CIMEO Working Papers

Centre for Investigation and Modelling
of Experimental Observations
Sapienza Universita di Roma

Working Paper No. 208 - 2026

When Inflation Takes Hold: Labor-Market Pressure and Price Transmission

Elton Beqiraj | Mirabela Bika | Giovanni Di Bartolomeo

2026

Working Papers Series 2026

Centre for Investigation and Modelling of Experimental Observations

When Inflation Takes Hold: Labor-Market Pressure and Price Transmission*

Elton Beqiraj[a] Mirabela Bik[a] Giovanni Di Bartolomeo[a,b,c]

This draft: August 6, 2026

Abstract

This paper develops and estimates a state-dependent Phillips curve in which labor-market pressure and price adjustment jointly determine when an inflationary disturbance becomes aggregate inflation. Tight labor markets select a wage branch that raises marginal-cost pressure, while the level and age profile of price adjustment determine how rapidly that pressure reaches prices and how long its effects survive. This interaction generates an *inflation impact–persistence frontier*. The pricing state with the strongest immediate pass-through need not be the state with the longest-lived response. Estimated on U.S. quarterly data, the model locates postwar inflation surges in a high-pass-through pricing environment and finds inflation most vulnerable when that environment coincides with strong labor-market pressure. Bayesian evidence favors the joint state structure over specifications that suppress either its labor-pressure margin or its pricing-state margin.

JEL codes: E31, E32, E52, J64.
Keywords: Inflation dynamics, nonlinear Phillips curve, labor-market tightness, pass-through, inflation surges, state-dependent Phillips curves.

*Earlier versions of this paper were presented at the DG ECFIN Internal Seminar at the European Commission, the DG Economics Seminar at the European Central Bank, and the International Conference on Economic Modeling and Data Science—ECOMOD 2026 in Luxembourg. We thank all participants for their helpful comments and suggestions. We are especially grateful to Pierpaolo Benigno, Philipp Pfeiffer, Béla Szörfi, and Claudia Foroni.

(*a*) Sapienza University of Rome; (b) Corvinus University, Budapest; (c) University of Antwerp.

1 Introduction

Why do some inflationary disturbances fade while others become visible aggregate inflation? The post-pandemic surge returned this question to the center of macroeconomics. Supply disruptions, demand support, monetary accommodation, and labor shortages all supplied plausible inflationary impulses. Yet the same forces need not produce the same inflation response in different macroeconomic environments. Explaining a surge therefore requires more than identifying the shock. It requires explaining the state through which the shock is transmitted.

Recent work on nonlinear Phillips curves emphasizes one part of that state. When vacancies are abundant relative to job seekers, new-hire wages and hiring costs become especially responsive to labor-market conditions (Benigno and Eggertsson, 2024, 2025). This mechanism explains why tightness can generate strong marginal-cost pressure after a long period in which the Phillips curve appeared flat. It does not by itself determine how much of that pressure reaches current prices or how the inflation response evolves once the initial impulse has passed.

This paper studies the joint determination of pressure and price transmission. Labor-market conditions select the wage entering marginal cost. The price-setting environment then maps marginal cost into current and future inflation. When reset rates are similar across price vintages, cost pressure passes into current inflation rapidly. When adjustment becomes increasingly sensitive to price age, the selection of firms that reprice changes the dynamic response. The resulting Phillips curve has two state dimensions. Labor-market pressure changes the force arriving at the pricing block, while the age profile of adjustment changes both impact pass-through and intrinsic persistence.

The interaction produces an impact–persistence frontier. A high initial reset rate combined with a comparatively flat age profile transmits marginal cost strongly on impact. A lower initial reset rate combined with a steeper age profile can instead generate greater intrinsic continuation. High inflation and high intrinsic persistence are therefore distinct estimated outcomes. This distinction is important for historical episodes. A surge can begin in a high-pass-through state even when the state with the strongest intrinsic persistence is elsewhere in the sample.

Figure 1 provides some motivating evidence. Panel (a) documents the familiar steepening of the inflation–tightness relation when vacancies become large relative to unemployment. Panel (b) shows that predetermined inflation persistence contains substantial information about the variation left unexplained by tightness. The figure does not identify the latent states used later in the paper. However, it shows that a one-dimensional measure of slack cannot jointly account for the initial response of inflation and its subsequent persistence.

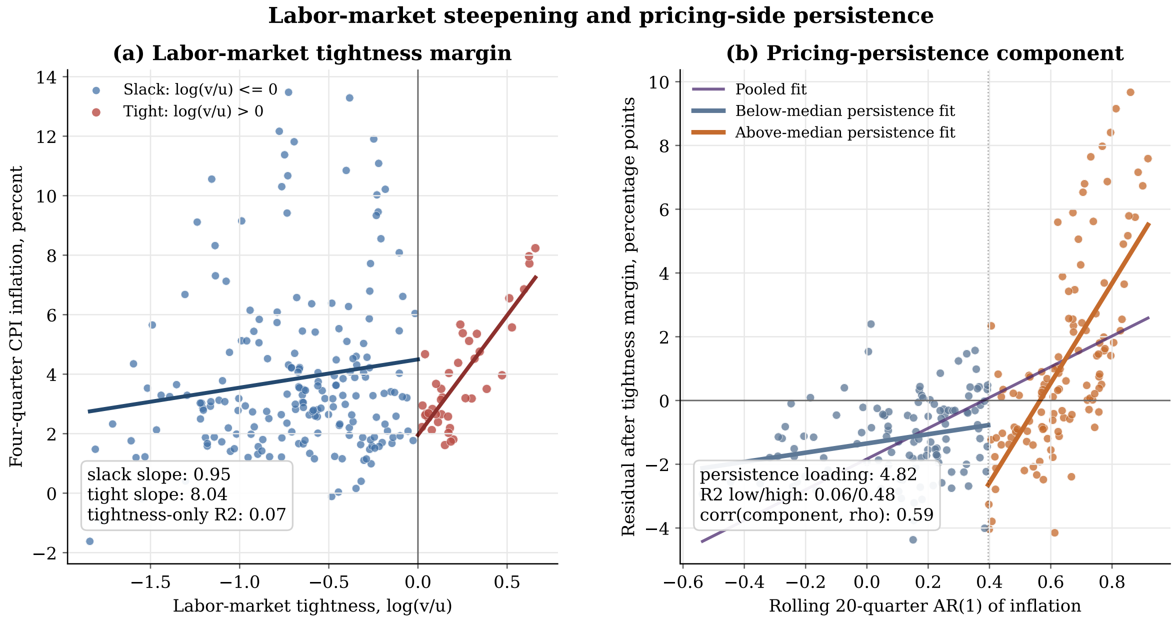


Figure 1: Labor-market pressure and pricing-side persistence.

Notes: The figure uses U.S. quarterly data. Inflation is four-quarter CPI inflation and labor-market tightness is the vacancy–unemployment ratio. Panel (a) reports the inflation–tightness relation on the two sides of the unitary tightness threshold. Panel (b) relates the residual left by that margin to a rolling measure of inflation persistence. The figure is descriptive and precedes the joint state-space estimation.

We explore and develop this argument in a New Keynesian model with search and matching, asymmetric wage adjustment, and duration-dependent pricing. In tight labor markets, the wage of new hires follows the flexible reservation wage and current tightness enters marginal cost directly. In slack labor markets, the incumbent wage anchors the wage of new hires and attenuates contemporaneous pressure. Price adjustment depends on the age of a price spell. This generates a Phillips curve in which the labor state changes marginal-cost forcing and the pricing state changes both the coefficient on that forcing and the intrinsic propagation of inflation.

The analytical results characterize three margins. First, wage selection creates a tight–slack difference in the local Phillips-curve slope. Second, the level and age profile of price adjustment jointly determine impact pass-through and intrinsic persistence, creating the impact–persistence frontier. Third, the two margins interact in the inflation-vulnerability threshold. Strong labor-market pressure lowers the impulse needed to reach high inflation, but its effect is largest in the pricing environment that transmits marginal cost most readily. Neither condition is sufficient by itself. A permissive pricing environment has no pressure to transmit in the absence of a disturbance, and tightness need not become a surge when pass-through is weak.

We estimate the model on U.S. quarterly data using a Bayesian regime-switching system. Vacancies and unemployment discipline the labor-pressure state. Inflation, activity, wages, unemployment, and the policy rate jointly reveal the pricing environment and the model innovations. The resulting state space separates the formation of cost pressure from its transmission into prices. The model comparison follows the same economic ordering, moving from a constant-state benchmark to a model with labor-state variation and then to the joint labor-pricing specification.

Our empirical analysis yields a coherent account of inflation vulnerability. Postwar surges occur predominantly in the flat, high-pass-through pricing state, while high-inflation quarters are especially concentrated when this environment coincides with strong labor-market pressure. Generalized threshold experiments confirm that this joint state requires the smallest disturbance to reach a common inflation boundary. The labor-market state operates primarily upstream, through wage selection and marginal-cost formation, whereas the pricing state determines how strongly that pressure is reflected in aggregate inflation. Historical shock accounting further shows that the vulnerable environment is reached through different disturbances across episodes. Demand pressure plays the leading role in the estimated decompositions of the late-1960s and Great Inflation episodes, whereas policy and markup innovations are more prominent after 2020.

The paper develops an integrated account of inflation vulnerability in which labor-market pressure and price transmission operate as complementary mechanisms rather than competing explanations. It shows analytically that a rotation in the price-adjustment profile generates an impact–persistence trade-off in which the pricing state producing the strongest immediate response need not be the one producing the most persistent response. The joint labor–pricing state space is then estimated and used to identify the conditions under which inflationary disturbances develop into aggregate inflation surges.

The analysis builds on the literature on nonlinear Phillips curves and labor-market tightness (Phillips, 1958; Benigno and Ricci, 2011; Benigno and Eggertsson, 2024, 2025; Hazell et al., 2022; Cerrato and Gitti, 2023; Gitti, 2023). This literature establishes that the marginal-cost pressure generated by a given disturbance can vary sharply with labor-market conditions.[1] Recent evidence further shows that monetary policy affects both labor-demand and labor-supply margins: contractionary shocks reduce hiring and vacancies, while increasing job-seeking among the non-employed and reducing quits to nonemployment (Graves et al., 2026). We complement this literature by studying how the price-adjustment environment translates the resulting marginal-cost pressure into inflation and governs its subsequent propagation.

The paper also contributes to the literature on intrinsic inflation persistence and the pricing mechanisms through which past price setting affects current inflation (Fuhrer and Moore, 1995; Galí and Gertler, 1999; Sheedy, 2010; Di Bartolomeo and Di Pietro, 2017; Dixon and Le Bihan, 2012). Micro price evidence documents heterogeneous and weakly synchronized adjustment, persistent reference prices, and regular-price hazards that are often close to flat once temporary sales are removed (Bils and Klenow, 2004; Nakamura and Steinsson, 2008; Klenow and Kryvtsov, 2008; Eichenbaum et al., 2011; Klenow and Malin, 2010). Motivated by this evidence, we characterize nominal rigidity through a state-conditioned aggregate adjustment profile rather than through a single invariant hazard or average price duration. Duration dependence provides a forward-looking source of continuation through the selection of price vintages. Allowing the adjustment profile to vary across states bridges intrinsic persistence and regime-dependent transmission and thereby separates impact pass-through from persistence within a Phillips curve whose marginal-cost forcing is itself state dependent.

Finally, the paper relates to work using regime-switching models to study changes in

[1] This mechanism is consistent with search-and-matching models, in which labor-market tightness plays a central role in the amplification of aggregate disturbances (Shimer, 2005; Hall, 2005), and with staggered-wage models linking wage determination to aggregate dynamics (Gertler and Trigari, 2009; Christiano et al., 2016). Evidence on newly formed matches further indicates that wage cyclicality depends on worker and match composition (Haefke et al., 2013; Gertler et al., 2020). These findings motivate the distinction, embedded in our wage-selection mechanism, between the wage relevant for marginal hiring decisions and the unadjusted average wage observed among new hires.

macroeconomic transmission (Bianchi, 2013; Baele et al., 2015; Beqiraj et al., 2026). Our framework distinguishes the state in which labor-market pressure is formed from the state governing its transmission into prices. Estimating these dimensions jointly allows us to assess their separate and combined roles in the onset of inflation surges.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 derives the state-dependent Phillips curve and its analytical properties. Section 4 develops the inflation-vulnerability frontier and the motivating matched contrasts. Section 5 describes the data and Bayesian estimation. Section 6 studies the estimated state space, posterior vulnerability, historical shock alignment, and state-specific transmission. Section 7 concludes.

2 The theoretical framework

The model is organized around a single state-dependent Phillips curve. It contains two discrete regime dimensions. The labor-market regime governs the wage mapping through which vacancy–unemployment conditions enter marginal cost, while the pricing regime determines how a marginal-cost impulse is transmitted and propagated through the duration structure of price adjustment. Alongside these regimes, the incumbent wage inherited from earlier contracts is a predetermined endogenous variable. The inherited wage is retained because it affects the slack-regime Phillips curve. We denote the labor-market regime by s_t^w and the pricing regime by s_t^p. The lagged incumbent wage, $\hat{w}_{t-1}^{ex}$, carries contractual history into the current inflation equation but is not an additional regime. Together, these elements deliver two Phillips curves. The first applies to tight labor markets, where marginal cost depends directly on the flexible reservation wage. The second applies to slack labor markets, where it depends on the inherited incumbent wage. The sequence is fixed throughout the paper. Labor-market conditions govern the marginal wage response, that wage enters marginal cost, price adjustment maps marginal cost into inflation, and inherited wages shape the slack-regime marginal-cost state.

2.1 Households

A representative household chooses consumption C_t, one-period nominal bonds B_t, and labor-force participation $F_t \in [0,1]$, where F_t denotes the fraction of household members who either work or actively search for a job. Preferences are of the Greenwood–Hercowitz–Huffman type (Greenwood et al., 1988).

$$\max_{\{C_t,B_t,F_t\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \, \xi_t \frac{X_t^{1-\sigma}-1}{1-\sigma}, \qquad X_t \equiv C_t - \chi_t \frac{F_t^{1+\omega}}{1+\omega} + \Psi_t, \tag{1}$$

where $\beta \in (0,1)$ is the discount factor, $\sigma > 0$ is the inverse intertemporal elasticity of substitution, ξ_t is an intertemporal demand shifter, $\chi_t > 0$ shifts the marginal disutility of participation, $\omega > 0$ governs the curvature of participation costs, and Ψ_t is an additive preference shifter. The GHH specification eliminates the wealth effect on labor supply, which is convenient when combined with a search-and-matching block whose only intensive labor margin is participation.

The household's nominal budget constraint is written with separate wage-bill components for incumbent workers and new hires.

$$P_t C_t + B_t + T_t = (1+i_{t-1})\, B_{t-1} + (1-s) W_t^{\text{ex}} F_t + s(1-\gamma_b) f(\theta_t) W_t^{\text{new}} F_t + Z_t^F + Z_t^E. \tag{2}$$

Here P_t is the aggregate price level, T_t is a lump-sum tax, i_{t-1} is the nominal interest rate set in the previous period, W_t^{ex} is the wage paid on inherited employment matches, W_t^{new} is the wage paid on newly formed matches, and Z_t^F, Z_t^E are profits rebated by firms and employment agencies. The parameter $s \in (0,1)$ is the fraction of participants who begin the period as job seekers, $f(\theta_t)$ is the job-finding probability, and $\gamma_b \in [0,1)$ is the proportional fee retained by the employment agency on the wage bill of new hires. The pool of job seekers at the beginning of period t is therefore $U_t^b \equiv s\, F_t$.

Standard manipulations of the first-order conditions yield the Euler equation

$$1 = (1+i_t)\,\mathbb{E}_t\Lambda_{t,t+1}, \qquad \Lambda_{t,t+1} \equiv \beta\left(\frac{X_{t+1}}{X_t}\right)^{-\sigma}\frac{\xi_{t+1}}{\xi_t}\,\Pi_{t+1}^{-1}, \tag{3}$$

with $\Pi_{t+1} \equiv P_{t+1}/P_t$, and the optimal participation condition

$$(1-s)w_t^{\text{ex}} + s(1-\gamma_b)f(\theta_t)w_t^{\text{new}} = \chi_t\, F_t^{\omega}, \tag{4}$$

where $w_t^j \equiv W_t^j/P_t$ for $j \in \{\text{ex}, \text{new}\}$.

Log-linearizing the Euler equation around the deterministic steady state and using a preference-shock representation of the natural rate of interest yields the IS equation that closes the demand side of the model.

$$\hat{Y}_t = \mathbb{E}_t\hat{Y}_{t+1} - \frac{1}{\sigma}\Big(\hat{\imath}_t - \mathbb{E}_t\hat{\pi}_{t+1} - \hat{r}_t^e\Big), \tag{5}$$

where $\hat{r}_t^e \equiv (1-\rho_\xi)\hat{\xi}_t$ is a preference (or natural-rate) shock and hatted variables denote log-deviations from steady state.

2.2 Production and marginal cost

Intermediate-good firms produce a differentiated variety using labor and an imported intermediate input O_t (energy/oil) according to a Cobb–Douglas technology.

$$y_t(i) = A_t\, N_t(i)^{\alpha_N}\, O_t(i)^{1-\alpha_N}, \qquad 0 < \alpha_N < 1, \tag{6}$$

where A_t is aggregate productivity and α_N is the labor share in production. Cost minimization gives the aggregate marginal-cost component

$$\widehat{mc}_t = \alpha_N\, \hat{w}_t^m + (1-\alpha_N)\,\hat{o}_t - \hat{A}_t, \tag{7}$$

where $\hat{o}_t = \log(P_t^O/P_t) - \log(\bar{o})$ is the log relative price of the imported input and $\hat{w}_t^m$ is the wage relevant for marginal employment expansion. In the baseline implementation we set $\hat{w}_t^m = \hat{w}_t(s_t^w)$, where the labor block below selects either the flexible new-hire wage or the inherited incumbent wage. This assumption is exact when employment adjustment occurs at the hiring margin and newly hired labor is the marginal input. More generally, the coefficient on the selected wage should be read as the reduced-form loading of the marginal employment cost on the wage selected by the labor-market regime. The quantitative specification calibrates α_N as the labor share. This parameter is not used to generate price-setting real rigidity.

The pass-through from aggregate marginal cost to the optimal reset price is governed by a separate real-rigidity parameter. To make this distinction explicit, suppose that a resetting firm's marginal cost contains a firm-specific component generated by relative output demand,

$$\widehat{mc}_{t+\ell|t}(i) = \widehat{mc}_{t+\ell} + \chi_p \, \widehat{y}_{t+\ell|t}(i), \qquad \chi_p \geq 0, \tag{8}$$

with demand $\widehat{y}_{t+\ell|t}(i) = -\epsilon \, (p_t^* - p_{t+\ell})$. Linearizing the reset-price condition then scales the response of the reset price to aggregate marginal cost by

$$\Theta_p \equiv \frac{1}{1+\epsilon\chi_p}, \qquad 0 < \Theta_p \leq 1. \tag{9}$$

Thus α_N measures the labor share in marginal cost, while Θ_p measures firm-level strategic complementarity in price setting. If no firm-specific real rigidity is imposed, $\chi_p = 0$ and $\Theta_p = 1$. In the empirical implementation Θ_p is fixed as a pass-through calibration entering k_p. It is conceptually distinct from the calibrated labor share α_N.

Micro evidence motivates separating the overall frequency of price changes from the age profile of adjustment. Price changes are frequent and heterogeneous, while reference prices can remain inertial and unconditional age hazards are often nearly flat (Bils and Klenow, 2004; Nakamura and Steinsson, 2008; Klenow and Kryvtsov, 2008; Eichenbaum et al., 2011; Klenow and Malin, 2010). In the fixed-pricing-environment benchmark, the probability of price adjustment depends on the duration since the last reset (Sheedy, 2010; Di Bartolomeo and Di Pietro, 2017). Let $\ell \geq 1$ denote the age of a price. A firm with a price of age ℓ reoptimizes with probability

$$\alpha_{p,1} = \alpha_p, \qquad \alpha_{p,\ell} = \min\left\{\alpha_p + \frac{\varphi_p}{1-\alpha_{p,\ell-1}}, 1\right\}, \quad \ell > 1. \tag{10}$$

The parameter $\alpha_p \in (0,1)$ governs the initial reset probability, while $\varphi_p \geq 0$ governs duration dependence. A positive φ_p implies a positively-sloped hazard — older prices are more likely to be reset than recently posted ones — which generates a positive selection effect and intrinsic inflation persistence. When the recursion reaches one, the hazard has a terminal spike and the associated price spell ends with probability one.

Define survival probabilities

$$\varsigma_{p,0} = 1, \qquad \varsigma_{p,\ell} = \prod_{h=1}^{\ell} \Big(1 - \alpha_{p,h}\Big). \tag{11}$$

The benchmark coefficient restrictions below require an admissible hazard sequence. Let $\tilde{\alpha}_{p,\ell}$ denote the same recursion before applying the cap. The formulas are exact when either the cap never binds and $\sum_{\ell\geq 0} \varsigma_{p,\ell} < \infty$, or when the uncapped recursion reaches the terminal value exactly, $\tilde{\alpha}_{p,L} = 1$, for some finite L. Under these conditions,

$$(\alpha_p + \varphi_p) \sum_{\ell \geq 0} \varsigma_{p,\ell} = 1, \tag{12}$$

with the finite-support sum ending at $L-1$ when a terminal spike occurs, and the stationary distribution of price vintages is

$$\theta_{p,\ell} = (\alpha_p + \varphi_p)\varsigma_{p,\ell}. \tag{13}$$

If instead the cap binds strictly, $\tilde{\alpha}_{p,L} > 1$, the stationary age distribution is

$$\theta_{p,\ell} = \frac{\varsigma_{p,\ell}}{\sum_{j\geq 0} \varsigma_{p,j}}, \tag{14}$$

and the compact mapping in (18) is no longer used as a benchmark restriction without recomputing the finite-support price-index and reset-price system. In the empirical model, (12) is treated only as an admissibility condition for the reported hazard-motivated coefficient mapping.

Within the same fixed-hazard benchmark, a firm resetting at time t chooses p_t^* to maximize

$$\mathbb{E}_t \sum_{\ell=0}^{\infty} \Lambda_{t,t+\ell}\, \varsigma_{p,\ell} \left[\left(p_t^* - MC_{t+\ell}\right) y_{t+\ell|t} \right]$$

subject to the demand schedule $y_{t+\ell|t} = (p_t^*/P_{t+\ell})^{-\epsilon}\, Y_{t+\ell}$, where $\epsilon > 1$ is the elasticity of substitution across varieties. A first-order approximation of the optimal reset-price condition implies that the log relative reset price $x_t \equiv \log(p_t^*/P_t)$ satisfies

$$x_t = \Theta_p \left[1 - \beta(1-\alpha_p) + \beta^2\varphi_p\right] \sum_{\ell=0}^{\infty} \beta^\ell\, \varsigma_{p,\ell}\, \mathbb{E}_t \left[\widehat{mc}_{t+\ell} + p_{t+\ell} - p_t\right]. \tag{15}$$

where $p_t \equiv \log P_t$. The second term is the cumulative inflation between the reset date and the period in which the price may still be in effect. The aggregate price index, together with the stationary vintage distribution, implies the static price-index identity

$$(\alpha_p + \varphi_p)x_t = (1 - \alpha_p - \varphi_p)\pi_t - \varphi_p\pi_{t-1}. \tag{16}$$

Combining the fixed-hazard reset-price condition (15) with the price-index identity (16) and eliminating x_t motivates the reduced-form inflation equation

$$\pi_t = \psi_p\, \pi_{t-1} + \beta\Big[1 + (1-\beta)\psi_p\Big]\, \mathbb{E}_t\pi_{t+1} - \beta^2\psi_p\, \mathbb{E}_t\pi_{t+2} + k_p \left(\widehat{mc}_t + \hat{\nu}_t\right), \tag{17}$$

where $\hat{\nu}_t$ collects markup and other cost-push disturbances and

$$\psi_p = \frac{\varphi_p}{(1-\alpha_p) - \varphi_p[1-\beta(1-\alpha_p)]}, \qquad k_p = \frac{(\alpha_p+\varphi_p)\Big[1-\beta(1-\alpha_p)+\beta^2\varphi_p\Big]}{(1-\alpha_p) - \varphi_p[1-\beta(1-\alpha_p)]} \cdot \Theta_p. \tag{18}$$

The coefficient ψ_p captures the intrinsic backward-looking component generated by duration dependence, while k_p scales the sensitivity of inflation to marginal cost. The propagation coefficient is implied by the price-vintage structure, while the impact coefficient also contains the separate reset-price pass-through calibration Θ_p. Lagged inflation is therefore not an independently appended indexation term. It is generated by the duration-dependent adjustment hazard. The price-index identity (16) keeps track of which price vintages survive, and the hazard determines which vintages are selected into adjustment. In that benchmark, the hazard pair, α_p and φ_p, determines the cross-sectional distribution of prices and the propagation coefficient ψ_p. Together with the separate pass-through calibration Θ_p, it determines the impact coefficient k_p. Persistence is not a free lag coefficient. It is tied to the same selection mechanism that changes how marginal-cost pressure reaches inflation on impact.

In the estimated model, the pricing state indexes the coefficient map implied by the fixed-hazard benchmark. The micro-founded interpretation of this map is local to a pricing environment and its stationary vintage distribution. A fully nonlinear vintage economy with changing hazards would additionally carry the inherited distribution of price ages as an endogenous state. The empirical specification instead uses a tractable

Markov system for pass-through and intrinsic-continuation coefficients, with cross-state restrictions motivated by duration-dependent pricing.[2]

Endogenizing the transition between pricing environments is not required for the present analysis, which focuses on how alternative adjustment profiles shape inflation transmission and propagation. A natural direction for future research would be to develop a theory of the mechanisms that induce firms to change their adjustment behavior and thereby make intrinsic persistence endogenous. Possible candidates include changes in price dispersion, trend inflation, uncertainty and information-acquisition incentives, or sufficiently large shocks that induce firms to move between predominantly time-dependent and state-dependent pricing.

2.3 Labor market and asymmetric wage adjustment

New hires are generated by a Cobb–Douglas matching function, as in the quantitative search-and-matching tradition (Shimer, 2005; Christiano et al., 2016),

$$M_t = m\,(U_t^b)^\eta\,V_t^{1-\eta}, \qquad m > 0, \quad \eta \in (0,1), \tag{19}$$

where M_t is the number of new matches, $U_t^b = sF_t$ is the pool of job seekers, V_t is the number of vacancies, m is matching efficiency, and η is the matching elasticity. Labor-market tightness is $\theta_t \equiv V_t/U_t^b$, and the job-finding probability is $f(\theta_t) = M_t/U_t^b = m\,\theta_t^{1-\eta}$. Employment evolves according to

$$N_t = (1-s)\,F_t + M_t = F_t\Big[1 - s + s\,f(\theta_t)\Big], \tag{20}$$

so that the unemployment rate is

$$u_t \equiv \frac{F_t - N_t}{F_t} = s\,\Big[1 - f(\theta_t)\Big]. \tag{21}$$

A first-order approximation around steady state gives

$$\hat{u}_t^c = -\varphi_u^c\,(1-\eta)\,\hat{\theta}_t,\ \ \varphi_u^c \equiv \frac{\bar{s}-\bar{u}}{\bar{s}} > 0,\ \ \hat{\theta}_t = -\nu_u^c\,\hat{u}_t^c,\ \ \nu_u^c \equiv \frac{\bar{s}}{(1-\eta)(\bar{s}-\bar{u})}. \tag{22}$$

Relation (22) provides the linear bridge between labor-market tightness and the unemployment scale used in the quantitative implementation. Expressing unemployment instead as a conventional log deviation only rescales the coefficient by the steady-state factor $\bar{u}/\bar{s}$. The tightness mapping and all state comparisons are unchanged once the unit of $\hat{u}_t$ is kept fixed. To keep notation light, the unemployment-based Phillips curves below write $\hat{u}_t$ and ν_u for this code-consistent scale and its inverse coefficient.

The labor-pressure state is governed by the observable vacancy–unemployment signal. Let $q_t^w \equiv \log(V_t/U_t^b)$ and denote the high- and low-pressure states by T and S. The quantitative model uses the conditional exit and entry hazards

$$\begin{aligned} p_{TS,t} &\equiv \Pr(s_t^w = S \mid s_{t-1}^w = T, q_t^w) = p_{\min}^\ell + \frac{(1-p_{\min}^\ell)p_{\text{scale}}^\ell}{1+\exp\{\kappa_B(q_t^w - \tau_B)\}}, \\ p_{ST,t} &\equiv \Pr(s_t^w = T \mid s_{t-1}^w = S, q_t^w) = p_{\min}^\ell + \frac{(1-p_{\min}^\ell)p_{\text{scale}}^\ell}{1+\exp\{-\kappa_B(q_t^w - \tau_B)\}}. \end{aligned} \tag{23}$$

[2] Appendix M distinguishes the literal hazard interpretation from this estimated coefficient representation.

High values of q^w_t make exits from T less likely and entries into T more likely, while the preceding state preserves the duration of labor-pressure spells. Let $\omega_{t|t} \equiv \Pr(s^w_t = T \mid \mathcal{Y}_t)$ denote the filtered high-pressure probability. Before the current observation updates the state, its prediction is

$$\omega_{t|t-1} = (1 - p_{TS,t})\omega_{t-1|t-1} + p_{ST,t}\Big(1 - \omega_{t-1|t-1}\Big). \tag{24}$$

The measurement likelihood then maps $\omega_{t|t-1}$ into $\omega_{t|t}$. The analytical results below evaluate the Phillips curve conditional on the labor-pressure state selected by this recursion.

Employment agencies post vacancies at a resource cost $\gamma_c > 0$ and charge a fee γ_b on the wage bill of new hires. Real profits are $Z^E_t = U^b_t\Big[\gamma_b w^{\text{new}}_t f(\theta_t) - \gamma_c \theta_t\Big]$, and the optimal vacancy-posting condition delivers the flexible (market-clearing) real wage

$$w^{\text{flex}}_t = \frac{1}{m\,(1-\eta)}\,\frac{\gamma_c}{\gamma_b}\,\theta^\eta_t. \tag{25}$$

At first order, $\hat{w}^{\text{flex}}_t = \eta\,\hat{\theta}_t$. The log-linear participation and employment relations are written in the wage-bill form used in the quantitative model.

$$\begin{aligned}
\omega \hat{F}_t &= \frac{1-\bar{s}}{1-\bar{u}}\,\hat{w}^{\text{ex}}_t + \frac{\bar{s}-\bar{u}}{1-\bar{u}}\,\hat{w}^{\text{new}}_t - \frac{\bar{u}}{1-\bar{u}}\,\hat{u}_t + \hat{\chi}_t,\\
\hat{N}_t &= \hat{F}_t - \frac{\bar{u}}{1-\bar{u}}\,\hat{u}_t,\\
\hat{u}_t &= -\frac{\bar{s}-\bar{u}}{\bar{s}}\,(1-\eta)\,\hat{\theta}_t.
\end{aligned} \tag{26}$$

Here $\bar{s}$ is the steady-state separation rate, $\bar{u}$ is the steady-state unemployment rate, and $\hat{\chi}_t$ collects participation and labor-market disturbances that are suppressed in the compact analytical expressions below. These expressions map labor-market conditions into marginal cost through (7).

The "incumbent" wage of workers already attached to firms evolves sluggishly according to a staggered rule with partial indexation to expected inflation.

$$\hat{w}^{\text{ex}}_t = \lambda\Big(\hat{w}^{\text{ex}}_{t-1} - \pi_t + \delta\,\mathbb{E}_t\pi_{t+1}\Big) + (1-\lambda)\,\hat{w}^{\text{flex}}_t, \tag{27}$$

where $\lambda \in [0,1]$ governs wage rigidity and $\delta \in [0,1]$ governs partial indexation (Gertler and Trigari, 2009; Benigno and Eggertsson, 2025). The estimated Phillips curve does not impose the Benigno–Eggertsson maximum condition as the regime-selection rule. Instead, it uses a vacancy–unemployment-driven wage equation. In high-pressure quarters the marginal wage is the flexible new-hire wage, while in low-pressure quarters the marginal wage is the inherited-wage expression implied by (27). Thus

$$\hat{w}_t(s^w_t) = \begin{cases} \eta\,\hat{\theta}_t, & s^w_t = 1,\\ \lambda\,\hat{w}^{\text{ex}}_{t-1} + (1-\lambda)\,\eta\,\hat{\theta}_t - \lambda\,(\pi_t - \delta\,\mathbb{E}_t\pi_{t+1}), & s^w_t = 0. \end{cases} \tag{28}$$

The maximum condition $w^{\text{flex}}_t > w^{\text{ex}}_t$ is therefore not a second state definition and does not override the vacancy–unemployment classification inside the likelihood. It is a validation benchmark for the economic interpretation of the high-pressure state. When the reconstructed maximum condition and the vacancy–unemployment state disagree, the estimated model uses (28). The high-pressure state should then be read as a reduced-form tightness-dependent wage branch rather than as the exact Benigno–Eggertsson maximum branch.

Combining (26), the incumbent-wage law, and the aggregate resource condition gives the reduced tightness closure used in the quantitative implementation.

$$\hat{\theta}_t = \frac{\lambda}{z_\theta}\Big(\pi_t - \delta\,\mathbb{E}_t\pi_{t+1}\Big) - \frac{\lambda}{z_\theta}\hat{w}^{\text{ex}}_{t-1} + \frac{\omega}{\alpha_N z_\theta}\Big(\hat{Y}_t - \hat{A}_t\Big) + \frac{1}{z_\theta}\mathcal{Z}^\theta_t, \tag{29}$$

where $\mathcal{Z}^\theta_t$ collects the participation, matching, and labor-market shocks omitted from the compact exposition, and

$$z_\theta \equiv (1-\lambda)\eta + \frac{\bar{s}-\bar{u}}{1-\bar{u}}\,\frac{\bar{u}}{\bar{s}}\,(1+\omega)(1-\eta). \tag{30}$$

The factor $\bar{u}/\bar{s}$ is not an extra free parameter. It comes from expressing unemployment in the same units as the wage-bill participation equation in (26). Equation (29) affects the mapping from output to tightness in the quantitative analysis. The local Phillips-curve formulas below remain conditional on $\hat{\theta}_t$. The inherited wage state enters the slack Phillips curve directly, while in the tight regime it only affects the aggregate tightness-output conversion.

2.4 Monetary policy

The monetary authority follows a smoothed Taylor rule.

$$\hat{\imath}_t = \rho_i\,\hat{\imath}_{t-1} + (1-\rho_i)\Big(\varphi_\pi\,\pi_t + \varphi_y\,\hat{Y}_t\Big) + \varepsilon^i_t. \tag{31}$$

The prior is centered on the standard Taylor (1999) benchmark, but the policy coefficients are estimated and then held fixed across regimes in the quantitative analysis.[3]

Our restriction is an identification discipline. In a regime-switching DSGE model with generalized time-dependent price and wage adjustment, Beqiraj et al. (2026) show that allowing intrinsic inflation persistence to switch materially changes the interpretation of estimated monetary-policy regimes. Variation otherwise attributed to good or bad policy may instead reflect state-dependent shock propagation generated by the price- and wage-setting environment. Their estimates identify policy switches as episodic and concentrated around recessions, while the Great Inflation is explained primarily by changes in price- and wage-setting hazards. The policy regimes also differ mainly in their degree of gradualism rather than in their systematic response to inflation.

Guided by this evidence, we do not introduce a separate monetary-policy regime once duration-dependent nominal rigidity is allowed to vary across pricing states. The Taylor rule therefore provides a common policy-feedback block, estimated over the full sample, while state dependence relevant for the Phillips curve is assigned to labor-market pressure and price adjustment. The monetary-policy innovation is white noise with a state-invariant standard deviation.[4]

[3] This keeps monetary policy from becoming a third switching mechanism. Changes in transmission come from labor-market pressure and pricing pass-through, while the posterior-mode policy coefficients are reported in Table I.2.

[4] Because the sample includes the effective-lower-bound period, the policy innovation should be interpreted as the residual innovation to the common estimated rule rather than narrowly as an unexpected change in the federal funds rate. Medium- and longer-term interest rates remained responsive to monetary-policy news at the lower bound, while forward guidance and large-scale asset purchases constituted distinct dimensions of policy (Swanson and Williams, 2014; Swanson, 2021). Since our specification does not identify these instruments separately, their effects may be reflected in both the estimated policy innovation and the other structural disturbances.

3 The integrated Phillips curve and its analytical properties

This section develops the analytical core of the paper. Section 3.1 states the integrated Phillips curve in its two regime-dependent compact forms. Section 3.2 characterizes the local steepening of the Phillips curve at the regime threshold (Proposition 1), the multiplier role of the pricing hazard (Proposition 2), the rotation needed to separate impact from persistence (Proposition 3), and the joint role of labor-market pressure and pricing pass-through in inflation onset (Proposition 4). Section 3.3 derives the unemployment-based representation of the Phillips curve as a direct empirical counterpart.[5]

3.1 The integrated Phillips curve in two regimes

The wage selected by labor-market conditions enters the duration-dependent price-setting equation through marginal cost. Solving this model mapping for current inflation yields two compact representations. The labor state changes the marginal-cost input, while the pricing block determines impact transmission through k_p and propagation through ψ_p.

Conditional on the high-pressure vacancy–unemployment state, $s_t^w = 1$, the new-hire wage coincides with the flexible reservation wage. Marginal cost therefore depends directly on labor-market tightness, and the Phillips curve is

$$\pi_t = \psi_p \pi_{t-1} + \kappa^{\text{tight}} \hat{\theta}_t + \kappa_\nu^{\text{tight}} (\upsilon_t + \nu_t) + \kappa_\beta^{\text{tight}} \mathbb{E}_t \pi_{t+1} - \beta^2 \psi_p \mathbb{E}_t \pi_{t+2}. \tag{32}$$

The associated coefficients are

$$\kappa_\beta^{\text{tight}} = \beta\Big(1 + (1-\beta)\psi_p\Big), \qquad \kappa^{\text{tight}} = k_p \alpha_N \eta, \qquad \kappa_\nu^{\text{tight}} = k_p. \tag{33}$$

The coefficient κ^{tight} captures the direct pass-through from tightness to the flexible wage, from the wage to marginal cost through the labor share α_N, and from marginal cost to inflation through the pricing slope k_p. The factor α_N reflects the Cobb–Douglas labor share of the production technology in (6), consistent with Benigno and Eggertsson (2025).

Conditional on the low-pressure vacancy–unemployment state, $s_t^w = 0$, the new-hire wage is anchored to the incumbent wage. Since the incumbent-wage law of motion contains current inflation, marginal cost contains a current-period inflation term. Solving the Phillips curve for π_t then gives

$$\pi_t = \widetilde{\psi} \psi_p \pi_{t-1} + \kappa_w \hat{w}_{t-1}^{\text{ex}} + \kappa \hat{\theta}_t + \kappa_\nu (\upsilon_t + \nu_t) + \kappa_\beta \mathbb{E}_t \pi_{t+1} - \beta^2 \psi_p \widetilde{\psi} \mathbb{E}_t \pi_{t+2}. \tag{34}$$

Here $\widetilde{\psi} \equiv (1 + k_p \alpha_N \lambda)^{-1}$, and

$$\kappa_w = \frac{k_p \alpha_N \lambda}{1 + k_p \alpha_N \lambda}, \qquad \kappa = \frac{k_p \alpha_N (1-\lambda)\eta}{1 + k_p \alpha_N \lambda},$$
$$\kappa_\nu = \frac{k_p}{1 + k_p \alpha_N \lambda}, \qquad \kappa_\beta = \frac{\beta\Big(1 + (1-\beta)\psi_p\Big) + k_p \alpha_N \lambda \delta}{1 + k_p \alpha_N \lambda}. \tag{35}$$

Equations (32)–(35) summarize the key asymmetry. In the tight regime, marginal cost depends on the flexible wage $\eta \hat{\theta}_t$, which does not contain current inflation. Hence, no

[5] The local steepening and complementarity results follow from the wage-selection and pricing maps. The substantive cross-state restriction is the impact–persistence rotation. The pricing state with the strongest contemporaneous pass-through need not be the one with the greatest intrinsic continuation.

additional denominator appears in (33). In the slack regime, by contrast, the incumbent-wage law of motion makes marginal cost depend on current inflation. Solving out this feedback introduces the common denominator $1 + k_p\alpha_N\lambda$ in the slack-regime coefficients. At the same time, tightness affects the new-hire wage only through the fraction $1-\lambda$ of wages that reset toward the flexible wage. These two attenuations compress the slack-regime slope relative to the tight-regime slope and are the proximate source of the local steepening characterized below.

3.2 Local steepening at the regime threshold

The slope of contemporaneous inflation with respect to labor-market tightness is given by the coefficient on $\hat{\theta}_t$ in equations (32) and (34). It is κ^{tight} in the tight regime and κ in the slack regime. The ratio of these two coefficients quantifies the local steepening of the Phillips curve when the economy crosses the Beveridge threshold.

The analytical results below focus on the interior case $0 < \lambda < 1$ and $k_p, \alpha_N, \eta > 0$.

Proposition 1 (Local steepening)**.** *Let κ^{tight} and κ denote the contemporaneous slope of inflation with respect to labor-market tightness in the tight and slack regimes, as defined in* (33) *and* (35)*, respectively. Then*

$$\frac{\kappa^{tight}}{\kappa} = \frac{1 + k_p\,\alpha_N\,\lambda}{1-\lambda}. \tag{36}$$

The proof is a direct calculation. The expression has three useful implications. First, the steepening ratio is greater than one whenever wages are staggered. Second, it rises with the wage-staggering parameter λ and with the pricing slope k_p, so wage rigidity and duration-dependent pricing reinforce the tight/slack slope gap. Third, it does not depend on matching elasticities, preference parameters, or monetary-policy coefficients. The local slope change is therefore a property of the joint wage-price mapping, not of the Taylor rule.

The economic content of Proposition 1 is the following. The contemporaneous response of inflation to labor-market tightness has two distinct channels in our model. In the tight regime, the new-hire wage equals the flexible wage, which is proportional to tightness with coefficient η. The impact of tightness on inflation is then $k_p\,\alpha_N\,\eta$, with the multiplier $k_p\,\alpha_N$ reflecting pass-through from wages to prices under duration-dependent adjustment. In the slack regime, the new-hire wage equals the incumbent wage, which loads the flexible wage with weight $1-\lambda$. The impact of tightness on the new-hire wage is therefore $(1-\lambda)\eta$, attenuated by wage staggering. A second attenuation arises because the slack-regime new-hire wage depends on current inflation through the indexation term. The slack-regime Phillips curve must therefore be solved for π_t explicitly, dividing through by $1 + k_p\,\alpha_N\,\lambda$ and further reducing the slope. The two attenuations together produce the steepening ratio of (36).

A natural empirical point of comparison is the estimate in Benigno and Eggertsson (2025), who report a much steeper reduced-form slope of $\ln\theta$ above the unitary threshold than below. Our analytical ratio $(1+k_p\alpha_N\lambda)/(1-\lambda)$ is not estimated from that reduced-form regression. It is the model-implied local steepening holding the pricing state fixed. Evaluated at the posterior mode of our estimated model, the ratio is 2.1 in the persistent state and 2.3 in the flat state. The quantitative comparison is taken up in Section 6.

Proposition 2 (Hazard-motivated coefficient multiplier)**.** *Let k_p and ψ_p denote the composite pricing slope and intrinsic propagation coefficient restricted by the fixed-hazard pricing benchmark. Holding labor-market parameters fixed, the coefficient map affects the Phillips curve through three distinct margins.*

$$\frac{\partial \kappa^{tight}}{\partial k_p} = \alpha_N \eta > 0, \tag{37}$$

$$\frac{\partial \kappa}{\partial k_p} = \frac{\alpha_N(1-\lambda)\eta}{(1+k_p\alpha_N\lambda)^2} > 0, \tag{38}$$

$$\frac{\partial \kappa_w}{\partial k_p} = \frac{\alpha_N\lambda}{(1+k_p\alpha_N\lambda)^2} > 0. \tag{39}$$

Thus, the hazard-motivated coefficient restrictions are not equivalent to adding an ad-hoc lag of inflation. They jointly discipline the impact slope, the wage-state feedback, and intrinsic propagation within the reduced-form Phillips curve.

The proposition is useful because it separates two roles of the hazard-motivated coefficient map. A larger k_p makes marginal-cost movements pass through more strongly into inflation. A larger ψ_p makes inflation more persistent. The two objects are restricted jointly by the fixed-hazard benchmark, but in the estimated model they are reduced-form coefficients rather than structural primitives of a switching-hazard economy. This is the key difference from a specification that simply appends lagged inflation to an otherwise unchanged Phillips curve. In that case, the persistence term can move without changing the price-setting selection margin. In the hazard model, the same duration structure governs impact pass-through and dynamic propagation. This is why the hazard matters for both local Phillips-curve slopes and the dynamic survival of inflationary impulses.

Proposition 3 (Impact–persistence separation)**.** *Define*

$$B(\alpha) \equiv 1-\beta(1-\alpha), \qquad D(\alpha,\varphi) \equiv 1-\alpha-\varphi B(\alpha), \tag{40}$$

and consider the admissible region $0<\alpha<1$, $\varphi \geq 0$, and $D(\alpha,\varphi)>0$. For a common $\Theta_p > 0$, $k_p(\alpha,\varphi)$ in (18) is strictly increasing in each pricing primitive, while $\psi_p(\alpha,\varphi)$ is strictly increasing in φ and weakly increasing in α (strictly whenever $\varphi > 0$). Consequently, let L and H be two pricing states labeled so that $\varphi_H \geq \varphi_L$. The impact–persistence ordering

$$k_L > k_H \qquad \text{and} \qquad \psi_L < \psi_H \tag{41}$$

requires $\alpha_L > \alpha_H$. More precisely, it holds if and only if

$$\varphi_H D_L > \varphi_L D_H, \tag{42}$$

$$(\alpha_L+\varphi_L)(B_L+\beta^2\varphi_L)D_H > (\alpha_H+\varphi_H)(B_H+\beta^2\varphi_H)D_L, \tag{43}$$

where $B_j = B(\alpha_j)$ and $D_j = D(\alpha_j,\varphi_j)$.

Proposition 3 rules out a scalar "more duration dependence" account of the frontier. Holding the initial reset rate fixed, a larger duration slope raises intrinsic continuation but also raises impact pass-through. The opposite ranking in (41) therefore requires the whole adjustment profile to rotate. The high-pass-through state combines a high initial reset rate with a flat age profile. The persistent state combines a lower initial reset rate with a steeper age profile. The posterior-mode estimates satisfy both frontier conditions and generate the crossed adjustment profiles shown in Figure 3. The estimated frontier is therefore a rotation in price adjustment across vintages.

Proposition 4 (Pressure-pass-through complementarity). *Let $z_t \equiv \hat{v}_t + \hat{\nu}_t$ collect non-labor cost-push disturbances, and fix expectations, lagged inflation, and the inherited wage state. The impact component of inflation generated by current pressure and cost-push terms can be written as*

$$B_t^T(s_t^p) = k_p(s_t^p)\Big[\alpha_N \eta \hat{\theta}_t + z_t\Big], \tag{44}$$

$$B_t^S(s_t^p) = \frac{k_p(s_t^p)}{1 + k_p(s_t^p)\alpha_N \lambda}\Big[\alpha_N(1-\lambda)\eta\hat{\theta}_t + \alpha_N \lambda \hat{w}_{t-1}^{\text{ex}} + z_t\Big], \tag{45}$$

for tight and slack labor markets, respectively. Then

$$\frac{\partial^2 B_t^T}{\partial \hat{\theta}_t \, \partial k_p} = \alpha_N \eta > 0, \qquad \frac{\partial^2 B_t^T}{\partial z_t \, \partial k_p} = 1 > 0, \tag{46}$$

$$\frac{\partial^2 B_t^S}{\partial \hat{\theta}_t \, \partial k_p} = \frac{\alpha_N(1-\lambda)\eta}{(1 + k_p \alpha_N \lambda)^2} > 0, \qquad \frac{\partial^2 B_t^S}{\partial z_t \, \partial k_p} = \frac{1}{(1 + k_p \alpha_N \lambda)^2} > 0. \tag{47}$$

Thus, pricing pass-through raises the inflationary value of both labor-market pressure and non-labor cost-push impulses. If a surge onset requires $B_t^r(s_t^p)$ to exceed a threshold net of lagged and expectational terms, neither the labor-market state nor the pricing state is sufficient on its own. The high-pass-through pricing state raises the inflationary content of a given pressure impulse, while labor tightness raises the pressure available to be transmitted.

Proposition 4 is the analytical version of the paper's central empirical claim. Tightness raises the marginal-cost pressure carried by an impulse. The flat pricing state raises the pass-through of that impulse into current inflation. The interaction between these two margins determines the local onset threshold. At the same time, the pricing states lie on an impact–persistence frontier. The state with the larger $k_p(s_t^p)$ transmits more pressure immediately, whereas the state with the larger $\psi_p(s_t^p)$ attaches more intrinsic continuation to inflation. Actual inflation paths move along this frontier as shocks and regimes evolve.

Proposition 5 (Threshold dominance at the onset margin). *Let L and H denote pricing states with $k_L \equiv k_p(L) > k_H \equiv k_p(H) > 0$, and let $\bar{B} > 0$ be a surge-onset threshold for the impact component of inflation, net of lagged and expectational terms. For fixed $\hat{\theta}_t$ and $\hat{w}_{t-1}^{\text{ex}}$, define z_{rj}^* as the smallest non-labor impulse z_t such that $B_t^r(j) \geq \bar{B}$, with $r \in \{T, S\}$ and $j \in \{L, H\}$. Then*

$$z_{Tj}^* = \frac{\bar{B}}{k_j} - \alpha_N \eta \hat{\theta}_t, \tag{48}$$

$$z_{Sj}^* = \frac{\bar{B}}{k_j} + \alpha_N \lambda \bar{B} - \alpha_N(1-\lambda)\eta\hat{\theta}_t - \alpha_N \lambda \hat{w}_{t-1}^{\text{ex}}. \tag{49}$$

If

$$\hat{w}_{t-1}^{\text{ex}} < \bar{B} + \eta \hat{\theta}_t, \tag{50}$$

then

$$z_{T,L}^* < \min\{z_{T,H}^*, z_{S,L}^*, z_{S,H}^*\}. \tag{51}$$

Hence there exists a non-empty interval of intermediate impulses,

$$z_t \in \Big[z_{T,L}^*, \min\{z_{T,H}^*, z_{S,L}^*, z_{S,H}^*\}\Big), \tag{52}$$

for which the surge threshold is crossed only in state (T, L), where tight labor-market pressure meets flat, high-pass-through pricing.

Proposition 5 formalizes the local threshold role of the two model margins. It is not a statement about every possible realization. A sufficiently large disturbance can cross the threshold in other states. The result is sharper. For intermediate impulses, the state that first turns an inflationary disturbance into a surge is (T, L), because labor-market pressure lowers the required shock and the flat pricing state raises impact pass-through. This is the analytical counterpart of the surge-ignition multipliers reported in Section 6.2.

3.3 Unemployment-based representation

For empirical work, it is often more convenient to express the Phillips curve in terms of an unemployment measure rather than labor-market tightness, since unemployment is directly observable while tightness requires constructing a vacancy series. Using the notation introduced after (22), the first-order relationship $\hat{\theta}_t = -\nu_u \hat{u}_t$ provides the bridge.

Substituting $\hat{\theta}_t = -\nu_u \hat{u}_t$ into the tight-regime wage $\hat{w}_t = \eta \hat{\theta}_t$ and then into the marginal-cost equation (7) yields

$$\widehat{mc}_t = -b_u^{\text{tight}} \hat{u}_t + \hat{v}_t, \qquad b_u^{\text{tight}} \equiv \alpha_N \eta \nu_u, \tag{53}$$

where $\hat{v}_t \equiv (1 - \alpha_N)\hat{o}_t - \hat{A}_t$ collects the imported-input price and the technology disturbance. Substituting (53) into the generalized Phillips curve (17) delivers the tight-regime unemployment-based Phillips curve

$$\pi_t = \psi_p \pi_{t-1} + \beta\Big[1 + (1 - \beta)\psi_p\Big] \mathbb{E}_t \pi_{t+1} - \beta^2 \psi_p \mathbb{E}_t \pi_{t+2} - k_p b_u^{\text{tight}} \hat{u}_t + k_p (\hat{v}_t + \hat{\nu}_t). \tag{54}$$

In the slack regime, substituting the wage law of motion (28) into (7) gives

$$\widehat{mc}_t = \alpha_N \lambda \Big(\hat{w}_{t-1}^{\text{ex}} - \pi_t + \delta \mathbb{E}_t \pi_{t+1}\Big) - b_u^{\text{slack}} \hat{u}_t + \hat{v}_t, \qquad b_u^{\text{slack}} \equiv \alpha_N (1 - \lambda) \eta \nu_u. \tag{55}$$

Using (17) and collecting current inflation terms gives the slack-regime unemployment-based Phillips curve

$$\begin{aligned}
\pi_t = {} & \widetilde{\psi} \psi_p \pi_{t-1} - \widetilde{\psi} \beta^2 \psi_p \mathbb{E}_t \pi_{t+2} \\
& + \widetilde{\psi} \Big[\beta\Big(1 + (1 - \beta)\psi_p\Big) + k_p \alpha_N \lambda \delta\Big] \mathbb{E}_t \pi_{t+1} \\
& + \widetilde{\psi} k_p \alpha_N \lambda \hat{w}_{t-1}^{\text{ex}} - \widetilde{\psi} k_p b_u^{\text{slack}} \hat{u}_t + \widetilde{\psi} k_p (\hat{v}_t + \hat{\nu}_t),
\end{aligned} \tag{56}$$

with $\widetilde{\psi} \equiv 1/(1 + k_p \alpha_N \lambda)$ as in (35).

Equations (54) and (56) provide a unified unemployment-based Phillips curve with regime-dependent slope and persistence. The slope coefficients $b_u^{\text{tight}} = \alpha_N \eta \nu_u$ and $b_u^{\text{slack}} = \alpha_N (1 - \lambda) \eta \nu_u$ measure the pass-through from unemployment to marginal cost in the two regimes. Since wage adjustment is more responsive when not anchored by the wage norm, $b_u^{\text{tight}}/b_u^{\text{slack}} = 1/(1-\lambda) > 1$, with the wage-staggering parameter λ as the sole driver of the ratio. The unemployment-based representation is therefore a direct empirical counterpart of the $(\hat{\theta}_t)$-based statements of Proposition 1. Inflation is more sensitive to fluctuations in unemployment in the tight labor-market regime than in the slack one, holding the pricing regime fixed.

4 The inflation-vulnerability frontier

The stylized facts in the introduction suggest that inflation episodes differ along two dimensions, not one. The model translates this regularity into a joint Phillips-curve state. Labor-market conditions govern the wage mapping in marginal cost, and that wage is priced in an environment whose age profile of adjustment jointly governs impact transmission and intrinsic persistence. The relevant empirical object is therefore not a sequence of independent labor and pricing mechanisms.

This integration matters because a one-state Phillips curve collapses two questions into one coefficient. The first is how much marginal-cost pressure a disturbance creates. The second is how the price-setting environment distributes that pressure between current and future inflation. The state-dependent wage mapping changes the first margin. The pricing state changes both the impact coefficient k_p and the intrinsic-continuation coefficient ψ_p. The inherited wage remains a predetermined component of the slack-state Phillips curve and hence part of the same state-dependent mapping.

The resulting pricing environments form an impact–persistence frontier. The flat state combines a high initial reset rate with little variation across price vintages, delivering high contemporaneous pass-through and relatively little intrinsic continuation. The persistent state combines a lower initial reset rate with a more age-sensitive adjustment profile, delivering lower impact transmission and greater continuation. As Proposition 3 shows, this ranking is generated by a rotation of the adjustment profile rather than by the duration slope alone. High inflation and high intrinsic persistence therefore need not occur in the same quarter or even in the same pricing state. Historical episodes trace paths across this frontier as their shocks and state probabilities change.

Inflation onset is one implication of this integrated state space. A disturbance must generate pressure, and the pricing environment must transmit enough of that pressure to cross the high-inflation boundary. The vacancy–unemployment signal raises the probability of the high-pressure labor state, in which the marginal wage and marginal cost respond more sharply. The high-pass-through pricing state raises the inflationary content of the same cost movement. Proposition 4 establishes their complementarity, and Proposition 5 identifies the range of impulses for which the joint high-pressure, high-pass-through state crosses the threshold while the other states do not.

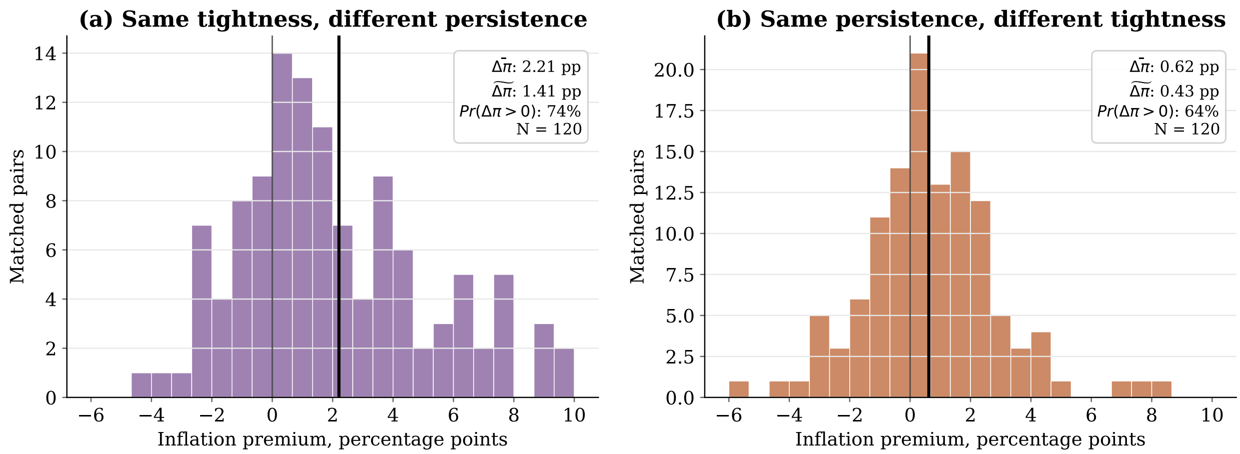
The quantitative analysis asks whether the U.S. evidence has this structure. It examines where inflation surges occur in the estimated state space, how much larger an impulse other states require to reach the same boundary, which shocks arrive in vulnerable environments, and whether duration-dependent price adjustment contains information beyond persistent disturbances. Section 6 answers these questions using estimated state histories, threshold comparisons, historical shock accounting, and state-dependent transmission. The above discussion implies that the data should not collapse onto a single state variable. If tightness and pricing persistence play distinct roles, then inflation should look different across persistence states, even at similar levels of tightness, and it should still look different across tightness states, even at similar levels of persistence. This section uses simple matched contrasts only as a descriptive screen before the full state-space estimation. They are not used as an identification device.

To avoid building the outcome mechanically into the persistence measure, inflation persistence is defined with information dated no later than $t-1$. Let

$$\widehat{\rho}^{\pi}_{t-1}(W) = \text{AR}(1) \text{ coefficient estimated from } \{\pi_q\}_{q=t-W}^{t-1}, \tag{57}$$

where W is the rolling-window length. The baseline diagnostic uses a 40-quarter window. Robustness checks use shorter and longer windows. The inflation outcome is kept separate from the construction of $\widehat{\rho}^{\pi}_{t-1}(W)$. Hence a high-persistence quarter means that inflation was persistent before the quarter being evaluated, not that the current realization is mechanically classified as persistent.

The matched comparison is correspondingly interpreted narrowly. Panel (a) of Figure 2 compares high- and low-predetermined-persistence quarters after matching on labor-market tightness. Panel (b) performs the reverse comparison, comparing high- and low-tightness quarters after matching on predetermined persistence. The matching rule is nearest-neighbor matching on the conditioning variable, on common support, with a caliper equal to one-quarter of that variable's sample standard deviation. The baseline uses replacement to avoid poor matches at the edge of the support. The relevant uncertainty is serially dependent. Accordingly, the figure reports the empirical distribution of matched differences and is interpreted as descriptive motivation, not as a causal estimate or a separate source of identification.



Notes: U.S. quarterly data, 1965Q1-2024Q4. $\Delta\pi$ is the matched inflation premium; bar, tilde, and *Pr* denote the mean, median, and positive-share of the premium. Inflation is four-quarter CPI inflation.

Figure 2: Inflation pressure beyond a single state variable.

Notes: Panel (a) matches high-predetermined-persistence quarters to low-predetermined-persistence quarters on labor-market tightness. Panel (b) matches high-tightness quarters to low-tightness quarters on predetermined inflation persistence. The distributions report the premium in four-quarter CPI inflation. A positive premium means that the first state in the comparison has higher inflation than its matched counterpart. The figure is a descriptive diagnostic. The identifying evidence comes from the estimated joint labor-pricing state model. The CPI contrast is used only to motivate the empirical mechanism.

The two matched contrasts point in the same direction. Predetermined persistence carries information after conditioning on tightness, and tightness carries information after conditioning on predetermined persistence. These descriptive facts motivate a state space in which the pressure reaching firms and the way firms transmit that pressure are jointly, rather than separately, determined.

The next section estimates that joint state structure and uses the resulting historical probabilities to study when labor-market pressure is transmitted into aggregate inflation.

5 Bayesian estimation

The empirical specification contains two regime dimensions within a single state-space system. The first captures labor-market pressure. Its observation-driven transition probabilities vary smoothly with the stationary component of the vacancy–unemployment ratio, assigning greater probability to the high-pressure labor state as labor-market pressure strengthens. The pricing chain is latent and estimated as a two-state Markov process for reduced-form pass-through and propagation coefficients. The resulting four-state system preserves the wage-selection asymmetry while allowing the data to reveal the pricing environment in which labor-market pressure is transmitted.

5.1 Data and measurement

The baseline estimation uses six quarterly U.S. observables. Inflation is measured with the GDP deflator, monetary policy with the federal funds rate, real activity with output, labor-market slack with unemployment, wage pressure with compensation, and labor-market pressure with a vacancy–unemployment signal. The measurement strategy follows the economics of the model. Nominal observables discipline inflation and policy, real and labor-market observables discipline marginal cost, and the vacancy–unemployment signal anchors the labor-pressure state. The use of nominal, real, and labor-market observables to exploit the model's cross-equation restrictions follows the empirical New Keynesian tradition (Smets and Wouters, 2007).[6]

Inflation and the policy rate anchor the nominal block. Output, unemployment, and wages discipline marginal cost and the search-and-matching labor market. The vacancy–unemployment signal moves the transition probabilities of the labor-pressure state. It does not replace the labor-market block. Search and matching, employment dynamics, and the asymmetric wage rule continue to map labor-market conditions into marginal cost. The pricing state is inferred from the joint behavior of nominal and real observables.

5.2 State discipline

The two regime dimensions draw their empirical content from different variation. The vacancy–unemployment signal shifts the transition probabilities of the labor state, while wages, unemployment, and activity discipline the marginal-cost consequences of wage selection. Conditional on those labor-market conditions, the joint dynamics of inflation and real activity reveal the pricing environment. Both states enter the same likelihood, so their separation is disciplined by the cross-equation restrictions of the model rather than by two independently estimated subsystems. Section 5.6 further asks whether the pricing-age structure survives when the persistence of every aggregate shock is re-estimated.

Historical filtering uses the observation-driven hazards in (23) and the recursive prediction in (24). The dynamic experiments answer a different question. They condition on the initial joint state. For the demand and policy innovations considered in the main analysis, the vacancy–unemployment transition signal follows its own process. Shocked and baseline simulations are paired on the same future regime sequence and the same non-target innovations. The reported responses therefore isolate transmission through the initial labor and pricing environment rather than an additional effect operating through state reclassification. This is not a fixed-state calculation because the policy functions come from the switching-model solution and incorporate the possibility of future state changes.

[6] Appendix Table B.1 reports the series definitions and transformations.

5.3 Calibration and prior specification

The quantitative specification separates parameters according to the source of their empirical discipline. Parameters that are directly informed by the six observables are estimated. Steady-state quantities and a small set of calibrated elasticities are fixed before estimation. Finally, auxiliary processes needed to retain vacancy creation, hiring, separation, and employment dynamics are calibrated rather than added to an already large estimated shock block. This division follows a standard principle in Bayesian DSGE and regime-switching models (Smets and Wouters, 2007; Herbst and Schorfheide, 2014). Priors should be sufficiently broad for the data to move common parameters, while asymmetry across states should be introduced only where it has a clear economic interpretation. It also preserves the parsimony emphasized by Baele et al. (2015) for the identification and interpretation of macroeconomic regimes and by Bianchi (2013) and Beqiraj et al. (2026) for Bayesian state-dependent models.

We begin with the baseline calibration. The quarterly discount factor is set to $\beta = 0.99$. The elasticity of substitution across differentiated goods is fixed at $\epsilon = 6$, implying a steady-state gross price markup of 1.20. The labor share is $\alpha_N = 0.90$, leaving a ten-percent share for the imported intermediate input in the production technology. The reset-price pass-through calibration is $\Theta_p = 0.60$, which corresponds to $\chi_p = (1/\Theta_p - 1)/\epsilon$ under (9). We set the matching elasticity to $\eta = 0.40$, the quarterly separation rate to $\bar{s} = 0.0723$, and steady-state unemployment to $\bar{u} = 0.04$. These values anchor the steady-state relations that map vacancies, effective job seekers, hires, and employment into labor-market tightness. The steady-state hiring-cost wedge is fixed at $\bar{\gamma}_b = 0.60$.

The two parameters governing inherited wages are held common across labor states. We set the wage-staggering parameter to $\lambda = 0.50$ and the expected-inflation term in the incumbent-wage law to $\delta = 0.50$. This symmetric baseline gives the inherited wage and the newly selected wage comparable weight in wage adjustment. The direct tightness loading in the slack branch is therefore $(1 - \lambda)\eta$, whereas the high-pressure branch loads on η. Keeping these parameters fixed and common isolates variation in the wage mapping from state-specific wage rigidity and anchors the upstream amplification measured below.

The preference and monetary-policy priors follow the locations used in the related Bayesian estimates. The intertemporal-substitution curvature σ has a Gamma prior with mean 1 and standard deviation 0.375. Monetary policy is common across the four joint states. The response to inflation has a Normal prior with mean 1.50 and standard deviation 0.30, where the mean is the midpoint of the two policy-state locations used by Beqiraj et al. (2026). The output response has a Normal prior with mean 0.125 and standard deviation 0.05, and interest-rate smoothing has a Beta prior with mean 0.60 and standard deviation 0.20. The latter is also the smoothing prior used by Bianchi (2013). Keeping the rule common is deliberate. The quantitative analysis attributes state variation to labor-market pressure and price adjustment, while allowing the common policy coefficients to be determined by the full sample. The monetary policy innovation is white noise and its standard deviation is common across states.

The estimated shock processes receive symmetric and relatively loose priors. The autoregressive coefficients of technology, markup, demand, and matching-efficiency disturbances have Beta priors with mean 0.50 and standard deviation 0.20. Their innovation standard deviations, together with that of the monetary-policy shock, have inverse-Gamma priors with scale 0.01 and shape 2. This follows the prior discipline in Beqiraj et al. (2026). Persistence is allowed to move substantially away from its prior center, while the shock scales remain positive and weakly informative. Shock persistence and volatility are common across pricing states in the benchmark.

The pricing priors distinguish the level of the reset hazard from its response to price age. The hazard level α_p has the same Beta prior in both states, with mean 0.132 and standard deviation 0.10. Thus the prior does not create state separation by assigning a different average reset frequency across regimes. State dependence enters through the duration slope φ_p. Its prior locations are 0.122 in state L and 0.322 in state H, the locations obtained from the duration-dependent pricing evidence in Sheedy (2010) and used in Beqiraj et al. (2026). The benchmark assigns both slopes a common standard deviation of 0.10. These locations give the states an economic orientation before observing the data, but leave both the hazard level and the strength of duration dependence to be estimated. We normalize the labels so that state H has the weakly larger duration slope. This removes the observationally equivalent relabeling of the two pricing states and has no effect on the likelihood under a permutation of labels.

The two pricing transition probabilities have identical Beta priors with mean 0.10 and standard deviation 0.05. At the prior mean, either pricing state has an expected duration of ten quarters. The symmetric transition prior does not favor one state in terms of persistence or ergodic frequency. This is the same prior discipline used for state transitions in the related duration-dependent estimates and is close in spirit to the persistent-regime prior used by Bianchi (2013).

The labor-pressure transition is disciplined separately because its economic timing is anchored by the vacancy–unemployment signal. The bounded hazards in (23) use $p^{\ell}_{\min} = 0.0005$, $p^{\ell}_{\text{scale}} = 0.95$, $\kappa_B = 12$, and $\tau_B = -0.05$. The first two parameters determine the floor and range of the transition probabilities, κ_B controls how sharply the state responds to the signal, and τ_B locates the transition threshold. Together with the lagged state, this specification ties the timing of high- and low-pressure spells to observed tightness before inflation is used to evaluate their consequences. Wages, unemployment, output, and the labor-market equations then determine how much marginal-cost pressure is associated with those states inside the joint likelihood.

To retain the full vacancy, hiring, separation, and employment dynamics without asking six observables to identify every auxiliary disturbance, the remaining labor-block autoregressive coefficients and innovation scales are fixed at the values obtained from the unrestricted version of the same labor block. The persistence and innovation scale of the vacancy–unemployment signal are instead obtained directly from its prepared quarterly series. Panel C of Table 1 reports these values. This leaves the technology, markup, demand, matching-efficiency, and monetary-policy innovations estimated, while preventing weakly identified auxiliary processes from absorbing variation assigned to the two economically central state dimensions.

5.4 Posterior modes

Table 2 reports the posterior mode of the estimated four-state specification. The estimates imply a standard monetary-policy response to inflation, a very small output-gap response, and persistent technology, preference, markup, and matching shocks. The estimated price-adjustment profile is central for the quantitative mechanism. The flat, high-pass-through state combines a relatively high initial reset probability with similar implied benchmark reset rates across price ages. In the persistent state, a lower initial reset probability is paired with a steeper response to price age. The associated transition probabilities imply economically meaningful spells in both pricing environments. These estimates locate the

Table 1: Calibration and prior specification.

Parameter	Description	Distribution	Location	Scale
Panel A: Common structural and shock parameters				
σ	Intertemporal substitution curvature	Gamma	1.00	0.375
ϕ_π	Taylor-rule response to inflation	Normal	1.50	0.300
ϕ_y	Taylor-rule response to activity	Normal	0.125	0.050
ρ_i	Interest-rate smoothing	Beta	0.60	0.200
ρ_A	Technology-shock persistence	Beta	0.50	0.200
ρ_μ	Markup-shock persistence	Beta	0.50	0.200
ρ_ξ	Demand-shock persistence	Beta	0.50	0.200
ρ_m	Labor/search-shock persistence	Beta	0.50	0.200
σ_A	Technology-shock standard deviation	Inv. gamma	0.01	2.00
σ_ξ	Demand-shock standard deviation	Inv. gamma	0.01	2.00
σ_μ	Markup-shock standard deviation	Inv. gamma	0.01	2.00
σ_m	Labor/search-shock standard deviation	Inv. gamma	0.01	2.00
σ_i	Monetary-policy shock standard deviation	Inv. gamma	0.01	2.00
Panel B: Duration-dependent pricing states				
$\alpha_{p,L}$	Price-reset hazard level in state L	Beta	0.132	0.100
$\alpha_{p,H}$	Price-reset hazard level in state H	Beta	0.132	0.100
$\varphi_{p,L}$	Duration dependence in state L	Normal	0.122	0.100
$\varphi_{p,H}$	Duration dependence in state H	Normal	0.322	0.100
p^p_{LH}	Transition probability from L to H	Beta	0.100	0.050
p^p_{HL}	Transition probability from H to L	Beta	0.100	0.050
Panel C: Calibrated state-selection and auxiliary-process parameters				
λ	Wage staggering parameter	Fixed	0.50	–
δ	Incumbent-wage indexation parameter	Fixed	0.50	–
$p^\ell_{\min}$	Lower bound of each labor-state transition hazard	Fixed	0.0005	–
p^ℓ_{scale}	Scale of the labor-state transition hazards	Fixed	0.95	–
κ_B	Sensitivity of labor-state transitions to the signal	Fixed	12.00	–
τ_B	Center of the labor-state transition threshold	Fixed	-0.05	–
ρ_q	Vacancy/posting-process persistence	Fixed	0.6162	–
ρ_{γ_c}	Consumption-preference persistence	Fixed	0.6285	–
ρ_{γ_b}	Bargaining/preference persistence	Fixed	0.9824	–
ρ_{ϕ_w}	Wage-distortion persistence	Fixed	0.5927	–
ρ_s	Separation-process persistence	Fixed	0.5211	–
ρ_{bev}	Vacancy–unemployment signal persistence	Fixed	0.9340	–
σ_q	Vacancy/posting-process innovation scale	Fixed	0.00277	–
σ_{γ_c}	Consumption-preference innovation scale	Fixed	0.00198	–
σ_{γ_b}	Bargaining/preference innovation scale	Fixed	0.01624	–
σ_{ϕ_w}	Wage-distortion innovation scale	Fixed	0.00122	–
σ_s	Separation-process innovation scale	Fixed	0.00365	–
σ_{bev}	Vacancy–unemployment signal innovation scale	Fixed	0.09312	–

Notes: Location and scale denote the prior mean and standard deviation for normal, beta, and gamma priors. For inverse-gamma priors, the table reports the scale and shape hyperparameters used for shock standard deviations. State L is the flat, high-pass-through pricing state and state H is the persistent pricing state. Panel C reports fixed quarterly values. The core steady-state calibration is described in the text.

two states on the impact–persistence frontier used below. The vacancy–unemployment signal disciplines labor pressure, while the pricing environment is inferred from the joint dynamics of the six observables.

Table 2: Posterior-mode estimates.

Parameter	Description	Mode	s.e.
Panel A: Common block			
σ	Intertemporal substitution curvature	4.751	0.489
ϕ_π	Taylor-rule response to inflation	2.015	0.076
ϕ_y	Taylor-rule response to activity	0.005	0.038
ρ_i	Interest-rate smoothing	0.780	0.019
ρ_A	Technology-shock persistence	0.991	0.006
ρ_μ	Markup-shock persistence	0.798	0.055
ρ_ξ	Demand-shock persistence	0.945	0.015
ρ_m	Labor/search-shock persistence	0.970	0.008
σ_A	Technology-shock standard deviation	0.0051	0.0003
σ_ξ	Demand-shock standard deviation	0.0306	0.0078
σ_μ	Markup-shock standard deviation	0.0064	0.0023
σ_m	Labor/search-shock standard deviation	0.0204	0.0013
σ_i	Monetary-policy shock standard deviation	0.0027	0.0002
Panel B: Pricing block			
$\alpha_{p,L}$	Price-reset hazard level in state L	0.451	0.035
$\varphi_{p,L}$	Duration dependence in state L	0.040	0.038
$\alpha_{p,H}$	Price-reset hazard level in state H	0.047	0.020
$\varphi_{p,H}$	Duration dependence in state H	0.239	0.028
p^p_{LH}	Transition probability from L to H	0.034	0.028
p^p_{HL}	Transition probability from H to L	0.053	0.029

Notes: The table reports posterior modes and local posterior standard errors at the selected four-state specification. Pricing state L denotes flat, high-pass-through pricing; pricing state H denotes the persistent pricing state. The estimated pricing transition probabilities imply persistent pricing states, while the labor-state transition is disciplined by the vacancy–unemployment signal.

Figure 3 opens the pricing-state mechanism. The economically revealing feature is that the two reset schedules cross. The flat state begins with a high reset probability and quickly approaches a nearly age-invariant profile. A firm is therefore likely to reset even when its price is young. Nearly one half of the stationary price distribution is concentrated among newly reset prices, and current marginal cost receives a large weight in reset-price decisions. Inflation moves quickly because a large share of firms incorporates current conditions without waiting for its price to become old.

The persistent state reorganizes this timing. Young prices are unlikely to change, while adjustment becomes progressively more likely as a price ages. Old vintages accumulate, and firms that eventually reset place more weight on the marginal-cost path beyond the current quarter. The resulting persistence does not come from attaching an arbitrary lag to the Phillips curve. It comes from the composition of prices still in effect and from the different dates at which those vintages return to the market.

Panel D summarizes the macroeconomic rotation. Impact pass-through is about 5.2 times larger in the flat state, while intrinsic continuation is about 3.4 times larger in the persistent state. The same marginal-cost impulse can consequently produce either a sharp initial movement or a more gradual, long-lived inflation response. This distinction is central for the historical analysis. The pricing environment in which inflation takes hold need not be the environment that carries it forward.[7]

[7] Appendix C reports posterior means, credible intervals, and multi-chain convergence diagnostics.

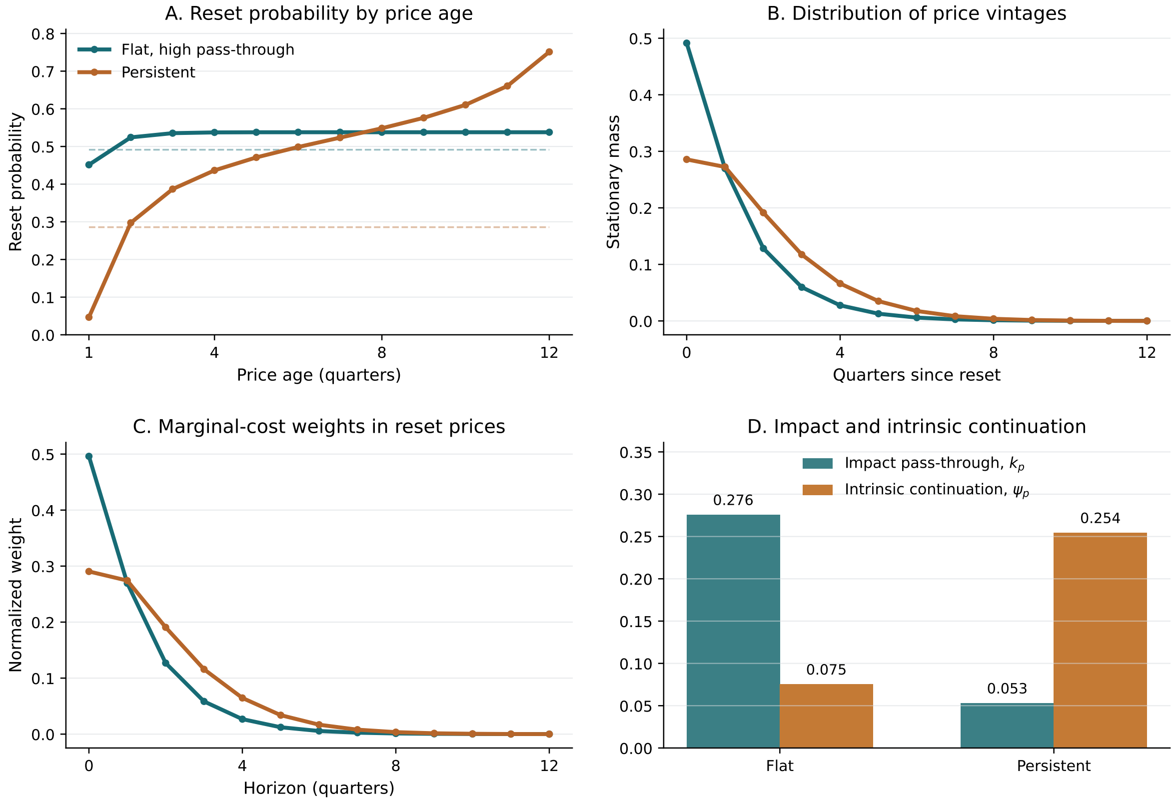


Figure 3: How the price-adjustment profile separates impact from persistence.

Notes: The figure uses posterior-mode pricing parameters. Panel A reports the reset probability by price age. Dashed lines are constant hazards with the same mean duration and are shown only as scale references. Panel B reports the stationary distribution of price vintages. Panel C reports the discounted marginal-cost weights in reset-price decisions, proportional to $\beta^{\ell}\varsigma_{p,\ell}$ and normalized to sum to one. Panel D reports impact pass-through k_p and intrinsic continuation ψ_p. The reset schedules cross at about eight quarters. The crossing rotates both the distribution of prices in effect and the horizon of marginal cost entering a reset decision, thereby separating immediate pass-through from intrinsic continuation.

5.5 Bayesian evidence for the joint state structure

The model comparison follows the economic construction of the state space. Table 3 begins from a constant-state benchmark, then allows labor-market pressure to move between slack and tight conditions, and finally adds variation in the pricing environment. The three specifications share the same observables, sample, policy rule, shock processes, model equations, and priors on common parameters. The constant-pricing specifications retain one generalized duration-dependent hazard. The comparison therefore asks whether the data support progressively richer state dependence while holding the common economic blocks fixed.

Table 3: Bayesian evidence across nested state structures.

Specification	Log MDD	Δ from benchmark	Δ from prior stage
Constant-state benchmark	4059.79	0.00	—
Labor-state model	4094.93	35.14	35.14
Joint labor-pricing model	4099.90	40.11	4.97

Notes: The table reports Laplace log marginal data densities for a nested sequence of models estimated on the same six U.S. observables over 1965Q1–2022Q4. Shared parameters retain common priors, and model prior probabilities are equal. The constant-state benchmark remains in the low-pressure, anchored-wage branch and retains one generalized duration-dependent pricing hazard. The labor-state model adds high–low pressure variation across the two wage-selection branches. The joint model also allows the pricing environment to vary. Log marginal data densities are normalized for the prior mass associated with pricing-state labeling. The third column reports cumulative gains relative to the benchmark; the final column reports the gain at each successive stage.

The first step in the sequence is quantitatively large. Allowing labor-market pressure to change state raises the log marginal data density from 4059.8 to 4094.9, a gain of 35.1 log points. This improvement is measured after the integrated likelihood penalizes the additional state parameters. The result supports the model's first economic margin. The mapping from aggregate conditions into marginal-cost pressure changes across labor-market states.

The pricing environment supplies the second step. Once the labor state is allowed to vary, adding pricing-state variation raises the log marginal data density by a further 5.0 points, from 4094.9 to 4099.9. The joint-state model therefore has the highest model evidence and exceeds the constant-state benchmark by 40.1 log points overall. The additional pricing state refines how changes in labor-market pressure are transmitted through price setting. Its contribution is incremental to the substantial gain already delivered by labor-state variation.

This ordered comparison gives the two margins a clear quantitative hierarchy without separating them economically. Labor-state variation provides the first-order improvement in fit, while pricing-state variation completes the transmission mechanism governing inflation pass-through and propagation. The historical and state-dependent evidence in the next section examines what this joint structure implies for the onset and propagation of inflation.

5.6 Price duration or persistent shocks?

Why does inflation inherit momentum? One possibility is that the disturbances moving marginal cost are themselves persistent. Another is that the price distribution carries memory because a firm's adjustment decision depends on how long its price has been in place. These explanations have different economic content. A persistent shock keeps pushing marginal cost in the same direction. Duration dependence instead changes which price vintages remain active and when their accumulated adjustment reaches the aggregate price level.

We distinguish these channels by re-estimating a nested model in which the two pricing states can differ in their reset frequency, but reset rates do not vary with price age. Technology, markup, demand, and labor-search persistence are reoptimized in both specifications. The age-invariant model is therefore free to place additional persistence in the shocks whenever that helps to replicate inflation dynamics. The comparison asks whether the distribution of price vintages still contains information after this reallocation has taken place.

It does. Allowing price age to shape adjustment raises the pre-2020 log marginal data density by 6.7 points and the joint 2020–22 log predictive density by 23.2 points. The four-quarter inflation forecast error falls by about 0.1 percentage point. At the same time, estimated markup-shock persistence falls from about 0.9 to 0.8. The model does not obtain the gain by relabeling a persistent markup disturbance as persistent pricing. Once the price-vintage distribution is allowed to propagate marginal cost, less persistence is required from the disturbance itself.

Table 4: Price duration and shock persistence.

Pricing specification	Log MDD	2020–22 score	$\text{RMSE}_{\pi^{(4)}}$	ρ_A	ρ_μ	ρ_ξ
Age-invariant reset rates	4199.6	-357.0	2.0	0.995	0.910	0.941
Duration-dependent reset rates	4206.3	-333.8	1.9	0.994	0.835	0.941

Notes: Log marginal data densities use observations through 2019Q4. Holdout scores and four-quarter inflation RMSEs use 2020Q1–2022Q4. The comparison keeps labor-state variation and pricing-state variation in both rows. The first row permits state-dependent reset frequencies but makes reset rates invariant to price age; the second allows duration dependence. All shock persistence parameters are re-estimated in each specification.

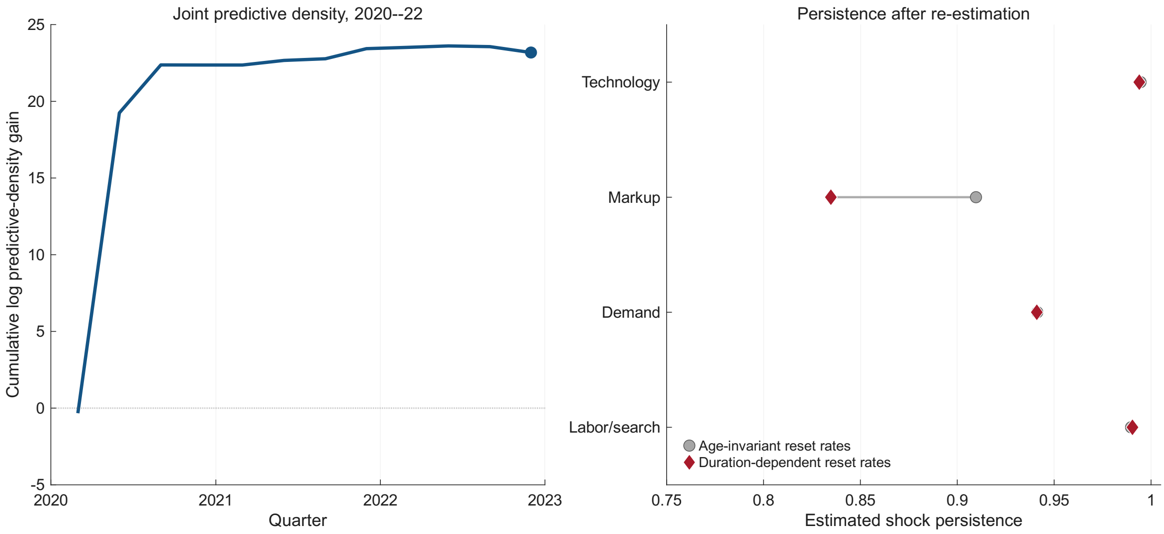


Figure 4: Price duration after re-estimating shock persistence.

Notes: The left panel cumulates the difference in joint one-step log predictive densities between duration-dependent and age-invariant reset rates over 2020–22. A positive value favors duration dependence. The gain emerges when the pandemic changes the joint distribution of the observables and remains through the inflation surge. The right panel compares the re-estimated shock-persistence coefficients. Both models use the same final-sample transformations, so the holdout evaluates predictive content under a common information treatment rather than a real-time data vintage.

The timing of the predictive gain is economically informative. It appears early in the holdout, before inflation reaches its eventual peak, and is not subsequently reversed. Price age therefore helps the model read the changing transmission environment, not merely fit the most visible part of the inflation surge after it occurs. The result gives the pricing frontier an independent empirical role: aggregate disturbances determine how marginal cost moves, while the age distribution of prices determines how that movement is converted into current inflation and carried forward.

The same conclusion is therefore not confined to full-sample model fit. Table 4 and Figure 4 show that the duration-dependent pricing specification also improves pre-2020 model evidence and the joint one-step predictive density over 2020–22, even after the aggregate-shock persistence parameters are re-estimated. This gives the pricing-age structure an independent empirical role relative to persistent aggregate disturbances.

6 Inflation vulnerability across estimated states

This section organizes the quantitative results around three questions. First, where in the estimated joint state space does inflation become most vulnerable? Second, how does the initial labor–pricing environment change the disturbance required to cross a common inflation boundary? Third, which estimated shocks supplied the inflationary pressure during the main postwar surges? We answer these questions using the historical state map and cross-state vulnerability statistics, a state-conditioned threshold experiment, and a compact historical shock decomposition.[8]

6.1 The historical location of inflation vulnerability

Throughout the historical analysis, a surge window refers to one of the post-1965 episodes dated in Benigno and Eggertsson (2025). These are 1968Q2–1971Q3, 1973Q1–1982Q4, and 2021Q2–2023Q2. Because the estimation sample ends in 2022Q4, the last window is truncated at that date in the benchmark results. The common surge boundary is 1.41 percentage points of four-quarter inflation above its estimation-sample mean, the smallest window-average deviation among these episodes. In the first-order threshold calculations below, changing the boundary rescales the absolute shock required to reach it but leaves the reported cross-state ratios unchanged.

The first question is where inflation surges begin. Figure 5 plots the two state variables that matter for this question. They are the probability of the high-pressure labor state and the probability of the persistent pricing state. The bottom panel combines them into the joint state. The striking pattern is that the surge windows begin in the flat, high-pass-through pricing environment rather than in the persistent pricing state. Labor-market pressure helps create the inflationary margin. The flat state transmits that pressure rapidly into prices. This is the high-impact side of the estimated impact–persistence frontier.[9]

[8] Additional variance decompositions, shock-by-shock threshold exercises, interaction accounting, dynamic paths, and auxiliary state-alignment diagnostics are reported in the appendix as supporting model diagnostics.

[9] Appendix G reports the surge-window alignment table. The main text keeps the historical state map and the full-sample vulnerability table as the two primary location diagnostics.

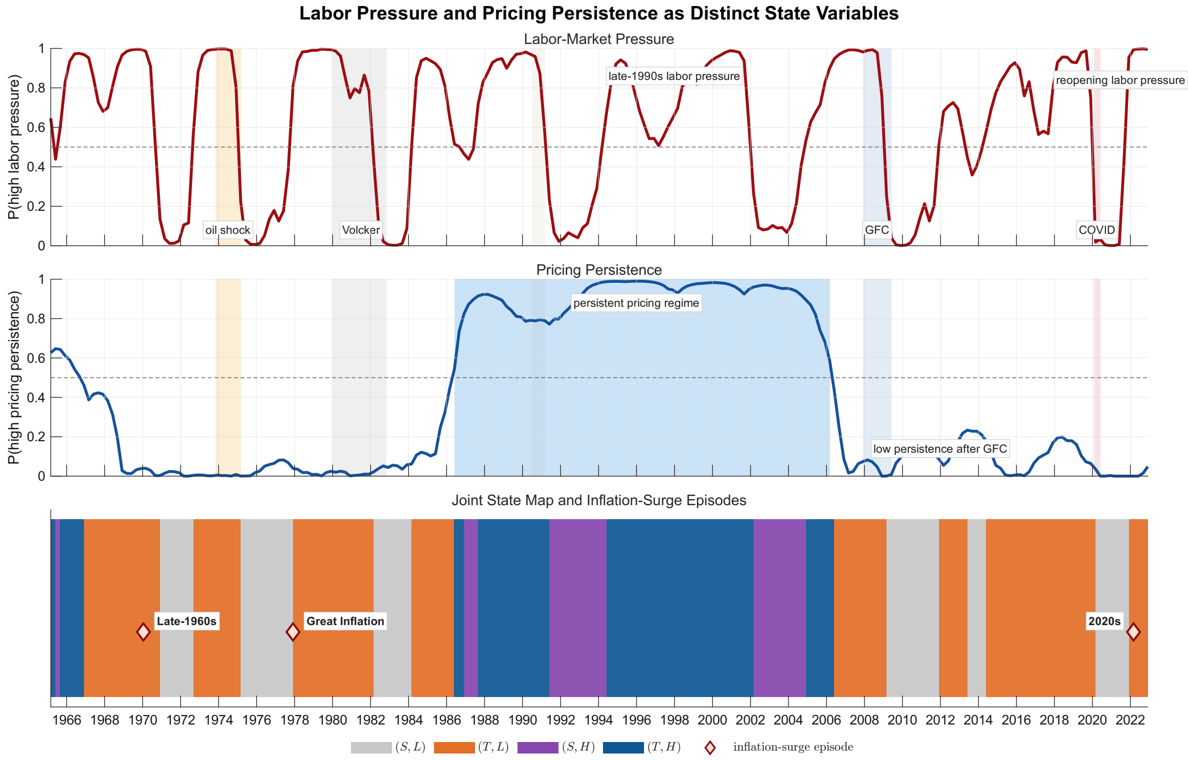


Figure 5: Estimated labor and pricing states.

Notes: Panel A reports the smoothed probability of the high-pressure labor-market state. Panel B reports the smoothed probability of the persistent pricing state. Panel C maps the joint state and marks the post-1965 U.S. inflation-surge episodes. The figure shows where labor-market pressure and pricing pass-through coincide in the estimated state space.

The next question is whether this vulnerable state is a rare label attached to a few famous episodes or a recurring macroeconomic environment. The distinction matters. A state that appears only during known surges would describe those episodes without identifying a broader economic environment. By contrast, a state that recurs throughout the sample but is disproportionately represented in the upper tail of inflation identifies a systematic change in vulnerability. Table 5 reports this comparison over the quarters for which four-quarter inflation is available.

The (T, L) state accounts for 42.4 percent of the sample, but for 61.4 percent of inflation volatility, 62.1 percent of high-inflation quarters, and 66.7 percent of surge-window quarters. It is therefore not merely frequent. It carries substantially more inflation risk than its time share would predict. The excess concentration is 19.7 percentage points for high-inflation quarters and 24.3 points for surge-window quarters. The (S, L) state is also over-represented, revealing the first economic hierarchy in the table. High pricing pass-through opens the inflation margin, while labor pressure determines how intensely the economy loads on that margin.

Table 5: Joint-state inflation vulnerability.

Panel A: Inflation vulnerability

(s_t^w, s_t^p)	Sample	Mean π_4	Std. π_4	Var. share	High-π share	High-π risk
(T, L)	42.4	0.71	2.60	61.4	62.1	1.47
(T, H)	25.3	-0.99	0.85	8.6	0.0	0.00
(S, L)	21.0	0.43	2.48	26.1	37.9	1.81
(S, H)	11.4	-1.21	0.49	3.9	0.0	0.00

Panel B: Real and labor-market volatility

(s_t^w, s_t^p)	Sample	Surge-window share	Surge risk	Std. Y	Std. u	Std. w
(T, L)	42.4	66.7	1.57	1.07	0.27	0.96
(T, H)	25.3	0.0	0.00	0.90	0.15	1.01
(S, L)	21.0	33.3	1.59	1.36	0.23	1.48
(S, H)	11.4	0.0	0.00	0.58	0.10	1.02

Notes: The state column reports (s_t^w, s_t^p), where T and S denote tight and slack labor-market states and L and H denote the flat, high-pass-through and persistent pricing states. States are assigned by dominant smoothed probabilities. π_4 is the four-quarter GDP-deflator inflation deviation from the estimation-sample mean, in percentage points. The high-inflation threshold is the sample top quartile, $\pi_4 \geq 1.10$. Var. share is the state's share of total squared inflation deviations. High-π share is the share of all high-inflation quarters occurring in the state. High-π risk is $P(\text{high inflation} \mid \text{state})/P(\text{high inflation})$; values above one indicate over-representation among high-inflation quarters. Surge-window share and surge risk use the post-1965 inflation-surge windows reported in the historical-state analysis. Y and w are reported in log points; u is the unemployment-rate deviation in percentage points.

The real-side columns sharpen the comparison within the high-pass-through pricing environment. Output and wage growth are more volatile in (S, L) than in (T, L), yet the latter carries twice as many surge-window observations. Conditional on high pricing pass-through, the concentration of surge observations therefore rises with labor-market pressure rather than with the unconditional volatility of output or wages. Conversely, the two persistent pricing states occupy 36.7 percent of the sample but contain none of the high-inflation or surge-window quarters. They are economically meaningful continuation states, not the usual location of inflation onset.[10]

6.2 State-dependent surge thresholds and transmission

The historical evidence describes where surges occur in the estimated state space. The next question is why those states are more vulnerable inside the model. Figure 6 asks how large a disturbance must be to push inflation above a common surge boundary in each state. This is a threshold experiment rather than a reconstruction of any one historical episode. It isolates the state dependence of inflation vulnerability under a common disturbance.

The economy is initialized in one of the four joint states of the estimated state-space model. For the focal demand and policy innovations, the vacancy–unemployment transition signal follows its own process. The reported generalized impulse responses pair shocked and baseline simulations on the same future labor-pricing regime sequence and the same non-target innovations, and average the shocked-minus-baseline difference over those common paths. They therefore measure how the initial environment shapes transmission without adding a separate reclassification margin. Because the policy functions

[10] Appendix G reports the graphical version of the same vulnerability comparison.

are obtained from the switching-model solution, equilibrium decisions continue to incorporate the possibility of future state changes. For each shock, the impulse size is chosen so that the (T, L) response just touches the surge boundary. The other states are then evaluated under the same disturbance. The boundary is the conservative post-1965 empirical cutoff used above. It equals the smallest surge-window mean four-quarter inflation deviation in the historical decomposition.

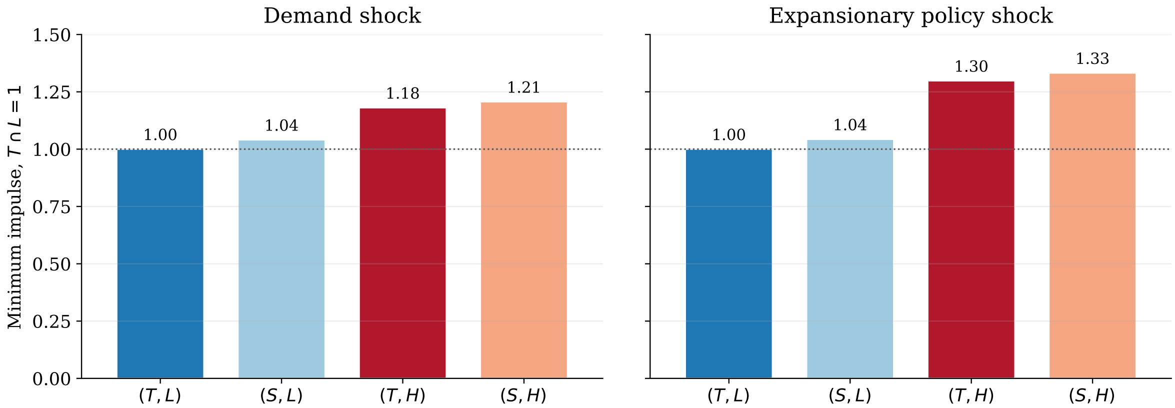


Figure 6: The impulse required to generate an inflation surge.

Notes: The bars report the minimum impulse required for peak four-quarter inflation to reach the empirical surge boundary. For each shock, the requirement in (T, L) is normalized to one. The economy is initialized in the indicated joint state, with future regime paths simulated under the estimated transition structure and held common across shocked and baseline paths. The demand shock organizes the Late-1960s and Great Inflation windows. The expansionary policy shock organizes the post-2020 window. All parameters are evaluated at their posterior modes.

Figure 6 gives the main threshold ranking. For a demand shock, the (S, L) state requires a shock about 4 percent larger than (T, L) to reach the same surge boundary, while (T, H) and (S, H) require shocks about 18 and 21 percent larger. For an expansionary policy shock, (S, L) requires a shock about 4 percent larger than (T, L), while (T, H) and (S, H) require shocks about 30 and 33 percent larger. These multipliers summarize the state dependence of peak four-quarter inflation under a common shock normalization. The final-inflation ranking is dominated by the price-adjustment environment, while labor-market pressure operates through the upstream margin examined next.

The threshold multiplier is deliberately a final-outcome statistic. It compresses the wage response, the formation of marginal-cost pressure, and the pricing of that pressure into one number. Figure 7 opens this transmission map. Holding the pricing state fixed, it normalizes the slack response to one at each stage and follows the tightness premium from the wage paid to new hires to the contemporaneous forcing term in the Phillips curve and, finally, to contemporaneous inflation. The middle object is computed directly from the estimated equation,

$$\mathcal{P}_t(s) = \kappa_w(s)\widehat{w}^{ex}_{t-1} + \kappa_\theta(s)\widehat{\theta}_t + \kappa_\nu(s)\widehat{\nu}_t,$$

where $\widehat{\nu}_t$ is the model's complete cost-push composite.

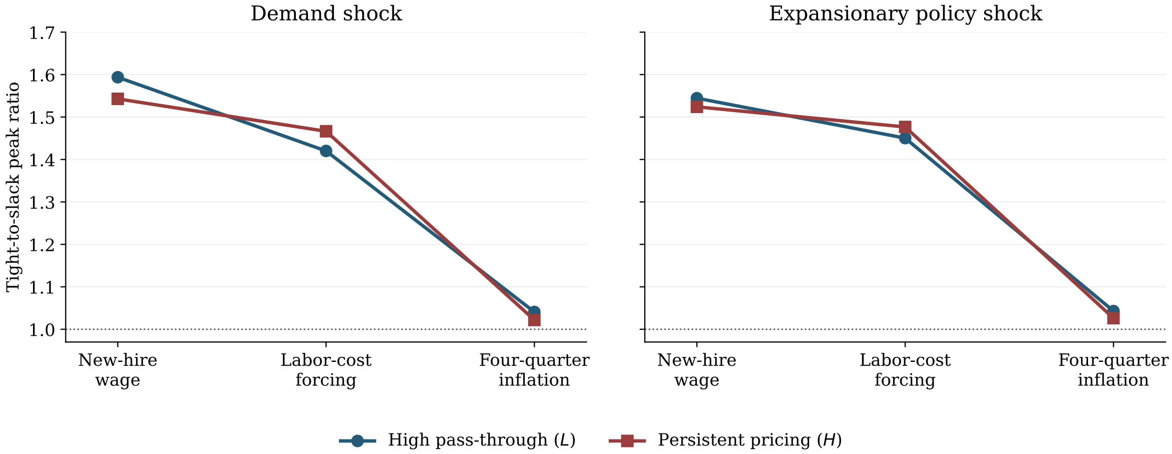


Figure 7: The pressure-transmission ladder.

Notes: Each point is the ratio of the tight-state impact response to the slack-state impact response within the same pricing environment. The first stage is the wage of new hires. The second is the exact contemporaneous labor-cost forcing $\mathcal{P}_t(s)$ in the Phillips curve. The third is the contemporaneous inflation response. Values above one measure amplification from labor-market tightness. The responses are common-path generalized impulse responses evaluated at posterior-mode parameters. Shocked and baseline paths share the same future regime realizations when future reduced-form coefficient regimes are simulated.

The ladder locates the quantitative role of tightness before final inflation is observed. Under the demand shock, the model-implied tight-state new-hire wage response is 59 percent larger than the slack response in L and 54 percent larger in H. The corresponding labor-cost forcing is 42 and 47 percent larger. For the expansionary policy shock, the wage premia are 54 and 52 percent, and the forcing premia are 45 and 48 percent. These magnitudes are equilibrium implications of the estimated system evaluated at the calibrated labor parameters. They quantify the pressure created by the state-dependent wage mapping before general-equilibrium feedback and price adjustment reshape it. At impact, the corresponding inflation premia are about 16 and 10 percent for the demand shock and 22 and 15 percent for the policy shock in states L and H, respectively. The premium narrows to 2–4 percent at the peak of four-quarter inflation because that later outcome incorporates intertemporal demand, policy feedback, expectations, and the price-adjustment coefficient map. The high-pressure labor state amplifies the wage and cost responses in the model. The pricing state determines how much of that pressure appears in current inflation and how it is distributed over time.

The passage from first-stage pressure to peak inflation is economically informative. A comparison of inflation paths alone places most of the visible separation at the final pricing margin. The ladder reveals the division of labor inside the mechanism. Labor-market pressure creates a large wage and marginal-cost impulse, while price adjustment selects the fraction that becomes current inflation and its distribution over time.[11]

Monetary policy also changes the supply of effective job search. Graves et al. (2026) show that a contraction induces non-employed workers to search more actively and employed workers to quit less often, cushioning the fall in employment. In our model this response enters upstream of price setting by changing the stock of effective job seekers and hence labor-market pressure. Appendix F combines this worker-side margin with the estimated state space. A stronger effective-search response cushions employment losses in

[11] Appendix E reports the complete dynamic paths behind the threshold multipliers.

every joint state, while its inflation consequences depend on the shape of price adjustment. The amount of adjustment absorbed by employment and the amount reaching inflation are therefore jointly determined along a single transmission chain.

The pressure-transmission ladder also clarifies how the paper complements the tightness view of nonlinear Phillips curves. Tight labor markets are essential to the pressure channel because they raise marginal-cost pressure. But pressure is most inflationary when prices are in the high-pass-through state. Conversely, high pass-through is most dangerous when the labor market supplies pressure. Across the state space, the multiplier calculation ranks vulnerability by the disturbance required to cross a common inflation boundary. The smallest impulse is required in (T, L).

Figures 6 and 7 focus on demand and expansionary policy shocks because those shocks discipline the postwar surge windows in the estimated historical accounting. Supply, markup, and wage-state disturbances load on different margins of the model and can produce different state rankings.[12] The atlas identifies (T, L) as the most vulnerable environment for the historically relevant demand, policy, and labor-pressure disturbances, while locating other shocks on the margins through which they operate.

6.3 Shock accounting across postwar inflation surges

Figure 5 also raises a sharper question. The high-pressure, high-pass-through state occurs outside the three surge windows. Why does inflation take hold in some of those quarters but not in others? The answer separates the state of the economy from the shocks arriving in it. A vulnerable state changes the conversion of pressure into prices. It does not create pressure in the absence of an inflationary disturbance.

We measure the force delivered by the estimated shocks using

$$I_t^+ = \sum_{j \in \mathcal{J}} \max\{HD_{j,t}^{\pi_4}, 0\}, \tag{58}$$

where $HD_{j,t}^{\pi_4}$ is shock block j's historical-decomposition contribution to four-quarter inflation and $\mathcal{J}$ contains technology, demand, markup, labor/search, and policy-rule innovations. The measure adds the positive forces reaching inflation before offsetting contributions are netted out. It preserves the economic distinction at issue. The estimated state governs transmission, while I_t^+ records how much inflationary force the realized disturbances deliver. Because the contributions already embed the model's historical propagation, I_t^+ is an ex post accounting measure. The fixed-shock comparisons below provide the corresponding transmission experiment.

[12] Appendix J reports an estimated-shock atlas that repeats the same normalization for all shocks whose persistence and standard deviations are estimated.

Table 6: When an inflation-prone state becomes a surge.

Panel A: Pricing environment and shock arrival					
Environment	Positive contribution	Obs.	Mean I_t^+	Mean π_4	High-π / surge incidence
Persistent pricing	Below median	69	0.4	-1.3	0.0 / 0.0
Persistent pricing	Above median	15	1.6	-0.1	0.0 / 0.0
Flat, high pass-through	Below median	45	0.6	-1.6	0.0 / 0.0
Flat, high pass-through	Above median	100	3.2	1.6	58.0 / 60.0
Panel B: Labor pressure within the high-pass-through environment					
Environment	Positive contribution	Obs.	Mean I_t^+	Mean π_4	High-π / surge incidence
Low pressure	Below median	6	0.8	-2.0	0.0 / 0.0
Low pressure	Above median	42	2.6	0.8	52.4 / 47.6
High pressure	Below median	39	0.5	-1.5	0.0 / 0.0
High pressure	Above median	58	3.6	2.2	62.1 / 69.0

Notes: I_t^+ is the sum of the positive technology, demand, markup, labor/search, and policy-rule contributions to four-quarter inflation in the estimated historical decomposition. The median cutoff is 1.1 percentage points. High inflation is the upper quartile of four-quarter inflation deviations, a cutoff of 1.1 percentage points. The final column reports, respectively, the percentage of observations classified as high inflation and the percentage falling in the post-1965 surge windows. States are assigned using dominant smoothed probabilities.

Table 6 produces two results. First, sizable positive contributions become high inflation only in the flat pricing environment. Above the median of I_t^+, average inflation is 1.6 percentage points in that state, and roughly 60.0 percent of observations belong to the high-inflation tail or a surge window. In the persistent state, the corresponding mean is close to zero and neither outcome occurs. The state with high pass-through is therefore the environment in which a sufficiently large disturbance can become an inflation surge.

Second, labor pressure is associated with the intensity of the historical realization. Within the flat pricing state and above the median contribution, mean I_t^+ is 3.6 percentage points under high pressure and 2.6 under low pressure. Average four-quarter inflation is correspondingly 2.2 and 0.8 points, while the share of surge-window quarters rises from 47.6 to 69.0 percent. The threshold experiments above hold the shock fixed and isolate how much of this difference comes from transmission rather than from the historical shock mix.

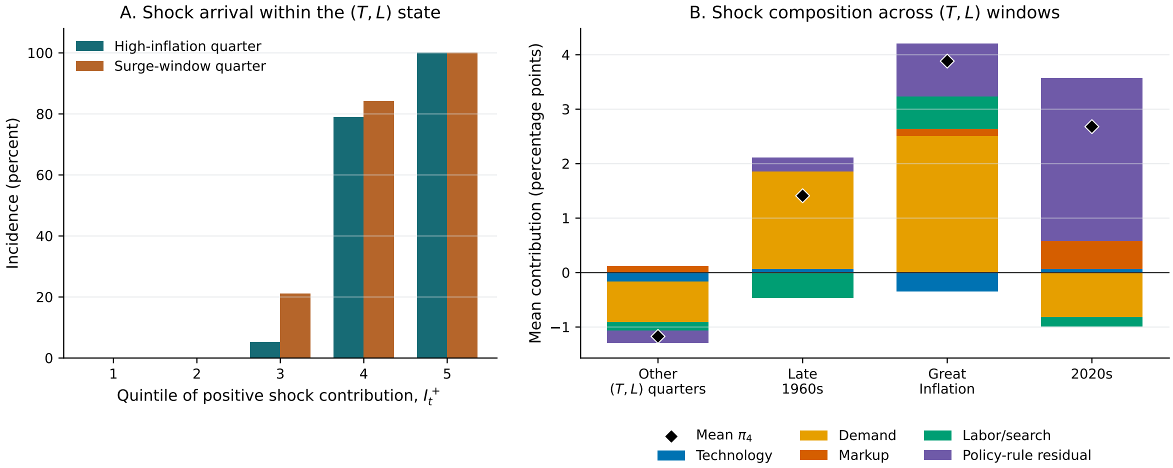


Figure 8: What turns an inflation-prone state into a surge.

Notes: Panel A restricts the sample to (T, L) quarters and sorts them into quintiles of I_t^+, the gross positive contribution defined in (58). Bars report the incidence of upper-quartile inflation and surge-window observations. Panel B compares the mean model-implied contributions in the three surge windows with all other (T, L) quarters. The black diamond is mean four-quarter inflation. Initial-condition and switching terms are omitted from the bars, so the diamond need not equal their sum. The policy-rule residual is the innovation in the estimated policy equation rather than a high-frequency monetary-policy shock.

Panel A of Figure 8 makes the sufficiency distinction visible while holding the joint state fixed. Neither of the lowest two impulse quintiles contains a high-inflation or surge-window quarter. Incidence rises sharply in the fourth quintile and reaches 100.0 percent in the fifth. The vulnerable state is therefore fertile ground, not an automatic outcome.

Panel B shows why the three historical episodes are not repetitions of one shock. Demand contributes 1.8 percentage points in the Late 1960s and 2.5 points during the Great Inflation. In the 2020s, the policy-rule residual contributes 3.0 points and markup innovations another 0.5, while the demand contribution is negative. Outside surge windows, the same (T, L) environment is met by contractionary demand, technology, labor/search, and policy contributions, and average inflation is negative. The common feature across surges is thus not a single disturbance. It is the arrival of unusually strong inflationary pressure in an environment that transmits it forcefully.[13]

The episodes line up with this mechanism in different ways. The Late-1960s and post-2020 surges are high-pressure, high-pass-through episodes. The estimated labor state assigns substantial probability to tightness and the pricing state is overwhelmingly flat. They are close to the (T, L) environment identified above as the most vulnerable onset state. The Great Inflation is different. Its dominant impulse is large and persistent, while labor-market tightness is less sharp. That episode shows why the paper treats tightness and pricing pass-through as complements rather than substitutes. Pass-through determines how pressure enters prices, while the size and persistence of shocks determine how much pressure there is to transmit.[14]

[13]Appendix H reports the episode-level table linking dominant shock blocks to median state probabilities.

[14]Appendix K reports the full-sample historical decomposition, while the main text focuses on the macroeconomic environments in which shocks are transmitted.

7 Conclusions

This paper asks when inflation takes hold. The estimated mechanism is layered. Tightness creates upstream pressure in the wage and marginal-cost block, while the pricing environment determines how much of that pressure becomes visible inflation and how long the response persists.

The integrated Phillips curve makes this interaction precise. The state-dependent wage mapping changes the pressure carried by an impulse, while the level and age profile of price adjustment place the economy on an impact–persistence frontier. The high-pass-through state converts more marginal cost into current inflation. The lower-intercept, age-sensitive state converts less on impact but gives inflation greater intrinsic continuation. Consequently, the state with the largest current response need not be the state with the longest-lived response.

The U.S. evidence supports this two-dimensional view. Labor-state variation delivers the largest improvement over a constant-state benchmark, and pricing variation adds further model evidence once labor pressure is allowed to move. More importantly, duration dependence improves both pre-2020 model evidence and the conditional joint one-step density over 2020–22, even after all aggregate-shock persistence parameters are re-estimated. The pricing-age structure therefore contains information beyond persistent disturbances.

Historically, postwar surges concentrate on the high-impact side of the pricing frontier, especially when it coincides with strong labor-market pressure. Yet the episodes differ in the shocks that reach that environment. This distinction reconciles a steep nonlinear Phillips curve with the fact that tightness alone neither predicts nor explains every surge. Inflation takes hold when the economy combines sufficient pressure with a pricing environment that transmits it.

References

Baele, L., Bekaert, G., Cho, S., Inghelbrecht, K., and Moreno, A. (2015). Macroeconomic regimes. *Journal of Monetary Economics*, 70:51–71.

Baxter, M. and King, R. G. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics*, 81(4):575–593.

Benigno, P. and Eggertsson, G. B. (2024). The slanted-L Phillips curve. *AEA Papers and Proceedings*, 114:84–98.

Benigno, P. and Eggertsson, G. B. (2025). It's baaack: The surge in inflation in the 2020s and the return of the non-linear Phillips curve. Revised version, October 2025.

Benigno, P. and Ricci, L. A. (2011). The inflation–output trade-off with downward wage rigidities. *American Economic Review*, 101(4):1436–1466.

Beqiraj, E., Ciccarone, G., and Di Bartolomeo, G. (2026). The U.S. economic dynamics and inflation persistence: A regime-switching perspective. *Journal of the European Economic Association*. Forthcoming.

Bianchi, F. (2013). Regime switches, agents' beliefs, and post-world war II U.S. macroeconomic dynamics. *Review of Economic Studies*, 80(2):463–490.

Bils, M. and Klenow, P. J. (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy*, 112(5):947–985.

Bjørnland, H. C., Larsen, V. H., and Maih, J. (2018). Oil and macroeconomic (in)stability. *American Economic Journal: Macroeconomics*, 10(4):128–151.

Cerrato, A. and Gitti, G. (2023). Inflation since COVID: Demand or supply? Working Paper, UC Berkeley.

Christiano, L. J., Eichenbaum, M. S., and Trabandt, M. (2016). Unemployment and business cycles. *Econometrica*, 84(4):1523–1569.

Di Bartolomeo, G. and Di Pietro, M. (2017). Intrinsic persistence of wage inflation in New Keynesian models of the business cycles. *Journal of Money, Credit and Banking*, 49(6):1161–1195.

Dixon, H. and Le Bihan, H. (2012). Generalised Taylor and Calvo pricing. *Economic Journal*.

Eichenbaum, M., Jaimovich, N., and Rebelo, S. (2011). Reference prices, costs, and nominal rigidities. *American Economic Review*, 101(1):234–262.

Fuhrer, J. C. and Moore, G. R. (1995). Inflation persistence. *Quarterly Journal of Economics*, 109(1):127–159.

Galí, J. and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44(2):195–222.

Gertler, M., Huckfeldt, C., and Trigari, A. (2020). Unemployment fluctuations, match quality, and the wage cyclicality of new hires. *Review of Economic Studies*, 87(4):1876–1914.

Gertler, M. and Trigari, A. (2009). Unemployment fluctuations with staggered Nash wage bargaining. *Journal of Political Economy*, 117(1):38–86.

Gitti, G. (2023). Nonlinearities in the regional Phillips curve with labor market tightness. Working Paper, Brown University.

Graves, S., Huckfeldt, C. K., and Swanson, E. T. (2026). The labor demand and labor supply channels of monetary policy. Working Paper 31770, National Bureau of Economic Research. Revised February 2026.

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

Haefke, C., Sonntag, M., and van Rens, T. (2013). Wage rigidity and job creation. *Journal of Monetary Economics*, 60(8):887–899.

Hall, R. E. (2005). Employment fluctuations with equilibrium wage stickiness. *American Economic Review*, 95(1):50–65.

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

Herbst, E. and Schorfheide, F. (2014). Sequential Monte Carlo sampling for DSGE models. *Journal of Applied Econometrics*, 29(7):1073–1098.

Klenow, P. J. and Kryvtsov, O. (2008). State-dependent or time-dependent pricing: Does it matter for recent U.S. inflation? *Quarterly Journal of Economics*, 123(3):863–904.

Klenow, P. J. and Malin, B. A. (2010). Microeconomic evidence on price-setting. In Friedman, B. M. and Woodford, M., editors, *Handbook of Monetary Economics*, volume 3, pages 231–284. Elsevier.

Maih, J. (2015). Efficient perturbation methods for solving regime-switching DSGE models. Working Paper 01/2015, Norges Bank.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics*, 34(4):574–589.

Nakamura, E. and Steinsson, J. (2008). Five facts about prices: A reevaluation of menu cost models. *American Economic Review*, 98(4):1415–1464.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861–1957. *Economica*, 25(100):283–299.

Sheedy, K. D. (2010). Intrinsic inflation persistence. *Journal of Monetary Economics*, 57(8):1049–1061.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, 95(1):25–49.

Smets, F. and Wouters, R. (2007). Shocks and frictions in U.S. business cycles: A Bayesian DSGE approach. *American Economic Review*, 97(3):586–606.

Swanson, E. T. (2021). Measuring the effects of Federal Reserve forward guidance and asset purchases on financial markets. *Journal of Monetary Economics*, 118:32–53.

Swanson, E. T. and Williams, J. C. (2014). Measuring the effect of the zero lower bound on medium- and longer-term interest rates. *American Economic Review*, 104(10):3154–3185.

Taylor, J. B. (1999). A historical analysis of monetary policy rules. In Taylor, J. B., editor, *Monetary Policy Rules*, pages 319–341. University of Chicago Press.

On-line Appendices

A Analytical proofs

Proof of Proposition 1. Using (33) and (35),

$$\frac{\kappa^{\text{tight}}}{\kappa} = \frac{k_p\alpha_N\eta}{k_p\alpha_N(1-\lambda)\eta/(1+k_p\alpha_N\lambda)} = \frac{1+k_p\alpha_N\lambda}{1-\lambda}. \tag{59}$$

Because $0 \le \lambda < 1$ and $k_p\alpha_N\lambda \ge 0$, the ratio is weakly above one and is strictly above one whenever $\lambda > 0$. Direct differentiation gives

$$\frac{\partial}{\partial k_p}\left(\frac{\kappa^{\text{tight}}}{\kappa}\right) = \frac{\alpha_N\lambda}{1-\lambda} > 0, \tag{60}$$

and

$$\frac{\partial}{\partial \lambda}\left(\frac{\kappa^{\text{tight}}}{\kappa}\right) = \frac{1+k_p\alpha_N}{(1-\lambda)^2} > 0. \tag{61}$$

□

Proof of Proposition 2. Let $a \equiv \alpha_N\lambda$, $c \equiv \alpha_N(1-\lambda)\eta$, and write k for k_p. Then

$$\kappa^{\text{tight}} = k\alpha_N\eta, \qquad \kappa = \frac{kc}{1+ka}, \qquad \kappa_w = \frac{ka}{1+ka}. \tag{62}$$

Differentiating with respect to k yields

$$\frac{\partial\kappa^{\text{tight}}}{\partial k} = \alpha_N\eta, \qquad \frac{\partial\kappa}{\partial k} = \frac{c}{(1+ka)^2}, \qquad \frac{\partial\kappa_w}{\partial k} = \frac{a}{(1+ka)^2}, \tag{63}$$

which are the expressions in the proposition. □

Proof of Proposition 3. Write $B = 1-\beta(1-\alpha)$, $D = 1-\alpha-\varphi B$, and $N = (\alpha+\varphi)(B+\beta^2\varphi)$, so that $\psi_p = \varphi/D$ and $k_p = \Theta_p N/D$. On the admissible region,

$$\frac{\partial\psi_p}{\partial\varphi} = \frac{1-\alpha}{D^2} > 0, \qquad \frac{\partial\psi_p}{\partial\alpha} = \frac{\varphi(1+\beta\varphi)}{D^2} > 0. \tag{64}$$

Moreover,

$$\frac{\partial k_p}{\partial\varphi} = \frac{\Theta_p}{D^2}\left[\Big(B+\beta^2\alpha+2\beta^2\varphi\Big)D + BN\right] > 0, \tag{65}$$

$$\frac{\partial k_p}{\partial\alpha} = \frac{\Theta_p}{D^2}\left[\Big(B+\beta^2\varphi+\beta(\alpha+\varphi)\Big)D + (1+\beta\varphi)N\right] > 0. \tag{66}$$

The terms are positive under the stated conditions, except that $\partial\psi_p/\partial\alpha = 0$ at the boundary $\varphi = 0$. If $\varphi_H \ge \varphi_L$ and $\alpha_H \ge \alpha_L$, monotonicity would imply $k_H \ge k_L$ and $\psi_H \ge \psi_L$, contradicting (41). Hence $\alpha_L > \alpha_H$ is necessary. Finally, because $D_L, D_H > 0$, cross-multiplying $\psi_H > \psi_L$ and $k_L > k_H$ gives (42) and (43), respectively. The two inequalities are also sufficient, completing the proof. □

Proof of Proposition 4. In the tight state, the coefficients on $\hat{\theta}_t$ and z_t in B_t^T are $k_p \alpha_N \eta$ and k_p, respectively. Their derivatives with respect to k_p are therefore $\alpha_N \eta$ and one. In the slack state, let $a \equiv \alpha_N \lambda$. The coefficients on $\hat{\theta}_t$ and z_t are

$$\frac{k_p \alpha_N (1-\lambda)\eta}{1+k_p a} \qquad \text{and} \qquad \frac{k_p}{1+k_p a}. \tag{67}$$

The derivative of $k_p/(1+k_p a)$ with respect to k_p is $(1+k_p a)^{-2}$. Multiplication by $\alpha_N(1-\lambda)\eta$ gives the first slack-state cross-partial. The second follows directly. Positivity holds for $k_p, \alpha_N, \eta > 0$ and $0 \le \lambda < 1$. □

Proof of Proposition 5. Solving $B_t^T(j) = \bar{B}$ and $B_t^S(j) = \bar{B}$ for z_t gives the two threshold expressions in the proposition. Because $k_L > k_H$,

$$z^*_{T,H} - z^*_{T,L} = \bar{B}\left(\frac{1}{k_H} - \frac{1}{k_L}\right) > 0. \tag{68}$$

For either pricing state j,

$$z^*_{S,j} - z^*_{T,j} = \alpha_N \lambda \left(\bar{B} + \eta \hat{\theta}_t - \hat{w}^{\text{ex}}_{t-1}\right). \tag{69}$$

Condition (50) makes this difference strictly positive. It follows that $z^*_{T,L}$ is below $z^*_{T,H}$, $z^*_{S,L}$, and $z^*_{S,H}$. The minimum of the latter three thresholds therefore exceeds $z^*_{T,L}$, so the interval stated in the proposition is non-empty. □

B Data and measurement details

The main text summarizes the economic role of the observed variables in the estimation. Appendix Table B.1 reports the exact series and transformations. The series are drawn from the FRED-QD database documented by McCracken and Ng (2016). The estimation sample starts in 1965Q1 and ends in 2022Q4 after endpoint trimming induced by the Baxter–King filter (Baxter and King, 1999). The variables enter in raw quarterly log or rate units, rather than percentage units, and are demeaned before estimation.

The bilateral Baxter–King transformation defines the empirical object as an ex post historical state decomposition. The 2020–22 density analysis therefore evaluates conditional fit under the same information transformation used in estimation.

Table B.1: Estimation observables.

Observable	FRED-QD series	Transformation
π_t	GDPCTPI	$\Delta \log P_t$, demeaned
i_t	FEDFUNDS	$\log(1 + i_t^{ann}/400)$, demeaned
Y_t	GDPC1	BK cycle of $\log Y_t$
u_t	UNRATE	$\log(u_t/100)$, demeaned
w_t	COMPRNFB	BK cycle of $\log w_t$
b_t	HWIURATIOx	BK cycle of $\log(v_t/u_t)$

Notes: The estimation sample is 1965Q1–2022Q4, after endpoint trimming from the Baxter–King filter. All observables are in raw quarterly log or rate units and are demeaned before estimation. The stationary component of the vacancy–unemployment ratio is observed and disciplines the transition of the labor-pressure state.

C Posterior simulation and convergence

At each posterior proposal, the four-regime equilibrium is solved, checked for mean-square stability, and filtered over the complete estimation sample. The solution follows the perturbation methods for regime-switching equilibrium models in Maih (2015). Posterior simulation and convergence assessment follow standard Bayesian DSGE practice (Herbst and Schorfheide, 2014).

Posterior inference uses four continuation chains initialized at the endpoints of the first-stage chains. Each chain contains 55,000 iterations; the first 5,000 are discarded and the remaining 50,000 are retained without thinning, yielding 200,000 posterior draws. The random-walk proposal scale is 0.20, and chain-specific acceptance rates range from 0.231 to 0.247.

Appendix Table C.1 reports posterior modes, means, and equal-tailed 90 percent credible intervals. Appendix Table C.2 reports rank-normalized split-chain $\widehat{R}$ statistics and effective sample sizes. The largest split $\widehat{R}$ is 1.013, for ρ_ξ; the minimum bulk and tail effective sample sizes are 651 and 492, respectively.

Table C.1: Posterior distributions of estimated parameters.

Parameter	Description	Mode	Mean	5th pct.	95th pct.
Panel A: Common block					
σ	Intertemporal substitution curvature	4.751	4.590	3.515	5.447
ϕ_π	Taylor-rule response to inflation	2.015	2.062	1.818	2.321
ϕ_y	Taylor-rule response to activity	0.0045	-0.0006	-0.0558	0.0551
ρ_i	Interest-rate smoothing	0.780	0.778	0.740	0.810
ρ_A	Technology-shock persistence	0.991	0.989	0.979	0.996
ρ_μ	Markup-shock persistence	0.798	0.777	0.677	0.867
ρ_ξ	Demand-shock persistence	0.945	0.947	0.920	0.970
ρ_m	Labor/search-shock persistence	0.970	0.964	0.947	0.979
σ_A	Technology-shock standard deviation	0.0051	0.0051	0.0047	0.0055
σ_ξ	Demand-shock standard deviation	0.0306	0.0344	0.0233	0.0515
σ_μ	Markup-shock standard deviation	0.0064	0.0067	0.0047	0.0094
σ_m	Labor/search-shock standard deviation	0.0204	0.0207	0.0190	0.0226
σ_i	Monetary-policy shock standard deviation	0.0027	0.0028	0.0025	0.0031
Panel B: Pricing block					
$\alpha_{p,L}$	Price-reset hazard level in state L	0.451	0.446	0.342	0.547
$\varphi_{p,L}$	Duration dependence in state L	0.040	0.053	-0.038	0.145
$\alpha_{p,H}$	Price-reset hazard level in state H	0.047	0.085	0.013	0.188
$\varphi_{p,H}$	Duration dependence in state H	0.239	0.204	0.093	0.323
p^p_{LH}	Transition probability from L to H	0.034	0.060	0.019	0.122
p^p_{HL}	Transition probability from H to L	0.053	0.079	0.032	0.143

Notes: The table reports the posterior mode, posterior mean, and equal-tailed 90 percent credible interval. The moments and quantiles are computed from 200,000 retained draws pooled across four chains. Pricing state L denotes flat, high-pass-through pricing; pricing state H denotes the persistent pricing state. The prior specification is reported in Table 1.

Table C.2: Posterior simulation and convergence diagnostics.

Parameter	Description	Split $\widehat{R}$	Bulk ESS	Tail ESS
Panel A: Common block				
σ	Intertemporal substitution curvature	1.004	1,164	2,352
ϕ_π	Taylor-rule response to inflation	1.002	2,587	5,045
ϕ_y	Taylor-rule response to activity	1.001	1,834	4,267
ρ_i	Interest-rate smoothing	1.002	1,460	2,892
ρ_A	Technology-shock persistence	1.001	1,633	2,201
ρ_μ	Markup-shock persistence	1.003	1,407	2,537
ρ_ξ	Demand-shock persistence	1.013	651	561
ρ_m	Labor/search-shock persistence	1.002	1,863	2,563
σ_A	Technology-shock standard deviation	1.004	2,235	3,806
σ_ξ	Demand-shock standard deviation	1.009	683	492
σ_μ	Markup-shock standard deviation	1.006	1,055	1,284
σ_m	Labor/search-shock standard deviation	1.002	2,011	3,966
σ_i	Monetary-policy shock standard deviation	1.002	1,883	3,365
Panel B: Pricing block				
$\alpha_{p,L}$	Price-reset hazard level in state L	1.002	1,353	2,523
$\varphi_{p,L}$	Duration dependence in state L	1.002	1,825	4,070
$\alpha_{p,H}$	Price-reset hazard level in state H	1.002	2,453	3,098
$\varphi_{p,H}$	Duration dependence in state H	1.007	1,199	1,637
p^p_{LH}	Transition probability from L to H	1.004	2,017	2,267
p^p_{HL}	Transition probability from H to L	1.002	1,487	1,903

Notes: Split $\widehat{R}$ is the rank-normalized split-chain statistic. Bulk and tail ESS denote effective sample sizes for the center and tails of each marginal posterior. Posterior simulation uses four chains of 55,000 iterations, discards the first 5,000 iterations of each chain, and retains 50,000 draws per chain without thinning. The proposal scale is 0.20; chain-specific acceptance rates range from 0.231–0.247.

The posterior distribution provides a useful distinction between the two parts of the pricing frontier. The impact ranking is sharply concentrated, whereas the intrinsic-continuation ranking carries greater uncertainty. Across the retained posterior, $\Pr(k_L > k_H) = 1.000$ and the probability of the joint ordering $k_L > k_H$ and $\psi_H > \psi_L$ is 0.846. The corresponding probability that the pricing primitives rotate in the direction $\alpha_L > \alpha_H$ and $\varphi_H \geq \varphi_L$ is 1.000. These probabilities use 200000 valid draws from the posterior simulation. Figure C.1 shows the joint posterior geometry. Panel A maps each draw into the two inequalities in (41). Panel B maps the same draws into the rotation of the underlying adjustment profile.

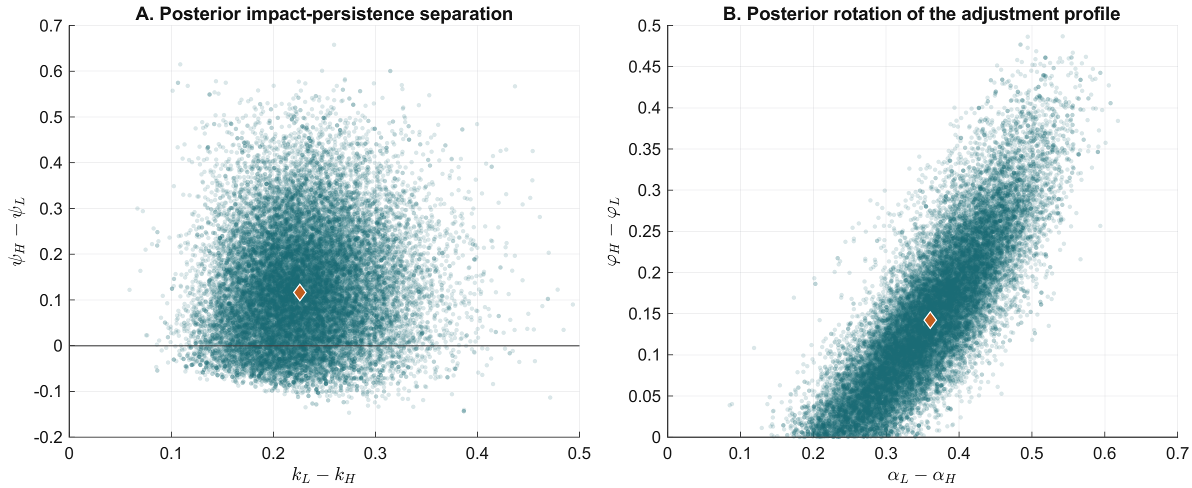


Figure C.1: Posterior evidence on the impact–persistence frontier.

Notes: Panel A plots the posterior differences $k_L - k_H$ and $\psi_H - \psi_L$. Panel B plots the primitive rotation $\alpha_L - \alpha_H$ and $\varphi_H - \varphi_L$. Zero lines delimit the impact–persistence and profile-rotation regions characterized in Proposition 3. Diamonds denote posterior medians. For legibility, the plotted cloud uses 25,000 draws selected evenly from the 200,000 draws used to compute the reported probabilities.

D State transitions and dynamic estimands

Historical filtering and dynamic transmission use the same estimated model but answer different economic questions. Historical filtering applies (23)–(24) and the joint likelihood to infer the realized state probabilities. The dynamic experiments instead condition on an initial joint state and compare two paths that differ only in the target innovation.

For simulation draw r, let $\mathcal{E}^{(r)}$ collect the future non-target innovations and let $\mathcal{S}^{(r)}$ denote the future joint-regime sequence. The state-conditioned response to an innovation of size a in shock j is

$$\mathrm{GIRF}_h^j(s_0, a) = \frac{1}{R}\sum_{r=1}^{R}\Big[y_{t+h}\Big(ae_j, \mathcal{E}^{(r)}, \mathcal{S}^{(r)} \mid s_0\Big) - y_{t+h}\Big(0, \mathcal{E}^{(r)}, \mathcal{S}^{(r)} \mid s_0\Big)\Big]. \tag{70}$$

Each shocked-baseline pair receives the same $\mathcal{E}^{(r)}$ and $\mathcal{S}^{(r)}$. The policy functions used on both paths come from the switching-model solution, so agents' decisions still incorporate the possibility of future state changes. What is excluded from the reported difference is an additional effect generated by assigning a different realized regime sequence to the shocked path.

Table D.1: Regime treatment in the quantitative analyses.

Analysis	Regime treatment	Reported object
Historical filtering	Observation-driven labor hazards, recursive state prediction, and joint likelihood update. Pricing transitions are estimated	Filtered and smoothed state probabilities
Surge GIRF	Initial joint state specified. Future regimes and non-target innovations are common to shocked and baseline paths	Conditional peak inflation and threshold multiplier
Pressure ladder	Same state-conditioned GIRF evaluated at impact. The wage and Phillips-curve forcing are read from the structural equations	Tight-to-slack impact ratios
Fixed-state diagnostic	The stated labor or pricing coefficient state is held unchanged at every horizon in both paths	Coefficient-channel accounting

Notes: The surge GIRFs and pressure ladder condition on the initial state. They do not hold that state fixed at later horizons. The label fixed-state is reserved for the final row.

Definition 1 (Threshold diagnostic)**.** *Fix an initial joint state $s_0 = (s_0^w, s_0^p)$, a shock type j, a horizon H, and an inflation boundary $\bar{\pi}$. Let $\pi_h^{(4),j}(s_0, a)$ denote the four-quarter inflation component of* (70)*. The threshold impulse is*

$$a^{\star}(s_0, j; \bar{\pi}, H) \equiv \inf \left\{ a > 0 : \max_{0 \leq h \leq H} \pi_h^{(4),j}(s_0, a) \geq \bar{\pi} \right\}. \tag{71}$$

The reported multiplier is $a^{\star}(s_0, j; \bar{\pi}, H)/a^{\star}((T, L), j; \bar{\pi}, H)$. In the first-order solution it is a normalized inverse peak-response statistic.

Definition 2 (Fixed-state transmission diagnostic)**.** *Fix an initial state s_0 and an innovation ε_0^j. A fixed-state diagnostic holds the stated labor or pricing coefficient state at its initial value at every horizon in both shocked and baseline paths. If only one state dimension is fixed, the other follows the transition treatment stated for that calculation. Integrating over future states is not described as fixed-state. These responses isolate coefficient channels and are distinct from the state-conditioned GIRFs in* (70)*.*

E Dynamic paths behind the surge multipliers

Figure E.1 reports the complete inflation paths underlying Figure 6. For each shock, the common impulse is scaled so that peak four-quarter inflation in (T, L) reaches the empirical surge boundary. The other three paths use exactly the same impulse. The figure makes clear that the threshold ranking is not generated by an isolated impact observation. The ordering is present over the onset window and then narrows as the four economies converge. The convergence also explains why the pressure-transmission ladder is most informative early in the response, when wage selection and cost formation differ most across labor states.

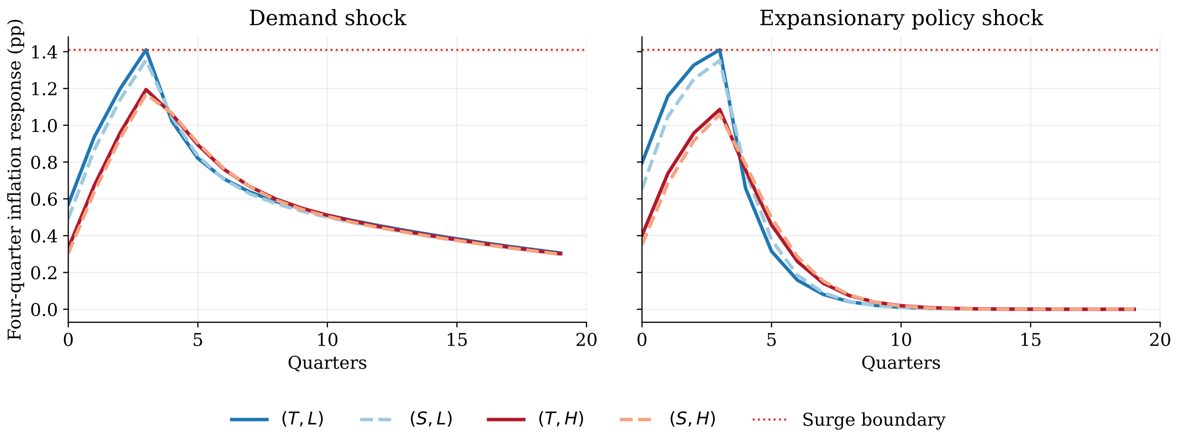


Figure E.1: Dynamic inflation paths behind the surge multipliers.

Notes: For each shock, the impulse is scaled so that the (T, L) peak four-quarter inflation response reaches the empirical surge boundary. The remaining joint states are evaluated under the same impulse. The economy is initialized in the indicated state. Shocked and baseline paths are paired on the same future labor-pricing regime sequence and the same non-target innovations. The responses therefore isolate transmission conditional on the initial state and exclude an additional reclassification margin. Parameters are evaluated at their posterior modes.

F Effective Search and the Allocation of Monetary Adjustment

A monetary contraction reduces vacancy demand, but it also changes the pool of workers available to fill jobs. Graves et al. (2026) document that workers respond to a contraction by reducing quits to nonemployment and increasing search from nonparticipation. These worker flows materially cushion the employment response. The labor block of our model contains the same upstream object through the stock of effective job seekers. This appendix asks how variation in that object is transmitted to employment and inflation across the four estimated labor-pricing environments.

Let O_t denote nonparticipation and define the broader search pool as

$$\widetilde{U}_t^{AS} = U_t + \alpha O_t, \qquad \alpha = 0.6. \tag{72}$$

The corresponding labor-pressure signal is

$$q_t^{AS} = \log\left(\frac{V_t}{U_t}\right) + \log\left(\frac{U_t}{\widetilde{U}_t^{AS}}\right) = \log\left(\frac{V_t}{\widetilde{U}_t^{AS}}\right). \tag{73}$$

We construct this signal from the same data vintage, apply the same filter, and retain the same 232-quarter estimation sample as in the benchmark. Its correlation with the benchmark labor-pressure signal is 0.98. The two measures agree on the sign of labor pressure in 96.1 percent of quarters and on membership in the upper-pressure quartile in 94.8 percent. Broadening the search pool therefore preserves the timing of the estimated labor state while compressing the amplitude of labor pressure. The standard deviation of q_t^{AS} is 59.3 percent of the benchmark signal. This moment supplies the external calibration below.

We scale the response of effective job seekers by χ_S. The benchmark normalization has $\chi_S = 1$, the broader-search calibration has $\chi_S = 0.593$, and $\chi_S = 0$ holds effective search fixed at its steady-state level. All remaining parameters and transition laws stay at their benchmark values. In every initial joint state, the monetary contraction is normalized to raise the annualized policy rate by 25 basis points on impact. Figure F.1 reports the peak employment loss and peak four-quarter disinflation over the subsequent twenty quarters.

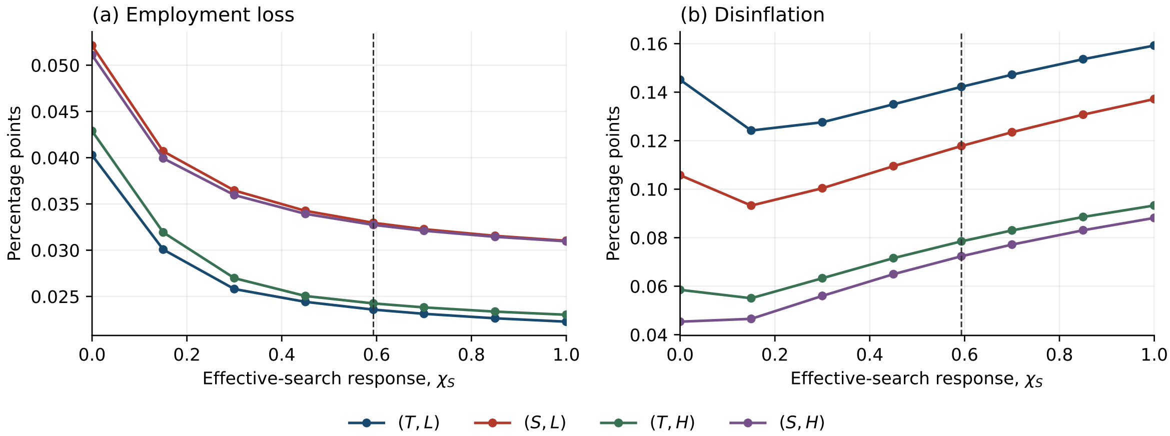


Figure F.1: Effective search and the allocation of monetary adjustment.

Notes: Panel (a) reports the peak decline in employment and panel (b) reports peak four-quarter disinflation after a monetary contraction that raises the annualized policy rate by 25 basis points on impact. The coefficient χ_S scales the response of effective job seekers. The benchmark is $\chi_S = 1$, the vertical dashed line is the broader-search calibration $\chi_S = 0.593$, and $\chi_S = 0$ holds effective search fixed. All remaining parameters and transition laws are held at their benchmark values. The four lines correspond to (T, L), (S, L), (T, H), and (S, H).

Moving from the benchmark to the broader-search calibration raises the peak employment loss by 5.3 to 6.2 percent across the four states and raises the cumulative output loss by 25.4 to 30.2 percent. Peak disinflation falls by 10.7 to 17.9 percent. These effects vary smoothly around the externally disciplined calibration. A targeted posterior audit based on 25 draws selected evenly across the five chains preserves the sign of all three effects in every state.

The analysis adds an inflation dimension to the worker-search response. Over the empirically disciplined interval between the broader-search calibration and the benchmark, a stronger response keeps more workers available to fill jobs, reduces the employment decline, and moves a larger share of monetary adjustment into disinflation. When that response is muted, adjustment shifts toward employment and output. The prevailing price-adjustment environment governs how this upstream labor-market response appears in inflation, so the same change in effective search produces different inflation paths across the four joint states. Endogenous search and state-dependent price adjustment are two successive margins of monetary transmission.

G Additional state-location and vulnerability diagnostics

This appendix reports state-location diagnostics that support Section 6.1 but are not separate identifying moments in the main text.

Table G.1: Inflation surges and estimated joint states.

Episode	Observed tightness evidence	Estimated joint state	State probabilities
Late-1960s surge	Near the threshold: mean $\theta = 1.03$, peak $\theta = 1.51$	Onset and window: (T, L)	$\widetilde{P}_T = 0.91$ $\widetilde{P}_L = 0.97$
Great Inflation	Below the threshold on average: mean $\theta = 0.60$, peak $\theta = 0.99$	Window: (T, L)	$\widetilde{P}_T = 0.81$ $\widetilde{P}_L = 0.98$
2020s surge	Clearly tight: mean $\theta = 1.66$, peak $\theta = 1.91$	Onset borderline, window: (T, L)	$\widetilde{P}_T = 0.99$ $\widetilde{P}_L = 1.00$

Notes: Observed tightness statistics report the episode mean and peak labor-market tightness used to classify postwar U.S. inflation surges in the related nonlinear Phillips-curve evidence. $\widetilde{P}_T$ and $\widetilde{P}_L$ are median smoothed probabilities within the surge window for the high-pressure labor state and the flat, high-pass-through pricing state. The median describes the typical quarter of the episode and is less sensitive than the mean to onset and unwinding quarters. Observed θ reports the raw vacancy–unemployment level, whereas $\widetilde{P}_T$ summarizes the labor-pressure state identified from its stationary component within the estimated system.

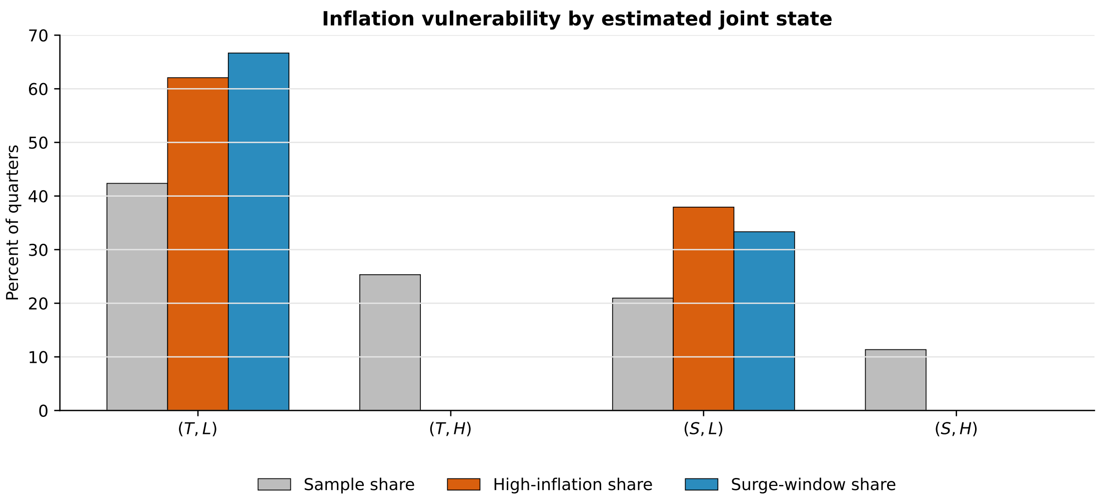


Figure G.1: Inflation vulnerability by estimated joint state.

Notes: Bars compare each state's sample share with its share of high-inflation quarters and inflation-surge-window quarters. If a state were unrelated to inflation vulnerability, the bars would move together. The flat pricing states are over-represented among high-inflation observations. (T, L) carries the largest surge-window mass, whereas the persistent pricing states retain substantial sample frequency but no upper-tail inflation observations. The figure measures the concentration of outcomes across states, not the probability that a state mechanically produces a surge.

H Historical impulse-state table

This appendix reports the episode-level table underlying the compact shock accounting in Section 6.3.

Table H.1: Historical impulses and transmission states.

Episode	Dominant impulse	$\widetilde{P}_T$	$\widetilde{P}_L$
Late-1960s surge	demand block (1.79 pp)	0.91	0.97
Great Inflation	demand block (2.51 pp)	0.81	0.98
2020s surge	policy-rule residual (2.98 pp)	0.99	1.00

Notes: The dominant impulse is the shock block with the largest absolute episode-average contribution to four-quarter GDP-deflator inflation deviations from the estimation-sample mean. $\widetilde{P}_T$ and $\widetilde{P}_L$ are median smoothed probabilities within the surge window for the high-pressure labor state and the flat, high-pass-through pricing state. The policy-rule residual is the contribution of the policy-rule innovation in the estimated state-space system, not a high-frequency monetary-policy shock.

I Auxiliary state-conditional transmission and interaction accounting

The following diagnostics support the semi-structural interpretation in Section 6. They are model-accounting exercises rather than additional sources of identification.

I.1 Estimated mechanism

This appendix collects the mechanism tables used to interpret the estimated states. They are placed outside the main text so that Section 6 keeps the sequence historical location, controlled threshold experiment, and historical shock accounting.

Table I.1: Estimated state frequencies and Phillips-curve margins.

Panel A: Smoothed joint-state probabilities

(s_t^w, s_t^p)	Expected share	Dominant share	Max spell	Mean $P(T)$	Mean $P(H)$
(S, L)	23.6	20.7	11	0.119	0.061
(S, H)	14.6	11.6	12	0.190	0.903
(T, L)	38.1	41.8	23	0.867	0.095
(T, H)	23.7	25.9	31	0.806	0.872

Panel B: Posterior-mode Phillips-curve margins

(s_t^w, s_t^p)	α_p	φ_p	Local slope	ψ_p
(S, L)	0.451	0.040	0.0442	0.0753
(S, H)	0.047	0.239	0.0093	0.2544
(T, L)	0.451	0.040	0.0993	0.0753
(T, H)	0.047	0.239	0.0191	0.2544

Notes: Panel A is computed from the smoothed probabilities of the selected four-state estimated model over 1965Q1–2022Q4. The state column reports (s_t^w, s_t^p), where T and S denote high- and low-pressure labor states and L and H denote the flat, high-pass-through and persistent pricing states. Expected share is the average joint probability; dominant share and max spell use the most likely joint state in each quarter. Panel B reports the posterior-mode pricing primitives and the implied local Phillips-curve margins used in the state-by-state mechanism analysis. The high- and low-pressure states map into the tight and slack wage-selection branches; their local slopes are κ^{tight} and κ, respectively.

Table I.2: Posterior-mode calibration and implied Phillips-curve margins.

Panel A: Baseline posterior-mode and fixed parameters

Parameter	Description	Value
β	Discount factor	0.990
σ	Intertemporal substitution curvature	4.751
ϵ	Steady-state goods-market elasticity	6.000
Θ_p	Reset-price pass-through / real rigidity calibration	0.600
α_N	Labor share in production	0.900
η	Tightness elasticity of the flexible wage	0.400
λ	Wage staggering parameter	0.500
δ	Incumbent-wage indexation parameter	0.500
φ_π	Taylor-rule response to inflation	2.015
φ_y	Taylor-rule response to activity	0.005
ρ_i	Interest-rate smoothing	0.780
$\alpha_{p,L}$	Price-reset hazard level in state L	0.451
$\varphi_{p,L}$	Duration dependence in state L	0.040
$\alpha_{p,H}$	Price-reset hazard level in state H	0.047
$\varphi_{p,H}$	Duration dependence in state H	0.239

Panel B: Implied pricing and Phillips-curve coefficients

s_t^p	Description	$k_p(s^p)$	$\psi_p(s^p)$	$\kappa^{\text{tight}}(s^p)$	$\kappa(s^p)$
L	Flat, high-pass-through pricing state	0.2757	0.0753	0.09926	0.04415
H	Persistent pricing state	0.0530	0.2544	0.01907	0.00931

Notes: Panel A reports selected posterior-mode and fixed parameters used in the state-by-state mechanism analysis. Panel B reports the composite margins that appear in the analytical Phillips curve.

I.2 Shock transmission across inflation states

Table I.3 reads the historical decomposition from the perspective of the state rather than the episode. It groups inflation movements by the estimated state in which they occur and therefore separates the shock mix from the labor-pricing environment through which shocks are transmitted.

Table I.3: Shock accounting by estimated Phillips-curve state.

(s_t^w, s_t^p)	Obs. share	Mean π_4	Technology	Demand	Markup	Labor/search	Monetary
(T, L)	42.4	0.71	8.5	50.4	9.0	14.4	17.7
(T, H)	25.3	-0.99	7.5	27.9	9.6	15.3	39.8
(S, L)	21.0	0.43	6.5	38.5	10.2	20.2	24.5
(S, H)	11.4	-1.21	6.9	57.8	11.4	11.4	12.5

Notes: The state column reports (s_t^w, s_t^p), where T and S denote high- and low-pressure labor states and L and H denote the flat, high-pass-through and persistent pricing states. States are assigned by dominant smoothed probabilities: high pressure if $P(T) \geq 0.5$ and persistent pricing if $P(H) \geq 0.5$. Mean π_4 is the mean four-quarter GDP-deflator inflation deviation from the estimation-sample mean, in percentage points. Shock columns report each structural shock group's share of the mean absolute historical-decomposition contribution within the state; columns are normalized over technology, demand, markup, labor/search, and monetary contributions. Initial-condition and switching terms are excluded from the normalization and reported in the companion CSV.

I.3 The same shocks in different states

State-specific variance accounting asks a model-accounting question rather than a new identification question. Holding the model shock distribution fixed, how does the initial labor-pricing environment rotate the composition of inflation risk?

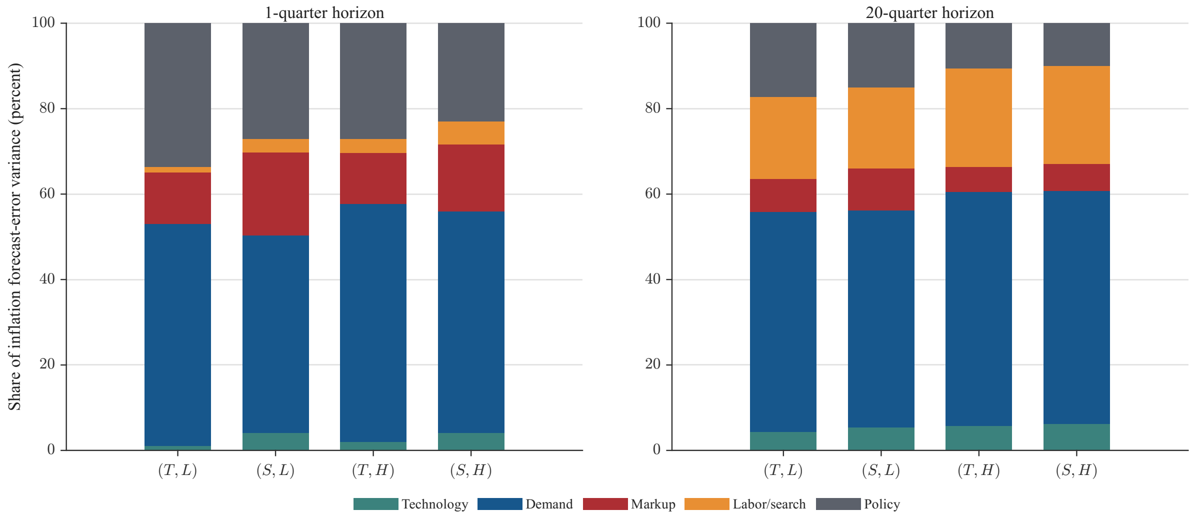


Figure I.1: The composition of inflation risk across joint states.

Notes: The bars decompose the forecast-error variance of quarterly inflation into five model shock blocks. The left panel conditions on the initial joint state and reports the one-quarter horizon. The right panel reports the twenty-quarter horizon. The model shock distribution is identical across states.

I.4 Interaction accounting

The interaction accounting decomposes the mechanism after the state-space model has already been selected. It shows how the estimated labor and pricing margins interact from wage selection to marginal-cost forcing and inflation.

Table I.4: Where labor–pricing complementarity enters transmission.

	New-hire wage premium		Forcing gap	Impact-inflation premium		Impact-inflation gap
Shock	L	H	L/H	L	H	L/H
Demand	59.4	54.3	3.4	15.7	9.5	2.7
Expansionary policy	54.4	52.4	3.7	21.6	14.6	2.8

Notes: Wage and impact-inflation premia are percentage differences between the high- and low-pressure labor states within the indicated pricing state. The forcing-gap ratio is the absolute high–low labor-cost forcing gap in state L divided by the corresponding gap in state H. The final column reports the absolute tight–slack impact-inflation gap in state L divided by the corresponding gap in state H. All moments are impact responses from state-conditioned generalized impulse responses evaluated at posterior-mode parameters. The inflation columns concern quarterly impact inflation and are therefore distinct from the peak four-quarter-inflation comparisons in the main text.

J Estimated-shock surge-ignition atlas

The main text focuses on the demand and expansionary policy shocks because those are the shock blocks that organize the post-1965 surge windows. This appendix asks the broader diagnostic question of whether the same threshold logic holds across the estimated shock space? For each estimated shock, the generalized impulse response is signed so that it is inflationary in (T, L) and then scaled so that the (T, L) peak four-quarter inflation response reaches the surge boundary. The other states are evaluated under that same shock normalization.

The purpose of the analysis is not to rank shocks by historical importance. It is to verify that the state labels have economic content. If (T, L) were merely a high-inflation label, it would mechanically dominate for every shock. Figure J.1 and Table J.1 show instead a more disciplined pattern. Demand, expansionary policy, and matching-efficiency shocks are most easily converted into a surge in (T, L). Markup and technology shocks can have a different ranking, because they enter marginal cost and the wage state through different margins. The main result is therefore not that (T, L) dominates every possible disturbance. It is that the historically relevant surge shocks become most dangerous when labor-market pressure meets high pricing pass-through.

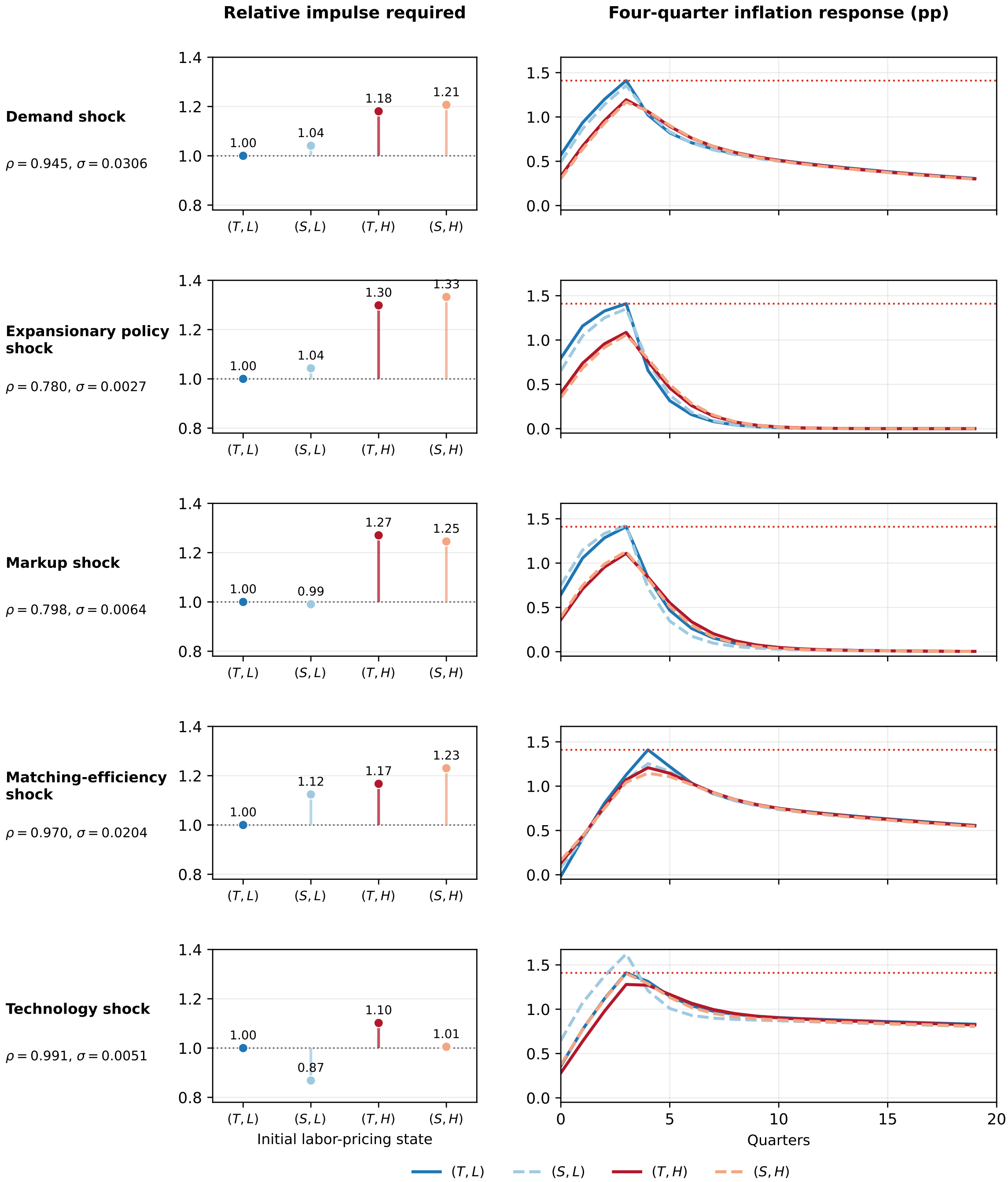


Figure J.1: Estimated-shock surge-ignition atlas.

Notes: Each row corresponds to one estimated shock. The left column reports the minimum impulse required in each initial state relative to the (T, L) requirement, with the dotted line normalized to one. The right column reports normalized four-quarter inflation paths and the red dotted line marks the surge boundary. Each shock is signed to be inflationary in (T, L) and scaled so that the (T, L) peak four-quarter inflation response reaches that boundary. The persistence and standard deviation beside each row label are posterior-mode estimates.

Table J.1: Estimated-shock surge-ignition atlas.

Shock	ρ	σ	Peak π_4 in (T, L)	Required impulse	(S, L)	(T, H)	(S, H)
Demand shock	0.945	0.0306	0.723	1.95	1.04	1.18	1.21
Expansionary policy shock	0.780	0.0027	0.417	3.38	1.04	1.30	1.33
Markup shock	0.798	0.0064	0.308	4.58	0.99	1.27	1.25
Matching-efficiency shock	0.970	0.0204	0.385	3.66	1.12	1.17	1.23
Technology shock	0.991	0.0051	0.165	8.55	0.87	1.10	1.01

Notes: Peak π_4 is the peak four-quarter inflation response, in percentage points, to a one-standard-deviation inflationary innovation in the generalized impulse response. Required impulse is the multiplier that makes the (T, L) response reach the surge boundary of 1.41 percentage points. The last three columns report the required impulse in each state relative to the (T, L) requirement for the same shock. Values above one mean that the state requires a larger impulse than (T, L) to reach the same boundary; values below one mean that another state is more vulnerable for that particular shock.

K Full-sample historical decomposition

Figure K.1 reports the full-sample version of the historical-decomposition evidence used in Table H.1. The decomposition is applied to the demeaned quarterly inflation observable and then cumulated into four-quarter percentage-point deviations from the estimation-sample mean. The first panel verifies the reconstruction. The second panel groups the contributions into the same shock blocks used in the main-text impulse table. The calculation is descriptive rather than causal in the high-frequency identification sense because it identifies which estimated state-space residuals account for inflation movements inside the model, while the estimated state probabilities determine the transmission environment.

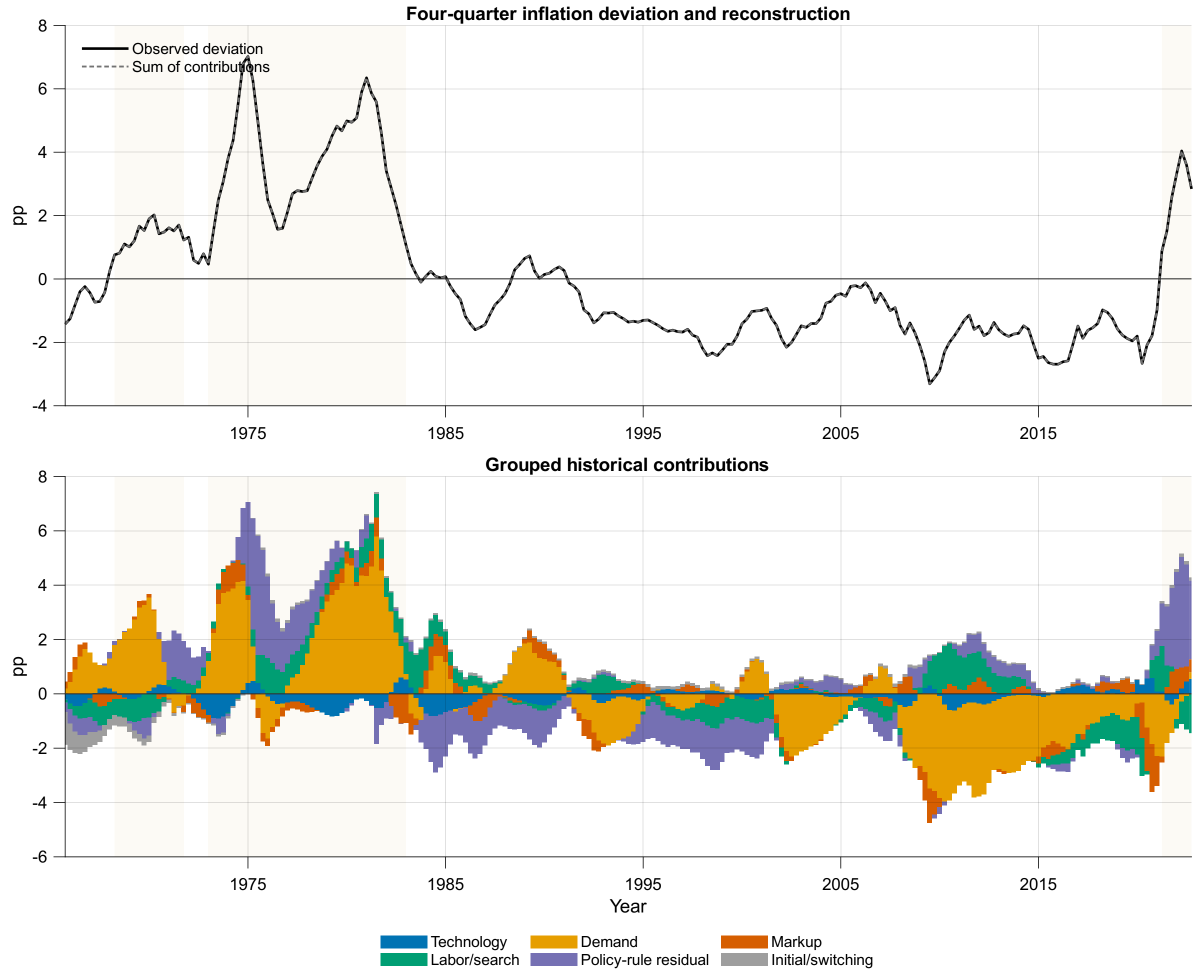


Figure K.1: Full-sample historical decomposition of four-quarter inflation deviations.

Notes: The black line is the smoothed four-quarter GDP-deflator inflation deviation from the estimation-sample mean, and the dashed line is the sum of the historical-decomposition contributions. The lower panel groups the contributions into technology, demand, markup, labor/search, policy-rule residual, and initial/switching components. Shaded regions are the inflation surge windows used in the main text.

L State-specific forecast-error variance decomposition

Table L.1 reports the numerical shares underlying Figure I.1. Unlike the historical decomposition, which conditions on the sequence of estimated shocks, this calculation holds the shock distribution fixed and varies the initial joint state. It therefore isolates how labor-market pressure and the pricing environment change the composition of inflation risk at impact and over the propagation horizon. The exercise adapts the regime-conditioned variance accounting of Bjørnland et al. (2018) to a two-dimensional labor-pricing state space.

Table L.1: Inflation forecast-error variance across joint states.

State	Technology	Demand	Markup	Labor/search	Policy
Panel A: Impact: one quarter					
(T, L)	1.1	52.0	12.0	1.4	33.6
(S, L)	4.1	46.2	19.3	3.2	27.1
(T, H)	2.1	55.6	12.0	3.3	27.0
(S, H)	4.1	51.9	15.7	5.4	23.0
Panel B: Propagation: twenty quarters					
(T, L)	4.4	51.4	7.7	19.1	17.3
(S, L)	5.4	50.7	9.8	19.0	15.0
(T, H)	5.8	54.7	5.8	23.0	10.6
(S, H)	6.2	54.5	6.4	22.9	10.0

Notes: Entries are percentages. The decomposition conditions on the initial joint state while holding the structural shock distribution fixed. Labor/search combines matching efficiency, Beveridge, separation, vacancy, participation, and wage-state innovations. Rows sum to 100 up to rounding.

Tables L.2 and L.3 extend the calculation to the remaining observables. At twenty quarters, technology and labor/search innovations account for virtually all output variation, while labor/search innovations explain more than 95 percent of unemployment variation and close to 90 percent of tightness variation in every joint state. The new-hire wage is the bridge between these real-side disturbances and inflation. At short horizons its shock composition moves substantially with the pricing environment, while at longer horizons technology and labor-market innovations again dominate. The decomposition therefore supports the economic separation used in the model. The labor block organizes real pressure, and the pricing environment governs how that pressure appears in inflation.

Table L.2: Shock incidence across observables at one quarter.

State	Technology	Demand	Markup	Labor/search	Policy
Panel A: Output					
(T, L)	60.4	1.5	1.6	34.6	1.9
(S, L)	60.1	1.9	1.7	34.0	2.3
(T, H)	57.7	3.1	0.9	34.8	3.4
(S, H)	57.6	3.2	1.0	34.6	3.6
Panel B: Unemployment					
(T, L)	2.0	1.4	1.4	93.5	1.7
(S, L)	3.5	3.0	2.2	87.8	3.6
(T, H)	1.8	2.7	0.8	91.7	3.0
(S, H)	2.9	4.7	1.2	86.1	5.1
Panel C: New-hire wage					
(T, L)	23.0	15.3	15.9	26.6	19.2
(S, L)	25.0	11.8	19.3	28.1	15.8
(T, H)	17.0	25.2	7.6	22.1	28.2
(S, H)	17.9	23.4	8.6	23.3	26.7
Panel D: Labor-market tightness					
(T, L)	4.1	2.7	2.8	87.0	3.4
(S, L)	6.6	5.6	4.1	77.1	6.7
(T, H)	3.5	5.2	1.6	83.8	5.8
(S, H)	5.3	8.5	2.3	74.5	9.4

Notes: Entries are percentages of the state-conditional forecast-error variance. The structural shock distribution is common across joint states. Labor/search combines matching efficiency, Beveridge, separation, vacancy, participation, and wage-state innovations. Rows sum to 100 up to rounding.

Table L.3: Shock incidence across observables at twenty quarters.

State	Technology	Demand	Markup	Labor/search	Policy
Panel A: Output					
(T, L)	64.4	0.1	0.5	34.8	0.1
(S, L)	64.7	0.2	0.5	34.5	0.2
(T, H)	64.4	0.3	0.4	34.6	0.3
(S, H)	64.5	0.3	0.4	34.5	0.3
Panel B: Unemployment					
(T, L)	3.3	0.1	0.5	95.9	0.1
(S, L)	3.4	0.3	0.5	95.5	0.3
(T, H)	3.3	0.3	0.4	95.8	0.3
(S, H)	3.4	0.4	0.4	95.4	0.4
Panel C: New-hire wage					
(T, L)	43.1	2.2	6.9	45.1	2.6
(S, L)	45.0	1.1	7.3	45.2	1.5
(T, H)	41.0	4.9	5.0	44.1	5.0
(S, H)	43.8	2.9	5.4	44.7	3.2
Panel D: Labor-market tightness					
(T, L)	7.4	0.4	1.2	90.6	0.5
(S, L)	7.7	0.9	1.3	89.1	1.0
(T, H)	7.2	0.9	0.9	90.1	0.9
(S, H)	7.5	1.5	0.9	88.6	1.5

Notes: Entries are percentages of the state-conditional forecast-error variance. The structural shock distribution is common across joint states. Labor/search combines matching efficiency, Beveridge, separation, vacancy, participation, and wage-state innovations. Rows sum to 100 up to rounding.

M Admissibility of the pricing hazard

Let $S_\ell \equiv \varsigma_{p,\ell}$. Before the cap binds, the hazard recursion implies

$$S_\ell = (1 - \alpha_p)S_{\ell-1} - \varphi_p S_{\ell-2}, \qquad \ell \geq 2, \tag{74}$$

with $S_0 = 1$ and $S_1 = 1-\alpha_p$. If the cap never binds and the survival sequence is summable, summing (74) over $\ell \geq 2$ gives

$$\sum_{\ell \geq 0} S_\ell = \frac{1}{\alpha_p + \varphi_p}. \tag{75}$$

If a terminal spike occurs at L and the uncapped recursion reaches exactly one, the same summation applies on the finite support $\ell = 0, \dots, L-1$. If instead the cap binds strictly at L, so that $\tilde{\alpha}_{p,L} > 1$, the finite sum satisfies

$$(\alpha_p + \varphi_p) \sum_{\ell=0}^{L-1} S_\ell = 1 + \varphi_p S_{L-2} - (1 - \alpha_p - \varphi_p)S_{L-1}, \tag{76}$$

which need not equal one. Equation (76) is the reason the strictly capped case must use the normalized distribution in (14) and recompute the pricing block rather than applying (18) directly.

The reported posterior modes and means used in the mechanism tables admit a literal hazard interpretation. In state L, the mode $(\alpha_{p,L}, \varphi_{p,L}) = (0.451, 0.040)$ produces no terminal cap and $(\alpha_p + \varphi_p) \sum_\ell \varsigma_{p,\ell} = 1$ up to machine precision. In state H, the mode $(0.047, 0.239)$ produces a terminal spike at age 13 using the rounded values in the text, with total benchmark mass 1.0000478. The discrepancy is 4.8×10^{-5} and vanishes at the unrounded parameter vector. The posterior means, $(0.446, 0.053)$ in state L and $(0.085, 0.204)$ in state H, also satisfy the normalization at numerical precision.

The estimation deliberately does not truncate the Normal priors for φ_p. It follows the same prior family used in the related Bayesian estimates and imposes only the ordering that labels state H as the state with weakly greater duration dependence. Consequently, posterior vectors are always interpreted as draws of the reduced-form coefficient system, whereas a literal probability-hazard interpretation is attached only to admissible vectors satisfying $0 < \alpha_p < 1$, $\varphi_p \geq 0$, and (12). A vector that triggers a strict terminal cap and fails the normalization would require the finite-support distribution in (14) and a recomputation of the micro-founded pricing block. No such micro-founded interpretation is assigned to that vector in the posterior analysis.