INTERNATIONAL MONETARY FUND

The Consequences of Falling Behind the Curve:

Inflation Shocks and Policy Delays under Rational and Behavioral Expectations

Mai Hakamada and Carl E. Walsh

WP/24/42

2024
MAR

WORKING PAPER

 WP/24/42

IMF Working Paper
Research Department

The Consequences of Falling Behind the Curve: Inflation Shocks and Policy Delays under Rational and Behavioral Expectations

Prepared by Mai Hakamada and Carl E. Walsh*

Authorized for distribution by Maria Soledad Martinez Peria
March 2024

ABSTRACT: Central banks in major industrialized economies were slow to react to the surge in inflation that began in early 2021. The proximate causes of this surge were the supply chain disruptions associated with the easing of COVID restrictions, fiscal policies designed to cushion the economic impact of COVID, and the impact on commodity prices and supply chains of the war in Ukraine. We investigate the consequences of policy delay in responding to inflation shocks. First, using a simple three-period model, we show how policy delay worsens inflation outcomes, but can mitigate or even reverse the output decline that occurs when policy responds without delay. Then, using a calibrated new Keynesian framework and two measures of loss that incorporate a "balanced-approach" to weigh inflation and the output gap, we find that loss is monotonically increasing in the length of the delay. Loss is reduced if policy, when it does react, is more aggressive. To investigate whether these results are sensitive to the assumption of rational expectations, we consider cognitive discounting as an alternative assumption about expectations. With cognitive discounting, forward guidance is less powerful and results in a reduction in the costs of delay. Under either assumption about expectations, the costs of a short delay can be eliminated by adopting a less inertial policy rule and a more aggressive response to inflation.

JEL Classification Numbers:	E31, E51, E52, E58, E61
Keywords:	monetary policy; inflation; policy delay; behavioral expectations; falling behind the curve
Author's E-Mail Address:	mhakamada@imf.org, walshc@ucsc.edu

* The authors would like to thank Athanasios Orphanides, Maria Soledad Martinez Peria, and the participants at the seminar organized by the IMF Research Department and Banco Central do Brasil for their comments.

WORKING PAPERS

The Consequences of Falling Behind the Curve:

Inflation Shocks and Policy Delays under Rational and Behavioral Expectations

Prepared by Mai Hakamada and Carl E. Walsh

1 Introduction

2021 provided vivid evidence of how large shocks to aggregate demand and supply can lead to sharp surges in inflation; Figure 1 illustrates the jump in CPI inflation experienced by the US, countries in the Euro area, Japan, and the UK. The belief that the rise in inflation was due to temporary factors, including reopenings after COVID shutdowns, supply chain disruptions, and large fiscal expansions designed to mitigate the impact of COVID, the war in Ukraine, and energy price spikes, combined with the belief that ten years of low inflation had firmly anchored inflation expectations, produced a view among central banks that inflation would quickly return to pre-COVID levels. As a consequence, many central banks, including the Federal Reserve, the Bank of England and the European Central Bank, delayed responding as inflation rose well above the banks' official targets. In the U.S., the Federal Reserve only moved to raise its policy rate in March 2022, a full year after PCE inflation had breached its 2-percent target.

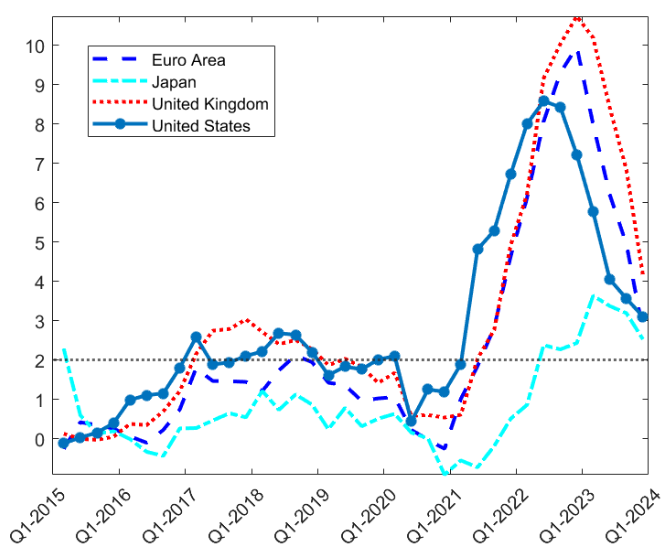

Figure 1: CPI inflation across countries (year-over-year percentage change. Source: IMF IFS, FRED.)

The inflation surge persisted for much longer than central banks expected. For the U.S. case, Figure 2 shows PCE inflation projections from the December and June meetings of the Federal Reserve Open Market Committee (FOMC), beginning in December 2020 and ending in December 2023. In Dec. 2020, the FOMC was projecting inflation at the end of 2021 would be less than its 2 percent target. By June 2021, this estimate had risen to 3.4. The actual year-over-year inflation for 2021 was 6.2 percent. The FOMC was still projecting inflation to return to roughly 2 percent by the end of 2022; it ended up equaling 5.4 percent by the end of 2022. The sequence of projections until June of 2023 showed a consistent upward shift at each new meeting. These projections reveal underestimates of inflation and overestimates of the speed with which inflation would decline. The initial 2021 surge in inflation persisted to a degree that policymakers failed to foresee.

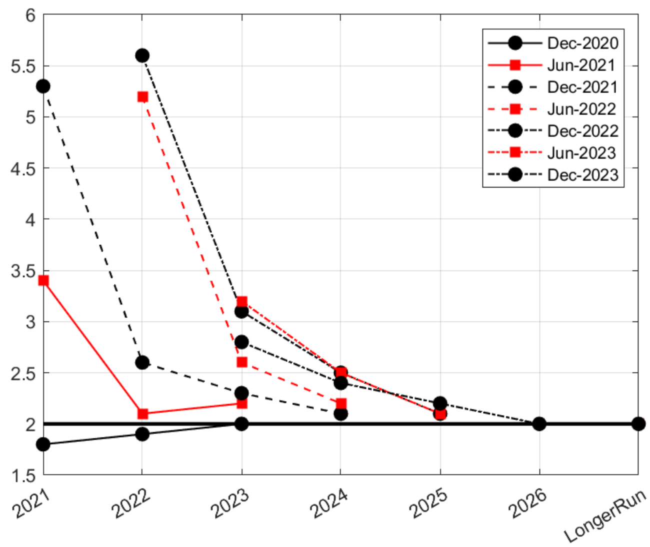

Figure 2: FOMC projections for 4-quarter PCE inflation at different meeting dates (Source: FOMC Summary of Economic Projections, various meeting minutes)

In this paper, we evaluate the costs and benefits of waiting to react in the face of a surge in inflation caused by a transitory shock. Standard analyses of monetary policy assume that the monetary authority reacts immediately when a shock to inflation occurs. The failure to react in 2021 raised the specter of a return to the high inflation of the 1960s and 1970s when disruptions in energy markets created surges in inflation and declines in economic activity – the era of "stagflation". Figure 3 shows U.S. inflation and unemployment from 1960 until 1985.[1] The two large oil price shocks in the 1970s generated increases in inflation that are clearly visible in the figure. Even though core PCE excludes food and energy prices, both inflation measures paint a similar picture. The red line shows a linear trend estimated using PCE inflation between 1960 and 1981 as inflation drifted upwards from around 2 percent to almost 10 percent.

[1] Inflation is measured by both the PCE price index and core PCE (PCE less food and energy); the unemployment rate gap is measured by civilian unemployment less the CBO's estimate of the natural rate of unemployment.

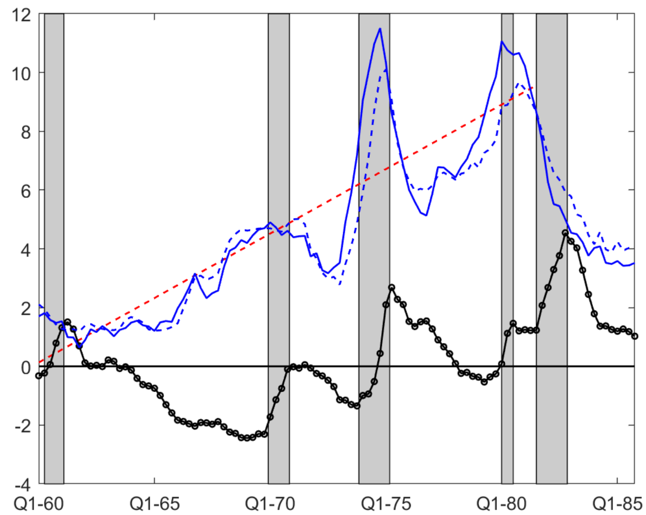

Figure 3: U.S. PCE inflation (year-over-year percentage change), core-PCE inflation (year-over-year percentage change), and the civilian unemployment rate (percent): 1960-1985.

In hindsight, the major policy error during the 1970s was not the volatility of inflation caused by oil shocks. It was allowing inflation to have a decades-long upward trend. Only with a major shift in monetary policy did countries bring inflation back to low levels, but the cost of doing so was reflected in significant recessions. In the case of the U.S., disinflation was associated with back-to-back recessions in 1980 and 1982. Central bankers learned much from this earlier experience. And unlike the post-financial crisis decade of zero nominal interest rates during which many central banks found themselves in an unfamiliar environment (Japan being the exception here), by the time of the recent inflation surge, central bankers should have been well-schooled in how to deal with a resurgence of inflation. Yet the slow reaction to inflation during 2021 raises several important questions: How costly is delay when fighting a surge in inflation? If a policy is delayed, should it then be more aggressive? Is a credible promise to eventually fight inflation a substitute for undertaking an immediate policy action?

We employ a new Keynesian model with endogenous sources of persistence to address these questions. Monetary policy is modelled by an inertial Taylor rule and agents hold rational expectations. We model the inflation surge of 2021 as the result of a negative innovation to the economy's flexible-price output that pushes up firms' real marginal costs. This is clearly a stylized representation of the factors, which included disruptions to supply chains, to energy markets, and fiscal policies, that played out during 2021 and 2022. We assume the shock to inflation is persistent but transitory; the policies we consider, which all satisfy the Taylor principle, ensure inflation will eventually revert to its steady-state level. Our approach neglects the COVID-related fiscal contributions to inflation. This reflects our main objective, which is not to model the specifics of recent inflation experiences but to address the more general issue of how a failure by the monetary authority to react promptly to a jump in inflation affects the subsequent paths of inflation and the output gap. For this same reason, we ignore issues associated with the zero lower bound on the nominal policy rate. Our experiment, consisting of a positive and persistent shock to inflation, never results in the policy rate dipping below its initial value.

We show that waiting to respond, as central banks did in 2021 and early 2022, can amplify the rise in inflation, but simultaneously it can boost output and dampen any subsequent recession. A metric is needed to evaluate the net effect of these different impacts on economic activity and inflation. We employ two measures to rank outcomes. One is based on a standard quadratic loss function. The second is based on the notion that, for a given disinflationary path for inflation, policymakers prefer paths that avoid recessions or even boost output, in keeping with the desire for a "soft landing".

We find that on both measures, delay is always costly, and the marginal cost of additional delay is increasing. Importantly, we find that adopting a more aggressive response to inflation in the sense of implementing a rule with a large coefficient on inflation reduces loss, as does adopting a less inertial policy rule. Conditional on delay, loss is lowest if policy reacts with both a stronger response to inflation and with less inertia.

Because expectations of future policy actions play an important role, a credible promise to respond in the future can serve as a substitute for an immediate reaction. We assess the consequences of policy delay when agents' expectations systematically differ from full information rational expectations.

To evaluate the consequences of non-rational expectations, we follow the behavioral-based approach of Gabaix (2014; 2020). He uses both micro and macro empirical findings to argue for cognitive discounting – the systematic tendency for agents to discount more heavily the future relative to what would occur under rational expectations. Pfäuti and Seyrich (2022) provide evidence of agents' underreaction due to cognitive discounting by estimating the discounting parameter using US consumer survey data. Hirose and Iiboshi (2022) estimate a New Keynesian model that incorporates bounded rationality and the effective lower bound on the nominal interest rate. Their Bayesian estimation results find their model with bounded rationality fits US data better than the version of the model that assumes rational expectations. Ilabaca, Meggiorini, and Milani (2020) also estimate a new Keynesian model that incorporates behavioral expectations in the form of cognitive discounting. They find a substantial degree of bounded rationality and estimate that this helped prevent the pre-Volcker economy from falling into indeterminacy. Andrade, Coredeiro, and Lambais (2019) integrated cognitive discounting into the new Keynesian framework and provide estimates that offer support for the presence of such discounting.

Cognitive discounting reduces the power of forward guidance. Both inflation and the output gap are less volatile when agents display cognitive discounting. However, the general conclusions obtained under rational expectations still hold. Both measures of loss are lowest when there is no delay in responding, and conditional on delay, loss is lower when policy reacts more aggressively.

Our paper is part of the large literature that evaluates the role of monetary policy rules in affecting the dynamic response of the economy to exogenous shocks. This literature has played a major role in the design and evaluation of monetary policy in the 30 years since the seminal work of Taylor (1993). This research has investigated the implications of adding or replacing variables in the original Taylor rule, estimating or optimizing the coefficients in the rule, and comparing non-inertial versus inertial rules. However, this work has assumed that the monetary authority reacts immediately to movements in the endogenous variables that appear in the rule. An exception is Walsh (2022), who investigated the consequences of a delayed policy reaction to a persistent inflation shock. Walsh's analysis had three important limitations. First, delay was modeled as involving both a failure to react when the shock occurs and backward-looking policy when the monetary authority finally did react. We instead assume that once the monetary authority does react, it does so by responding to current inflation and the output gap. Second, he only examined the case of rational expectations. Beliefs that the policy authority will respond in the future can have large effects immediately through the power of forward guidance. We evaluate the consequences of deviations from rational expectations to investigate how the costs of delay are effected when forward guidance is less powerful. Third, delay can affect the behavior of inflation and unemployment differently, but Walsh did not provide a metric which could be used to assess the costs of policy delay. We adopt two measures of loss that allow us to evaluate the costs of delay.

The cognitive discounting approach has been used by Budianto, Nakata, and Schmidt (2023) to evaluate average inflation targeting relative to inflation targeting. They find that while price level targeting dominates finite-period average inflation targeting under rational expectations, this is no longer the case with boundedly rational expectations, unless the deviation of expectations from rational expectations is small. If such deviations are large, inflation targeting can perform better than either price level of average inflation targeting. Nakata, Ogaki, Schmidt, and Yoo (2019) show that forward guidance at the effective lower bound is less effective under cognitive discounting and that this implies it is optimal for the central bank to keep the policy rate at the effective bound for longer.

The organization of the rest of the paper is as follows. In section 2, we use a simple three-period model to illustrate how policy delay has opposing effects on the paths of inflation and the output gap. Section 3 develops the basic model we use to analyze policy delay. The model incorporates indexation to trend inflation, habits in consumption and inertia in the policy rule. In section 4 we compare outcomes when these beliefs are formed under rational expectations. To investigate how sensitive these results are to the rational expectations assumption, we then consider a behavioral approach to expectations formation in section 5. Conclusions are summarized in section 7.[2]

2 A simple example

In this section, we set out a simple, three-period example designed to illustrate the effects of policy delay. The economy is characterized by an Euler condition and a new Keynesian Phillips curve. Let π_i and x_i denote inflation and the output gap in period i, $i = 1, 2, 3$. The economy experiences a positive inflation shock $e_1 > 0$ in period 1; the shock persists but decays and equals $e_2 = \rho e_1$, $0 \leq \rho < 1$, in period 2. Note that ρ can also be interpreted as the probability that the shock in period 2 is either e_1 or 0. There are no shocks in period 3, and $\pi_3 = x_3 = 0$.

In period 2, the equilibrium conditions are

$$\pi_2 = \beta \mathrm{E}_2 \pi_3 + \kappa x_2 + e_2 = \kappa x_2 + e_2,$$

$$x_2 = \mathrm{E}_2 x_3 - \frac{1}{\sigma}\left(i_2 - \mathrm{E}_2 \pi_3\right) = -\frac{1}{\sigma} i_2,$$

and

$$i_2 = \gamma_2 \phi \pi_2.$$

The last equation is the policy rule in which ϕ is the response to inflation, conditional on the central bank responding, and γ_2 is an index variable that equals 1 if policy responds in period 2 and 0 otherwise. Note that the assumption that $\pi_3 = x_3 = 0$ has been used to set $\mathrm{E}_2\pi_3 = \mathrm{E}_2 x_3 = 0$. It follows that

$$\pi_2 = -\frac{\gamma_2 \kappa \phi_2}{\sigma}\pi_2 + e_2 \Rightarrow 0 < \left(\frac{\sigma}{\sigma + \gamma_2 \kappa \phi}\right) e_2 \leq e_2 \tag{1}$$

and

$$x_2 = -\frac{1}{\sigma} i_2 = -\frac{\gamma_2 \phi}{\sigma}\pi_2 \Rightarrow -e_2 \leq -\left(\frac{\gamma_2 \phi}{\sigma + \gamma_2 \kappa \phi}\right) e_2 \leq 0 < 0. \tag{2}$$

When policy does act, inflation is partially stabilized and rises less than e_2, while output is destabilized as a consequence. Note that if $\gamma_2 = 0$ (no policy response in period 2), $\pi_2 = e_2$ and $x_2 = 0$.

In period 1, equilibrium is determined by

$$\pi_1 = \beta \mathrm{E}_1 \pi_2 + \kappa x_1 + e_1,$$

and

$$x_1 = \mathrm{E}_1 x_2 - \frac{1}{\sigma}\left(\gamma_1 \phi \pi_1 - \mathrm{E}_1 \pi_2\right).$$

The parameter γ_1 is an index that equals 0 if policy does not respond in period 1 and 1 if it does.

[2] The appendix provides details on derivations and robustness results.

Using (1) and (2) in the period 1 equilibrium conditions yields algebraic expressions for π_1 and x_1. See Appendix 1 for details.

We assume that if the central bank responds in period 1, it also responds in period 2. That leaves three alternative policy specifications to consider. First, with $\gamma_1 = \gamma_2 = 0$, the central bank does not respond in either period 1 or 2. Second, with $\gamma_1 = 0$ but $\gamma_2 = 1$, the central bank delays until period 2 but does respond at that point. Finally, the third case to consider is when $\gamma_1 = \gamma_2 = 1$; policy responds without any delay. Let $\pi_s(\lambda_1, \lambda_2)$ and $x_s(\lambda_1, \lambda_2)$ denote the equilibrium values of inflation and the output gap in period s as a function of the policy choices. The appendix shows that

$$\begin{aligned}
\pi_1(0,0) &= \left[1 + \rho_e \left(\beta + \frac{\kappa}{\sigma}\right)\right] e_1, \\
\pi_1(0,1) &= \left[1 + \rho_e \left(\beta + \frac{\kappa}{\sigma}(1-\phi)\right)\right] e_1, \\
\pi_1(1,1) &= \delta \left[1 + \delta\rho_e \left(\beta + \frac{\kappa}{\sigma}(1-\phi)\right)\right] e_1,
\end{aligned}$$

where

$$\delta = \left(\frac{\sigma}{\sigma + \kappa\phi}\right) \leq 1.$$

it follows that

$$\pi_1(0,0) > \pi_1(0,1) > \pi_1(1,1),$$

and the rise in inflation in period 1 is increasing in delay.

For period 2,

$$\begin{aligned}
\pi_2(0,0) &= \rho_e e_1, \\
\pi_2(0,1) &= \delta\rho_e e_1 = \delta\pi_2(0,0) < \pi_2(0,0), \\
\pi_2(1,1) &= \delta\rho_e e_1 = \delta\pi_2(0,0) < \pi_2(0,0),
\end{aligned}$$

implying

$$\pi_2(0,0) > \pi_2(0,1) = \pi_2(1,1).$$

The rise in period 2 is also increasing in delay. Delay is detrimental to controlling inflation.

For the output gap in period 1,

$$x_1(0,0) = \frac{\rho_e}{\sigma} e_1 > 0,$$

$$x_1(0,1) = \frac{\rho_e}{\sigma}(1-\phi) e_1 < 0 < x_1(0,0),$$

and

$$x_1(1,1) = \delta\left[-\frac{\phi}{\sigma} + \delta\frac{\rho_e}{\sigma}(1-\phi-\beta\phi)\right]e_1 = \delta^2 x_1(0,1) - \delta\frac{\phi}{\sigma}(1+\delta\rho_e\beta)\,e_1 < x_1(0,1) < 0.$$

For period 2,

$$x_2(0,0) = 0,$$
$$x_2(0,1) = -\delta\frac{\rho_e\phi}{\sigma}e_1 < 0 = x_2(0,0),$$
$$x_2(1,1) = x_2(0,1) < x_2(0,0).$$

Thus, the absence of a policy reaction ($\lambda_1 = \lambda_2 = 0$) leads to an output expansion in the first period due to the fall in the real interest rate caused by the rise in expected future (i.e., period 2) inflation coupled with the lack of a policy response. The output gap then returns to zero in the second period. When policy delays buts then does respond in period 2, output in the second period falls as policy responds more than one-to-one to the rise in inflation ($\phi > 1$). This, in turn, causes output to fall in period 1. Absent any delay ($\lambda_1 = \lambda_2 = 1$), output falls in period 1 by more than when $\lambda_1 = \lambda_2 = 0$ or when $\lambda_1 = 0$ and $\lambda_2 = 0$, while output in period 2 is independent of any period 1 policy response and declines the least if policy does not respond in period 2.

This simple example suggests that the consequences of delay depend, in a significant way, on the persistence of the shock, as measured by ρ_e, and on the strength of the policy reaction, as measured by ϕ. For example,

To illustrate the properties of the equilibrium and the consequences of policy delay, we numerically evaluate the model. We chose standard parameter values for the model. Specifically, $\beta = 0.995$, $\kappa = 0.17$, $\sigma = 1$, and $\phi = 2$. We consider a one-unit inflation shock, $e_1 = 1$, and set $\rho = 0.9$.

Figure 4 shows the outcomes for the policy interest rate (left panel), inflation (center panel), and the output gap (right panel). The standard analysis would have policy respond when the shock first occurs in period 1; this is shown by the black (crosses) line in the left panel ($\gamma_1 = 1$, $\gamma_2 = 1$). The policy interest rate is increased in response to the positive inflation shock. Since the shock persists into period 2 ($\rho = 0.9$), i_2 remains high. When the shock disappears in period 3, $i_3 = 0$. As a consequence, the center panel shows that this policy does a good job of stabilizing inflation. However, it does so at the cost of a sharp drop in the output gap, which remains negative in period 2.

Next, consider a policy that delays one period before responding to the shock ($\gamma_1 = 0$, $\gamma_2 = 1$). This scenario is denoted by red (diamonds). Because this rate hike in period 2 is anticipated in period 1, it succeeds in limiting the rise in π_1 through the impact of expected period 2 inflation on π_1. Inflation behaves similarly to the $\gamma_1 = \gamma_2 = 1$ case. The output gap does as well, though the

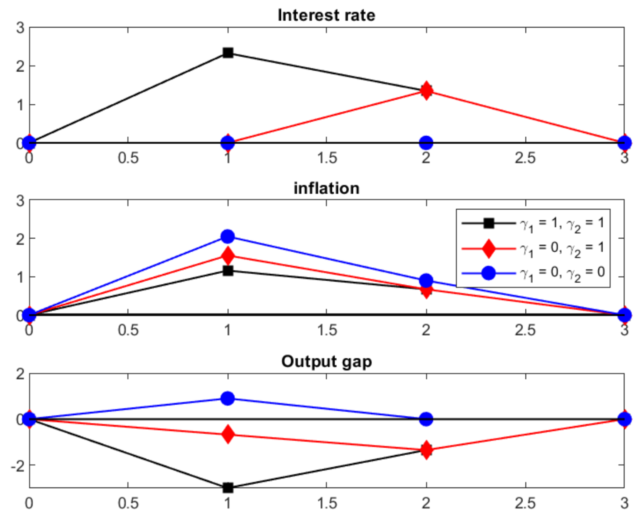

Figure 4: Three period example: shock e_1 occurs in period 1 and ρe_1 in period 2. If $\gamma = 0$ ($q = 0$), there is no policy response in period 1 (period 2); if $\gamma = 1$ ($q = 1$), policy does respond in period 1 (period 2).

drop in output in period 2 is somewhat smaller.

Finally, the blue (circles) lines show outcomes in the face of no policy response in either period 1 or 2 ($\gamma_1 = 0$, $\gamma_2 = 0$). Inflation is much more volatile than in the other two cases. However, the output gap rises in period 1 rather than falls, and has returned to its steady state by period 2. This result is due to the decline of the real interest rate when the policy rate is unchanged while inflation falls.

We take away two conclusions from this simple example. First, delay affects inflation and the output gap differently. Greater delay causes a larger rise in inflation, while it leads to a smaller decline and potentially even a rise in the output gap. Second, if policy outcomes are evaluated using a standard quadratic loss function in inflation and output gap volatility, the loss associated with policy delay will depend critically on the relative weight placed on output gap volatility in the loss function. If it is large, delay may be optimal; if it is small, delay will be costly.

We next turn to a more fully specified NK model to see whether the basic insights from this simple example continue to hold.

3 A NK model with indexation and habit formation

We work with a calibrated Keynesian model incorporating habits in consumption choice and indexation to the central bank's inflation target in firm pricing decisions. This latter assumption implies transitory inflation shocks will not affect trend inflation as long as the policy rule ensures a unique determinate equilibrium.[3]. Key components of the model are an intertemporal optimality condition derived from household consumption choice, a new Keynesian Phillips curve, and a rule specifying the behavior of monetary policy. Because we will be considering both rational and non-rational expectations, we use $\tilde{\mathrm{E}}$ to denote a general expectations operator, while E denotes rational expectations.

The demand side of the economy is characterized by a log-linearized Euler equation consistent with the presence of external habits in consumption. The preferences of household i are given by

$$U_t^i = \nu_t \left[\frac{\left(C_t^i - hC_{t-1}\right)^{1-\sigma}}{1-\sigma} - \chi_t \frac{\left(N_t^i\right)^{1+\eta}}{1+\eta} \right],$$

where ν_t and χ_t are taste shocks, C_t^i is the standard Dixit-Stiglitz basket of differentiated finals goods and N_t represents hours of labor supplied. Habits are external, depending on aggregate consumption C_t, though in equilibrium, $C_t^i = C_t$. Agents with higher h display a stronger effect of others' consumption on their own consumption choice. The household's problem is to choose C_{t+i}^i and N_{t+i}^i to maximize

$$\sum_{i=0}^{\infty} \beta^i \mathrm{E}_t U_{t+i}^i \left(C_{t+i}^i, C_{t+i-1}, N_{t+i}^i\right),$$

where $0 < \beta < 1$, subject to the budget constraint

$$W_t N_t^i + \left(1 + i_t^n\right) B_{t-1}^i = P_t C_t^i + B_t^i,$$

where P_t is the price of a unit of C_t^i, W_t is the nominal wage, i_t^n is the nominal interest rate on one period bonds, and B_t^i is the quantity of bonds held from t to $t+1$. Nominal wages are assumed to be flexible. The two optimality conditions implied by the household problem are the intertemporal condition

$$\nu_t \left(C_t^i - hC_{t-1}\right)^{-\sigma} = \beta \left(1 + i_t^n\right) \mathrm{E}_t \left(\frac{P_t}{P_{t+1}}\right) \nu_{t+1} \left(C_{t+1}^i - hC_t\right)^{-\sigma} \tag{3}$$

and the intratemporal optimal condition equating the real wage to the marginal rate of substitution between leisure and consumption given by

$$\frac{W_t}{P_t} = \frac{\chi_t \left(N_t^i\right)^{\eta}}{\left(C_t^i - hC_{t-1}\right)^{-\sigma}}. \tag{4}$$

[3]The capital stock and nominal wage stickiness are additional factors affecting dynamics in empirical DSGE models such as Christiano, Eichenbaum, and Evans (2005), Smets and Wouter (2007) Del Negro, Schorfheide, Smets and Wouter (2007), Justiniano, Primiceri, and Tambalotti (2013), and Cai, Del Negro, Herbst, Matlin, Sarfati, and Schorfheide (2021)

Define π_t as the deviation of inflation around steady-state inflation, π, and let $\imath_t = i_t^n - (\rho + \pi)$ be the deviation of the nominal interest rate around its steady-state, where $\rho = \beta^{-1} - 1$. For all other variables, we use a hat to denote log-deviation around steady state; for example, $\hat{c}_t \equiv \log(C_t/C)$, where C without a time subscript is the variable's non-stochastic steady-state value.

Taking log deviations around steady state and imposing the equilibrium condition that $\hat{c}_t^i = \hat{c}_t$ and goods clearing $\hat{c}_t = \hat{y}_t$, (3) becomes

$$\hat{y}_t = \left(\frac{1}{1+h}\right)\tilde{\mathrm{E}}_t\hat{y}_{t+1} + \left(\frac{h}{1+h}\right)\hat{y}_{t-1} - \frac{1}{\sigma}\left(\frac{1-h}{1+h}\right)\left(\imath_t - \tilde{\mathrm{E}}_t\pi_{t+1} + \mathrm{E}_t\hat{\nu}_{t+1} - \hat{\nu}_t\right). \tag{5}$$

Let $\hat{y}_t^f$ denote output (as a log-deviation around steady state) in a flex-price, zero inflation equilibrium when appropriate subsidies ensure an efficient steady state. Then (5) can be written in terms of the output gap $x_t \equiv \hat{y}_t - \hat{y}_t^f$ as

$$x_t = \left(\frac{1}{1+h}\right)\tilde{\mathrm{E}}_t x_{t+1} + \left(\frac{h}{1+h}\right)x_{t-1} - \frac{1}{\sigma}\left(\frac{1-h}{1+h}\right)\left(\imath_t - \tilde{\mathrm{E}}_t\pi_{t+1} - \hat{r}_t^*\right), \tag{6}$$

where

$$\hat{r}_t^* = -\left(\mathrm{E}_t\hat{\nu}_{t+1} - \hat{\nu}_t\right) - \sigma\left(\frac{1+h}{1-h}\right)\left[\Delta\hat{y}_t^f - \left(\frac{1}{1+h}\right)\tilde{\mathrm{E}}_t\Delta\hat{y}_{t+1}^f - \left(\frac{h}{1+h}\right)\Delta\hat{y}_{t-1}^f\right]. \tag{7}$$

We derive an explicit expression for $\hat{y}_t^f$ below as a function of lagged output, the labor taste ($\hat{\chi}_t$), and productivity ($\hat{z}_t$) shocks; see (10).

There is a continuum of firms, each of which is the monopoly supplier of a differentiated final consumption good. Firms operate with a common production technology. The output of the firm j is

$$Y_{j,t} = Z_t N_{j,t}^{1-\alpha}, \tag{8}$$

where Z_t is a stochastic, aggregate productivity factor. Firm j faces a downward-sloping demand curve with price elasticity ϵ and receives a subsidy at rate $\tau > 0$. With flexible prices, firm j would set its price $P_{j,t}$ as a markup $\epsilon/(\epsilon - 1) > 1$ over nominal marginal cost $(1-\tau)P_t MC_{j,t}$. With all firms charging the same price, $P_t(j)$ equals the average price P_t and the markup is equal to one over real marginal cost. Real marginal cost is equal to the real wage divided by the marginal product of labor, or

$$MC_t = \frac{W_t/P_t}{Z_t N_t^{-\alpha}} = \frac{1}{1-\tau}\frac{\epsilon-1}{\epsilon}, \tag{9}$$

where we drop the j index as, with all firms setting the same price, employment is equalized across firms. Assume the subsidy τ is set at $1/\epsilon$ to offset the steady-state distortion due to the markup.

With flexible prices, the real wage is also equal to the marginal rate of substitution of the household given by (4). Setting $C_t = Y_t$, and letting a superscript f denote variables when prices and wages are flexible, the equilibrium real wage ω_t^f is given by

$$\omega_t^f = \frac{\chi_t \left(N_t^f\right)^\eta}{\left(Y_t^f - hY_{t-1}^f\right)^{-\sigma}} = Z_t N_t^{f-\alpha}.$$

Expressing this condition in terms of log deviations around the steady state and then using the production function (26), which implies to first-order $\hat{n}_t^f = (1-\alpha)^{-1}\left(y_t^f - \hat{z}_t\right)$, one obtains

$$\hat{y}_t^f = \left[\frac{\sigma}{1-h} + \frac{\alpha+\eta}{1-\alpha}\right]^{-1}\left[\left(\frac{\sigma h}{1-h}\right)\hat{y}_{t-1}^f + \left(\frac{1}{1-\alpha}\right)\hat{z}_t - \hat{\chi}_t\right]. \tag{10}$$

In a flexble-price, flexible-wage equilibrium, output follows an AR(1) process with innovations arising from productivity and labor supply shocks.[4]

When prices are sticky, we assume a standard Calvo adjustment process in which each period a randomly chosen fraction $1-\omega$ of firms optimally adjust their price, while the remaining fraction ω udaptes their price based on steady-state inflation.[5] The reduced form inflation adjustment equation takes the form

$$\pi_t = \beta\tilde{\mathrm{E}}_t\pi_{t+1} + \kappa\left(\widehat{mc}_t + \hat{e}_t\right), \tag{11}$$

where π_t is inflation, $\widehat{mc}_t$ is real marginal cost, and $\hat{e}_t$ is an inflation shock reflecting inefficiencies that result in stochastic fluctuations around the efficient steady-state, and

$$\kappa = \frac{(1-\omega)(1-\omega\beta)}{\omega}\frac{1-\alpha}{1-\alpha+\epsilon\alpha}.$$

In the expression for κ, $\epsilon > 1$ denotes the elasticity of substitution among varieties of goods and α denotes the degree of decreasing returns to labor in production. While we specify the inflation process to be consistent with nonzero trend inflation, we do not consider changes in trend inflation and therefore normalize trend inflation to equal zero: $\pi = 0$.

Average real marginal cost in log-deviation form is the real wage minus the marginal product of labor, or, using (26),

$$\widehat{mc}_t = \hat{\omega}_t - \hat{z}_t + \alpha\hat{n}_t. \tag{12}$$

The average marginal rate of substitution (MRS) between leisure and consumption depends on

[4]The process is stable; the coefficient on $\hat{y}_{t-1}^f$ is positive and less than one.

[5]See Cogley and Sbordone (2008, p. 2105). As surveyed in Ascari and Sbordone (2014), introducing non-zero trend inflation into a standard Calvo model of price adjustment without indexation (consistent with micro firm-level data) leads to a NKPC in which the reduced form coefficients are functions of trend inflation.

habits. Log-linearizing (4) yields

$$\hat{\omega}_t = \eta \hat{n}_t + \frac{\sigma}{1-h}\left(\hat{c}_t - h\hat{c}_{t-1}\right) + \hat{\chi}_t. \tag{13}$$

Using (13) in (12), together with the log-linearized production function, $\hat{n}_t = (\hat{y}_t - \hat{z}_t) / (1-\alpha)$ and goods market clearing, $\hat{c}_t = \hat{y}_t$, imply

$$\widehat{mc}_t = \left[\frac{\sigma}{1-h} + \frac{\alpha+\eta}{1-\alpha}\right]\hat{y}_t - \frac{\sigma h}{1-h}\hat{y}_{t-1} + \hat{\chi}_t - \left(\frac{1+\eta}{1-\alpha}\right)\hat{z}_t.$$

Combining these results with (10), one obtains

$$\pi_t = \beta \tilde{\mathrm{E}}_t \pi_{t+1} + \kappa\left[\left(\frac{\sigma}{1-h} + \frac{\alpha+\eta}{1-\alpha}\right)x_t - \left(\frac{\sigma h}{1-h}\right)x_{t-1} + \hat{e}_t\right]. \tag{14}$$

The final component of our baseline model is a specification of monetary policy. We allow for policy inertia ($0 < \rho < 1$) and policy delay in which policy responds to inflation and the output gap with a k-period delay. Letting t_0 denote the period in which the shock occurs. Policy is represented by an instrument rule of the form

$$i_t = \begin{cases} 0 \text{ for } t = t_0, t_0+1, ..., t_0+k-1 \\ \rho_i i_{t-1} + (1-\rho_i)\left(\phi_\pi \pi_t + \phi_x x_t\right) \text{ for } t \geq t_0 + k. \end{cases} \tag{15}$$

When $t = t_0 + k$, the monetary authority begins to implement an inertial Taylor rule.

To summarize, our basic model consists of equations (6), (7), (10), (14), and (15), plus the definition of the output gap as $x_t = \hat{y}_t - \hat{y}_t^f$ and the specification of the exogenous shocks. The model, though simple, contains aspects that will affect the dynamic response to shocks. Consumption habits, inertia in the policy rule, and serial correlation in the shock are fairly standard sources of persistence, One additional property affecting the model dynamics is not standard: delay, as distinct from inertia, in the reaction of policy to shock realizations.

3.1 Basic Calibration

Table 1 summarizes the parameters appearing in the model equations. We set the discount factor $\beta = 0.995$, implying a steady-state real interest rate ρ equal to 2% at an annual rate. From Galí (2015), $\sigma = 1$, $\eta = 5$ and $\alpha = 0.25$. Setting $\epsilon = 9$ implies a steady-state price markup of 12.5%. For the Calvo adjustment parameter, we set $\omega = 0.75$ which is consistent with an average duration of price spells of one year (four periods in the model). These values imply $\kappa = 0.169$ in (14).

In the policy rule (15), we set $\phi_\pi = 1.5$, $\phi_x = 0.5/4 = 0.125$, consistent with a basic Taylor rule. For policy inertia, we set $\rho_i = 0.85$, a common value in the literature. We allow policy delay to be zero to 6 periods, $k = \{0\ 2\ 4\ 6\}$. To represent a more aggressive response to inflation, we evaluate

Table 1: Baseline Parameters

β	0.995	Discount Rate	Galí (2015)
σ	1	Goods Elasticity	Galí (2015)
η	5	Frisch Elasticity of Labor	Galí (2015)
h	0.761	Habit Parameter	Afsar et al. (2023)
α	0.25	Capital Share	Galí (2015)
ϵ	9	Markup Parameter	Steady-state markup of 12.5%
ω	0.75	Calvo Parameter	Average duration of one year Galí (2015)
ρ_i	0.85	Persistence in the Policy Rule	Dennis (2009)
ϕ_π	1.5	Coefficient on Inflation in the Policy Rule	Galí (2015)
ϕ_x	0.5/4	Coefficient on the Output Gap in the Policy Rule	Galí (2015)

outcomes when ϕ_π is doubled from 1.5 to 3.0. As a second way to capture a more aggressive policy response to inflation, we also reduce policy inertia from 0.85 to 0.5. When $\phi_\pi = 1.5$, this increases the impact effect of inflation on the policy rate from 0.225 to 0.75.

We calibrate the variance of the inflation shock innovation to generate a peak rise of inflation of 9 percent when expressed as a deviation from the target and at an annual rate when policy delays four periods before reacting. This delay corresponds to the wait of a year before the Fed reacted. With the inflation target equal to 2 percent, this roughly captures the jump in inflation to over 11 percent that occurred in early 2022. We set the AR(1) coefficient for the inflation shock at 0.85.

Next, we present the responses of key variables to an inflation shock based on the calibration of Table 1. This is done first under the assumption of rational expectations and then for cognitive discounting.

4 Delay under rational expectations

In this section, we adopt the assumption of rational expectations: $\tilde{\mathrm{E}} = \tilde{\mathrm{E}}^{\mathrm{RE}}$. We define an inflation surge as an unexpected positive innovation to ϵ_t, the disturbance term in the inflation equation. The resulting impulse responses for the nominal interest rate (the policy rate), the real interest rate, the output gap, and inflation are shown in Figure 5. The results are shown for k = 0 (black circles), k = 1 (cyan squares), k = 2 (blue dashed diamonds), k = 3 (red asterisks), $k = 4$ (blue dashed diamonds), and $k = 5$ (black circles). Interest rates and inflation are reported at annual rates. Section A.2 of the appendix outlines the solution method to obtain the impulse responses as well as a link to the Matlab code employed to obtain the solutions.

The results for $k = 0$ in each panel serve as a reference, since standard analyses assume policy reacts contemporaneously to any shock. The policy rate, shown in the upper left panel, rises immediately and peaks two periods after the shock hits. However, the real interest rate falls (upper right panel) reflecting the jump in inflation (lower right panel) that is much larger than the rise in the nominal rate. Inflation peaks immediately, reaching just over 3 percent at an annual rate.

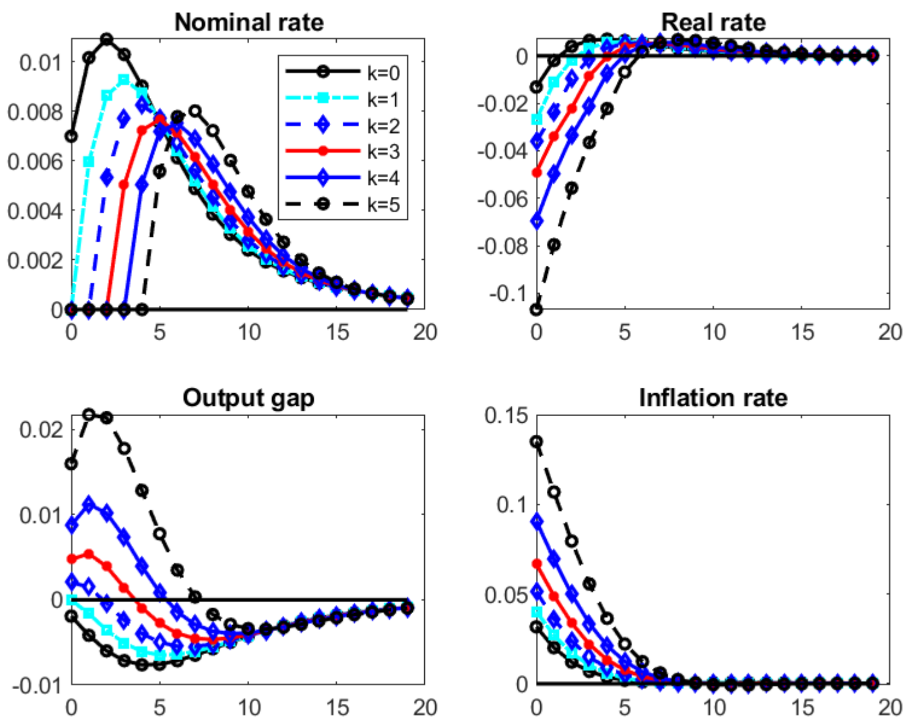

Figure 5: Impulse Responses to Positive Inflation Shock under Rational Expectations when policy delays k periods after the shock (interest rates and inflation at annual rates).

Inflation is less than 10 basis points above its steady state after seven periods. The output gap (lower left panel) initially declines, reaching a trough four periods after the shock before slowly returning to steady state from below. The rise in inflation is limited but at the cost of a recession.

Now consider the consequences of a delay in the response of policy. The upper left panel shows the path of the nominal rate when it is kept constant for k periods before responding. When the delay is more than 1 period, the output gap in the period of the shock *increases*, rather than decreases. As k increases, the increase in the output gap also rises at an increasing rate; the effect of delaying from 4 to 5 is much greater than the impact of increasing k from 3 to 4. A similar effect for inflation is seen in the lower right panel. Delaying four periods increases the peak rise in inflation from 3 percent (annual rate) when $k = 0$ to 9.1 percent. Delaying one additional period, to $k = 5$, raises peak inflation to 13.5 percent. Delay leads to a rise in the output gap and larger increases in the initial rise in inflation, results consistent with the simple example of section 2 and with figure 4.

The higher output gap as policy reacts with a longer delay contributes to explaining the higher path of inflation. The output gap rises because delay produces a fall in the real interest rate (upper left panel). This is also consistent with the marginal increases of the output gap in delayed policy in simple three-periods example in section 2. The output gap will depend on the entire path of expected future real interest rates. As can be seen in the figure, policy delay results in a larger and more persistent fall in the real interest rate. The lower future path of the real interest rate stimulates output in the period of the shock due to the forward-looking nature of demand.

In the section A.2 of the appendix, we show that the responses from $t+k$ forward depend only on the state at time $t+k-1$. A long delay by the monetary authority in reacting to the inflation shock leaves both the output gap and inflation higher at $t+k-1$, ensuring that inflation and the output gap persist in exceeding outcomes obtained when policy reacts immediately to the shock.

In section 4.2, we assess the costs of delay using a metric for evaluating alternative policies. Before doing so, notice that in Figure 5, the $k=3$ policy might almost be described as an immaculate disinflation. Inflation is higher than when delay is shorter, and the output gap is positive for a few periods before experiencing a milder recession than that achieved when policy reacts immediately.

4.1 Reacting more strongly to inflation

Does delay call for a stronger response to inflation once policy does act? This question was raised during 2022 as the Fed began lifting its policy rate, and the FOMC did embark on a series of rate increases that were larger than what had been its normal practice of moving in 25 basis point increments. While the first move in March 2022 was a 25 basis point increase, this was followed by a 50 basis point hike in May 2022. Four further increases of 75 basis points followed.[6]

We consider two interpretations of a more aggressive policy response. The first interpretation treats an aggressive policy as one that assigns a larger value to ϕ_π, the coefficient on inflation in the policy rule. The second interpretations views a more aggressive policy as one that assigns a smaller weight ρ_i to lagged inflation and that, therefore, displays less inertia. The contemporaneous response of the policy rate to inflation is equal to $(1-\rho_i)\phi_\pi$, which increases with a rise in ϕ_π, (i.e., a stronger response to inflation) or a fall in ρ_i, (i.e., less inertia).

We begin with the first form of aggressive policy by doubling the coefficient on inflation in the policy rule from 1.5 to 3.0. Results are shown in Figure 6. To reduce the clutter in the figure, we only show the results for $k=0$ and $k=4$. The solid lines for $k=0$ (black circles) and $k=4$ (blue diamonds) replicate those in the previous figure with $\phi_\pi = 1.5$. The dashed lines are obtained when $\phi_\pi = 3.0$. When $k=0$, reacting more strongly to inflation has the expected results; inflation is stabilized more, and output declines rather than increases.

When $k=0$, the stronger response to inflation causes the policy rate to peak at a higher level. The real interest rate is higher, leading to a larger fall in the output gap. As a consequence, inflation

[6] In January 2023, the funds rate target was raised by 50 basis points, and in March 2023, financial instability brought about in the wake of the failure of Silicon Valley Bank and Signature Bank led the FOMC to increase the funds rate target by only 25 basis points.

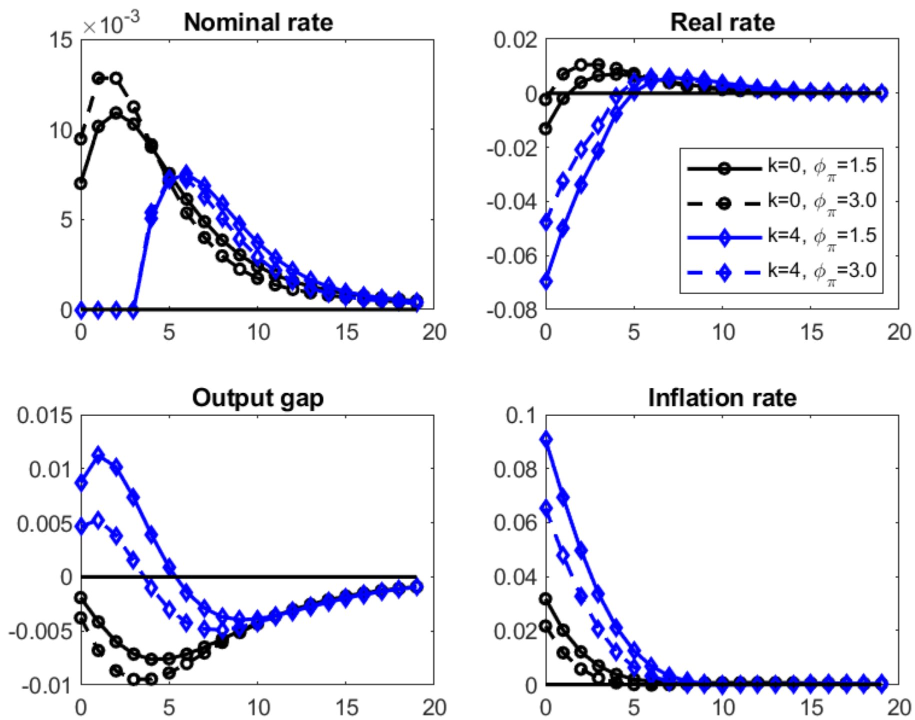

Figure 6: Impulse Responses to Positive Inflation Shock under Rational Expectations – Comparing $\phi_\pi = 1.5$ and $\phi_\pi = 3.0$

rises less. Lower inflation and a lower output gap reduce the upward pressure on the policy rate, leading the policy rate to fall and end up slightly lower than under the less aggressive response to inflation after $t + 5$.

When $k = 4$, the dampening effect on the output gap and inflation when $\phi_\pi = 3.0$ is apparent, even though the path of the nominal rate ends up little affected (see upper left panel). Lower inflation produces a smaller fall in the real interest rate and a smaller expansion in the output gap. The latter helps account for the lower path of inflation. Offsetting effects on the nominal policy rate leave the path of the nominal rate little affected. Conditional on delay, a more aggressive policy response, interpreted as a higher value for ϕ_π, results in lower output and inflation paths. Because both π and x are too high under the baseline value of ϕ_π, the more aggressive response moves both both closer to their steady-state values for several periods compared to the less aggressive response to inflation. Furthermore, the delayed case shows larger changes (toward the steady-state) with the larger value ofϕ_π, implying that the stronger policy responses under the delayed policy would help stabilize both inflation and the output gap.

Next, we consider the second form of aggressive policy by holding ϕ_π at its baseline value of 1.5

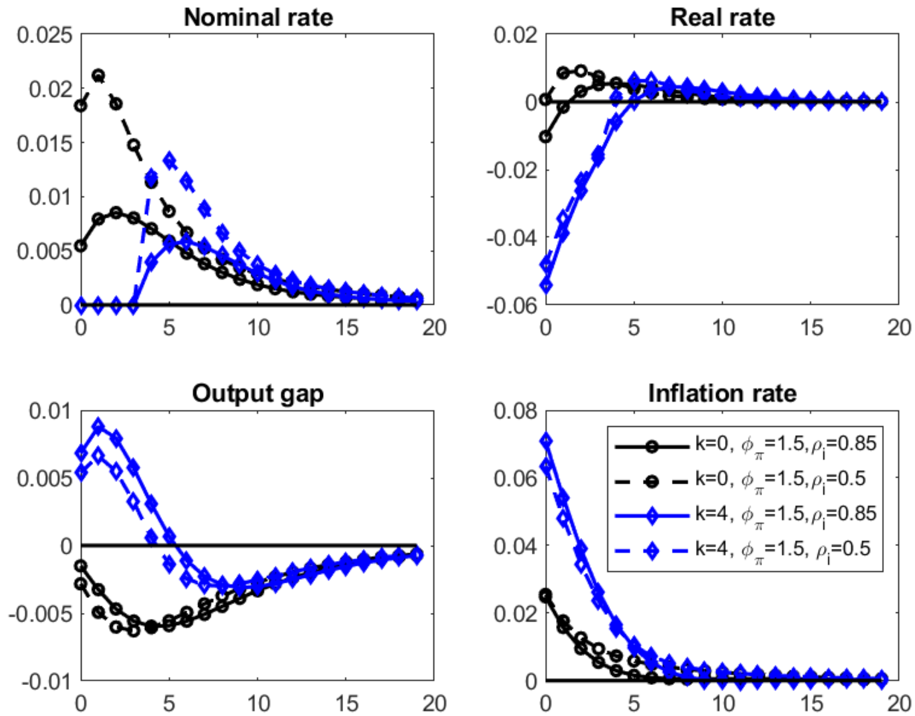

Figure 7: Impulse Responses to Positive Inflation Shock under Rational Expectations – Comparing $\rho_i = 0.85$ and $\rho_i = 0.5$ when $\phi_\pi = 1.5$

while reducing inertia by lowering ρ_i from 0.85 to 0.5. Results are shown in Figure 7. We again show results only for $k = 0$ and $k = 4$. Thus, the solid lines are the same as those in the previous figure. The dashed lines show impulse responses when $\rho_i = 0.5$. Starting with the no-delay case ($k = 0$), a reduction in policy inertia leads to a much larger rise in the nominal rate, both initially and throughout the convergence back to steady state. This is also the case when $k = 4$; less inertia front-loads the interest rate increase.

Absent any policy delay, the less inertial policy rule produces a higher real interest rate path and has a small effect in lowering the path of the output gap until $t + 5$. The effect on inflation of the less inertial policy is small.

Figure 9 illustrates the effects of less inertia combined with a stronger response to inflation. Setting $\phi_\pi = 3.0$, the figure compares response when $\rho_i = 0.85$, the baseline degree of inertia, to $\rho_i = 0.5$. When $k = 4$, this more aggressive policy dampens the rise in the output gap and in the inflation rate. However, the output gap then turns negative sooner and for longer than under a more inertial policy. Whether this represents an improvement will depend on how lower inflation is balanced against a deeper and more prolonged period during which the output gap is negative.

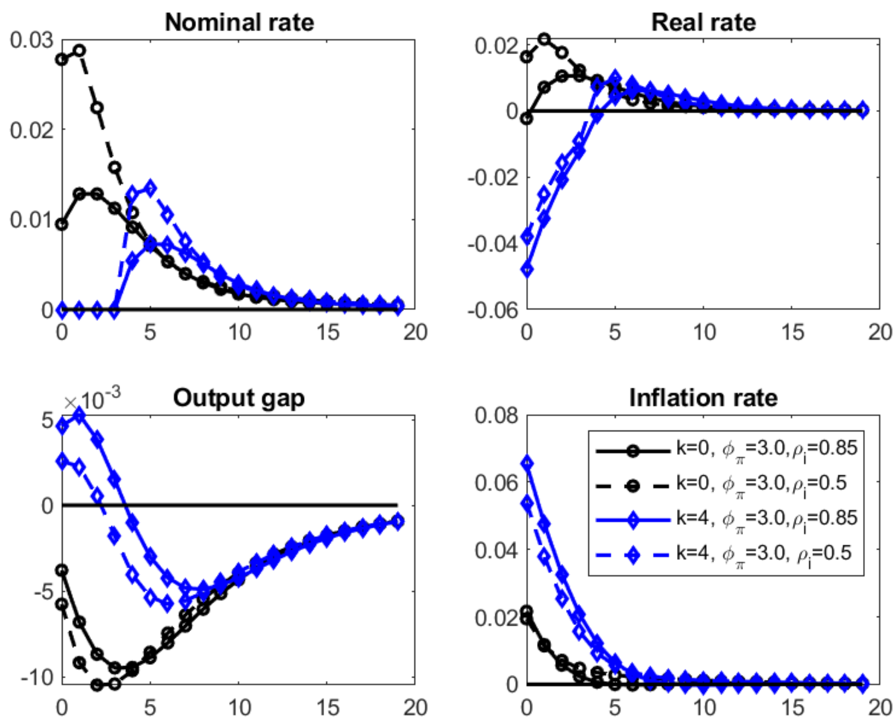

Figure 8: Impulse Responses to Positive Inflation Shock under Rational Expectations – Comparing $\rho_i = 0.85$ and $\rho_i = 0.5$ when $\phi_\pi = 3.0$

We return to this trade-off in section 4.2.

Under the baseline calibration of the policy rule, extending the delay in the policy reaction beyond 4 periods significantly increases the peak rise in the output gap and inflation. This is illustrated in Figure 9; the blue line (diamonds) shows the impulse response for $k = 4$ and corresponds to the similar line in Figure 6. When delay is increased further to $k = 6$, peak inflation rises from 9.1 percent to 24.6 percent. The peak output gap jumps from 0.9 percent to 3.4 percent. This implies that significantly delayed policy responses – more than one year of delay – can substantially increase the fluctuations of the output gap and inflation in the economy.

4.2 Measuring the costs of delay under rational expectations

To evaluate alternative policy rules and the consequences of delayed reaction to inflation surges, we need a metric with which to rank the results under different assumptions about k. One approach would be to employ a second-order approximation to the welfare of the representative agent, an approach pioneered by Woodford (2003) and common in the literature on optimal monetary policy. In a basic sticky price new Keynesian model with Calvo pricing but without habits in consumption

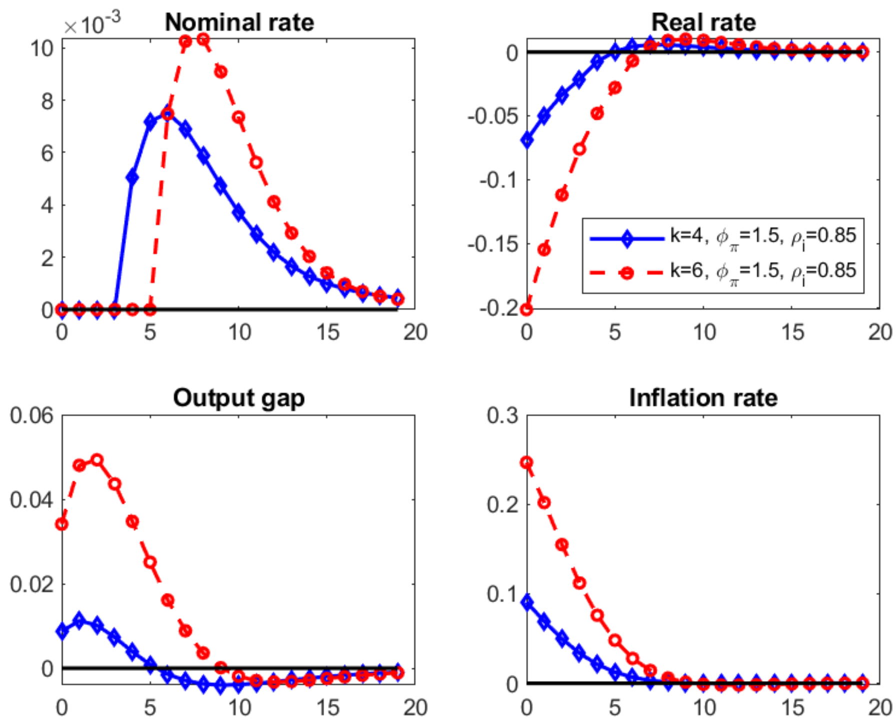

Figure 9: Impulse Responses to Positive Inflation Shock under Rational Expectations – $k = 4$ and 6 when $\phi_\pi = 3.0$ and $\rho_i = 0.5$

or price indexation, Woodford shows that the period-loss in such a welfare approximation depends on the volatility of inflation and the output gap. The simple example of section 2 indicated that delay may worsen inflation outcomes but improve output gap performance. Thus, any ranking will depend on the relative weight applied to the volatility of π and x. If this weight is chosen based on a second-order approximation to the welfare of the representative agent in a baseline NK model, this weight would be quite small; for example, under the calibration adopted by Woodford (2003), if the weight on inflation volatility is normalized to equal 1, the weight on the volatility of the output gap would be 0.048.[7]

However, the recent macroeconomic research on heterogeneity and the distributional impact of monetary policy has highlighted the limits of relying on the fiction of a representative agent to design aggregate economic policies. In addition, modern approaches to monetary policy implementation stress the importance of being transparent and communicating clearly about policy. This task becomes more difficult if a central bank attempts to balance conflicts between multiple conflicting

[7] Amato and Laubach (2004) and Leith, Moldovan, and Rossi (2012) derive welfare approximations in new Keynesian models characterized by habit persistence and sticky prices, components of the model we use. Both employ Rotemberg price adjustment models.

objectives. Fewer objectives simplify this communications challenge. For formal inflation targeting central banks, inflation is usually taken to be the primary objective, but, in practice, inflation targeters are flexible inflation targeters, implicitly if not explicitly incorporating a concern for stabilizing real economic activity as an appropriate objective.

Therefore, we choose to adopt an ad hoc loss function that involves inflation volatility and the volatility of an output gap measure. This leaves the choice of the relative weight to place on the two objectives to be determined. Debortoli, Kim, Lindé, and Nunes (2019) show that policies designed to minimize a simple loss function, such as

$$\mathrm{L} = \frac{1}{2}\left(\frac{\epsilon}{\kappa}\right)\left[var\left(\pi_t\right) + \lambda var\left(x_t\right)\right]. \tag{16}$$

can generate outcomes in empirically relevant models such as that of Smets and Wouter (2007) that compare favorably to fully-optimal policies based on maximizing the model-consistent welfare of the model's representative agent as long as λ is optimally chosen.[8] Achieving good outcomes using a simple loss function requires that more weight be placed on output stabilization than the welfare-approximation approach would typically imply. This suggests using a value for λ that exceeds Woodford's 0.048. For example, Debortoli, et al (2019, p. 2021) find that for the Smets and Wouter (2007) model, a policy with $\lambda = 0.25$ achieves a welfare loss that is one-third that obtained using $\lambda = 0.048$. As Debortoli, et al note, a weight of 0.25 on the output gap approximates the loss function Yellen (2012) characterized as reflecting a dual mandate, one with equal weights on output volatility and unemployment rate volatility. Assuming an Okun's Law coefficient of 2, $var\left(\pi_t\right) + var\left(u_t\right) = var\left(\pi_t\right) + 0.25 var\left(x_t\right)$. In this case, (16) becomes

$$\mathrm{L} = \frac{1}{2}\left(\frac{\epsilon}{\kappa}\right)\left[var\left(\pi_t\right) + 0.25 var\left(x_t\right)\right]. \tag{17}$$

Equation (17) provides a measure of loss when the economy is subject to new realizations of the shock processes each period. Because our focus is on how policy affects the dynamic adjustment of real activity and inflation after a *single* realization of the inflation shock innovation, ranking policies based on the performance of an instrument rule in a stochastic equilibrium in which new shocks occur each period may fail to capture the impact of an episode of surging inflation during which the monetary authorities waits as much as a year to react.

For that reason, we define two metrics that we use to rank outcomes. Each is based on the impulse responses from a single realization of the innovation to the inflation equation. The first metric is an extension of a balanced loss function in output gap and inflation volatility. It focuses

[8] The factor $(1/2)(\epsilon/\kappa)$ in (16) expresses loss as a fraction of steady-state consumption.

specifically on the paths of x and π in response to the inflation shock. This measure is defined as

$$\mathrm{L}_T^1(k, \phi_\pi, \rho_i) = \frac{1}{2}\left(\frac{\epsilon}{\kappa}\right)\mathrm{E}_t\left[\frac{1}{T}\sum_{j=0}^{T}\beta^j \pi_{t+j}^2 + \lambda\frac{1}{T}\sum_{j=0}^{T}\beta^j x_{t+j}^2\right], \tag{18}$$

where π_{t+j} and x_{t+j} are both functions of the parameters in the monetary policy rule that we will vary (i.e., ϕ_π, ρ_i, and k). L_T^1 is the cumulative average discounted squared deviation from steady state from the time of shock to T periods after the shock. It depends on the inflation and output gap realizations as functions of the policy rule parameters k, ϕ_π, and ρ_i, as well as the size of the shock innovation. In line with the balanced loss approach, we set $\lambda = 0.25$.

L_T^1 gives equal weight to positive and negative deviations from steady state. In 2022, policymakers expressed the desire to bring inflation down while also aiming for a "soft landing" or, even better, an "immaculate disinflation." We interpret these sentiments as reflecting a desire for a policy that avoids or minimizes negative output gaps while also reducing above-target inflation rates. The measure of loss given by (18), however, treats an immaculate disinflation, defined as a decline in inflation while maintaining a zero output gap as worse than the same disinflation accompanied by an economic expansion because x_t^2 is positive in the latter case. If the natural rate of output used to define the output gap is inefficiently low, policymakers should rank the latter outcome as better than the former.

This suggests our second measure, defined as

$$\mathrm{L}_T^2(k, \phi_\pi, \rho_i) = \frac{1}{2}\left(\frac{\epsilon}{\kappa}\right)\mathrm{E}_t\left[\frac{1}{T}\sum_{j=0}^{T}\beta^j \pi_{t+j} - \lambda\frac{1}{T}\sum_{j=0}^{T}\beta^j x_{t+j}\right]. \tag{19}$$

This measure equals the average cumulative discounted deviations of the level of inflation from steady state from t to $t+T$ minus the average cumulative discounted deviations over the same period of the output gap. It penalizes positive deviations of inflation and negative deviations of the output gap. Equation (19) is a simple way of allowing positive values of the output gap to lower the measure of loss. We employ $\lambda = 0.25$ for both (18) and (19).

Figure 10 reports loss as measured by (18) for four alternative specifications of the monetary policy rule. In each quadrant, the horizontal axis gives the number of periods since the shock (T), and each line represents L_T^1 for a different k. The upper left quadrant reports results for the baseline policy rule with $\phi_\pi = 1.5$ and $\rho_i = 0.85$. Values are shown for delay lags of $k = [0\ 1\ 3\ 4\ 5]$. The black solid line marked with circles is for $k = 0$, the case in which policy reacts immediately to the shock. In the period the shock occurs, volatility is lowest when policy reacts immediately and is then increasing with k. Delaying just one period before reacting ($k = 1$) leads to only a small increase in the loss measure. The marginal increase in loss as k rises from 3 to 5 is increasing. A

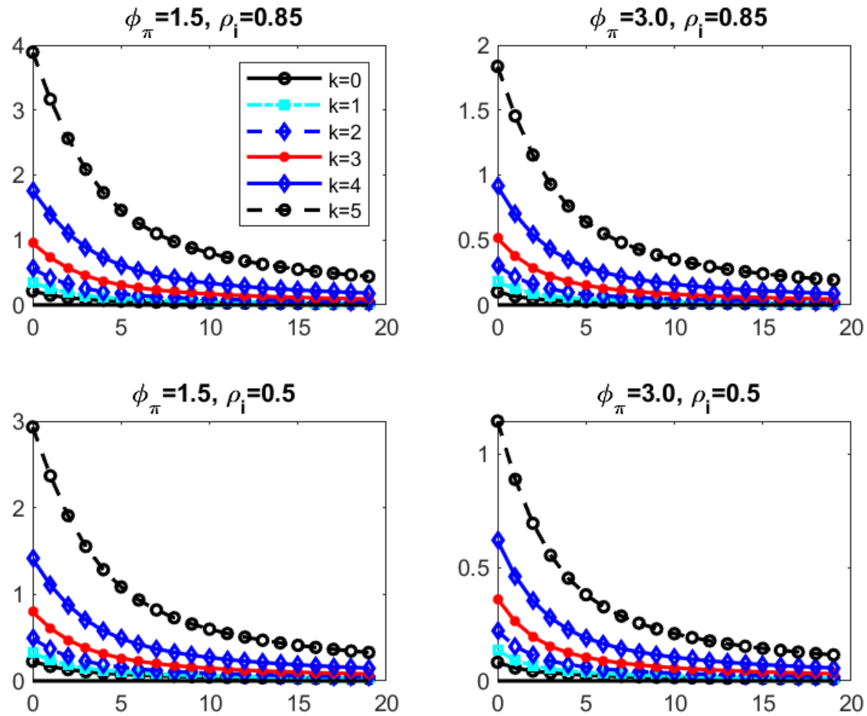

Figure 10: Average Volatility Defined by (18) under Rational Expectations: Periods since shock on horizontal axis

delay of more than 4 quarters causes a very significant deterioration in performance.

Are there advantages to adopting a more aggressive response? The upper right quadrant of Figure 10 illustrates the volatility measure given by (18) when ϕ_π is doubled from 1.5 to 3. Comparing the upper left and right quadrants reveals that reacting more strongly to inflation reduces overall loss for all k at each horizon T. Volatility again increases monotonically in k; at each horizon T, volatility is lowest when $k = 0$, and highest when $k = 5$.

The second notion of a more aggressive policy is one that displays less inertia. The lower two quadrants of Figure 10 illustrate the impact of adopting a less inertial policy rule, with ρ_i reduced from 0.85 to 0.5. Comparing a lower quadrant with the corresponding upper quadrant reveals that reducing inertia lowers volatility for all k and T. This is the case whether ϕ_π equals 1.5 or 3.0. As with the more inertial policy in the right upper panel, less inertia in the policy rule indicates that a very short delay ($k = 1$) roughly approximates outcomes when there is no delay. Longer delays significantly increase loss.

Finally, comparing all four panels shows that adopting both notions of an aggressive policy response by increasing ϕ_π and reducing ρ_i produces the lowest losses for all k and all T without affecting the relative rankings across values of k. If $k = 4$, moving from the baseline policy to one with a stronger response to inflation lowers the value of L_0^1 by 48 percent (from 0.82 to 0.43) while combining a stronger response to inflation with a less inertial rule results in a further reduction from 0.43 to 0.29. The total percentage reduction in moving from the baseline rule to the most aggressive rule when $k = 4$) is 65 percent (from 0.82 to 0.29). When reacting late, it pays to be aggressive.

These results support the idea that a more aggressive policy, one that both responds more strongly to inflation and displays less inertia, should be adopted if the central bank has initially been slow to respond. What is not supported is any argument for delay; given the parameters of the policy rule, loss is increasing in delay. However, aggression can compensate for delay; the loss for $k = 1$ under the most aggressive policy outperforms the baseline policy with $k = 0$, and the aggressive policy with a delay of $k = 2$ yields essentially the same loss as the baseline rule with no delay. The best outcomes, as measured by (18), are obtained by responding aggressively, with less inertia and with at most a short delay.

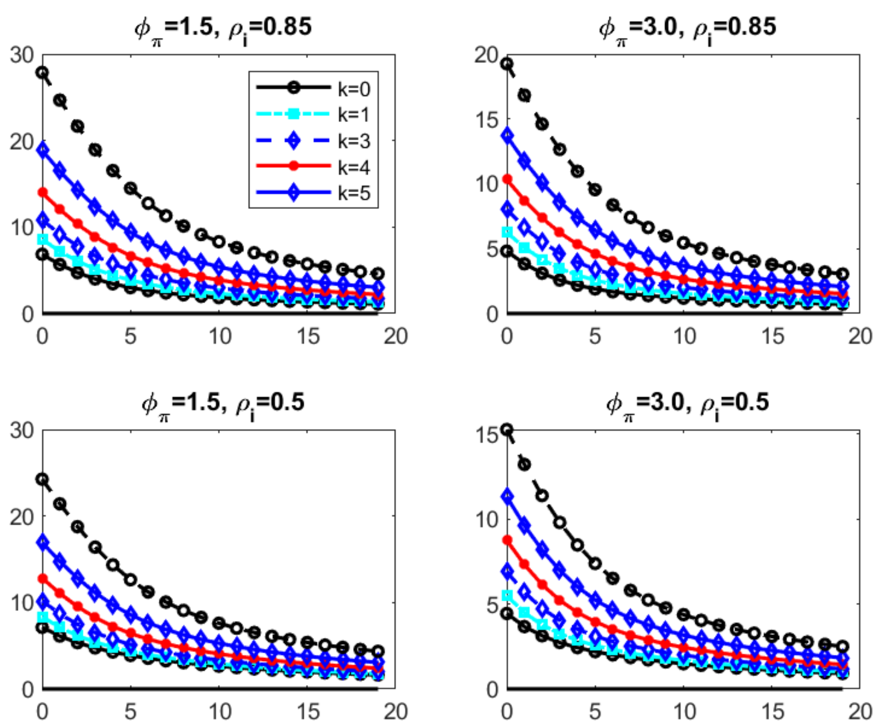

Figure 11: Average loss defined by (19) under Rational Expectations: Periods since the shock are shown on the horizontal axis

Figure 11 shows the values of our second loss measure given by (19) for different policy parameters, following the same organization as Figure 10. This measure of loss reveals a similar ranking as (18); loss is increasing in k. Even though delays increase the output component of (19), that is not enough to offset the poorer inflation performance. This would still be the case even if λ, the weight on the output term, were increased from 0.25 to 2. The overall behavior of both (18) and (19) is dominated by the inflation component.

To summarize, under rational expectations and holding the policy rule fixed, any delay increases loss. The marginal cost of delaying an additional period are increasing as k rises. A short delay of one period can improve over the baseline, no delay policy rule if delay is accompanied with a stronger response to inflation and a less inertial response.

These results are all based on a model that incorporates backward-looking aspects through habits in consumption and policy inertia, but the model also incorporates forward-looking components; expectations of future policy are very important. As is well known, in basic new Keynesian models, the credible promise of a future increase in the policy rate can be more powerful than an actual rise in the current rate (see McKay, Nakamura, and Steinsson 2016). The power of forward guidance plays a role in explaining why even a delay as long as a year ($k = 4$) might not be very costly if combined with an aggressive policy rule. In the next section, we drop the assumption that expectations are formed rationally and consider how the adoption of a behavioral-based alternative model of expectations formation due to Gabaix (2014; 2020) that reduces the effectiveness of forward guidance affects the costs of delay.

5 Deviations from rational expectations: Cognitive Discounting

In this section, we adopt the assumption of cognitive discounting of future events: $\tilde{\mathrm{E}} = \mathrm{E}^{\mathrm{CD}}$, where the superscript CD denotes cognitive discounting as developed in Gabaix (2014; 2020). Gabaix models expectations of a future variable q_{t+d} as given by

$$\mathrm{E}_t^{CD} q_{t+d} = \bar{m}^d \mathrm{E}_t q_{t+d},\ 0 < \bar{m} \leq 1,$$

where $\bar{m}$ denote the degree of cognitive discounting and E denotes rational expectation.

When $\bar{m} = 1$, expectations correspond to the standard assumption of rational expectations. As Gabaix showed, policies such as forward guidance become less powerful in the face of cognitive discounting. This suggests that a policy that delays in responding to inflation promises to eventually react may lead to a less immediate dampening of inflation and a larger initial rise in output than would be the case with rational expectations.

With rational expectations, the model was given by (6), (14) and (15). This model is modified under cognitive discounting to take the form[9]

$$x_t = \left(\frac{1}{1+h}\right)\bar{m}\mathrm{E}_t x_{t+1} + \left(\frac{h}{1+h}\right)x_{t-1} - \left(\frac{1}{\sigma}\right)\left(i_t - \bar{m}\mathrm{E}_t\pi_{t+1} - r_t^*\right), \tag{20}$$

$$\pi_t = \beta M^f \mathrm{E}_t\pi_{t+1} + \kappa\left[\left(\frac{\sigma}{1-h} + \frac{\alpha+\eta}{1-\alpha}\right)x_t - \left(\frac{\sigma h}{1-h}\right)x_{t-1} + \hat{e}_t\right], \tag{21}$$

and

$$\hat{i}_t = \rho_i\hat{i}_{t-1} + (1-\rho_i)\left[\phi_\pi\pi_{t-k} + \phi_x x_{t-k}\right], \tag{22}$$

where

$$M^f = \bar{m}\left[\omega + (1-\omega)\left(\frac{1-\beta\omega}{1-\beta\omega\bar{m}}\right)\right] \leq 1. \tag{23}$$

The additional discounting of expected future inflation in (21) as measured by M^f arises from general equilibrium considerations. In resetting their price, an individual firm discounts the future but also takes into account that other firms are discounting the future as well.

As is clear from (20) and (21), the role that forward guidance plays through future expectations can be significantly reduced by cognitive discounting, depending on the value of $\bar{m}$. Gabaix (2020) suggests $\bar{m} = 0.85$, based on indirect evidence from both micro and macro sources. Using typical parameter values employed in calibrated NK models ($\omega = 0.75$, $\beta = 0.995$), setting $\bar{m} = 0.85$ would imply $M^f = 0.79$. Ilabaca, Meggiorini, and Milani (2020) estimate a new Keynesian model with cognitive discounting and find $\bar{m} = 0.71$ from their post-1982 sample estimates of the Euler relationship and $M^f = 0.41$ from the Phillips curve. We adopt $\bar{m} = 0.85$ for our baseline and use (23) to calibrate M^f to be consistent with the underlying theory. However, Andrade, Coredeiro, and Lambais (2019) integrated cognitive discounting into the New Keynesian framework and obtained a robust confidence range for $\bar{m}$ of between 0.013 and 0.645, suggesting a high level of uncertainty as to the value of $\bar{m}$. Therefore, we report robustness results for alternative values of $\bar{m}$ in the appendix.

5.1 Results under cognitive discounting

To assess the effects of cognitive discounting, we first compare them to the results under rational expectations for the baseline policy rule with no delay. The impulse responses are shown in Figure 12. The black circles denote outcomes under rational expectations[10], while the blue diamonds denote results with cognitive discounting. If agents apply cognitive discounting, both the nominal and real interest rates rises more and remain higher for longer, the output gap falls less, and

[9] For the derivation of M^f under constant returns to scale, see the proofs of Lemma 2 and Proposition 2 in appendix X.B of Gabaix 2020, p. 2322. For the adaptation to the case of decreasing returns to scale ($\alpha < 1$), see Billi and Walsh (2022). This generalization only affects the mapping from real marginal cost to the output gap.

[10] These IRFs are the same as those shown as $k = 0$ in Figure 5

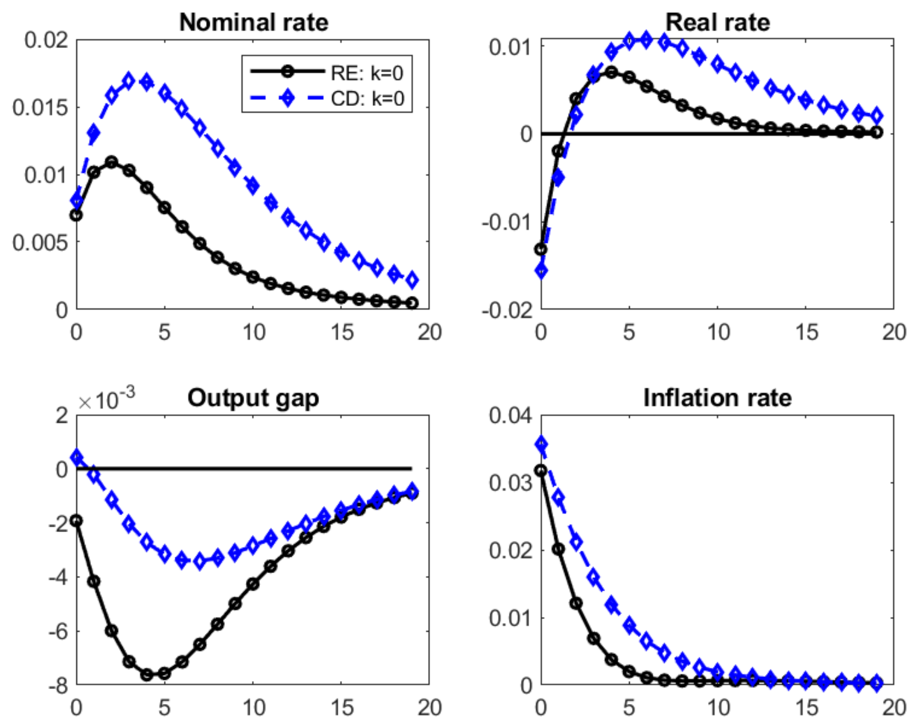

Figure 12: Impulse Responses to Positive Inflation Shock with Baseline Policy Rule under Rational Expectations (black solid line) and Cognitive Discounting (dashed blue diamonds)

inflation is somewhat higher and more persistent than with rational expectations. The more muted recession and the higher inflation rate help account through the policy rule for the higher nominal interest rate. The muted recession contributes to higher inflation. The difference in the paths of the output gap under alternative assumptions about expectations is a reflection of the dampened impact of expected future output on current output under cognitive discounting. Under rational expectations, the expected decline in the output gap causes the output gap to fall immediately when the shock occurs. This effect of expected future output is dampened with cognitive discounting, and the output gap actually rises slightly. The higher level of the output gap contributes to the higher path of inflation.

Figure 13 shows the effects under cognitive discounting for different values of k when the baseline policy rule is implemented after delays of zero to five periods. This figure should be compared to the responses under rational expectations shown in Figure 5. With cognitive discounting, the nominal interest rate reaches its highest peak when $k = 0$. Delay reduces the policy rate peak, with the peak decreasing as k increases. The nominal rate peaks at higher levels than under rational expectations. Both the output gap and inflation are more stable when future expectations are discounted. For

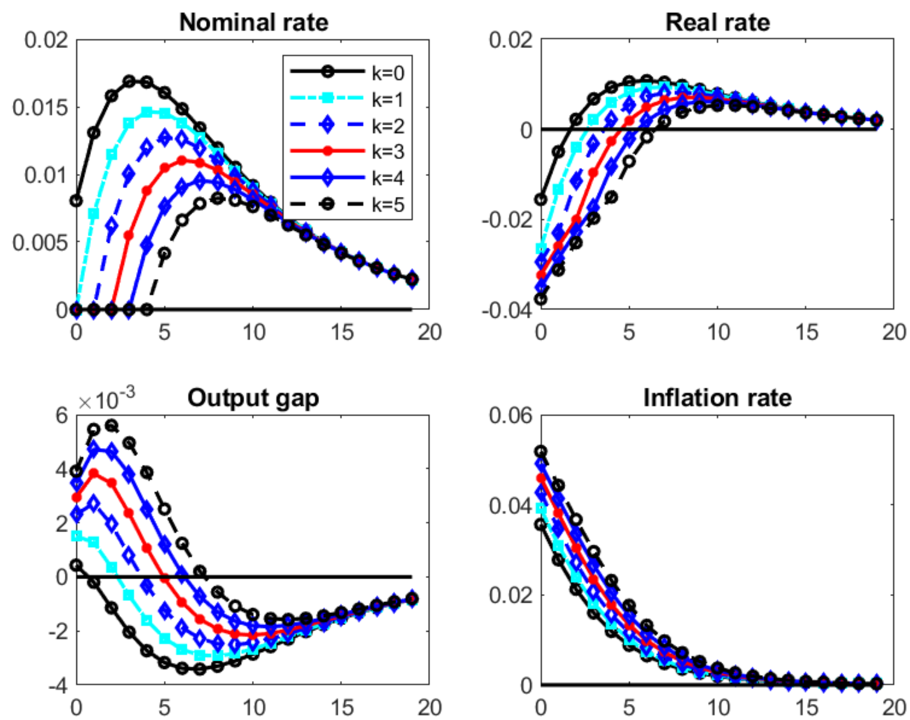

Figure 13: Impulse Responses to Positive Inflation Shock under Cognitive Discounting (interest rates and inflation at annual rates)

example, with cognitive discounting, x varies between a high of just under 0.006 when $k = 5$ and a low of -0.003 when $k = 0$. In contrast, under rational expectations the highs and lows are 0.021 ($k = 5$) and -0.008 ($k = 0$). The baseline policy under cognitive discounting results in increases in inflation that are similar to those with the aggressive policy under rational expectations seen in Figure 8.

The other noticeable effect of cognitive discounting is to reduce the marginal impact of an increase in k. This is especially true of the impulse responses of the output gap. An increase of k from 4 to 5 has about the same impact as an increase from 3 to 4. This was not the case under rational expectations, where the consequences of increasing k from 4 to 5 were much larger than going from 3 to 4. Forward guidance implies changes in the future path of the policy rate can have large effects on current output, and the effects are larger the further into the future the rate change occurs. Cognitive discounting reduces the power of forward guidance and reduces the marginal impact of increasing the delay before policy acts.

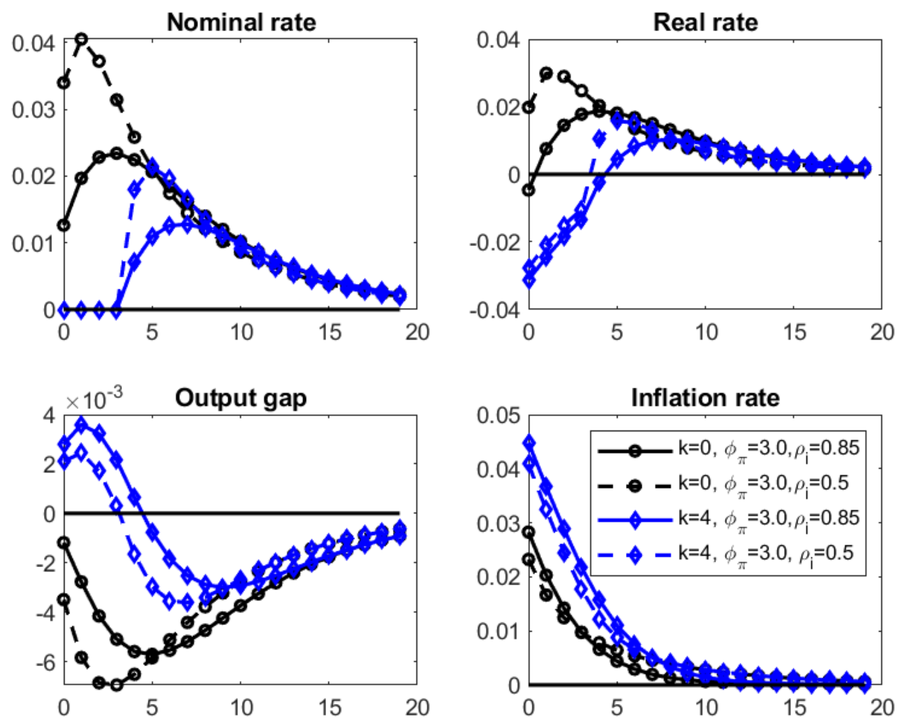

Figure 14: Impulse Responses to Positive Inflation Shock under Cognitive Discounting – $\phi_\pi = 3.0$, comparing $\rho_i = 0.85$ and $\rho_i = 0.5$

5.2 Reacting more strongly to inflation

The implications of policy delay under cognitive discounting combined with a stronger response to inflation when policy finally acts are qualitatively similar to the findings under rational expectations. The amplitude of output swings is smaller and the rise in inflation is dampened under cognitive discounting. Figures corresponding to Figures 6 to 7 are reported in section A.3 of the appendix. Figure 14, which corresponds to Figure 8 for the case of rational expectations, illustrates the effect under cognitive discounting of a policy rule that responds strongly to inflation, $\phi_\pi = 3.0$ and displays less inertia $\rho_i = 0.5$. Both the output gap and inflation are less volatile when expectations of the future are discounted, reflecting the dampening of future expectations under cognitive discounting.

We directly compare outcomes for rational expectations and cognitive discounting in Figure 15 for $k = 0$ and $k = 4$ under the most aggressive policy rule that sets $\phi_\pi = 3.0$ and $\rho_i = 0.5$. The rational expectations outcomes are denoted by black circles, cognitive discounting by blue diamonds. Solid lines indicate the case with $k = 0$, dashed lines $k = 4$. Recall that the Fed waiting 4-quarters before raising its policy rate in March 2022.

It is clear that the aggressive policies with no delay result in the policy rate rising sooner and

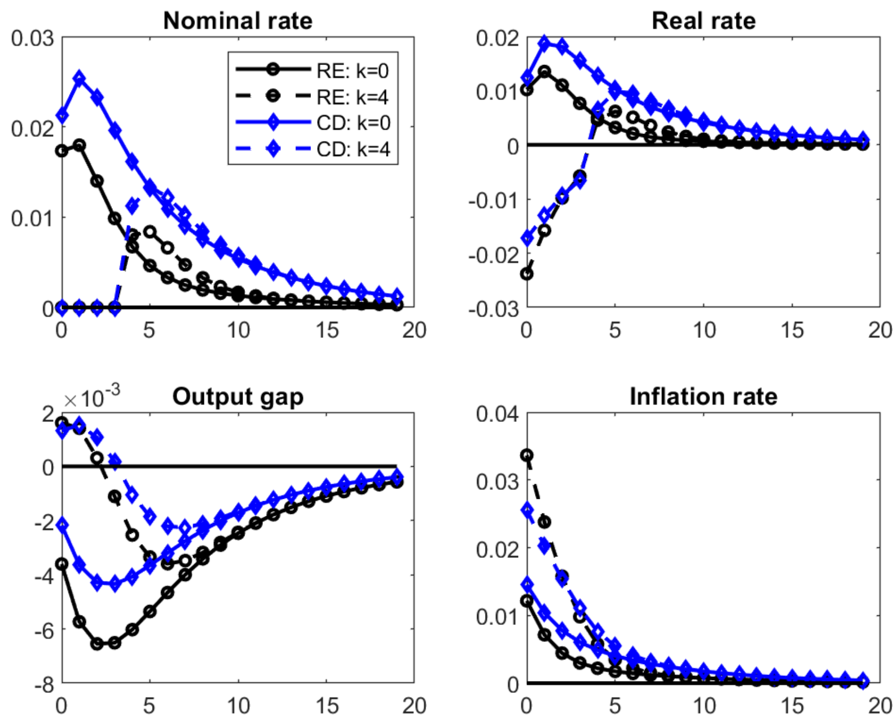

Figure 15: impulse Responses under Different Expectations and an Aggressive Policy ($\phi_\pi = 3.0$ and $\rho_i = 0.5$)

higher, particularly under cognitive discounting. This succeeds in stabilizing inflation, though it also leads to persistent declines in the output gap. This decline is smaller when expectations of the future play a smaller role as they do under cognitive discounting. Under either assumption about expectations, a delay of 4 periods leads to higher inflation, but it also results in a brief rise in the output gap and a subsequent recession that is smaller than that generated by the policy that responds immediately to the inflation shock.

5.3 Measuring the costs of delay under cognitive discounting

The measures of loss given by (18) and (19) that were used to evaluate loss under rational expectations are also used when we assume agents display cognitive discounting. One might argue that a policymaker, knowing private agents form non-rational expectations, should also discard rational expectations in assessing the welfare consequences of alternative policies. In the related context of the literature on robust control, Sims (2001) argued "...the criteria for acceptable shortcuts in decision-making by a central bank should generally be much stricter than those applying to, say, a consumer buying a new washing machine. On the other hand, a 'representative agent that sum-

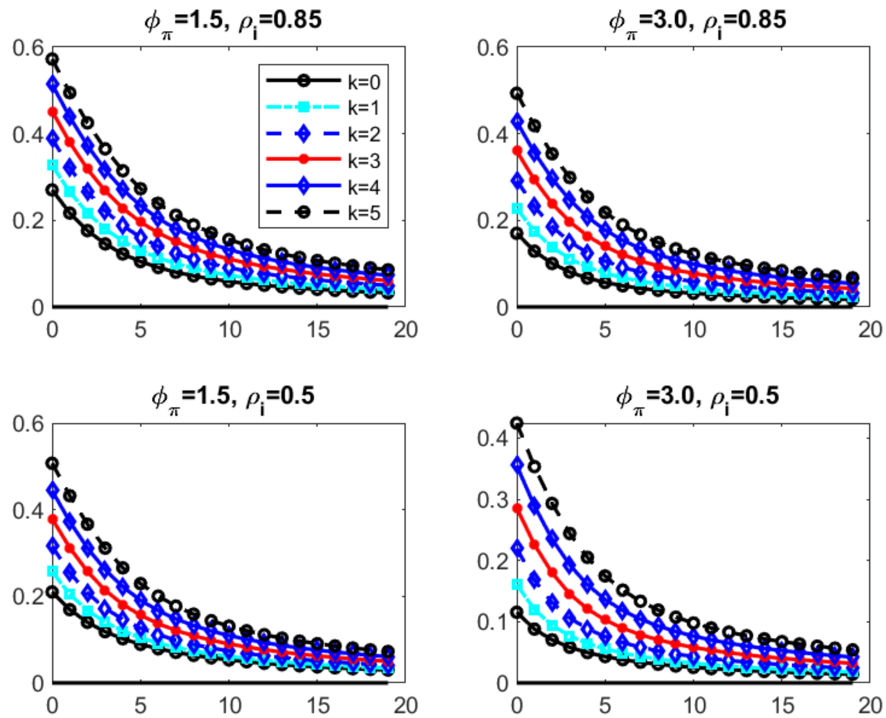

Figure 16: Average Volatility Defined by (18) with Cognitive Discounting: Periods since shock on horizontal axis

marizes the behavior of many individuals with disparate information sources, coordinated through many markets, may be well modeled as having fewer computational constraints than a monetary policy maker. In either case, the criteria for good descriptive modeling and good normative policy advice ought to be kept distinct." Sims contrasted this view with that of Hansen and Sargent (2002) who he characterized as "...recommending to policy-makers the same sub-rational behavior that they postulate in private agents." We are sympathetic with Sims' position, and note that Gabaix (2020) employs the standard loss appropriate under rational expectations in his analysis of optimal policy when private agents display cognitive discounting, as do Budianto, Nakata, and Schmidt (2023) in their analysis of average inflation targeting.

Figure 16 shows volatility as defined by (18) as a function of T for various k and different parameters for the policy rule. It should be compared to Figure (10) obtained under rational expectations. The scales on the two figures differ as the measure of volatility is always lower under cognitive discounting. Cognitive discounting dampens the impact of future deviations from steady state on current inflation and the output gap and reduces overall volatility, regardless of the particular values used in the policy rule. Compare, for example, Figure 5 to Figure 13. In each of the four quadrants, volatility is increasing in k at each T. For each k, loss is lowest under the aggressive policy rule (lower right quadrant). The lowest loss as measured by (18) is achieved if $k = 0$, that is, if policy reacts immediately to the shock.

The measure of loss given by (19) is shown in Figure 17. This figure also shows that the policy that responds immediately achieves the lowest loss. As with Figure 16, loss is lowest with the most aggressive of the four policy rules, and loss is increasing with k. Because delay tends to lead, initially, to a positive output gap and to a shallower subsequent decline in output, (19), which depends negatively on the cumulative discounted path of x could do better under a policy that incorporates a delay if λ is large enough. However, the absolute deviations of x from zero are much smaller than the deviations of π expressed at annual rates and, in practice, the inflation term dominates the measure of loss.

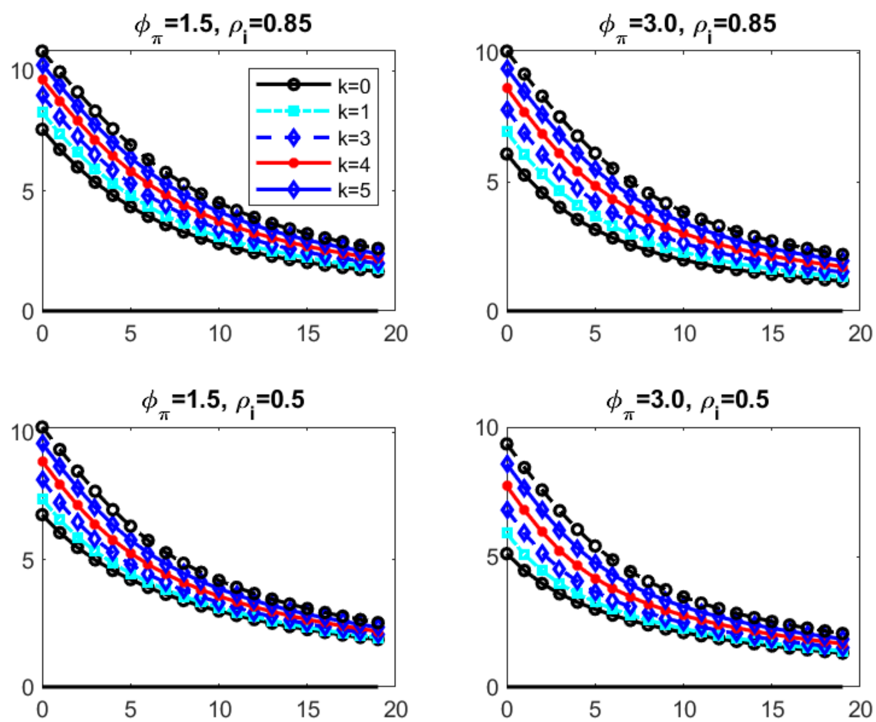

Figure 17: Average Volatility Defined by (19) with Cognitive Discounting: Periods since shock on horizontal axis

6 Can aggression compensate for delay?

The previous results suggest that a central bank that has delayed in reacting to an inflation shock might compensation by adopting a more aggressive policy rule. In this section, we compare the value of the two measures of loss under the baseline policy with no delay ($k = 0$) and $\phi_\pi = 1.5$ and $\rho_i = 0.85$, with more aggressive polices that are implemented with delay. We focus on the policy rule with $\phi_\pi = 3.0$ and $\rho_i = 0.5$ and values of $k = 1, 2, 3$.[11]

Figure 18 shows (18) as a function periods since the shock for various policies. The large (red) circles represent the baseline policy implemented without delay. The other lines reflect the aggressive policy rule and delays of one period (diamonds, solid line), two period (cirles, dashed line), three periods (squares, dot-dashed line), and four periods (diamonds, dotted line). The top panel, based on the assumption of rational expectations, shows that a delay of two periods, combined with the more aggressive respond to inflation and less inertia exactly matches the baseline, no delay policy. A shorter delay ($k = 1$) performs better, longer delays ($k = 3$) do worse. Thus, the central bank can compensate for a late start in reacting to an inflation shock if it adopts a more aggressive policy rule, but not if the policymaker has waited too long.

If private sector expectations are characterized by cognitive discounting, the bottom panels show that the baseline, no delay policy can be replicated by an aggressive policy rule that delays for three periods. Similarly results (not shown) are obtained using the loss function defined in 19; with rational expectations a delay of two periods combined with an aggressive policy rule replicates the baseline policy rule implemented without delay.

[11] Larger values of k produce higher losses than the baseline policy rule without delay and so are not shown.

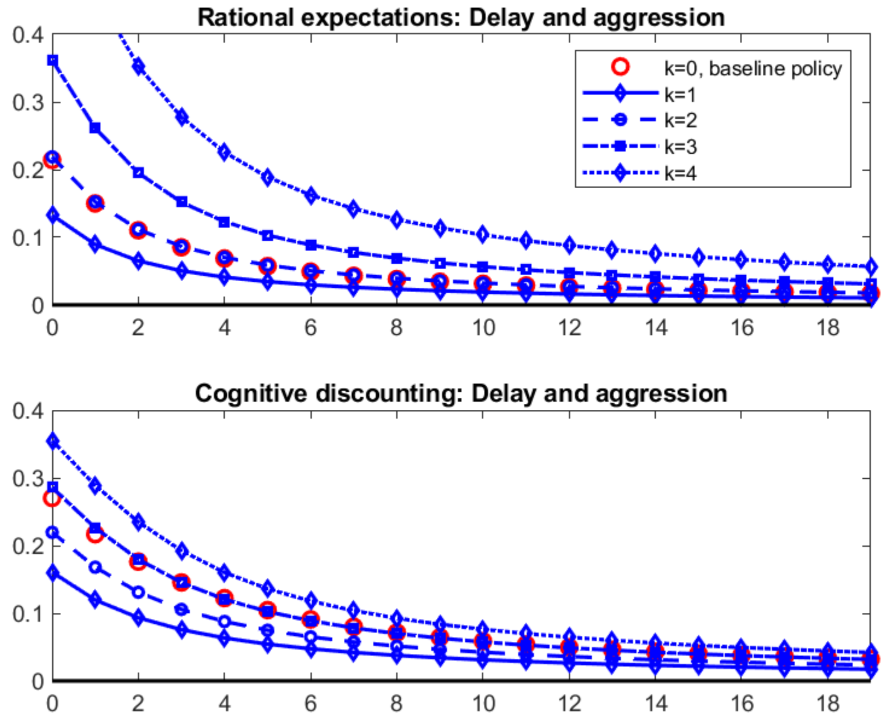

Figure 18: Baseline policy without delay and aggressive policies with delay. Loss measured by (18)

7 Conclusions

In this paper, we first used a simple three-period model to show how policy delay worsens inflation outcomes but can mitigate or even reverse the output decline that occurs when policy responds without delay. Then, using a calibrated new Keynesian framework that incorporates habits in consumption and an inertial Taylor rule, we investigated the dynamic responses of the economy to an inflation shock when monetary policy reacts with a delay. We further expanded the analyses to the cognitive discounting model to investigate the robustness of the results under non-rational expectation.

Under the assumption of both rational expectation and cognitive discounting, we find that delay increases peak inflation and leads the output gap to rise initially, due to the fall in the real interest rate. This is consistent with the simple model. We find that delay mitigates subsequent declines in the output gap during the adjustment back to steady state. This produces a trade-off: reacting more quickly initially boosts the rise in the output gap but exacerbates the rise and persistence of inflation.

To evaluate the impacts of policy delay on welfare, we employ a quadratic, balanced-approach loss function to rank outcomes. This allows us to address the three questions raised in the Introduction. First, how costly is a delay when fighting a surge in inflation? Not very, if policy delays by only one or two period; it can be much more costly if delay extends to four periods or longer. Loss is minimized when policy responds without delay. The volatility of inflation increases monotonically with delay. For the output gap however, volatility is minimized when delay is two periods.

The answer to our second question – If policy is delayed, should it be more aggressive? – is yes. We explored the implications of two definitions of what it might mean to adopt a more aggressive policy response. The first definition focuses on the inflation coefficient in the policy rule; the second focuses on the degree of inertia in the policy response. A more aggressive policy is one that places a larger coefficient on inflation or displays less inertia. Conditional on falling behind the curve, deviations are smallest when policy reacts strongly to inflation and with less inertia. This finding is true under rational expectations and under cognitive discounting.

The answer to our third question – Is a credible promise to fight inflation in the future a substitute for undertaking current policy actions? – is yes, but only if the policymaker does not wait too long and responds aggressively when eventually reacting. Under rational expectations, loss under the no delay, baseline policy rule is replicated with an aggressive response that involves responding more strongly to inflation and acting with less inertia if the policymakers wait one period before responding. Under cognitive discounting, a three-period delay and an aggressive response can replicate the loss achieved under the no delay, baseline rule.

We found that stronger policy responses combined with less inertia moderate the overall magnitude of losses due to delayed policy under both rational expectations and cognitive discounting. However, excessive delay is particularly costly when agents hold rational expectations.

References

AMATO, J. D., AND T. LAUBACH (2004): "Implications of habit formation for optimal monetary policy," *Journal of Monetary Economics*, 66(4), 305–325.

ANDRADE, J., P. CORDEIRO, AND G. LAMBAIS (2019): "Estimating a Behavioral New Keynesian Model," *Working Paper*, pp. 1–23.

ASCARI, G., AND A. M. SBORDONE (2014): "The Macroeconomics of Trend Inflation," *Journal of Economic Literature*, 52(January 2012), 679–739.

BIANCHI, F., C. ILUT, AND H. SAIJO (2021): "Diagnostic Business Cycles," *Working Paper*, (August).

BILLI, R. AND C. E. WALSH (2022): "Seeming Irresponsible but Welfare Improving Fiscal Policy at the Lower Bound," *Sveriges Riksbank Working Paper No. 410.*

BLANCO, A., P. OTTONELLO, AND T. RANOSOVA (2022): "The Dynamics of Large Inflation Surges," *NBER Working Paper No. 30555.*

BUDIANTO, F., T. NAKATA, AND S. SCHMIDT (2023): "Average Inflation Targeting and the Interest Rate Lower Bound," *European Economic Review*, 152, p.104384.

CAI M., DEL NEGRO M., HERBST E., MATLIN E., SARFATI R. AND SCHORFHEIDE F. (2021): "Online estimation of DSGE models.," *The Econometrics Journal*, 24(1), pp.C33-C58.

CHRISTIANO L. J., M. E., AND C. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.

COGLEY, B. T., AND A. M. SBORDONE (2008): "Trend Inflation , Indexation, and Inflation Persistence in the New Keynesian Phillips Curve," *American Economic Review*, 98(5), 2101–2126.

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(621), 2010–2038.

DEL NEGRO, M., F. SCHORFHEIDE, F. SMETS, AND R. WOUTERS (2007): "On the fit of new Keynesian models," *Journal of Business and Economic Statistics*, 25(2), 123–143.

DENNIS, R., K. LEITEMO, AND U. SÖDERSTRÖM (2009): "Methods for robust control," *Journal of Economic Dynamics and Control*, 33(8), 1604–1616.

GABAIX, X. (2014): "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 129(4), 1661–1710.

Gabaix, X. (2020): "A Behavioural New Keynesian Model," *American Economic Review*, 110(8), 2271–2327.

Galí, J. (2015): *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton University Press, Princeton, 2nd edn.

Hansen, L. P., and T. J. Sargent (2002): "Robust Control of Forward-Looking Models," *Journal of Monetary Economics*, 50, 581–604.

Hirose, Y., and Y. Hirose (2022): "Estimating a Behavioral New Keynesian Model with the Zero Lower Bound," *CARF Working Paper F-535*.

Holbrook, R. S.(1972): "Optimal Economic Policy and the Problem of Instrument Instability," *American Economic Review*, 62(1), 57–65.

Ilabaca, F., G. Meggiorini, and F. Milani (2020): "Bounded rationality, monetary policy, and macroeconomic stability," *Economics Letters*, 186, 108522.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2013): "Is There a Trade-Off between Inflation and Output Stabilization?," *American Economic Journal: Macroeconomics*, 5(2), 1–31.

L'Hullier, J.-P., S. Singh, and D. Yoo (2021): "Incorporating Diagnostic Expectations into the New Keynesian Framework," *Working Paper*.

Lane, P. (1984): "Instrument Instability and Short-term Monetary Control" *Journal of monetary Economics*, 14(2), 209–224.

Leith, C., I. Moldovan, and R. Rossi (2012): "Optimal monetary policy in a New Keynesian model with habits in consumption," *Review of Economic Dynamics*, 15(3), 416–435.

Levin A. T. and J. C. Williams (2003): "Robust Monetary Policy with Competing Reference Models," *Journal of Monetary Economics*, 50(15), 945–975.

McKay, A., E. Nakamura, and J. Steinsson (2016): "The power of forward guidance revisited," *American Economic Review*, 106(10), 3133–3158.

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.

Pfäuti, O. and Seyrich, F. (2022): "A behavioral heterogeneous agent New Keynesian model," *DIW Berlin Discussion Paper*.

Sims, C. A. (2001): "Pitfalls of a minimax approach to model uncertainty," *American Economic Review*, 91(2), 51–54.

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian Approach," *American Economic Review*, 97(3), 586–606.

Taylor, J. B. (1993): "Discretion versus Policy Rules in Practice," *Carnegie Rochester Conference Series on Public Policy*, 39(1), 195–214.

Walsh, Carl E. (2022): "Inflation Surges and Monetary Policy," *Bank of Japan Monetary and Economic Studies*, 40, 30–65.

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

Woodford, Michael (2011): "Simple analytics of the government expenditure multiplier," *American Economic Journal: Macroeconomics*, 3(1), 1–35.

Yellen, Janet L. (2012): "The Economic Outlook and Monetary Policy," *Speech to the Money Marketers of New York University*, April 11th.

A Appendix

A.1 The simple three period example

Remark 1 *See simple$_e$xample.tex and simple$_e$xample.m.*

Equilibrium in period 2 is given by (1) and (2), written here as

$$\begin{bmatrix} \pi_2 \\ x_2 \end{bmatrix} = \left(\frac{\sigma}{\sigma + \gamma_2 \kappa \phi} \right) \begin{bmatrix} 1 \\ -\sigma^{-1} \gamma_2 \phi \end{bmatrix} e_2 \tag{24}$$

In period 1, the system is

$$\begin{bmatrix} 1 & -\kappa \\ \sigma^{-1} \gamma_1 \phi & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ x_1 \end{bmatrix} = \begin{bmatrix} \beta & 0 \\ \sigma^{-1} & 1 \end{bmatrix} \begin{bmatrix} \mathrm{E}_1 \pi_2 \\ \mathrm{E}_1 x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} e_1.$$

Using (24),

$$\begin{bmatrix} 1 & -\kappa \\ \sigma^{-1} \gamma_1 \phi & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ x_1 \end{bmatrix} = \left(\frac{\sigma}{\sigma + \gamma_2 \kappa \phi} \right) \begin{bmatrix} \beta & 0 \\ \sigma^{-1} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -\sigma^{-1} \gamma_2 \phi \end{bmatrix} e_2 + \begin{bmatrix} 1 \\ 0 \end{bmatrix} e_1.$$

Solving, after noting that $e_2 = \rho e_1$, one obtains

$$\begin{bmatrix} \pi_1 \\ x_1 \end{bmatrix} = \left(\frac{\sigma}{\sigma + \kappa \phi \gamma_1} \right) \begin{bmatrix} 1 + \rho \frac{\sigma}{\sigma + \kappa \phi \gamma_1} \left(\sigma \beta - \kappa \left(\phi \gamma_2 - 1 \right) \right) \\ -\frac{\gamma_1 \phi}{\sigma} - \rho \frac{1}{\sigma + \kappa \phi \gamma_1} \left(\beta \phi \gamma_1 - 1 + \phi \gamma_2 \right) \end{bmatrix} e_1.$$

We consider three alternative cases.

1. Policy does not respond in either periods 1 or 2: $\gamma_1 = \gamma_2 = 0$.

$$\begin{bmatrix} \pi_1 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 + \rho \left(\sigma \beta + \kappa \right) \\ \rho \frac{1}{\sigma} \end{bmatrix} e_1.$$

$$\begin{bmatrix} \pi_2 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} e_2$$

2. Policy does not respond in period 1 but does in period 2: $\gamma_1 = 0$, $\gamma_2 = 1$.

$$\begin{bmatrix} \pi_1 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 + \rho \frac{\sigma}{\sigma} \left(\sigma \beta - \kappa \left(\phi - 1 \right) \right) \\ \rho \frac{1}{\sigma} \left(1 - \phi \right) \end{bmatrix} e_1.$$

$$\begin{bmatrix} \pi_2 \\ x_2 \end{bmatrix} = \left(\frac{\sigma}{\sigma + \kappa \phi} \right) \begin{bmatrix} 1 \\ -\sigma^{-1} \phi \end{bmatrix} e_2$$

3. Policy responds in periods 1 and 2: $\gamma_1 = \gamma_2 = 1$.

$$\begin{bmatrix} \pi_1 \\ x_1 \end{bmatrix} = \left(\frac{\sigma}{\sigma + \kappa\phi}\right) \begin{bmatrix} 1 + \rho\frac{\sigma}{\sigma+\kappa\phi}\left(\sigma\beta + \kappa\left(1-\phi\right)\right) \\ -\frac{\phi}{\sigma} - \rho\frac{1}{\sigma+\kappa\phi}\left(\beta\phi - 1 + \phi\right) \end{bmatrix} e_1.$$

$$\begin{bmatrix} \pi_2 \\ x_2 \end{bmatrix} = \left(\frac{\sigma}{\sigma + \kappa\phi}\right) \begin{bmatrix} 1 \\ -\sigma^{-1}\phi \end{bmatrix} e_2$$

A.2 Solution method

These notes lay out how we solve the model in which i_t is fixed at zero for k periods. The policy rate then switches to follow an inertial Taylor rule:

$$i_t = \begin{cases} 0 \text{ for } t = 0, ..., t+k-1 \\ \rho_i i_{t-1} + (1-\rho_i)\, q_t \text{ for } t \geq t+k \end{cases}, \tag{25}$$

where $q_t = \phi_\pi \pi_t + \phi_x x_t$.

We set out the general approach to solving the model in section A.2.1 before turning to our specific model in section A.2.4.

A.2.1 The general model

Assume the model can be written as

$$A_1 Z_t = A_0 Z_{t-1} + A_2 E_t Z_{t+1} + A_3 q_t + A_4 v_t,$$

where Z_t is an $N \times 1$ vector of endogenous variables, q_t is the control variable and v_t is the shock innovation. The matrices A_0, A_1, and A_2 are $N \times N$, while A_3 and A_4 are $N \times 1$.[12] For our application, there is a shock at time t only, so $v_{t+i} = 0$ for all $i > 0$.

The solution is obtain in two steps. First, the equilibrium for $Z_{t+k}, ...$ is obtained as functions of Z_{t+k-1} and v_t. Second, the equilibrium from t to $t+k-1$ when $q_t = 0$ is characterized by a set of k equations in k unknowns that are solved to obtain $Z_t, ..., Z_{t+k-1}$, given the initial condition Z_{t-1}.

A.2.2 Step 1: From $t+k$ forward

Let the policy instrument q_t be given by

$$q_t = F Z_t. \tag{26}$$

[12] If there are S shocks, A_4 would be $N \times S$.

From $t+k$ forward, the equilibrium of the model satisfies

$$A_1 Z_t = A_0 Z_{t-1} + A_2 E_t Z_{t+1} + A_3 F Z_t + A_4 v_t. \tag{27}$$

This is a standard linear, rational expectations model and can be solved in dynare. We chose instead to adopt a more direct approach that is outlined here.

Collecting the terms in Z_t appearing in (27) yields

$$(A_1 - A_3 F) Z_t = A_0 Z_{t-1} + A_2 E_t Z_{t+1} + A_4 v_t. \tag{28}$$

Guess that the solution takes the form

$$Z_t = H_0 Z_{t-1} + H_1 v_t, \tag{29}$$

where H_0 is $N \times N$ and H_1 is $N \times 1$. Equation (29) implies

$$E_t Z_{t+1} = H_0 Z_t$$

under rational expectations. Using this to eliminate $E_t Z_{t+1}$ in (28) and solving for Z_t gives

$$Z_t = (A_1 - A_3 F - A_2 H_0)^{-1} (A_0 Z_{t-1} + A_4 v_t).$$

For this to agree with our initial guess in (29), it must be the case that

$$H_0 = (A_1 - A_3 F - A_2 H_0)^{-1} A_0 \tag{30}$$

and

$$H_1 = (A_1 - A_3 F - A_2 H_0)^{-1} A_4. \tag{31}$$

One can solve (30) for H_0 by iterating on H_0 to find the matrix H_0^* such that

$$H_0^* = (A_1 - A_3 F - A_2 H_0^*)^{-1} A_0.$$

H_1 is then obtained from (31) as

$$H_1^* = (A_1 - A_3 F - A_2 H_0^*)^{-1} A_4.$$

Because $v_{t+k} = 0$ for $k > 0$, for $s \geq k$,

$$Z_{t+s} = H_0^* Z_{t+s-1}. \tag{32}$$

A.2.3 Step 2: From $t+k$ backward

Given (32), the next step is to solve backward to find $Z_{t+k-1}, ..., Z_t$, where t is the period in which the shock occurs.

From t to $t+k-1$, $q_t = 0$. Hence, (27) becomes

$$A_1 Z_t = A_0 Z_{t-1} + A_2 E_t Z_{t+1} + A_4 v_t \tag{33}$$

When the model is expressed in terms of deviations from the steady state, the initial condition becomes $Z_{t-1} = 0$. Since no further shocks after period t, $E_t Z_{t+1} = Z_{t+1}$. Therefore

$$A_1 Z_t - A_2 Z_{t+1} = A_4 v_t. \tag{34}$$

For $j = 1, ...k-2$,

$$A_1 Z_{t+j} - A_0 Z_{t+j-1} - A_2 Z_{t+j+1} = 0. \tag{35}$$

For $j = k-1$,

$$A_1 Z_{t+k-1} - A_0 Z_{t+k-2} = A_2 Z_{t+k} = A_2 H_0^* Z_{t+k-1},$$

where (32) has been used for Z_{t+k}. Rewrite this as

$$(A_1 - A_2 H_0^*) Z_{t+k-1} - A_0 Z_{t+k-2} = 0. \tag{36}$$

Equations (34), (35) and (36) provide k simultaneous equations for solving for $Z_t, ..., Z_{t-k}$.

For example, suppose $k = 4$. Stacking (34), (35) and (36) produces

$$\begin{bmatrix} A_1 & -A_2 & 0 & 0 \\ -A_0 & A_1 & -A_2 & 0 \\ 0 & -A_0 & A_1 & -A_2 \\ 0 & 0 & -A_0 & A_1 - A_2 H_0^* \end{bmatrix} \begin{bmatrix} Z_t \\ Z_{t+1} \\ Z_{t+2} \\ Z_{t+3}. \end{bmatrix} = \begin{bmatrix} A_4 v_t \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

or

$$RQ_t = S,$$

which gives us k equations for obtaining the k unknowns $Z_t, ..., Z_{t+k-1}$, where R is a $kN \times kN$ matrix and Q_t as a $kN \times 1$ stacked vector such that

$$Q_t = [Z_t; Z_{t+1}; Z_{t+2}; ...; Z_{t+k-1}]'.$$

Equilibrium before the monetary authority reacts is

$$Q_t = R^{-1} S.$$

A.2.4 Application to our model

$$e_t = \rho_e e_{t-1} + v_t$$

$$i_t = \rho_i i_{t-1} + (1 - \rho_i) z_t$$

$$q_t = \phi_\pi \pi_t + \phi_x x_t.$$

$$r_t = i_t - E_t \pi_{t+1}$$

$$(1 + \beta\mu) \pi_t = \mu \pi_{t-1} + \beta E_t \pi_{t+1} + \kappa_1 x_t + \kappa_2 x_{t-1} + \kappa_0 e_t$$

$$(1 + h) x_t = h x_{t-1} + E_t x_{t+1} - \frac{1}{\sigma} (1 - h) (i_t - E_t \pi_{t+1})$$

where

$$\kappa_0 = \frac{(1-\omega)(1-\beta\omega)}{\omega} \frac{1-\alpha}{1-\alpha+\epsilon\alpha}$$

$$\kappa_1 = \kappa_0 \left(\frac{\sigma}{1-h} + \frac{\alpha+\eta}{1-\alpha} \right)$$

$$\kappa_2 = -\kappa_0 \left(\frac{\sigma h}{1-h} \right)$$

Define

$$Z_t = [e_t \ i_t \ r_t \ \pi_t \ x_t]'.$$

This system can be rewritten in the form given by (27) with

$$A_0 = \begin{bmatrix} \rho_e & 0 & 0 & 0 & 0 \\ 0 & \rho_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & \kappa_2 \\ 0 & 0 & 0 & 0 & h \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ -\kappa_0 & 0 & 0 & 1+\beta\mu & -\kappa_1 \\ 0 & \frac{1-h}{\sigma} & 0 & 0 & 1+h \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \frac{1-h}{\sigma} & 1 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0 \\ (1-\rho_i) \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$F_t = \begin{bmatrix} 0 & 0 & 0 & \phi_\pi & \phi_x \end{bmatrix}.$$

In section 5, we reformulate the model to incorporate cognitive discounting in expectations formation. The only modification this requires to our solution methods is in the specification of the matrix A_2, which becomes

$$A_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & \beta M^f & 0 \\ 0 & 0 & 0 & \frac{1-h}{\sigma} & \bar{m} \end{bmatrix},$$

A.2.5 The code

`delay_model.m` calls `solvemodel.m` and `nopolicyv.m` to obtain the model solutions and `refigures` and `cdfigures` to generate figures. These are all available at https://people.ucsc.edu/~walshc/HakamadaWalsh/.

A.3 Reacting more aggressively under cognitive discounting

Figures are reported here for cognitive discounting correspond to Figures 6 - 8 for rational expectations.

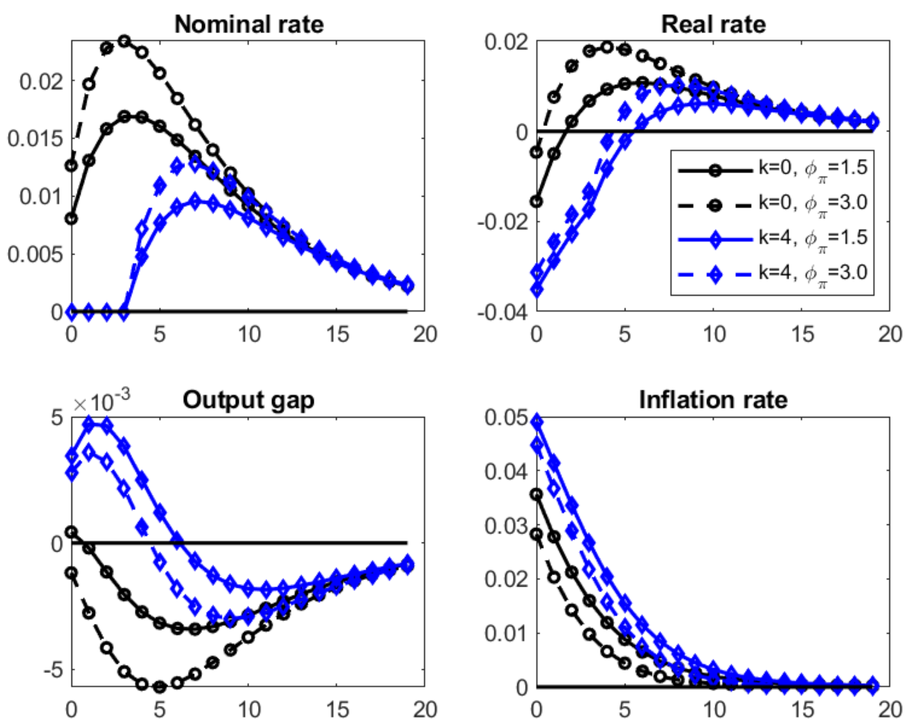

Figure 19: Impulse Responses to Positive Inflation Shock under Cognitive Discounting: Comparing $\phi_\pi = 1.5$ and $\phi_\pi = 3.0$

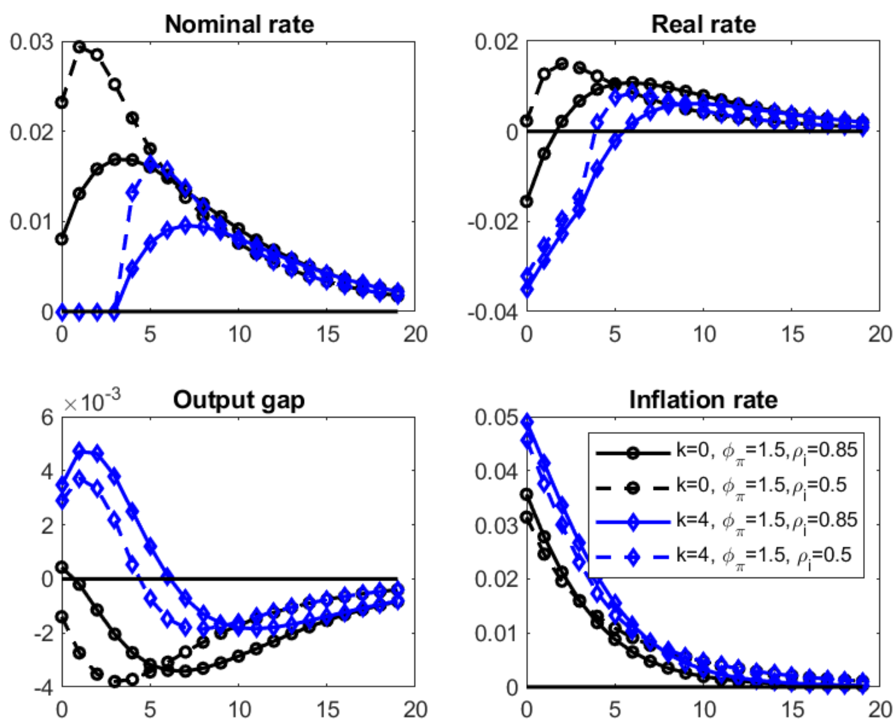

Figure 20: Impulse Responses to Positive Inflation Shock under Cognitive Discounting – $\phi_\pi = 1.5$, comparing $\rho_i = 0.85$ and $\rho_i = 0.5$

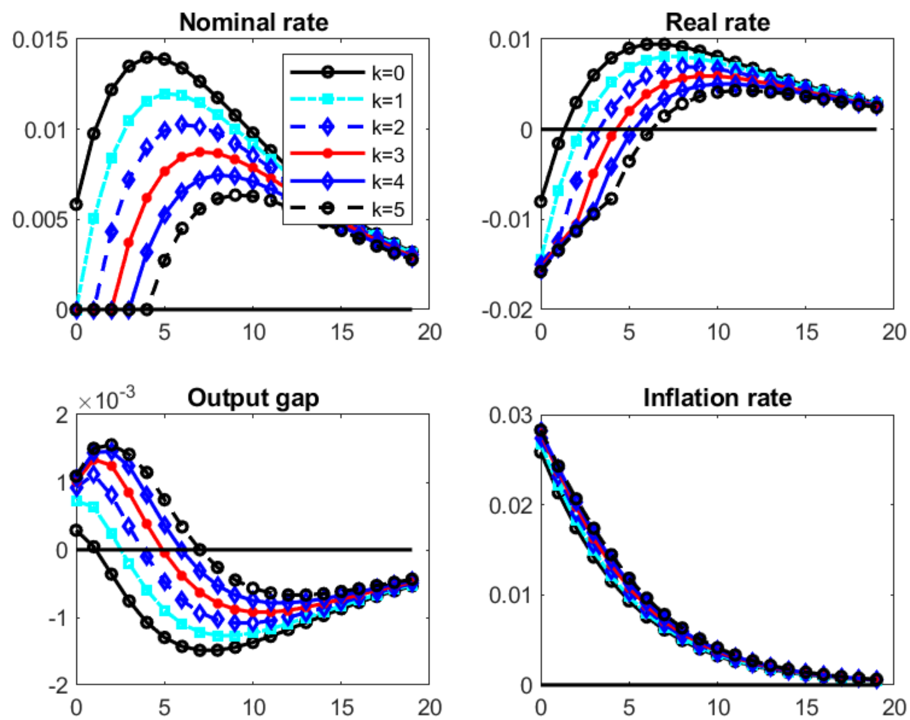

Figure 21: Impulse Responses to Positive Inflation Shock under Cognitive Discounting: $\bar{m} = 0.75$

A.4 The effects of $\bar{m}$ with cognitive discounting

The results under cognitive discounting depend of the values assumed for $\bar{m}$. When $\bar{m} = 1$, one obtains the rational expectations case, so increasing $\bar{m}$ from our baseline value of 0.85 reduces the gap between the outcomes with cognitive discounting and those under rational expectations. However, this does not affect any of the basic conclusions of the paper.

If $\bar{m}$ is reduced, the power of forward guidance is reduced relative to our baseline. This reduces the economy's initial reaction to the inflation as future inflation and the output gap have a smaller impact on the current equilibrium. This can be seen in Figure 21. The response of both the output gap and inflation are muted relative to those with $\bar{m} = 0.85$, shown in Figure 13 or to those under rational expectations shown in Figure 5. As a result, the measures of loss are correspondingly smaller. The values of 18 for various k when $\bar{m}$ can be seen in Figure 22. This should be compared to Figure 10 for rational expectations and Figure 16 for cognitive discounting using the baseline value of $\bar{m}$.

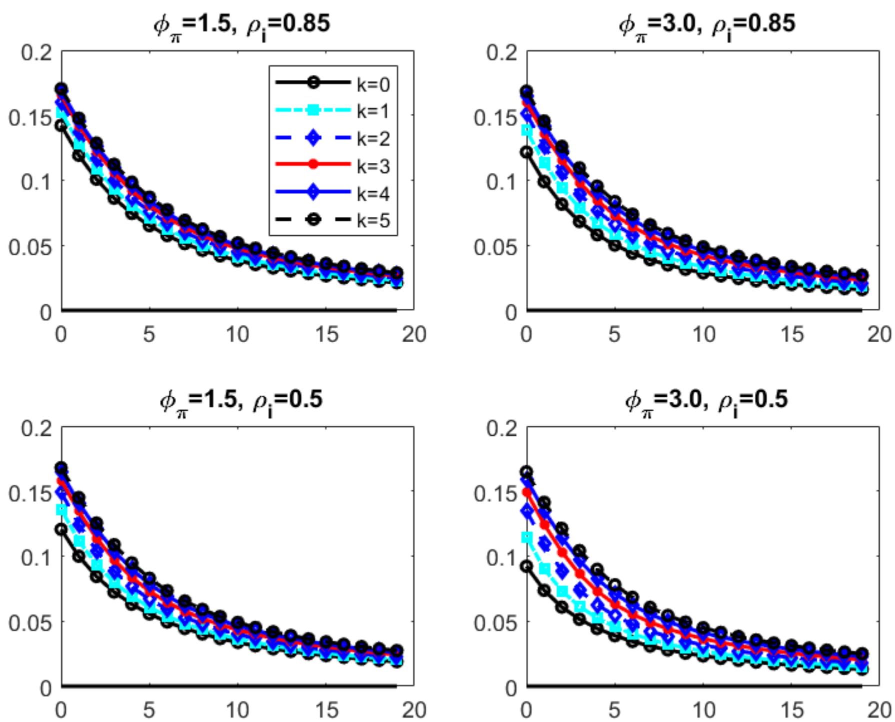

Figure 22: Impulse Responses to Positive Inflation Shock under Cognitive Discounting: $\bar{m} = 0.75$

PUBLICATIONS

The Consequences of Falling Behind the Curve: Inflation Shocks and Policy Delays under Rational and Behavioral Expectations

Working Paper No. WP/2024/042