Markup Distribution and Inflation Dynamics*

Mario Giarda
Central Bank of Chile

Antonio Martner
Central Bank of Chile

September 2026

Abstract

We study how the distribution of firm-level markups shapes inflation. Using Chilean administrative data, we show aggregate markup cyclicality is driven by incumbent firms at the top of the markup distribution. A model with Kimball demand and common Rotemberg price adjustment costs accounts for this pattern: high-markup firms have flatter pricing slopes, so larger gaps are associated with firms whose prices adjust less, adding a covariance term to the Phillips curve. Using a sufficient-statistic approach, we find that the covariance is negative, comparable in magnitude to the average markup-gap term, and attenuates inflation fluctuations associated with that term.

*Previous versions of the paper were circulated under the title ”Markup Distribution and Aggregate Dynamics”. We thank seminar participants at CREI, SED, Summer SED, CEBRA, LACEA, CEMLA, and the Central Bank of Chile for excellent suggestions. The views expressed are those of the authors and do not necessarily represent the views of the Central Bank of Chile or its board members. This study was developed within the scope of the research agenda conducted by the Central Bank of Chile (CBC) in economic and financial affairs of its competence. The CBC has access to anonymized information from various public and private entities, by virtue of collaboration agreements signed with these institutions. To secure the privacy of workers and firms, the CBC mandates that the development, extraction and publication of the results should not allow the identification, directly or indirectly, of natural or legal persons. Officials of the Central Bank of Chile processed the disaggregated data. All the analysis was implemented by the authors and did not involve nor compromise the Servicios de Impuestos Internos de Chile. The information contained in the databases of the Chilean IRS is of a tax nature originating in self-declarations of taxpayers presented to the Service; therefore, the veracity of the data is not the responsibility of the Service. Authors emails: mgiarda@bcentral.cl, amartner@bcentral.cl

1 Introduction

Markup cyclicality is central to New Keynesian theories of inflation. In them, a contractionary demand shock lowers marginal costs relative to prices, raising markups because prices adjust only gradually due to price rigidities. In the canonical formulation, the strength of this force is summarized by an aggregate markup gap, defined as the observed minus the flexible-price markup (Galí, 2015). However, that setup abstracts from whether the distribution of markup gaps across firms affects inflation dynamics. In this paper, we study this question with Chilean microdata and, using a sufficient statistic approach, show that the distributional component attenuates the inflation response associated with the average markup gap by about 40%.

We use monthly administrative tax records from Chile to estimate firm-level markups using balance-sheet information and transaction-level data on prices and quantities, following Hall (1988) and De Loecker and Warzynski (2012). Our key empirical object, however, depends only on the joint distribution of firm-level markups and sales: these are sufficient to characterize the average markup-gap term and the distributional covariance in the New Keynesian Phillips curve (NKPC). Transaction prices enter only in the construction of physical-quantity production functions used to measure markups.

We document three findings about markup cyclicality in Chile. First, aggregate markups are countercyclical, consistent with a large literature studying the behavior of markups over the business cycle (Bils et al., 2018; Nekarda and Ramey, 2020). Second, aggregate markup movements are driven primarily by the intensive margin, as continuing firms change their own markups, while reallocation and composition generate a smoother offsetting component. Third, markup responses differ systematically across the markup distribution. Following identified contractionary monetary and financial shocks, firms in the top decile of the persistent markup distribution raise their markups more than firms in the bottom decile, and this difference persists over the two-year horizon.

These findings point to a pricing mechanism that operates within continuing firms and varies with firms' positions in the markup distribution. To answer the relevance of these findings, we present a model using Kimball demands and Rotemberg price adjustment costs (Kimball, 1995; Rotemberg, 1982), building on the demand structure of Edmond et al. (2023). Under Kimball demand, firms differ in their steady-state markups but face the same price-adjustment technology. Firms with higher steady-state markups have lower desired-price pass-through and flatter profit functions around their optimum. High-markup firms therefore have flatter firm-level Phillips curves and adjust prices less in response to a marginal-cost shock. Their realized markups consequently absorb more of the shock, generating stronger markup responses.

Then we aggregate these firm-level Phillips curves. We show that aggregate inflation depends on the average markup gap times an average slope, and the covariance between firm-level markup gaps and firm pricing slopes, which we call the distributional term. We find that the sign of this covariance determines the role of heterogeneity: when markup gaps are concentrated among firms with lower pricing slopes, the covariance is negative and the distributional term offsets part of the inflation response associated with the average markup gap. This is consistent with the pattern in markup cyclicality: firms with higher steady-state markups have lower pricing slopes in the model and exhibit larger markup responses as in the data.

We use a sufficient statistic approach to quantify the contribution of the distributional term by recovering the cross-sectional distribution of firm-level markup gaps. Following the Kimball demand mapping, we use observed markups and market shares to recover a model-implied flexible-price markup associated with each firm's position on its residual demand curve. This provides a measure of the desired markup and hence of the markup gap. Steady-state markups, together with the estimated curvature of demand, determine firms' pricing slopes. Using these objects, we construct both terms of the aggregate Phillips curve. In our baseline accounting, the covariance term offsets roughly 40% of the inflation response associated with the average markup-gap term.

The resulting framework provides a parsimonious link between firm-level markup heterogeneity and aggregate inflation. It uses firm-level markups and sales to construct the distribution of markup gaps and their location across firms with different pricing slopes, without estimating firm-specific nominal rigidities. The approach is therefore portable to other settings with firm-level markups and sales.

Our results relate to the literature on markup cyclicality and heterogeneous monetary transmission. A large literature studies the cyclical behavior of markups and how it depends on markup measurement and the underlying shock (Bils et al., 2018; Nekarda and Ramey, 2020; Anderson et al., 2018; Burstein et al., 2020). We take advantage of the granularity of the data to use measures of quantities in the production function estimation and construct monthly markups across all industries and firm sizes in the firm population.

Other work studies heterogeneous responses to monetary policy across firms with different degrees of market power or price rigidity (Meier and Reinelt, 2024; Chiavari et al., 2021; Kouvavas et al., 2021; Höynck et al., 2023). We show that aggregate markup cyclicality is driven primarily by within-firm movements and that these movements differ systematically across the persistent markup distribution. Together, these findings motivate an NKPC with a distributional term that depends on the location of markup gaps across firms.

Our mechanism is closely related to recent work on how markup heterogeneity shapes the aggregate effects of monetary policy. Baqaee et al. (2024) and Meier and Reinelt (2024) emphasize how monetary shocks reallocate activity across firms with different markups, pass-through, or price rigidities, with consequences for aggregate productivity and monetary transmission. The object here is different. Our distributional term does not require changes in firms' aggregate weights. Holding those weights fixed, inflation depends on which firms carry the realized markup gaps and on those firms' pricing slopes. This distinction is motivated directly by the data: the cyclical movement in aggregate markups comes primarily from continuing firms' own markup changes rather than from reallocation and composition.

2 Firm-Level Anatomy of Markup Cyclicality

We study the firm-level origins of aggregate markup fluctuations and document three findings. Aggregate markups are countercyclical; their movements are driven primarily by incumbent firms changing their own markups, while reallocation partly offsets them; and firms higher in the persistent markup distribution exhibit larger responses to contractionary demand shocks.

Data and Markup Measurement. We use monthly administrative tax data from Chile for 2007–2024, covering over 6 million firm-month observations across 626 six-digit ISIC industries. The data combine firm-level sales, intermediate input expenditures, employment, wages, investment, and capital with the universe of electronic tax invoices, which report product-level prices and quantities for firm-to-firm transactions. We estimate firm-level markups following De Loecker and Warzynski (2012). Let θ_j^M denote the time-invariant output elasticity of materials in industry j, and α_{it}^M firm i's material expenditure share in revenue, so that $\mu_{it} = \theta_j^M(\alpha_{it}^M)^{-1}$.

We estimate Cobb–Douglas production functions in labor, capital, and materials at the six-digit industry level using Ackerberg et al. (2015). Output and inputs are deflated by firm-specific Tornqvist price indices built from invoice-level prices and quantities, so the production functions are estimated using physical quantities rather than revenue measures. This distinction matters for identification. As emphasized by Bond et al. (2021), when output elasticities are estimated from revenue production functions, the material elasticity is not separately identified from the markup. That problem does not arise here because θ_j^M is estimated from physical output and input quantities.[1]

Because θ_j^M is fixed within six-digit industries, all within-firm markup variation comes from the material expenditure share. This rules out within-industry cyclical variation in the material elasticity but allows factor intensity to vary across industries.[2]

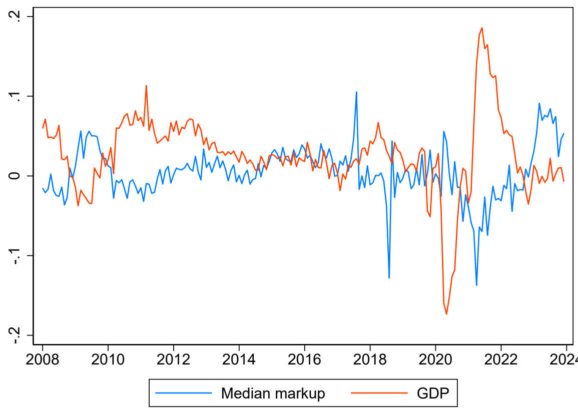
Aggregate markup cyclicality and the distribution of firm markups. We first examine the unconditional comovement between GDP and moments of the firm-level markup distribution. Figure 1 shows that markups are countercyclical. The twelve-month log change in the median firm-level markup has a correlation of -0.44 with twelve-month GDP growth. The cyclicality is concentrated in the upper tail of the markup distribution. The twelve-month log change in the 90th percentile of markups has a correlation of -0.52 with GDP growth, while the corresponding correlation for the 10th percentile is close to zero. Thus, markup cyclicality is concentrated in the upper tail.

[1]The invoice data begin in 2014, so we estimate production functions on 2014–2024 and apply the resulting industry-level elasticities to the full 2007–2024 panel.

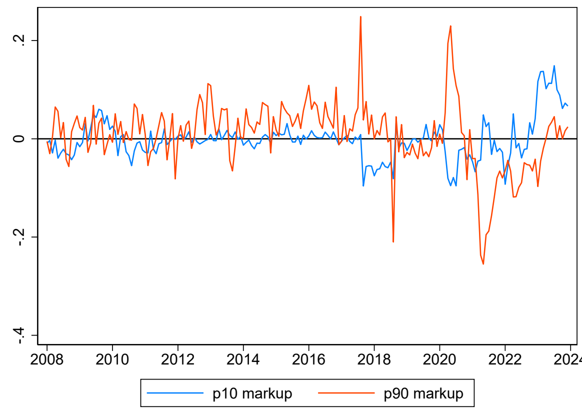
[2]Adjustment frictions in material inputs can also generate countercyclical material shares, and hence countercyclical measured markups, independently of pricing behavior. This is an alternative interpretation of the unconditional cyclicality documented below and one reason we do not rely on that evidence alone. Explaining the later findings through this channel would additionally require such frictions to vary systematically with firms' persistent markups and with the within-firm versus reallocation margin.

Figure 1: Anatomy of Aggregate Markup Cyclicality

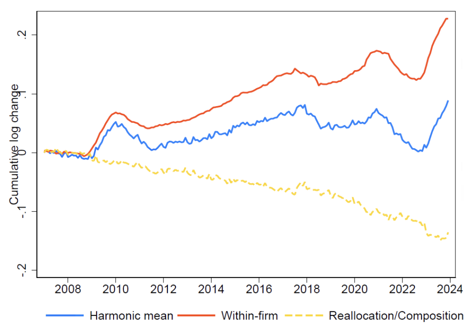
(a) Median markup and GDP



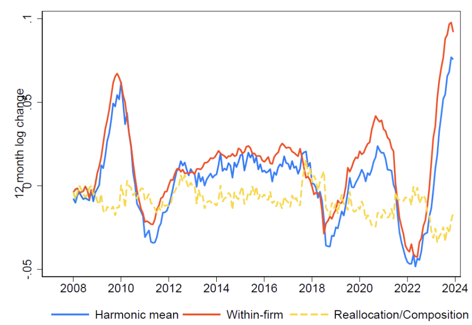
(b) Markup P10 and P90



(c) Cumulative aggregate markup decomposition



(d) 12-month aggregate markup decomposition



Notes: Panel (a) plots the 12-month log change in the median firm-level markup and the 12-month growth rate of Chile's monthly GDP index (IMACEC). Panel (b) plots the corresponding 12-month log changes in the 10th and 90th percentiles of the firm-level markup distribution. Panels (c) and (d) decompose the aggregate harmonic markup, $H_t = (\sum_i s_{it}/\mu_{it})^{-1}$, where s_{it} is firm i's sales share, into a within-firm component and a residual reallocation/composition component. The within-firm component changes continuing firms' markups holding firm weights fixed, while the residual captures changes in firm weights and in the composition of active firms. Panel (c) reports cumulative log changes and Panel (d) 12-month log changes. Markups are constructed from firm-level tax and invoice data using physical quantity-based production functions.

Aggregate fluctuations are primarily within firms. Aggregate markup movements may reflect continuing firms changing their own markups, reallocation across firms, or changes in the composition of active firms. We decompose the aggregate harmonic markup into a within-firm component and a residual reallocation/composition component (Figure 1).

Panel (c) shows that the cumulative rise in aggregate markups is driven by the intensive margin. Over the sample, the aggregate harmonic markup rises by about 9%.

Continuing firms' own markup changes contribute 22%, while reallocation/composition contributes -13%, partly offsetting the increase generated by incumbents.

Panel (d) shows the same pattern at business-cycle frequencies: year-over-year aggregate markup fluctuations track the within-firm component, while reallocation/composition is smaller and often moves in the opposite direction. Thus, both the trend and cyclical movements in aggregate markups are driven primarily by incumbent firms' own markup changes rather than by shifts in firm weights or composition.

Markup responses to aggregate demand shocks. The previous evidence does not isolate the shocks behind markup movements. We therefore use identified contractionary demand shocks to test whether markups rise and whether the response is stronger among firms higher in the persistent markup distribution.

We use two shocks. Monetary policy surprises (MPS) from Aruoba et al. (2021) are the difference between the policy rate chosen by the Central Bank of Chile and market expectations on the day of the policy meeting. The shock have been shown to lower output and inflation in Chile, so we interpret positive innovations as contractionary demand shocks. We also use financial shocks based on the Excess Bond Premium (EBP) following Gilchrist and Zakrajšek (2012) and interpreted homologous to the MPS.

For each horizon h, we estimate

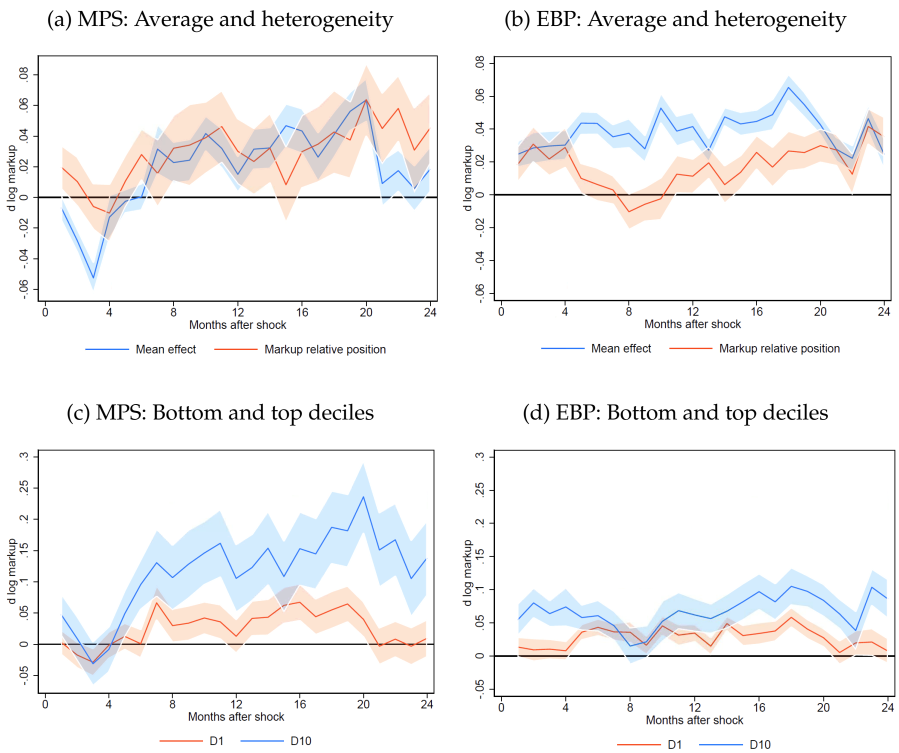
$$\log \mu_{i,t+h} - \log \mu_{i,t-1} = \alpha_h + \beta_h shock_t + \gamma_h(\mu_i^P - \overline{\mu}^P) shock_t + \Gamma_h' Z_{it} + \theta_{im} + \epsilon_{i,t+h}. \tag{1}$$

The dependent variable is the cumulative change in firm i's log markup from the month before the shock to horizon h. The coefficient β_h is the response at the economy-wide mean persistent markup, while γ_h captures how the response varies with persistent markup.[3] The fixed effects θ_{im} are firm-by-calendar-month effects, absorbing firm-specific seasonality, and Z_{it} includes employment measured before the shock. Standard errors are clustered at the six-digit industry level, allowing for arbitrary correlation across firms and over time within industries. We use the full sample for the EBP shock, while restricting the MPS sample to 2010–2019, after the Great Recession and before COVID-19, to exclude episodes in which monetary policy decisions were dominated by unusually large macroeconomic disturbances.

[3] We estimate $\log \mu_{it} = \alpha + \theta_i + \theta_{ms} + \theta_{ys} + \upsilon_{it}$, where θ_i is a firm effect, θ_{ms} a sector-by-calendar-month effect, and θ_{ys} a sector-by-year effect. We define $\mu_i^P = \exp(\alpha + \theta_i)$ and use it to rank firms in the economy-wide persistent markup distribution.

Figure 2 reports the results. Contractionary demand shocks raise firm-level markups, and the response is stronger among firms higher in the persistent markup distribution.

Figure 2: Markup Responses to Demand Shocks



Notes: The figure reports panel local projections of firm-level markups following identified contractionary demand shocks over a 24-month horizon. Panels (a) and (b) report the response at the economy-wide mean persistent markup, β_h, and its interaction with persistent markup, γ_h. Panels (c) and (d) report responses for firms in the bottom and top deciles of the persistent markup distribution. Panels (a) and (c) use monetary policy surprises; panels (b) and (d) use EBP shocks. Shaded areas are 95% confidence intervals. Standard errors are clustered at the six-digit industry level.

Panels (a) and (b) show that markups rise following both shocks. For a firm at the mean persistent markup, the response to a monetary policy surprise peaks around 6–7% between months 12 and 16 and remains elevated at 24 months. Following an adverse EBP shock, markups rise by about 5–6% and also remain above baseline.

The response also varies across the markup distribution. In panels (a) and (b), γ_h is predominantly positive at medium horizons, indicating stronger markup responses among firms with higher persistent markups. Panels (c) and (d) show the same pattern in the tails of the distribution: top-decile firms exhibit larger and more persistent markup increases after both shocks, while bottom-decile firms respond weakly or not at all. The difference is more pronounced following monetary policy surprises.

Combined with the decomposition above, these results show that markup cyclicality operates primarily within firms and is stronger among firms with persistently high markups. This pattern motivates the model.

3 Model

The empirical evidence points to a pricing mechanism operating within continuing firms and varying with firms' positions in the markup distribution. We therefore build a model with Kimball demand and a common price-adjustment technology (we abstract from reallocation and entry). Heterogeneity in steady-state markups generates heterogeneity in pricing responses: high-markup firms have lower desired-price pass-through and flatter profit functions, so their prices adjust less and their realized markups fluctuate more following a common marginal-cost shock.

In the aggregate, this heterogeneity adds a distributional component to the New Keynesian Phillips curve. Inflation depends both on an average markup gap and on a covariance between firms' markup gaps and their pricing slopes. The model also provides a parsimonious way to quantify both components using firm-level markups and sales: steady-state markups and demand curvature discipline firms' pricing slopes, while observed markups and sales determine the distribution of markup gaps and their aggregate weights. Thus, the same average markup gap can imply different inflation responses depending on where those gaps are located across firms.

Demand Environment. We use the Kimball demand specification in Klenow and Willis (2016) and Edmond et al. (2023). Firm i faces residual demand with elasticity

$$\sigma_i(\tilde{p}_i) \equiv -\frac{\partial \log y_i}{\partial \log \tilde{p}_i}, \qquad \tilde{p}_i \equiv \frac{p_i}{P}, \tag{2}$$

where $\tilde{p}_i$ is the firm's price relative to the aggregate price index. Locally, the Kimball specification satisfies

$$\frac{\partial \log \sigma_i}{\partial \log \tilde{p}_i} = b\sigma_i, \qquad b > 0, \tag{3}$$

so a higher relative price raises the residual demand elasticity. Under CES demand, $b = 0$ and the elasticity is constant.[4]

The firm's flexible-price desired markup satisfies $\mu_{it}^* = \sigma_{it}/(\sigma_{it} - 1)$. Because residual demand elasticity varies with the firm's position on its demand curve, the desired markup varies across firms and over time. Evaluating this condition at the firm's steady-state relative position defines $\bar{\mu}_i$, its steady-state desired markup. Defining the realized markup as $\mu_{it} = p_{it}/mc_{it}$, we get the log of markup gap

$$\tilde{\mu}_{it} \equiv \log \mu_{it} - \log \mu_{it}^*,$$

which measures the gap between the firm's realized markup and the markup it would choose under flexible prices.

Pass-through. Under Kimball demand, desired markups vary with relative prices, so flexible-price pass-through from marginal costs to prices is generally below one. Define

$$\rho_i \equiv \frac{\partial \log p_i^*}{\partial \log mc_i},$$

where p_i^* is the firm's flexible-price desired price, holding the aggregate price index fixed. The flexible-price condition $p_i^* = \mu_i^*(p_i^*/P)mc_i$ implies

$$d \log p_i^* = \frac{\partial \log \mu_i^*}{\partial \log p_i^*} d \log p_i^* + d \log mc_i.$$

Using $\mu_i^* = \sigma_i/(\sigma_i - 1)$ and (3), $\partial \log \mu_i^* / \partial \log \tilde{p}_i = -b\mu_i^*$. Evaluating at the firm's steady-state position gives

$$\rho_i = \frac{1}{1 + b\bar{\mu}_i}. \tag{4}$$

Pass-through therefore decreases with the firm's steady-state markup: following a change in marginal cost, high-markup firms move their desired prices less.

Profit function curvature. Profit curvature is the second object linking the markup distribution to price adjustment. Holding the aggregate price index fixed, firm i's static

[4] In Edmond et al. (2023), $\sigma_i = \bar{\sigma} q_i^{-b}$, where q_i denotes relative size. Since $d \log \tilde{p}_i / d \log q_i = -1/\sigma_i$, this implies (3).

profits are

$$\Pi_i(p_i) = (p_i - mc_i)\, y_i\Big(\frac{p_i}{P}\Big), \qquad \frac{\partial \Pi_i}{\partial \log p_i} = p_i y_i \left[1 - \sigma_i \left(1 - \frac{mc_i}{p_i}\right)\right].$$

At the flexible-price optimum, the term in brackets is zero, which delivers $\mu_i^* = \sigma_i/(\sigma_i - 1)$. Differentiating once more and normalizing by firm revenue, define the curvature of the profit loss around the optimum as

$$C_i \equiv -\frac{1}{p_i y_i} \frac{\partial^2 \Pi_i}{\partial (\log p_i)^2}\bigg|_{p_i = p_i^*} = (\sigma_i - 1) + \frac{\partial \log \sigma_i}{\partial \log \tilde{p}_i}.$$

Using (3) and evaluating at the steady-state position gives

$$C_i = \frac{1 + b\bar{\mu}_i}{\bar{\mu}_i - 1} = \frac{\sigma_i - 1}{\rho_i}. \tag{5}$$

Thus, profit curvature decreases with the firm's steady-state markup.

Pass-through and curvature capture different margins. ρ_i determines how far the firm's desired price moves when marginal cost changes, while C_i determines how rapidly profits fall as the firm's price moves away from that optimum. Their product is

$$C_i \rho_i = \sigma_i - 1 = \frac{1}{\bar{\mu}_i - 1}, \tag{6}$$

so the local pricing incentive generated by a marginal-cost change depends only on the firm's steady-state markup. Higher-markup firms therefore have lower desired-price pass-through, flatter profit functions, and weaker local incentives to adjust prices after a marginal-cost shock. Following a contractionary shock that lowers marginal costs, they have weaker incentives to cut prices and absorb more of the shock through higher realized markups.

Price adjustment. Around the flexible-price optimum, the static profit loss from setting p_{it} instead of p_{it}^* is, up to second order, $\frac{1}{2} C_i (\log p_{it} - \log p_{it}^*)^2$, normalized by firm revenue. Suppose firms face Rotemberg adjustment costs $(\phi_R/2)(\Delta \log p_{it})^2$, with ϕ_R common across firms (Rotemberg, 1982; Woodford, 2003). The firm's pricing problem is

$$\min_{\{p_{it}\}} \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \left[\frac{C_i}{2} \Big(\log p_{i,t+s} - \log p^*_{i,t+s} \Big)^2 + \frac{\phi_R}{2} \Big(\Delta \log p_{i,t+s} \Big)^2 \right].$$

Its first-order condition is

$$\pi_{it} - \beta \mathbb{E}_t \pi_{i,t+1} = -\frac{C_i}{\phi_R} \Big(\log p_{it} - \log p^*_{it} \Big), \tag{7}$$

where $\pi_{it} = \Delta \log p_{it}$. The effective pricing slope is therefore

$$\lambda_i^R = \frac{C_i}{\phi_R}.$$

With a common adjustment-cost parameter, differences in pricing slopes come entirely from profit curvature: firms with lower C_i respond less to a given price gap.

The same condition determines the speed of price adjustment. Consider a permanent change in the desired price and let $\Phi_i \in (0, 1)$ denote the fraction of the remaining price gap closed each period. Solving (7) gives

$$C_i(1 - \Phi_i) = \phi_R \Phi_i \Big[1 - \beta(1 - \Phi_i) \Big]. \tag{8}$$

A larger C_i implies a larger Φ_i. Since profit curvature decreases with the steady-state markup, high-markup firms close price gaps more slowly even though all firms face the same adjustment-cost parameter.

Markup cyclicality. Adjustment speed maps directly into markup cyclicality. A marginal-cost shock moves firm i's flexible-price desired price by $\rho_i \, d \log mc_t$. On impact, the firm implements a fraction Φ_i of this movement, so $d \log p_{it} / d \log mc_t = \Phi_i \rho_i$. Since $\log \mu_{it} = \log p_{it} - \log mc_t$,

$$\frac{d \log \mu_{it}}{d \log mc_t} = -\Big(1 - \Phi_i \rho_i \Big) = \underbrace{\Big(\rho_i - 1 \Big)}_{\text{demand}} - \underbrace{\Big(1 - \Phi_i \Big) \rho_i}_{\text{price adjustment}} . \tag{9}$$

Markup cyclicality therefore reflects two forces. Lower pass-through makes the flexible-price markup more countercyclical, while slower adjustment makes the realized markup absorb more of the marginal-cost shock. Since both ρ_i and Φ_i decrease with the steady-state markup, high-markup firms move their prices less and their markups more following a common shock, even though all firms face the same adjustment-cost parameter.[5]

The same mechanism operates beyond impact. Following a permanent marginal-cost shock, the cumulative price response after h periods is $\Theta_i(h) = \rho_i[1-(1-\Phi_i)^{h+1}]$, implying

$$\frac{d\log\mu_{i,t+h}}{d\log mc_t} = (\rho_i - 1) - \rho_i(1-\Phi_i)^{h+1}. \tag{10}$$

At short horizons, nominal rigidity makes firms absorb more of the shock through markups. As the price gap closes, this component decays and markup cyclicality converges to the flexible-price response $\rho_i - 1$. Because high-markup firms have both lower pass-through and slower adjustment, their markups respond more at every horizon. This is the ordering observed in Figure 2, where firms in the top decile exhibit larger markup responses than firms in the bottom decile.[6]

An aggregate Phillips Curve. We now aggregate the firm-level pricing conditions. Since realized and desired markups are evaluated at the same marginal cost,

$$\tilde{\mu}_{it} = \log\mu_{it} - \log\mu^*_{it} = \log p_{it} - \log p^*_{it},$$

so (7) can be written as

$$\pi_{it} - \beta\mathbb{E}_t\pi_{i,t+1} = -\lambda_i^R\tilde{\mu}_{it}, \qquad \lambda_i^R = \frac{C_i}{\phi_R}. \tag{11}$$

A positive markup gap means that the firm's price is above its flexible-price desired price. The force pushing inflation down is therefore stronger when the gap is larger and when the firm's pricing slope is steeper.

[5]The same demand-side force is present under other nominal frictions. Under Calvo pricing, Kimball demand generates heterogeneous pass-through, while the reset probability is exogenous. Under menu costs, lower profit curvature also reduces the gain from adjustment, lowering the probability of resetting. Rotemberg pricing preserves this adjustment incentive while delivering a closed-form characterization of markup dynamics.

[6]Equation (10) describes the markup response to a given marginal-cost path and is monotone in h for a permanent shock. The empirical responses are hump-shaped because monetary policy affects marginal costs gradually, so the estimated impulse response combines these marginal-cost dynamics with the pricing response characterized here. The model speaks to the cross-sectional comparison rather than to the common marginal-cost path.

Let ω_{it} denote firm i's weight in the aggregate price index, with $\sum_i \omega_{it} = 1$, and define $\mathbb{E}_{\omega,t}[x_{it}] \equiv \sum_i \omega_{it} x_{it}$. Aggregating the firm-level pricing conditions gives, to first order,

$$\pi_t - \beta \mathbb{E}_t \pi_{t+1} = -\bar{\lambda}_t^R \, \mathbb{E}_{\omega,t}[\tilde{\mu}_{it}] - \mathrm{Cov}_{\omega,t}\Big(\lambda_i^R, \tilde{\mu}_{it}\Big), \tag{12}$$

where $\bar{\lambda}_t^R = \mathbb{E}_{\omega,t}[\lambda_i^R]$. [7]

The first term is the contribution of the average markup gap. The second is a distributional term that captures where markup gaps are located across firms with different pricing slopes. Whenever the covariance is nonzero, the average markup gap alone is not sufficient to characterize inflation dynamics.

The model and empirical evidence together discipline the sign of this covariance. Following a contractionary shock, firms with higher steady-state markups exhibit larger increases in realized markups, while the model implies that these firms have lower pricing slopes. Markup gaps are therefore concentrated among firms with lower pricing slopes, implying $\mathrm{Cov}_{\omega,t}(\lambda_i^R, \tilde{\mu}_{it}) < 0$. The distributional term then offsets part of the inflation response associated with the average markup gap. Following an expansionary shock the signs reverse, but the same mechanism attenuates the inflationary response.

This mechanism differs from the reallocation channel emphasized by Baqaee et al. (2024) and Meier and Reinelt (2024). In those papers, markup heterogeneity matters because monetary policy reallocates activity across firms with different markups and pass-throughs. Here, the distributional term operates holding firms' aggregate weights fixed: it depends on which firms carry the markup gaps and on those firms' pricing slopes. This distinction matches the evidence in Section 2, where aggregate markup movements are driven primarily by within-firm adjustment rather than reallocation.

4 Quantification

The model above shows that inflation depends both on the average markup gap and a distributional term given by the covariance between firm pricing slopes and markup gaps. We quantify both terms with a deliberately simple accounting exercise that requires no calibration of nominal rigidity. We recover profit curvature and the markup gap from observed markups and sales alone and study whether the recovered covariance is neg-

[7] To first order, $\pi_t = \mathbb{E}_{\omega,t}[\pi_{it}]$. Aggregating the firm-level NKPC gives $-\mathbb{E}_{\omega,t}[\lambda_i^R \tilde{\mu}_{it}]$, which can be decomposed into the product of means and the covariance. We treat the weights ω_{it} as fixed when passing the expectation term through the sum, as appropriate to first order around a steady state with constant weights.

ative and whether its magnitude is of the same order as the average markup-gap term. The covariance is negative throughout the sample, and the implied attenuation is of the same order of magnitude, at roughly 40 percent of the average-gap term.

A sufficient statistic. Under Rotemberg pricing, $\lambda_i^R = C_i/\phi_R$. Because ϕ_R is common across firms, it cancels when the distributional term is measured relative to the average markup-gap term:

$$-\frac{D_t}{M_t} \equiv -\frac{\mathrm{Cov}_{\omega,t}\left(\lambda_i^R, \tilde{\mu}_{it}\right)}{\bar{\lambda}_t^R\, \mathbb{E}_{\omega,t}[\tilde{\mu}_{it}]} = -\frac{\mathrm{Cov}_{\omega,t}\left(C_i, \tilde{\mu}_{it}\right)}{\mathbb{E}_{\omega,t}[C_i]\, \mathbb{E}_{\omega,t}[\tilde{\mu}_{it}]}. \tag{13}$$

A positive value of $-D_t/M_t$ means that the distributional term offsets part of the inflation response associated with the average markup-gap term.

Kimball demand maps markups and sales into the two firm-level objects in (13). Given the steady-state markup $\bar{\mu}_i$ and demand curvature b, profit curvature follows from (5). Recovering the markup gap additionally requires the flexible-price desired markup μ_{it}^*. Under Kimball demand, a firm's market share determines its position on the residual demand curve and hence the elasticity and desired markup associated with that position. For the specification of Edmond et al. (2023), defining

$$g(\mu) = 1/\mu + \log(1 - 1/\mu)$$

gives

$$g(\mu_{it}^*) = a_i + a_{st} + b \log \omega_{it}^m, \tag{14}$$

where ω_{it}^m is firm i's within-market sales share, a_i is a firm fixed effect, and a_{st} is a market-time fixed effect. Since g is monotone for $\mu > 1$, this relationship maps market shares into desired markups and hence into markup gaps,

$$\tilde{\mu}_{it} = \log \mu_{it} - \log \mu_{it}^*.$$

Measurement. We construct each object in (13) from the monthly firm panel. The structural mapping requires $\mu_{it} > 1$, so we restrict the quantification to firm-month observations with estimated markups above one. This sample accounts for 74.2% of aggregate value added. We use two sales shares for different purposes: ω_{it}^m, firm i's sales share within its market, is used to estimate the demand system, while ω_{it}, its share of aggregate sales, is used to aggregate the firm-level objects in (12).[8]

We first estimate the Kimball curvature parameter b, which governs how the elasticity of residual demand changes along the demand curve. Using the markup–market-share relationship in (14), we estimate

$$g(\mu_{it}) = a_i + a_{st} + b \log \omega_{it}^m + u_{it}, \tag{15}$$

where a_i absorbs persistent differences across firms and a_{st} conditions common to firms in the same market and month. The estimate uses within-firm variation in market share relative to other firms in the same market and month. Standard errors are clustered by firm.

We then use the fitted systematic component

$$\widehat{g}_{it}^* = \widehat{a_i} + \widehat{a_{st}} + \widehat{b} \log \omega_{it}^m,$$

as the model-implied value of $g(\mu_{it}^*)$. For fitted values satisfying the structural restriction $\widehat{g}_{it}^* < 0$, which holds for 97% of the sample, we numerically invert g to obtain $\widehat{\mu}_{it}^*$ and construct[9]

$$\widehat{\tilde{\mu}}_{it} = \log \mu_{it} - \log \widehat{\mu}_{it}^*. \tag{16}$$

The remaining object is profit curvature, which depends on the steady-state markup $\bar{\mu}_i$. We proxy $\bar{\mu}_i$ with the persistent component of firm markups we used in section 2. We regress log markups on firm fixed effects, industry-by-calendar-month fixed effects, and industry-by-year fixed effects, and define $\bar{\mu}_i$ as the exponential of the estimated firm component, including the regression constant.[10] We then construct

$$\widehat{C}_i = \frac{1 + \widehat{b}\bar{\mu}_i}{\bar{\mu}_i - 1}. \tag{17}$$

These objects determine M_t and D_t in (13). Because the common Rotemberg parameter cancels from their ratio, the quantification does not require calibrating nominal rigidity.

[8] We define markets in this exercise as one-digit ISIC industries.

[9] Because the inverse mapping becomes steep as $\widehat{g}_{it}^*$ approaches zero, we winsorize $\widehat{\mu}_{it}^*$ at percentiles 1 and 99.

[10] We winsorize $\bar{\mu}_i$ at the 1 and 99 percentiles.

Three caveats limit the interpretation of the exercise. First, following Edmond et al. (2023), we use the relationship between markups and market shares implied by Kimball demand to estimate the demand curvature parameter b. In their flexible-price environment, this relationship directly links observed markups to firms' positions on the demand curve. In our setting, observed markups may additionally contain a markup gap due to price adjustment frictions. We therefore use the fitted systematic component of (15) as the model-implied flexible-price component, while allowing residual variation to reflect deviations from the flexible-price optimum and other sources of variation. In addition, because firm revenue enters both the measured markup and the market-share measure, part of their empirical comovement may arise mechanically. Second, because the inverse mapping from g to μ is nonlinear, residual variation in $g(\mu)$ can mechanically affect the level of the recovered markup gap. Third, we use aggregate sales shares as empirical counterparts of the price-index weights. We therefore interpret the exercise as model-based accounting of the sign and order of magnitude of the distributional term, rather than as point identification of a structural attenuation parameter.

Results. Table 1 reports the estimated demand curvature, the recovered firm-level objects, and the resulting aggregate statistics.

The recovered covariance has the sign implied by the mechanism. The sales-weighted correlation between profit curvature and the recovered markup gap is negative in all 204 months, with a median of -0.25 and an interquartilerange of -0.28 to -0.21. Thus, larger recovered markup gaps are concentrated among firms with flatter profit functions and, in the model, lower pricing slopes.

The distributional term is also of the same order of magnitude as the average-gap term. In the baseline accounting, it offsets roughly 40 percent of the latter: the monthly attenuation ratio has a mean of 0.41, a median of 0.43, and an interquartile range of 0.35 to 0.48. Taking the ratio after aggregating D_t and M_t over time gives the same value, 0.41. Panel C also shows that the magnitude is stable under progressively tighter restrictions on the markup distribution. As the minimum markup increases from one to 1.20, the structural sample falls from 74 to 61 percent of aggregate value added and the estimated demand curvature falls from 0.11 to 0.04, while the attenuation remains between 0.40 and 0.42.

Table 1: Quantification of the Distributional Component

	Estimate / Mean	S.E.	Median	IQR
Panel A. Demand curvature				
Kimball curvature, $\widehat{b}$	0.1055	0.0015		
Panel B. Model-implied firm-level objects				
Desired markup, $\widehat{\mu}^*_{it}$	1.713		1.405	[1.226, 1.801]
Markup gap, $\widehat{\tilde{\mu}}_{it}$	0.226		0.132	[−0.042, 0.413]
Persistent markup, $\bar{\mu}_i$	2.252		1.726	[1.415, 2.489]
Profit curvature, $\widehat{C}_i$	2.202		1.628	[0.848, 2.773]
Panel C. Aggregate implication				
Correlation, $\text{Corr}_{\omega,t}(\widehat{C}_i, \widehat{\tilde{\mu}}_{it})$	−0.244		−0.250	[−0.281, −0.208]
Distributional attenuation, baseline	0.409		0.426	[0.348, 0.479]
Minimum markup > 1.05	0.418		0.435	[0.352, 0.504]
Minimum markup > 1.10	0.417		0.436	[0.358, 0.485]
Minimum markup > 1.20	0.403		0.409	[0.342, 0.480]

Notes: Panel A reports the estimate of the Kimball curvature parameter from equation (15); the standard error is clustered by firm. Panel B summarizes the model-implied firm-level objects on the common sample used in the baseline aggregate quantification. Desired and persistent markups are winsorized at percentiles 1 and 99. Panel C reports the monthly distribution of the attenuation statistic in equation (13). The baseline statistic is positive in all 204 months. The ratio of the time-series sums, $-\sum_t D_t / \sum_t M_t$, is 0.409. The remaining rows reestimate the full procedure after imposing the indicated minimum markup. The structural sample requires positive sales and estimated markups above one; it accounts for 74.2% of aggregate value added.

More generally, this representation offers a parsimonious strategy to assess when the distribution of markups matters for inflation. It requires only a small number of objects constructed from firm-level markups and sales, together with the demand-curvature parameter, rather than a full structural estimation of nominal rigidity. It therefore allows us to assess whether, and in which direction, the location of markup movements across firms affects their aggregate inflation implications.

5 Conclusion

Using monthly firm-level microdata for Chile, including information on prices and quantities, we find that aggregate markup cyclicality is primarily an intensive-margin phenomenon, and that the firms driving it are those at the top of the persistent markup distribution. In our model, these firms have lower pass-through and flatter profit functions, so they adjust prices less and absorb a larger share of aggregate shocks through their markups. Aggregating these heterogeneous pricing responses gives an average markup-gap term and a distributional term. As a result, the inflationary implication of a given average markup gap depends on where that gap is located across firms: when markup gaps are concentrated among high-markup incumbents with flatter pricing slopes, the distributional term attenuates the inflation response associated with the average markup-gap term.

References

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2015): "Identification properties of recent production function estimators," *Econometrica*, 83, 2411–2451.

ANDERSON, E., S. REBELO, AND A. WONG (2018): "Markups Across Space and Time," Working Paper 24434, National Bureau of Economic Research.

ARUOBA, B., A. FERNÁNDEZ, D. GUZMÁN, E. PASTÉN, AND F. SAFFIE (2021): "Monetary Policy Surprises in Chile: Measurement & Real Effects," Working Papers Central Bank of Chile 921, Central Bank of Chile.

BAQAEE, D. R., E. FARHI, AND K. SANGANI (2024): "The supply-side effects of monetary policy," *Journal of Political Economy*, 132, 1065–1112.

BILS, M., P. J. KLENOW, AND B. A. MALIN (2018): "Resurrecting the role of the product market wedge in recessions," *American Economic Review*, 108, 1118–1146.

BOND, S., A. HASHEMI, G. KAPLAN, AND P. ZOCH (2021): "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data," *Journal of Monetary Economics*, 121, 1–14.

BURSTEIN, A., V. M. CARVALHO, AND B. GRASSI (2020): "Bottom-up Markup Fluctuations," Working Paper 27958, National Bureau of Economic Research.

CHIAVARI, A., D. SMIRNOV, AND M. MORAZZONI (2021): "Heterogeneous markups cyclicality and monetary policy," Tech. rep., Working paper.

De Loecker, J. and F. Warzynski (2012): "Markups and firm-level export status," *American economic review*, 102, 2437–71.

Edmond, C., V. Midrigan, and D. Y. Xu (2023): "How costly are markups?" *Journal of Political Economy*, 131, 1619–1675.

Galí, J. (2015): *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press.

Gilchrist, S. and E. Zakrajšek (2012): "Credit spreads and business cycle fluctuations," *American economic review*, 102, 1692–1720.

Hall, R. E. (1988): "The relation between price and marginal cost in US industry," *Journal of political Economy*, 96, 921–947.

Höynck, C., M. Li, and D. Zhang (2023): "The (Unequal) Rise of Market Power, Nominal Rigidities, and Monetary Non-Neutrality," .

Kimball, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 27, 1241–1277.

Klenow, P. J. and J. L. Willis (2016): "Real rigidities and nominal price changes," *Economica*, 83, 443–472.

Kouvavas, O., C. Osbat, T. Reinelt, and I. Vansteenkiste (2021): "Markups and inflation cyclicality in the euro area," .

Meier, M. and T. Reinelt (2024): "Monetary policy, markup dispersion, and aggregate tfp," *Review of Economics and Statistics*, 106, 1012–1027.

Nekarda, C. J. and V. A. Ramey (2020): "The cyclical behavior of the price-cost markup," *Journal of Money, Credit and Banking*, 52, 319–353.

Rotemberg, J. J. (1982): "Sticky prices in the United States," *Journal of political economy*, 90, 1187–1211.

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