The Central Bank's Dilemma:
Look Through Supply Shocks or Control Inflation Expectations?*

Paul Beaudry — University of British Columbia
Thomas J. Carter — IMF
Amartya Lahiri — University of British Columbia

March 7, 2025

Abstract

When countries are hit by supply shocks, central banks often face the dilemma of either looking through such shocks or reacting to them to ensure that inflation expectations remain anchored. In this paper, we propose a tractable framework to capture this dilemma and explore optimal policy under a range of assumptions on how expectations are formed, including a form of bounded rationality involving level-k thinking (LKT). Despite modelling LKT in a way that nests both adaptive and rational expectations as special cases, we show that the optimal policy under LKT is qualitatively different and involves abrupt pivots in the policy stance. In particular, it is optimal for the central bank to initially look through supply shocks until a threshold is reached, then pivot discontinuously to a more hawkish anti-inflationary stance. We find that such pivots can, if optimally executed, be compatible with soft landings in the sense that most (or even all) of the reduction in inflation occurs through re-anchoring of expectations rather than economic slack. We also discuss risks and why policy errors in terms of tightening too late or too slowly can be especially costly in such an environment.

Keywords: Inflation expectations, supply shocks, monetary policy

JEL Classification: E3, E5

*The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada. We would like to thank Roc Armenter, Jenny Chan, José Dorich, Frank Smets, Ludwig Straub, George-Marios Angeletos and participants at the 2022 Dutch National Bank Research conference, the 2022 ISI Growth conference, 2023 joint BoE-IMF-ECB-BIS conference on high inflation and spillovers, the 2023 European Economics Association meetings, the 2023 joint ECB-Cleveland Fed conference "Inflation: Drivers and Dynamics", and UBC for helpful comments and discussions. A previous version of this paper circulated under the title "Looking through supply shocks versus controlling inflation expectations: understanding the central bank dilemma".

1 Introduction

How best to respond to supply shocks is a classic topic in monetary economics, and one that advanced economies' recent experiences with inflation have brought back to the forefront. Starting in mid-2021, a series of supply shocks led central banks in many countries to ask similar questions:[1] should they look through supply shocks – even temporarily – at the risk of de-anchoring inflation expectations; or should they immediately tighten monetary policy to keep expectations anchored, albeit at the risk of hindering the recovery from the COVID pandemic? A range of developments on the demand side of the economy further complicated this choice.

To adequately address the question of the appropriate policy response to supply shocks, it is important to tackle them in a framework rich enough to capture both the costs and benefits of looking through these shocks. In our view, this requires two key modelling ingredients: (i) an explicit mechanism for why looking through supply shocks can be attractive to policy-makers under certain circumstances; and (ii) a flexible approach to modelling inflation expectations that allows for the possibility that looking through supply shocks could lead to de-anchoring. While each of these elements is present in the literature, the aim of this paper is to bring them together in a setup that is sufficiently tractable to allow for an explicit characterization of the relevant policy trade-offs and their optimal resolution under a range of assumptions on the expectation formation process.

To preview our model's main features, the two key ingredients that we just described will respectively take the form of wage rigidities and a model of bounded rationality that involves level-k thinking but nests both adaptive and rational expectations as special cases. With regard to the first of these key ingredients, one might be tempted to instead turn to the canonical three-equation New Keynesian (NK) model, which assumes a combination of sticky prices and flexible wages and plays a central role in many monetary policy discussions. However, as we are reminded in Reis (2022), an incentive to look through supply shocks only arises in this context when these shocks are conflated with mark-up shocks, which we view as inappropriate given that many of the shocks contributing to the recent inflation surge were "true" supply shocks with direct bearing on the productive capacities of many economies (e.g., supply chain disruptions, commodity price shocks).[2] When facing such shocks, the prescription of the canonical NK model is instead that

[1]Recent work in di Giovanni et al. (2022) provides empirical evidence for the significant role of supply shocks in accounting for the recent inflation surges in the euro area and the United States.

[2]In the model, supply shocks are specifically assumed to take the form of productivity shocks, but these can be interpreted as subsuming a range of potential supply shocks. For example, in the Online Appendix, we establish a formal equivalence between productivity shocks and oil price shocks in the context of a small, oil-importing economy.

policy-makers should *never* look through them, even if they are temporary, since a policy that adjusts output one-for-one with its natural level will achieve a "divine coincidence" under which the inflation and output gaps are simultaneously closed. As a result, we view the canonical model as an inadequate laboratory for our analysis but recognize that wage rigidities are well known to overturn its stark prescriptions.[3] This makes a model with sticky wages (or more generally with stickier wages than prices) a more promising laboratory – in addition to offering more realism to the extent that empirical estimates generally favour higher degrees of wage rigidity than price rigidity.

To get more specific, we model wage rigidity by considering an economy subject to productivity shocks and assuming that wages are set before observing the current shock, while prices are set after. This creates potential for wage-price spirals in the sense that wage-setters set wages based on the prices they expect firms to set, and these wages then become part of the costs that firms factor into their pricing decisions. This configuration also gives rise to a Phillips curve in which productivity shocks enter as a residual, and policy-makers will generally have an incentive to look through these shocks to avoid variation in employment and output.

However, the cost of looking through supply shocks will be that it could lead to a de-anchoring in inflation expectations, and this brings us to the second of the model's key ingredients. Here our goal is to understand how outcomes and optimal policy are shaped by the degree of sophistication with which agents form expectations – especially with regard to the way that they account for the central bank's announced policies in their inflation expectations. To do so, we rely on level-k thinking (LKT), which is a form of bounded rationality that, as we detail below, enjoys significant support in the experimental literature and has recently been applied to a range of topics in macroeconomics. We specifically model LKT in such a way that expectations are adaptive at level $k = 0$, while the "maximum sophistication" limit $k \to \infty$ corresponds to the case of rational expectations, where agents can fully account for the effects of monetary policy. In contrast, intermediate values of k lead to a situation where agents' expectations are based on a combination of backward- and forward-looking elements, with the latter reflecting an imperfect understanding of the way that monetary policy impacts the economy.

With these ingredients in place, we examine how monetary policy should respond to supply shocks when policy-makers aim to minimize a loss function that penalizes deviations of inflation from target and employment from its natural level. To set the stage, we first show that under

[3] This is an example of how models with sticky prices and sticky wages often deliver similar prescriptions with regard to the management of demand shocks, but can differ considerably in the case of supply shocks.

rational expectations (RE), it is optimal for central banks to fully look through supply shocks. In contrast, the optimal policy under adaptive expectations (AE) is generally tighter but involves a constant degree of "look-through" in the precise sense that the rate at which policy-makers make trade-offs between the inflation and employment gaps never changes over time.

Our main analysis focuses on the intermediate case where agents form expectations under LKT. In this case, the optimal policy that emerges is qualitatively very different from those emerging under RE and AE. In particular, one of our main results is to identify conditions under which the optimal policy under LKT involves a type of "pivoting" behaviour in the sense that policy-makers initially opt to mostly look through supply shocks, but then switch suddenly to a very aggressive policy response if shocks cumulate beyond some threshold level.

These abrupt pivots occur because LKT gives rise to a non-convexity in the central bank's loss function. The context for this non-convexity is that the employment impacts of policy tightening work through two off-setting channels in the model economy: (i) a "direct" channel whereby tighter policy reduces economic activity at a given level of private-sector expectations; and (ii) an "indirect" channel whereby tighter policy helps to re-anchor inflation expectations, thus reducing the amount of *actual* economic slack that policy-makers have to engineer to stabilize realized inflation. All else equal, the indirect channel will tend to be stronger during periods when the economy has recently experienced large inflation overshoots, since in this case expectations will already be elevated due to the backward-looking elements on which they are partially based under LKT. However, it is also possible for the indirect channel to get stronger *as policy-makers tighten*, since we find that a key property of private-sector expectations under LKT is that they are more sensitive to changes in monetary policy if policy is already tight to begin with.[4] This is the underlying source of non-convexity, and it contributes to a situation where there are benefits to tightening by a sufficiently wide margin, but doing so only after the economy has experienced sufficiently large overshoots.

Another key result is that with the direct and indirect channels above both in play, pivots need not be associated with hard landings in the sense of triggering large contractions. In particular, if the indirect channel is strong enough, then pivots could help to re-anchor expectations without requiring a significant amount of slack. In this respect, we highlight a special case where the two channels offset one another exactly in expectation, implying that pivots have no impact on the expected level of employment. Outside this special case, we also find that expected impacts on

[4] In the Online Appendix, we show that this key property also holds under models of bounded rationality involving reflective expectations (García-Schmidt and Woodford, 2019) and cognitive hierarchies (Camerer et al., 2004).

employment are consistently small in numerical exercises. In this sense, the model predicts that optimally-executed pivots should be compatible with soft landings, at least in expectation.

However, this is not to say that pivots are costless, since they also increase the variance of employment by making monetary policy more responsive to supply shocks. As a result, pivots are risky and involve accepting some probability that poor draws of the supply shock could lead to hard landings. In simulations, we further show that the costs of mis-timing pivots can be substantial, as can the costs of attempting to achieve similar degrees of tightening in more gradual ways.

Three points regarding our approach and results are in order at this stage. First, to make maximal use of the LKT framework as a way of exploring the risk of de-anchoring, our approach downplays the intertemporal coordination frictions that typically arise in NK models due to staggered price- and wage-setting, instead putting the focus on *intratemporal* coordination frictions. In particular, we assume that prices and wages reset each period, though in the case of wages this occurs before agents have an opportunity to observe the current productivity shock. As a result, the key coordination issue in our model is that wage-setters have to form expectations on prices, which will then be set by firms following a mark-up rule that partly depends on the average wage. Wage-setters are thus effectively forced to forecast each other's expectations. This gives rise to a static coordination game to which LKT can readily be applied and within which monetary policy serves as a coordination mechanism for nominal choices.

Second, in order to focus on the policy dilemmas presented by supply shocks, we mostly disregard the role of demand shocks in our analysis. We adopt this focus not to be dismissive of the likelihood that demand forces played an important role in the recent inflation surges experienced in many countries. Rather, the choice reflects the fact that our approach has nothing novel to contribute regarding the demand side of the economy. Accordingly, one should view our results as describing the optimal adjustments that supply shocks dictate, over and above the central bank's underlying response to demand shocks. We discuss and formalize this interpretation in Section 6.

Third, it is important to bear in mind that the issue of how to respond to supply-driven inflation shocks is not unique to recent times, nor to industrial countries. Arguably, the question is of even greater importance in emerging economies, where food and fuel expenditures comprise a much larger share of consumption expenditure. Given that food and energy prices are volatile and often driven by local or global supply shocks, inflation management becomes a much trickier exercise for central banks in these countries. Questions related to whether a central bank should look through or react to inflation movements that are driven by supply-side developments are thus recurrent and highly germane in emerging economies.[5]

Our work connects with four distinct but related strands of the literature. The first two of these strands are the voluminous literature on NK macroeconomic models and the nascent literature on the recent global inflation episode. Comprehensive summaries of the former can be found in Galí (2015) and Woodford (2004), amongst many others, while Reis (2022) provides an excellent overview of the various hypotheses regarding the recent episode, the role of the inflation anchor and the associated challenges for central banks.

Another related literature focuses on bounded rationality, LKT, and their implications for macroeconomics. In experimental game theory, an extensive body of laboratory experiments offer evidence for LKT, with most estimates pointing to low k values that place agents far from the RE benchmark (among many others, see Arad and Rubinstein, 2012; Costa-Gomes et al., 2013; Kneeland, 2015). Moreover, LKT and the closely related concept of "reflective expectations" have now been applied to a wide range of macroeconomic questions in both monetary contexts (e.g., Farhi and Werning, 2019; García-Schmidt and Woodford, 2019; Iovino and Sergeyev, 2023) and fiscal ones (Bianchi Vimercati, Eichenbaum, and Guerriro, 2021).[6]

Finally, our findings regarding soft landings connect with the literature on factors that contribute to successful disinflations (e.g., Hazell et al., 2022; Sargent, 1982). Within this literature, a common theme is that effective expectation management can make it possible for disinflations to occur faster than slack measures would suggest,[7] with the expectational mechanisms commonly emphasized having to do with perceptions around the long-term credibility of policy-makers. Though such mechanisms clearly contributed to past disinflations, their applicability to the recent episode is less clear given that long-run inflation expectations were relatively stable over this episode. In this sense, one of our contributions is to show how LKT gives rise to a complementary mechanism for how disinflations can be achieved without large shifts in *both* slack *and* long-term expectations.[8]

The rest of the paper is organized as follows. Section 2 presents the model, while Section 3 formalizes the optimal policy problem and presents special cases of the problem that can be solved

[5] As noted in a previous footnote, the productivity shocks on which the model focuses can be interpreted as the mirror image of oil price shocks, which are one of the more volatile and frequent shocks hitting both emerging and advanced economies. See the Online Appendix for details.

[6] In the Online Appendix, we confirm that our main results hold under reflective expectations and also do so under the "cognitive hierarchy" framework proposed by Camerer et al. (2004).

[7] See also Reis (2021) on this issue.

[8] Our paper is also related to some recent work by Lorenzoni and Werning (2023), who study the phenomenon of wage-price spirals in a general setting and explore implications for policy. One way in which our work is distinct from theirs is that we focus on an environment with boundedly rational agents.

analytically. Section 4 presents the full dynamic policy problem and numerical simulations of the optimal policy. Section 5 explores the consequences of policy errors, while Section 6 discusses demand shocks. Section 7 then elaborates on the model's interpretation of the recent inflation surge, while Section 8 concludes. Most of our proofs are collected in an Appendix, and additional material is provided in an Online Appendix.

2 Model

We consider a simple economy in which the key players are: (i) a set of households that supply labour services on a monopolistically competitive basis; (ii) a set of firms that supply goods on a monopolistically competitive basis; and (iii) a central bank that sets monetary policy with an eye toward minimizing an ad-hoc loss function.

As we stressed in the introduction, an important element will be that wages are set one period in advance, while prices are flexible, and this will be key to delivering a Phillips curve that differs from canonical NK models and helps to rationalize why looking through supply shocks can sometimes be warranted. The absence of multi-period nominal rigidities will also allow us to focus on the cross-sectional coordination problems that arise among price- and wage-setters at each point in time. In this sense, the model can be seen as being closer in spirit to the original environment of Blanchard and Kiyotaki (1987).

Before we turn to the model's components, it is important to note that expectation operators should not be assumed to be rational unless this is specifically indicated, since our ultimate goal is to understand the policy implications of different expectation formation processes. To be as clear as possible in this area, we will attach superscripts to all expectations to convey relevant information about *whose* expectations they represent, or *how* they are being formed.

2.1 Firms

We assume a two-stage production process, with distinct final and intermediate goods. The details on each stage are as follows.

2.1.1 Final good producers

Final goods are produced from a unit measure of intermediate goods using a technology of the form

$$Y_t = \left(\int_0^1 Y_{ft}^{\frac{\gamma-1}{\gamma}} df \right)^{\frac{\gamma}{\gamma-1}},$$

where Y_t denotes output of final goods; Y_{ft} denotes purchases of intermediate good $f \in [0,1]$; and $\gamma > 1$ denotes the elasticity of substitution among intermediate goods. Assuming that final goods producers are perfectly competitive and make zero profits, their demand for intermediate good f is then given by

$$Y_{ft} = Y_t \left(\frac{P_{ft}}{P_t} \right)^{-\gamma}, \tag{2.1}$$

where

$$P_t \equiv \left(\int_0^1 P_{ft}^{1-\gamma} df \right)^{\frac{1}{1-\gamma}}$$

measures the aggregate price level.

2.1.2 Intermediate good producers

Each intermediate good is generated from a unit measure of labour services using a technology of the form

$$Y_{ft} = \theta_t \left(\int_0^1 N_{sft}^{\frac{\rho-1}{\rho}} ds \right)^{\frac{\rho}{\rho-1}},$$

where N_{sft} denotes hiring of service $s \in [0,1]$; $\rho > 1$ denotes the elasticity of substitution among services; and θ_t denotes an aggregate productivity level, which we assume to be stochastic and the only source of randomness in the economy. While the particular form of the process governing θ_t is not critical to our results, it will sometimes be convenient to assume that θ_t follows a random walk in logarithms – specifically,

$$\ln \theta_t = \ln \theta_{t-1} + \epsilon_t,$$

where ϵ_t is an i.i.d. normal shock with mean zero and variance σ_θ^2.

We assume that intermediate good producers take wages as given but set their own prices on a monopolistically competitive basis. As a result of the wage-taking assumption, a given producer's demand for labour service s is given by

$$N_{sft} = \left(\frac{Y_{ft}}{\theta_t}\right)\left(\frac{W_{st}}{W_t}\right)^{-\rho},$$

where

$$W_t \equiv \left(\int_0^1 W_{st}^{1-\rho} ds\right)^{\frac{1}{1-\rho}} \tag{2.2}$$

measures the aggregate wage level. The producer's implied profits are in turn given by

$$P_{ft}Y_{ft} - \int_0^1 W_{st}N_{sft}ds = \left(P_{ft} - \frac{W_t}{\theta_t}\right)Y_{ft}. \tag{2.3}$$

Turning to the pricing behaviour of intermediate good producers, our key assumption here is that prices are flexible in the sense that firms can adjust them every period and do so after observing the current level of productivity. As a result, the optimization problem facing a given firm involves choosing (Y_{ft}, P_{ft}) to maximize the profit function on line 2.3, subject to the demand curve on line 2.1. This leads to a standard mark-up rule of the form

$$P_{ft} = \frac{\gamma}{\gamma - 1}\frac{W_t}{\theta_t}, \tag{2.4}$$

implying an inverse relationship between prices and productivity that will carry over to an inverse relationship between productivity and inflation at the aggregate level. Another corollary of this mark-up rule is that all intermediate good producers behave identically in the sense that $(P_{ft}, Y_{ft}) = (P_t, Y_t)\ \forall f \in [0, 1]$, so overall demand for labour service s is given by

$$\int_0^1 \left(\frac{Y_{ft}}{\theta_t}\right)\left(\frac{W_{st}}{W_t}\right)^{-\rho} df = \left(\frac{Y_t}{\theta_t}\right)\left(\frac{W_{st}}{W_t}\right)^{-\rho}. \tag{2.5}$$

2.2 Households

We assume a unit measure of households, which are indexed by $h \in [0, 1]$ and each supply a range of labour services. A given household's utility is given by

$$\mathbb{E}_t^{HH(h)}\left[\sum_{\tau=0}^{\infty}\beta^{\tau}\left(\ln C_{h,t+\tau} - \eta\int_0^1 \tilde{N}_{s,h,t+\tau}ds\right)\right], \tag{2.6}$$

where $\mathbb{E}_t^{HH(h)}(\cdot)$ denotes the expectations of the particular household in question; $\beta \in (0,1)$ is a discount factor; $\eta > 0$ is a parameter scaling the disutility of labour; C_{ht} denotes the household's consumption; and $\tilde{N}_{sht}$ denotes the total hours of work that the household supplies under labour service s. The tilde above this last variable is thus meant to distinguish a "supply-side" labour quantity from the "demand-side" quantities considered in our characterization of firms' behaviour, and this notational convention will be maintained throughout the remainder of this section.

When optimizing, households face a budget constraint of the form

$$\int_0^1 W_{st}\tilde{N}_{sht}ds + D_t + T_t + B_{h,t-1} = P_t C_{ht} + \frac{B_{ht}}{1+i_t}, \tag{2.7}$$

where D_t denotes dividends received from firms (which are assumed to be distributed equally across households); T_t denotes transfers received from the government (which are distributed on a similar basis); B_{ht} denotes the face value of the household's purchases of risk-free, one-period bonds; and i_t denotes the nominal interest rate on those bonds.

We assume that supply decisions for each labour service are made by service-specific unions, while the household's overall consumption-saving decision is delegated to a household head. The details on these decisions are as follows.

2.2.1 Consumption-saving decision

The consumption-saving decision is assumed to take place after observing the current level of productivity in the economy. This leads to an Euler equation of the form

$$\beta \mathbb{E}_t^{HH(h)}\left(\frac{C_{ht}}{C_{h,t+1}}\frac{1+i_t}{P_{t+1}/P_t}\right) = 1, \tag{2.8}$$

where the subscript in the expectation $\mathbb{E}_t^{HH(h)}(\cdot)$ indicates that the current productivity level θ_t is part of the information set to which the household head had access when forming this expectation.

2.2.2 Labour-supply decision

Turning to labour unions, we assume that they set the wages for their respective services on a monopolistically competitive basis, taking as given the behaviour of all other unions, along with that of household heads. Unlike our treatment of prices, we also assume a modest degree of wage stickiness in the sense that unions can adjust their wages every period but must do so before observing the current level of productivity.

For the union representing a given service $s \in [0,1]$, the relevant program involves choosing a nominal wage rate W_{st}, which the labour demand curve on line 2.5 will then associate with a certain number of hours $\tilde{N}_{st}$ to be worked by each of the union's members – i.e., $\tilde{N}_{hst} = \tilde{N}_{st} = (Y_t/\theta_t)(W_{st}/W_t)^{-\rho}\ \forall h \in [0,1]$. When making this decision, the union is assumed to value wages and the disutility of labour based on the marginal preferences of its members. This leads to an objective of the form

$$\mathbb{E}_{t-1}^{WS(s)} \left(W_{st}\tilde{N}_{st} \int_0^1 \frac{1}{P_t C_{ht}} dh - \eta \tilde{N}_{st} \right), \tag{2.9}$$

where $\mathbb{E}_{t-1}^{WS(s)}(\cdot)$ denotes the expectation of the union setting wages for service s, conditional on an information set that excludes θ_t; and the integral $\int_0^1 [1/(P_t C_{ht})]dh$ measures the average household's marginal utility of nominal income. The implied wage-setting rule can then be written as

$$W_{st} = \frac{\rho\eta}{\rho - 1} \frac{\mathbb{E}_{t-1}^{WS(s)}\left(\frac{Y_t}{\theta_t}\right)}{\mathbb{E}_{t-1}^{WS(s)}\left(\frac{Y_t}{\theta_t}\int_0^1 \frac{1}{P_t C_{ht}} dh\right)}. \tag{2.10}$$

The expectations on the right-hand side of this last expression will be key objects in our analysis, and we will entertain a range of assumptions on how they are formed. However, a key feature common to all of the models of expectation formation that will be considered in the main text is that they each predict that all wage-setters should share common, *though potentially incorrect*, expectations on all aggregate outcomes in the economy[9] – i.e., $\mathbb{E}_{t-1}^{WS(s)}(X_t) = \mathbb{E}_{t-1}^{WS(\tilde{s})}(X_t)$ for an arbitrary aggregate variable X_t and any pair of wage-setters $(s, \tilde{s}) \in [0,1]^2$. Given the form of the expectations on the right-hand side of line 2.10, this implies that all wage-setters should behave identically in the sense of selecting a common wage rate – i.e., $W_{st} = W_t\ \forall s \in [0,1]$. This allows us to re-express the wage-setting condition on line 2.10 in terms of the aggregate wage rate and a set of common expectations $\mathbb{E}_{t-1}^{WS}(\cdot)$ that describe the subjective beliefs of all wage-setters in the economy:

$$W_t = \frac{\rho\eta}{\rho - 1} \frac{\mathbb{E}_{t-1}^{WS}\left(\frac{Y_t}{\theta_t}\right)}{\mathbb{E}_{t-1}^{WS}\left(\frac{Y_t}{\theta_t}\int_0^1 \frac{1}{P_t C_{ht}} dh\right)}. \tag{2.11}$$

[9] At the same time, we stress that common expectations are not critical to our results. For example, we show in the Online Appendix that our main results carry over to a version of the model in which we allow for heterogeneity in wage-setters' expectations using the "cognitive hierarchy" framework of Camerer et al. (2004).

2.3 Phillips and IS curves

Before turning to the model's policy side, it is now convenient to derive the Phillips and IS curves associated with the private-sector outcomes above. As a first step, we also solve for the "natural" levels of output and employment that would prevail in a counterfactual economy with flexible wages.

2.3.1 A symmetric, flexible-wage benchmark

Consider a counterfactual economy in which wages are flexible in the sense of being set after observing the current level of productivity, and in which consumption is equally distributed across households. This is equivalent to replacing expectation operators with identity maps on the right-hand side of the wage-setting condition on line 2.11, while also setting $C_{ht} = Y_t \; \forall h \in [0, 1]$. This leads to a new, flexible wage-setting condition of the form

$$W_t = \frac{\rho\eta}{\rho - 1} P_t Y_t.$$

At the same time, the mark-up rule on line 2.4 can be written as follows after multiplying both sides by total output and defining $N_t \equiv Y_t/\theta_t$:

$$P_t Y_t = \frac{\gamma}{\gamma - 1} W_t N_t.$$

Combining these last two expressions then allows us to pin down the natural level of employment:

$$N_t = \frac{\rho - 1}{\rho\eta} \frac{\gamma - 1}{\gamma} \equiv \bar{N}.$$

The natural level of employment is thus constant and intuitively decreasing in the disutility of labour, as well as the degrees of monopoly distortion assumed in the goods and labour markets. The implied level of potential output is $\bar{Y}_t \equiv \theta_t \bar{N}$.

2.3.2 Phillips curve

We now turn our attention back to the "true" wage-setting condition on line 2.11. By combining this expression with the mark-up rule on line 2.4 and then linearizing around the symmetric, flexible-wage benchmark derived above, we reach a Phillips curve of the form

$$\hat{\pi}_t = \mathbb{E}^{WS}_{t-1}(\hat{N}_t) + \mathbb{E}^{WS}_{t-1}(\hat{\pi}_t) - \hat{\theta}_t, \tag{2.12}$$

where $\hat{\pi}_t \equiv \ln P_t - \ln P_{t-1} - \ln(1+\pi^*)$ denotes the deviation of inflation from some target value π^*; $\hat{N}_t \equiv \ln N_t - \ln \bar{N}$ denotes the employment gap; and $\hat{\theta}_t \equiv \ln \theta_t - \mathbb{E}_{t-1}^{WS}(\ln \theta_t)$ denotes the deviation of productivity from its expected value and can therefore be interpreted as a supply shock.

Some important features of this Phillips curve should be noted. The first is that the residual in equation 2.12 is a "true" supply shock rather than a mark-up shock. This is a key dimension in which this Phillips curve differs from the canonical NK Phillips curve and is a consequence of our focus on wage rigidities; had we instead assumed sticky prices but flexible wages, then it can easily be confirmed that this term would not appear in the Phillips curve. At the same time, the fact that the expression on line 2.12 can be derived without taking a specific stand on the way that wage-setters form their expectations is a consequence of our focusing on a setting where nominal frictions arise from a Lucas (1972)-style assumption that wages are set one period in advance. As explained in Preston (2005), this is a convenient property of Lucas-style nominal rigidities, whereas the Calvo (1983)- or Rotemberg (1982)-style rigidities typically assumed in the NK literature make the precise form of the implied supply curve more sensitive to the details of how private-sector expectations are formed.

2.3.3 IS curve

Turning to the model's demand side, the same arguments presented in Preston (2005) make it difficult to derive a single IS curve that holds across a range of potential choices on the way that one could specify the household expectations entering into the Euler equation on line 2.8. As a result, we pursue a two-pronged strategy under which we: (i) first focus on the standard case where household expectations are fully rational for the purposes of the main text; then (ii) use Section A of the Online Appendix to show how our results generalize to a broader range of expectational assumptions. For the purposes of the main text, this leads to a standard IS curve of the following form after linearizing equation (2.8) and accounting for market clearing:

$$\hat{N}_t = \mathbb{E}_t^{RE}(\hat{N}_{t+1}) - [\hat{\imath}_t - \mathbb{E}_t^{HH}(\hat{\pi}_{t+1})], \tag{2.13}$$

where $\mathbb{E}_t^{RE}(\cdot)$ denotes a rational expectation; and $\hat{\imath}_t \equiv \ln(1+i_t) - [\ln(1+\pi^*) - \ln \beta]$ denotes the deviation of the nominal rate from its steady-state value. However, we stress from the outset that the particular assumptions made with regard to household expectations have no bearing on our main results (and will shortly expand on this point in Subsection 2.5).

2.4 Monetary and fiscal policy

We now close the model by specifying the behaviour of the central bank and fiscal authority. The details on each are as follows.

2.4.1 Central bank

On the monetary policy side, we could immediately specify an interest rate rule and later optimize on the parameters in that rule. However, it will be more convenient to think of the central bank as directly controlling the level of employment and setting it as a (potentially time-varying) function of the prevailing rate of inflation.

This will involve a feedback rule of the form

$$\hat{N}_t = -\phi_t \hat{\pi}_t, \tag{2.14}$$

where ϕ_t is a policy stance reflecting the rate at which policy-makers are prepared to tighten in response to off-target inflation. As we detail in our next section, policy-makers are assumed to update this policy stance each period, announcing its new value before observing the current level of productivity, and doing so with an eye toward minimizing an ad-hoc loss function. By allowing for time variation in the policy stance and the implied level of feedback from $\hat{\pi}_t$ to $\hat{N}_t$, an approach along these lines will ultimately enable us to identify the conditions under which policy-makers find it optimal to "pivot" in the sense of shifting from low values of ϕ_t at which they do not respond much to inflation to high values associated with much stronger reactions.

One way to interpret the feedback rule above is by forward-solving the IS curve – i.e.,

$$\hat{N}_t = -\mathbb{E}_t^{RE}\left[\sum_{\tau=0}^{\infty}\left(\hat{\imath}_{t+\tau} - \hat{\pi}_{t+\tau+1}\right)\right].$$

This equation allows us to express the current employment gap as a negative function of the cumulative sum of all expected future deviations of the real interest rate from its steady-state value. As a result, this cumulative sum can be viewed as a measure of the total stimulus being provided by monetary policy, and equation (2.14) can then be interpreted as the product of a rule according to which policy-makers adjust this measure of stimulus as a time-varying function of the prevailing inflation rate – that is,

$$\phi_t \hat{\pi}_t = \mathbb{E}_t^{RE} \left[\sum_{\tau=0}^{\infty} \left(\hat{\imath}_{t+\tau} - \hat{\pi}_{t+\tau+1} \right) \right]. \tag{2.15}$$

Given the importance that real-world policy-makers place on communicating a policy stance that goes beyond the current value of the policy rate, an interpretation along these lines can be viewed as more realistic and encompassing relative to more standard policy rules.

2.4.2 Fiscal authority

Turning finally to fiscal policy, we assume for concreteness that the government operates a balanced budget each period and keeps bonds in zero net supply. However, none of these assumptions are essential to our results, which go through for a wide range of Ricardian fiscal policies.

2.5 Summary of equilibrium conditions

To summarize, the model's key variables are the inflation gap $\hat{\pi}_t$, employment gap $\hat{N}_t$, nominal rate gap $\hat{\imath}_t$, and policy stance ϕ_t. In equilibrium, these four unknowns are jointly pinned down by the Phillips and IS curves on lines 2.12 and 2.13, the policy feedback rule on line 2.14, and the solution to a policy problem on which we elaborate in the next section.

A key property of this system is that the model can be solved recursively in the sense that one can first solve for $(\hat{\pi}_t, \hat{N}_t, \phi_t)$ independent of the IS curve, then use the IS curve to back out a value for $\hat{\imath}_t$. This is the approach that we take below, and it importantly makes all of our main results independent of the details of the IS curve, including the way that households are assumed to form the expectation $\mathbb{E}_t^{HH}(\cdot)$ at the time they make their consumption-saving decisions. Instead, the key private-sector expectations in our analysis will be those of wage-setters, and much of our attention will be on how these are influenced by monetary policy under different models of expectation formation.

3 Expectations and optimal policy

We now consider the design of optimal monetary policy in the model economy, and how it is shaped by different assumptions on the expectation formation process in the private sector. However, before tackling these issues, it will be useful to lay out the theories of expectation formation that we will consider and the outcomes to which they would lead for arbitrary values of the policy stance ϕ_t. Two additional theories of boundedly rational expectation formation are considered in the Online Appendix, where we show that our main results also carry over to these cases.

3.1 Expectations

3.1.1 Rational expectations (RE)

If we assume that wage-setters' expectations are rational, then taking $(t-1)$-dated expectations on both sides of the Phillips curve on line 2.12 gives $\mathbb{E}_{t-1}^{WS}(\hat{N}_t) = 0$. Combining this result with the feedback rule on line 2.14 then further yields $\mathbb{E}_{t-1}^{WS}(\hat{\pi}_t) = 0$ so long that $\phi_t > 0$. Substituting these results back into equations (2.12) and (2.14), we see that

$$(\hat{\pi}_t, \hat{N}_t) = (-\hat{\theta}_t, \phi_t\hat{\theta}_t),$$

so inflation collapses to an i.i.d. process and lies above (below) its target value whenever the supply shock is negative (positive). Any $\phi_t > 0$ thus suffices to fully anchor inflation expectations at target, and employment is also fully stabilized in the limit $\phi_t \to 0$.

3.1.2 Adaptive expectations (AE)

For this case, we consider the simplest possible version of adaptive expectations, with

$$[\mathbb{E}_{t-1}^{WS}(\hat{\pi}_t), \mathbb{E}_{t-1}^{WS}(\hat{N}_t)] = (\hat{\pi}_{t-1}, \hat{N}_{t-1}). \tag{3.16}$$

Under this assumption, the inflation and employment gaps are respectively given by the following:

$$\hat{\pi}_t = \hat{N}_{t-1} + \hat{\pi}_{t-1} - \hat{\theta}_t$$
$$\hat{N}_t = -\phi_t(\hat{N}_{t-1} + \hat{\pi}_{t-1} - \hat{\theta}_t).$$

As a result, inflation collapses to a random walk in the special case where policy-makers opt to fully look through inflation pressures in the sense of setting $\phi_t = 0$. Contrasting this result with the rational case above illustrates how the mapping between policy and inflation outcomes can vary drastically depending on the nature of private-sector expectations.

Of course, the very stark form of adaptive expectations on line 3.16 could be replaced with a variety of more sophisticated backward-looking rules, including ones based on constant- or decreasing-gain learning, or endogenous-gain learning in the recent spirit of Carvalho et al. (2023) and Gatí (2023). In Section E of the Online Appendix, we explore a range of these alternatives and confirm that our main results are robust to these extensions.

3.1.3 Level-k thinking (LKT)

Turning finally to LKT, our goal here is to consider a more general framework for expectation formation that can accommodate RE and AE within a broader range of possibilities. LKT provides one such framework.

To motivate LKT, it is useful to note from the wage-setting condition on line 2.10 that wage-setters in this economy face an important coordination problem. This is because they each need to forecast aggregate outcomes that depend on the choices of all other wage-setters via the aggregate wage W_t entering into firms' mark-up rule. As a result, each wage-setter must "forecast the forecasts" of all other wage-setters.

Under LKT, "mutual forecasting" problems of this sort are formalized as iterative processes. At the first stage $\tilde{k} = 1$, agents posit some initial guess on the expectations of all other agents and compute the aggregate outcomes that would emerge in this case. At the next stage $\tilde{k} = 2$, the guess on others' expectations is updated to reflect the aggregate outcomes implied by the previous guess and used as the initial guess of a new iteration. This process is then repeated up to some finite $\tilde{k} = k$, with each iteration thus accounting for a new layer of higher-order expectations. The restriction to a finite number of iterations reflects some bounded computing power on the part of individual agents on account of limited resources or capacity for forecasting others' behaviour.

In the case of our model economy, the first stage of this process involves positing some initial (level-zero) guesses $[\mathbb{E}^0_{t-1}(\hat{\pi}_t), \mathbb{E}^0_{t-1}(\hat{N}_t)]$, which are then updated via the Phillips curve and the central bank's feedback rule – i.e.,

$$\mathbb{E}^1_{t-1}(\hat{\pi}_t) = \mathbb{E}^0_{t-1}(\hat{N}_t) + \mathbb{E}^0_{t-1}(\hat{\pi}_t)$$
$$\mathbb{E}^1_{t-1}(\hat{N}_t) = -\phi_t[\mathbb{E}^0_{t-1}(\hat{N}_t) + \mathbb{E}^0_{t-1}(\hat{\pi}_t)],$$

where we have assumed that LKT has no impact on agents' expectations on purely exogenous variables, so $\mathbb{E}^0_{t-1}(\hat{\theta}_t) = \mathbb{E}^1_{t-1}(\hat{\theta}_t) = 0$. Iterating on this recursion up to the final level k then yields expectations of the following form:

$$\mathbb{E}^k_{t-1}(\hat{\pi}_t) = (1-\phi_t)^{k-1}[\mathbb{E}^0_{t-1}(\hat{N}_t) + \mathbb{E}^0_{t-1}(\hat{\pi}_t)] \tag{3.17}$$

$$\mathbb{E}^k_{t-1}(\hat{N}_t) = -\phi_t(1-\phi_t)^{k-1}[\mathbb{E}^0_{t-1}(\hat{N}_t) + \mathbb{E}^0_{t-1}(\hat{\pi}_t)]. \tag{3.18}$$

Throughout, we will assume that the initial guesses seeding this process take the form $\mathbb{E}^0_{t-1}(\hat{\pi}_t) = \hat{\pi}_{t-1}$ and $\mathbb{E}^0_{t-1}(\hat{N}_t) = \hat{N}_{t-1}$ – i.e., agents use a simple adaptive benchmark as the starting point for their reasoning. Under this assumption, feeding the final, level-k expectations above into the model's "true" Phillips curve and feedback rule leaves us with the following equilibrium system:

$$\hat{\pi}_t = (1-\phi_t)^k(\hat{N}_{t-1} + \hat{\pi}_{t-1}) - \hat{\theta}_t$$
$$\hat{N}_t = -\phi_t[(1-\phi_t)^k(\hat{N}_{t-1} + \hat{\pi}_{t-1}) - \hat{\theta}_t].$$

Comparing this system with the equilibrium system that was shown to emerge under AE, we see that the two coincide when $k = 0$. On the other hand, when k goes to infinity, the system collapses to $(\hat{\pi}_t, \hat{N}_t) = (-\hat{\theta}_t, \phi_t\hat{\theta}_t)$ so long that $\phi_t \in (0, 1]$, and we thus recover the RE solution that we described earlier. In these senses, LKT can be viewed as nesting AE and RE as opposite ends of a common iterative spectrum.[10]

3.2 Optimal policy

We now turn our attention to the policy problem that the central bank faces when selecting the policy stance ϕ_t. Here we assume that policy-makers announce their stance before observing the current productivity level θ_t and do so with eye toward minimizing an ad-hoc, quadratic loss function of the form

$$\mathbb{E}^{RE}_{t-1}\left[\sum_{\tau=0}^{\infty}\beta^{\tau}_{CB}\left(\hat{\pi}^2_{t+\tau} + \mu\hat{N}^2_{t+\tau}\right)\right], \tag{3.19}$$

where $\beta_{CB} \in [0, 1)$ is a discount factor that may differ from that of the private sector; $\mu > 0$ is a parameter reflecting policy-makers' relative weight on employment stabilization; and the rational expectation $\mathbb{E}^{RE}_{t-1}(\cdot)$ reflects our assumption that policy-makers are not subject to the expectational

[10] We note here that the a key property of LKT is that each wage-setter's inflation expectation will generally fail to coincide with their expectation of the inflation expectations of all other wage-setters – i.e., $\mathbb{E}^{WS(s)}_{t-1}(\hat{\pi}_t) \neq \mathbb{E}^{WS(s)}_{t-1}[\overline{\mathbb{E}}^{WS}_{t-1}(\hat{\pi}_t)]$, where $\overline{\mathbb{E}}^{WS}_{t-1}(\hat{\pi}_t) \equiv \int_0^1 \mathbb{E}^{WS(\tilde{s})}_{t-1}(\hat{\pi}_t)d\tilde{s}$ denotes the average expectation in the economy. This discrepancy arises because each wage-setter believes that they are more sophisticated than all other wage-setters under LKT. A key advantage of LKT is that it offers a simple, tractable way to capture such tensions in agents' expectations relative to many alternative models of higher-order expectations.

frictions that may arise in the private sector. The optimization is subject to the Phillips curve on line 2.12 and feedback rule on line 2.14 and thus implicitly assumes that the monetary authority can commit to carrying out the policy actions prescribed by ϕ_t even if it may not be optimal to do so once $\hat{\theta}_t$ has been revealed.

This problem can be simplified by defining $x_t \equiv \hat{\pi}_t + \hat{N}_t$ as a measure of the total overheating that the economy experiences in a given period. In this case, the policy problem collapses to choosing ϕ_t to minimize

$$\mathbb{E}_{t-1}^{RE}\left[\sum_{\tau=0}^{\infty}\beta_{CB}^{\tau}(1+\mu\phi_{t+\tau}^{2})\left[(1-\phi_{t+\tau})^{2k}x_{t+\tau-1}^{2}+\sigma_{\theta}^{2}\right]\right], \tag{3.20}$$

subject to the following law of motion for the overheating measure:

$$x_t = (1-\phi_t)[(1-\phi_t)^k x_{t-1} - \hat{\theta}_t]. \tag{3.21}$$

With this simplified policy problem in hand, we will now characterize policy-makers' optimal choice on ϕ_t under RE ($k \to \infty$), AE ($k = 0$), and LKT ($k \in \{1, 2, \ldots\}$).

3.2.1 Optimal policy under RE ($k \to \infty$)

For any sequence of positive policy stances, the loss function under RE reads as

$$\mathbb{E}_{t-1}^{RE}\left[\sum_{\tau=0}^{\infty}\beta_{CB}^{\tau}(1+\mu\phi_{t+\tau}^{2})\sigma_{\theta}^{2}\right],$$

while the law of motion for the overheating measure reads as $x_t = -(1-\phi_t)\hat{\theta}_t$.

With losses thus independent of x_t and strictly increasing in ϕ_t^2, it is clearly optimal for the central bank to make the latter as small as possible. However, since uniqueness issues arise when $\phi_t = 0$ exactly,[11] we interpret the RE economy as one in which it is optimal for policy-makers to set ϕ_t at a positive but "vanishingly low" level, and stress that this leads to full anchoring of inflation expectations while keeping the variance of employment also vanishingly low.

In this sense, RE effectively allows the central bank to fully look through any deviations of inflation from target when setting monetary policy. Intuitively, this reflects the fact that rational

[11] Another way to see the uniqueness issues that arise under RE when $\phi_t = 0$ is that in this case the nominal rate setting consistent with the central bank's preferred "zero gaps" equilibrium would be constant at $\hat{\imath}_t = 0$, but merely announcing this would lead to indeterminacy. When $\phi_t = 0$, these issues could be overcome by supplementing the policy feedback rule with additional information regarding the central bank's intended off-equilibrium behaviour.

agents fully understand that all deviations from the inflation target are transient for even vanishingly positive values of ϕ_t, so their inflation expectations are always anchored at target – even if the economy has recently experienced a long sequence of negative supply shocks. Put differently, even a very weak commitment to keeping inflation at target is enough to keep *expected* inflation at target when the private sector is fully rational.

3.2.2 Optimal policy under AE ($k = 0$)

Under AE, the policy problem can be represented as a Bellman equation of the form

$$V(x_{t-1}) = \min_{\phi_t} \left\{ \left(1 + \mu\phi_t^2\right)\left(x_{t-1}^2 + \sigma_\theta^2\right) + \beta_{CB} E_{t-1}^{RE}\left[V(x_t)\right] \right\},$$

where the minimization is subject to the following law of motion for the overheating measure: $x_t = (1 - \phi_t)(x_{t-1} - \hat{\theta}_t)$.

We solve this problem using a guess-and-verify approach. Specifically, we conjecture a value function of the form $V(x) = a_1 x^2 + a_2 \sigma_\theta^2$. This leads to a solution of the following form:

$$\phi_t = \frac{\beta_{CB} a_1}{\mu + \beta_{CB} a_1}$$
$$a_2 = \frac{1 + \mu\phi_t^2 + \beta_{CB} a_1 (1 - \phi_t)^2}{1 - \beta_{CB}},$$

where a_1 corresponds to the unique positive solution of the following equation:

$$a_1 = 1 + \frac{\mu \beta_{CB} a_1}{\mu + \beta_{CB} a_1}.$$

Several features of this solution are noteworthy. First, as under RE, the optimal policy involves a constant stance and thus precludes any pivoting. Moreover, in the special case of a myopic central bank with $\beta_{CB} = 0$, it is optimal for policy-makers to fully look through supply shocks in the sense of setting $\phi_t = 0$, much as under the RE case. In contrast, a forward-looking central bank will opt for a more hawkish policy under which $\phi_t > 0$. This last result reflects the fact that policy-makers cannot influence the private sector's *contemporaneous* expectations under AE, but they can influence the expectations that agents will carry into *future* periods, and a forward-looking central bank will find it optimal to exploit this channel to at least some extent.

3.2.3 Optimal policy under LKT ($k \in \{1, 2, \ldots\}$)

Under LKT, the policy problem does not admit a closed-form solution, nor easy qualitative characterizations. As a result, we will proceed in two steps. First, we will use the remainder of this section to examine the policy problem when the central bank is assumed to be myopic in the sense that $\beta_{CB} = 0$. Although this is a special case, it will allow us to derive several important features of policy-makers' optimal choice on ϕ_t when viewed as a function of key parameters and the state variable x_{t-1}. In the following section, we then turn our attention to the case of a forward-looking central bank with $\beta_{CB} > 0$. This involves solving the full dynamic version of the policy problem using numerical methods and will allow us to establish the robustness of the qualitative results derived under the myopic special case.

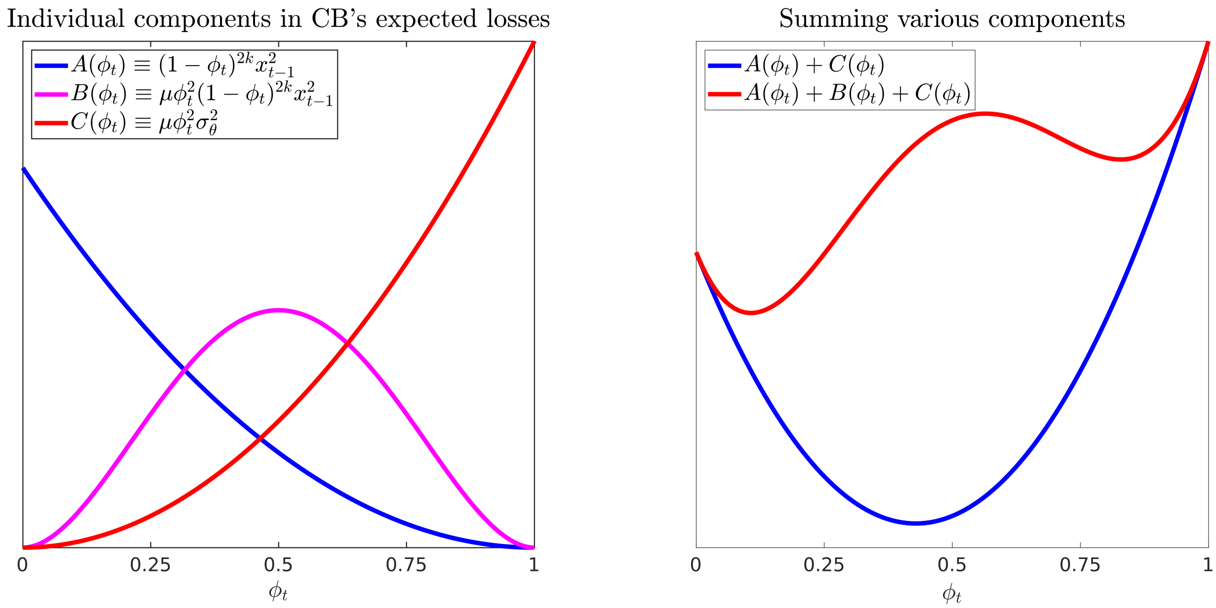
To understand why it is difficult to solve the policy problem even in the myopic case, it is useful to decompose the central bank's expected losses in a given period as follows:

$$(1 + \mu\phi_t^2)\left[(1 - \phi_t)^{2k} x_{t-1}^2 + \sigma_\theta^2\right] = \overbrace{(1 - \phi_t)^{2k} x_{t-1}^2 + \sigma_\theta^2}^{=\mathbb{E}_{t-1}^{RE}(\hat{\pi}_t^2)} + \underbrace{\mu\phi_t^2(1 - \phi_t)^{2k} x_{t-1}^2 + \mu\phi_t^2\sigma_\theta^2}_{=\mu\mathbb{E}_{t-1}^{RE}(\hat{N}_t^2)}. \tag{3.22}$$

Since the solution for private-sector expectations on lines 3.17 and 3.18 implies that $\mathbb{E}_{t-1}^{WS}(x_t) = (1 - \phi_t)^k x_{t-1}$, the term $(1 - \phi_t)^{2k} x_{t-1}^2$ in the decomposition above can be interpreted as the cost of inflation gaps driven by poor anchoring of private-sector expectations, while the term σ_θ^2 can be interpreted as the cost of inflation gaps driven by supply shocks. By extension, the term $\mu\phi_t^2(1 - \phi_t)^{2k} x_{t-1}^2$ can be interpreted as the employment cost of central bank tightening *in response to inflation gaps driven by poor anchoring*, while the final term $\mu\phi_t^2\sigma_\theta^2$ can be interpreted as the employment cost of tightening *in response to inflation gaps driven by supply shocks.*

As we illustrate in Figure 1, this decomposition is useful because it identifies the "cross-term" $\mu\phi_t^2(1 - \phi_t)^{2k} x_{t-1}^2$ as the only obstacle to a simple solution, since suppressing this term would lead to a situation where total expected losses are a single-troughed function of the policy stance. In contrast, the cross-term itself is hump-shaped as a function of the policy stance, since the costs that it captures are best mitigated either by setting ϕ_t very low (and thus committing not to tighten much in the first place), or by setting ϕ_t very high (and thus ensuring that expectations are very well-anchored). In cases where the economy has recently experienced significant overheating (i.e., $x_{t-1}^2 >> 0$) and the policy preference parameter μ is sufficiently large, this cross-term can get large

Figure 1: Decomposition of the central bank's expected losses



Notes. The left-hand panel depicts the components in the decomposition on line 3.22 that depend on the policy stance, while the right-hand panel sums these components with and without the cross-term described in the main text. The specific values assumed in the panels are $(\mu, k, x_{t-1}^2/\sigma_\theta^2) = (10, 1, 7.5)$.

enough to leave a significant signature on the shape of the central bank's overall losses, leading to a W-shaped profile along the lines illustrated in the right-hand panel of Figure 1. Taken altogether, this means that a key challenge when solving for optimal policy under LKT will be to account for the possibility of a non-convex loss function with multiple local extrema.

With these points in mind, we proceed as follows. Letting $\mathcal{L}(\phi_t, x_{t-1}^2)$ denote the expected losses on line 3.22, the first-order condition for a myopic central bank can be written as

$$0.5\mathcal{L}_1(\phi_t, x_{t-1}^2) = \mu\phi_t \left[(1-\phi_t)^{2k} x_{t-1}^2 + \sigma_\theta^2\right] - k(1-\phi_t)^{2k-1} x_{t-1}^2 (1 + \mu\phi_t^2) = 0. \tag{3.23}$$

The richness of the policy problem is then immediately evident, since our earlier points about the shape of the loss function suggest that this equation will generally take the form of a correspondence, rather than a function, when viewed as a mapping from the (squared) overheating measure x_{t-1}^2 to the policy stance ϕ_t. For example, even in the simplest case where $k = 1$, the first-order condition will be cubic in the policy stance and may thus admit multiple solutions.

For these reasons, it will be more convenient to focus on the inverse mapping that the first-order condition implies from ϕ_t back to x_{t-1}^2, since this mapping will take the form of a function:

$$\frac{\mu\phi_t}{(1-\phi_t)^{2k-1}\left[k(1+\mu\phi_t^2)-\mu\phi_t(1-\phi_t)\right]} = \frac{x_{t-1}^2}{\sigma_\theta^2}. \tag{3.24}$$

Letting $f(\phi_t)$ denote the left-hand side of this expression, most of our analysis will therefore focus on exploring the shape of the function $f(\cdot)$.

However, before doing so, it is useful to establish some additional restrictions that the optimal choice on ϕ_t should respect, apart from satisfying equation (3.24):

Lemma 3.1. *The optimal policy stance should satisfy both $\phi_t \in [0,1]$ and $f'(\phi_t) > 0$.*

Proof. To place the optimal stance in the unit interval, we note that all $\phi_t \in (1,\infty)$ satisfy $\mathcal{L}(\phi_t, x_{t-1}^2) > (1+\mu)\sigma_\theta^2 = \mathcal{L}(1, x_{t-1}^2)$, while all $\phi_t \in (-\infty, 0)$ satisfy $\mathcal{L}(\phi_t, x_{t-1}^2) > x_{t-1}^2 + \sigma_\theta^2 = \mathcal{L}(0, x_{t-1}^2)$.

To further see that the optimal stance should satisfy $f'(\phi_t) > 0$, note that equation (3.24) can only hold at points where the denominator on its left-hand side is positive, which is equivalent to $\mathcal{L}_{12}(\phi_t, x_{t-1}^2) < 0$. Totally differentiating equations (3.23) and (3.24) around any such point then yields $d\phi_t/dx_{t-1}^2 = 1/\left[\sigma_\theta^2 f'(\phi_t)\right] = (-1)\mathcal{L}_{12}(\phi_t, x_{t-1}^2)/\mathcal{L}_{11}(\phi_t, x_{t-1}^2)$, so $f'(\phi_t)$ and $\mathcal{L}_{11}(\phi_t, x_{t-1}^2)$ must share a common sign. As a result, the central bank's second-order condition holds only if $f'(\phi_t) > 0$. ■

With these restrictions in mind, we will now use a pair of lemmata to build up a characterization of the function $f(\cdot)$, relegating all remaining proofs to the Appendix. As a first step, we define $d(\phi_t)$ as the term in square brackets in the denominator on the left-hand side of equation (3.24) – i.e., $f(\phi_t) = \mu\phi_t/(1-\phi_t)^{2k-1}d(\phi_t)$. $d(\cdot)$ is a quadratic function whose behaviour over the unit interval can be summarized as follows:

Lemma 3.2. *If policy-makers' relative weight on employment is sufficiently low in the sense that $\mu < 4k(1+k)$, then $d(\phi_t) > 0\ \forall \phi_t \in [0,1]$. Otherwise, the function $d(\cdot)$ has two real roots,*

$$\phi_1^d = \frac{1}{2(1+k)} - \sqrt{\frac{1}{4(1+k)^2} - \frac{k}{\mu(1+k)}} > 0, \ \textit{and}$$
$$\phi_2^d = \frac{1}{2(1+k)} + \sqrt{\frac{1}{4(1+k)^2} - \frac{k}{\mu(1+k)}} \in [\phi_1^d, 1),$$

with $d(\phi_t) < 0$ i.f.f. $\phi_t \in (\phi_1^d, \phi_2^d)$.

Our next step is to turn our attention to the derivative

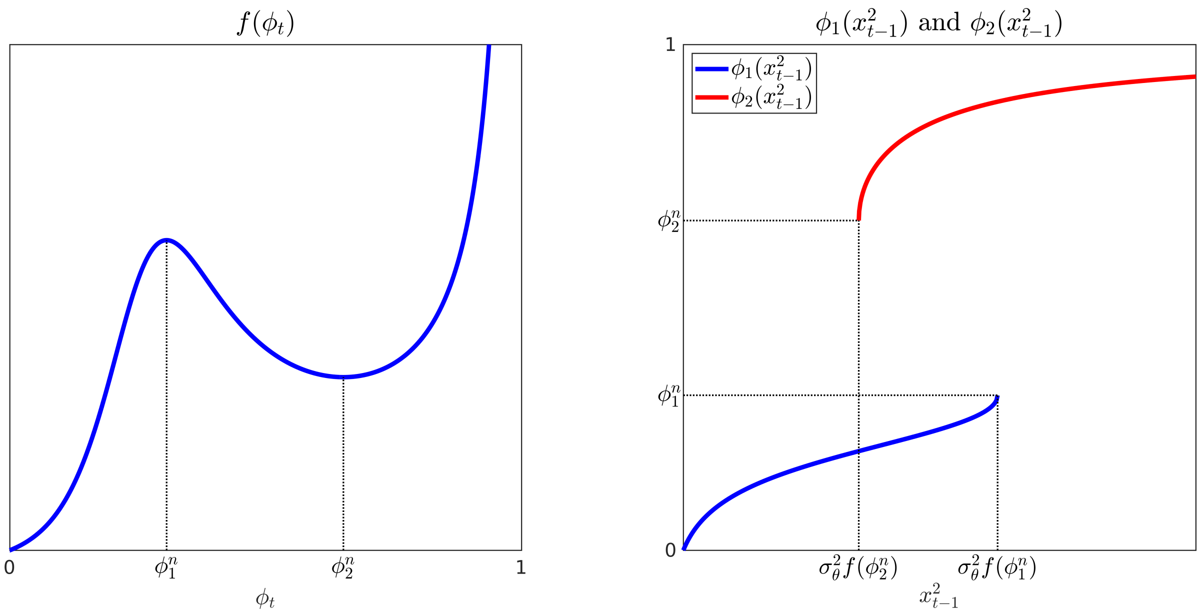
$$f'(\phi_t) = \frac{\mu k \left[2\mu(k+1)\phi_t^3 - 3\mu\phi_t^2 + 2(k-1)\phi_t + 1\right]}{(1-\phi_t)^{2k}\left[d(\phi_t)\right]^2}, \tag{3.25}$$

which shares its sign with the cubic function in its numerator, $n(\phi_t) \equiv 2\mu(k+1)\phi_t^3 - 3\mu\phi_t^2 + 2(k-1)\phi_t + 1$. The key features of this cubic can be summarized as follows:

Lemma 3.3. *There exists some $\overline{\mu} \in (0, 4k(1+k))$ with the following properties:*

- *If $\mu < \overline{\mu}$, then $n(\phi_t) > 0 \ \forall \phi_t \in [0,1]$;*
- *If instead $\mu \geq \overline{\mu}$, then the function $n(\cdot)$ has two roots in the unit interval, $\phi_1^n > 0$ and $\phi_2^n \in [\phi_1^n, 1)$. These roots have the property that $n(\phi_t) < 0 \ \forall \phi_t \in (\phi_1^n, \phi_2^n)$, with $n(\phi_t) > 0 \ \forall \phi_t \in [0, \phi_1^n) \cup (\phi_2^n, 1]$, and are distinct when $\mu > \overline{\mu}$;*
- *Finally, if $\mu \geq 4k(1+k)$ as well, then the roots described above can be ordered as follows relative to those described in Lemma 3.2: $\phi_1^d \leq \phi_1^n \leq \phi_2^d < \phi_2^n$.*

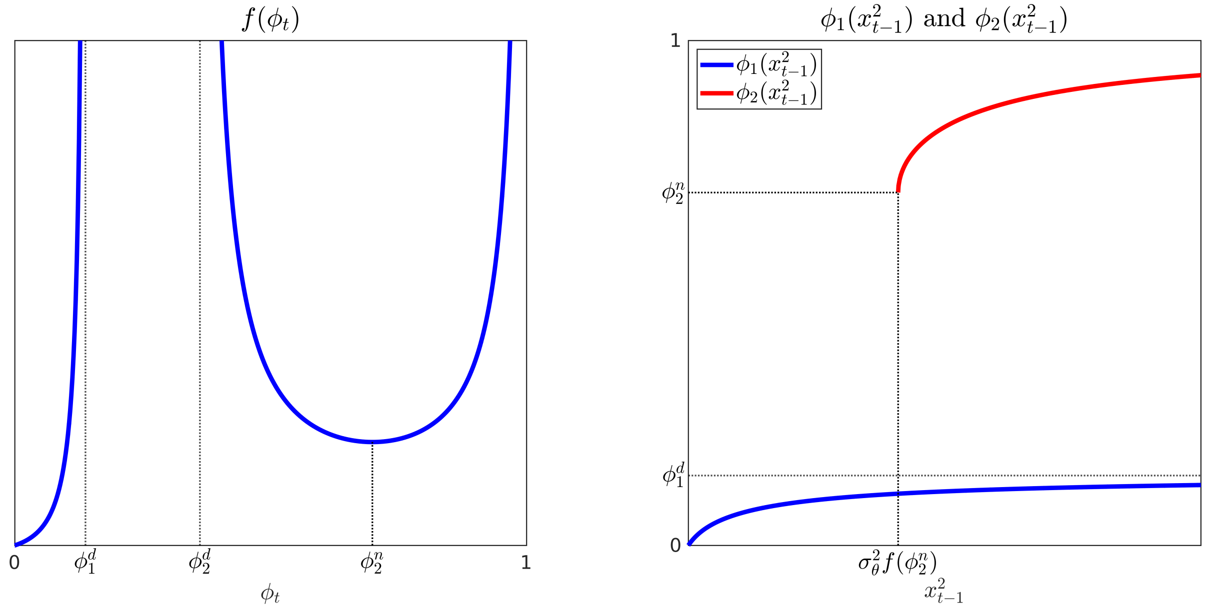
Figure 2: Case $\mu \in (\overline{\mu}, 4k(1+k))$



Notes. The left-hand panel depicts the function on the left-hand side of equation (3.24). The right-hand panel depicts the functions $\phi_1(\cdot)$ and $\phi_2(\cdot)$ described in the main text, each of which corresponds to a local optimum. The parameter values assumed in the panels are $(\mu, k) = (6, 1)$.

Based on these lemmata and the limiting values $f(0) = 0$ and $\lim_{\phi_t \nearrow 1}\{f(\phi_t)\} = \infty$, we can conclude that one of three scenarios must obtain with regard to the shape of the function $f(\cdot)$,

Figure 3: Case $\mu \geq 4k(1+k)$

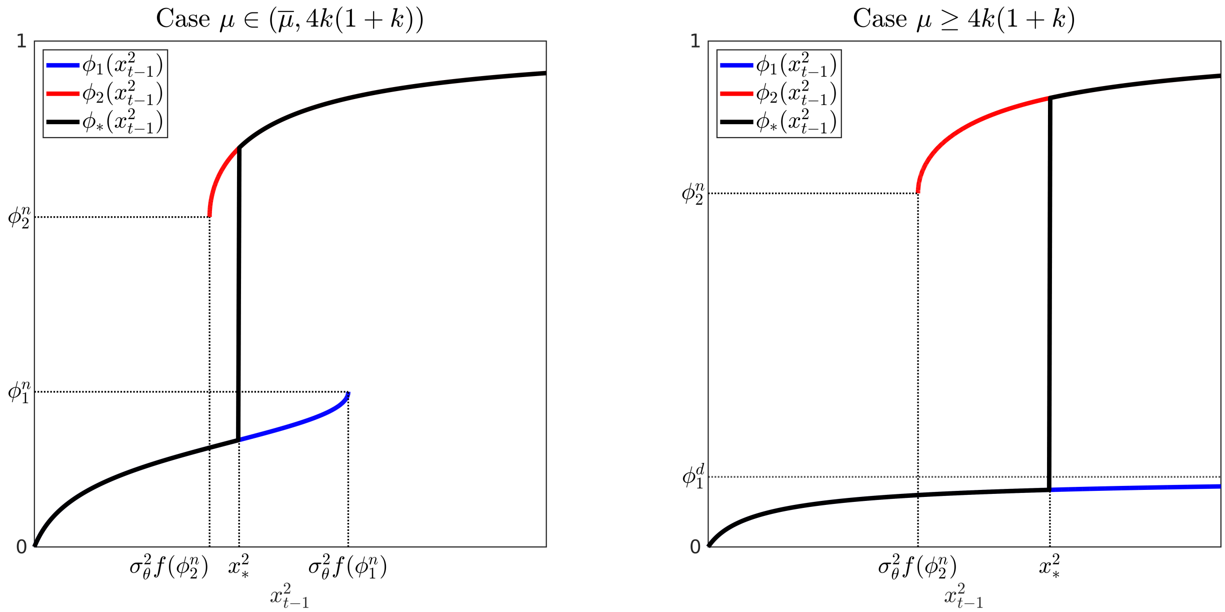


Notes. The left-hand panel depicts the function on the left-hand side of equation (3.24). The right-hand panel depicts the functions $\phi_1(\cdot)$ and $\phi_2(\cdot)$ described in the main text, each of which corresponds to a local optimum. The parameter values assumed in the panels are $(\mu, k) = (10, 1)$.

whose domain we will now restrict to the unit interval. If $\mu \leq \overline{\mu}$, then $f(\cdot)$ is continuous and strictly increasing, and the uniquely optimal policy stance at any level of past overheating x_{t-1} is given by the inverse $f^{-1}(x_{t-1}^2/\sigma_\theta^2)$. If instead $\mu \in (\overline{\mu}, 4k(1+k))$, then $f(\cdot)$ is continuous but non-monotonic, and the optimal policy stance will lie on one of the two branches along which this function is strictly increasing: either a lower branch mapping from $[0, \phi_1^n]$ to $[0, f(\phi_1^n)]$, or an upper branch that maps from $[\phi_2^n, 1)$ to $[f(\phi_2^n), \infty)$ and thus involves strictly higher values for the policy stance. A similar situation obtains if $\mu \geq 4k(1+k)$, except in this case $f(\cdot)$ is discontinuous, and the lower branch now maps from $[0, \phi_1^d)$ to $[0, \infty)$. The latter two scenarios are illustrated in Figures 2 and 3, with $\phi_1(\cdot)$ denoting the implied one-to-one mapping from x_{t-1}^2 to ϕ_t along the lower branch, while $\phi_2(\cdot)$ denotes the analogous mapping along the upper branch.

When $\mu > \overline{\mu}$, this characterization opens up the key possibility that the globally optimal policy stance could respond to changes in x_{t-1}^2 by jumping discretely between the local minima represented by the two branches that we just described. In this respect, it is straightforward to show which branches will be selected when x_{t-1}^2 is either very small or very large. For example, policy-makers will strictly prefer the upper branch when x_{t-1}^2 is very large, since in this case the terms $(1-\phi_t)^{2k}x_{t-1}^2$ and $\mu\phi_t^2(1-\phi_t)^{2k}x_{t-1}^2$ will dominate all others in the loss function on line 3.22.

Figure 4: Optimal pivots around $x_{t-1}^2 = x_*^2$



Notes. The panels depict the globally optimal policy stance, $\phi_*(x_{t-1}^2)$, for a range of values for the state variable x_{t-1}^2 and two cases involving different values for the policy preference parameter μ. The locally optimal stances $\phi_1(x_{t-1}^2)$ and $\phi_2(x_{t-1}^2)$ are also depicted. The parameter values assumed in the left-hand panel are the same as in Figure 2, while those assumed in the right-hand panel are the same as in Figure 3.

If $\phi_*(x_{t-1}^2)$ denotes the globally optimal policy stance at an arbitrary value of x_{t-1}^2, we can thus conclude that $\phi_*(x_{t-1}^2) = \phi_2(x_{t-1}^2)$ when x_{t-1}^2 is sufficiently large. At the same time, we must have $\phi_*(x_{t-1}^2) = \phi_1(x_{t-1}^2)$ when $x_{t-1}^2 \in [0, \sigma_\theta^2 f(\phi_2^n))$, since the upper branch is inaccessible in this range.

Based on these limiting outcomes, we can conclude that when $\mu > \overline{\mu}$ there must exist at least one value for x_{t-1}^2 around which it is optimal for policy-makers to pivot in the sense of jumping from the lower branch to the upper branch. In the Appendix, we further establish a single-crossing property that makes this pivot point unique, leading to our main result:

Proposition 3.1. *The globally optimal policy stance $\phi_*(x_{t-1}^2)$ is strictly increasing in x_{t-1}^2. If $\mu \leq \overline{\mu}$, then it is also continuous. Otherwise, there exists some cutoff $x_*^2 \in [\sigma_\theta^2 f(\phi_2^n), \infty)$ around which the optimal policy stance jumps discontinuously in the sense that $\phi_*(x_{t-1}^2) = \phi_1(x_{t-1}^2)$ $\forall x_{t-1}^2 \in [0, x_*^2)$, but $\phi_*(x_{t-1}^2) = \phi_2(x_{t-1}^2)$ $\forall x_{t-1}^2 \in (x_*^2, \infty)$.*

Figure 4 illustrates the implied policy profile when $\mu > \overline{\mu}$, with policy-makers initially opting for low values of ϕ_t at which they are mostly looking through supply shocks, then suddenly jumping to a much more aggressive stance if shocks cumulate up to a point that drives the overheating measure beyond some threshold level.

The intuition for Proposition 3.1 is that as x_{t-1}^2 rises and puts more pressure on private-sector expectations, it is optimal for policy-makers to compensate by increasing ϕ_t to at least some extent. When μ is relatively low, this can be done in a smooth way. However, when μ is high, a key concern for policy-makers is that smooth approaches will eventually run afoul of the "cross-term" that we emphasized in our discussion of Figure 1 – that is, the term $\mu\phi_t^2(1-\phi_t)^{2k}x_{t-1}^2$ capturing the employment costs of tightening in response to poorly anchored expectations. This is because smoothly ramping up the policy stance will eventually lead to situations where ϕ_t takes intermediate values that are still too low to effectively re-anchor private-sector expectations, but high enough that policy-makers have to engineer significant amounts of slack to offset the impact of those expectations on realized inflation outcomes. Pivots give central banks with high μ values a way to avoid such costly scenarios, and one of the key benefits of the more decisive re-anchoring with which they are associated is precisely that this allows policy-makers to re-stabilize realized inflation without necessitating large employment gaps.

With these points in mind, we use the remainder of this section to elaborate on some additional key features of our solution. In particular, we start with our attention on the sensitivity of private-sector expectations to the policy stance, as measured by the (unsigned) elasticity

$$\left| \frac{d\mathbb{E}_{t-1}^{WS}(x_t)}{d\phi_t} \frac{\phi_t}{\mathbb{E}_{t-1}^{WS}(x_t)} \right| = \frac{k\phi_t}{1-\phi_t} \equiv \zeta(\phi_t), \tag{3.26}$$

and elaborate on the important role that this object plays in driving our results, along with the related role being played by the parameter k controlling the sophistication of private-sector expectations. We then turn our attention to the way that the employment gap behaves around the pivot point and whether pivots can be compatible with soft landings.

Role of the elasticity $\zeta(\phi_t)$ and parameter k. To better understand the role being played by the elasticity $\zeta(\phi_t)$, it is useful to re-express the central bank's first-order condition in a way that allows us to distinguish between the "direct" channels via which the policy stance impacts the economy at a given level of private-sector expectations, versus the "indirect" channels that are intermediated by changes in these expectations. Dividing line 3.23 by $\phi_t^{-1}\mathcal{L}(\phi_t, x_{t-1}^2)$ and re-arranging terms, we get

$$\frac{\mu\phi_t^2}{1+\mu\phi_t^2} = \zeta(\phi_t)\frac{(1-\phi_t)^{2k}x_{t-1}^2}{(1-\phi_t)^{2k}x_{t-1}^2+\sigma_\theta^2}, \tag{3.27}$$

where the left-hand ratio $\mu\phi_t^2/(1+\mu\phi_t^2)$ can be interpreted as the elasticity of the central bank's expected losses with respect to the policy stance, holding private-sector expectations constant, while the right-hand ratio $(1-\phi_t)^{2k}x_{t-1}^2/[(1-\phi_t)^{2k}x_{t-1}^2+\sigma_\theta^2] \equiv r(\phi_t, x_{t-1}^2)$ can be interpreted as the elasticity of expected losses with respect to expectations, now holding direct effects constant. The right-hand side of line 3.27 thus captures the marginal benefits of tightening in terms of better anchoring expectations, while the left-hand side captures the marginal costs in terms of making employment more volatile at a given level of expectations.

What is key here is that while marginal costs and the ratio $r(\phi_t, x_{t-1}^2)$ are respectively increasing and decreasing in ϕ_t, marginal benefits are generally non-monotonic due to the fact the elasticity $\zeta(\phi_t)$ is increasing in ϕ_t – i.e., expectations tend to be more responsive to changes in the policy stance when ϕ_t is high. This critical property of private-sector expectations under LKT is the underlying source of non-convexity in the economy and leads to a situation where there are benefits to tightening by a sufficiently wide margin. Moreover, since the ratio $r(\phi_t, x_{t-1}^2)$ is increasing in x_{t-1}^2, the marginal benefits of tightening are amplified when the economy has recently experienced significant overheating, and this gives policy-makers an incentive to delay large increases in ϕ_t until the overheating measure has reached sufficiently high levels.

At the same time, the fact that the elasticity on line 3.26 is increasing in the "levels of thinking" parameter k helps to explain why monetary policy is so powerful in the model's RE limit. It also suggests that policy-makers facing very large but finite values for k may be able to make do without pivots, while others facing lower values for k may find pivots highly valuable as a way of engineering large changes in private-sector expectations. In the Appendix, we are able to verify and formalize this intuition as follows:

Proposition 3.2. *The threshold $\overline{\mu}$ described in Lemma 3.3 is strictly increasing in the parameter k and satisfies $\lim_{k\to\infty}\{\overline{\mu}\} = \infty$ in the RE limit. As a result, increasing k narrows the range of values for μ within which pivots emerge as a feature of optimal policy.*

Pivots and soft landings. We finally turn to the way that employment behaves around the pivot point. While pivots are associated with large increases in the policy stance, their actual impact on employment is unclear and depends on the balance of the direct and indirect channels that we emphasized when discussing equation (3.27): if the latter channel were very strong, then the adjustment in inflation expectations that pivoting induces could reduce inflation pressures enough that the amount of *actual* slack that policy-makers have to engineer would be relatively low. This tension is reflected in the fact that the (rationally) expected employment gap is given by

$$\mathbb{E}^{RE}_{t-1}(\hat{N}_t) = -\phi_t(1-\phi_t)^k x_{t-1}, \tag{3.28}$$

so its behaviour around the pivot point is driven by how the jump to a higher policy stance impacts the product $\phi_t(1-\phi_t)^k$.

While this impact is difficult to characterize as a general matter, we show in the Appendix that the special case with $k = 1$ level of thinking has the property that the direct and indirect channels offset one another exactly in expectation, leading to no change in $\phi_t(1-\phi_t)^k$ or the expected level of employment around the pivot point. In other words, pivots are compatible with soft landings when $k = 1$, *at least in expectation*, and essentially amount to a "tough talk" strategy that re-anchors expectations in a way that eliminates the need for actual slack.

However, if pivots are expected to lead to soft landings, then why wait to implement them? The answer has to do with the impact of pivots on the variance of the employment gap,

$$\text{var}^{RE}_{t-1}(\hat{N}_t) = \phi_t^2\sigma_\theta^2, \tag{3.29}$$

which increases discontinuously around the pivot point. This reflects the fact that an aggressive policy stance makes employment more sensitive to supply shocks by committing the central bank to respond strongly to off-target inflation, regardless of whether that inflation is driven by expectations or shocks. Optimally-timed pivots under $k = 1$ are thus associated with "risky" or "narrow" soft landings in the sense that they re-stabilize inflation without triggering changes in the *expected* level of employment, but do so at the cost of increasing the uncertainty around employment outcomes.

To summarize:

Proposition 3.3. *If $\mu > \overline{\mu}$, then the pivots described in Proposition 3.1 are associated with discontinuous increases in the variance of the employment gap, as given by equation* (3.29) *above. However, if $k = 1$ as well, then they have no effect on the expected level of employment, as given by equation* (3.28) *above.*

4 Dynamic policy problem: numerical solution

Having shown in our previous section how pivots can emerge as a feature of optimal policy when policy-makers are myopic, we now ask if pivots are robust to extensions in which policy-makers are assumed to be forward-looking. This will involve using numerical methods to solve the full dynamic problem facing a central bank with a non-zero discount factor $\beta_{CB} > 0$.

However, before doing this, it will be useful to generalize the Phillips curve derived in our previous section to allow for an arbitrary slope on the expected employment gap, rather than the slope of one implicit in equation (2.12). This will be helpful in calibrating the numerical model and interpreting its results, in addition to allowing for greater generality.

4.1 Generalizing the Phillips curve

To allow for a non-unit slope in the Phillips curve, we replace the household preferences in the previous section with ones in line with Greenwood, Hercowitz, and Huffman (1988). Specifically, we assume that utility is now given by

$$\mathbb{E}_t^{HH}\left[\sum_{\tau=0}^{\infty}\beta^{\tau}\ln\left(C_{t+\tau}-\eta\theta_t\int_0^1 N_{s,t+\tau}^{1+\chi}ds\right)\right],$$

where $\chi > 0$ is the inverse Frisch elasticity of labour supply. In the Online Appendix, we show that this leads to a modified Phillips curve of the following form, where $\lambda \equiv \chi/(1+\rho\chi)$:

$$\hat{\pi}_t = \lambda\mathbb{E}_{t-1}^{WS}(\hat{N}_t) + \mathbb{E}_{t-1}^{WS}(\hat{\pi}_t) - \hat{\theta}_t,$$

In the same appendix, we further show how non-unit values for the slope parameter λ impact the problem facing policy-makers. This involves re-defining our overheating measure as $x_t \equiv \hat{\pi}_t + \lambda\hat{N}_t$, in addition to defining $\tilde{\mu} = \mu/\lambda^2$ as policy-makers' "effective weight" on stabilizing employment, and $\tilde{\phi}_t \equiv \lambda\phi_t$ as their "effective policy stance" in a given period. Under these definitions, the central bank's problem amounts to choosing $\tilde{\phi}_t$ to minimize

$$\mathbb{E}_{t-1}^{RE}\left[\sum_{\tau=0}^{\infty}\beta_{CB}^{\tau}(1+\tilde{\mu}\tilde{\phi}_{t+\tau}^2)\left[(1-\tilde{\phi}_{t+\tau})^2x_{t+\tau-1}^2+\sigma_\theta^2\right]\right], \tag{4.30}$$

subject to the following law of motion:

$$x_t = (1-\tilde{\phi}_t)[(1-\tilde{\phi}_t)^k x_{t-1} - \hat{\theta}_t]. \tag{4.31}$$

Comparing this program with the original program on lines 3.20 and 3.21 of the previous section, we see that the two coincide when we replace (μ, ϕ_t) in the latter program with $(\tilde{\mu}, \tilde{\phi}_t)$. In this sense, solving the policy problem in the general case of a non-unit slope is equivalent to solving our original program with a modified weight on the central bank's employment objective, though one must be mindful to infer the true policy stance via $\tilde{\phi}_t = \lambda\phi_t$.

This equivalence has several key implications. Most importantly, it implies that all of the qualitative results established in the previous section readily carry over to the case of an arbitrarily sloped Phillips curve. Moreover, since the literature favours values for λ much less than one, it implies a significant widening in the range of values for the "true weight" μ within which the pivoting behaviour described in Proposition 3.1 emerges. To the extent that λ takes values less than one, pivots should also be associated with wider swings in the true policy stance $\phi_t = \tilde{\phi}_t/\lambda$.

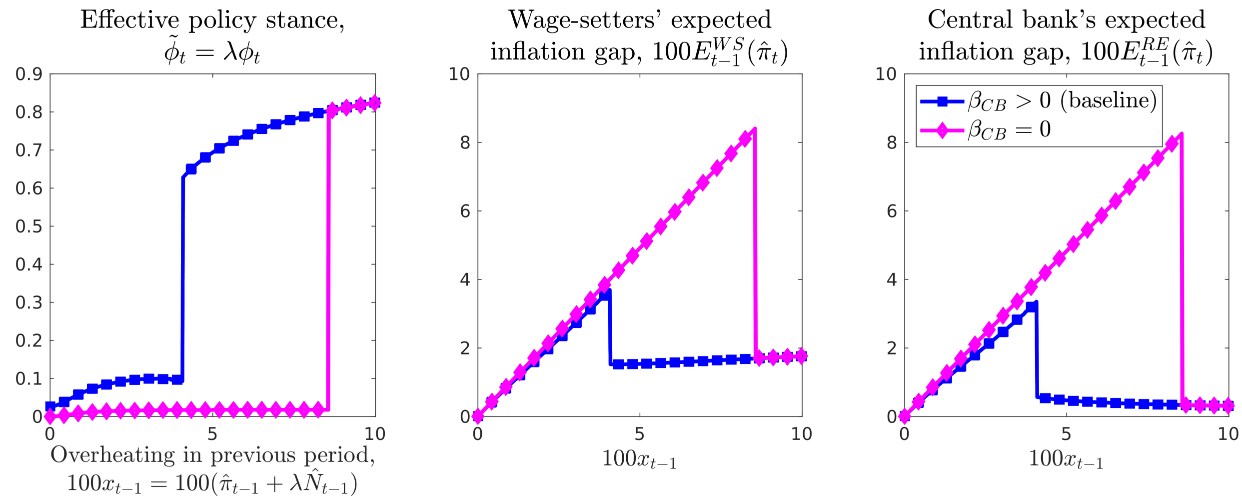
4.2 Baseline parameterization and solution method

We are now ready to solve the generalized policy problem on lines B.45 and B.46 using numerical methods. Given the model's emphasis on wage-setting, coupled with its property that wages are sticky within periods but fully flexible across periods, it is natural to focus on parameterizations that identify model periods with years. We therefore set policy-makers' discount factor to $\beta_{CB} = 0.995$, aligning with the midpoint of Bank of Canada staff's current assessed range of 0 to 1% for the real neutral rate in Canada (Champagne et al., 2023).

As for the model's other parameters, our baseline parameterization will set the relative weight on employment in the central bank's loss function to $\mu = 1$, a natural benchmark consistent with policy-makers placing equal weight on deviations of inflation from target and employment from its natural level. At the same time, we set the Phillips curve slope to $\lambda = 0.092$, a value broadly in line with Canadian estimates in Djeutem et al. (2022) and Wagner et al. (2022). We also assume $k = 2$ levels of thinking in the private sector, placing us in the middle of the range of $k = 1$ to 4 considered in Farhi and Werning (2019) and in line with the experimental literature (which generally estimates k values in the low single digits). The only remaining parameter is then the supply-shock variance, σ_θ^2, which we calibrate to match the variance of CPI inflation in Canada, computed using a data sample that begins when the country's current 2% inflation target first formally started applying in the mid-1990s.

For a given set of parameters, we solve the model by value function iteration. Details on our solution algorithm are given in the Online Appendix.

Figure 5: Optimal pivots under the baseline parameterization and $\beta_{CB} = 0$



4.3 Results

Results for the baseline parameterization are given by the blue lines in Figure 5, with the left-hand panel depicting the optimal policy stance as a function of the past overheating measure x_{t-1}, and the other panels depicting the implied profiles for the inflation expectations of wage-setters and policy-makers. Pivots clearly remain a key feature of optimal policy, with a threshold value for x_{t-1} around 4% triggering a sudden increase in the policy stance and a sharp re-normalization in inflation expectations.

In fact, pivoting behaviour is not merely *robust* to allowing for a forward-looking central bank – there is a precise sense in which it is actually *strengthened* by higher values for the central bank's discount factor. To see this, note that the magenta lines in Figure 5 report results for a myopic parameterization under which we set $\beta_{CB} = 0$ while keeping all other parameters at their baseline values. Comparing these results against their baseline counterparts, we see that the shift to assuming a non-myopic central bank "pulls forward" the threshold around which policy-makers are prepared to pivot. This greater willingness to pivot is a natural consequence of the fact that a forward-looking central bank internalizes the benefits that stabilizing inflation in the current period will generate in future periods through better-anchored expectations.

With these points in mind, we use the remainder of this section to explore the roles that two additional parameters play in driving our numerical results, namely: (i) the "levels of thinking" parameter k; and (ii) the ratio $\tilde{\mu} = \mu/\lambda^2$ representing the effective weight that policy-makers place on their employment objective.

Role of the parameter k. Figure 6 reports results from varying k between one and three while keeping all other parameters at their baseline values. The figure also reports illustrative results for an extreme parameterization under which we set $k = 40$, well outside the range supported by experimental studies. While the cases involving $k \leq 3$ all lead to pivots around thresholds in the 3-5% range, the case with $k = 40$ has no pivots occurring even when x_{t-1} is as high as 10%. This is consistent with Proposition 3.2's message that policy-makers can make do without pivots when facing a sufficiently sophisticated private sector. In cases where pivots do occur, it is also notable that Figure 6 associates higher values for k with smaller jumps in the policy stance around the pivot point. This pattern reflects the fact that policy-makers facing higher values for k have greater leverage over private-sector expectations and thus do not need to tighten as aggressively to re-anchor expectations.

These points are important partly because a smaller jump in the policy stance around a given pivot point should lead to a situation where employment outcomes are less sensitive to supply shocks following pivots, all else being equal. This suggests that the "risky soft landing" problem that we discussed earlier should be less of an issue when k is high. To confirm this intuition, we use Figure 7 to zoom in on the model's behaviour around the pivot points associated with the $k = 1$ through 3 parameterizations described earlier. In this figure, solid lines depict (rationally) expected inflation and employment outcomes before observing the supply shock $\hat{\theta}_t$, while the surrounding bands depict realized outcomes for a range of potential shock values. Since the horizontal axes in both panels have been normalized so that pivots occur at a value of zero for all parameterizations, this makes it easy to compare parameterizations in terms of how the first and second moments of the employment gap respond to pivots.

In particular, we note that pivots have negligible impacts on expected employment for all parameterizations considered, and that the main employment impact of pivots is instead to increase the uncertainty surrounding employment outcomes – especially when k is small. This indicates that risky soft landings remain a key feature of the model economy, and that higher values for k open up a better policy trade-off between stabilizing expectations and mitigating the variance of employment and associated risk of hard landings.

Role of the parameter $\tilde{\mu}$. We now turn our attention to the effective weight that policy-makers place on minimizing deviations of employment from its natural level, as measured by the ratio $\tilde{\mu} = \mu/\lambda^2$. In Figure 8, we report results for a range of values for this ratio, holding all other

Figure 6: Impact of changing the levels of private-sector thinking, k

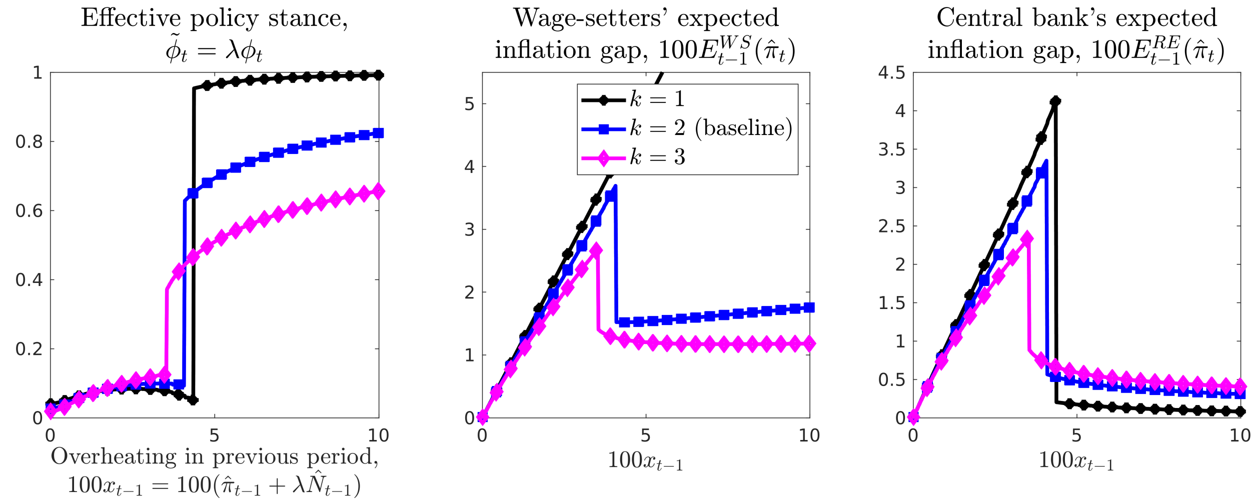


Figure 7: Risky soft landing under different levels of private-sector thinking (k)

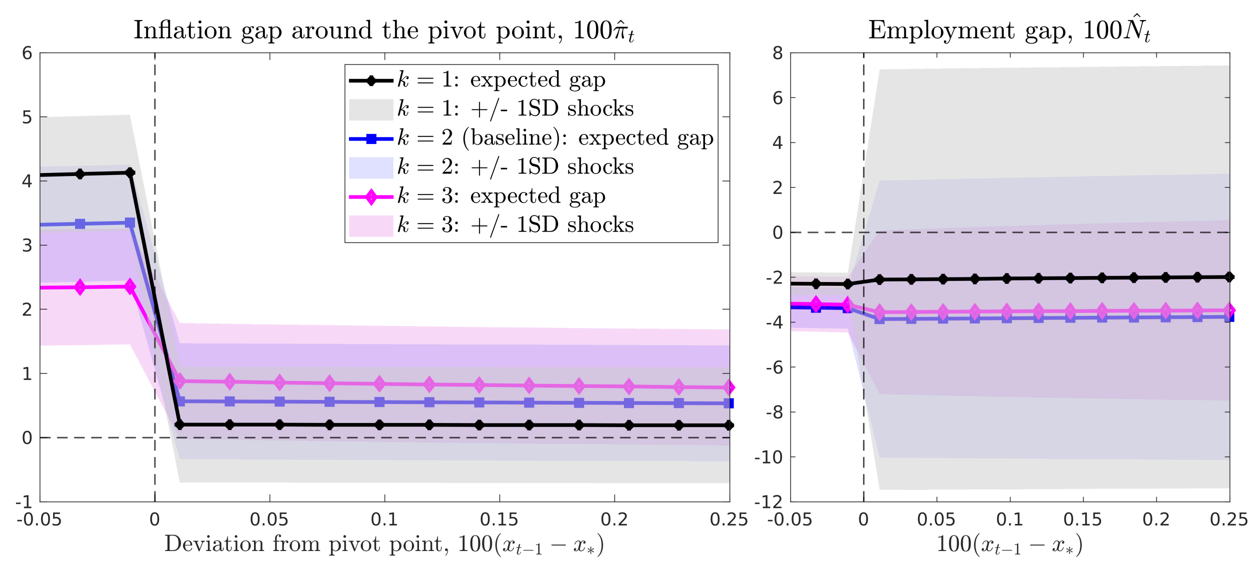
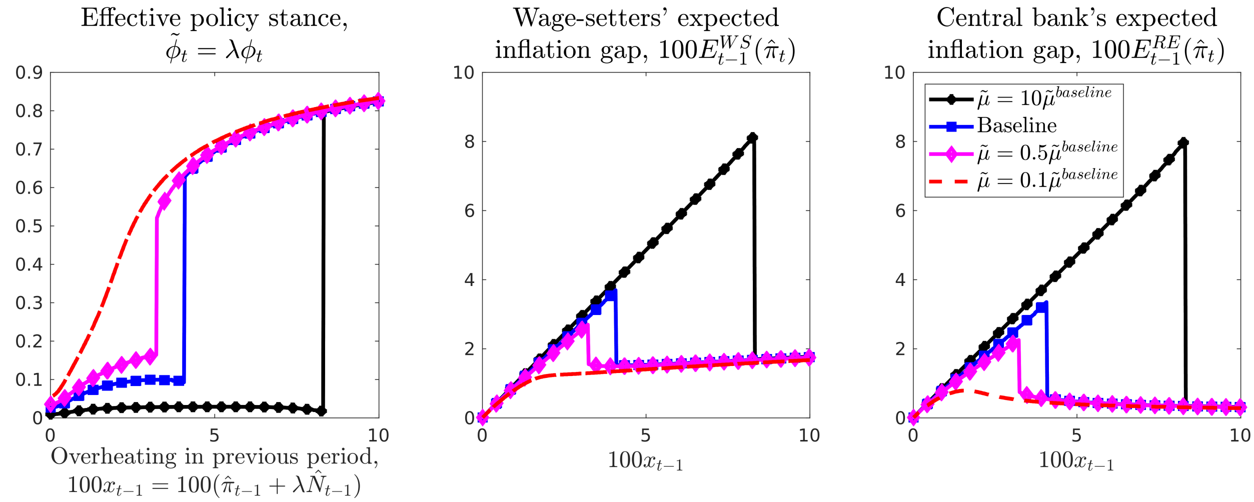


Figure 8: Impact of changing policy-makers' effective weight on employment, $\tilde{\mu} = \mu/\lambda^2$



parameters at their baseline values.

Consistent with our analytical results, the optimal policy stance is discontinuous when $\tilde{\mu}$ is large but takes on a smooth profile for small values of $\tilde{\mu}$, suggesting that $\tilde{\mu}$ must exceed some threshold in order for pivoting to occur. Moreover, conditional on being above this threshold, higher values for $\tilde{\mu}$ are associated with higher pivot points, reflecting a greater willingness to look through periods of high inflation when policy-makers place greater weight on their employment objectives.

While policy-makers thus tend to delay pivots when $\tilde{\mu}$ is high, it is important to note from Figure 8 that in this case pivots will be associated with larger swings in the policy stance should x_{t-1} eventually reach levels high enough to trigger them. In this sense, pivots "come later", but they are also more pronounced.[12]

5 Costs of deviating from the optimal policy

While the analysis in the last two sections identified conditions under which discontinuous pivots represent an optimal response to sufficient levels of overheating, such pivots are clearly difficult to time as a practical matter, and concerns that an abrupt tightening could trigger economic disruptions often lead policy-makers to take more gradual approaches. While gradual approaches

[12] In this respect, we note that recent pivots in advanced economies occurred at a time when (i) some central banks had recently announced framework changes placing greater emphasis on employment outcomes; and (ii) a range of studies suggested that Phillips curves in advanced economies had flattened significantly. To the extent that these factors respectively contributed to a higher value for μ and lower value for λ, they would have increased the ratio $\tilde{\mu} = \mu/\lambda^2$, setting the scene for a later but more pronounced pivot when viewed through the lens of this framework.

are generally sub-optimal when viewed through the lens of the model, we use this section to get more specific about their costs, as well of those of poorly timed pivots.

Of course, there are many ways in which the monetary authority could deviate from the optimal policy in our model. We will focus on two possible deviations in simulations where we initialize the baseline parameterization of the model at a modest level of recent overheating and then assume a supply shock large enough to trigger pivoting under the optimal policy. Figure 9 depicts the implied time paths for key variables under the optimal policy, as well as the two deviant policies.

The first of these deviant policies aims to capture a scenario where the policy pivot is mis-timed and occurs too late. To construct this "too late" scenario, we make use of the fact that the central bank's loss function generally admits multiple, locally optimal values for the policy stance, with pivots involving jumps from the lowest of these optima to a higher one. In particular, we assume that policy-makers (i) initially stay on their lowest local optimum after the supply shocks arrives, even though it is no longer globally optimal to do so; and (ii) only recognize and correct this mistake after three periods. As shown in the middle and bottom panels of Figure 9, this leads to significantly worse inflation and employment outcomes relative to the globally optimal policy.

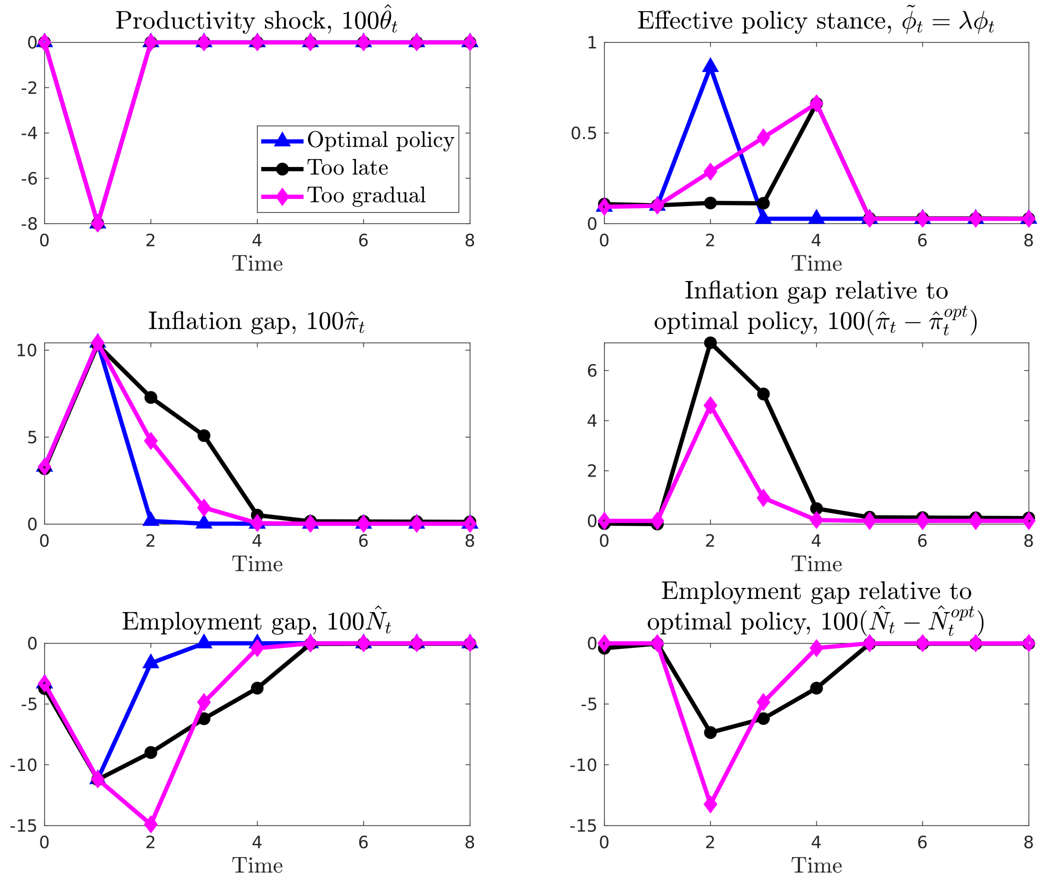
The second deviation that we consider aims to capture a scenario where policy-makers no longer pivot and instead tighten gradually in response to the supply shock. For this scenario, we assume that the central bank smoothly increases its stance in the three periods following the supply shock, eventually reaching the same peak value as in the "too late" scenario; once they reach this peak, policy-makers are assumed to revert back to the globally optimal policy. As shown in Figure 9, a key feature of this "too gradual" scenario is that it leads to worse employment outcomes relative to *both* the optimal policy and the "too late" scenario. This illustrates the risk of hard landings associated with intermediate values for the policy stance and more generally highlights the rich, non-monotonic mapping that the model admits between the policy stance and aggregate outcomes.

6 What about demand shocks?

Up to now, we have ignored demand shocks, despite the important role that they played in the recent surge in advanced economies and other real-world inflation episodes. In this section, we briefly discuss the consequences of introducing demand shocks and show how our results can be interpreted in this broader context.

We will focus on the standard case where demand shocks arise from shocks to the household

Figure 9: Simulating an optimal pivot and two possible deviations



discount factor, leading the IS curve on line 2.13 to include a stochastic intercept d_t following an arbitrary process. Forward-solving this equation then yields

$$\hat{N}_t = -\mathbb{E}_t^{RE}\left[\sum_{\tau=0}^{\infty}\left(\hat{\imath}_{t+\tau} - \hat{\pi}_{t+\tau+1}\right)\right] + \mathbb{E}_t^{RE}\left(\sum_{\tau=0}^{\infty} d_{t+\tau}\right), \tag{6.32}$$

so the appropriate measure of the stimulus being provided by the central bank is now the cumulative sum of all expected future differences between the real rate and a time-varying natural rate determined by the demand shock. In the special case where the central bank is able to adjust the rate path as needed to ensure an average difference of zero going forward, this would imply full stabilization of the employment gap.

With these points in mind, we can still interpret the feedback rule on line 2.14 as the product of a situation where policy-makers set the level of stimulus as a function of the prevailing rate of inflation, though the appropriate measure of stimulus should now be based on equation (6.32):

$$\phi_t \hat{\pi}_t = \mathbb{E}_t^{RE}\left[\sum_{\tau=0}^{\infty}\left(\hat{\imath}_{t+\tau} - \hat{\pi}_{t+\tau+1}\right)\right] + \mathbb{E}_t^{RE}\left(\sum_{\tau=0}^{\infty} d_{t+\tau}\right).$$

Under this interpretation, the core policy problem laid out in Section 3 is unaltered relative to its form in an economy without demand shocks, and the only impact of demand shocks is thus to shift the nominal rate settings that the solution to this problem would imply.

Put differently, when the model economy is subject to both demand and supply shocks, the policy-setting process can be conceptualized as a two-stage process where policy-makers first gauge the rate path needed to fully insulate employment from demand shocks, and then judge if adjustments are needed to address supply shocks or poor anchoring of wage-setters' expectations. All of our results thus carry over to this context but should be interpreted as informing the second of the two stages that we just described.

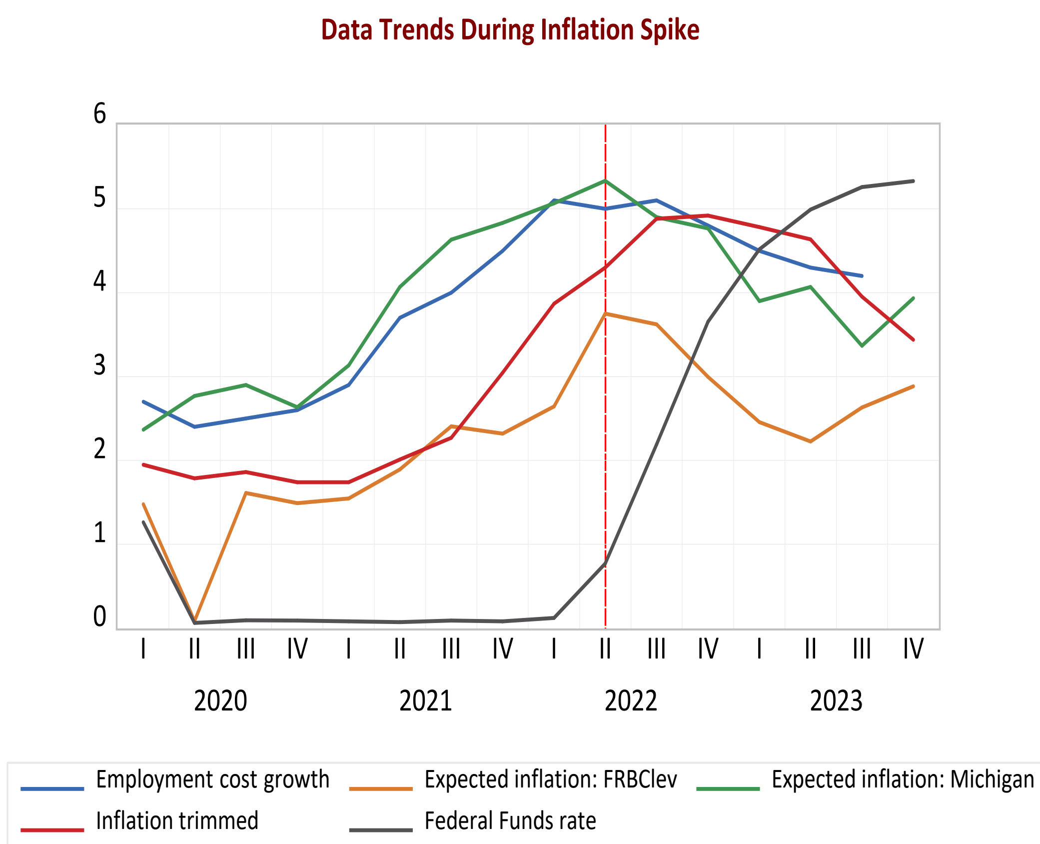
Of course, this assumes that policy-makers are able to distinguish between demand and supply shocks, and this is a non-trivial assumption. If, for example, a central bank mistakenly interpreted some demand shocks as supply shocks, then this could lead to an overly stimulative stance to the extent that the two-stage process above will generally involve fully offsetting demand shocks but looking through supply shocks to at least some extent. Similar issues would arise if policy-makers simply mis-judged the strength of demand. "Policy errors" of this sort could be incorporated into our analysis but would involve a signal-extraction problem that we leave for future work.

7 Model Mechanism and the Recent Inflation Surge

Our model has the implication that it is optimal for the policymaker to initially look-through inflation shocks but pivot sharply to aggressive tightening if the shocks cumulate above a threshold level. While this prediction squares with the experience during the inflation surge 2021, is there any evidence to support the mechanisms formalized in the model? We turn to this issue here.

We start by noting that our model's primary mechanism works through expected inflation. With unchanging policies inflation expectations tend to co-move with lagged inflation. A pivot to a tight monetary policy however induces an immediate moderation of expected inflation independent of the specific inflation realization in that period. Similarly, wage growth in the model drifts up and down with expected inflation expectation.

Figure 10:



Is there any evidence of such relationships in the data? Figure 10 shows the movements from

2020Q1to 2023Q4 of two measures of inflation expectations (Michigan survey and the FRB Cleveland's expected 1-year inflation measure), wage growth proxied by changes in the employment cost index, and the trimmed inflation rate. The figure shows that both measures of expected inflation as well as wage growth kept trending up until 2022Q2 which marked the first tightening of monetary policy by the FOMC. After the tightening, expected inflation and wage growth both turned around even though inflation continued trending up for a couple of quarters longer. These patterns are consistent with the model mechanism.

More formally, there are three key equilibrium relationships in the model: (a) inflation depends on wage growth and productivity growth; (b) wage growth depends on expected inflation; and (c) expected inflation depends on past inflation. Specifically, the model yields the following key conditions for inflation, wage growth and expected inflation:

$$\hat{\pi}_t = \hat{W}_t - \hat{\theta}_t \tag{7.33}$$

$$\hat{W}_t = (1 - \lambda\phi_t)^k \left(\hat{\pi}_{t-1} + \lambda\hat{N}_{t-1}\right) \tag{7.34}$$

$$\mathbb{E}^{WS}_{t-1}\hat{\pi}_t = (1 - \lambda\phi_t)^{k-1} \left(\hat{\pi}_{t-1} + \lambda\hat{N}_{t-1}\right) \tag{7.35}$$

For periods with a constant policy stance $\bar{\phi}$, equation 7.35 can be rewritten as

$$\mathbb{E}^{WS}_{t-1}\hat{\pi}_t = \left(1 - \lambda\bar{\phi}\right)^{k-1} \left(\hat{\pi}_{t-1} + \lambda\hat{N}_{t-1}\right) + \epsilon^e_t \tag{7.36}$$

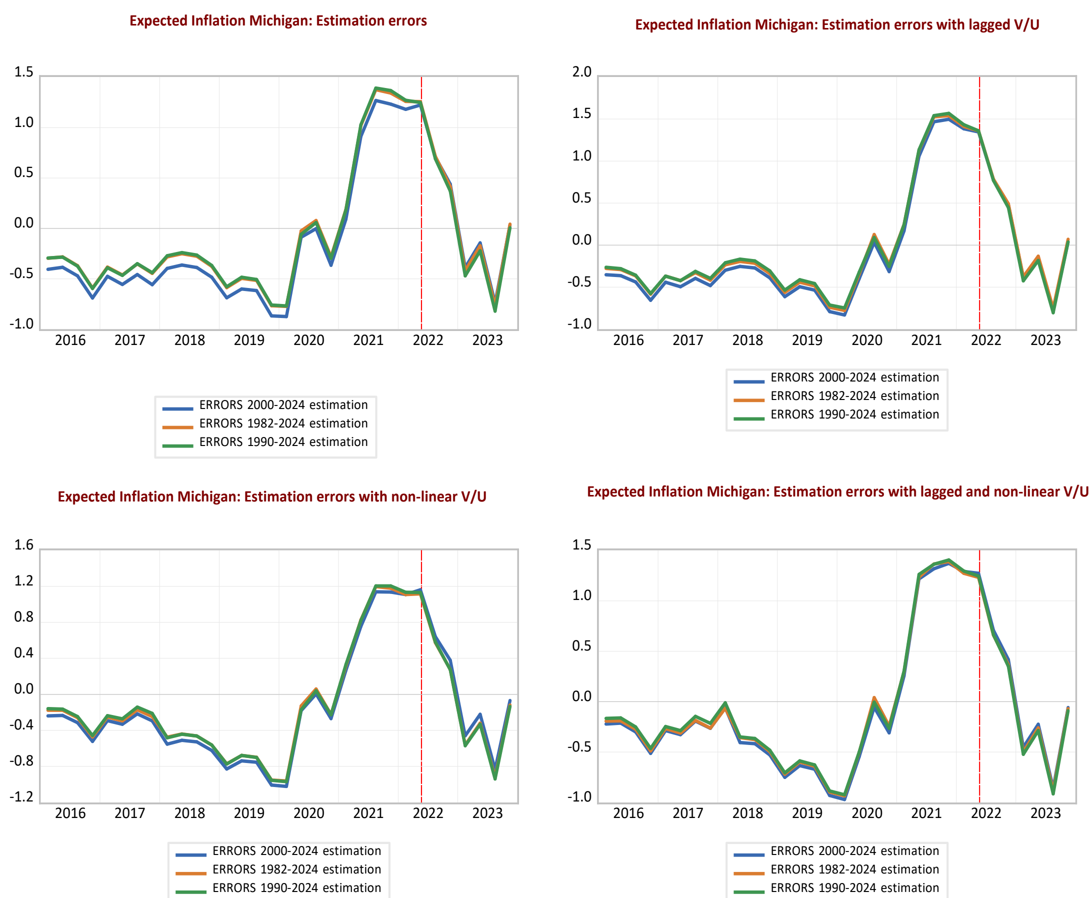
where $\epsilon^e_t = \left[(1 - \lambda\phi_t)^{k-1} - \left(1 - \lambda\bar{\phi}\right)^{k-1}\right] \left(\hat{\pi}_{t-1} + \lambda\hat{N}_{t-1}\right)$. During periods with an overheating economy so that $\hat{\pi}_{t-1} + \lambda\hat{N}_{t-1} > 0$, a policy of looking through would imply that the constant policy is such that $\bar{\phi} > \phi_t$. Hence, one would expect an overheating economy along with a monetary policy that looks through it to be characterized by a rising residual ϵ^e in the expected inflation equation. Moreover, when the policy pivot occurs so that $\phi_t > \bar{\phi}$, the model predicts that the errors ϵ^e should revert towards zero.

To test this prediction, we use quarterly data to run a regression of expected inflation on lagged inflation and the vacancy/unemployment ratio as a proxy for the labor market variable N. This regression is based on equation 7.36. We use the one-year ahead inflation expectation reported in the Michigan survey of household expectations as our measure of inflation expectation while we use the trimmed inflation rate which is tighter measure of core inflation.

To test for robustness of the results in multiple ways. First, we estimate the regression for three different sample periods: 1982Q1:2023Q4, 1990Q1:2023Q4 and 2000Q1:2023Q4. Second, we include a non-linear labor market effects of the vacancy-unemployment ratio (V/U) by including a dummy for periods when $V/U > 1$. This is motivated by recent work in Benigno-Eggertson who find evidence of such non-linear effects in the recent inflation episode. Third, we include their contemporaneous values of the V/U ratio instead of their lagged values.

Figure 11 shows the computed errors ϵ_t^e from all these different specifications. The results are remarkably similar. Across specifications, we find that the regression errors start rising around 2020Q2 and keep rising until 2022Q2 which marked the beginning of the FOMC monetary tightening. Remarkably, the errors under all specifications and for all the different sample estimation periods began declining back towards zero from 2022Q2. We view this as strikingly supportive evidence of the key mechanism embedded in our model.

Figure 11: Expected inflation regression errors



8 Conclusion

The appropriate monetary response to inflation driven by supply shocks is a question that has captured the attention of many policy-makers and analysts over the last few years. In this paper, we have aimed to contribute to this discussion by examining how best to stabilize inflation and employment under a range of assumptions on how agents form expectations, including cases where agents are sufficiently sophisticated to understand that policy decisions affect their environment but are not able to fully internalize all of the implications. The environment we examined was one where sticky wages can favour looking through supply shocks, while the sensitivity of inflation expectations to the policy stance and realized inflation makes that strategy risky.

We showed that under level-k thinking, where agents only partially think through the way that monetary policy affects their environment, optimal policy generally takes the form of a threshold rule. If inflation has been pushed only slightly above target, then policy should mainly look through supply shocks. However, if a series of supply shocks cumulate up to a specific threshold, we showed that policy needs to shift discontinuously to a more hawkish stance in order to control inflation expectations.

While the hawkish policy stance involves reacting strongly to any subsequent inflation, the pivot itself – if executed optimally – can lead to a soft landing since its aim is to directly decrease inflation expectations and thereby inflation itself. Declining inflation expectations induce an expansionary effect on employment due to its moderating effect on wage growth. This indirect effect on employment can largely counteract the negative effect of policy tightening, thereby generating a soft-landing.

At the beginning of 2022, many central banks changed their policy stance in a manner that could be referred to as a pivot. Viewed through the lens of this model, this response represented a step in the right direction, but whether it was sufficiently aggressive is not clear. Especially in cases where policy-makers opted for less aggressive and/or more gradual pivots, our analysis suggests that a more decisive shift in the policy stance could potentially have led to a shorter tightening cycle and better odds of a soft landing.

References

Arad, A. and A. Rubinstein (2012). The 11-20 money request game: a level-k reasoning study. *American Economic Review 102*(7), 3561–73.

Bianchi Vimercati, R., M. S. Eichenbaum, and J. Guerriro (2021, August). Fiscal stimulus with imperfect expectations: Spending vs. tax policy. Working Paper 29134, National Bureau of Economic Research.

Blanchard, O. and N. Kiyotaki (1987). Monopolistic competition and the effects of aggregate demand. *American Economic Review 77*(4), 647–66.

Calvo, G. (1983). Staggered prices in the utility-maximizing framework. *Journal of Monetary Economics 12*(3), 383–398.

Camerer, C. F., T. Ho, and J. Chong (2004, August). A cognitive hierarchy model of games. *Quarterly Journal of Economics 119*(3), 861–898.

Carvalho, C., S. Eusepi, E. Moench, and B. Preston (2023, January). Anchored inflation expectations. *American Economic Journal: Macroeconomics 15*(1), 1–47.

Champagne, J., C. Hajzler, D. Matveev, H. Melinchuk, A. Poulin-Moore, K. Ozhan, Y. Park, and T. Taskin (2023, April). Potential output and the neutral rate in Canada: 2023 assessment. Working Paper 2023-6, Bank of Canada.

Costa-Gomes, M., V. P. Crawford, and N. Iriberri (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature 51*(1), 5–62.

di Giovanni, J., S. Kalemli-Özkan, A. Silva, and M. A. Yildirim (2022, July). Global supply chain pressures, international trade, and inflation. Working Paper 30240, National Bureau of Economic Research.

Djeutem, E., M. He, A. Reza, and Y. Zhang (2022, March). Household heterogeneity and the performance of monetary policy frameworks. Working Paper 2022-12, Bank of Canada.

Farhi, E. and I. Werning (2019, November). Monetary policy, bounded rationality, and incomplete markets. *American Economic Review 109*(11), 3887–3928.

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

García-Schmidt, M. and M. Woodford (2019, January). Are low interest rates deflationary? A paradox of perfect-foresight analysis. *American Economic Review 109*(1), 86–120.

Gatí, L. (2023). Monetary policy & anchored expectations – an endogenous gain learning model. *Journal of Monetary Economics* (forthcoming).

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988, June). Investment, capacity utilization, and the real business cycle. *American Economic Review 78*(3), 402–417.

Hazell, J., J. Herreño, E. Nakamura, and J. Steinsson (2022, August). The slope of the Phillips curve: evidence from U.S. states. *Quarterly Journal of Economics 137*(3), 1299–1344.

Iovino, L. and D. Sergeyev (2023). Central bank balance sheet policies without rational expectations. *Review of Economic Studies* (forthcoming).

Kneeland, T. (2015). Identifying higher-order rationality. *Econometrica 83*(5), 2065–79.

Lorenzoni, G. and I. Werning (2023, January). Wage price sprials. Working paper, MIT.

Lucas, R. (1972). Expectations and the neutrality of money. *Journal of Economic Theory 4*(2), 103–124.

Preston, B. (2005, September). Learning about monetary policy rules when long-horizon expectations matter. *International Journal of Central Banking 1*(2), 81–126.

Reis, R. (2021, October). Losing the inflation anchor. Discussion Paper 16664, Centre for Economic Policy Research.

Reis, R. (2022). The burst of high inflation: How and why did we get here? Working paper, London School of Economics.

Rotemberg, J. (1982). Staggered prices in the utility-maximizing framework. *Journal of Political Economy 90*(6), 1187–1211.

Sargent, T. J. (1982). The ends of four big inflations. In R. E. Hall (Ed.), *Inflation: Causes and Effects*, pp. 41–98. Chicago: Chicago University Press.

Wagner, J., T. Schlanger, and Y. Zhang (2022, Febraury). A horse race of alternative monetary policy regimes under bounded rationality. Working Paper 2022-4, Bank of Canada.

Woodford, M. (2004). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.

Appendix

Proof of Lemma 3.2. All claims follow immediately from the fact that $d(\phi_t) = \mu(1+k)\phi_t^2 - \mu\phi_t + k$, making $d(\cdot)$ a convex quadratic function with a discriminant of the form $\mu^2 - 4\mu k(1+k)$. ■

Proof of Lemma 3.3. $n(\cdot)$ is a cubic function satisfying $n(0) > 0$, $n'(0) \geq 0$, $n(1) > 0$, and $n'(1) > 0$. As a result, it is either strictly positive over the unit interval, assuming that its discriminant is strictly negative, or otherwise admits two (possibly duplicate) roots ϕ_1^n and ϕ_2^n in this interval. Noting that the discriminant of a generic cubic $ax^3 + bx^2 + cx + d$ is given by $18abcd - 4b^3d + b^2c^2 - 4ac^3 - 27a^2d^2$, the relevant discriminant can be shown to take the form

$$\mu\left[108\mu^2 + B(k)\mu + C(k)\right], \tag{.37}$$

where:

$$B(k) \equiv 36(k-1)^2 - 216(k+1)(k-1) - 108(k+1)^2 \tag{.38}$$

$$C(k) \equiv -64(k+1)(k-1)^3 \leq 0. \tag{.39}$$

Since the function $B(\cdot)$ is decreasing and satisfies $B(1) < 0$, we can conclude that $B(k) < 0$. As a result, the expression in square brackets on the right-hand side of line .37 must, when viewed as a function of μ, take the form of a convex quadratic with two real roots: one strictly positive, and the other weakly negative. Letting $\overline{\mu}$ denote the positive root, we can conclude that the discriminant on line .37 is strictly negative when $\mu \in (0, \overline{\mu})$, but strictly positive with $\phi_2^n > \phi_1^n$ when $\mu > \overline{\mu}$. Since the term in square brackets on line .37 can be shown to be strictly positive when $\mu = 4k(1+k)$, we can also conclude that $\overline{\mu} < 4k(1+k)$.

All that remains at this point is to verify the ordering of the set $\{\phi_1^n, \phi_2^n, \phi_1^d, \phi_2^d\}$, assuming that $\mu \geq 4k(1+k)$. Suppose first that this inequality holds strictly. In this case, it is useful to note from the definitions of the functions $f(\cdot)$ and $n(\cdot)$ that

$$n(\phi_t) = \frac{[1 + 2\phi_t(k-1)]d(\phi_t) - \phi_t(1-\phi_t)d'(\phi_t)}{k},$$

so $n(\phi_t)$ and $d'(\phi_t)$ must have opposite signs when $\phi_t \in \{\phi_1^d, \phi_2^d\}$. However, with $\mu > 4k(1+k)$, Lemma 3.2 implies that $d'(\phi_1^d) < 0 < d'(\phi_2^d)$, so it must be the case that $n(\phi_1^d) > 0 > n(\phi_2^d)$, which is only possible if $\phi_1^d < \phi_1^n < \phi_2^d < \phi_2^n$ based on the shape of the function $n(\cdot)$. By continuity, it must therefore be the case that $\phi_1^d = \phi_1^n = \phi_2^d < \phi_2^n$ when $\mu = 4k(1+k)$ exactly, since in this case we have $\phi_1^d = \phi_2^d$, but we still have $\phi_1^n < \phi_2^n$ by virtue of the fact that $4k(1+k) > \overline{\mu}$. ■

Proof of Proposition 3.1. The claims made regarding the case $\mu \le \overline{\mu}$ were established in the main text – likewise the continuity and strict monotonicity of the functions $\phi_1(\cdot)$ and $\phi_2(\cdot)$ when $\mu > \overline{\mu}$. As a result, we will focus on establishing the uniqueness of the pivot point x_*^2 when $\mu > \overline{\mu}$.

As a first step in this direction, let

$$\mathcal{L}^i(x_{t-1}^2) \equiv \mathcal{L}[\phi_i(x_{t-1}^2), x_{t-1}^2] = \Big[1 + \mu[\phi_i(x_{t-1}^2)]^2\Big]\Big[[1 - \phi_i(x_{t-1}^2)]^{2k} x_{t-1}^2 + \sigma_\theta^2\Big]$$

denote the central bank's expected losses along the lower branch ($i = 1$) or upper branch ($i = 2$). Note that $\mathcal{L}^i(\cdot)$ is continuous in both cases, as is the function $\mathcal{L}(\phi_t, \cdot)$ $\forall \phi_t \in [0,1]$. In light of these points, coupled with the fact that

$$\mathcal{L}^1(x_{t-1}^2) < \mathcal{L}[\phi_2[\sigma_\theta^2 f(\phi_2^n)], x_{t-1}^2] \;\; \forall x_{t-1}^2 \in [0, \sigma_\theta^2 f(\phi_2^n)),$$

we can conclude that $\mathcal{L}^1[\sigma_\theta^2 f(\phi_2^n)] \le \mathcal{L}^2[\sigma_\theta^2 f(\phi_2^n)]$. If $\mu < 4k(1+k)$, then similar reasoning will confirm that $\mathcal{L}^1[\sigma_\theta^2 f(\phi_1^n)] \ge \mathcal{L}^2[\sigma_\theta^2 f(\phi_1^n)]$, so there must exist some $x_*^2 \in [\sigma_\theta^2 f(\phi_2^n), \sigma_\theta^2 f(\phi_1^n)]$ at which

$$\mathcal{L}^1(x_*^2) = \mathcal{L}^2(x_*^2). \tag{.40}$$

If instead $\mu \ge 4k(1+k)$, then some $x_*^2 \in [\sigma_\theta^2 f(\phi_2^n), \infty)$ satisfying equation (.40) should still exist, since we know from the main text that $\mathcal{L}^1(x_{t-1}^2) > \mathcal{L}^2(x_{t-1}^2)$ when x_{t-1}^2 is very large.

In both cases, it is now useful to note that $\phi_2(x_*^2) \ge \phi_2^n > \phi_1^n \ge \phi_1(x_*^2)$, where the middle inequality follows from the fact that $\mu > \overline{\mu}$. As a result, equation (.40) can only hold if

$$\Big[1 + \mu[\phi_1(x_*^2)]^2\Big][1 - \phi_1(x_*^2)]^{2k} > \Big[1 + \mu[\phi_2(x_*^2)]^2\Big][1 - \phi_2(x_*^2)]^{2k},$$

which is equivalent to $\left(\mathcal{L}^1\right)'(x_*^2) > \left(\mathcal{L}^2\right)'(x_*^2)$ by the envelope theorem. Conclude that any point of intersection between the functions $\mathcal{L}^1(\cdot)$ and $\mathcal{L}^2(\cdot)$ must have the former crossing strictly from below, thus precluding multiple intersections. ■

Proof of Proposition 3.2. In our proof of Lemma 3.3, we showed that $\overline{\mu}$ is the upper root of a convex quadratic of the form

$$108\mu^2 + B(k)\mu + C(k), \tag{.41}$$

where $B(k)$ and $C(k)$ are respectively defined on lines .38 and .39 and satisfy $B'(k) < 0$ and $C'(k) < 0$ $\forall k \in [1, \infty)$. Evaluating line .41 at $\mu = \overline{\mu}$ and totally differentiating with respect to k thus yields

$$(d\overline{\mu}/dk)\underbrace{[216\overline{\mu} + B(k)]}_{>0} + \underbrace{B'(k)\overline{\mu} + C'(k)}_{<0} = 0,$$

so $d\overline{\mu}/dk > 0$. In the RE limit, lines .38 and .39 further imply that

$$\lim_{k\to\infty}\{\overline{\mu}\} = \lim_{k\to\infty}\left\{\frac{-B(k) + \sqrt{[B(k)]^2 - 432C(k)}}{216}\right\} = \lim_{k\to\infty}\left\{(1+k)^2\right\} = \infty. \quad \blacksquare$$

Proof of Proposition 3.3. That the variance of the employment gap is strictly increasing around the pivot point was already established in the main text, so we will focus on showing that the expected level of the employment gap is unchanged by pivots when $\mu > \overline{\mu}$ and $k = 1$.

Suppose first that $\mu > \overline{\mu}$. In this case, our previous findings imply that the triplet $(\phi_1, \phi_2, z) = [\phi_1(x_*^2), \phi_2(x_*^2), x_*^2]$ is the unique solution to the following system of equations:

$$\mathcal{L}(\phi_1, z) = \mathcal{L}(\phi_2, z)$$
$$\mathcal{L}_1(\phi_1, z) = 0$$
$$\mathcal{L}_1(\phi_2, z) = 0$$

– i.e., when $x_{t-1}^2 = z$, the central bank's first-order condition must hold at both $\phi_t = \phi_1$ and $\phi_t = \phi_2$, and expected losses at these two points should also be equal.

Let us now impose $k = 1$. In this case, the system above can be written as follows:

$$(1 + \mu\phi_1^2)(1 - \phi_1)^2 z + (1 + \mu\phi_1^2) = (1 + \mu\phi_2^2)(1 - \phi_2)^2 z + (1 + \mu\phi_2^2)$$
$$[\mu\phi_1(1 - \phi_1)^2 - (1 - \phi_1)(1 + \mu\phi_1^2)]z + \mu\phi_1 = 0$$
$$[\mu\phi_2(1 - \phi_2)^2 - (1 - \phi_2)(1 + \mu\phi_2^2)]z + \mu\phi_2 = 0.$$

Using the second of these equations to eliminate z then yields the following system:

$$\frac{\mu\phi_1(1+\mu\phi_1^2)(1-\phi_1)^2}{(1-\phi_1)(1+\mu\phi_1^2)-\mu\phi_1(1-\phi_1)^2}+(1+\mu\phi_1^2)=\frac{\mu\phi_1(1+\mu\phi_2^2)(1-\phi_2)^2}{(1-\phi_1)(1+\mu\phi_1^2)-\mu\phi_1(1-\phi_1)^2}+(1+\mu\phi_2^2)$$
$$\frac{[\mu\phi_2(1-\phi_2)^2-(1-\phi_2)(1+\mu\phi_2^2)]\mu\phi_1}{(1-\phi_1)(1+\mu\phi_1^2)-\mu\phi_1(1-\phi_1)^2}+\mu\phi_2=0.$$

At this point, it can be verified that the pair

$$(\phi_1,\phi_2)=\left[\frac{\mu-\sqrt{\mu^2-4\mu}}{2\mu},\frac{\mu+\sqrt{\mu^2-4\mu}}{2\mu}\right]$$

comprises a solution to the two-dimensional system above.[13] Note that this pair must be real, since the characterization of $\overline{\mu}$ in our proof of Lemma 3.3 implies that $\overline{\mu}=4$ when $k=1$. Moreover, since the pair satisfies $\phi_2=1-\phi_1$, it must be the case that the expected employment level $\mathbb{E}_{t-1}^{RE}(\hat{N}_t)=-\phi_t(1-\phi_t)x_{t-1}$ is constant around the pivot point. ■

[13]One way of seeing that these values for ϕ_1 and ϕ_2 satisfy the two equations in question is to begin by conjecturing that $\phi_2=1-\phi_1$ and then use this to replace ϕ_2. In this case, each of these two equations individually becomes an equation in ϕ_1 only, and they can both be reduced to the same cubic equation, namely: $2\mu\phi_1^3-3\mu\phi_1^2+(2+\mu)\phi_1-1=0$. This cubic equation can then be factored as $(\phi_1-.5)(2\mu\phi_1^2-2\mu\phi_1+2)=0$, with the claimed values for ϕ_1 and ϕ_2 corresponding to the roots of the term $(2\mu\phi^2-2\mu\phi+2)$, which indeed satisfy our initial conjecture.

Online Appendix

A Demand curve under non-rational expectations

In this section, we revisit the household Euler equation on line 2.8 of the main text while adapting methods first developed in Preston (2005) to allow for the possibility that households' expectations may not be rational. In lieu of full rationality, this approach requires the weaker condition that household expectations be consistent with the law of iterated expectations – i.e., $\mathbb{E}_t^{HH(h)}(\cdot) = \mathbb{E}_t^{HH(h)}[\mathbb{E}_T^{HH(h)}(\cdot)]$ for each household $h \in [0,1]$ and each horizon $T \in \{t+1, t+2, \ldots\}$. It is also convenient but non-essential to assume that households are rational *when forming expectations on fully exogenous variables*, echoing a similar assumption that was imposed on wage-setters in the main text.

Under these assumptions, linearizing and forward-solving the household Euler equation allows us to express a given household's consumption plans at a given horizon as follows:

$$\mathbb{E}_t^{HH(h)}(\hat{C}_{hT}) = \hat{C}_{ht} + \mathbb{E}_t^{HH(h)}\left[\sum_{\tau=t}^{T-1}(\hat{\imath}_\tau - \hat{\pi}_{\tau+1})\right],$$

where $\hat{C}_{ht} \equiv \ln C_{ht} - \ln \overline{Y}_t$ denotes the deviation of consumption from the symmetric, flexible-wage benchmark described in the main text. At the same time, the household's expected lifetime budget constraint can be expressed as follows after linearizing and forward-solving the flow constraint on line 2.7 of the main text:

$$\frac{B_{h,t-1}}{P_t Y_t} = \mathbb{E}_t^{HH(h)}\left[\sum_{T=t}^{\infty}\beta^{T-t}(\hat{C}_{hT} - \hat{N}_T)\right].$$

Together, these expressions lead to a final demand curve of the following form after summing across households and accounting for goods and bond market clearing:

$$\hat{N}_t = (1-\beta)\overline{\mathbb{E}}_t^{HH}\left(\sum_{T=t}^{\infty}\beta^{T-t}\hat{N}_t\right) - \beta\overline{\mathbb{E}}_t^{HH}\left[\sum_{T=t}^{\infty}\beta^{T-t}(\hat{\imath}_T - \hat{\pi}_{T+1})\right], \tag{A.42}$$

where $\overline{\mathbb{E}}_t^{HH}(\cdot) \equiv \int_0^1 \mathbb{E}_t^{HH(h)}(\cdot)dh$ denotes an average expectation across households. This allows us to express the overall level of demand in the economy partly as a function of households' expectations regarding the real policy rate path, $\overline{\mathbb{E}}_t^{HH}\left[\sum_{T=t}^{\infty}\beta^{T-t}(\hat{\imath}_T - \hat{\pi}_{T+1})\right]$, and partly as a function of fluctuations in households' expected permanent income, $\overline{\mathbb{E}}_t^{HH}\left(\sum_{T=t}^{\infty}\beta^{T-t}\hat{N}_t\right)$.

Though more complicated than the IS curve considered in the main text, this demand curve remains consistent with the recursive solution structure emphasized in Subsection 2.5 – i.e., one can continue solving for the key triplet $(\hat{\pi}_t, \hat{N}_t, \phi_t)$ using other blocks of the model, then back out the policy rate path needed to ratify these outcomes using equation (A.42). It is also possible to continue interpreting the policy feedback rule on line 2.14 as the product of a rule where the central bank adjusts the policy rate path as needed to regulate the overall level of demand as a function of the prevailing rate of inflation (though in this case the central bank must account for potential permanent income effects in its assessment of overall demand conditions).

B Phillips curve under GHH preferences

In contrast to the analysis in Section 3 of the main text, we now assume that households have preferences of a form in line with Greenwood, Hercowitz, and Huffman (1988). Specifically, we assume a utility function of the form

$$\mathbb{E}_t^{HH}\left[\sum_{\tau=0}^{\infty}\beta^{\tau}\ln\left(C_{t+\tau}-\eta\theta_{t+\tau}\int_0^1 N_{s,t+\tau}^{1+\chi}ds\right)\right],$$

where $\chi > 0$ is the inverse Frisch elasticity of labour supply. In this section, we first characterize the Phillips curve to which these preferences give rise, then turn our attention to the implications for the problem facing policy-makers.

Phillips curve. Under the preferences above, the problem facing the union setting wages for service $s \in [0,1]$ involves choosing (W_{st}, N_{st}) to maximize an objective of the form

$$\mathbb{E}_{t-1}^{WS(s)}\left(\frac{\frac{W_{st}N_{st}}{P_t}-\eta\theta_t N_{st}^{1+\chi}}{C_t-\eta\theta_t\int_0^1 N_{\tilde{s}t}^{1+\chi}d\tilde{s}}\right),$$

subject to the labour demand curve on line 2.5 of the main text. The first-order condition associated with this problem reads as

$$W_{st}^{1+\rho\chi}=\frac{\eta\rho(1+\chi)}{\rho-1}\frac{\mathbb{E}_{t-1}^{WS(s)}\left[\frac{\theta_t\left[(Y_t/\theta_t)W_t^{\rho}\right]^{1+\chi}}{C_t-\eta\theta_t\int_0^1 N_{\tilde{s}t}^{1+\chi}d\tilde{s}}\right]}{\mathbb{E}_{t-1}^{WS(s)}\left[\frac{\frac{(Y_t/\theta_t)W_t^{\rho}}{P_t}}{C_t-\eta\theta_t\int_0^1 N_{\tilde{s}t}^{1+\chi}d\tilde{s}}\right]}.$$

Recalling from the main text that all wage-setters share common, though potentially incorrect, expectations on all aggregate outcomes, the first-order condition above can be re-expressed as

$$W_t^{1+\rho\chi} = \frac{\eta\rho(1+\chi)}{\rho-1}\frac{\mathbb{E}_{t-1}^{WS}\left[\frac{\theta_t\left(N_t W_t^{\rho}\right)^{1+\chi}}{C_t-\eta\theta_t\int_0^1 N_{\tilde{s}t}^{1+\chi}d\tilde{s}}\right]}{\mathbb{E}_{t-1}^{WS}\left[\frac{\frac{N_t W_t^{\rho}}{P_t}}{C_t-\eta\theta_t\int_0^1 N_{\tilde{s}t}^{1+\chi}d\tilde{s}}\right]}. \tag{B.43}$$

Taking logarithms on both sides of this equation and ignoring second- and higher-order terms associated with Jensen's inequality further yields

$$(1+\rho\chi)\ln W_t = \ln\left[\frac{\eta\rho(1+\chi)}{\rho-1}\right] + \mathbb{E}_{t-1}^{WS}(\ln\theta_t + \chi\ln N_t + \rho\chi\ln W_t + \ln P_t),$$

at which point the mark-up rule on line 2.4 of the main text can be used to eliminate $\ln W_t$. This yields the following after some algebra:

$$(1+\rho\chi)\hat{\pi}_t = \chi\mathbb{E}_{t-1}^{WS}(\ln N_t) - \ln\left[\frac{\rho-1}{\eta\rho(1+\chi)}\frac{\gamma-1}{\gamma}\right] + (1+\rho\chi)[\mathbb{E}_{t-1}^{WS}(\hat{\pi}_t) - \hat{\theta}_t] \tag{B.44}$$

With this last expression in hand, it is now useful to note that the wage-setting condition on line B.43 would read as follows if wages were flexible:

$$W_t = \frac{\eta\rho(1+\chi)}{\rho-1}\theta_t N_t^{\chi} P_t.$$

At the same time, the mark-up rule on line 2.4 of the main text can be written as

$$P_t\theta_t N_t = \frac{\gamma}{\gamma-1}W_t N_t.$$

Together, these two expressions pin down the natural rate of employment:

$$N_t = \left[\frac{\rho-1}{\eta\rho(1+\chi)}\frac{\gamma-1}{\gamma}\right]^{1/\chi} \equiv \bar{N}.$$

Combining this expression with equation (B.44) then finally gives us a Phillips curve of the form

$$\hat{\pi}_t = \left(\frac{\chi}{1+\rho\chi}\right)\mathbb{E}_{t-1}^{WS}(\hat{N}_t) + \mathbb{E}_{t-1}^{WS}(\hat{\pi}_t) - \hat{\theta}_t,$$

as claimed in the main text.

Implications for the policy problem. Letting $\lambda \equiv \chi/(1+\rho\chi)$ denote the slope of the Phillips curve above, private-sector expectations are now given by the following:

$$\mathbb{E}_{t-1}^{k}(\hat{\pi}_t) = (1-\lambda\phi_t)^{k-1}(\hat{\pi}_{t-1}+\lambda\hat{N}_t)$$
$$\mathbb{E}_{t-1}^{k}(\hat{N}_t) = -\phi_t(1-\lambda\phi_t)^{k-1}(\hat{\pi}_{t-1}+\lambda\hat{N}_t)].$$

Realized inflation and employment outcomes are in turn given by:

$$\hat{\pi}_t = (1-\lambda\phi_t)^{k}(\hat{\pi}_{t-1}+\lambda\hat{N}_t)$$
$$\hat{N}_t = -\phi_t(1-\lambda\phi_t)^{k}(\hat{\pi}_{t-1}+\lambda\hat{N}_t).$$

So, if we define $x_t \equiv \hat{\pi}_t + \lambda\hat{N}_t$, then the central bank's expected lifetime losses can now be written as

$$\mathbb{E}_{t-1}^{RE}\left[\sum_{\tau=0}^{\infty}\beta_{CB}^{\tau}(\hat{\pi}_{t+\tau}^2+\mu\hat{N}_{t+\tau}^2)\right] = \mathbb{E}_{t-1}^{RE}\left[\sum_{\tau=0}^{\infty}\beta_{CB}^{\tau}(1+\mu\phi_{t+\tau}^2)\left[(1-\lambda\phi_{t+\tau})^2x_{t+\tau-1}^2+\sigma_\theta^2\right]\right],$$

while the law of motion for the overheating measure can be written as

$$x_t = (1-\lambda\phi_t)[(1-\lambda\phi_t)^k x_{t-1} - \hat{\theta}_t].$$

Further defining $\tilde{\mu} \equiv \mu/\lambda^2$ and $\tilde{\phi}_t \equiv \lambda\phi_t$, this program finally collapses to the form claimed in the main text – i.e., policy-makers choose $\tilde{\phi}_t$ to minimize

$$\mathbb{E}_{t-1}^{RE}\left[\sum_{\tau=0}^{\infty}\beta_{CB}^{\tau}(1+\tilde{\mu}\tilde{\phi}_{t+\tau}^2)\left[(1-\tilde{\phi}_{t+\tau})^2x_{t+\tau-1}^2+\sigma_\theta^2\right]\right], \tag{B.45}$$

subject to the following constraint:

$$x_t = (1-\tilde{\phi}_t)[(1-\tilde{\phi}_t)^k x_{t-1} - \hat{\theta}_t]. \tag{B.46}$$

C Numerical solution algorithm

For a given set of parameters, we convert the policy problem on lines B.45 and B.46 to a Bellman equation that we then solve by value function iteration. Briefly, this involves the following five steps:

1. Fix some grid X_G of values for the state variable x_{t-1};

2. Make some guess on the values that the central bank's value function takes at each $x_{t-1} \in X_G$;

3. Use a fine grid search to find the central bank's implied optimal choice on $\tilde{\phi}_t$ at each $x_{t-1} \in X_G$, evaluating any relevant expectations using a combination of Gauss-Hermite quadrature and linear interpolation – i.e.,

$$\tilde{\phi}_*(x_{t-1}) \in \arg\min_{\tilde{\phi}_t \in \tilde{\Phi}} \left\{ \begin{array}{l} (1+\tilde{\mu}\tilde{\phi}_t^2) \left[\left(1-\tilde{\phi}_t\right)^{2k} x_{t-1}^2 + \sigma_\theta^2 \right] \\ \\ +\beta_G \sum_{n=1}^N \omega_n V \left[\left(1-\tilde{\phi}_t\right)^{k+1} x_{t-1} - \left(1-\tilde{\phi}_t\right) \hat{\theta}_n \right] \end{array} \right\},$$

   where $\tilde{\Phi}$ is a fine grid; $\{\omega_n\}_{n=1}^N$ and $\{\hat{\theta}_n\}_{n=1}^N$ respectively denote the weights and nodes of an N-point quadrature rule approximating a normal shock with mean zero and variance σ_θ^2; and $V(\cdot)$ denotes a linear interpolant constructed from our current guesses on the central bank's value function;

4. For each $x_{t-1} \in X_G$, use $\tilde{\phi}_*(x_{t-1})$ from our previous step to update our guess on the value that the central bank's value function takes at the grid point in question;

5. Repeat steps three and four until successive guesses fall within some small tolerance of one another.

Further details on this algorithm and the accuracy of the solution it delivers are available on request.

D Alternate models of non-rational expectations

D.1 Reflective expectations

The reflective expectation concept was developed by García-Schmidt and Woodford (2019) and is similar to LKT, except that the reasoning underlying agents' expectations is assumed to involve a continuous process rather than discrete steps.

As under LKT, agents with reflective expectations are assumed to form their expectations by making guesses on the expectations of others and updating those guesses based on the aggregate outcomes that their current guesses would imply if true. However, rather than involving a finite number of discrete updates, the process is assumed to occur at an infinite number of points along a continuum $[0, k]$, with updating at each point governed by smooth differential equations.

To get more specific, fix some wage-setter $s \in [0,1]$ at some point $\tilde{k} \in [0,k]$ in his reasoning process. Let $\hat{\pi}^g(\tilde{k})$ denote his guess on the inflation expectations of others at this point – i.e., his current guess is that $\mathbb{E}_{t-1}^{WS(\tilde{s})}(\hat{\pi}_t) = \hat{\pi}^g(\tilde{k})\ \forall \tilde{s} \in [0,1] \setminus \{s\}$. Define $\hat{N}^g(\tilde{k})$ analogously and let $[\hat{\pi}^o(\tilde{k}), \hat{N}^o(\tilde{k})]$ denote the outcomes that would obtain on average if these guesses were true – that is:

$$\hat{\pi}^o(\tilde{k}) = \hat{\pi}^g(\tilde{k}) + \hat{N}^g(\tilde{k})$$
$$\hat{N}^o(\tilde{k}) = -\phi_t[\hat{\pi}^g(\tilde{k}) + \hat{N}^g(\tilde{k})].$$

For each $z \in \{\pi, N\}$, we assume that updating is governed by a differential equation of the form

$$\frac{d\hat{z}^g(\tilde{k})}{d(\tilde{k})} = \hat{z}^o(\tilde{k}) - \hat{z}^g(\tilde{k})$$

– i.e., guesses are updated in the direction of the aggregate outcomes that they would imply if true.

If we define $x^g(\cdot) \equiv \hat{\pi}^g(\cdot) + \hat{N}^g(\cdot)$, then this problem collapses to a single differential equation of the form

$$\frac{dx^g(\tilde{k})}{d\tilde{k}} = -\phi_t x^g(\tilde{k}), \tag{D.47}$$

which we will solve subject to the "adaptive" initial conditional $x^g(0) = x_{t-1}$. This yields a solution of the form $x^g(\tilde{k}) = \exp(-\tilde{k}\phi_t)x_{t-1}$. Since all wage-setters are assumed to stop updating at the final level k, this means that they ultimately share common expectations of the form

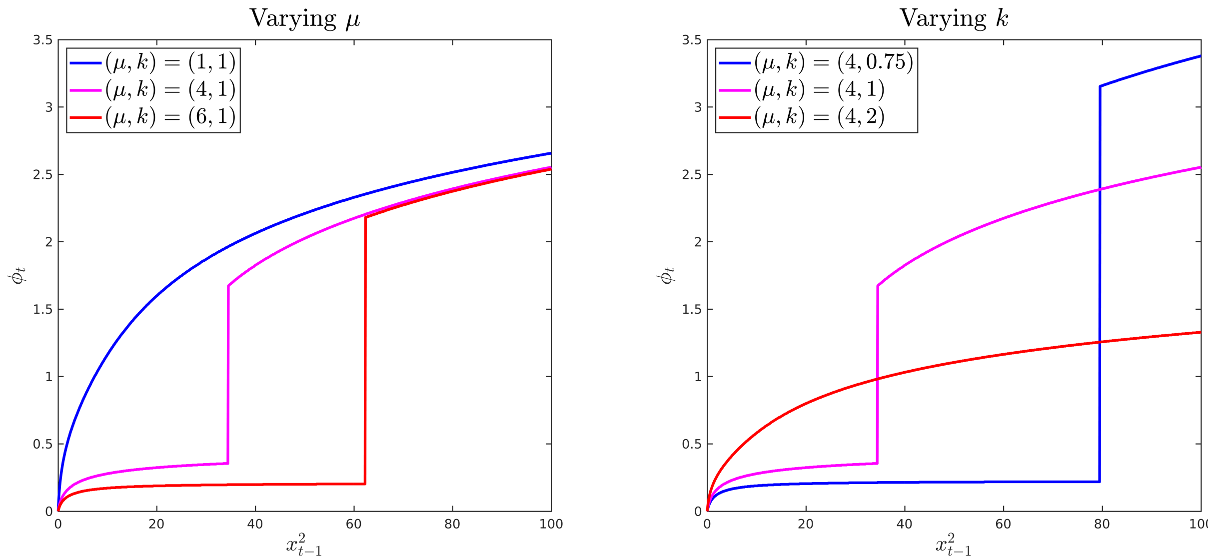
$$\mathbb{E}_{t-1}^{WS}(x_t) = \exp(-k\phi_t)x_{t-1}. \tag{D.48}$$

Assuming $\phi_t > 0$, RE and AE are thus respectively nested as the special cases $k \to \infty$ and $k = 0$, as was the case under LKT. Since $\left|\frac{d\mathbb{E}_{t-1}^{WS}(x_t)}{d\phi_t}\frac{\phi_t}{\mathbb{E}_{t-1}^{WS}(x_t)}\right| = k\phi_t$, another key commonality with LKT is that private-sector expectations are more sensitive to changes in the policy stance when ϕ_t is high.

Turning to actual inflation outcomes, they are given by the following expression when expectations take the form on line D.48:

$$\hat{\pi}_t = \exp(-k\phi_t)x_{t-1} - \hat{\theta}_t.$$

Figure 12: Optimal pivots with reflective expectations



Notes. The panels depict the policy stance that globally minimizes the loss function on line D.49 for a range of values for the state variable x_{t-1}^2 and various cases involving different values for the parameters μ and k. The only remaining parameter σ_θ^2 has been normalized to one in both panels.

Combining this expression with the central bank's feedback rule, we find that policy-makers' expected losses in period t are given by

$$\mathbb{E}_{t-1}^{RE}(\hat{\pi}_t^2 + \mu \hat{N}_t^2) = (1 + \mu\phi_t^2)\left[\exp(-2k\phi_t)x_{t-1}^2 + \sigma_\theta^2\right], \tag{D.49}$$

so a myopic central bank will choose ϕ_t to minimize this objective. Figure 12 depicts numerical solutions to this problem for several illustrative calibrations and a range of values for the state variable x_{t-1}^2. As under LKT, pivots emerge when the policy preference parameter μ is sufficiently large (left-hand panel) and the parameter k is sufficiently small (right-hand panel).

D.2 Cognitive hierarchy

Cognitive hierarchies are a generalization of LKT that allows for heterogeneity in the sophistication with which agents form their expectations. The concept was first developed by Camerer et al. (2004).

In this subsection, we replace LKT with a cognitive hierarchy comprising $k + 1$ levels, where $k \in \{1, 2, \ldots\}$. Wage-setters at level $\tilde{k} = 0$ have adaptive expectations of the same form described in the main text. At any given higher level $\tilde{k} \in \{1, 2, \ldots, k\}$, wage-setters form their expectations based on an assumption that all other wage-setters are on level $(\tilde{k} - 1)$ or lower. A wage-setter at level $\tilde{k} = 1$ thus believes that all other wage-setters are adaptive, while a wage-setter at level $\tilde{k} = 2$ believes that the economy is populated by a mix of level-0 and level-1 wage-setters, etc.

To get more specific, let $f^{\tilde{k}} \geq 0$ denote the true share of wage-setters on level $\tilde{k} \in \{0, 1, \ldots, k\}$, while $g^{\tilde{k}}(\underset{\sim}{k})$ denotes these wage-setters' subjective belief on the share of wage-setters on the lower level $\underset{\sim}{k} \in \{0, 1, \ldots, \tilde{k} - 1\}$. LKT would thus correspond to the special case where $f^k = 1$ and $g^k(k-1) = 1$. Instead, we will now leave the true distribution $\{f^{\tilde{k}}\}_{\tilde{k}\in\{0,1,\ldots,k\}}$ generic, though we follow Camerer et al. in restricting wage-setters' subjective beliefs to be consistent with this distribution in the sense that

$$g^{\tilde{k}}(\underset{\sim}{k}) = \frac{f^{\underset{\sim}{k}}}{\sum_{\hat{k}=0}^{\tilde{k}-1} f^{\hat{k}}} \quad \forall \underset{\sim}{k} \in \{0, 1, \ldots, \tilde{k} - 1\} \text{ and all } \tilde{k} \in \{1, 2, \ldots, k\} \text{ s.t. } \sum_{\hat{k}=0}^{\tilde{k}-1} f^{\hat{k}} > 0.$$ [14]

For technical reasons that will become clear in a moment, it will also be convenient to define $g^{k+1}(\underset{\sim}{k}) \equiv f^{\underset{\sim}{k}}\ \forall \underset{\sim}{k} \in \{0, 1, \ldots, k\}$.

With these assumptions and conventions in hand, we introduce some notation. For each $\tilde{k} \in \{0, 1, \ldots, k\}$, let $W_t^{\tilde{k}}$ denote the nominal wage set by wage-setters on level $\tilde{k}$. At the same time, let $\overline{W}_t^{\tilde{k}}$ denote the aggregate wage that would prevail in a counterfactual economy in which the distribution of wage-setters is given by $\{g^{\tilde{k}+1}(\underset{\sim}{k})\}_{\underset{\sim}{k}\in\{0,1,\ldots,\tilde{k}\}}$ – i.e., the counterfactual economy is one in which the most sophisticated wage-setters are on level $\tilde{k}$, with population shares for them and all lower-level wage-setters proportional to the corresponding population shares in the true economy. It will also be useful to let $P_t^{\tilde{k}}$ and $\hat{\pi}_t^{\tilde{k}} \equiv \ln P_t^{\tilde{k}} - \ln P_{t-1} - \pi^*$ respectively denote the price level and inflation gap prevailing under the counterfactual that we just described, and to define the following normalized analogues for $W_t^{\tilde{k}}$ and $\overline{W}_t^{\tilde{k}}$:

$$\left(w_t^{\tilde{k}}, \overline{w}_t^{\tilde{k}}\right) \equiv \left[\frac{W_t^{\tilde{k}}}{\exp(\pi^*)P_{t-1}\theta_{t-1}} \frac{\gamma - 1}{\gamma}, \frac{\overline{W}_t^{\tilde{k}}}{\exp(\pi^*)P_{t-1}\theta_{t-1}} \frac{\gamma - 1}{\gamma}\right]. \tag{D.50}$$

Under this notation, wage-setters at level $\tilde{k} \in \{1, 2, \ldots, k\}$ believe that the aggregate nominal wage and price level will respectively be given by $W_t = \overline{W}_t^{\tilde{k}-1}$ and $P_t = P_t^{\tilde{k}-1}$. This leads to the following recursion:

Lemma D.4. *Wages in successive counterfactual economies are linked by the following expression,*

[14] If instead $\sum_{\hat{k}=0}^{\tilde{k}-1} f^{\hat{k}} = 0$, we adopt a convention that $g^{\tilde{k}}(\underset{\sim}{k} - 1) = 1$ – i.e., wage-setters at level $\tilde{k}$ in the cognitive hierarchy have the same expectations as would level-$\tilde{k}$ wage-setters under LKT.

which holds up to second- and higher-order terms:

$$\ln \overline{w}_t^{\tilde{k}} = \left(1 - \phi_t \frac{f^{\tilde{k}}}{\sum_{\underset{\sim}{k}=0}^{\tilde{k}} f^{\underset{\sim}{k}}}\right) \ln \overline{w}_t^{\tilde{k}-1},$$

where $\tilde{k} \in \{1, 2, \dots, k\}$*, with* $\overline{w}_t^0 = \exp(x_{t-1})$.

Proof. Fix some $\tilde{k} \in \{0, 1, \dots, k\}$. Based on the wage-setting condition on line 2.10 of the main text, we know that

$$\ln W_t^{\tilde{k}} = \ln\left(\frac{\rho\eta}{\rho - 1}\right) + \mathbb{E}_{t-1}^{\tilde{k}}(\ln P_t) + \mathbb{E}_{t-1}^{\tilde{k}}(\ln N_t) + \mathbb{E}_{t-1}^{\tilde{k}}(\ln \theta_t),$$

up to second- and higher-order terms associated with Jensen's inequality. Combining this expression with the normalization on line D.50 and the random walk property of θ_t then yields

$$\ln w_t^{\tilde{k}} = \mathbb{E}_{t-1}^{\tilde{k}}(\hat{\pi}_t + \hat{N}_t). \tag{D.51}$$

So, if $\tilde{k} = 0$, we immediately have $w_t^0 = \overline{w}_t^0 = \exp(x_{t-1}) > 0$ as claimed.

Suppose instead that $\tilde{k} \in \{1, \dots, k\}$ and fix some generic $\overline{W}_t^{\tilde{k}-1} > 0$. In this case, we have $\mathbb{E}_{t-1}^{\tilde{k}}(\hat{N}_t) = -\phi_t \mathbb{E}_{t-1}^{\tilde{k}}(\hat{\pi}_t)$, so equation (D.51) collapses to the following:

$$\ln w_t^{\tilde{k}} = (1 - \phi_t)\mathbb{E}_{t-1}^{\tilde{k}}(\hat{\pi}_t). \tag{D.52}$$

Note here that wage-setters at level $\tilde{k}$ in the cognitive hierarchy believe that prices are given by

$$P_t = P_t^{\tilde{k}-1} = \frac{\gamma}{\gamma - 1} \frac{\overline{W}_t^{\tilde{k}-1}}{\theta_t}.$$

After accounting for the normalization on line D.50, this means that level-$\tilde{k}$ wage-setters believe that inflation is given by

$$\hat{\pi}_t = \hat{\pi}_t^{\tilde{k}-1} = \ln\left(\frac{\overline{w}_t^{\tilde{k}-1}}{\theta_t/\theta_{t-1}}\right),$$

yielding

$$\mathbb{E}_{t-1}^{\tilde{k}}(\hat{\pi}_t) = \ln \overline{w}_t^{\tilde{k}-1},$$

again up to second- and higher-order terms. Equation (D.52) thus finally becomes

$$\ln w_t^{\tilde{k}} = (1 - \phi_t) \ln \overline{w}_t^{\tilde{k}-1}. \tag{D.53}$$

At this point, it is useful to note that the expression for the aggregate wage on line 2.2 of the main text can be re-written as follows:

$$\left(\overline{w}_t^{\tilde{k}}\right)^{1-\rho} = \left(1 - \frac{f^{\tilde{k}}}{\sum_{\underset{\sim}{k}=0}^{\tilde{k}} f^{\underset{\sim}{k}}}\right) \left(\overline{w}_t^{\tilde{k}-1}\right)^{1-\rho} + \left(\frac{f^{\tilde{k}}}{\sum_{\underset{\sim}{k}=0}^{\tilde{k}} f^{\underset{\sim}{k}}}\right) \left(w_t^{\tilde{k}}\right)^{1-\rho}. \tag{D.54}$$

Based on this expression, line D.53, and our previous finding that $\overline{w}_t^0 = \exp(x_{t-1})$, we can conclude that $\overline{w}_t^{\tilde{k}} = \overline{w}_t^{\tilde{k}-1} = w_t^{\tilde{k}} = 1$ when $x_{t-1} = 0$. Linearizing equation D.54 around this point and using line D.53 to eliminate $w_t^{\tilde{k}}$ from the resulting expression then gives the recursion being claimed. ■

This lemma implicitly defines a function $\overline{w}(\cdot, \cdot)$ satisfying $\overline{w}_t^k = \overline{w}\left(\phi_t, x_{t-1}\right)$ for arbitrary values for ϕ_t and x_{t-1}. Given this function, the inflation gap in the true economy is given by

$$\hat{\pi}_t = \hat{\pi}_t^k = \ln\left[\frac{\overline{w}\left(\phi_t, x_{t-1}\right)}{\theta_t/\theta_{t-1}}\right].$$

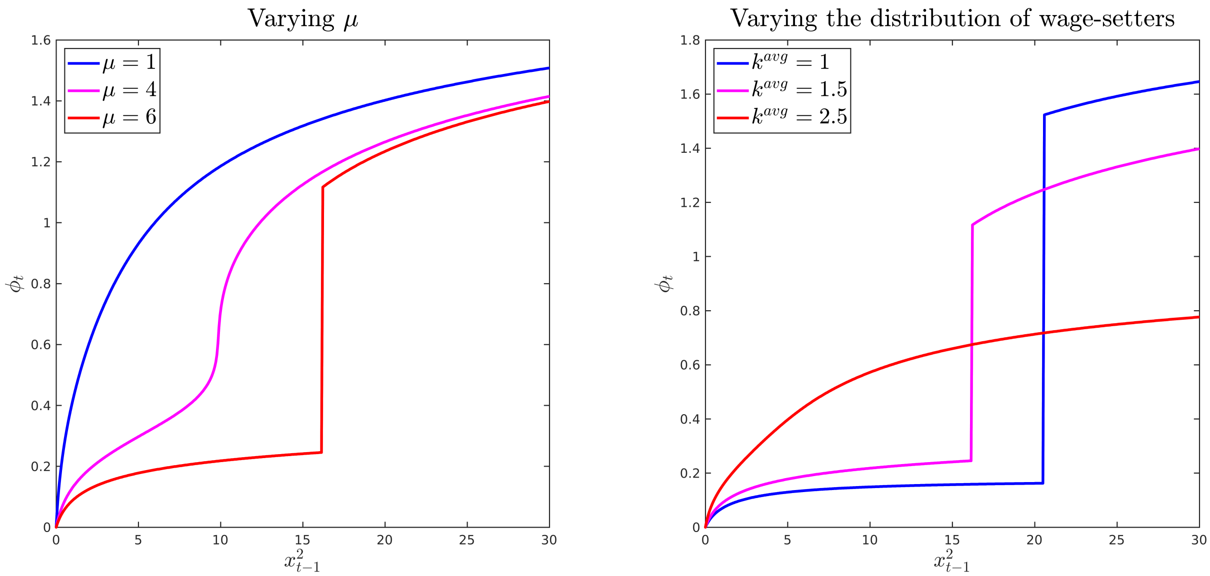
Combining this expression with the central bank's feedback rule, we find that policy-makers' expected losses in period t are given by

$$\mathbb{E}_{t-1}^{RE}(\hat{\pi}_t^2 + \mu \hat{N}_t^2) = (1 + \mu\phi_t^2)\left[\left[\ln\left\{\overline{w}\left(\phi_t, x_{t-1}\right)\right\}\right]^2 + \sigma_\theta^2\right], \tag{D.55}$$

and that a myopic central bank would thus choose ϕ_t to minimize this objective.

Figure 12 depicts numerical solutions to this problem for several illustrative calibrations and a range of values for the state variable x_{t-1}^2. The left-hand panel indicates that pivots emerge when the policy preference parameter μ is sufficiently large, as under LKT and reflective expectations. On the other hand, the right-hand panel focuses on the consequences of varying the distribution of wage-setters across the cognitive hierarchy and specifically assumes that $k = 4$, corresponding to the top of the range considered in Farhi and Werning (2019). Given this upper bound on wage-setters' sophistication, the blue line in the panel assumes population weights $(f^0, f^1, f^2, f^3, f^4) = (1/3, 1/3, 1/3, 0, 0)$, yielding an "average k" of one. In contrast, the magenta and red lines respectively assume $(f^0, f^1, f^2, f^3, f^4) = (1/4, 1/4, 1/4, 1/4, 0)$ and $(f^0, f^1, f^2, f^3, f^4) = (0, 1/4, 1/4, 1/4, 1/4)$.

Figure 13: Optimal pivots with cognitive hierarchies

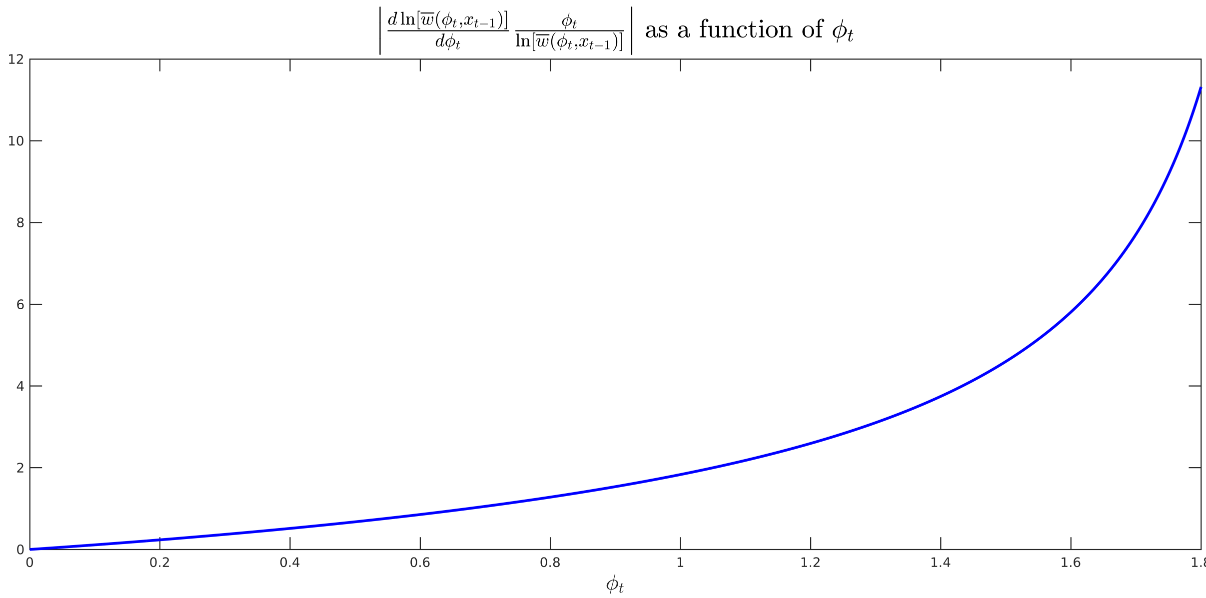


Notes. The panels depict the policy stance that globally minimizes the loss function on line D.55 for a range of values for the state variable x^2_{t-1} and various cases involving different values for the parameter μ and the distribution of wage-setters across the cognitive hierarchy. The left-hand panel assumes $k = 3$, with equal population shares at all four levels in the implied hierarchy. The right-panel assumes $(\mu, k) = (6, 4)$, with population shares of the form described in the main text. The only remaining parameter σ^2_θ has been normalized to one in both panels.

Comparing the results for these three cases, we see that pivots only occur for sufficiently low levels of sophistication, again as under LKT and reflective expectations.

Moreover, comparing equation (D.55) against its analogues under LKT (equation 3.22) and reflective expectations (equation D.49), we see that the term $\ln[\overline{w}(\phi_t, x_{t-1})]$ in equation (D.55) plays a role analogous to the private sector's expected level of overheating under the other two models and effectively represents an aggregate of the expectations distributed across the cognitive hierarchy. Though the elasticity of this term with respect to the policy stance is difficult to characterize analytically, it can be approximated using finite differences and is depicted in Figure 14 for an illustrative parameterization and a range of values for ϕ_t. As under LKT and reflective equilibria, the (unsigned) elasticity is found to be increasing in ϕ_t, suggesting that cognitive hierarchies share the other two models' property that the policy-relevant measure of private-sector expectations is more sensitive to changes in the policy stance when ϕ_t is high.

Figure 14: Sensitivity of $\ln\left[\overline{w}\left(\phi_t, x_{t-1}\right)\right]$ to changes in the policy stance under cognitive hierarchies



Notes. The profile depicted is invariant to the value assumed for the state variable x_{t-1}. The other values assumed in the figure are $(\mu, k, \sigma_\theta^2) = (6, 4, 1)$, and $(f^0, f^1, f^2, f^3, f^4) = (1/4, 1/4, 1/4, 1/4, 0)$.

E Alternate assumptions on agents' level-0 guesses

In this section, we retain the baseline assumption that wage-setters form their expectations using LKT and explore the consequences of different assumptions on the level-zero guesses at which they initialize their reasoning. More specifically, while the main text consistently assumed that level-zero guesses were based on simple adaptive rules of the form

$$\mathbb{E}^0_{t-1}(\hat{z}_t) = \hat{z}_{t-1}, \quad \forall z \in \{\pi, N\}, \tag{E.56}$$

we will now entertain cases where these guesses are instead based on a range of backward-looking learning rules. In the special case $k = 0$, this leads to a situation where expectations are purely based on learning, while setting $k \in \{1, 2, \ldots\}$ leads to a hybrid of LKT and learning.

E.1 Constant-gain learning

To implement LKT with constant-gain learning at level zero, we drop line E.56 and instead assume that wage-setters base their initial, level-zero guesses on a learning rule of the form

$$\mathbb{E}^0_{t-1}(\hat{z}_t) = \mathbb{E}^k_{t-2}(\hat{z}_{t-1}) + \gamma_z[\hat{z}_{t-1} - \mathbb{E}^k_{t-2}(\hat{z}_{t-1})], \quad \forall z \in \{\pi, N\}, \tag{E.57}$$

where $\gamma_z \in [0, 1]$ denotes a fixed gain rate.

In this case, wage-setters' final, level-k expectations are still governed by equations (3.17) and (3.18) in the main text, and the Phillips curve and feedback rule that respectively appear on lines 2.12 and 2.14 continue to hold. Using these equations, the loss function on line 3.19 can be expressed as

$$\mathbb{E}_{t-1}^{RE}\left[\sum_{\tau=0}^{\infty}\beta_{CB}^{\tau}(1+\mu\phi_{t+\tau}^{2})\left[(1-\phi_{t+\tau})^{2}\left[\mathbb{E}_{t+\tau-1}^{0}\left(x_{t+\tau}\right)\right]^{2}+\sigma_{\theta}^{2}\right]\right], \tag{E.58}$$

where $x_t \equiv \hat{\pi}_t + \hat{N}_t$ as in the main text. At the same time, the law of motion for the new state variable $\mathbb{E}_{t-1}^{0}\left(x_t\right)$ can be expressed as

$$\mathbb{E}_{t}^{0}(x_{t+1}) = \mathbb{E}_{t-1}^{0}(x_t) + \gamma_{\pi}[\hat{\pi}_t - \mathbb{E}_{t-1}^{0}(\hat{\pi}_t)] + \gamma_N[\hat{N}_t - \mathbb{E}_{t-1}^{0}(\hat{N}_t)]$$

$$= (1-\phi_t)^k\mathbb{E}_{t-1}^{0}(x_t) - (\gamma_{\pi} - \gamma_N\phi_t)[\phi_t(1-\phi_t)^k\mathbb{E}_{t-1}^{0}(x_t) + \hat{\theta}_t], \tag{E.59}$$

so optimal policy now involves choosing ϕ_t to minimize line E.58, subject to this constraint.

Figure 15 depicts numerical solutions to this problem for several illustrative calibrations and a range of values for the state variable $E_{t-1}^{0}(x_t)$. As shown in the left-hand panel, pivots emerge when the policy preference parameter μ is sufficiently large, as was the case under LKT. On the other hand, the right-hand panel entertains a range of choices on the parameter k and again points to patterns similar to those previously documented under LKT. In particular, pivots do not occur when expectations are fully backward-looking ($k = 0$) and otherwise occur only so long that k is sufficiently is small.

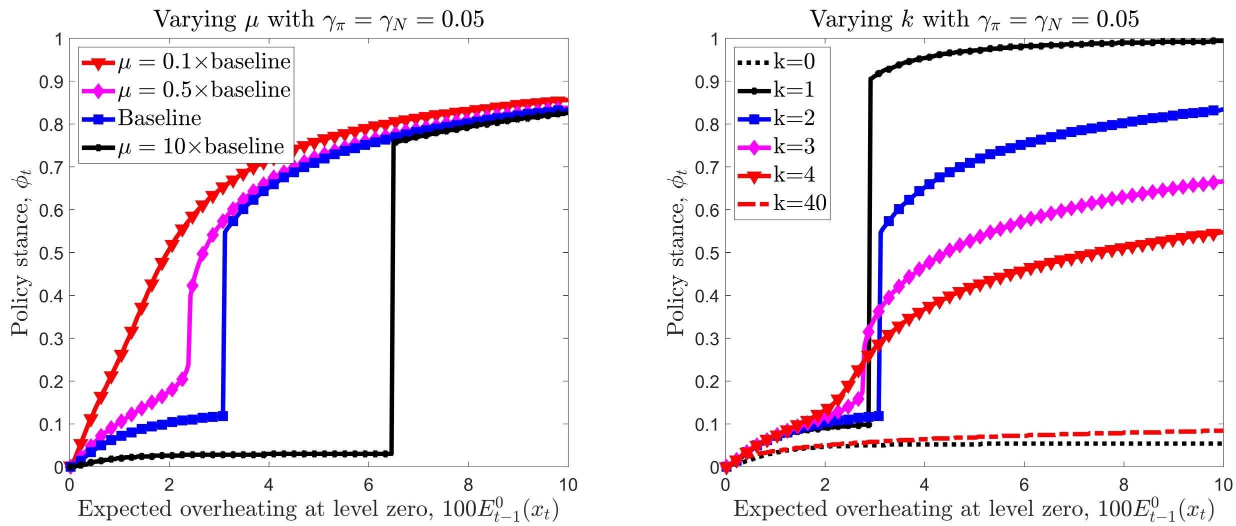
E.2 Decreasing-gain learning

LKT with decreasing-gain learning at level zero can be implemented along lines similar to the constant-gain case above, except now we allow wage-setters' gain rates to fall over time – i.e.,

$$\mathbb{E}_{t-1}^{0}(\hat{z}_t) = \mathbb{E}_{t-2}^{k}(\hat{z}_{t-1}) + (\overline{\gamma}_z/t)[\hat{z}_{t-1} - \mathbb{E}_{t-2}^{k}(\hat{z}_{t-1})], \quad \forall z \in \{\pi, N\}, \tag{E.60}$$

where $\overline{\gamma}_z \in [0, 1]$ denotes the gain rate in some initial period $t = 1$. In this case...

Figure 15: Optimal pivots with constant-gain learning



Notes. The panels depict the policy stance that globally minimizes the loss function on line E.58 subject to the law of motion on line E.59, considering a range of values for the state variable $\mathbb{E}^0_{t-1}(x_t)$ and various cases involving different values for the parameters μ and k. The baseline calibration in both figures is the same as in section 4 of the main text, except that λ has been normalized to one while μ has been adjusted to keep the ratio μ/λ^2 unchanged (thus leaving the location of any pivot points unchanged).

F Oil price shocks as productivity shocks

One of the objectives of this paper has been to examine optimal monetary policy when an economy faces supply-side shocks, with oil price shocks being one especially relevant example. However, in the model, the only source of uncertainty is a productivity shock. We therefore use this section to establish an equivalence between oil price shocks and productivity shocks in the context of a small, oil-importing economy.

Let the gross output of this economy be given by

$$Q_t = F(Z_t N_t, O_t), \tag{F.61}$$

where Z_t is a labour-augmenting productivity factor; N_t is a labour input; O_t is an oil input; and the function $F(\cdot,\cdot)$ is assumed to be strictly concave in both its arguments, and to exhibit constant returns to scale (CRS) in these two arguments. The corresponding value-added function is

$$\frac{Y_t}{N_t} = F\left(Z_t, \frac{O_t}{N_t}\right) - P^O_t \frac{O_t}{N_t},$$

where P^O_t denotes the price of a unit of oil. Note here that we have used the CRS property of F

to write the value-added function in per-worker terms.

The implied first-order condition with respect O_t reads as

$$F_2\left(Z_t, \frac{O_t}{N_t}\right) = P_t^O,$$

where $F_2(\cdot,\cdot)$ denotes the derivative of F with respect to its second argument. Moreover, since F is CRS in its two arguments, we have

$$F_2\left(Z_t, \frac{O_t}{N_t}\right) = F_2\left(1, \frac{O_t}{Z_t N_t}\right) = f'\left(\frac{O_t}{Z_t N_t}\right),$$

where $f(\cdot) \equiv F(1,\cdot)$, and thus $f''(\cdot) < 0$. As a result, the first-order condition above can be re-written as a downward-sloping oil demand curve of form

$$\frac{O_t}{N_t} = Z_t (f')^{-1}(P_t^O). \tag{F.62}$$

Plugging this demand curve back into the value-added production function then gives

$$Y_t = N_t Z_t h(P_t^O),$$

where $h(P_t^O) \equiv f[(f')^{-1}(P_t^O)] - P_t^O (f')^{-1}(P_t^O)$ and thus

$$h'(P_t^O) = -(f')^{-1}(P_t^O) < 0.$$

Defining $\theta_t \equiv Z_t h(P_t^O)$, we finally get a value-added production function of the form

$$Y_t = \theta_t N_t, \tag{F.63}$$

where labour productivity θ_t is a decreasing function of the oil price P_t^O.

References cited in the online appendix

Camerer, C. F., T. Ho, and J. Chong (2004, August). A cognitive hierarchy model of games. *Quarterly Journal of Economics 119*(3), 861–898.

Farhi, E. and I. Werning (2019, November). Monetary policy, bounded rationality, and incomplete markets. *American Economic Review 109*(11), 3887–3928.

García-Schmidt, M. and M. Woodford (2019, January). Are low interest rates deflationary? A paradox of perfect-foresight analysis. *American Economic Review 109*(1), 86–120.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988, June). Investment, capacity utilization, and the real business cycle. *American Economic Review 78*(3), 402–417.