Price Gaps and Inflation Dynamics*

Miguel Bandeira† Laura Castillo-Martínez‡ Shiyuan Wang§

April 2026

Abstract

Non-linear inflation dynamics are a key feature of menu cost models. Recent work has mostly focused on one dimension of this non-linearity: larger shocks have disproportionately larger effects. Yet the impact of a given shock also depends on the state of the economy at the time it hits. This paper studies how the cross-sectional distribution of price gaps—the misalignment between firms' actual and reset prices—shapes inflation dynamics. A central challenge is that this distribution is unobservable. We address this by recasting a random menu cost model in state-space form and developing Bayesian methods to recover quote-level price gaps from observed prices. Applying this approach to UK CPI microdata from 1997 to 2023, we construct a time-varying distribution of price gaps and document substantial variation over time, with the economy often far from its ergodic benchmark. We show that this variation matters: it gives rise to state-dependent transmission of monetary policy and serves as a leading indicator of future inflation.

*We are especially grateful to Ricardo Reis for his guidance and encouragement throughout this project. We also thank Andrea Alati, Isaac Baley, Andres Blanco, Marco Bonomo, Adrien Bussy, Patrick Coen, Thomas Drechsel, EAGLS, Mike Elsby, Friedrich Geiecke, Elena Manresa (discussant), William Matcham, Ben Moll, Rafael Rocha, Kevin Sheedy, Javier Turen; as well as participants at Barcelona GSE Summer Forum, FRB St. Louis, "Inflation: Causes, Consequences, and Policy Implications" Workshop in Milan, Insper, LSE, PUC-Rio, SED Barcelona, the University of Minnesota, the University of São Paulo, the Virtual Israel Macro Meeting, the Wharton School and the 7th Annual Women in Macroeconomics Conference for insightful comments. A previous version of this paper circulated under the title "Frictionless Inflation". The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

†Insper and Centre for Macroeconomics. Email: mbandeira@insper.edu.br

‡Duke University, FRB Minneapolis and CEPR. Email: l.castillo-martinez@duke.edu

§University of Minnesota. Email: wan02274@umn.edu

1 Introduction

The surge and retreat of post-Covid inflation has renewed interest in the non-linear dynamics of price adjustment. Menu cost models are the natural framework for thinking about them: firms tolerate some price misalignment due to adjustment costs, and how many of them adjust in response to a given shock is endogenous. The literature has mostly focused on one dimension of this non-linearity, shock size, showing that large shocks generate disproportionate price responses. But these models feature a second dimension that has received less attention: the state of the economy when the shock arrives. This is captured by the cross-sectional distribution of price gaps—the misalignment between firms' actual and reset prices.

Other settings with lumpy adjustment, such as investment and durable goods purchases, have explored this source of state dependence more extensively. There, gaps between actual and target levels can be constructed from observable fundamentals or proxied by the age of the stock. By contrast, measuring gaps in pricing requires knowledge of firms' optimal prices, which depend on marginal costs that are difficult to measure at the firm level. We develop methods to recover the time-varying distribution of price gaps from observed prices and show that this distribution is a powerful tool for understanding non-linear inflation dynamics.

This paper makes three contributions. First, using a random menu cost model, we revisit the gap distribution as a source of state dependence and show how it matters for shock propagation, optimal policy, and future inflation. Second, we recast the model in a state-space framework and develop Bayesian methods to recover quote-level price gaps from observed prices. Third, we apply this approach to UK CPI microdata from 1997 to 2023 and document that the gap distribution varies over time, with the economy spending extended periods far from its ergodic distribution. We show that this variation matters: it gives rise to state-dependent transmission of monetary policy and serves as a leading indicator of future inflation, beyond what past inflation alone can predict.

Section 2 develops the framework. We lay out a canonical random menu cost model à la Nakamura and Steinsson (2010) in which firms face a fixed cost to change their price, which collapses to zero with some probability. When a firm adjusts, it resets to a target level that evolves as a random walk with drift; between adjustments, the widening gap between the actual price and this target generates a profit loss. The optimal pricing policy takes the form of an Ss rule: the firm leaves its price unchanged as long as the gap remains inside an inaction region, and resets whenever the gap hits a threshold or a free adjustment arrives. Because the policy parameters are common across firms, the model aggregates cleanly: observed inflation reflects both changes in target prices and shifts in the cross-sectional distribution of gaps, G_t, which becomes the key state variable.

This distribution captures the full state of price misalignment in the economy at each

point in time. Because firms adjust intermittently, G_t encodes information about past shocks that have not yet fully passed through to prices, and its moments have direct macroeconomic content. The mean reflects accumulated inflationary or deflationary pressure not yet released through price changes. The variance measures the welfare cost of misalignment across firms. Higher-order moments carry additional information: the distribution may be asymmetric, with more firms poised to raise than to lower prices, or heavy-tailed, with an unusual concentration of firms near their adjustment thresholds.

We show that the same economy can respond very differently to the same aggregate shock depending on the state of G_t when it hits. If many firms have accumulated large gaps near their adjustment thresholds, even a small shock can generate a wave of price changes. Conversely, if most firms have recently adjusted and their gaps are clustered near the reset point, even a large shock may trigger few adjustments. We illustrate these effects numerically using simulations from the calibrated model. This state dependence matters for optimal policy: the welfare cost of a given inflation rate depends on the underlying distribution that generates it, so the optimal policy response must account for the shape of G_t.

Beyond its role in shock propagation, G_t also contains information about the future path of inflation. Under infrequent adjustment, the density across the inaction region acts as a schedule of upcoming price changes: firms near the lower threshold will soon reprice upward, while those deeper in the interior will follow in later periods. This predictable unwinding of pent-up pressure is not captured by the history of inflation itself, because many distinct gap distributions can generate the same inflation record. We formalize this intuition in a deterministic special case of the model, showing that G_t Granger-causes inflation even after conditioning on its full history. More broadly, constructing and using the gap distribution empirically requires recovering price gaps from observed data.

Price gaps are not directly observable, which brings us to the methodological contribution of the paper. Section 3 recasts the pricing dynamics implied by the random menu cost model in state-space form. Given a panel of observed price trajectories, each price serves as a measurement variable, and the state vector for each quote-line includes its reset price and an indicator for the arrival of a free adjustment opportunity. The *Ss* rule serves as the measurement equation, mapping past observed prices and current latent states into current observed prices. The model's assumptions about the evolution of reset prices and adjustment opportunities dictate the transition equations, and an initial distribution over the latent states completes the specification. The goal is to recover the sequence of latent states for each quote-line and the model parameters that govern them.

To recover the latent states conditional on a given set of parameters, we solve the filtering and backward recursions implied by the state-space representation. Although the *Ss* rule introduces kinks that preclude a closed-form solution, we exploit its structure to simplify the problem. In particular, the rule tightly bounds the latent state. When a firm's price remains unchanged, we know with certainty that no free adjustment opportunity arrived.

When prices do change, the new price reveals the reset price exactly. At the same time, the size and direction of the adjustment are informative about what triggered it: if the observed change is smaller than the width of the inaction region, the firm must have received a free adjustment; if it is large enough that the gap moved outside the inaction region, the firm may have paid the menu cost, though the trigger itself remains unknown. These features mean that for many observations the latent state is either directly observed or tightly constrained, reducing the effective dimensionality of the forward-filtering backward-sampling (FFBS) problem. We take advantage of this by using grid-based methods that concentrate computational effort on the subset of states where inference remains nontrivial.

To jointly estimate the model parameters and latent states, we embed the FFBS algorithm in a Gibbs sampler. The parameters of the reset price process (trend drift and idiosyncratic volatility) and the free-adjustment probability admit conjugate full conditional distributions and can be updated directly. The latent states and inaction thresholds, by contrast, are tightly linked by the *Ss* rule: changing the thresholds changes which state trajectories are admissible, and vice versa. We therefore update them jointly through a Metropolis-Hastings step, proposing candidate states from the grid-based FFBS approximation and candidate thresholds from truncated normal random walks. A Monte Carlo experiment confirms that the algorithm recovers both parameters and price gaps with negligible bias and well-calibrated posterior uncertainty, even with panels as small as 100 quote-lines observed over 24 months.

Section 4 applies the method to the universe of locally collected UK CPI price quotes, covering over 23 million price observations across nearly 1,000 distinct items. We estimate the model separately by item, allowing parameters to vary across them. The estimated price gaps allow us to construct, for the first time, a time-varying empirical distribution of price gaps for a major economy. The distribution varies substantially over time: the economy spends extended periods far from its ergodic distribution, and the moments of G_t shift persistently in response to changing macroeconomic conditions. Two dates illustrate the range of variation. In December 2008, a temporary VAT cut triggered a synchronized wave of price adjustments that left the distribution tightly concentrated near the reset point. In March 2021, a year of pandemic-era supply disruptions and rising input costs had pushed nearly half of all price setters into negative territory without repricing. Cumulative inflation over the following twelve months was nearly three times as large after March 2021 as after December 2008.

We put the estimated gap distribution to work in Section 5 and establish two main empirical results. First, the gap distribution is a source of state dependence in the transmission of monetary policy. We test this using state-dependent local projections with high-frequency identification of monetary policy shocks based on Braun *et al.* (2025). The inflationary effect of an expansionary monetary shock depends strongly on the shape of G_t at the time of impact, with skewness as the dominant moment: a negatively skewed distribution, with a thick left tail of firms whose prices have fallen behind, amplifies the response. Evaluating

the estimated impulse responses under the December 2008 and March 2021 distributions, the cumulative price response to the same shock differs by more than a factor of two.

Second, the gap distribution is a leading indicator for inflation. Panel Granger-causality tests confirm our theoretical prediction that lagged moments of the gap distribution predict future inflation, but not the reverse. We then quantify the gain through a horse-race forecasting regression: gap moments contain significant predictive content beyond past inflation at horizons up to six months. The gains hold out of sample as well, with forecasts augmented by gap moments producing lower root mean squared errors than a lagged-inflation benchmark at all horizons and statistically significant improvements at one and three months ahead. Section 6 concludes.

Relation to the literature. This paper builds on the tradition of menu cost models (Golosov and Lucas, 2007; Nakamura and Steinsson, 2010). A central challenge in this literature is that the distribution of price gaps, the key state variable, is both infinite-dimensional and unobservable. Much of the literature therefore treats the gap distribution as an object to circumvent rather than measure. One approach is structural: solving the full model either by imposing assumptions that reduce the dimensionality of the state space (Dotsey *et al.*, 1999) or by applying numerical approximations that track the distribution directly. Another approach is analytical: characterizing the response of inflation without solving for the full distribution, from the generalized *Ss* framework of Caballero and Engel (2007) to the sufficient statistics approach of Alvarez *et al.* (2016), which expresses propagation in terms of moments of the *steady-state* distribution of price changes. This paper takes a different route by estimating the time-varying gap distribution directly from micro data.

Within the menu cost literature, a growing body of work studies non-linear inflation dynamics. Early work stressed the role of shock size, showing that large shocks generate disproportionately large price-level responses (Golosov and Lucas, 2007), with first-order implications for optimal stabilization policy (Karadi *et al.*, 2025). In standard models, however, this channel is quantitatively limited once the prevalence of small price changes is taken into account (Blanco *et al.*, 2024a,b). The state-dependence channel we emphasize is complementary: it operates through the shape of the gap distribution rather than the magnitude of the shock. While this insight goes back to the generalized *Ss* framework of Caballero and Engel (1993, 2007), it has been difficult to study empirically in pricing because reset prices are unobservable.[1] A recent exception is Gagliardone *et al.* (2025), who estimate price gaps using data on costs and markups, though their focus is on how the pass-through of cost shocks varies non-linearly with shock size. Our approach recovers price gaps from observed prices alone and studies how the state of price misalignment shapes inflation dynamics.

Methodologically, this paper relates to the literature on filtering and smoothing in non-

[1] In other *Ss* applications such as investment and durable goods, the analogous state-dependence channel has been explored more extensively. See, for example, Caballero *et al.* (1995), Bachmann *et al.* (2013), and Berger and Vavra (2015).

linear, non-Gaussian state-space models. Engineering and signal processing use these techniques extensively, but macroeconomics has largely confined them to the estimation of nonlinear DSGE models (Herbst and Schorfheide, 2015). While filtering and smoothing admit closed-form solutions in the linear-Gaussian case (Kalman, 1960; Rauch, Tung and Striebel, 1965), nonlinear and non-Gaussian settings generally require numerical approximation.[2] Given the structure of our problem, we rely on grid-based methods following Kitagawa (1987) and Kramer and Sorenson (1988). Our contribution is to bring these tools to a new setting: *Ss* pricing models.

Finally, this paper connects the estimated gap distribution to two applied literatures. A large empirical literature documents that the effects of monetary policy vary with the business cycle (Tenreyro and Thwaites, 2016), firm balance sheets (Ottonello and Winberry, 2020), and credit conditions (Òscar Jordà *et al.*, 2020); we show that the cross-sectional distribution of price gaps is another source of this state dependence. We also contribute to inflation forecasting (see Faust and Wright, 2013, for a survey), showing that moments of the estimated gap distribution contain predictive content beyond past inflation, both in sample and out of sample.

2 Price gaps: concept and relevance

This section introduces the distribution of price gaps: what it is and why it matters. We begin by laying out the framework we use throughout, a canonical menu cost model with random free adjustment in discrete time, following Auclert *et al.* (2024) but allowing for trend inflation.[3] Appendix A.1 provides microfoundations for this framework. After defining price gaps and their cross-sectional distribution, we turn to why the gap distribution is important, highlighting its role in the non-linear transmission of aggregate shocks and as a source of predictive content for future inflation.

2.1 A simple model of price setting

A firm i produces a differentiated good and sets a log price $p_{i,t}$ in each period $t = 0, 1, 2, \ldots$. Its *reset price*, denoted $p^r_{i,t}$—the price the firm would choose upon adjustment—evolves according to a random walk with drift:

$$p^r_{i,t} = \mu_t + p^r_{i,t-1} + \varepsilon_{i,t}, \tag{1}$$

[2]The literature offers a wide range of methods, including extended and unscented Kalman filters, grid-based methods, and sequential Monte Carlo techniques such as particle filters and smoothers. See Särkkä (2013) for a comprehensive treatment and Doucet and Johansen (2011) for a detailed survey of particle filters. Extended Kalman filters are unsuitable here due to the non-differentiability of the *Ss* rule. Particle filters can handle non-differentiabilities but require tracking latent gaps for thousands of quote-lines simultaneously, and the number of particles needed to cover the joint state space grows exponentially. More sophisticated variants (guided or auxiliary particle filters) mitigate this by steering particles toward likely regions of the state space. Our grid-based approach takes this logic to its limit: the *Ss* rule provides exact analytical bounds on the latent state, so we sample directly from the truncated distribution rather than relying on stochastic approximation.

[3]This is the "CalvoPlus" model widely used in the analytical pricing literature.

where μ_t is the trend inflation rate and $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2)$ is an idiosyncratic shock, independently and identically distributed across firms and over time. Changes in μ_t capture the combined effect of aggregate fundamental shocks and the monetary policy response to them. We model these changes as unexpected.

Each period, the firm incurs a loss from deviating from its reset price. Adjusting the price requires paying a fixed cost $c_{i,t}$, independently and identically drawn from $\{0, \bar{c}\}$ each period: with probability $\lambda \in [0, 1]$, the firm receives a free adjustment ($c_{i,t} = 0$); with probability $1 - \lambda$, it must pay $\bar{c} > 0$.

The firm's problem can be written in terms of the *price gap* $x_{i,t} \equiv p_{i,t} - p_{i,t}^r$. The firm chooses a sequence $\{x_{i,t}\}_{t=0}^{\infty}$ to minimize expected discounted losses:

$$\min_{\{x_{i,t}\}_{t=0}^{\infty}} \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{2} x_{i,t}^2 + c_{i,t} \cdot \mathbf{1}_{\{x_{i,t} \neq x_{i,t-1} - \mu_t - \varepsilon_{i,t}\}} \right], \tag{2}$$

where the adjustment cost $c_{i,t}$ is incurred whenever the firm changes its price, i.e., when the price gap differs from its no-adjustment evolution $x_{i,t-1} - \mu_t - \varepsilon_{i,t}$.[4]

The optimal pricing policy takes the form of a standard *Ss* rule:[5]

$$x_{i,t} = \begin{cases} 0 & \text{with prob } \lambda \text{ or if } x_{i,t-1} - \mu_t - \varepsilon_{i,t} \notin [\underline{x}_t, \bar{x}_t] \\ x_{i,t-1} - \mu_t - \varepsilon_{i,t} & \text{otherwise} \end{cases} \tag{3}$$

The firm adjusts its price in one of two cases: either it receives a free adjustment opportunity, or the price gap falls outside an inaction region $[\underline{x}_t, \bar{x}_t]$, with $\underline{x}_t < 0 < \bar{x}_t$. Upon adjustment, the firm closes its gap entirely ($x_{i,t} = 0$), setting its price to the current reset level. Otherwise, the gap evolves passively, drifting with trend inflation and the idiosyncratic shock. In general, the reduced-form parameters $[\underline{x}_t, \bar{x}_t]$ vary over time.

2.2 The distribution of price gaps

Let G_t denote the cross-sectional distribution of price gaps across firms at time t, with density g_t over the inaction region $[\underline{x}_t, \bar{x}_t]$ and a mass point at the reset point $x = 0$. This distribution is the key state variable of the model: it encodes the full cross-sectional state of price misalignment. Under the *Ss* rule (3), G_t evolves through two opposing forces. Between adjustments, each firm's gap drifts with the common trend and its idiosyncratic

[4] Equation (2) is a reduced-form representation of the firm's pricing problem; Appendix A.1 derives it from a standard CES demand structure with random menu costs. This formulation is standard in the analytical menu cost literature, e.g., Alvarez and Lippi (2009), Alvarez *et al.* (2016), and Auclert *et al.* (2024).

[5] Auclert *et al.* (2024) prove that the *Ss* rule is the unique optimal policy when $\mu = 0$ in the steady state of the discrete-time model, and their contraction mapping argument guarantees that this structure persists in a neighborhood of $\mu = 0$. Working at the monthly frequency, even double-digit annual trend inflation yields a comfortably small μ. In addition, the *Ss* structure under positive drift is well established in continuous-time models. Even though a formal proof in the discrete-time case is challenging, Appendix A.2 lays out the case, covering both stationary and non-stationary environments.

shock, spreading mass toward the boundaries of the inaction region. At adjustment dates, triggered either by a free opportunity or by a gap crossing a threshold, the firm resets to zero, concentrating mass near the center and truncating the tails. In steady state, these forces balance to produce a stationary distribution with a characteristic shape: a concentration of mass near zero from recent adjusters, with tails extending toward both thresholds from firms that have not adjusted recently.

While G_t is an infinite-dimensional object, its key features can be read from a small number of moments. The *mean gap*, $E[x_t] \equiv \int x\, dG_t$, captures the average direction of price misalignment: when negative, firms' prices sit on average below their reset levels, reflecting accumulated inflationary pressure. The *variance*, $\text{Var}(x_t)$, measures how dispersed the misalignment is across firms; under the quadratic loss in (2), the aggregate welfare cost of price misalignment is

$$\mathcal{W}_t = \frac{1}{2}\left[\text{Var}(x_t) + (E[x_t])^2\right], \tag{4}$$

so that dispersion has direct welfare consequences even when the mean gap is zero.[6] The *skewness* captures directional asymmetry in the tails: whether more mass sits near the lower threshold (inflationary pressure) or the upper one (deflationary pressure). The *kurtosis* reflects how concentrated mass is near the adjustment thresholds versus the center of the inaction region.

From price gaps to inflation. Since $p_{i,t} = p^r_{i,t} + x_{i,t}$, each firm's inflation rate decomposes as

$$\pi_{i,t} = \pi^r_{i,t} + \Delta x_{i,t}, \tag{5}$$

where $\pi^r_{i,t} \equiv p^r_{i,t} - p^r_{i,t-1} = \mu_t + \varepsilon_{i,t}$ is the firm's *reset inflation* and $\Delta x_{i,t} \equiv x_{i,t} - x_{i,t-1}$ is the wedge introduced by adjustment frictions. Aggregating across firms, idiosyncratic shocks wash out and aggregate inflation reduces to

$$\pi_t = \mu_t + \Delta E[x_t], \tag{6}$$

where μ_t is aggregate reset inflation and $\Delta E[x_t]$ is the change in the mean gap.

Equation (6) reveals a key limitation of headline inflation as a summary of the economy's pricing state. Two economies can exhibit the same π_t while harboring very different gap distributions: one in which inflation arises from smooth, widespread small adjustments (low gap variance, prices close to their reset levels) and another in which the same aggregate number masks large adjustments by a few firms alongside many others whose prices remain stuck far from their reset levels. The two economies look identical through the lens of π_t, but differ sharply in their welfare costs (4), their vulnerability to subsequent shocks, and the directional pressure embedded in higher moments of G_t. The gap distribution thus provides

[6] As shown in Appendix A.3, this expression arises from a second-order approximation to the Ramsey planner's problem: the gap dispersion term captures misallocation of demand across varieties.

information about the state of the economy that headline inflation alone cannot reveal.

2.3 Why the gap distribution matters

We now turn to how we can use the information on the state of the economy that the gap distribution reveals. We show that it determines the non-linear transmission of aggregate shocks to inflation, and that it carries predictive content for future inflation beyond what the inflation history alone can provide.

2.3.1 Non-linearity and the role of initial conditions

In menu cost models, the inflation response to aggregate shocks is inherently non-linear. One well-understood dimension of this non-linearity is shock size: large shocks push more firms past their adjustment thresholds, so that the per-unit inflation response rises with the magnitude of the shock (Golosov and Lucas, 2007). We highlight a complementary dimension that has received less attention: the shape of the gap distribution at the time the shock arrives.

To illustrate both channels, we calibrate the microfounded model in Appendix A.1 to match three moments of UK CPI micro data reported by Blanco (2021): a monthly frequency of price change of 12.6%, a mean absolute price change of 15.3%, and a kurtosis of 3.8. Parameter values are reported in Table A.1 in the appendix. We then subject the economy to an unexpected, one-off level shock to the drift, $\mu_1 = \mu + \Delta\mu$ with $\mu_t = \mu$ for $t \geq 2$, and ask how the impact excess inflation $\pi_1 - \mu$ varies with the size of the shock and the state of G_0 when it hits.

Panel A of Figure 1 isolates the shock-size channel. Starting from the ergodic distribution, we vary the magnitude of the shock and plot the resulting impact excess inflation. The dashed line shows a linearized benchmark, extrapolated from the slope at small shocks. The relationship is convex: the impact response grows more than proportionally with shock size, reflecting the increasing share of firms pushed past their adjustment thresholds. For small shocks, however, the relationship is approximately linear: the model and the linearized benchmark nearly coincide. This is a well-known result in the literature.

Panel B turns to the less-studied dimension. We fix the shock at 6 pp annualized, still in the approximately linear region of Panel A, and vary only the initial gap distribution G_0. To generate economically meaningful variation in G_0, rather than constructing distributions arbitrarily, we subject the economy to a large persistent perturbation to the drift and record the gap distribution at successive dates as it transitions back to the steady state. Each snapshot provides a different initial condition for the shock. These distributions vary in location, dispersion, and tail behavior, reflecting the accumulated effects of the perturbation, while the structural parameters of the model remain identical throughout.

To isolate the genuine interaction between the shock and the initial conditions, we con-

struct the *net amplification ratio*. For each displaced G_0, we compute the impact excess inflation both with and without the shock, subtract the counterfactual to remove the mechanical unwinding of pent-up pressure that G_0 would have generated on its own, and normalize by the ergodic response. The ergodic point equals one by construction.

To summarize each distribution by a single number, we regress the net amplification ratio on each of the first four moments of G_0 separately. The mean gap $E[x_t]$ provides the best univariate fit ($R^2 = 0.93$); Panel B plots the net amplification ratio against this statistic. The relationship is monotone: a more negative mean gap amplifies the shock (ratio above one), while a more positive mean gap dampens it (ratio below one). The intuition is that a negative mean gap places more firms near the lower adjustment threshold, which is where a positive drift shock triggers repricing. The magnitudes are economically meaningful: a displacement of -5 percentage points in the mean gap, relative to the ergodic, raises the amplification ratio to about 1.06, a 6% increase in the inflation response. This variation arises even though the shock itself is small enough to generate no non-linearity in the size dimension (Panel A). With larger shocks, the two channels reinforce each other and the amplification increases.

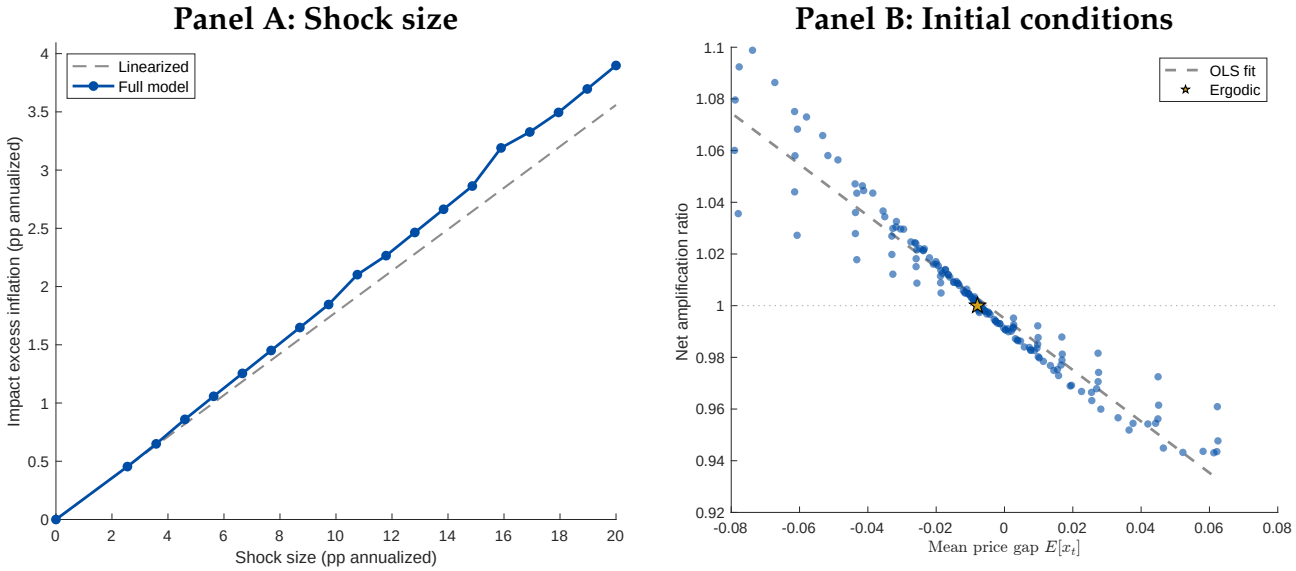
The displaced distributions in Panel B are generated by past perturbations to the drift μ. These primarily shift the mean of G_0 while leaving its shape relatively unchanged. If instead the displacement comes from past changes in idiosyncratic volatility σ_ε, the mean gap stays near zero but the variance changes substantially. In that case, the variance of G_0 becomes the relevant summary statistic: higher variance places more mass near the adjustment thresholds, increasing the economy's sensitivity to the shock. Figure A.1 in the appendix reports this exercise.

Implications for monetary policy. The non-linearity documented above has implications for the design of monetary policy. In a linearized New Keynesian model, the Phillips curve slope is constant and the welfare cost of a cost-push shock scales with the square of its magnitude. In the Ss model, the extensive margin (the additional firms pushed past their adjustment thresholds by larger shocks) makes the inflation response convex, so that the welfare cost of large shocks exceeds what a linearized model would predict. This is the mechanism underlying the "striking while the iron is hot" result of Karadi *et al.* (2025): since the marginal cost of inflation rises steeply with the shock, optimal policy calls for more aggressive intervention after large shocks, when the benefit of reducing inflation is greatest.

Our results add a complementary dimension to this logic. Even for a shock of fixed size, Panel B shows that the inflation response varies substantially with the initial gap distribution G_0. This means the welfare stakes of getting the policy response right depend not only on the size of the shock, as in Karadi *et al.* (2025), but also on the state of the gap distribution when the shock arrives. A central bank that monitors the gap distribution can in principle time and calibrate its response to exploit these state-dependent trade-offs.[7]

[7] A full characterization of optimal policy in this setting requires solving a Ramsey problem in which each candidate policy alters the drift, which changes the firm's pricing rule, which reshapes the distribution, a

Figure 1: Non-linearity in the inflation response



Notes: Panel A: starting from the ergodic distribution, we subject the economy to unexpected one-off level shocks of varying magnitude (x-axis) and plot the impact excess inflation $\pi_1 - \mu$ (y-axis). The dashed line extrapolates linearly from the slope at small shocks. Panel B: holding the shock fixed at 6 pp annualized, we vary the initial gap distribution G_0 by its mean price gap $E[x_t]$ (x-axis) and plot the net amplification ratio (y-axis). The net amplification ratio subtracts the no-shock counterfactual and normalizes by the ergodic response. The dashed line is the OLS fit. The star marks the ergodic distribution.

2.3.2 Predictive content of the gap distribution

The analysis in the preceding section was conditional: given a specific shock, the inflation response depends on the gap distribution at the time of impact. But the gap distribution matters for future inflation even without conditioning on any particular shock. In menu-cost models, inflation persistence is endogenous: whenever the gap distribution departs from its ergodic shape, whether due to past shocks, policy changes, or any other history, firms that find themselves near their adjustment thresholds will reprice in coming periods, generating inflation (or deflation) even in the absence of new disturbances. The pricing pressure stored in the current shape of G_t unwinds gradually as successive cohorts of firms reach their thresholds, producing a predictable component in future inflation.

A natural question is whether this information is already captured by the history of inflation itself, or whether the distribution carries additional predictive content. From equation (6), aggregate inflation is $\pi_t = \mu_t + \Delta E[x_t]$, a scalar projection of the infinite-dimensional state G_t. Since many distinct gap distributions can generate the same inflation rate, $\{\pi_s\}_{s \leq t}$ generically cannot recover G_t, and since the shape of G_t determines the future response of the price level to ongoing drift, this unrecovered information predicts future inflation. The gap distribution therefore Granger-causes inflation; a prediction we formalize below and test empirically in Section 5.

computationally demanding exercise. See Karadi *et al.* (2025) for a quantitative treatment in a calibrated menu cost model with general equilibrium. Formalizing how the state of the economy shapes the optimal policy problem is work in progress.

Consider the pure menu cost case, $\lambda = 0$, with negligible idiosyncratic volatility ($\sigma_\varepsilon \to 0$) and constant positive drift $\mu > 0$. Under these assumptions, each firm's gap drifts deterministically: $x_{i,t+1} = x_{i,t} - \mu$ in the absence of adjustment. A firm adjusts when its gap reaches the lower threshold $\underline{x}$, resetting to zero.

Proposition 1 *In the deterministic pure menu cost model:*

(i) The path of aggregate inflation is determined by the density g_t across the inaction region. For horizons $h = 1, 2, \ldots$,

$$\pi_{t+h} \;=\; |\underline{x}|\;\mu\; g_t\big(\underline{x} + (h-1)\mu\big) \;+\; O(\mu^2). \tag{7}$$

(ii) The history of aggregate inflation does not recover this information: for any window K shorter than the transit time across the inaction region, there exist gap distributions $G_t \neq G_t'$ that generate identical inflation histories $\pi_{t-k} = \pi'_{t-k}$ for $k = 0, \ldots, K$, yet differ in future inflation at some horizon $h \geq 1$.[8]

The proof is in Appendix A.5. The intuition for part (i) is that under constant drift the inaction region is partitioned into bands of width μ, each mapping to a future horizon: firms in the h-th band will reach the lower threshold and reset to zero h periods from now, each contributing a price increase of $|\underline{x}|$. The density thus acts as a schedule of upcoming price adjustments. For part (ii), inflation at any date reveals only the density near the threshold, while firms deeper in the interior—whose positions govern inflation at longer horizons—remain invisible to the inflation record.

In the general model ($\sigma_\varepsilon > 0$, $\lambda > 0$), threshold crossings are smoothed by idiosyncratic shocks and random free adjustments, but the core logic persists. The mass of firms *near* the boundaries of the inaction region governs the expected flow of future adjustments and is not recoverable from the inflation history alone. Indeed, the case for predictive content is arguably stronger in the stochastic model: idiosyncratic shocks continuously reshuffle the distribution, so that no sequence of aggregate inflation rates, each a scalar summary of G_t, can pin down the full infinite-dimensional state. The bound on K in Proposition 1 is an artifact of the deterministic limit.

A formal test of whether the gap distribution Granger-causes inflation requires taking the model to the data. To do so, we first need to recover the latent price gaps from observed prices. We turn to this in the next section.

[8]The transit time is $(\bar{x} - \underline{x})/\mu$: the number of periods for a firm to drift from the upper to the lower threshold. Under the calibration in Table A.1, this exceeds 500 months.

3 State-space methods for a random menu cost model

This section presents our methodology for estimating the random menu cost model using micro price data. We begin by casting the pricing dynamics of Section 2 in state-space form, then develop a Bayesian estimation procedure to jointly recover the model parameters and latent states. A Monte Carlo experiment assesses the finite-sample performance of the algorithm.

3.1 Random menu cost model in state-space form

To take the model of Section 2 to data, we cast it in state-space form. In the data, the unit of observation is a *quote-line*, the price of a specific product tracked at a specific outlet over time, grouped into *items*, fine product categories defined by the statistical office. For each quote-line i, we observe a sequence of prices $\{p_{i,t}\}$. From these, we seek to recover two latent objects: the reset price $p^r_{i,t}$ and an indicator $\ell_{i,t}$ for whether a free adjustment opportunity arrived. The *Ss* rule (3) provides the link between observables and latent states.

Measurement equation. The two-sided *Ss* pricing rule implies that the observed price is either unchanged (under inaction) or equal to the reset price (upon adjustment):

$$p_{i,t} = p_{i,t-1}\, d(\mathbf{x}_{i,t}, p_{i,t-1}, \boldsymbol{\theta}) + p^r_{i,t}\ (1 - d(\mathbf{x}_{i,t}, p_{i,t-1}, \boldsymbol{\theta}))\,, \tag{8}$$

where $d(\mathbf{x}_{i,t}, p_{i,t-1}, \boldsymbol{\theta})$ equals one when inaction is optimal and zero otherwise:

$$d(\mathbf{x}_{i,t}, p_{i,t-1}, \boldsymbol{\theta}) = \mathbb{1}\{p_{i,t-1} - p^r_{i,t} \in [\underline{s}, \bar{s}]\}(1 - \ell_{i,t}) + \mathbb{1}\{p_{i,t} = p^r_{i,t}\}\, \ell_{i,t}. \tag{9}$$

The first term captures inaction when no free adjustment opportunity arrives and the price gap lies within the inaction region. The second captures inaction when a free adjustment opportunity arrives but the gap is already closed.

Transition equations. The reset price follows the random walk with drift in (1):

$$p^r_{i,t} = \mu + p^r_{i,t-1} + \varepsilon_{i,t}, \tag{10}$$

where $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$, independently across quote-lines and time. Free adjustment opportunities arrive independently with probability λ:

$$\ell_{i,t} = \mathbb{1}\{\nu_{i,t} \leqslant \lambda\}, \tag{11}$$

where $\nu_{i,t} \sim \text{Uniform}(0, 1)$, independently across quote-lines and time.

Initial conditions. The initial latent states are independent across quote-lines:

$$f(\mathbf{x}_{1:n,1} \mid \mathbf{p}_{1:n,1}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(\mathbf{x}_{i,1} \mid p_{i,1}, \boldsymbol{\theta}). \tag{12}$$

We specify the functional form for $f(\mathbf{x}_{i,1} \mid p_{i,1}, \boldsymbol{\theta})$ in Section 3.2.2.

Summary. Equations (8)–(12) define a nonlinear, non-Gaussian state-space system for the latent state vector $\mathbf{x}_{i,t} \equiv [p^r_{i,t}, \ell_{i,t}]$, governed by the parameter vector $\boldsymbol{\theta} = \{\mu, \sigma_\varepsilon, \lambda, \underline{s}, \bar{s}\}$. The representation is semi-structural: the measurement equation follows directly from the optimal *Ss* rule, whereas the threshold parameters $\{\underline{s}, \bar{s}\}$ are reduced-form objects, each a function of deeper structural primitives such as preferences, technology, and menu costs. We estimate the model separately for each item, with $\boldsymbol{\theta}$ common across quote-lines within an item and constant within each item-period estimation window.[9]

3.2 Bayesian estimation

We now turn to the Bayesian estimation procedure. We first describe how to sample the latent states conditional on the model parameters and observed data, then use this conditional sampling step to construct a Gibbs sampler for the joint posterior.

3.2.1 Sampling the latent states

To sample the latent states conditional on parameters and observed prices, we use a forward-filtering backward-sampling (FFBS) algorithm. This step is challenging because the state-space representation is nonlinear and non-Gaussian and the panel contains a large number of price trajectories.

A key observation is that, under the two-sided *Ss* rule and the assumed initial and transition distributions, the latent-state posterior factorizes across price trajectories. The result is formalized in the following lemma.

Lemma 1 (FFBS). *Given the state-space representation in* (8)–(12)*, the forward-filtering backward-sampling recursions factorize across price trajectories.*

Forward filtering. *The filtered and one-step-ahead predictive densities satisfy*

$$f\left(\mathbf{x}_{1:n,t} \mid \mathbf{p}_{1:n,1:t}, \boldsymbol{\theta}\right) = \prod_{i=1}^{n} f\left(\mathbf{x}_{i,t} \mid \mathbf{p}_{i,1:t}, \boldsymbol{\theta}\right), \tag{13}$$

$$f\left(\mathbf{x}_{1:n,t} \mid \mathbf{p}_{1:n,1:t-1}, \boldsymbol{\theta}\right) = \prod_{i=1}^{n} f\left(\mathbf{x}_{i,t} \mid \mathbf{p}_{i,1:t-1}, \boldsymbol{\theta}\right), \tag{14}$$

[9] Section 4 discusses the choice of estimation windows and how they are adapted to different sample periods.

with trajectory-level recursions given by:

$$f\left(\mathbf{x}_{i,t} \mid \mathbf{p}_{i,1:t-1}, \boldsymbol{\theta}\right) = \int f\left(\mathbf{x}_{i,t} \mid \mathbf{x}_{i,t-1}, \boldsymbol{\theta}\right) f\left(\mathbf{x}_{i,t-1} \mid \mathbf{p}_{i,1:t-1}, \boldsymbol{\theta}\right) \, d\mathbf{x}_{i,t-1}, \tag{15}$$

$$f\left(\mathbf{x}_{i,t} \mid \mathbf{p}_{i,1:t}, \boldsymbol{\theta}\right) \propto f\left(p_{i,t} \mid \mathbf{x}_{i,t}, p_{i,t-1}, \boldsymbol{\theta}\right) f\left(\mathbf{x}_{i,t} \mid \mathbf{p}_{i,1:t-1}, \boldsymbol{\theta}\right). \tag{16}$$

Backward sampling. *The posterior over latent state trajectories also factorizes:*

$$f\left(\mathbf{x}_{1:n,1:T} \mid \mathbf{p}_{1:n,1:T}, \boldsymbol{\theta}\right) = \prod_{i=1}^{n} f\left(\mathbf{x}_{i,1:T} \mid \mathbf{p}_{i,1:T}, \boldsymbol{\theta}\right), \tag{17}$$

and each trajectory is sampled according to the backward recursion:

$$f\left(\mathbf{x}_{i,1:T} \mid \mathbf{p}_{i,1:T}, \boldsymbol{\theta}\right) = f\left(\mathbf{x}_{i,T} \mid \mathbf{p}_{i,1:T}, \boldsymbol{\theta}\right) \prod_{t=1}^{T-1} f\left(\mathbf{x}_{i,t} \mid \mathbf{x}_{i,t+1}, \mathbf{p}_{i,1:t}, \boldsymbol{\theta}\right), \tag{18}$$

$$f\left(\mathbf{x}_{i,t} \mid \mathbf{x}_{i,t+1}, \mathbf{p}_{i,1:t}, \boldsymbol{\theta}\right) \propto f\left(\mathbf{x}_{i,t+1} \mid \mathbf{x}_{i,t}, \boldsymbol{\theta}\right) f\left(\mathbf{x}_{i,t} \mid \mathbf{p}_{i,1:t}, \boldsymbol{\theta}\right). \tag{19}$$

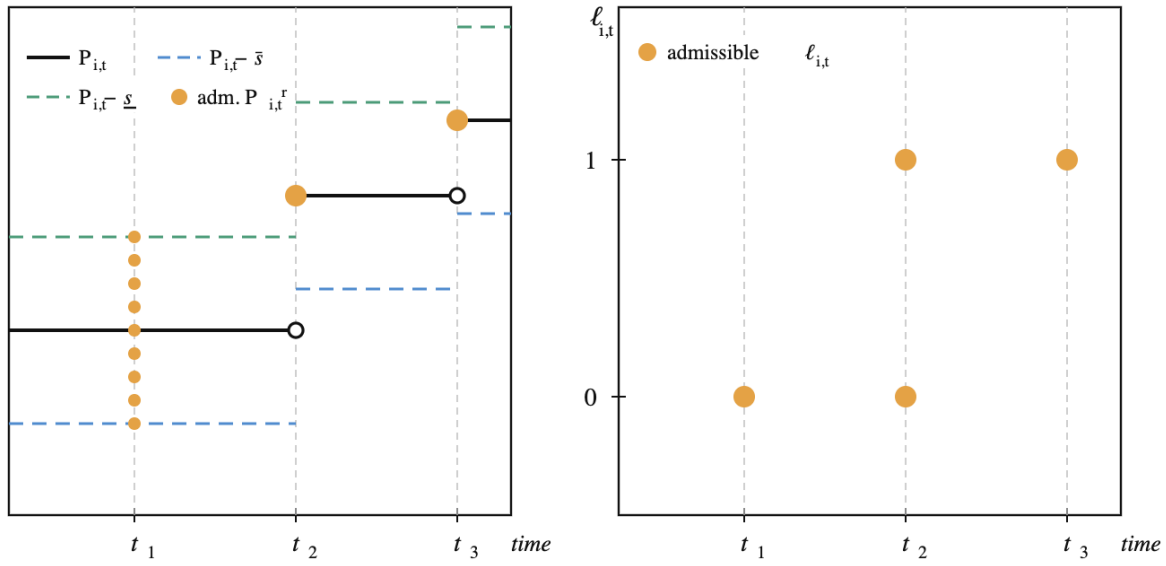
The proof is in Appendix B.1. The joint filtered, predictive, and smoothed densities all factorize into products of trajectory-level densities. This follows from two features of the model. Under the *Ss* rule (8)–(9), each observed price depends only on its own lagged price, latent state, and the model parameters, so the measurement equation links observables to latent states trajectory by trajectory. And under (10)–(12), latent states are initialized and evolve independently across quote-lines. The key computational implication is that latent-state sampling can proceed trajectory by trajectory rather than jointly over the full panel, reducing the problem to a collection of low-dimensional systems that can be processed in parallel.

Given Lemma 1, latent-state sampling reduces to solving the trajectory-level FFBS recursions in (15), (16), (18), and (19). To do so, we build on numerical methods for nonlinear, non-Gaussian state-space systems (Kitagawa, 1987; Kramer and Sorenson, 1988) to develop a grid-based FFBS procedure. The grid-based approach is particularly well suited to our setting because the trajectory-level state space is only two-dimensional, and the absence of measurement error under the two-sided *Ss* rule makes observed price changes highly informative about admissible latent states over time. Together, these features sharply restrict the relevant support for the filtering and smoothing distributions, making grid-based approximation both accurate and computationally efficient.

Figure 2 illustrates by showing how a hypothetical price trajectory restricts the set of admissible latent states over time. Depending on the observed price change, the admissible sets take three forms. Under inaction, as at t_1, the latent reset price must lie inside the inaction region, so $p^r_{i,t} \in [p_{i,t} - \bar{s},\, p_{i,t} - \underline{s}]$ and $\ell_{i,t} = 0$.[10] If there is an adjustment from outside

[10] We ignore the case of a free-adjustment opportunity when the reset price coincides exactly with the pre-

Figure 2: Admissible latent states given an observed price trajectory



Notes: The black solid line shows a hypothetical observed price trajectory. Panel (a): shaded regions indicate the admissible values of the latent reset price $p^r_{i,t}$ at each date, determined by the inaction region and the observed price change. Panel (b): admissible values of the free-adjustment indicator $\ell_{i,t}$. Under inaction (t_1), $\ell_{i,t} = 0$ and $p^r_{i,t}$ lies in the inaction band. Under adjustment from outside the inaction region (t_2), $p^r_{i,t} = p_{i,t}$ but $\ell_{i,t}$ is not identified. Under adjustment from inside the inaction region (t_3), $p^r_{i,t} = p_{i,t}$ and $\ell_{i,t} = 1$.

the inaction region, as at t_2, the reset price must coincide with the observed price, $p^r_{i,t} = p_{i,t}$, while the free-adjustment indicator is not identified, so both $\ell_{i,t} = 0$ and $\ell_{i,t} = 1$ are admissible. Finally, if there is an adjustment from within the inaction region, as at t_3, the reset price again coincides with the observed price, but in this case the adjustment must reflect a free opportunity, so $\ell_{i,t} = 1$. We use this classification to construct date-specific grids as a function of the observed price change $\Delta p_{i,t}$, and solve the trajectory-level FFBS recursions over the resulting sequence of grids. Appendix B.2 provides the details.

3.2.2 Gibbs sampler for the joint posterior

Building on the latent-state sampler described above, we construct a Gibbs sampler for the joint posterior of the model parameters and latent states. Table 1 reports the prior distributions.

We choose prior functional forms to obtain conjugate full conditionals wherever possible, with weakly informative hyperparameters. Two components of the prior are tied more directly to the data. First, we bound the prior support for the inaction bands using the range of observed price changes. The likelihood identifies the bands only in samples containing at least one price change from inside and one from outside the inaction region on each side of the price-change distribution, so the prior rules out regions of the parameter space that

vious observed price, since under a continuous distribution for the reset price this is a probability-zero event.

Table 1: Prior distributions for parameters and initial states

Panel A: Model parameters

μ	Normal$(m_\mu^0,\, v_\mu^0)$	$m_\mu^0 = 0,\ v_\mu^0 = 1$
$1/\sigma_\varepsilon^2$	Gamma$(s_\sigma^0,\, r_\sigma^0)$	$s_\sigma^0 = r_\sigma^0 = 10^{-3}$
λ	Beta$(s_{1,\lambda}^0,\, s_{2,\lambda}^0)$	$s_{1,\lambda}^0 = s_{2,\lambda}^0 = 1$
$\underline{s}$	Uniform $\left(\underline{s}_{\text{LB}}^0,\, \underline{s}_{\text{UB}}^0\right)$	$\underline{s}_{\text{LB}}^0 = \min -\Delta\mathbf{p}^+ \,, \underline{s}_{\text{UB}}^0 = \max -\Delta\mathbf{p}^+$
$\bar{s}$	Uniform $\left(\bar{s}_{\text{LB}}^0,\, \bar{s}_{\text{UB}}^0\right)$	$\bar{s}_{\text{LB}}^0 = \min -\Delta\mathbf{p}^- \,, \bar{s}_{\text{UB}}^0 = \max -\Delta\mathbf{p}^-$

Panel B: Initial latent states

$\mathbf{x}_{i,1} \mid p_{i,1}, \boldsymbol{\theta}$	$(p_{i,1}, 1)$ w/ probability $\kappa_0\, \omega_0$	$\kappa_0 = \text{freq}(\Delta\mathbf{p}_t)$
	$(p_{i,1}, 0)$ w/ probability $\kappa_0(1-\omega_0)$	$\omega_0 = 0.6 \times \kappa_0$
	$(p_{i,1}^r, 0),\ \ p_{i,1}^r \sim \text{Uniform}(I_{i,1}),\ \ \text{w/ probability } 1-\kappa_0$	

Notes: For the Gamma prior, s_σ^0 and r_σ^0 denote the shape and rate parameters, respectively. $\Delta\mathbf{p}$ denotes the set of all observed price changes, while $\Delta\mathbf{p}^+$ and $\Delta\mathbf{p}^-$ denote the subsets of positive and negative price changes. The initial state is $\mathbf{x}_{i,1} = (p_{i,1}^r, \ell_{i,1})$. The parameter κ_0 governs the probability that the initial price gap is closed, and ω_0 the conditional probability that, given a closed initial gap, the closure reflects a free-adjustment opportunity. Finally, $I_{i,1} \equiv [p_{i,1} - \bar{s},\ p_{i,1} - \underline{s}]$ denotes the inaction region for the initial reset price.

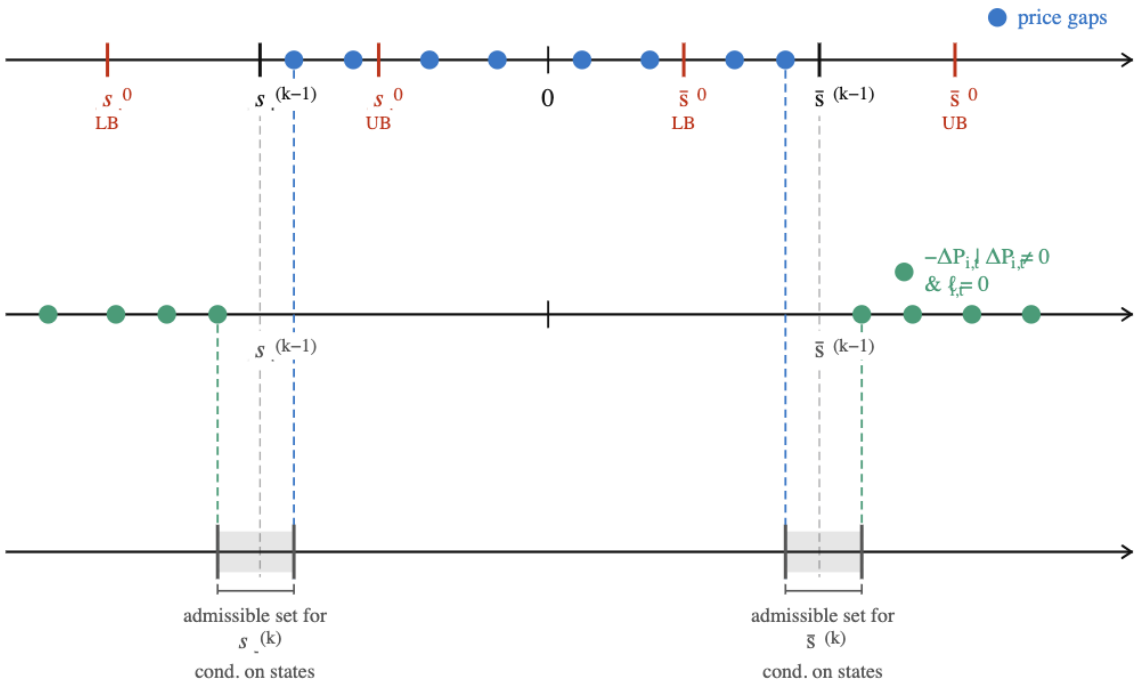
the data cannot identify.[11] Second, the prior for the initial latent states reflects the fact that we do not observe whether the first price in each quote-line results from an adjustment. We set the probability of a closed initial gap equal to the empirical frequency of price changes, and conditional on closure, assign a 60% probability that it reflects a free adjustment, in line with the average Calvoness estimated by Gautier and Le Bihan (2022) for French micro price data.

Given these priors, the full conditionals for μ, σ_ε^2, and λ belong to standard conjugate families: Normal-Inverse-Gamma for $(\mu, \sigma_\varepsilon^2)$ and Beta-Bernoulli for λ. The remaining unknowns, the latent states and the inaction thresholds, do not admit closed-form full conditionals. We draw them jointly using a Metropolis-Hastings step within the Gibbs, motivated by the tight link the *Ss* rule imposes between admissible thresholds and admissible state trajectories, as Figure 3 illustrates.

At each iteration, conditioning on the current states restricts the admissible thresholds, and vice versa, so separate updates lead to poor mixing. We propose latent states from their smoothed conditional (approximated via the grid-based FFBS procedure above) and inaction thresholds from truncated normal random walks centered at the previous draw, with standard deviations tuned during burn-in via a Robbins–Monro stochastic approximation

[11] A natural concern is what happens when this condition fails and the prior support excludes the true band values. Appendix B.4 reports a Monte Carlo exercise based on two extreme data-generating processes. In those cases, our prior choice induces bias in the estimates of the bands and of λ, but the drift, the innovation standard deviation, the latent reset prices, and the moments of the price-gap distribution remain well estimated.

Figure 3: Thresholds update conditional on latent states



Notes: The figure illustrates the mutual dependence between the inaction thresholds and the latent states. Conditioning on the current latent-state trajectories restricts the set of admissible threshold values, since alternative values would imply zero likelihood for some observed price changes. Conversely, conditioning on the current thresholds restricts the admissible latent-state trajectories.

(Robbins and Monro, 1951) targeting a 33% acceptance rate (Roberts and Rosenthal, 2001). Appendix B.3 provides the details.

3.3 Monte Carlo

We assess the finite-sample properties of the estimator using a Monte Carlo experiment. To obtain empirically plausible parameter values for the data-generating process, we calibrate the model to item-level moments from the UK CPI micro data described in Section 4, targeting the frequency of price changes, the mean absolute price change, and the kurtosis of price changes at the median level of Calvoness across items.[12] We consider two sample sizes that correspond approximately to the 10th and 20th percentiles of the panel dimensions in the UK data: 100 quote-lines observed over 24 months (small sample) and 100 quote-lines over 60 months (large sample). For each, we generate 200 replications, estimate the model using the Gibbs sampler, and evaluate posterior mean bias, posterior standard deviation, and coverage of the 95% credible interval. Table 2 reports the results.

[12] Appendix B.4 provides a detailed description of the calibration procedure.

Table 2: Monte Carlo results

Panel A: Parameters

	Bias		Standard Deviation		95% Coverage	
	Small Sample	Big Sample	Small Sample	Big Sample	Small Sample	Big Sample
$\underline{s}$	-0.00	0.00	0.13	0.05	94.80	97.20
$\bar{s}$	-0.01	-0.01	0.14	0.05	96.00	94.80
μ	-0.01	-0.01	0.13	0.08	96.40	92.40
σ_e	0.02	0.01	0.14	0.09	94.80	95.60
λ	-0.18	-0.06	0.82	0.52	97.20	94.00

Panel B: Price Gaps

	Bias		Standard Deviation		95% Coverage	
	Small Sample	Big Sample	Small Sample	Big Sample	Small Sample	Big Sample
Ind	0.03	-0.01	5.55	5.47	95.04	95.02
Mean	0.02	-0.01	0.65	0.64	94.45	95.02
Std	-0.00	-0.00	0.43	0.42	94.52	94.53
Skew	-0.00	-0.00	0.16	0.16	95.44	95.37
Kurt	0.05	0.04	0.23	0.23	94.71	95.28

Notes: Entries report Monte Carlo averages across 200 replications. Bias is defined as the difference between the posterior mean and the true parameter value. Standard deviation refers to the posterior standard deviation. Coverage reports the fraction of replications in which the true value lies within the 95% posterior credible interval. Bias and standard deviation are expressed in percentage points, except for skewness and kurtosis, which are scale-free moments. Coverage is reported in percent. The small and large-sample designs correspond to $(N, T) = (100, 24)$ and $(N, T) = (100, 60)$, respectively.

Across both sample sizes, posterior mean bias is negligible for all structural parameters and coverage of the 95% credible intervals is close to the nominal level (Panel A). The cross-sectional moments of the gap distribution that matter for the aggregate analysis (mean, standard deviation, skewness, kurtosis) are similarly well recovered (Panel B), even though individual gap trajectories carry more posterior uncertainty. Appendix B.4 presents additional experiments under varying degrees of Calvoness, pure menu-cost and pure Calvo models, and alternative algorithmic settings, confirming the stability of the procedure across a wide range of environments.

4 Application to UK micro-price data

This section puts the algorithm to work using micro-level price data from the UK. We begin by describing the dataset and the cleaning procedure. We then present our estimation results and show that the estimated parameters align with those reported in the literature.

Finally, we discuss the estimated distribution of price gaps and document its time variation.

4.1 Data description

We use publicly available data on locally collected price quotes underlying the construction of the UK Consumer Price Index (CPI). In order to produce the CPI, the Office for National Statistics (ONS) collects monthly prices for a broad set of goods and services selected to represent household expenditure across the UK. Price collection follows two methods: central or local. Central collection is used for items with uniform pricing nationwide or where regional variation can be captured remotely, for example through internet, telephone, or email. To protect confidentiality, centrally collected prices that could reveal the identity of the price setter are excluded from the public release. The remaining items, accounting for roughly 60% of the aggregate CPI by weight, are collected in person by ONS field agents, who visit thousands of retail outlets across more than 140 UK locations each month, recording over 100,000 prices using handheld devices. Our baseline sample consists of these locally collected price quotes from 1996m1 to 2023m12.[13]

Our cleaning procedure has two stages. In the first stage, we apply three filters: we exclude price quotes that did not enter the actual CPI calculation because they failed the ONS internal validation procedures; we remove quotes referring to research items that were not part of the official CPI; and we exclude any quotes that cannot be uniquely identified by the combination of month, shop code, region, and item identifier.[14] In the second stage, we clean the raw data to obtain quote-lines of regular prices: we remove product substitutions, impute regular prices during sale spells, and split quote-lines at gaps to ensure contiguous observations. Details on these procedures are provided in Appendix C.1. Our final sample includes over 23 million price quote observations, more than 2.2 million quote-lines, and 979 unique items.

4.2 Posterior estimates: model parameters

As a first pass, model parameters are estimated separately for each item and held constant over the full sample period. The assumption of time-invariant parameters is a reasonable starting point, particularly for the period of moderate and relatively stable inflation that preceded the pandemic; in Section 5, we discuss how it can be relaxed through time-varying estimation. Table 3 reports summary statistics of the posterior distributions for each parameter across items, and a graphical representation of the cross-item distributions is provided in Figure A.2 in the appendix.

[13] Bunn and Ellis (2012) were the first to document stylized facts from this dataset, which has since been widely used in the pricing literature. Since this is not the first paper to use the data, we focus on the features most relevant for constructing our final sample.

[14] The validation procedures are described in ONS (2014, Chapter 6). In the ONS internal system, quote-lines are uniquely identified by month, shop code, location, and item identifier. Because the published data omit the location variable for confidentiality reasons, we use region as a proxy. In rare cases, two outlets in the same region share a shop code, making their price trajectories indistinguishable.

Table 3: Summary statistics for estimated parameters

	$\hat{\underline{x}}$	$\hat{\bar{x}}$	$\hat{\mu}$	$\hat{\sigma}_\varepsilon$	$\hat{\lambda}$
Mean	-2.691	2.271	0.001	0.09	0.139
Median	-0.596	0.568	0.002	0.072	0.114
Std Dev	8.633	6.676	0.011	0.06	0.109
IQR	1.87	1.466	0.004	0.064	0.092
Minimum	-130.84	0	-0.137	0.006	0.001
Maximum	0	130.262	0.058	0.4	0.868
N	979	979	979	979	979

Notes: Summary statistics are computed across items using the posterior means for each estimated parameter. All values are rounded to three decimal places. Numbers smaller than 5×10^{-4} are displayed as zero.

The estimates reveal substantial variation across items. A key dimension of interest is the asymmetry of the inaction region, which we measure by $\mathcal{A} = \hat{\bar{x}} + \hat{\underline{x}}$. This statistic ranges from -171% to 206%, with 58% of items displaying negative asymmetry.[15] Consistent with menu cost models under positive trend inflation (Ball and Mankiw, 1994), Panel (a) of Figure 4 shows that the asymmetry correlates negatively with the estimated drift, with a correlation of -0.41.

Estimated reset inflation trends $\hat{\mu}$ range from -14.7% to 5.8% per month, with a median of 0.2%, consistent with average monthly CPI inflation of 0.17% over the sample period. As Panel (b) of Figure 4 shows, these trends correlate positively with the average size of price changes, with a correlation of 0.63. Roughly one-quarter of items display negative trends, concentrated in categories such as recreation and culture, household equipment, and clothing and footwear.[16]

As Panel (c) of Figure 4 shows, differences in estimated idiosyncratic volatility $\hat{\sigma}_\varepsilon$ drive most of the cross-sectional variation in the dispersion of price changes, with a correlation of 0.86. The median estimated volatility is 7.2%, consistent with the 5–9% range reported by Gautier and Bihan (2018) for French CPI data.

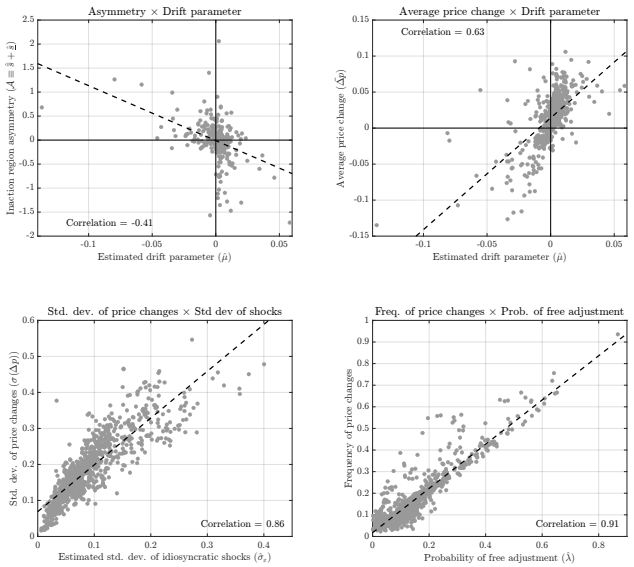
As Panel (d) of Figure 4 shows, the free-adjustment probability $\hat{\lambda}$ accounts for much of the heterogeneity in price change frequency, with a correlation of 0.91. Overall, 88% of price changes in the sample occur inside the inaction region, indicating adjustment triggered by costless opportunities rather than threshold crossings.[17] This is in line with prior estimates: Nakamura and Steinsson (2010) report 75% for US data, Gautier and Bihan (2018) 80% for

[15] The asymmetry measure is computed only for items that experience at least one price increase and one price decrease triggered by crossing each respective boundary (total of 424 items).

[16] Examples include personal CD players, cassette radios, washing machines, vacuum cleaners, and boys' shirts.

[17] Formally, for each item category, this is computed as $\mathcal{C} = \sum_i \sum_t \mathbb{1}\{\Delta p_{i,t} \in (-\hat{\bar{x}}, 0) \cup (0, -\hat{\underline{x}})\} \,/\, \sum_i \sum_t \mathbb{1}\{\Delta p_{i,t} \neq 0\}$.

Figure 4: Pricing moments and reduced form parameters



Notes: In all the scatter plots, each grey dot represents a given 6-digit item in the sample. In the first scatter plot only items for which there is at least one price change triggered by crossing the upper and the lower boundaries of the inaction region are considered (total of 424 items). In all the scatter plots, the dashed black line is the best fit line obtained from a bivariate regression of the variable on the y-axis on the variable on the x-axis.

France, and Blanco (2021) 93% for the UK when including fat-tailed shocks.

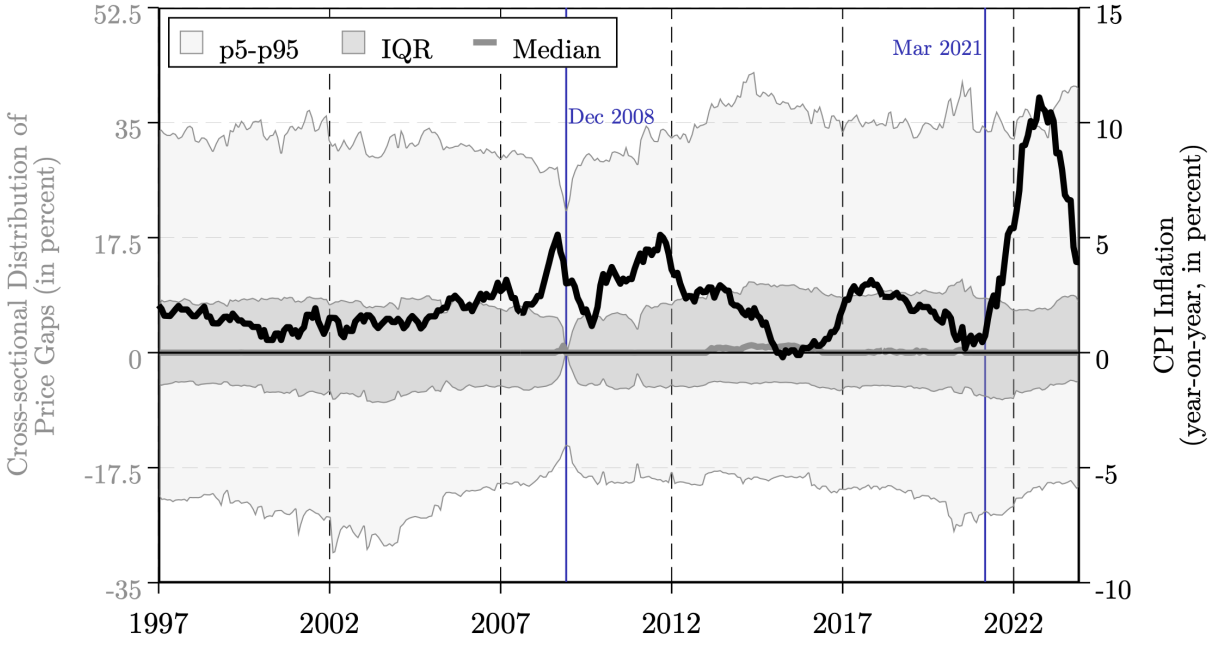
4.3 The estimated distribution of price gaps

The Bayesian algorithm recovers the full sequence of latent states separately for each quote-line, yielding an estimated gap $x_{i,t} = p_{i,t} - p^r_{i,t}$ for each quote-line i and month t. Since the gap resets to zero upon adjustment for every quote-line, the estimated gaps are directly comparable across items and can be pooled to construct an aggregate distribution. We weight each quote-line by its CPI weight, so that the aggregate distribution is representative of the price index.[18] We denote this aggregate distribution G_t throughout. Panel (a) of Figure 5 plots the evolution of G_t between January 1997 and December 2023 alongside published year-on-year CPI inflation. The shaded regions report the interquartile range and the 5th-to-95th percentile range of the monthly distribution, while the thick gray line reports its median.

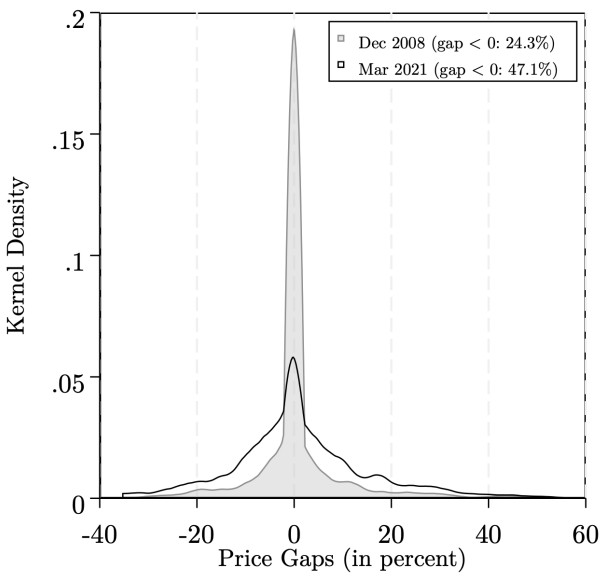
Figure 5 reveals substantial time variation in the shape of the gap distribution. Year-on-year inflation ranges from -0.2 to 11 percent over the sample, but the underlying distribution of price gaps varies even more dramatically: the median gap ranges from 0 to 1.15 percentage points, the interquartile range from 0 to 17 percentage points, and the 5th-to-95th percentile range from 36 to 68 percentage points. Characterizing the gap distribution through its ergodic counterpart alone would miss this variation entirely: the moments of G_t shift over time and take time to revert as firms gradually adjust their prices.

[18] See Appendix C.2 for details on the weight construction.

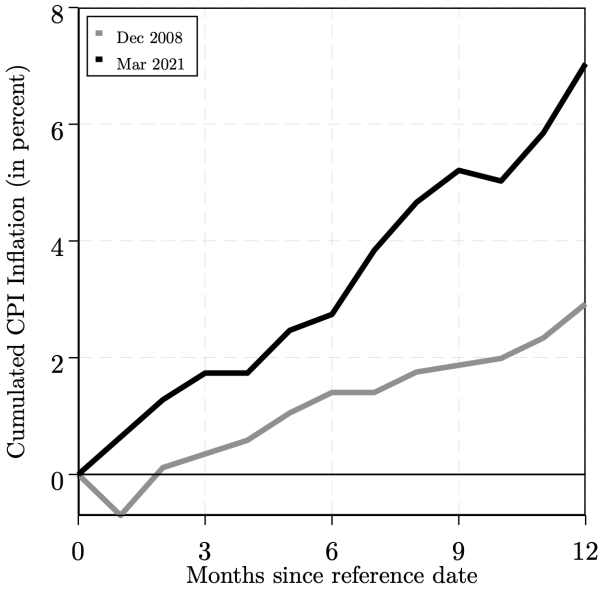
Figure 5: Price gaps and inflation over time



(a) Moments of the Price-Gap Distribution and Inflation



(b) Price-Gap Distributions



(c) Cumulative Inflation

Notes: Each quote-line is weighted by its CPI weight (Appendix C.2). Panel (a): the thick gray line reports the median of the weighted cross-sectional price-gap distribution; dark and light shaded regions show the interquartile and 5th–95th percentile ranges, respectively (left axis). The black line is published year-on-year CPI inflation (source: ONS; right axis). Vertical dashed lines mark December 2008 and March 2021. Panel (b): kernel density estimates of the weighted price-gap distribution at the two marked dates. Panel (c): cumulative inflation over the 12 months following each date.

Episodes in which the distribution widens tend to precede large movements in inflation. Panels (b) and (c) zoom into two contrasting dates. In December 2008, following the temporary VAT cut during the financial crisis, the distribution was tightly concentrated around zero, consistent with a synchronized wave of price adjustments. In March 2021, amid pandemic-era supply disruptions, the distribution had substantially wider tails, with nearly 48 percent of price setters exhibiting negative gaps. As Panel (c) shows, cumulative inflation over the following 12 months was nearly three times as large after March 2021 as after December 2008. This comparison is descriptive, but it illustrates the information embedded in the shape of G_t that headline inflation alone does not capture. In the next section, we test more formally whether the gap distribution shapes shock propagation and helps forecast inflation.

5 Price gaps in action

The previous section established that the distribution of price gaps varies substantially over time. Does this variation matter? We address this question using the estimated price gap distribution in three exercises: a historical decomposition of post-Covid inflation into fundamentals and frictions (work in progress), a test of state-dependent shock propagation, and an assessment of the gap distribution's predictive content for future inflation.

5.1 State-dependent shock propagation

The theory predicts that the same aggregate shock can generate very different inflation responses depending on when it hits. The distribution of price gaps summarizes the state of the economy right before the shock arrival. As discussed in Section 2.3.1, when many firms sit near their adjustment thresholds, even a moderate shock to μ triggers a wave of price changes; when the distribution is concentrated near the reset point, the same shock passes largely unnoticed. We test this prediction by studying the transmission of monetary policy shocks, for which credible identification is available, using state-dependent local projections in the spirit of Jordà (2005).

For each product class c and horizon h, we estimate

$$\log P^c_{t+h} = \beta_h \Delta i_t + \sum_{j=1}^{4} \gamma_{j,h}\, \Delta i_t\, m^c_{j,t-1} + \mathbf{w}_t' \Gamma_h + e_{t+h}, \tag{20}$$

where P^c_{t+h} is the published price index for class c, Δi_t is the change in the one-year gilt yield, and $m^c_{j,t-1}$ denotes the four lagged moments of the estimated price gap distribution for that class: the weighted mean, standard deviation, skewness, and kurtosis. The control vector $\mathbf{w}_t$ includes class fixed effects, class-specific linear trends, lags of the dependent variable, and additional macroeconomic controls standard in UK monetary SVARs.[19] The interaction

[19] Specifically, we include log GDP, the log exchange rate, corporate bond spreads, unemployment, and the

terms $\gamma_{j,h}$ are the objects of interest: jointly significant coefficients would indicate that the gap distribution shapes shock propagation, as the theory predicts.

Identification relies on high-frequency surprises around Bank of England monetary policy announcements. We use the shock series estimated by Braun *et al.* (2025), who decompose movements in asset prices within narrow windows around each announcement into three components: a target shock capturing unexpected changes in the current policy rate, a path shock capturing revisions to the expected future path of rates, and a QE shock capturing surprise changes in the scale of asset purchases. We use these three shocks as instruments for Δi_t, and their products with the lagged gap moments as instruments for the corresponding interaction terms.

Table 4 reports the estimated coefficients. Skewness stands out among the individual moments: its interaction with the monetary policy shock is positive and significant at the 5 percent level at nearly all horizons. A more negatively skewed gap distribution, with a thicker left tail of firms whose prices have fallen behind, amplifies the inflationary effect of an expansionary shock. In the *Ss* model, skewness captures the balance between inflationary and deflationary pressure stored in the tails of the distribution, and firms closest to their adjustment thresholds are the most likely to reprice when a shock arrives. The kurtosis interaction is negative and significant at the 3-month horizon, while the mean and standard deviation do not enter significantly on their own. Most importantly, the joint test rejects the null that all four interaction coefficients are zero at most horizons, confirming that the gap distribution shapes the transmission of monetary policy shocks.

To gauge the quantitative implications, we evaluate the estimated impulse responses under the two contrasting gap distributions documented in Figure 5: December 2008 and March 2021. Figure 6 reports the implied cumulative price response to a 100 basis point decline in the one-year gilt yield under each set of initial conditions. The contrast is stark. Under March 2021 conditions, the response is significant on impact and from horizon 9 onward, building to a cumulative effect more than twice as large as under December 2008 conditions at 12 months. Under December 2008 conditions, the response is imprecisely estimated throughout, consistent with a concentrated distribution dampening the transmission of aggregate shocks.

These results confirm that the distribution of price gaps is a source of state dependence in the transmission of monetary policy to inflation. The same expansionary shock that produces a large inflationary response when the gap distribution is dispersed and asymmetric has a negligible effect when the distribution is concentrated near the reset point. More broadly, the gap distribution can serve as a real-time indicator of the economy's fragility to new shocks: a dispersed, asymmetric distribution signals that latent pricing pressure has built up and that even a moderate shock can trigger a significant inflationary response. The same features that generate this state dependence should also contain information about

log FTSE as in Cesa-Bianchi *et al.* (2019) and Braun *et al.* (2025).

Table 4: State-dependent transmission of monetary policy shocks

	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Δi_t	-0.063	1.812	-0.709	-0.363	0.394
	(0.526)	(2.493)	(1.565)	(1.554)	(2.149)
$\Delta i_t \times$ Mean	-0.007	0.075	-0.088	0.027	0.094
	(0.071)	(0.075)	(0.091)	(0.084)	(0.075)
$\Delta i_t \times$ Std. Dev.	-0.085**	-0.226	-0.122	-0.149	-0.306
	(0.038)	(0.213)	(0.145)	(0.134)	(0.192)
$\Delta i_t \times$ Skewness	0.689**	0.878***	1.360*	1.292**	1.055**
	(0.337)	(0.318)	(0.723)	(0.633)	(0.415)
$\Delta i_t \times$ Kurtosis	-0.003	-0.017**	-0.020	-0.022	-0.029
	(0.004)	(0.008)	(0.015)	(0.018)	(0.019)
Panel FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Joint F-stat	6.525	3.901	1.322	3.175	3.277
	[0.000]	[0.006]	[0.270]	[0.019]	[0.016]

Notes: Each column reports IV estimates from the state-dependent local projection at horizon h (in months). The dependent variable is the log published price index at $t + h$. Δi_t is the change in the one-year gilt yield, instrumented using the target, path, and QE shocks from Braun *et al.* (2025). Mean, Std. Dev., Skewness, and Kurtosis refer to the lagged moments of the estimated price gap distribution for each product class. Controls include class fixed effects, class-specific linear trends, 12 lags of the dependent variable, and macroeconomic controls (log GDP, log exchange rate, corporate bond spreads, unemployment, log FTSE). Standard errors clustered by product class in parentheses. Joint F-statistic tests H_0: all four interaction coefficients are zero; p-values in brackets. *, **, *** denote significance at the 10%, 5%, and 1% levels.

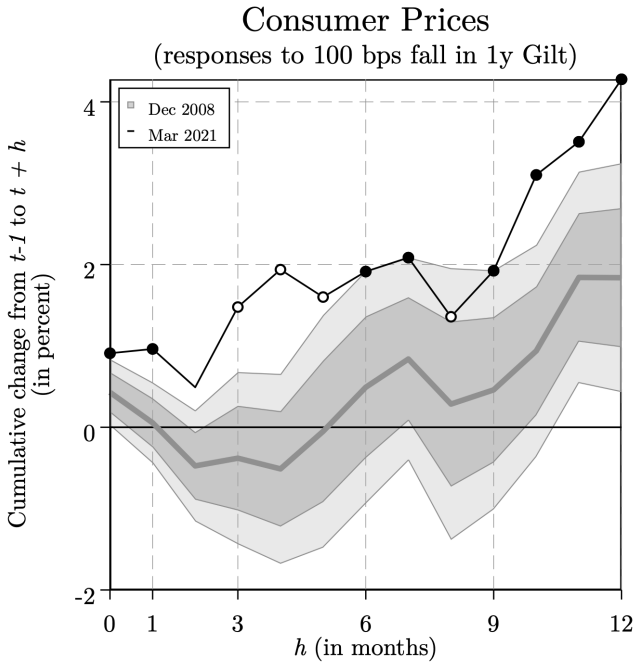
future inflation. We explore this next.

5.2 Predictive content of price gaps

Inflation at any date is a scalar projection of the infinite-dimensional state G_t: many distinct gap distributions can generate the same current inflation rate while embedding very different implications for future price adjustment. Proposition 1 formalized this intuition, showing that the gap distribution Granger-causes inflation. According to the theory, the distribution should serve as a leading indicator. Whether the estimated distribution retains this property in practice is an empirical question: the model may be misspecified, or the estimated gaps may be too noisy to carry useful predictive content. We now test whether they do. We proceed in two steps: a panel Granger-causality test establishes the direction of predictability, and a forecasting exercise quantifies the gain.

Granger causality. To test whether movements in the price gap distribution precede inflation, we exploit both time-series and cross-sectional variation through panel Granger-causality tests. For each COICOP class c, let π_t^c denote monthly inflation and let $\mathbf{m}_t^c$ denote a vector containing the weighted mean, standard deviation, skewness, and kurtosis of the

Figure 6: Inflation response to monetary policy shocks



Notes: Cumulative impulse response of the log price index to a 100 basis point decline in the one-year gilt yield, computed from the state-dependent local projection in Table 4 evaluated at the moments of the gap distribution observed at each reference date. For December 2008 (gray), dark and light shaded areas denote 68 and 90 percent confidence intervals, respectively. For March 2021 (black), solid circles denote that the 90 percent confidence interval excludes zero, open circles that the 68 percent interval excludes zero.

estimated price gap distribution for that class. We estimate the panel VAR

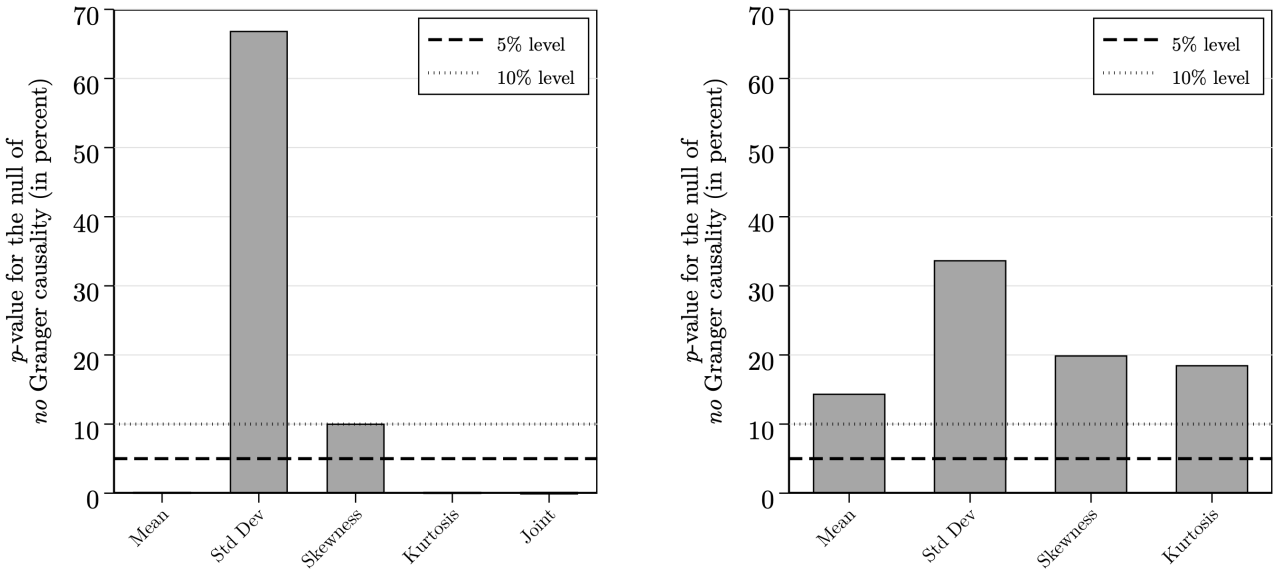
$$\mathbf{A}(L)\mathbf{y}_t^c = \boldsymbol{\alpha}_c + \boldsymbol{\delta}_t + \mathbf{e}_t^c, \tag{21}$$

where $\mathbf{y}_t^c = [\pi_t^c, \mathbf{m}_t^{c\prime}]'$, $\mathbf{A}(L)$ is a lag polynomial, and $\boldsymbol{\alpha}_c$ and $\boldsymbol{\delta}_t$ denote class and time fixed effects. We test whether lagged moments of the price gap distribution predict inflation, and conversely whether lagged inflation predicts the moments of the price gap distribution.

Figure 7 reveals a clear asymmetry. Panel (a) shows that lagged moments of the gap distribution predict future inflation: with the exception of the standard deviation, all moments Granger-cause inflation at least at the 10 percent level, with the strongest evidence coming from the mean and skewness. Panel (b) shows that the reverse does not hold: inflation does not predict any moment of the gap distribution at conventional significance levels. This one-directional pattern is what the theory implies: the gap distribution is a state variable that shapes the path of inflation, not the other way around. The mean gap captures the average direction of misalignment, while skewness reflects the directional asymmetry in latent pricing pressure, echoing its prominent role in the state-dependent propagation results of the previous subsection.

Forecasting. Having established the direction of predictability, we now ask how large the predictive gain is. Following Blinder and Reis (2005), we estimate a horse-race regression

Figure 7: Granger causality tests



(a) Do price gaps Granger cause inflation?

(b) Does inflation Granger cause price gaps?

Notes: p-values from Granger-causality tests based on the panel VAR in (21), estimated at the COICOP class level with 2 lags over the full sample period (1997–2023). Panel (a) tests whether lagged moments of the price gap distribution (mean, standard deviation, skewness, kurtosis) predict inflation. Panel (b) tests the reverse direction. Horizontal dashed lines indicate the 5 and 10 percent significance levels.

that nests lagged inflation and gap moments as competing predictors:

$$\pi^c_{t,t+h} = \alpha_c + \delta_t + \beta_h \, \pi^c_{t-h,t} + \sum_{j=1}^{4} \gamma_{j,h} \, m^c_{j,t} + v^c_{t+h}, \tag{22}$$

where $\pi^c_{t,t+h}$ denotes cumulative inflation between periods t and $t+h$ for product class c, and $m^c_{j,t}$ denotes the moments of the estimated price gap distribution for that class at time t. The specification includes class and time fixed effects, α_c and δ_t. If the gap distribution adds no information beyond what is already captured by past inflation, the coefficients $\gamma_{j,h}$ should be jointly zero.

Table 5 reports the results at horizons of 1, 3, 6, and 12 months. Lagged inflation enters negatively, consistent with mean reversion at the product-class level: classes that have just experienced above-average inflation tend to see lower inflation in subsequent periods as prices catch up to fundamentals. The mean gap coefficient is also negative: a more negative mean gap, indicating that firms' prices have on average fallen further behind their reset levels, predicts higher subsequent inflation. Skewness reinforces this pattern, with a negative and significant coefficient at all horizons, consistent with its role in capturing directional pricing pressure. Kurtosis enters positively at short horizons, indicating that heavier tails, with more firms near their adjustment thresholds, are associated with higher subsequent inflation. The standard deviation does not enter significantly. The joint F-test for the gap moments rejects at horizons up to 6 months ($p < 0.01$), indicating that the distribution con-

Table 5: Forecasting inflation with price gap moments

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Inflation	-0.116^{**}	-0.126^{*}	-0.038	-0.032
	(0.057)	(0.067)	(0.033)	(0.101)
Mean	-0.037^{***}	-0.066^{***}	-0.088^{**}	-0.124^{*}
	(0.011)	(0.024)	(0.039)	(0.070)
Std. Dev.	0.002	-0.006	-0.009	-0.025
	(0.010)	(0.025)	(0.038)	(0.066)
Skewness	-0.218^{**}	-0.348^{**}	-0.326^{**}	-0.343^{*}
	(0.106)	(0.164)	(0.154)	(0.173)
Kurtosis	0.002^{**}	0.003^{**}	0.002	0.004
	(0.001)	(0.001)	(0.001)	(0.003)
Panel FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Joint F-stat (inflation)	4.100	3.546	1.285	0.099
	[0.047]	[0.064]	[0.261]	[0.754]
Joint F-stat (gaps)	6.708	3.593	5.541	1.811
	[0.000]	[0.010]	[0.001]	[0.137]
R^2	0.115	0.180	0.237	0.344
Adj. R^2	0.098	0.164	0.222	0.332
Observations	21,888	21,624	21,228	20,437

Notes: Each column reports panel fixed-effects estimates from regression (22) at horizon h (in months). The dependent variable is cumulative CPI inflation between t and $t + h$ (in percent). *Inflation* is past h-month cumulative inflation. Mean, Std. Dev., Skewness, and Kurtosis are the contemporaneous moments of the estimated price gap distribution for each COICOP class. Standard errors clustered by product class in parentheses. Joint F-statistics test H_0: all inflation (resp. gap moment) coefficients are zero; p-values in brackets. R^2 and Adj. R^2 are the within R^2 from the fixed-effects estimator. *, **, *** denote significance at the 10%, 5%, and 1% levels.

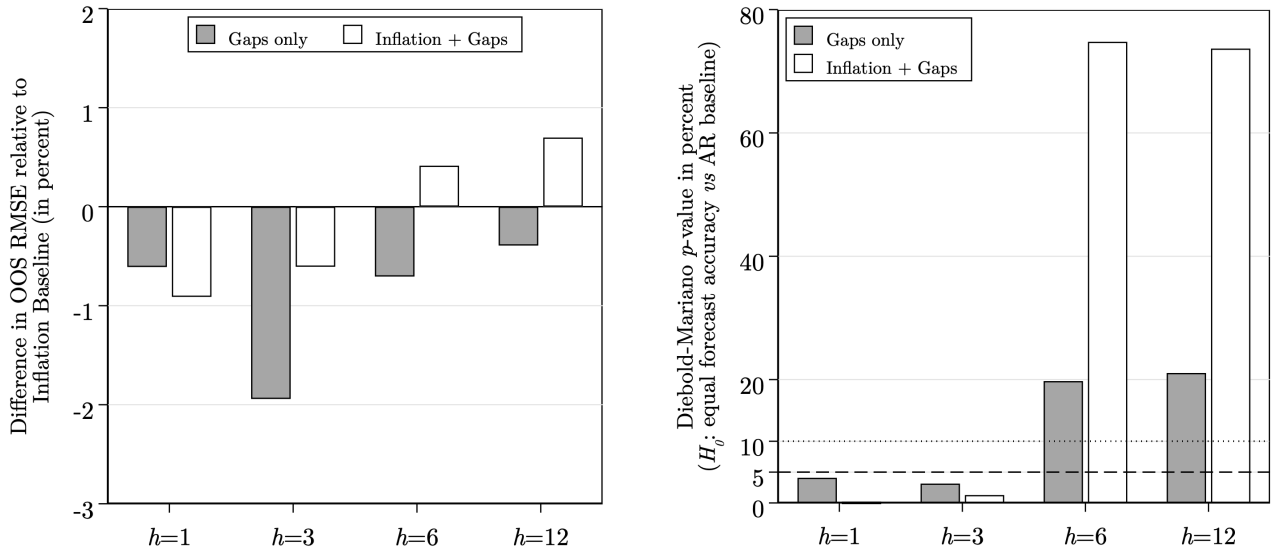
tains predictive information beyond what past inflation provides.

In-sample significance does not guarantee out-of-sample predictive power. To assess whether the gains survive in a genuinely predictive setting, we conduct a pseudo out-of-sample exercise: for each month from 2014m1 to 2023m12, we estimate (22) using only data available up to that date and produce forecasts at horizons of 1, 3, 6, and 12 months.[20]

Figure 8 summarizes the out-of-sample results. As Panel (a) shows, a model using price gap moments alone produces lower root mean squared errors than the lagged-inflation benchmark at all horizons, with the largest improvement at three months ahead. Combining gap moments with lagged inflation yields further gains, concentrated at shorter horizons. The Diebold–Mariano tests in Panel (b) confirm that the improvements are statistically significant at the 1-month and 3-month horizons. That the gains fade at longer horizons is

[20] At present, the price gap measures are based on full-sample estimation. Constructing rolling-window estimates to generate real-time vintages of the price gap distribution is work in progress.

Figure 8: Out-of-sample forecast evaluation



(a) Difference in RMSEs across models

(b) Diebold-Mariano test

Notes: Pseudo out-of-sample forecast evaluation over the period 2014m1–2023m12. At each date, the forecasting regression (22) is estimated on an expanding window using only data available up to that month. Panel (a): percentage difference in root mean squared forecast errors (RMSE) between the lagged-inflation benchmark and two alternative models (one using price gap moments only, one combining gap moments with lagged inflation) at horizons of 1, 3, 6, and 12 months. Positive values indicate that gap moments improve forecast accuracy. Panel (b): two-sided Diebold–Mariano p-values for the null of equal predictive accuracy between each gap-augmented model and the lagged-inflation benchmark, with Newey–West HAC standard errors (lag truncation equal to the forecast horizon).

consistent with the economics of menu-cost pricing: as time passes, firms gradually reprice and the information embedded in the initial gap distribution is released into observed inflation, reducing its marginal predictive value.[21]

6 Conclusion

The cross-sectional distribution of price gaps contains rich information about the state of the economy that headline inflation alone cannot reveal. This paper developed Bayesian methods to recover this distribution from micro price data, and our application to UK CPI data from 1997 to 2023 shows that the estimated distribution varies substantially over time, with the economy spending extended periods far from its ergodic benchmark. This variation has direct macroeconomic consequences. The gap distribution shapes the transmission of monetary policy: the same expansionary shock generates an inflationary response more than twice as large when the distribution is dispersed and asymmetric as when it is concentrated near the reset point. The gap distribution also Granger-causes inflation and forecasts it out of sample, with significant predictive gains over a lagged-inflation benchmark at horizons up to six months. For policymakers, the estimated gap distribution offers a real-time

[21]Our forecasting specification is deliberately parsimonious. The goal is not to build an optimal inflation forecasting model but to demonstrate that the cross-sectional distribution of price gaps contains information about future inflation that standard predictors miss.

reading of the inflationary pressure already embedded in the price-setting process and of the economy's vulnerability to new shocks.

The methodology we develop is general: it applies wherever agents follow *Ss*-type adjustment rules and the latent target is unobserved. This paper applies it to pricing, but the same forward-filtering backward-sampling approach can recover unobserved gaps in other lumpy-adjustment settings, from investment to durable goods purchases and inventory management. Within the pricing context, additional layers in the Gibbs sampler can accommodate richer models. We focus on the canonical framework here; incorporating aggregate uncertainty and strategic complementarities is a natural next step.

References

ALVAREZ, F., BIHAN, H. L. and LIPPI, F. (2016). The Real Effects of Monetary Shocks in Sticky Price Models: A Sufficient Statistic Approach. *American Economic Review*, **106** (10), 2817 – 2851.

— and LIPPI, F. (2009). Financial Innovation and the Transactions Demand for Cash. *Econometrica*, **77** (2), 363–402.

— and — (2014). Price Setting With Menu Cost for Multiproduct Firms. *Econometrica*, **82** (1), 89–135.

AUCLERT, A., RIGATO, R., ROGNLIE, M. and STRAUB, L. (2024). New Pricing Models, Same Old Phillips Curves? *The Quarterly Journal of Economics*, **139** (1), 121–186.

BACHMANN, R., CABALLERO, R. J. and ENGEL, E. M. R. A. (2013). Aggregate Implications of Lumpy Investment: New Evidence and a DSGE Model. *American Economic Journal: Macroeconomics*, **5** (4), 29–67.

BALEY, I. and BLANCO, A. (2021). Aggregate Dynamics in Lumpy Economies. *Econometrica*, **89** (3), 1235–1264.

BALL, L. and MANKIW, N. G. (1994). Asymmetric Price Adjustment and Economic Fluctuations. *The Economic Journal*, **104**, 247–261.

BERGER, D. and VAVRA, J. (2015). Consumption Dynamics During Recessions. *Econometrica*, **83** (1), 101–154.

BLANCO, A. (2021). Optimal Inflation Target in an Economy with Menu Costs and a Zero Lower Bound. *American Economic Journal: Macroeconomics*, **13** (3), 108–141.

—, BOAR, C., JONES, C. and MIDRIGAN, V. (2024a). Nonlinear Dynamics in Menu Cost Economies? Evidence from U.S. Data, NBER Working Paper 32748.

—, —, — and — (2024b). Nonlinear Inflation Dynamics in Menu Cost Economies, NBER Working Paper 32094.

BLINDER, A. S. and REIS, R. (2005). Understanding the Greenspan Standard. In *Proceedings - Economic Policy Symposium - Jackson Hole*, pp. 11 – 96.

BRAUN, R., MIRANDA-AGRIPPINO, S. and SAHA, T. (2025). Measuring Monetary Policy in the UK: The UK Monetary Policy Event-Study Database. *Journal of Monetary Economics*, **149**, 103645.

BUNN, P. and ELLIS, C. (2012). Examining the Behaviour of Individual UK Consumer Prices. *The Economic Journal*, **122** (558), F35 – F55.

CABALLERO, R. and ENGEL, E. (2007). Price Stickiness in Ss Models: New Interpretations of Old Results. *Journal of Monetary Economics*, **54** (Supplement), 100–121.

—, — and HALTIWANGER, J. (1995). Plant-Level Adjustment and Aggregate Investment Dynamics. *Brookings Papers on Economic Activity*, **2**.

CABALLERO, R. J. and ENGEL, E. M. R. A. (1993). Microeconomic Adjustment Hazards and Aggregate Dynamics. *Quarterly Journal of Economics*, **108** (2), 359–383.

CESA-BIANCHI, A., THWAITES, G. and VICONDOA, A. (2019). Monetary Policy Transmission in the United Kingdom: A High Frequency Identification Approach, Working Paper.

DOTSEY, M., KING, R. G. and WOLMAN, A. L. (1999). State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output. *The Quarterly Journal of Economics*, **114** (2), 655–690.

DOUCET, A. and JOHANSEN, A. M. (2011). *A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later*, Oxford University Press, pp. 656 – 704.

FAUST, J. and WRIGHT, J. H. (2013). Forecasting Inflation. In G. Elliot and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 2A, *1*, Elsevier, pp. 2–56.

GAGLIARDONE, L., GERTLER, M., LENZU, S. and TIELENS, J. (2025). Micro and Macro Cost-Price Dynamics in Normal Times and During Inflation Surges, NBER Working Paper 33478.

GAUTIER, E. and BIHAN, H. L. (2018). Shocks vs Menu Costs: Patterns of Price Rigidity in an Estimated Multi-Sector Menu-Cost Model, Banque de France Working Paper No. 682.

— and LE BIHAN, H. (2022). Shocks versus Menu Costs: Patterns of Price Rigidity in an Estimated Multisector Menu-Cost Model. *Review of Economics and Statistics*, **104** (4), 668–685.

GOLOSOV, M. and LUCAS, R. E. (2007). Menu Costs and Phillips Curves. *Journal of Political Economy*, **115** (2), 171–199.

HERBST, E. P. and SCHORFHEIDE, F. (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press, 1st edn.

JORDÀ, O. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, **95** (1), 161 – 182.

KALMAN, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering*, **82** (Seies D), 35–45.

KARADI, P., NAKOV, A., NUÑO, G., PASTÉN, E. and THALER, D. (2025). Strike While the Iron is Hot: Optimal Monetary Policy with a Nonlinear Phillips Curve, ECB Working Paper 3068.

KITAGAWA, G. (1987). Non-Gaussian State-Space Modeling of Nonstationary Time Series. *Journal of the American Statistical Association*, **82** (400), 1032–1041.

KRAMER, S. C. and SORENSON, H. W. (1988). Recursive Bayesian Estimation Uusing Piecewise Constant Approximations. *Automatica*, **24** (6), 789–801.

NAKAMURA, E. and STEINSSON, J. (2010). Monetary Non-Neutrality in a Mutisector Menu Cost Model. *The Quarterly Journal of Economics*, **125** (3), 961–1013.

ONS (2014). *Consumer Price Indices: Technical Manual*. Ofiice for National Statistics, 2014th edn.

ÒSCAR JORDÀ, SCHULARICK, M. and TAYLOR, A. M. (2020). The Effects of Quasi-Random Monetary Experiments. *Journal of Monetary Economics*, **112**, 22–40.

OTTONELLO, P. and WINBERRY, T. (2020). Financial Heterogeneity and the Investment Channel of Monetary Policy. *Econometrica*, **88** (6), 2473–2502.

RAUCH, H. E., TUNG, F. and STRIEBEL, C. T. (1965). Maximum Likelihood Estimates of Linear Dynamic Systems. *AIAA Journal*, **3** (8), 1445–1450.

ROBBINS, H. and MONRO, S. (1951). A Stochastic Approximation Method. *Annals of Mathematical Statistics*, **22** (3), 400–407.

ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Optimal Scaling for Various Metropolis-Hastings Algorithms. *Statistical Science*, **16** (4), 351–367.

SÄRKKÄ, S. (2013). *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks, Cambridge University Press, 1st edn.

TENREYRO, S. and THWAITES, G. (2016). Pushing on a String: US Monetary Policy Is Less Powerful in Recessions. *American Economic Journal: Macroeconomics*, **8** (4), 43–74.

Appendix to

Price Gaps and Inflation Dynamics

by Miguel Bandeira, Laura Castillo-Martínez and Shiyuan Wang

A Model Appendix

This appendix provides the theoretical foundations for the reduced-form model used in the main text. Appendix A.1 presents the microfoundations: the household and firm problems, the law of motion for the desired price, the firm's pricing problem, and aggregation. Appendix A.2 establishes conditions under which the optimal pricing policy takes the *Ss* form. Appendix A.3 derives the second-order welfare loss from price misalignment and states the Ramsey planner's problem. Appendix A.4 maps the desired-price notation used in Appendices A.1–A.3 to the reset-price notation used in the main text. Appendix **??** presents an additional simulation exercise on amplification under volatility shocks. Appendix A.5 contains the proof of Proposition 1.

Notation. Appendices A.1–A.3 work with the *desired-price gap* $s_{i,t} \equiv p_{i,t} - p^*_{i,t}$, where $p^*_{i,t}$ is the firm's static profit-maximizing log price. The cross-sectional distribution of $s_{i,t}$ is denoted H_t with density h_t. Appendix A.4 establishes the connection to the *reset-price gap* $x_{i,t} \equiv p_{i,t} - p^r_{i,t}$ and the distribution G_t used in the main text.

A.1 Microfoundations

Households. A representative household maximizes expected discounted lifetime utility:

$$\mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\log C_t - N_t\right],$$

where C_t is consumption, N_t is labor supply, and $\beta \in (0,1)$ is the discount factor. Consumption is a CES aggregate of differentiated varieties:

$$C_t = \left(\int_0^1 c_t(i)^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}},$$

where $\epsilon > 1$ is the elasticity of substitution. The household faces a cash-in-advance constraint:

$$P_t C_t = M_t,$$

where P_t is the aggregate price index and M_t is aggregate nominal spending, which is controlled by monetary policy.

The household's optimality conditions yield a labor supply condition and a demand func-

tion for each variety:

$$W_t = M_t, \tag{23}$$

$$c_t(i) = \left(\frac{P_t(i)}{P_t}\right)^{-\epsilon} C_t, \tag{24}$$

where (23) combines the intratemporal optimality condition $W_t = P_t C_t$ with the cash-in-advance constraint.

Firms. Each firm $i \in [0,1]$ produces a differentiated variety using labor with a linear technology:

$$y_t(i) = A_t(i) \cdot N_t(i),$$

where $A_t(i)$ is idiosyncratic productivity. The log of productivity follows a random walk:

$$\log A_t(i) = \log A_{t-1}(i) + a_{i,t}, \quad a_{i,t} \overset{iid}{\sim} \mathcal{N}(0, \sigma_a^2),$$

with $a_{i,t}$ independently distributed across firms and over time. We allow for an aggregate cost-push shock $\phi_t \in [0,1)$ that drives a wedge between wages and marginal cost. The nominal marginal cost of firm i is therefore:

$$MC_t(i) = (1-\phi_t)\frac{W_t}{A_t(i)} = (1-\phi_t)\frac{M_t}{A_t(i)},$$

where the second equality uses (23). The static profit-maximizing (desired) price is:

$$P_t^*(i) = \frac{\epsilon}{\epsilon - 1} \cdot (1-\phi_t) \cdot \frac{M_t}{A_t(i)}, \tag{25}$$

where $\epsilon/(\epsilon-1)$ is the constant desired markup over marginal cost.

Desired price dynamics. Taking logs of (25) and first-differencing yields:

$$p_t^*(i) - p_{t-1}^*(i) = \Delta \log(1-\phi_t) + \Delta \log M_t - a_{i,t}.$$

Define the common drift of the desired price as the sum of the cost-push shock (in logs) and the growth in nominal spending, $\nu_t \equiv \Delta \log(1-\phi_t) + \Delta \log M_t$ and the reduced-form idiosyncratic shock as $\varepsilon_{i,t} \equiv -a_{i,t}$. Then the above equation becomes:

$$p_t^*(i) = \nu_t + p_{t-1}^*(i) + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \tag{26}$$

where $\sigma_\varepsilon = \sigma_a$.

The firm's pricing problem. In each period, firm i chooses whether to adjust its price $P_t(i)$, trading off the gain from realigning with the desired price against the random menu cost. Using the demand function (24), the production technology, and the desired price (25), the

firm's normalized real profit in period t can be decomposed as:

$$\frac{\Pi_t(i)}{P_t C_t} = B_t(i) \cdot \tilde{\pi}(s_{i,t}), \tag{27}$$

where $s_{i,t} \equiv p_{i,t} - p^*_{i,t}$ is the desired-price gap,

$$\tilde{\pi}(s) \equiv e^{(1-\epsilon)s} - \frac{1}{\mathcal{M}} e^{-\epsilon s} \tag{28}$$

is the profit expressed purely as a function of the gap with $\mathcal{M} \equiv \epsilon/(\epsilon-1)$ denoting the desired markup, and $B_t(i) \equiv (P^*_t(i)/P_t)^{1-\epsilon} > 0$ is a time-varying weight that depends on aggregate conditions and the firm's idiosyncratic productivity but not on its pricing decision.[22]

Following Alvarez and Lippi (2014), we assume that the resource cost of price adjustment is proportional to the firm's desired profit $\tilde{\pi}(0) \cdot B_t(i) \cdot P_t C_t = \frac{1}{\epsilon} B_t(i) \cdot P_t C_t$. The normalized adjustment cost is therefore $B_t(i) \cdot c_{i,t}$, where $c_{i,t} \in \{0, \hat{c}\}$ with $\Pr(c_{i,t} = 0) = \lambda$.[23]

The firm minimizes expected discounted profit losses net of adjustment costs:

$$\min_{\{s_{i,t}\}_{t=0}^{\infty}} \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t B_t(i) \left[L(s_{i,t}) + c_{i,t} \cdot \mathbf{1}_{\{s_{i,t} \neq s_{i,t-1} - \nu_t - \varepsilon_{i,t}\}} \right], \tag{29}$$

where $L(s) \equiv \tilde{\pi}(0) - \tilde{\pi}(s)$ is the per-period profit loss from maintaining a desired-price gap s instead of the optimal gap 0.

Reduction to the quadratic formulation. By construction, $s = 0$ maximizes $\tilde{\pi}$, so $\tilde{\pi}'(0) = 0$. A second-order Taylor expansion of the profit loss around the optimal gap gives:

$$L(s) = \tilde{\pi}(0) - \tilde{\pi}(s) \approx \frac{\epsilon - 1}{2} s^2. \tag{30}$$

Substituting into (29), the firm's problem becomes:

$$\min_{\{s_{i,t}\}_{t=0}^{\infty}} \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t B_t(i) \left[\frac{\epsilon - 1}{2} s_{i,t}^2 + c_{i,t} \cdot \mathbf{1}_{\{s_{i,t} \neq s_{i,t-1} - \nu_t - \varepsilon_{i,t}\}} \right]. \tag{31}$$

Since $B_t(i) > 0$ multiplies both the flow loss and the adjustment cost, dividing through by $B_t(i)(\epsilon - 1)$ and setting $\tilde{\beta}_t \equiv \beta \, B_{t+1}(i)/B_t(i) \approx \beta$ yields the rescaled problem with flow loss $\frac{1}{2}s^2$ and menu cost $\bar{c} \equiv \hat{c}/(\epsilon - 1)$.[24]

[22] The derivation of (27) proceeds by writing $P_t(i)/P_t = (P^*_t(i)/P_t)\, e^{s_{i,t}}$, substituting into the profit expression, and using $MC_t(i) = P^*_t(i)/\mathcal{M}$ to simplify.

[23] A similar scaling of fixed adjustment costs by firm size is standard in the lumpy investment literature (Baley and Blanco, 2021).

[24] The approximation $\tilde{\beta}_t \approx \beta$ requires $B_{t+1}(i)/B_t(i) \approx 1$, which holds in a stationary environment or when aggregate conditions change slowly relative to the discount frequency.

Bellman equation and *Ss* rule. The rescaled value function $V_t(s) \equiv J_t(s)/[B_t(i)(\epsilon-1)]$ satisfies the Bellman equation:

$$V_t(s) = \frac{1}{2}s^2 + \tilde{\beta}_t \, \mathbf{E}_t\left[(1-\lambda)\min\Big(V_{t+1}(s-\nu_{t+1}-\varepsilon),\ \bar{c}+\min_{s'} V_{t+1}(s')\Big) + \lambda \min_{s'} V_{t+1}(s')\right], \tag{32}$$

where $\bar{c} \equiv \hat{c}/(\epsilon-1)$ is the rescaled menu cost and $\tilde{\beta}_t \equiv \beta\, B_{t+1}(i)/B_t(i)$ is the effective discount factor. In a stationary environment (constant drift $\nu_t = \nu$ and slowly varying $B_t(i)$) we have $\tilde{\beta}_t \approx \beta$ and the value function becomes time-invariant.[25]

As discussed in Appendix A.2, the solution to (32) takes the form of an *Ss* rule: for each drift ν_t, there exist thresholds $\underline{s}_t < s_t^* < \bar{s}_t$ such that the firm resets its desired-price gap to s_t^* whenever it falls outside $[\underline{s}_t, \bar{s}_t]$ or when adjustment is free.

Aggregation. Let H_t denote the cross-sectional distribution of desired-price gaps $s_{i,t} \equiv p_{i,t} - p_{i,t}^*$ across firms, with density h_t. Given the *Ss* pricing rule, H_t evolves according to:

$$H_t = \mathcal{T}(H_{t-1};\, \nu_t, \sigma_\varepsilon, \underline{s}_t, \bar{s}_t, s_t^*),$$

where $\mathcal{T}$ maps the distribution from period $t-1$ to period t, given the drift, shock volatility, and the parameters of the pricing rule. The CES price index implies the consistency condition:

$$1 = \int_0^1 B_t(i)\, e^{(1-\epsilon)s_{i,t}}\, di,$$

where $B_t(i) = (P_t^*(i)/P_t)^{1-\epsilon}$ is the firm-specific weight from (27). Integrating the inflation identity $\pi_{i,t} = (p_{i,t}^* - p_{i,t-1}^*) + (s_{i,t} - s_{i,t-1})$ across firms yields:

$$\pi_t = \nu_t + \int (s_{i,t} - s_{i,t-1})\, di,$$

where ν_t is the common drift of the desired price. Finally, the aggregate resource constraint is:

$$N_t = \frac{1}{\mathcal{M}(1-\phi_t)} \int_0^1 B_t(i)\, e^{-\epsilon s_{i,t}}\, di + \hat{c} \int_0^1 B_t(i)\, \mathbf{1}_{\{c_{i,t}=\hat{c},\ \text{adjust}\}}(i)\, di,$$

where the first term is aggregate production labor and the second is the real resource cost of price adjustment, with $\hat{c}$ denoting the structural menu cost. The resource cost is incurred only by firms that draw a positive menu cost ($c_{i,t} = \hat{c}$, probability $1-\lambda$) and whose gap falls outside the inaction region; firms receiving free adjustment opportunities pay no resource cost.

[25] If changes in ν_t are unanticipated, the firm treats the current drift as permanent and the time-invariant Bellman equation applies conditional on ν_t. When the firm instead anticipates the future path of drift, V_t is genuinely non-stationary; see the non-stationary case in Appendix A.2.

A.2 The *Ss* pricing rule

This appendix establishes conditions under which the optimal pricing policy for (32) takes the *Ss* form. The argument follows the approach in Auclert *et al.* (2024) (Online Appendix B.2), who prove the result analytically for $\nu = 0$, and extends it to $\nu > 0$ via a quasiconvexity condition.

A.2.1 Stationary case

Consider the stationary version of (32) with constant drift ν and discount factor β:

$$V(s) = \frac{1}{2}s^2 + \beta\Big\{(1-\lambda)\,\mathbf{E}\Big[\min\big(V(s-\nu-\varepsilon),\ \bar{c} + V(s^*)\big)\Big] + \lambda\, V(s^*)\Big\}, \tag{33}$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and $s^* = \arg\min V$. The Bellman operator is a contraction (modulus β), so V exists and is unique. Standard arguments yield the following properties:

(i) $V(s) \geq \frac{1}{2}s^2$ for all s, and $V(s^*) \leq \beta(1-\lambda)\bar{c}/(1-\beta)$.

(ii) Setting $M = \sqrt{2\bar{c}/(1-\beta)}$, the firm strictly prefers to adjust for $|s| > M$, so it suffices to track V on $[-M, M]$.

(iii) Since the expectation in (33) is a convolution with a Gaussian kernel, V is C^∞.

(iv) Differentiating under the integral gives

$$V'(s) = s + \beta(1-\lambda)\int_{\underline{s}}^{\bar{s}} V'(y)\, f(s-\nu-y)\, dy, \tag{34}$$

where f is the density of ε and $[\underline{s}, \bar{s}]$ is the inaction set. Since the integral is bounded, $V'(s)/s \to 1$ as $|s| \to \infty$, so V has at least one interior minimizer.

Quasiconvexity implies *Ss*. Say V is *strictly quasiconvex* if it has a unique minimizer s^* and is strictly decreasing on $(-\infty, s^*]$, strictly increasing on $[s^*, \infty)$. If V is strictly quasiconvex, then every sublevel set $\{s : V(s) \leq c\}$ is a closed interval. In particular, the inaction set $\mathcal{I} = \{s : V(s) \leq \bar{c} + V(s^*)\}$ is a closed interval $[\underline{s}, \bar{s}]$ with $\underline{s} < s^* < \bar{s}$ (since $\bar{c} > 0$). Outside this interval the firm strictly prefers to pay $\bar{c}$ and reset to the unique minimizer s^*; inside, it weakly prefers inaction. This is the *Ss* rule.

Note that quasiconvexity is weaker than convexity: V can have $V'' < 0$ near the thresholds, as long as it remains monotone on each side of s^*.

Verification. When $\nu = 0$, Auclert *et al.* (2024) show that V is symmetric ($V(s) = V(-s)$) and $V'(s) > 0$ for $s > 0$, which gives strict quasiconvexity with $s^* = 0$. When $\nu > 0$, symmetry breaks and the direct analytical proof is not available. Quasiconvexity can be checked ex-post from the computed value function.

A.2.2 Non-stationary case

Under MIT shocks to ν_t, the properties (i)–(iv) and the quasiconvexity-implies-Ss argument apply to V_t at each date t, treating the continuation value V_{t+1} as given and replacing ν with ν_{t+1}. Hence the Ss structure holds at date t whenever V_t is strictly quasiconvex.

If the shock is permanent, V_t is simply the stationary fixed point at the current ν_t, and verification reduces to the stationary case. If ν_t reverts along a deterministic path $\nu_t \to \nu_\infty$, the firm anticipates future changes and V_t is genuinely non-stationary. In this case V_t is obtained by backward induction from the terminal stationary solution V_∞, and quasiconvexity should be checked at each backward step.

A.3 Welfare loss and the Ramsey problem

This appendix derives the welfare loss from price misalignment and states the Ramsey planner's problem. The derivation follows standard steps in the menu cost literature and provides the formal justification for equation (4) in the main text.

Setup. Using the microfoundations in Appendix A.1, the representative household maximizes:

$$\mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\log C_t - N_t\right],$$

subject to the cash-in-advance constraint $P_t C_t = M_t$, where M_t is aggregate nominal spending controlled by monetary policy.

The Ramsey problem. The planner chooses a path for nominal spending $\{M_t\}_{t=0}^{\infty}$ to maximize household welfare subject to:

$$C_t = M_t / P_t \qquad \text{(cash-in-advance)}, \tag{35}$$

$$1 = \int e^{(1-\epsilon)s}\, h_t(s)\, ds \qquad \text{(price index)}, \tag{36}$$

$$N_t = C_t \int e^{-\epsilon s}\, h_t(s)\, ds + \hat{c}(1-\lambda)\omega_t \qquad \text{(labor market clearing)}, \tag{37}$$

$$H_t = \mathcal{T}(H_{t-1};\, \nu_t,\, \sigma_\varepsilon,\, \underline{s}_t,\, \bar{s}_t,\, s_t^*) \qquad \text{(distribution evolution)}, \tag{38}$$

$$\nu_t = \Delta \log(1-\phi_t) + \Delta \log M_t \qquad \text{(drift)}, \tag{39}$$

$$(\underline{s}_t, \bar{s}_t, s_t^*) \ \text{ from firm value function } V_t(s) \qquad \text{(optimal pricing)}, \tag{40}$$

where ω_t denotes the fraction of firms that pay the menu cost $\hat{c}$ in period t, and ϕ_t is an exogenous cost-push shock.

This problem is challenging because the planner must optimize over policy paths $\{M_t\}$, where each candidate policy alters the drift ν_t via (39), which changes the evolution of H_t via (38), which in turn changes the firm's value function and thus the Ss thresholds $(\underline{s}_t, \bar{s}_t, s_t^*)$ via (40). This nested feedback between policy, the distribution, and individual pricing rules creates a formidable computational challenge. Karadi *et al.* (2025) provide a solution using

sequence-space Jacobians that linearize the distribution's response to policy perturbations.

Second-order welfare approximation. We can nonetheless characterize the welfare cost of price misalignment by approximating the period utility loss relative to the flexible-price allocation. Using the constraints above, the period utility of the household can be written as:

$$u_t = \log C_t - N_t = \log \frac{M_t}{P_t} - \frac{M_t}{P_t} \int e^{-\epsilon s}\, h_t(s)\, ds - \hat{c}(1-\lambda)\omega_t.$$

In the flexible-price economy ($s_{i,t} = 0$ for all i), the price index constraint becomes $P_t = P_t^*(i)$ for all i, the integral $\int e^{-\epsilon s}\, h_t(s)\, ds = 1$, and the adjustment cost vanishes. Denoting the flexible-price utility by u_t^e and letting $\tilde{c}_t \equiv \log(C_t / C_t^e)$ denote the output gap, a second-order expansion of the utility loss yields:[26]

$$\mathcal{L}_t \equiv u_t^e - u_t \approx \frac{1}{2}\tilde{c}_t^2 + \frac{\epsilon^2}{2} \int s^2\, h_t(s)\, ds + \hat{c}(1-\lambda)\omega_t. \tag{41}$$

The second term is, up to a constant, the welfare loss from price misalignment: $\frac{\epsilon^2}{2} \int s^2\, h_t(s)\, ds = \frac{\epsilon^2}{2}[\mathrm{Var}(s_t) + (E[s_t])^2]$. This is the expression summarized in equation (4) of the main text, where the factor ϵ^2 is absorbed into the normalization and the desired-price gap s is replaced by the reset-price gap x (see Appendix A.4). The decomposition has a clear economic interpretation:

- The *output gap* term $\frac{1}{2}\tilde{c}_t^2$ captures the welfare cost of aggregate demand deviating from its efficient level.
- The *gap dispersion* term $\frac{\epsilon^2}{2} \int s^2\, h_t(s)\, ds$ captures the misallocation of demand across varieties. Firms with $s_{i,t} \neq 0$ produce quantities that deviate from the efficient level, generating welfare losses even if aggregate output is at its efficient level. This term depends on the full cross-sectional distribution of gaps and cannot be inferred from aggregate inflation alone.
- The *adjustment cost* term $\hat{c}(1-\lambda)\omega_t$ captures the real resources consumed by firms that pay the menu cost.

The key implication is that monitoring the variance and mean of the gap distribution provides a direct measure of one component of the welfare loss from price misalignment. This is the basis for the welfare metric $\mathcal{W}_t$ used in the main text.

A.4 Connection to the main text

The main text works with the *reset price* and the *reset-price gap*, rather than the desired price and its gap. This appendix establishes the mapping and derives equations (1)–(3) and (4) of the main text.

[26] The derivation uses the fact that $e^{(1-\epsilon)s} \approx 1 + (1-\epsilon)s + \frac{(1-\epsilon)^2}{2}s^2$ and $e^{-\epsilon s} \approx 1 - \epsilon s + \frac{\epsilon^2}{2}s^2$ around $s = 0$, combined with the price index constraint (36) to eliminate the mean gap term.

Reset price. When a firm adjusts, it sets its desired-price gap to the optimal reset point $s_t^* \equiv \arg\min V_t$, which is generically nonzero under positive drift, though quantitatively small ($s^* \approx 0.008$ under the calibration in Table A.1). The price the firm chooses upon adjustment is therefore $p_{i,t}^r \equiv p_{i,t}^* + s_t^*$, which we call the *reset price*. First-differencing and using (26):

$$p_{i,t}^r - p_{i,t-1}^r = \underbrace{(\nu_t + \varepsilon_{i,t})}_{\Delta p_{i,t}^*} + \Delta s_t^* = \mu_t + \varepsilon_{i,t},$$

where

$$\mu_t \equiv \nu_t + \Delta s_t^* \tag{42}$$

is the *reset-price drift*—the drift used throughout the main text. This is equation (1). The reset-price drift μ_t differs from the desired-price drift ν_t only when the optimal reset point s_t^* changes. For small drift changes, Δs_t^* is negligible and $\mu_t = \nu_t$ to an excellent approximation. Even under large shocks (e.g., a 20 pp increase in annual trend inflation), Δs_t^* remains small relative to the drift itself, so the approximation is reasonable in percentage terms.

Reset-price gap and the firm's problem. Define the *reset-price gap* $x_{i,t} \equiv p_{i,t} - p_{i,t}^r = s_{i,t} - s_t^*$. In the absence of adjustment, the gap evolves as:

$$x_{i,t}^{\text{na}} = \underbrace{s_{i,t-1} - \nu_t - \varepsilon_{i,t}}_{\text{no-adjustment } s_{i,t}} - s_t^* = x_{i,t-1} + s_{t-1}^* - \nu_t - \varepsilon_{i,t} - s_t^* = x_{i,t-1} - \mu_t - \varepsilon_{i,t},$$

using $s_{i,t-1} = x_{i,t-1} + s_{t-1}^*$ and $\mu_t = \nu_t + \Delta s_t^*$. The flow loss in the rescaled problem transforms as $\frac{1}{2}s^2 = \frac{1}{2}(x + s_t^*)^2 = \frac{1}{2}x^2 + s_t^* \, x + \frac{1}{2}(s_t^*)^2$. The constant is irrelevant; the linear term is negligible for s_t^* small (≈ 0.008). Up to this approximation, the rescaled firm's problem (31) written in terms of $x_{i,t}$ yields equation (2) in the main text.

Ss **rule.** The *Ss* rule resets the desired-price gap to s_t^*, which corresponds to resetting the reset-price gap to zero: $x_{i,t} = 0$. The inaction region becomes $[\underline{x}_t, \bar{x}_t]$ with

$$\underline{x}_t \equiv \underline{s}_t - s_t^* < 0 < \bar{s}_t - s_t^* \equiv \bar{x}_t.$$

This is equation (3) in the main text.

Distribution. The cross-sectional distribution of reset-price gaps, denoted G_t with density g_t in the main text, is H_t shifted by s_t^*:

$$g_t(x) = h_t(x + s_t^*).$$

In particular, $E[x_t] = E[s_t] - s_t^*$ and $\text{Var}(x_t) = \text{Var}(s_t)$.

Welfare. The welfare loss from price misalignment (41), expressed in reset-price gaps, is:

$$\mathcal{W}_t = \tfrac{1}{2}[\text{Var}(x_t) + (E[x_t])^2] = \tfrac{1}{2}[\text{Var}(s_t) + (E[s_t] - s_t^*)^2],$$

which approximates the exact desired-price welfare cost $\frac{1}{2}[\text{Var}(s_t) + (E[s_t])^2]$ up to terms involving s_t^*, negligible under our calibration ($s^* \approx 0.008$). This is equation (4) in the main text.

A.5 Proof of Proposition 1

Proof. For part (i), in the deterministic limit the only firms adjusting at $t+h$ are those with gaps at time t in the interval $[\underline{x} + (h-1)\mu,\ \underline{x} + h\mu)$: exactly the firms whose gaps will drift to the lower threshold over the next h periods. Their mass is approximately $g_t(\underline{x} + (h-1)\mu) \cdot \mu$, and upon adjustment each resets from approximately $\underline{x}$ to zero, contributing a price increase of $|\underline{x}|$. Non-adjusting firms hold prices fixed. Aggregating yields (7).

For part (ii), note that inflation at date s reveals the density at the lower threshold one period earlier: $\pi_s \propto g_{s-1}(\underline{x})$. The density at interior points $g_t(\underline{x} + h\mu)$ for $h \geq 1$, which by part (i) governs inflation at horizon $h+1$, corresponds to firms that have not recently adjusted and whose positions are not reflected in any past inflation rate. Starting from the ergodic distribution, perturb the density at some interior point $\underline{x} + h^*\mu$ with $h^* > K$, while preserving the density on $[\underline{x},\ \underline{x} + K\mu]$. The perturbed economy produces the same inflation for horizons 1 through K but differs at horizon $h^* + 1$. □

B State Space Methods Appendix

B.1 Forward Filtering Backward Sampling: Proof

To be added.

B.2 Grid-based FFBS

To be added.

B.3 Gibbs Sampler

To be added.

B.4 Details on Monte Carlo

To be added.

C Data Appendix

C.1 Data cleaning: product substitutions, sales and quote-line gaps

The model in Section 2 does not feature sales or product substitutions and, hence, the state-space representation above should be interpreted as a data generating process for *regular prices*. To ensure consistency between the model and the data, we remove product substitutions and temporary sales from the data in order to obtain quote-lines of regular prices.

Product substitutions. For each outlet where prices are collected, price collectors begin by selecting a single product - among those that match the item's specification - that best reflects what consumers typically purchase in that area. Once selected, the same product is priced each month at that outlet. However, substitutions are necessary if the original product is no longer available or if its physical characteristics, such as weight or size, change. In such cases, the price collector flags the quote accordingly. To account for these substitutions, we treat each flagged instance as a break in the price trajectory.

Product sales. The prices collected by the price collectors are shelf prices, meaning they may reflect temporary sales at the time of collection. If a price is affected by a sale, it is flagged using a sales indicator. We use these sales flags to construct quote-lines of regular prices. First, we treat any price not flagged as a sale as the regular price for that period. Second, if a price is flagged as a sale, we use the most recent regular price observed before the sale. If no such prior regular price exists, we use the first regular price observed after the sale. Third, if no regular prices are observed either before or after the sale, we treat the quote-line as composed exclusively of sale prices and exclude it from the sample.

Quote-line gaps. After applying the cleaning steps described above, some quote-lines may contain gaps, for example, due to price quotes that failed ONS internal validation. Since estimating reset inflation requires quote-lines with contiguous observations of regular prices, we split any quote-line containing gaps into separate segments of contiguous observations. We then exclude any resulting singleton quote-lines from the sample.

C.2 From price quotes to price indices

To calculate aggregate measures of inflation, both actual and frictionless, from quote-line data, we adapt the methodology used to construct the official CPI as described in ONS (2014). In the data, each firm is equivalent to a quote-line which is uniquely identified by a triplet of subscripts (i, j, t) where i refers to a particular outlet, j a particular product and t a month.

From price quotes to elementary aggregates. The lowest level at which price quotes are aggregated into indices is the stratum. Each item in the data may be stratified in one of four ways: by region, by shop type, by both region and shop type, or not stratified at all.[27] Within

[27]There are 12 regions (London, Southeast, Southwest, East Anglia, East Midlands, Yorks and Humber, Northwest, North, Wales, Scotland and Northern Ireland) and two shop types: (i) multiple if the shop has 10

each stratum, individual price quotes are aggregated to form an *elementary aggregate index*. The primary method used to construct these aggregates in the CPI is the geometric mean, also known as the Jevons formula. Specifically, the elementary aggregate for stratum s of item j in month t, with base period t^b is given by:

$$\mathcal{I}^{EA}_{s,j,t|t^b} = \left(\prod_{i=1}^{N_{s,j}} \frac{P_{i,j,t}}{P_{i,j,t^b}} \right)^{\frac{1}{N_{s,j}}}, \tag{43}$$

where $P_{i,j,t}$ and P_{i,j,t^b} are the prices of the same product collected in a particular outlet in period t and in the base period t^b, and $N_{s,j}$ is the number of price quotes in stratum s of item j, adjusted using central shop weights as replication factors. Using the previous month as the base period and applying the definition of cumulated individual inflation, the expression above can be rewritten as:

$$\mathcal{I}^{EA}_{s,j,t|t-1} = \exp\left\{ \frac{1}{N_{s,j}} \sum_{i=1}^{N_{s,j}} \Delta Z_{i,j,t} \right\}. \tag{44}$$

The regular price elementary aggregates are obtained from (44) with $\Delta Z^{\text{reg}}_{i,j,t}$. Similarly, the frictionless elementary aggregates are obtained from (44) with $\Delta \widehat{Z}^{\star}_{i,j,t}$.[28]

From elementary aggregates to higher level indices. Indices for higher levels are weighted averages of the elementary aggregate indices. First, we combine elementary aggregates using stratum weights to construct item-level indices. Second, we aggregate item-level indices using item weights to produce Classification of Individual Consumption by Purpose (COICOP) indices at the class, group, and division levels, as well as the overall CPI.

or more outlets or (ii) independent if it has fewer than 10 outlets).

[28] A key difference between the elementary aggregates produced following (44) and the elementary aggregates underlying the published CPI is that the later uses the previous month of January as the base period. Because of the breaks in quote-lines created to deal with product substitutions and quote-line gaps, using the previous January as the base period would require to either: (i) drop any observations that do not have information for $\Delta Z^{\text{reg}}_{i,j,t}$ in the previous month of January which are almost a half of the total sample or (ii) impute values for $\Delta Z^{\text{reg}}_{i,j,t^b}$ and $\Delta Z^{\star}_{i,j,t^b}$ at the base period. We choose to change the base period and chain-link the elementary aggregate indices every month over either of these two options.

E Additional Tables

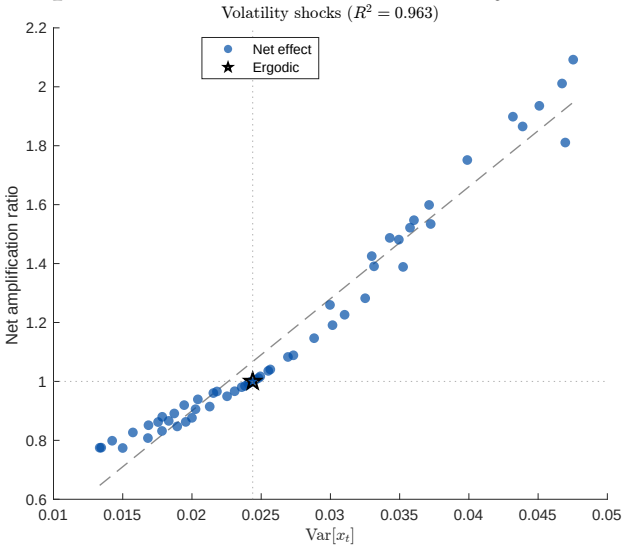
Table A.1: Calibration for simulated illustrations

Parameter	Description	Value	Source	Data	Model
Set externally					
β	Discount factor (monthly)	0.997	4% annual real rate		
μ	Steady-state drift (monthly)	0.00165	2% annual inflation		
Calibrated internally					
σ_ε	Idiosyncratic shock s.d.	0.073	Freq. of price change	12.6%	12.6%
λ	Free adjustment probability	0.120	Mean $\lvert\Delta p\rvert$	15.3%	15.3%
ξ	Menu cost	0.547	Kurtosis of Δp	3.8	3.8
Implied Ss rule					
$\underline{x}$	Lower threshold	-0.488			
$\bar{x}$	Upper threshold	0.480			

Notes: Target moments are from UK CPI micro data as reported by Blanco (2021). The parameters σ_ε, λ, and ξ are jointly calibrated to match the three target moments. The *Ss* rule parameters are obtained from the optimal policy solution given the structural parameters. Thresholds are expressed in terms of the reset price gap.

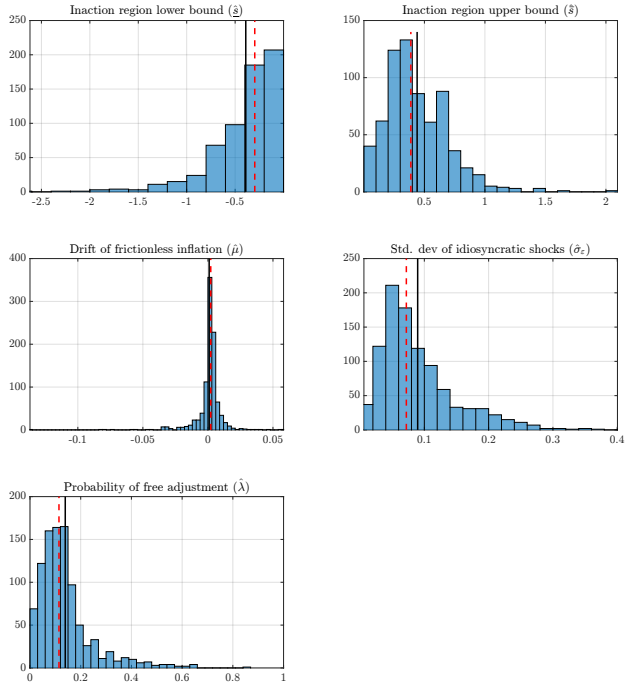
D Additional Figures

Figure A.1: Amplification ratio under volatility-shock backgrounds



Notes: Analogous to Panel B of Figure 1, but with distributions displaced by persistent perturbations to idiosyncratic volatility σ_ε rather than to the drift. Each dot represents a distribution sampled at a different point along the transition back to the steady state. The x-axis plots the variance $\mathrm{Var}(x_t)$. The y-axis plots the net amplification ratio: impact excess inflation minus the no-shock counterfactual, normalized by the ergodic response. The star marks the ergodic distribution. Same calibration as Figure 1.

Figure A.2: Heterogeneity in posterior means of estimated parameters



Notes: Each panel shows the cross-item distribution of posterior means for a given estimated parameter. The histogram for the lower bound of the inaction region excludes items with no observed price changes triggered by crossing that bound (358 out of 979 items). Similarly, the histogram for the upper bound excludes items without upper-bound-triggered price changes. Solid black and red dashed lines indicate, respectively, the mean and median of posterior means across the subset of items shown. Summary statistics for the full item sample are reported in Table 3.