors: $\overline{\mathbf{S}} = \mathbf{s}^1 \otimes \mathbf{s}^2 \otimes \mathbf{s}^3$. The size of tensor $\overline{\mathbf{S}}$ is then given by $N = n_1 \times n_2 \times n_3 = 129 \times 3 \times 30$. The *p.d.f.* of the natural rate is assumed to be normal and discretized into 13 values using the Gaussian quadrature.

Let $V(\mathbf{S}_t)$ and $\mathbf{h}_t \equiv (x_t, \pi_t)^\top$ denote a real-valued value function and the vector of forward-looking variables at time t, respectively, where $\mathbf{S}_t \equiv [i_{t-1}, \delta_{t-1}, r_t^n]^\top \subset \overline{\mathbf{S}}$. We compute the value function V and a policy function $\mathbf{h}$ as time-invariant functions of $\mathbf{S} \subset \overline{\mathbf{S}}$. The procedure is as follows:

1. Given a particular set of grids for the three state variables, $\mathbf{S}_t^j \subset \overline{\mathbf{S}}$, and the initial guess for functions V and $\mathbf{h}$, respectively denoted by V^0 and $\mathbf{h}^0$, compute $V^1(\mathbf{S}_t^j)$, $\mathbf{h}^1(\mathbf{S}_t^j)$ and a policy function $i(\mathbf{S}_t^j)$ as a solution to the problem (27). A cubic-spline function is used to approximate $V^0(\mathbf{S}_{t+1}^j)$ and $\mathbf{h}^0(\mathbf{S}_{t+1}^j)$.

2. Repeat step 1 for all $j = 1, \ldots N$.

3. Stop if $\|V^1(\overline{\mathbf{S}}) - V^0(\overline{\mathbf{S}})\|_\infty / \|V^0(\overline{\mathbf{S}})\|_\infty < 1.5 \times 10^{-6}$. Otherwise, update the functions as $V^0 \equiv V^1$ and $\mathbf{h}^0 \equiv \mathbf{h}^1$, and go to step 1.

In our computation, convergence is reached within four hours using Matlab with Xeon 3.60GHz and a 32GB memory.

Appendix B   Robustness check

As a robustness check, we will now examine how the optimality of reversal aversion could be affected by changing the baseline setting. We first consider the influences of varying the values of the steady-state interest rate, i^*, and the weight on the output gap, λ_x. Then, we compare the welfare gains under alternative policy regimes.

Appendix B.1   Steady-state interest rate

As discussed above, the desirability of policy reversal aversion depends heavily on the likelihood that the ZLB constraint is binding. The most important parameter that governs the frequency of hitting the ZLB is the steady state interest rate i^*. If i^* is high enough, even a very large negative shock could not cause the nominal interest rate to reach the ZLB. In contrast, a lower i^* itself makes it more likely for the central bank to implement a ZIRP.
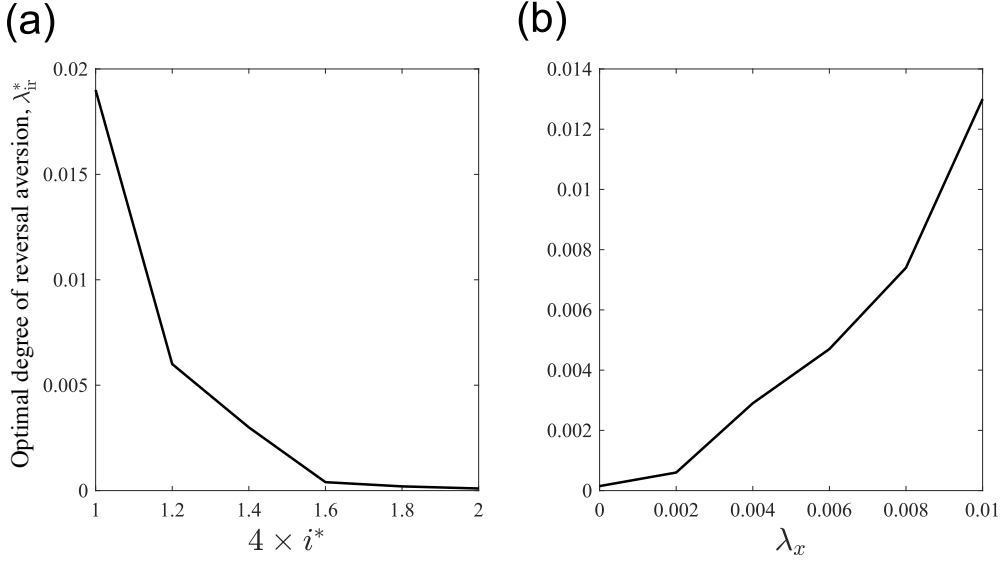
Figure 11: Optimal degree of reversal aversion. λ^*_{ir} as a function of (a) the steady state interest rate and (b) the weight on output gap.

Fig. 11a shows that there is a negative relationship between i^* and λ^*_{ir}. This is expected from Fig. 9, which plots a positive correlation between the welfare gain and the likelihood of being trapped at the ZLB, measured by the number of periods hitting the ZLB under pure discretion. As a rise in i^* reduces the frequency of reaching the ZLB, imposing a penalty on policy reversals would become less desirable. It turns out that if the steady-state interest rate is close to 2%, the advantage of reversal aversion would virtually disappear (i.e., $\lambda^*_{\text{ir}} \approx 0$).

Appendix B.2   Weight on output gap

Because the degree of reversal aversion λ_{ir} is given as a weight in the loss function, its optimal value necessarily depends on the weights on the central bank's target variables– namely, inflation and output. Fig. 11b plots λ^*_{ir} as a function of the weight on output λ_x. It shows a positive relationship between λ_x and λ^*_{ir}; as the weight on output increases, the desirable degree of reversal aversion also rises. This is intuitive because λ^*_{ir} is essentially the weight on policy reversals relative to inflation and output. λ^*_{ir} needs to be positively correlated with λ_x to keep an appropriate level of "punishment" for reversals. In addition, an increase in λ_x itself worsens the trade-off between output gap and inflation rate both at and away from the ZLB (Nakata and Schmidt, 2019a), so λ_{ir} has to be larger to offset the destabilizing effect coming from a rise in λ_x.

31

References

Adam, K. and R. M. Billi (2006) "Optimal monetary policy under commitment with a zero bound on nominal interest rates," *Journal of Money, Credit and Banking* 38, 1877–1905.

Adam, K. and R. M. Billi (2007) "Discretionary monetary policy and the zero lower bound on nominal interest rates," *Journal of Monetary Economics* 54, 728–752.

Ball, L. (1999) "Efficient rules for monetary policy," *International Finance* 2, 63–83.

Bernanke, B. S. (1983) "Irreversibility, uncertainty, and cyclical investment," *Quarterly Journal of Economics* 98, 85–106.

Bilbiie, F. O. (2014) "Delegating optimal monetary policy inertia," *Journal of Economic Dynamics and Control* 48, 63–78.

Billi, R. M. (2011) "Optimal inflation for the US economy," *American Economic Journal: Macroeconomics* 3, 29–52.

Blanchard, O. and J. Galí (2007) "Real wage rigidities and the New Keynesian model," *Journal of Money, Credit and Banking* 39, 35–65.

Blinder, A. S. (2006) "Monetary policy today: Sixteen questions and about twelve answers," Working Papers 73, Princeton University, Department of Economics, Center for Economic Policy Studies.

Bodenstein, M., J. Hebden, and R. Nunes (2012) "Imperfect credibility and the zero lower bound," *Journal of Monetary Economics* 59, 135–149.

Boneva, L., R. Harrison, and M. Waldron (2018) "Threshold-based forward guidance," *Journal of Economic Dynamics and Control* 90, 138–155.

Calvo, G. A. (1983) "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics* 12, 383–398.

Carlstrom, C. T., T. S. Fuerst, and M. Paustian (2015) "Inflation and output in New Keynesian models with a transient interest rate peg," *Journal of Monetary Economics* 76, 230–243.

Chen, H., V. Cúrdia, and A. Ferrero (2012) "The macroeconomic effects of large-scale asset purchase programmes," *Economic Journal* 122, 289–315.

Christiano, L., M. Eichenbaum, and S. Rebelo (2011) "When is the government spending multiplier large?" *Journal of Political Economy* 119, 78–121.

Coibion, O. and Y. Gorodnichenko (2012) "Why are target interest rate changes so persistent?" *American Economic Journal: Macroeconomics* 4, 126–162.

Coibion, O., Y. Gorodnichenko, and J. Wieland (2012) "The optimal inflation rate in New Keynesian models: Should central banks raise their inflation targets in light of the zero lower bound?" *Review of Economic Studies* 79, 1371–1406.

Consolo, A. and C. A. Favero (2009) "Monetary policy inertia: More a fiction than a fact?" *Journal of Monetary Economics* 56, 900–906.

Debortoli, D., J. Kim, J. Lindé, and R. Nunes (2019) "Designing a simple loss function for central banks: Does a dual mandate make sense?" *Economic Journal* 129, 2010–2038.

Del Negro, Marco., D. Giannone, M. P. Giannoni, and A. Tambalotti (2017) "Safety, liquidity, and the natural rate of interest," *Brookings Papers on Economic Activity* 48, 235–316.

Dixit, A. K. and R. S. Pindyck (1994) *Investment under Uncertainty*, Princeton University Press.

Eggertsson, G. B. and M. Woodford (2003) "Zero bound on interest rates and optimal monetary policy," *Brookings Papers on Economic Activity* 34, 139–233.

English, W. B., W. R. Nelson, and B. P. Sack (2003) "Interpreting the significance of the lagged interest rate in estimated monetary policy rules," *Contributions in Macroeconomics* 3.

English, W. B., J. D. López-Salido, and R. J. Tetlow (2015) "The Federal Reserve's framework for monetary policy: Recent changes and new questions," *IMF Economic Review* 63, 22–70.

Evans, C., J. Fisher, F. Gourio, and S. Krane (2015) "Risk management for monetary policy near the zero lower bound," *Brookings Papers on Economic Activity* 46, 141–219.

Federal Reserve Bank of New York, `https://www.newyorkfed.org/markets/programs-archive/large-scale-asset-purchases`, accessed on Dec 23, 2020.

Fujiwara, I., T. Nakajima, N. Sudo, and Y. Teranishi (2013) "Global liquidity trap," *Journal of Monetary Economics* 60, 936–949.

Gertler, M. and P. Karadi (2013) "QE 1 vs. 2 vs. 3. . . : A framework for analyzing large-scale asset purchases as a monetary policy tool," *International Journal of Central Banking* 9, 5–53.

Hills, T. S., Nakata, T. and S. Schmidt (2019) "Effective lower bound risk," *European Economic Review* 120, 103321.

Holston, K., T. Laubach, and J. C. Williams (2017) "Measuring the natural rate of interest: international trends and determinants," *Journal of International Economics* 108, S39-S75.

Jeanne, O. and L. E. O. Svensson (2007) "Credible commitment to optimal escape from a liquidity trap: The role of the balance sheet of an independent central bank," *American Economic Review* 97, 474–490.

Jung, T., Y. Teranishi, and T. Watanabe (2005) "Optimal monetary policy at the zero-interest-rate bound," *Journal of Money, Credit, and Banking* 37, 813–835.

Kobayashi, T. (2008) "Incomplete interest rate pass-through and optimal monetary policy," *International Journal of Central Banking* 4, 77–118.

Kobayashi, T. (2009) "Announcements and the effectiveness of monetary policy: A view from the US prime rate," *Journal of Banking & Finance* 33, 2253–2266.

Kobayashi, T. (2010) "Policy irreversibility and interest rate smoothing," *The B.E. Journal of Macroeconomics* 10, 1–29.

Laubach, T. and J. C. Williams (2003) "Measuring the natural rate of interest," *Review of Economics and Statistics* 85, 1063-1073.

Lei, X. and M. C. Tseng (2019) ""Wait-and-See" monetary policy," *Macroeconomic Dynamics* 23, 1793-1814.

Lowe, P. and L. Ellis (1997) "The smoothing of official interest rates," in *Monetary Policy and Inflation Targeting*, ed. by P. Lowe, Reserve Bank of Australia, RBA Annual Conference Volume.

Luenberger, D. G. and Y. Ye (2016) *Linear and Nonlinear Programming*, Springer, 3rd ed.

Mendes, R., S. Murchison, and C. A. Wilkins (2017) "Monetary policy under uncertainty: Practice versus theory," Discussion Papers 2017-13, Bank of Canada.

Miranda, M. J. and P. L. Fackler (2004) *Applied Computational Economics and Finance*, MIT Press.

Nakata, T. (2015) "Credibility of optimal forward guidance at the interest rate lower bound," Feds notes, Board of Governors of the Federal Reserve System.

Nakata, T., R. Ogaki, S. Schmidt, and P. Yoo (2019) "Attenuating the forward guidance puzzle: Implications for optimal monetary policy," *Journal of Economic Dynamics and Control* 105, 90-106.

Nakata, T. and S. Schmidt (2019a) "Conservatism and liquidity traps," *Journal of Monetary Economics* 104, 37–47.

Nakata, T. and S. Schmidt (2019b) "Gradualism and liquidity traps," *Review of Economic Dynamics* 31, 182-199.

Nakov, A. (2008) "Optimal and simple monetary policy rules with zero floor on the nominal interest rate," *International Journal of Central Banking* 4, 73–127.

Rogoff, K. (1985) "The optimal degree of commitment to an intermediate monetary target," *Quarterly Journal of Economics* 100, 1169–1189.

Rotemberg, J. J. and M. Woodford (1998) "An optimization-based econometric framework for the evaluation of monetary policy: expanded version," NBER Technical Working Papers 0233, National Bureau of Economic Research.

Rudebusch, G. D. (2002) "Term structure evidence on interest rate smoothing and monetary policy inertia," *Journal of Monetary Mconomics* 49, 1161–1187.

Rudebusch, G. D. (2006) "Monetary policy inertia: Fact or fiction?" *International Journal of Central Banking* 2, 85–135.

Sack, B. and V. Wieland (2000) "Interest-rate smoothing and optimal monetary policy: a review of recent empirical evidence," *Journal of Economics and Business* 52, 205–228.

Svensson, L. E. O. (1997) "Inflation forecast targeting: Implementing and monitoring inflation targets," *European Economic Review* 41, 1111–1146.

Swanson, E. T. (2020) "Measuring the effects of Federal Reserve forward guidance and asset purchases on financial markets," *Journal of Monetary Economics*, in press.

Teranishi, Y. (2015) "Smoothed interest rate setting by central banks and staggered loan contracts," *Economic Journal* 125, 162–183.

Thornton, D. L. (2006) "When did the FOMC begin targeting the federal funds rate? What the verbatim transcripts tell us," *Journal of Money, Credit and Banking* 38, 2039–2071.

Trehan, B. and T. Wu (2007) "Time-varying equilibrium real rates and monetary policy analysis," *Journal of Economic Dynamics and Control* 31, 1584–1609.

Walsh, C. (2003) "Speed limit policies: the output gap and optimal monetary policy," *American Economic Review* 93, 265–278.

Walsh, C. E. (2017) *Monetary Theory and Policy*, MIT Press, 4th ed.

Woodford, M. (2003) *Interest and Prices: Foundation of a Theory of Monetary Policy*, Princeton University Press.

Yun, T. (1996) "Nominal price rigidity, money supply endogeneity, and business cycles," *Journal of Monetary Economics* 37, 345–370.
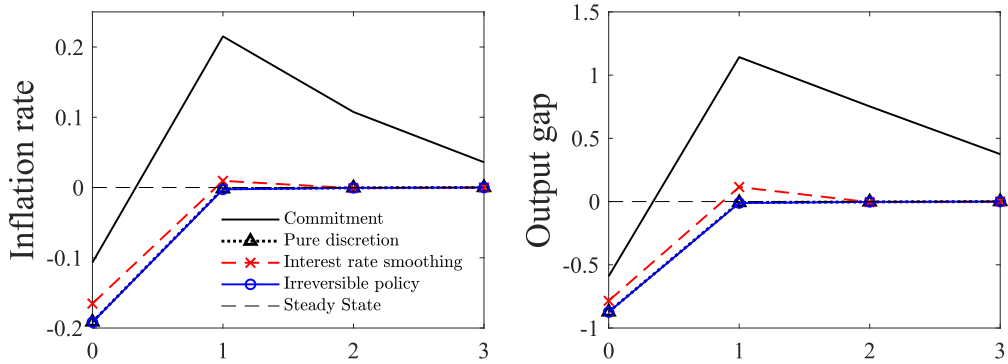
Yun, T. (2005) "Optimal monetary policy with relative price distortions," *American Economic Review* 95, 89–109.

"Optimal irreversible monetary policy"

Kohei Hasui, Teruyoshi Kobayashi, Tomohiro Sugo

(a) $p = 0.01$

(b) $p = 0.4$
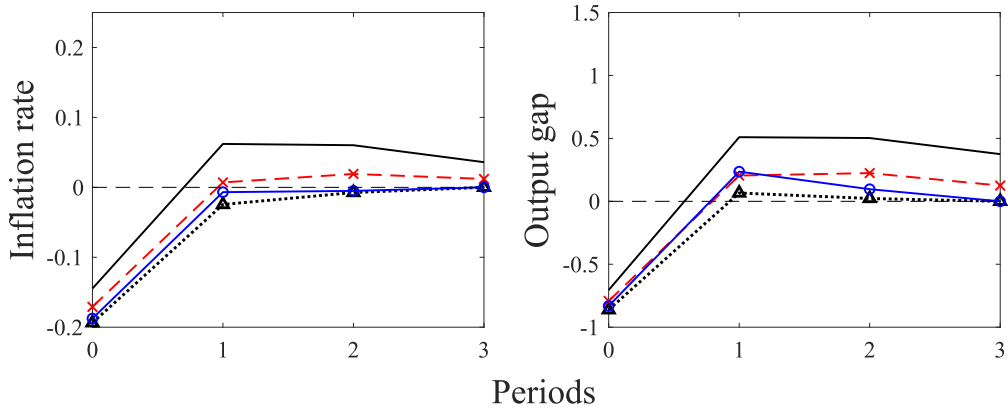
Figure S1: Output gap and inflation rate under different policy regimes. (a) r_H is nearly an absorbing state: $p = \text{Prob}(r_t^n = r_L | r_{t-1}^n = r_H) = 0.01$. (b) There is a reasonable probability of the natural rate returning to r_L from r_H: $p = 0.4$.
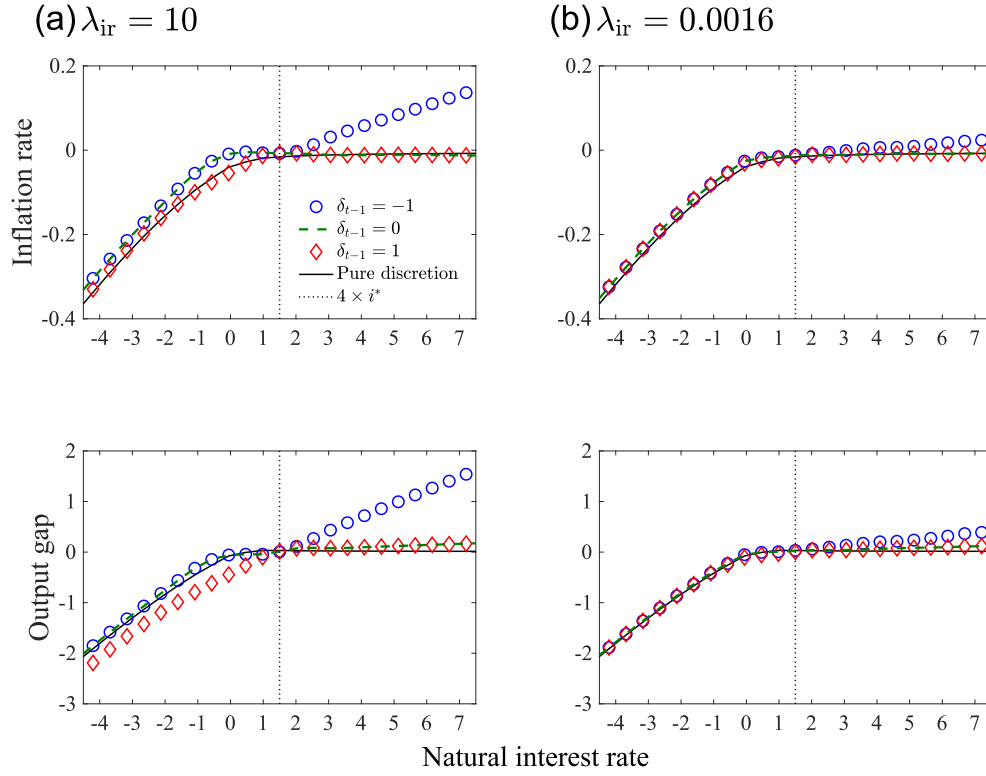
(a) $\lambda_{\mathrm{ir}} = 10$

(b) $\lambda_{\mathrm{ir}} = 0.0016$

Inflation rate

Output gap

Natural interest rate

$\delta_{t-1} = -1$
$\delta_{t-1} = 0$
$\delta_{t-1} = 1$
Pure discretion
$4 \times i^*$

Figure S2: Policy functions for inflation and output gap. Black solid line denotes a pure discretionary policy without reversal aversion (i.e., $\lambda_{\mathrm{ir}} = 0$). Blue circle, green dashed and red diamond respectively denote irreversible policies when the sign of policy shift in the previous period is -1, 0 and 1. We set $i_{t-1} = i^* = 1.5/4$ (vertical dotted line).
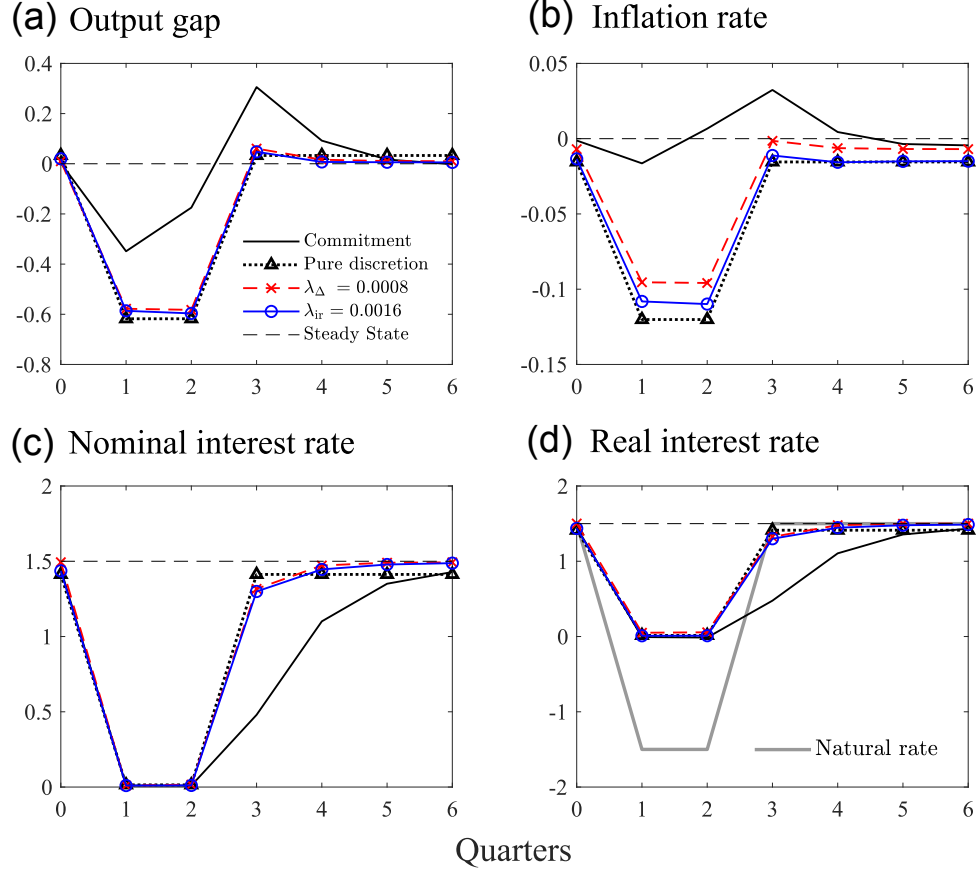
Figure S3: Impulse responses to a persistent -3% natural rate shock in the infinite-horizon model. We set $\rho_r = 0$ to remove exogenous policy inertia.
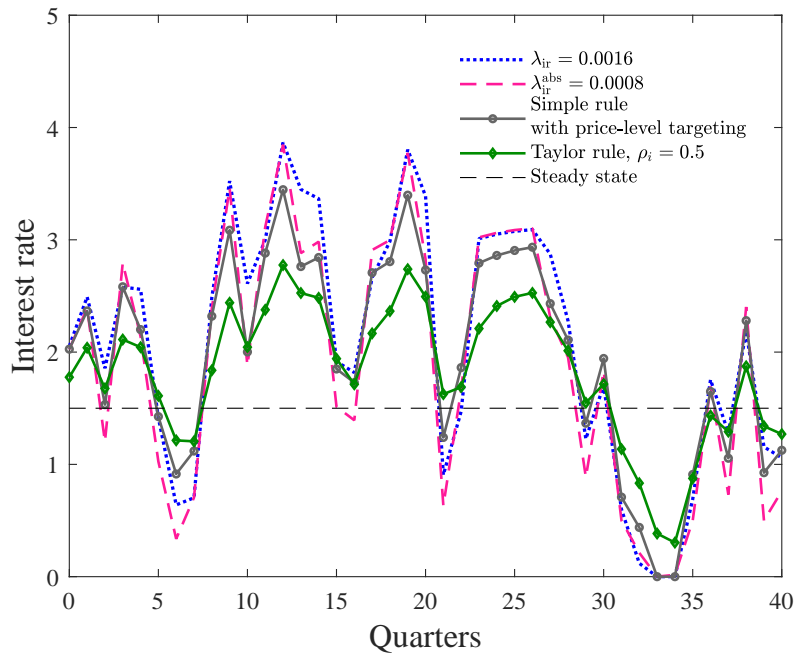
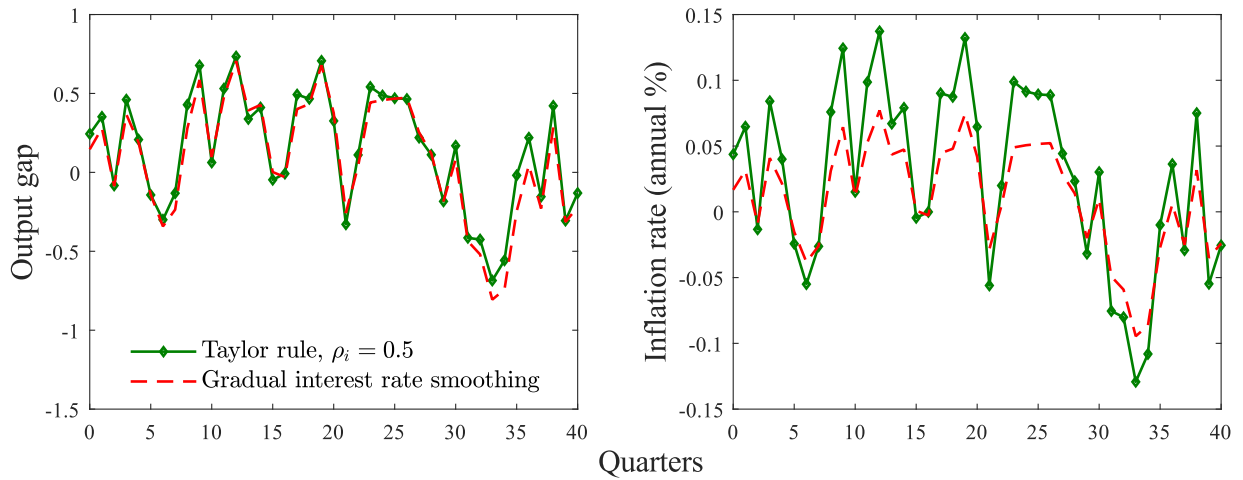Figure S4: Simulated paths of the nominal interest rates under alternative policy regimes.

Figure S5: Simulated paths of the output gap and inflation rate under the Taylor rule with $\rho_i = 0.5$ and gradual interest rate smoothing.