Price Gaps and Inflation Dynamics*

Miguel Bandeira† Laura Castillo-Martínez‡ Shiyuan Wang§

March 2026

(*click here for the latest version*)

Abstract

Non-linear inflation dynamics are a key prediction of menu cost models. The literature has mostly focused on shock size as the source of these non-linearities, leaving the role of the state of the economy largely unexplored. This paper brings the focus back to the cross-sectional distribution of price gaps—the misalignment between firms' actual and desired prices—and shows that it is a powerful source of non-linearity too. Using a random menu cost model, we show that economies with different gap distributions respond very differently to the same shock. We recast the model in state-space form and develop Bayesian methods to recover firm-level price gaps from observed prices. Applying this approach to UK CPI microdata from 1997 to 2023, we document that the gap distribution varies substantially over time, with the economy spending extended periods far from its ergodic distribution. We show that this variation matters: the gap distribution shapes the transmission of monetary policy shocks and serves as a leading indicator for 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, "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 these non-linearities: firms tolerate some price misalignment due to adjustment costs, and how many of them adjust in response to a given shock is endogenous, generating non-linear dynamics in the aggregate. 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 source of non-linearity that has received less attention: the state of the economy when the shock arrives as summarized by the cross-sectional distribution of price gaps, i.e., the misalignment between firms' actual and desired prices.

In other settings where adjustment is lumpy, such as investment or durable goods purchases, state dependence from the cross-sectional distribution has been explored more extensively. This is partly because gaps between actual and desired levels can be constructed from observable fundamentals or proxied by the age of the stock. In pricing, by contrast, measuring price gaps requires knowledge of firms' desired prices, which depend on marginal costs that are difficult to observe at the firm level. This paper develops methods to recover the distribution of price gaps from observed prices and shows 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 firm-level price gaps from observed prices, estimating each quote-line's desired price at the monthly frequency. 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 has consequences: the gap distribution shapes the transmission of monetary policy shocks and serves as a leading indicator for future inflation beyond what past inflation alone can predict.

We begin by laying out a canonical random menu cost model à la Nakamura and Steinsson (2010). Each firm's desired price evolves as a random walk with drift, and adjusting the actual price requires paying a fixed menu cost. With some probability, however, the firm receives a costless adjustment opportunity, introducing a time-dependent component alongside the state-dependent one. The optimal pricing policy takes the form of an Ss rule: the firm adjusts its price to a reset point whenever the gap between its actual and desired price exits an inaction region, or whenever a free adjustment opportunity arrives. Aggregating across firms, observed inflation reflects both changes in desired prices and changes in the distribution of gaps. The cross-sectional distribution of price gaps, G_t, therefore becomes the key state variable of the model: it 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 that has not yet been released through price changes. The variance measures the welfare cost of misalignment across firms. And the density near the adjustment thresholds, which we call fragility, determines how sensitive the economy is to the next shock.

These properties of the gap distribution have direct consequences for how shocks propagate through the economy. We show that economies with identical parameters but different gap distributions respond very differently to the same aggregate shock. If most firms have recently adjusted and their gaps are clustered near the reset point, even a large shock may trigger few adjustments. Conversely, if many firms have accumulated large gaps near their adjustment thresholds, even a small shock can generate a wave of price changes. This state dependence also has implications for optimal policy: the welfare cost of a given inflation rate depends on the underlying distribution that generates it, and the optimal policy response must therefore account for the shape of G_t.

The gap distribution also carries information about the future path of inflation. Firms that have accumulated large gaps will eventually reprice, and the distribution tells us how much of this adjustment is still to come. A distribution with substantial mass near the lower threshold, for instance, harbors firms whose prices have fallen behind their desired levels; as these firms catch up, inflation will rise even in the absence of new shocks. The gap distribution therefore serves as a leading indicator for inflation, and we prove that it Granger-causes observed inflation in the model.

Price gaps are not directly observable, however, which brings us to the core methodological contribution of the paper. We show that the pricing dynamics implied by the random menu cost model can be expressed in state-space form. Observed prices are the measurement variables and the state vector includes the reset price, defined as the price that would bring the gap to its optimal reset point, and an indicator for the arrival of a free adjustment opportunity. Since the bands and the reset point are not separately identified, we work with reset prices rather than desired prices directly. The *Ss* rule serves as the measurement equation, mapping past observed prices and current latent states into current observed prices. The transition equations are dictated by the model's assumptions about the evolution of desired prices and adjustment opportunities. An initial distribution over the latent states completes the specification. The goal is then 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 costless adjustment opportunity arrived, since

any such opportunity would have been taken. When prices do change, the size and direction of the adjustment are informative: if the observed change is smaller than the width of the inaction region, adjustment must have occurred without paying the menu cost, and the new price reveals the reset price exactly. If the change is large enough to span the inaction region, we know the firm crossed a threshold, and the reset price is again pinned down. These features mean that for many observations the latent state is either directly observed or tightly constrained, reducing the effective dimensionality of the filtering 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 use a Gibbs sampler. Given the choice of prior distributions, the drift, innovation variance, and adjustment-arrival parameter admit standard conjugate full conditional distributions and can therefore be updated directly. The latent states and inaction thresholds, by contrast, are updated jointly through a Metropolis-Hastings step. This joint update is motivated by the two-sided Ss rule, which makes the admissible thresholds depend tightly on the latent states and, conversely, the feasible latent-state trajectories depend on the thresholds. Candidate states are proposed from the grid-based FFBS approximation to their smoothed conditional distribution, and candidate thresholds from truncated normal random walks. Repeating these steps yields draws from the joint posterior distribution.

We apply our methodology 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 goods. The typical item exhibits monthly trend inflation of 0.2%, idiosyncratic price shocks with standard deviation around 7%, and a costless adjustment arrival rate of 14%. A key result is that 88% of price changes occur within the inaction region, suggesting that time-dependent adjustment is the dominant mode of price setting.

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. We put this object to work in two applications that demonstrate its macroeconomic content.

The first application tests directly whether the gap distribution is a source of state dependence in the transmission of monetary policy. Using state-dependent local projections with high-frequency identification, we show that the inflationary effect of an expansionary monetary shock depends strongly on the shape of the gap distribution at the time of impact. Skewness is the dominant moment: a negatively skewed distribution, with a thick left tail of firms whose prices have fallen behind, amplifies the response. Evaluated at the contrasting gap distributions observed in December 2008 and March 2021, the cumulative price response to the same shock differs by more than a factor of two. These results provide

direct empirical evidence for the state-dependence mechanism emphasized by the theory.

The second application asks whether the gap distribution serves as a leading indicator for inflation. Panel Granger-causality tests confirm the theoretical prediction: lagged moments of the gap distribution predict future inflation, but inflation does not predict the gap distribution. A horse-race forecasting regression shows that gap moments contain significant predictive content beyond past inflation, improving forecast accuracy at horizons up to six months. For policymakers and market participants, the gap distribution offers a window into the inflationary pressure already embedded in the price-setting process: even in the absence of further shocks or policy changes, the current distribution of price gaps tells us something about where inflation is heading next.

Relation to the literature This paper builds on the tradition of menu cost models (Golosov and Lucas, 2007; Nakamura and Steinsson, 2010), and in particular adopts the baseline framework of Auclert *et al.* (2024). 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, allowing the evolution of price misalignment across firms to be measured rather than inferred.

Within the menu cost literature, a growing body of work studies non-linear inflation dynamics. Early work stressed the role of shock size, showing through the selection effect 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 desired prices are unobservable.[1] A recent exception is Gagliardone *et al.* (2025), who estimate desired prices using data on costs and markups, though their focus is on the non-linear pass-through of cost shocks of different sizes. Our approach recovers price gaps from observed prices alone and studies how the state of price misalignment shapes in-

[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).

flation dynamics.

Methodologically, this paper relates to the literature on filtering and smoothing in nonlinear, non-Gaussian state-space models. Although these techniques are widely used in engineering and signal processing, their use in macroeconomics has largely been confined 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 contributes to two applied literatures by bringing the estimated gap distribution to bear on questions typically studied without it. First, we contribute to the empirical literature on the state-dependent transmission of monetary policy. While a number of papers have documented that the effects of monetary policy vary with the business cycle or with financial conditions, the mechanism through which this occurs is often left unspecified. We provide direct evidence that the cross-sectional distribution of price gaps is one such mechanism, linking the state-dependence to the microeconomics of price setting. Second, we contribute to the inflation forecasting literature, as surveyed by Faust and Wright (2013). Model-based approaches have relied on Phillips curve specifications (Stock and Watson, 1999, 2009) and DSGE models (Edge and Gürkaynak, 2010); we show that summary statistics of the estimated gap distribution serve as leading indicators that contain predictive content beyond past headline inflation.

Structure of the paper The remainder of the paper is structured as follows. Section 2 introduces a simple model of price setting with both state- and time-dependent adjustment, characterizes the distribution of price gaps and its macroeconomic relevance, and examines these properties using simulated data. Section 3 presents a state-space representation of the model, discusses the challenges posed by the forward filtering problem and how to address them, and outlines a Bayesian approach to estimate both parameters and latent states. Section 4 applies the method to micro-price data from the UK Consumer Price Index. Section 5 puts the estimated gap distribution to work, testing for state-dependent shock propagation and assessing the predictive content of gap moments for future inflation. Section 6 concludes.

[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. Some of these approaches are unsuitable in our context due to differentiability requirements (e.g., extended Kalman filters), while others are computationally infeasible at the scale required to reconstruct the full CPI (e.g., particle filters).

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, commonly known as a "CalvoPlus" model, in discrete time, following Auclert *et al.* (2024) but allowing for trend inflation. Microfoundations for this framework are provided in Appendix A.1. 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 desired log price, denoted $p^*_{i,t}$, evolves according to a random walk with drift:

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

where μ_t is the trend inflation rate and $\varepsilon_{i,t} \sim N(0, \sigma^2_\varepsilon)$ 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 quadratic loss of $\frac{1}{2}(p_{i,t} - p^*_{i,t})^2$ from deviating from its desired log 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 $x_{i,t} \equiv p_{i,t} - p^*_{i,t}$. The firm chooses a sequence $\{x_{i,t}\}_{t=0}^{\infty}$ to minimize expected discounted losses:

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

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 $x_{i,t-1} - \mu_t - \varepsilon_{i,t}$.[3]

[3] As derived in Appendix A.1, the structural model features a time-varying, firm-specific weight $B_t(i) > 0$ that multiplies both the flow loss and the adjustment cost. When menu costs are proportional to the firm's desired profits, this weight cancels from the decision problem up to an effective discount factor $\tilde{\beta}_t = \beta \cdot B_{t+1}(i)/B_t(i)$. At monthly frequency and with slowly changing aggregate conditions, $\tilde{\beta}_t \approx \beta$, so that (2) is the relevant reduced-form problem. This is the standard formulation in the analytical menu cost literature, e.g., Alvarez and Lippi (2009), Alvarez *et al.* (2016), and Auclert *et al.* (2024).

The optimal pricing policy takes the form of a standard *Ss* rule:[4]

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

The firm adjusts its price in one of two cases: either it receives a free adjustment opportunity, or the price gap $x_{i,t}$ falls outside an inaction region $[\underline{x}_t, \bar{x}_t]$. Upon adjustment, the firm resets its gap to x_t^*. Otherwise, the gap evolves passively, drifting with trend inflation and the idiosyncratic shock. In general, the reduced-form parameters $[x_t^*, \underline{x}_t, \bar{x}_t]$ 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_t^*. 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 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 x_t^*, 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 the reset point 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 desired 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.[5] The *skewness* captures directional asymmetry in pricing pressure, while the *kurtosis* reflects the heaviness of the tails, i.e., how much mass sits near the adjustment thresholds rather than near the center of the inaction region.

[4] 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.

[5] 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.

Skewness and kurtosis both speak to the tail behavior of G_t: how much mass sits near the adjustment thresholds, and on which side. We define a scalar that captures the net pricing pressure from firms near the boundaries, which we term the *net fragility index*:

$$\mathcal{F}_t \equiv g_t(\underline{x}_t) \cdot (x_t^* - \underline{x}_t) \; - \; g_t(\bar{x}_t) \cdot (\bar{x}_t - x_t^*). \tag{5}$$

Each term weights the density near a threshold by the size of the price adjustment a firm crossing it would make: the first captures inflationary pressure from firms near $\underline{x}_t$ whose prices have fallen behind, the second captures deflationary pressure from firms near $\bar{x}_t$. When $\mathcal{F}_t > 0$, the economy harbors more latent upward than downward pricing pressure; when $\mathcal{F}_t < 0$, the reverse holds.

From price gaps to inflation. Each firm's inflation rate decomposes as

$$\pi_{i,t} = \pi_{i,t}^* + \Delta x_{i,t}, \tag{6}$$

where $\pi_{i,t}^* \equiv p_{i,t}^* - p_{i,t-1}^* = \mu_t + \varepsilon_{i,t}$ is the rate of price change that would prevail absent menu costs—what we call *frictionless 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

$$\pi_t = \mu_t - \Delta E[x_t], \tag{7}$$

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

Equation (7) 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 desired 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 desired levels. The two economies look identical through the lens of π_t, but differ sharply in their welfare costs (4), their fragility to subsequent shocks (5), and the directional pressure embedded in higher moments of G_t. The gap distribution thus provides 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 unexpected 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. Crucially, the two interact—the per-unit response to a given shock depends on the shape of G_t when it hits. To illustrate both we use numerical simulations from the model.

We calibrate parameters to match moments from UK CPI micro data reported by Blanco (2021): a monthly frequency of price change of approximately 12.6%, a mean absolute price change of approximately 15.3%, and a kurtosis of price changes of 3.8. Parameter values are reported in Table A.1 in the appendix. The experiment is as follows. We subject the economy to an unexpected, one-off level shock to the drift: $\mu_1 = \mu + \Delta\mu$, with $\mu_t = \mu$ for all $t \geq 2$. We measure the *impact excess inflation*, $\pi_1 - \mu$, defined as the deviation of first-period inflation from the pre-shock trend.

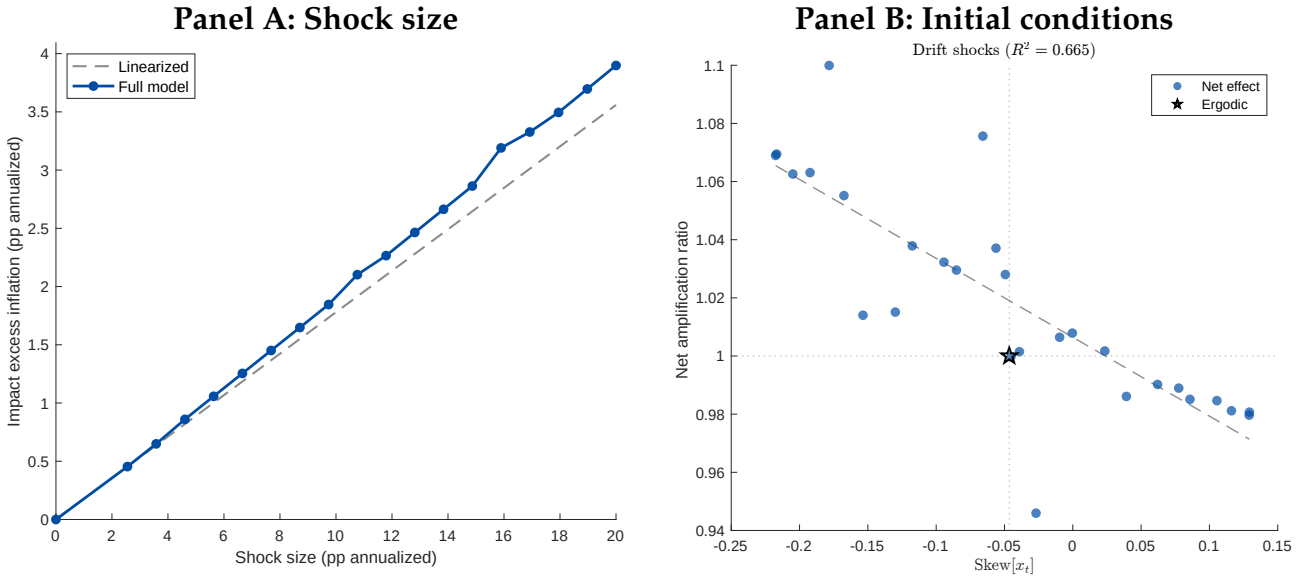
Panel A of Figure 1 isolates the shock-size channel. Starting from the ergodic distribution, we vary the magnitude of the shock and plot the resulting impact excess inflation. The dashed line shows a linearized benchmark, extrapolated from the slope at small shocks. The relationship is convex: the impact 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.

Panel B turns to the less-studied dimension. We fix the shock at 4 pp annualized—well within the approximately linear region of Panel A, so that the shock-size non-linearity is negligible—and vary only the initial gap distribution G_0. To generate economically meaningful distributions, rather than constructing them arbitrarily, we subject the *same economy* to different histories of drift shocks and record the resulting distribution at various points along the transition path back to the ergodic. These distributions differ in shape, reflecting the accumulated effects of past shocks, while the structural parameters of the model remain identical throughout.

We then hit each distribution with the same 4 pp shock and plot the *net amplification ratio*: the impact excess inflation minus the inflation that the displaced distribution would have generated even without the shock, normalized by the ergodic response. This strips out the mechanical unwinding of pent-up pressure and isolates the pure interaction between the shock and the initial conditions—how much the distribution amplifies or dampens the shock itself. The ergodic point (marked with a star) equals one by construction, since the ergodic distribution has no pent-up pressure to unwind.

In this exercise, where the background histories involve drift shocks, the relevant sum-

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 4 pp annualized, we vary the initial gap distribution G_0 (x-axis). The y-axis shows the net amplification ratio: impact excess inflation minus the no-shock counterfactual, normalized by the ergodic response. The star marks the ergodic distribution.

mary statistic turns out to be the mean price gap $E[x_t]$: the OLS fit shown in the figure is nearly perfect. Intuitively, a more negative mean gap signals more firms with prices below their desired level, creating pent-up inflationary pressure that amplifies a positive drift shock. For other types of disturbances that alter the shape of the distribution without displacing the mean, such as changes in idiosyncratic volatility, other moments become the relevant predictor. We present this exercise in Appendix A.5, where the variance ratio replaces the mean gap and explains the vast majority of the variation in the amplification ratio.

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): because 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—well within the approximately linear region of Figure 1, Panel A—the inflation response varies substantially with the initial gap distribution G_0 (Panel B). 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.[6]

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 (7), 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, the history $\{\pi_s\}_{s \le t}$ generically cannot recover the current shape of G_t. This suggests that the gap distribution Granger-causes inflation; a prediction we formalize below and test empirically in Section 5.

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 x^*.

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 $h = 1, 2, \ldots$,

$$\pi_{t+h} \;=\; (x^* - \underline{x})\;\mu\; g_t\big(\underline{x} + (h-1)\mu\big) \;+\; O(\mu^2). \tag{8}$$

(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

[6] A full characterization of optimal policy in this setting requires solving a Ramsey problem in which the planner's choices, the firm's pricing rule, and the distribution co-evolve—a computationally demanding exercise. See Karadi *et al.* (2025) for a quantitative treatment in a calibrated menu cost model with general equilibrium. Formalizing how initial conditions shape the optimal policy problem is work in progress.

at some horizon $h \geq 1$.[7]

The proof is in Appendix A.6. 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 threshold and adjust h periods from now. 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.

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 implied by the random menu cost model in Section 2 in state-space form. We then develop a Bayesian estimation procedure to jointly estimate the model parameters and latent states. The section concludes with a Monte Carlo experiment that assesses the finite-sample performance of the algorithm.

3.1 Random menu cost model in state-space form

To cast the pricing model of Section 2 in state-space form, we reparameterize the *Ss* rule and introduce the notation and assumptions used for estimation.

Notation and assumptions. Since the reset point x^* cannot be separately identified from the thresholds $(\underline{x}, \bar{x})$, we work with recentered quantities. Define the *reset price* as $p^r_{i,t} \equiv p^*_{i,t} + x^*$ and the *recentered inaction region* as the interval $(\underline{s}, \bar{s})$, where $\underline{s} \equiv \underline{x} - x^*$ and $\bar{s} \equiv \bar{x} - x^*$. Henceforth, *price gap* refers to $p_{i,t} - p^r_{i,t}$, and *inaction region* to $(\underline{s}, \bar{s})$. Let $\ell_{i,t}$ denote an indicator equal to one when firm i receives a free adjustment opportunity. The vector of latent states is $\mathbf{x}_{i,t} \equiv [p^r_{i,t}, \ell_{i,t}]$, and the parameter vector is $\boldsymbol{\theta} = \{\mu, \sigma_\varepsilon, \lambda, \underline{s}, \bar{s}\}$.

[7] 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.

We estimate the model under the assumption that $\boldsymbol{\theta}$ is common across firms and constant within each product-period estimation window.[8] The state-space representation consists of a measurement equation, two transition equations, and an initial condition.

Measurement equation. The two-sided *Ss* pricing rule (3) implies

$$p_{i,t} = p_{i,t-1}\, d(\mathbf{x}_{i,t}, p_{i,t-1}, \boldsymbol{\theta}) + p^r_{i,t}\ (1 - d(\mathbf{x}_{i,t}, p_{i,t-1}, \boldsymbol{\theta}))\,. \tag{9}$$

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

$$d(\mathbf{x}_{i,t}, p_{i,t-1}, \boldsymbol{\theta}) = \mathbb{1}\{p_{i,t-1} - p^r_{i,t} \in [\underline{s}, \bar{s}]\}(1 - \ell_{i,t}) + \mathbb{1}\{p_{i,t} = p^r_{i,t}\}\, \ell_{i,t}. \tag{10}$$

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

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

$$p^r_{i,t} = \mu + p^r_{i,t-1} + \varepsilon_{i,t}, \tag{11}$$

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

$$\ell_{i,t} = \mathbb{1}\{\nu_{i,t} \leqslant \lambda\}, \tag{12}$$

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

Initial distribution for the states. The initial firm-level states are assumed to be independent across firms:

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

A specific functional form for $f(\mathbf{x}_{i,1} \mid p_{i,1}, \boldsymbol{\theta})$ is introduced in Section 3.2.2.

Equations (9)–(13) define a nonlinear, non-Gaussian state-space representation of the firm-level pricing dynamics described in Section 2. This representation is semi-structural: the measurement equation is derived directly from the optimal *Ss* rule, whereas the band parameters $\{\underline{s}, \bar{s}\}$ are reduced-form objects, each given by a function of deeper structural primitives such as preferences, technology, and menu costs.

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

3.2 Bayesian estimation

We now turn to the Bayesian estimation procedure. For expositional convenience, we begin by describing how to sample the latent states conditional on the model parameters and the observed data. We then use this conditional sampling step to construct a Gibbs sampler for the joint posterior distribution of the parameters and latent states.

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. In the random menu cost model, this step is challenging because the state-space representation is nonlinear and non-Gaussian, and because 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* (9)–(13)*, the forward-filtering backward-sampling recursions factorize across price trajectories.*

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

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

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

with trajectory-level recursions recursions given by:

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

$$f\left(\mathbf{x}_{i,t} \mid \mathbf{p}_{i,1:t}, \boldsymbol{\theta}\right) \propto f\left(p_{i,t} \mid \mathbf{x}_{i,t}, p_{i,t-1}, \boldsymbol{\theta}\right) f\left(\mathbf{x}_{i,t} \mid \mathbf{p}_{i,1:t-1}, \boldsymbol{\theta}\right). \tag{17}$$

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

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

and each trajectory is sampled according to the backward recursion:

$$f\left(\mathbf{x}_{i,1:T} \mid \mathbf{p}_{i,1:T}, \boldsymbol{\theta}\right) = f\left(\mathbf{x}_{i,T} \mid \mathbf{p}_{i,1:T}, \boldsymbol{\theta}\right) \prod_{t=1}^{T-1} f\left(\mathbf{x}_{i,t} \mid \mathbf{x}_{i,t+1}, \mathbf{p}_{i,1:t}, \boldsymbol{\theta}\right), \tag{19}$$

$$f\left(\mathbf{x}_{i,t} \mid \mathbf{x}_{i,t+1}, \mathbf{p}_{i,1:t}, \boldsymbol{\theta}\right) \propto f\left(\mathbf{x}_{i,t+1} \mid \mathbf{x}_{i,t}, \boldsymbol{\theta}\right) f\left(\mathbf{x}_{i,t} \mid \mathbf{p}_{i,1:t}, \boldsymbol{\theta}\right). \tag{20}$$

Proof. See Appendix B.1. □

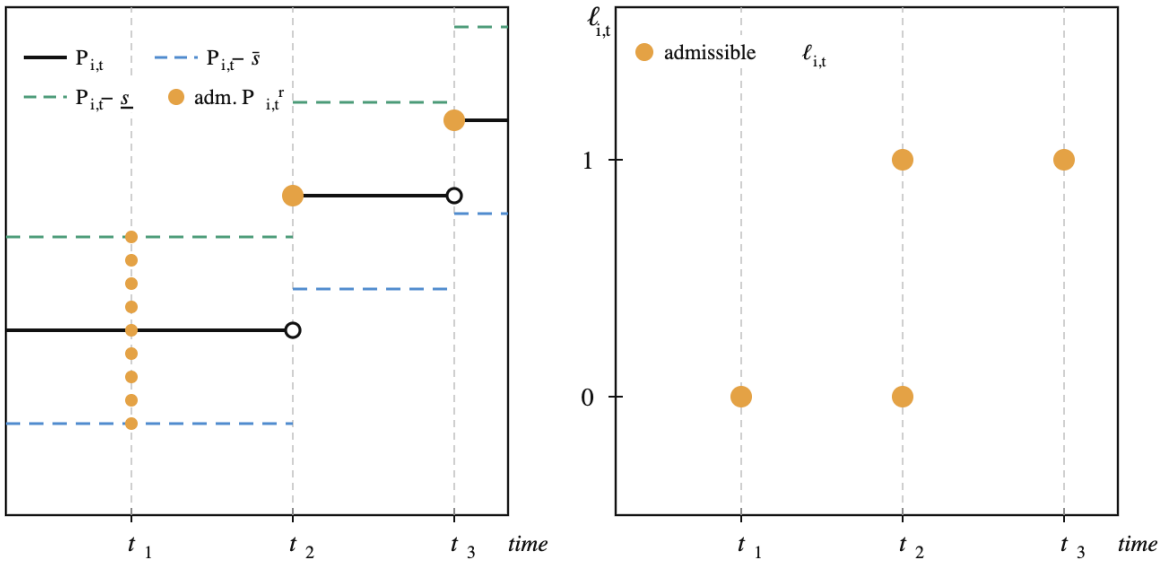
In words, the lemma shows that the joint filtered, predictive, and smoothed densities factorize into products of trajectory-level densities. This result follows from two basic facts. First, under the two-sided *Ss* rule summarized by (9)–(10), each observed price is determined solely by its own lagged price, latent state, and the model parameters. The measurement equation therefore links observables to latent states trajectory by trajectory. Second, under the assumptions in (11)–(13), latent states are initialized independently and evolve independently thereafter. From a computational standpoint, the implication of Lemma 1 is that latent-state sampling can be carried out trajectory by trajectory rather than jointly over the full panel. This reduces the problem to a collection of low-dimensional state-space systems that can also be processed in parallel.

Given Lemma 1, latent-state sampling reduces to solving the trajectory-level FFBS recursions in (16), (17), (19), and (20). 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 this intuition by showing how a hypothetical price trajectory restricts the set of admissible latent states over time.

Panel (a) plots the admissible values of the latent reset price $p^r_{i,t}$ for the hypothetical observed price trajectory shown by the black solid line, while panel (b) plots the corresponding admissible values of the free-adjustment indicator $\ell_{i,t}$. 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 band, so $p^r_{i,t} \in [p_{i,t} - \bar{s},\, p_{i,t} - \underline{s}]$ and $\ell_{i,t} = 0$.[9] If there is an adjustment from outside 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

[9] We ignore the case of a free-adjustment opportunity when the reset price coincides exactly with the previous observed price, since under a continuous distribution for the reset price this is a probability-zero event.

Figure 2: Admissible latent states given an observed price trajectory



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. Full details are given in Appendix B.2.

3.2.2 Gibbs sampler for the joint posterior

We now turn to posterior simulation of the full model. Building on the latent-state sampler described above, we construct a Gibbs sampler for the joint posterior distribution of the model parameters and latent states.

Priors. Table 1 reports the priors for the model parameters and latent states.

We choose prior functional forms to obtain conjugate full conditionals whenever possible, and set hyperparameters so that the resulting priors are weakly informative. Two components of the prior specification are instead tied more directly to the data. The first is the prior support for the inaction bands, which we bound using the range of observed price changes. This choice reflects an identification consideration: the likelihood only identifies the bands 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. We therefore use the prior to rule out regions of the band parameter space that are not identified by the data.[10] The second is the prior for the initial latent states. Because at the initial time period we do not observe whether the current price results from an adjustment, we assume that the initial price gap

[10] 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.

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 $(\underline{s}_{\text{LB}}^0,\, \underline{s}_{\text{UB}}^0)$	$\underline{s}_{\text{LB}}^0 = \min - \Delta\mathbf{p}^+ \,, \underline{s}_{\text{UB}}^0 = \max - \Delta\mathbf{p}^+$
$\bar{s}$	Uniform $(\bar{s}_{\text{LB}}^0,\, \bar{s}_{\text{UB}}^0)$	$\bar{s}_{\text{LB}}^0 = \min - \Delta\mathbf{p}^- \,, \bar{s}_{\text{UB}}^0 = \max - \Delta\mathbf{p}^-$

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.

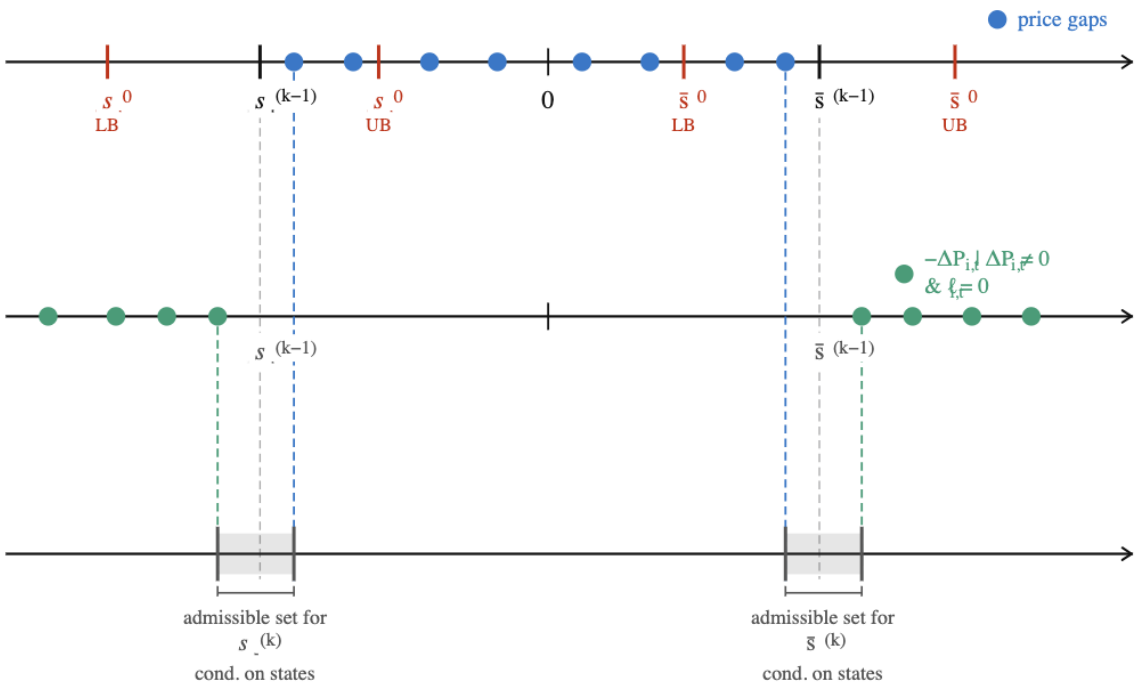
is closed with probability equal to the empirical frequency of observed price changes in the data. Conditional on a closed initial gap, we assign a 60% probability that this reflects a free-adjustment opportunity, in line with the average degree of Calvoness estimated by Gautier and Le Bihan (2022) using French micro price data.

Full conditionals. Given the priors in Table 1, 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, namely the latent states and the inaction bands, do not admit closed-form full conditionals. We draw them jointly from their full conditional distribution using a Metropolis-Hastings step within the Gibbs. The reason for this joint update is the tight link imposed by the measurement equations (9)–(10) under the two-sided *Ss* rule, illustrated in Figure 3.

For a given Gibbs iteration, conditioning on the current latent states restricts the set of admissible band values, since alternative values would imply zero likelihood. Conversely, conditioning on the current bands restricts the admissible latent-state trajectories. As a result, separate updates can lead to poor mixing of the Markov chain. For the Metropolis-Hastings step, we propose latent states from their full conditional, that is, the smoothed distribution approximated via the grid-based FFBS procedure described in the previous section. The proposal for the inaction bands is given by truncated normal random walks centered at the previous draw, with support equal to that of the prior and with standard

Figure 3: Bands update conditional on latent states



deviations tuned during burn-in via a Robbins-Monro stochastic approximation algorithm (Robbins and Monro, 1951), targeting an acceptance rate of 33% in line with the optimal scaling results of Roberts and Rosenthal (2001). The full Gibbs sampler is described in detail in Appendix B.3.

3.3 Monte Carlo

To conclude this section, we use a Monte Carlo experiment to assess the finite-sample properties of the estimator and the performance of the proposed algorithm.

3.3.1 Design

To calibrate the Monte Carlo design with empirically plausible parameter values, we use the Nakamura and Steinsson (2010) calibration reported in Auclert *et al.* (2024). We first map their reported structural parameters into the parameterization of the price-setting model described in Section 2.1.[11] Based on this mapping, we solve the model numerically to obtain the full set of parameters entering the state-space representation in Section 3.1, which we use as the data-generating process in the simulations.

For this calibration, we consider two alternative sample sizes, $(N, T) = (100, 24)$ and

[11] Appendix B.4 provides a detailed description of the mapping between the parameterizations.

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 250 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.

$(N, T) = (100, 60)$, which we refer to as the small- and large-sample designs, respectively. These sample sizes correspond approximately to the 10th and 20th percentiles of the panel dimensions observed in the UK micro price data used in the empirical analysis. For each sample size, we generate 200 Monte Carlo replications.

3.3.2 Results

For each Monte Carlo sample, we estimate the model parameters and latent states using the Gibbs sampler described in the previous subsection and compute three performance measures: (i) posterior mean bias, (ii) posterior standard deviation, and (iii) an indicator for whether the true parameter value lies within the 95% posterior credible interval. Table 2 reports results averaged across Monte Carlo replications.

Panel A summarizes the finite-sample performance of the estimator for the structural parameters. Posterior mean bias is negligible across all parameters and declines further as sample size increases, indicating accurate parameter recovery even in relatively small sam-

ples. Posterior dispersion decreases systematically with sample size, consistent with improved identification and increasing posterior concentration around the true parameter values. Empirical coverage of the 95% posterior credible intervals is close to the nominal level for most parameters, suggesting well-calibrated posterior uncertainty.

Panel B reports the same performance measures for objects derived from price gaps, including individual price gaps and moments of the cross-sectional distribution of price gaps averaged over time. The results closely parallel those in Panel A and provide evidence of posterior contraction around the true data-generating values. Posterior mean bias is negligible across all reported quantities, and posterior dispersion declines with sample size, indicating increasing concentration of the posterior distribution. As expected, posterior uncertainty is substantially larger for individual price-gap trajectories than for cross-sectional moments, which are more tightly identified by the data. Coverage rates are close to the nominal level for all reported moments, suggesting that posterior uncertainty is well calibrated.

Appendix B.4 presents additional Monte Carlo experiments that further assess the robustness of our results. These include post-estimation Bayesian diagnostics, simulations under data-generating processes with lower and higher degrees of Calvoness, and experiments based on pure menu-cost and pure Calvo models. We also study the sensitivity of the results to key algorithmic hyperparameters. Taken together, these additional exercises confirm the stability of the proposed estimation procedure across a wide range of model environments and numerical settings.

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—e.g., 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, record-

ing over 100,000 prices using handheld devices. Our baseline sample consists of these locally collected price quotes from 1996m1 to 2023m12.[12]

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.[13]

In the second stage, since the model in Section 2 does not feature sales or product substitutions, 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 32 million price quote observations, more than 2.2 million quote-lines, and 1317 unique items.

4.2 Posterior estimates: model parameters

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.

Results reveal substantial variation across items in the inaction region bounds, frictionless inflation trends, shock volatility, and adjustment intensities. A key dimension of interest is the asymmetry in the inaction region. A simple measure of asymmetry is given by the sum of the estimated upper and lower thresholds, $\mathcal{A} = \hat{\bar{s}} + \hat{\underline{s}}$. This statistic ranges from -171% to 206% in the data, with 58% of items displaying negative asymmetry.[14] A natural explanation for asymmetric pricing policies is trend inflation: in a menu cost model with positive inflation, price increases tend to be larger than decreases (Ball and Mankiw, 1994). Consistent with this, the asymmetry measure correlates negatively with the estimated drift in frictionless inflation (correlation = -0.41, Figure 4, top-left panel). However, this mechanism does not fully explain the observed dispersion in asymmetry. Other potential drivers

[12] 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.

[13] 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.

[14] 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).

Table 3: Summary statistics for estimated parameters

	$\hat{\underline{s}}$	$\hat{\bar{s}}$	$\hat{\mu}$	$\hat{\sigma}_\varepsilon$	$\hat{\lambda}$
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.

include asymmetric menu costs or profit losses.[15]

The distribution of estimated frictionless inflation trends $\hat{\mu}$ ranges from -1.4% to 1.2% per month, with a median of 0.2%. This is consistent with average monthly CPI inflation of 0.17% over the sample period. These estimated trends are positively correlated with the average size of price changes across items (correlation = 0.63; Figure 4, top-right panel). Roughly one-quarter of items display negative inflation trends, consistent with individual items becoming cheaper over time even as aggregate prices increase. These are concentrated in categories such as recreation and culture, household equipment, and clothing and footwear.[16]

Cross-sectional variation in the dispersion of price changes is primarily driven by differences in the estimated volatility of idiosyncratic shocks, $\hat{\sigma}_\varepsilon$. The correlation between this parameter and the observed standard deviation of price changes is 0.86 (Figure 4, bottom-left panel). The median estimated volatility is 7.2%, which aligns well with results reported in Gautier and Bihan (2018) for French CPI data (5–9%, depending on model structure).

The probability of a costless adjustment opportunity, $\hat{\lambda}$, accounts for much of the heterogeneity in price change frequency. Across items, $\hat{\lambda}_j$ correlates with observed frequency at 0.91 (Figure 4, bottom-right panel), underscoring that in the UK data, most price changes are estimated to occur due to the arrival of costless opportunities.

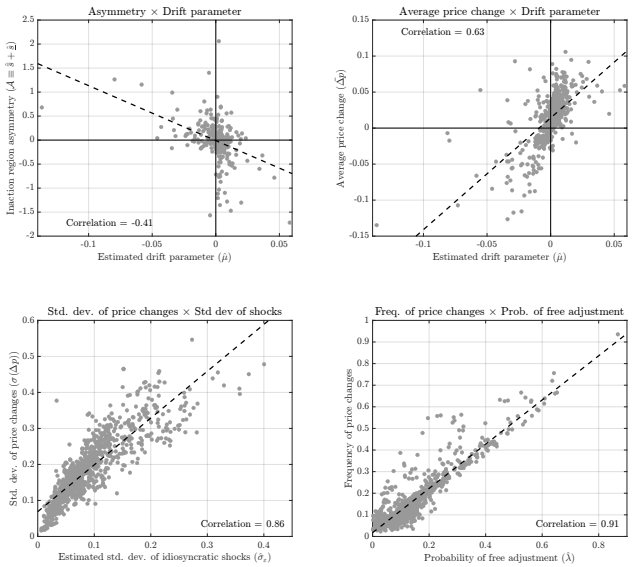
To quantify the relative importance of state-dependent and time-dependent adjustments, we compute the share of price changes that occur inside the inaction region—that is, not triggered by boundary crossing but by costless adjustment opportunities.[17] For the full sample,

[15]Symmetric pricing policies often arise in models with zero trend inflation and quadratic losses (Dixit, 1991), though these assumptions appear inconsistent with observed UK pricing behavior.

[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 $\mathcal{C} = \sum_i \sum_t \mathbb{1}\{\Delta p_{i,t} \in (-\hat{\underline{s}}, 0) \cup$

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.

this share is 88%, with item-level values ranging from 7.5% to 100% and an average of 87%. Across COICOP divisions, the share ranges from 66% (restaurants and hotels) to 94% (food and non-alcoholic beverages). These results are in line with prior work: Nakamura and Steinsson (2010) report a share of 75% for US data; Gautier and Bihan (2018), 80% for France; and Blanco (2017), 93% for the UK when including fat-tailed shocks. Despite methodological differences, the broad agreement across studies suggests that most price changes observed in consumer microdata are time-dependent in nature.

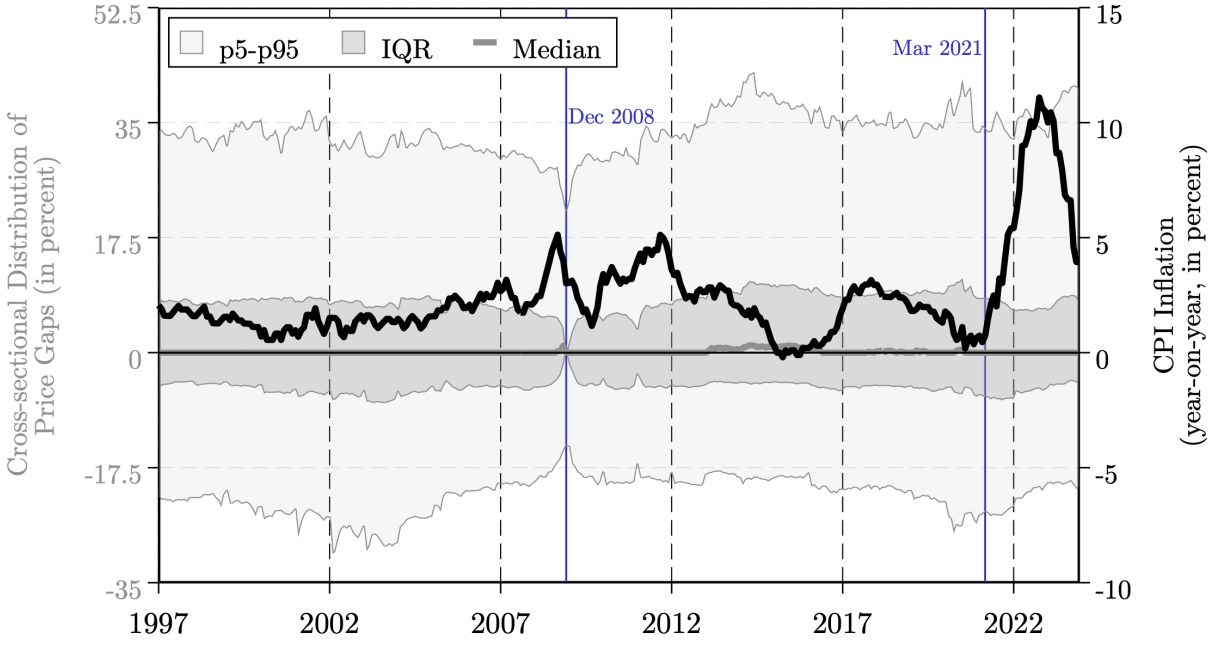
4.3 The estimated distribution of price gaps

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. To construct an aggregate distribution, we pool these gaps across items. Because the reset point x^* is not identified—only the gap $x_{i,t}$ is recovered—and differs across items, we demean within each item before pooling: the contribution of quote-line i in item j at date t is $x_{i,t} - \bar{x}_j$, where $\bar{x}_j$ is the time-series mean of gaps within item j. This removes the unidentified item-specific location shift while preserving the dispersion, shape, and time variation that are our objects of interest. We then weight each quote-line by its CPI weight, so that the aggregate distribution is representative of the price index (see Appendix C.2 for details on the weight construction). We denote this aggregate distribution G_t throughout.

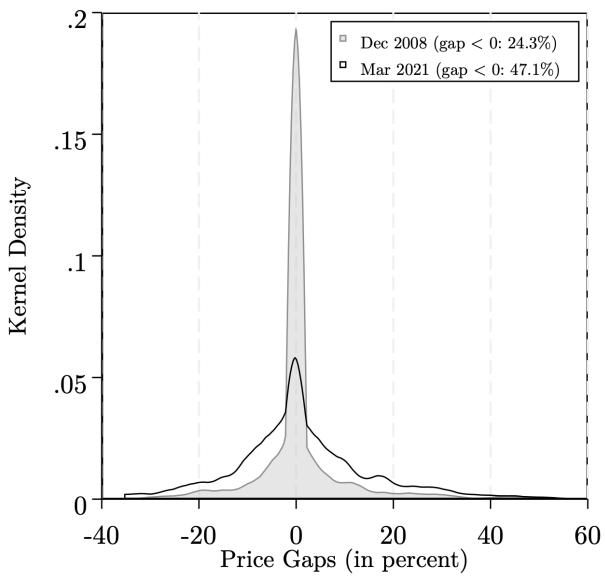
Panel (a) of Figure 5 plots the evolution of the cross-sectional distribution of price gaps

$(0, -\hat{\underline{s}})\} \,/\, \sum_i \sum_t \mathbb{1}\{\Delta p_{i,t} \neq 0\}$.

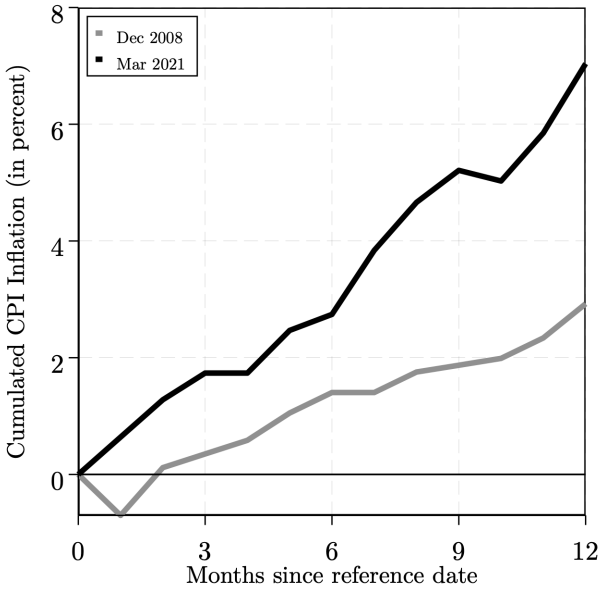
Figure 5: Price gaps and inflation over time



(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.

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 price-gap distribution, while the thick gray line reports its median.

The figure 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 persistently and take time to revert as firms gradually adjust their prices.

Even from this bird's-eye view, the variation appears systematic. Episodes in which the distribution widens, indicating a growing mass of firms whose prices have drifted from their desired levels, tend to precede large movements in inflation. To illustrate this more concretely, panels (b) and (c) zoom into two dates: December 2008 and March 2021. In December 2008, in the aftermath of the temporary VAT cut introduced during the global financial crisis, the distribution was tightly concentrated around zero, consistent with a synchronized wave of price adjustments. In March 2021, one year into the pandemic and amid rising input costs and supply-chain disruptions, the picture was very different: the distribution had substantially wider tails, with nearly 48 percent of price setters exhibiting negative price gaps, indicating that a growing mass of firms had yet to adjust to the buildup in cost pressures.

These two shapes correspond to strikingly different subsequent inflation outcomes: panel (c) shows that cumulative inflation over the following 12 months was nearly three times as large after March 2021 as after December 2008. While this comparison is descriptive rather than causal, it illustrates the kind of information embedded in the shape of G_t that headline inflation alone does not capture. In the next section, we show more formally how the estimated gap distribution helps decompose, explain, and forecast inflation dynamics.

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

A key implication of the theory is 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

$$\log P^c_{t+h} = \beta_h \Delta i_t + \sum_{j=1}^{4} \gamma_{j,h}\, \Delta i_t\, m^c_{j,t-1} + \mathbf{w}_t' \Gamma_h + e_{t+h}, \tag{21}$$

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.[18] The interaction 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. These three shocks, together with their interactions with the lagged gap moments, serve as instruments for Δi_t and its interactions.

Table 4 reports the estimated coefficients. Among the individual moments, skewness stands out: 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. This is natural in the context of the *Ss* model: skewness captures the balance between inflationary and deflationary pressure stored in the tails, and it is the firms closest to their adjustment thresholds that are 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

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

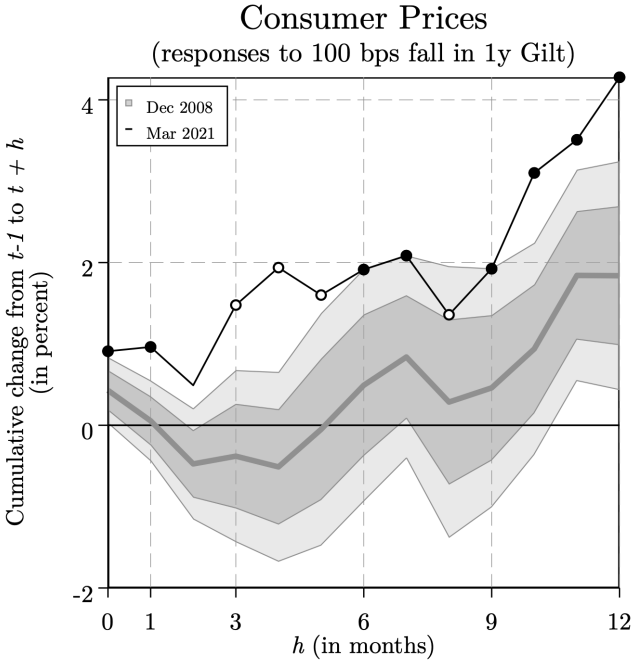
conditions at a 12-month horizon. Under December 2008 conditions, the response is imprecisely estimated throughout, consistent with a concentrated distribution near the reset point dampening the transmission of aggregate shocks.

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 fundamental 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 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. The distribution should there-

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.

fore serve as a leading indicator, and we now test whether it does. We proceed in two steps: a panel Granger-causality test establishes the direction of predictability, and a forecasting exercise quantifies the gain.

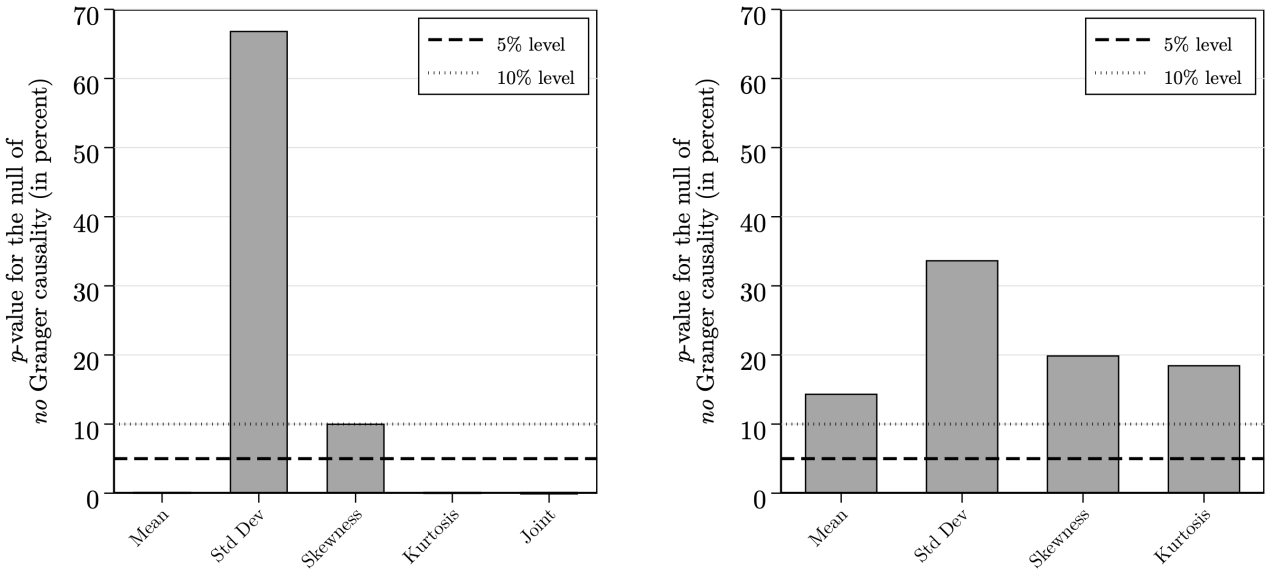
Granger causality. To test whether movements in the price-gap distribution precede inflation, we exploit both time-series and cross-sectional variation through panel Granger-causality tests. For each COICOP class c, let π_t^c denote monthly inflation and let $\mathbf{m}_t^c$ denote a vector containing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

$$\mathbf{A}(L)\mathbf{y}_t^c = \boldsymbol{\alpha}_c + \boldsymbol{\delta}_t + \mathbf{e}_t^c, \tag{22}$$

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.

The results, reported in Figure 7, reveal a clear asymmetry. Lagged moments of the gap distribution predict future inflation (panel a): 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. But inflation does not predict the gap distribution (panel b): none of the moments is Granger-caused by inflation at conventional significance levels. This one-directional pattern is exactly what the theory implies: the gap

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 (22), 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.

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 that nests lagged inflation and gap moments as competing predictors:

$$\pi^c_{t,t+h} = \alpha_c + \delta_t + \beta_h \, \pi^c_{t-h,t} + \sum_{j=1}^{4} \gamma_{j,h} \, m^c_{j,t} + v^c_{t+h}, \tag{23}$$

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. The coefficients carry the expected signs. Lagged inflation enters negatively, consistent with mean reversion. The mean gap coefficient is also negative: a more negative mean gap, indicating that firms' prices have on average fallen further behind their desired levels, predicts higher subsequent infla-

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 (23) 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.

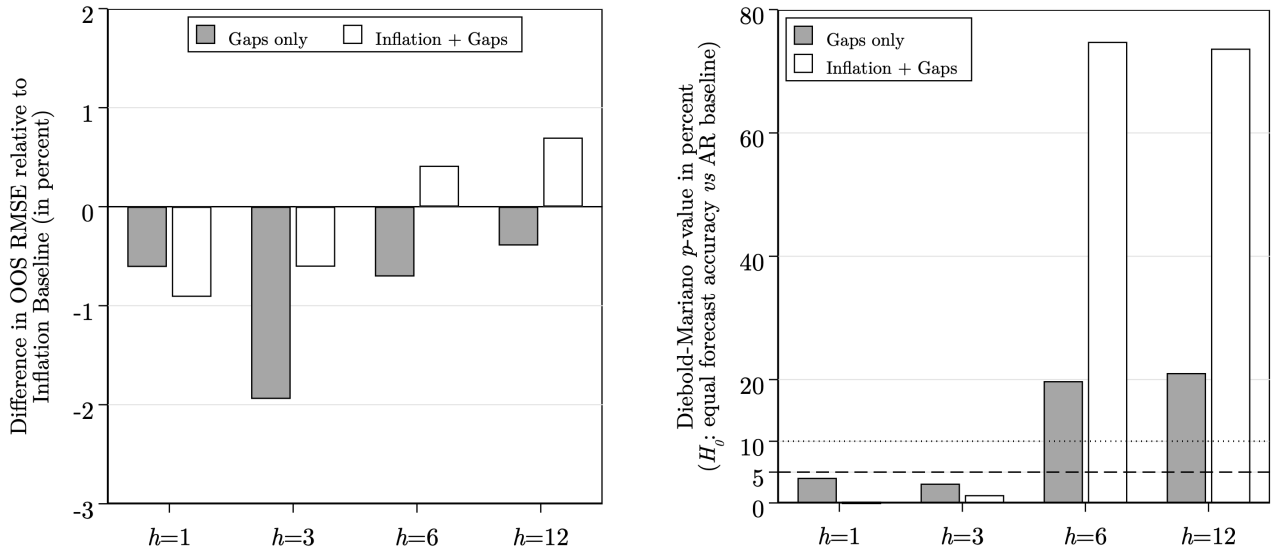
tion. Skewness reinforces this pattern, with a negative and significant coefficient at all horizons, consistent with its role in capturing directional pricing pressure. Kurtosis is significant at the shorter horizons while the standard deviation does not enter significantly. The bottom panel confirms what matters most: the joint F-test for the gap moments rejects at horizons up to 6 months ($p < 0.01$), indicating that the distribution contains 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 (23) using only data available up to that date and produce forecasts at horizons of 1, 3, 6, and 12 months.[19]

Figure 8 summarizes the results. Panel (a) shows that a model using price-gap moments

[19] 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 (23) 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).

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, though these are concentrated at shorter horizons. Panel (b) reports Diebold–Mariano test statistics: the improvements are statistically significant at the 1-month and 3-month horizons. That the gains fade at longer horizons is 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.

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. For policymakers and market participants, the gap distribution offers a window into the inflationary pressure already embedded in the price-setting process: even in the absence of further shocks or policy changes, the current distribution of price gaps tells us something about where inflation is heading next

6 Conclusion

This paper has developed methods to estimate the cross-sectional distribution of price gaps from micro price data and shown that it has substantial macroeconomic content. Using state-dependent local projections, we found that the gap distribution shapes the transmission of monetary policy shocks: the same expansionary shock that produces a large inflationary response when the distribution is dispersed and asymmetric has a negligible effect when it is concentrated near the reset point. The gap distribution also serves as a leading indicator for inflation, with significant predictive content beyond past inflation at horizons up to six months. For policymakers, the estimated gap distribution offers a real-time reading of the inflationary pressure already embedded in the price-setting process and of the economy's fragility to fundamental shocks.

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. (2017). Optimal inflation target in an economy with menu costs and a zero lower bound.

— (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.

DIXIT, A. K. (1991). Analytical approximations in models of hysteresis. *Review of Economic Studies*, **58** (1), 141–151.

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.

EDGE, R. and GÜRKAYNAK, R. S. (2010). How useful are estimated dsge model forecasts for central bankers. *Brookings Papers on Economic Activity*, **2010** (2), 209–244.

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 using 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.

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.

STOCK, J. H. and WATSON, M. W. (1999). Forecasting inflation. *Journal of monetary economics*, **44** (2), 293–335.

— and — (2009). Phillips curve inflation forecasts. In J. Fuhrer, Y. Kodrzycki, J. Little and G. Olivei (eds.), *Understanding Inflation and the Implications for Monetary Policy*, Cambridge, MA: MIT Press, pp. 101–186.

Appendix to

Price Gaps and Inflation Dynamics

by Miguel Bandeira, Laura Castillo-Martínez and Shiyuan Wang

A Model Appendix

A.1 Microfoundations

This appendix presents the microfoundations underlying the reduced-form model of price setting in Section 2.1. We describe the household and firm problems, derive the law of motion for the desired price, and provide the aggregation results that connect individual pricing decisions to aggregate inflation.

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 function for each variety:

$$W_t = M_t, \tag{24}$$

$$c_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-\epsilon} C_t, \tag{25}$$

where (24) 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 (24). The static profit-maximizing (desired) price is:

$$P_t^*(i) = \frac{\epsilon}{\epsilon - 1} \cdot (1 - \phi_t) \cdot \frac{M_t}{A_t(i)}, \tag{26}$$

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

Desired price dynamics. Taking logs of (26) 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 as the sum of the cost-push shock (in logs) and the growth in nominal spending, $\mu_t \equiv \Delta \log(1 - \phi_t) + \Delta \log M_t$ and the reduced-form idiosyncratic shock as $\varepsilon_{i,t} \equiv -a_{i,t}$. Then the above equation becomes:

$$p_t^*(i) = \mu_t + p_{t-1}^*(i) + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \tag{27}$$

where $\sigma_\varepsilon = \sigma_a$. This yields the law of motion for desired prices in (1).

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 (25), the production technology, and the desired price (26), the firm's normalized real profit in period t can be decomposed as:

$$\frac{\Pi_t(i)}{P_t C_t} = B_t(i) \cdot \tilde{\pi}(x_{i,t}), \tag{28}$$

where $x_{i,t} \equiv p_{i,t} - p_{i,t}^*$ is the log price gap,

$$\tilde{\pi}(x) \equiv e^{(1-\epsilon)x} - \frac{1}{\mathcal{M}} e^{-\epsilon x} \tag{29}$$

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

cision.[20]

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 $\Pr(c_{i,t} = 0) = \lambda$.[21]

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

$$\min_{\{x_{i,t}\}_{t=0}^{\infty}} \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t B_t(i) \left[L(x_{i,t}) + c_{i,t} \cdot \mathbf{1}_{\{x_{i,t} \neq x_{i,t-1} - \mu_t - \varepsilon_{i,t}\}} \right], \tag{30}$$

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

Reduction to the price-gap formulation. We now show how the firm's problem (30) reduces to the formulation in (2). By construction, $x = 0$ maximizes $\tilde{\pi}$, so $\tilde{\pi}'(0) = 0$. A second-order Taylor expansion of the profit loss around the optimal price gap gives:

$$L(x) = \tilde{\pi}(0) - \tilde{\pi}(x) \approx \frac{\epsilon - 1}{2} x^2. \tag{31}$$

Substituting into (30), the firm's problem becomes:

$$\min_{\{x_{i,t}\}_{t=0}^{\infty}} \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t B_t(i) \left[\frac{\epsilon - 1}{2} x_{i,t}^2 + c_{i,t} \cdot \mathbf{1}_{\{x_{i,t} \neq x_{i,t-1} - \mu_t - \varepsilon_{i,t}\}} \right]. \tag{32}$$

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 (2), with $\bar{c} \equiv \hat{c}/(\epsilon - 1)$ denoting the rescaled menu cost.[22]

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

$$V_t(x) = \frac{1}{2} x^2 + \tilde{\beta}_t \, \mathbf{E}_t \left[(1 - \lambda) \min\Big(V_{t+1}(x - \mu_{t+1} - \varepsilon), \; \bar{c} + \min_{x'} V_{t+1}(x') \Big) + \lambda \min_{x'} V_{t+1}(x') \right], \tag{33}$$

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 $\mu_t = \mu$ and slowly varying $B_t(i)$) we have $\tilde{\beta}_t \approx \beta$ and the value function becomes time-invariant.[23]

As discussed in Appendix A.2, the solution to (33) takes the form of an *Ss* rule (3): for each

[20] The derivation of (28) proceeds by writing $P_t(i)/P_t = (P_t^*(i)/P_t)\, e^{x_{i,t}}$, substituting into the profit expression, and using $MC_t(i) = P_t^*(i)/\mathcal{M}$ to simplify.

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

[22] 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. See footnote 3 for further discussion.

[23] Under the unexpected permanent shock assumption in Section 2.1, each realization of μ_t defines a new stationary environment, so the time-invariant Bellman equation applies conditional on the current drift.

drift μ_t, there exist thresholds $\underline{x}_t < x_t^* < \bar{x}_t$ such that the firm resets its price gap to x_t^* whenever it falls outside $[\underline{x}_t, \bar{x}_t]$ or when adjustment is free.

Aggregation. Let $g_t(x)$ denote the cross-sectional distribution of price gaps $x_{i,t} \equiv p_{i,t} - p_{i,t}^*$ across firms. Given the *Ss* pricing rule in (3), this distribution evolves according to:

$$g_t = \mathcal{T}(g_{t-1};\, \mu_t, \sigma_\varepsilon, \underline{x}_t, \bar{x}_t, x_t^*),$$

where $\mathcal{T}$ maps the distribution of price gaps 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:

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

where $B_t(i) = (P_t^*(i)/P_t)^{1-\epsilon}$ is the firm-specific weight from (28). Integrating the inflation identity (6) across firms yields:

$$\pi_t = \pi_t^* + \int (x_{i,t} - x_{i,t-1})\, di,$$

where aggregate frictionless inflation is $\pi_t^* \equiv \int \pi_{i,t}^*\, di = \mu_t$. Finally, the aggregate resource constraint is:

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

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.

A.2 The *Ss* pricing rule

This appendix establishes conditions under which the optimal pricing policy for (33) takes the *Ss* form (3). The argument follows the approach in Auclert *et al.* (2024) (Online Appendix B.2), who prove the result analytically for $\mu = 0$, and extends it to $\mu > 0$ via a quasiconvexity condition.

A.2.1 Stationary case

Consider the stationary version of (33) with constant drift μ and discount factor β:

$$V(x) = \frac{1}{2}x^2 + \beta\Big\{(1-\lambda)\,\mathbf{E}\Big[\min\big(V(x-\mu-\varepsilon),\ \bar{c} + V(x^*)\big)\Big] + \lambda\, V(x^*)\Big\}, \tag{34}$$

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

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

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

(iii) Since the expectation in (34) is a convolution with a Gaussian kernel, V is C^∞.

(iv) Differentiating under the integral gives

$$V'(x) = x + \beta(1-\lambda) \int_{\underline{x}}^{\bar{x}} V'(y)\, f(x - \mu - y)\, dy, \tag{35}$$

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

Quasiconvexity implies *Ss*. Say V is *strictly quasiconvex* if it has a unique minimizer x^* and is strictly decreasing on $(-\infty, x^*]$, strictly increasing on $[x^*, \infty)$. If V is strictly quasiconvex, then every sublevel set $\{x : V(x) \leq c\}$ is a closed interval. In particular, the inaction set $\mathcal{I} = \{x : V(x) \leq \bar{c} + V(x^*)\}$ is a closed interval $[\underline{x}, \bar{x}]$ with $\underline{x} < x^* < \bar{x}$ (since $\bar{c} > 0$). Outside this interval the firm strictly prefers to pay $\bar{c}$ and reset to the unique minimizer x^*; inside, it weakly prefers inaction. This is exactly the *Ss* rule (3).

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 x^*.

Verification. When $\mu = 0$, Auclert *et al.* (2024) show that V is symmetric ($V(x) = V(-x)$) and $V'(x) > 0$ for $x > 0$, which gives strict quasiconvexity with $x^* = 0$. When $\mu > 0$, 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 μ_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 μ with μ_{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 μ_t, and verification reduces to the stationary case. If μ_t reverts along a deterministic path $\mu_t \to \mu_\infty$, 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 in our framework. 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:

$$C_t = M_t / P_t \qquad \text{(cash-in-advance)}, \tag{36}$$

$$1 = \int e^{(1-\epsilon)x} g_t(x)\, dx \qquad \text{(price index)}, \tag{37}$$

$$N_t = C_t \int e^{-\epsilon x} g_t(x)\, dx + \bar{c}(1-\lambda)\omega_t \qquad \text{(labor market clearing)}, \tag{38}$$

$$g_t = \mathcal{T}(g_{t-1};\, \mu_t,\, \sigma_\varepsilon,\, \underline{x}_t,\, \bar{x}_t,\, x_t^*) \qquad \text{(distribution evolution)}, \tag{39}$$

$$\mu_t = \Delta \log(1-\phi_t) + \Delta \log M_t \qquad \text{(drift)}, \tag{40}$$

$$(\underline{x}_t, \bar{x}_t, x_t^*) \text{ from firm value function } V_t(x) \qquad \text{(optimal pricing)}, \tag{41}$$

where ω_t denotes the fraction of firms that pay the menu cost $\bar{c}$ 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 μ_t via (40), which changes the evolution of g_t via (39), which in turn changes the firm's value function and thus the Ss bands $(\underline{x}_t, \bar{x}_t, x_t^*)$ via (41). 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. We do not solve the Ramsey problem in this paper; instead, as discussed in Section 2.3.1, we evaluate simple parametric rules that condition on estimated moments of G_t.

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:

$$u_t = \log C_t - N_t = \log \frac{M_t}{P_t} - \frac{M_t}{P_t} \int e^{-\epsilon x} g_t(x)\, dx - \bar{c}(1-\lambda)\omega_t.$$

In the flexible-price economy ($x_{i,t} = 0$ for all i), the price index constraint becomes $P_t = P_t^*(i)$ for all i, the integral $\int e^{-\epsilon x} g_t(x)\, dx = 1$, and the adjustment cost vanishes. Denoting the

flexible-price utility by u_t^e and letting $\hat{c}_t \equiv \log(C_t/C_t^e)$ denote the output gap, a second-order expansion of the utility loss yields:[24]

$$\mathcal{L}_t \equiv u_t^e - u_t \approx \frac{1}{2}\hat{c}_t^2 + \frac{\epsilon^2}{2}\int x^2\, g_t(x)\, dx + \bar{c}(1-\lambda)\omega_t. \tag{42}$$

The second term is, up to a constant, the welfare loss from price misalignment: $\frac{\epsilon^2}{2}\int x^2\, g_t(x)\, dx = \frac{\epsilon^2}{2}[\mathrm{Var}(x_t) + \bar{x}_t^2]$. This is the expression summarized in equation (4) of the main text (where the factor ϵ^2 is absorbed into the normalization). The decomposition has a clear economic interpretation:

- The *output gap* term $\frac{1}{2}\hat{c}_t^2$ captures the welfare cost of aggregate demand deviating from its efficient level.
- The *gap dispersion* term $\frac{\epsilon^2}{2}\int x^2\, g_t(x)\, dx$ captures the misallocation of demand across varieties. Firms with $x_{i,t} \neq 0$ 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 $\bar{c}(1-\lambda)\omega_t$ 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 and in the applications of Section 5.

A.4 Calibration for simulated illustrations

Table A.1 reports the parameter values used in the simulated illustrations of Section 2.3.1. The model is calibrated at a monthly frequency to match two moments from US CPI micro data reported by Nakamura and Steinsson (2010): a monthly frequency of price change of approximately 8.7% and a mean absolute non-zero price change of approximately 8.5%. We set the discount factor and steady-state drift to standard values and solve for the remaining parameters (σ_ε, λ, and the menu cost ξ) to match the target moments. The *Ss* rule parameters ($\underline{x}$, $\bar{x}$, x^*) are outputs of the optimal policy solution.

A.5 Amplification under volatility shocks

Section 2.3.1 shows that when initial distributions are generated by different histories of drift shocks, the mean price gap is the relevant summary statistic for predicting the amplification ratio. Here we consider an alternative source of variation in the gap distribution: changes in idiosyncratic volatility σ_ε.

[24] The derivation uses the fact that $e^{(1-\epsilon)x} \approx 1 + (1-\epsilon)x + \frac{(1-\epsilon)^2}{2}x^2$ and $e^{-\epsilon x} \approx 1 - \epsilon x + \frac{\epsilon^2}{2}x^2$ around $x = 0$, combined with the price index constraint (37) to eliminate the mean gap term.

We subject the same calibrated economy to different histories of volatility shocks—AR(1) perturbations to σ_ε that raise or lower idiosyncratic uncertainty relative to the steady state—and sample the resulting distributions at various points during the transition. These distributions differ from the ergodic primarily in their variance and kurtosis, while their mean gap remains close to zero. We then apply the same 4 pp level shock used in Figure 1 and compute the amplification ratio.

Figure A.1 reports the results. The amplification ratio varies from approximately 0.5 to nearly 5, and the variance ratio $\text{Var}(x_t)/\text{Var}_{ss}(x)$ explains the vast majority of this variation. Distributions with higher variance have more mass near the adjustment thresholds, making the economy more sensitive to the test shock. The "no shock" counterfactual is close to zero for all distributions, confirming that the amplification arises from the interaction between the shock and the distribution shape rather than from mechanical reversion of pent-up pressure.

A.6 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 x^*, contributing a price increase of $(x^*-\underline{x})$. Non-adjusting firms hold prices fixed. Aggregating yields (8).

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 frictionless 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.[25] Within

[25]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 $\Delta \widehat{Z}^{\star}_{i,j,t}$.[26]

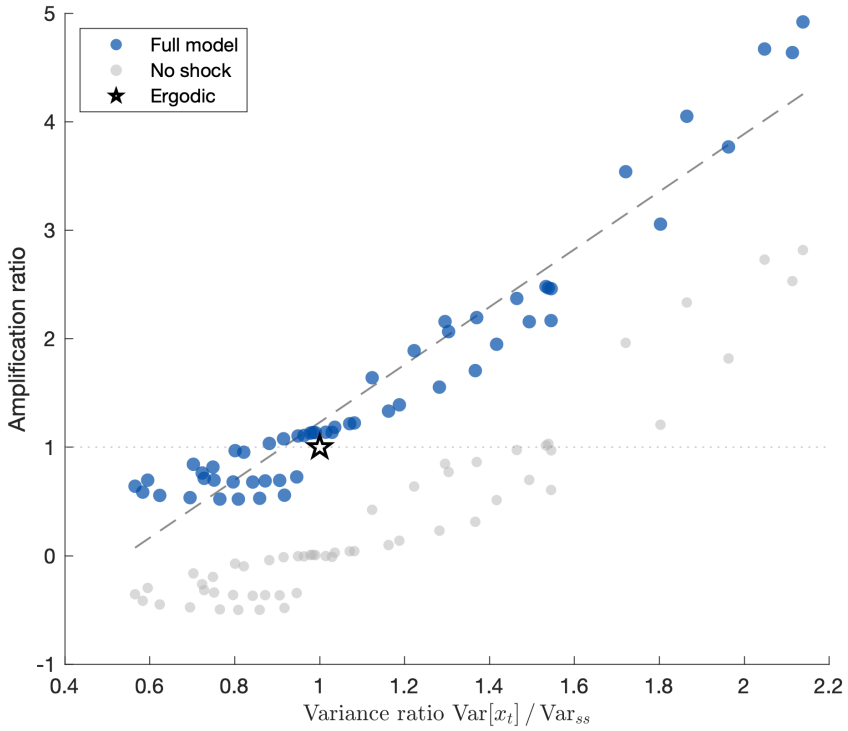
From elementary aggregates to higher level indices Indices for higher levels are weighted averages of the elementary aggregate indices. First, we combine elementary aggregates using stratum weights to construct item-level indices. Second, we aggregate item-level indices using item weights to produce 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).

[26] 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.

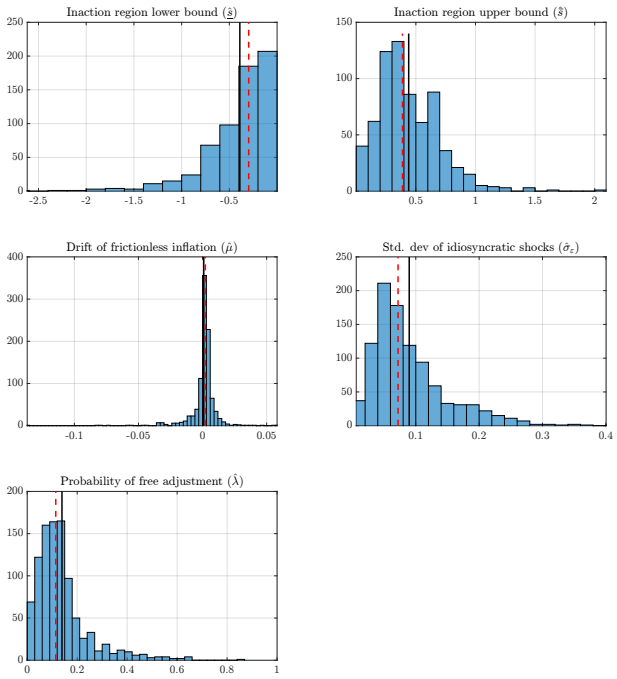
D Additional Figures

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



Notes: Each dot represents a distribution obtained by subjecting the model to a different history of idiosyncratic volatility shocks and sampling the gap distribution at a point during the transition. All distributions share the same structural parameters and are hit with the same 4 pp one-off level shock to drift. The y-axis shows the amplification ratio (impact excess inflation normalized by the ergodic response). Gray dots show the counterfactual response without the test shock. The star marks the ergodic distribution. Same calibration as Figure 1; see Table A.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.

E Additional Tables

Table A.1: Calibration for simulated illustrations

Parameter	Description	Value	Source
Set externally			
β	Discount factor (monthly)	0.997	Standard
μ	Steady-state drift (monthly)	0.00165	$\approx$ 2% annual inflation
Calibrated to match target moments			
σ_ε	Idiosyncratic shock s.d.	0.071	
λ	Free adjustment probability	0.120	
ξ	Menu cost	1.094	
Implied Ss rule			
$\underline{x}$	Lower threshold	-0.477	
$\bar{x}$	Upper threshold	0.485	
x^*	Reset point	0.008	
Target moments			
	Freq. of price change (monthly)	12.6%	
	Mean $\lvert\Delta p\rvert$ (non-zero)	15.3%	
	Kurtosis of Δp	3.8	

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.