The (Unequal) Rise of Market Power, Nominal Rigidities, and Monetary Non-Neutrality*

Christian Höynck † Minghao Li ‡ Donghai Zhang §

August 2026

Abstract

Market power has risen substantially and unevenly across U.S. sectors. We study how this change affects monetary non-neutrality. In a menu-cost model, higher desired markups flatten firms' profit losses, widen inaction regions, and reduce the frequency of price adjustment, thereby strengthening the real effects of monetary shocks. Empirically, contractionary monetary shocks have larger output effects in high-markup states, and higher markups are associated with lower frequencies of both price increases and decreases, even after controlling for inflation. A quantitative model disciplined by micro-price evidence implies that monetary non-neutrality rose by 42% between 1980 and 2016, with changing markups accounting for 75.9% of the increase. About 90% of the markup effect operates through the endogenous decline in price-adjustment frequency.

Keywords: Market power, Monetary non-neutrality, Menu cost model, Heterogeneous market power, Multi-sector model
JEL Classification: E31, E32, E52, E58

*An earlier version of this paper has been circulated under the title "The Markup Elasticity of Monetary Non-Neutrality". We thank Jinhui Bai, Alessandro Ferrari, John Grigsby, Chenyu Hou (Discussant), Francesco Lippi, Matthias Meier, Michael Weber, Shengliang Ou, Jingyi Zhang (Discussant), as well as conference participants at China International Conference in Macroeconomics 2022, The Macro/International Economics Workshop at CHUK (Shenzhen), Workshop at Shandong University, and 2023 CEBRA Annual Meeting for their suggestions and discussions. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Bank of Italy. All remaining errors are our own.

†European Central Bank, chhoeynck@gmail.com
‡National School of Development, Peking University, mhli@nsd.pku.edu.cn
§National University of Singapore and University of Bonn, donghai.d.zhang@gmail.com

1. Introduction

Does an economy with greater market power respond differently to monetary policy? This question has become more relevant as U.S. markups have risen over the past four decades (De Loecker et al., 2020; Edmond et al., 2023). At the same time, a large literature has shown that the real effects of monetary policy depend critically on how frequently firms adjust prices. These observations suggest a connection that has received less attention: changes in market power may themselves alter firms' incentives to adjust prices and thereby change the degree of nominal rigidity in the economy.

This paper studies this connection. Our central mechanism is simple. When price adjustment is costly, a higher desired markup makes profits less sensitive to deviations of the current price from its frictionless optimum. Firms are therefore willing to tolerate larger price gaps before paying the cost of adjustment. Their inaction regions widen, prices adjust less frequently, and nominal disturbances are absorbed more slowly by prices. Monetary shocks consequently have larger real effects. The mechanism exhibits diminishing returns: a given markup increase has a larger effect when the initial markup is low. As a result, monetary transmission depends not only on the aggregate level of market power, but also on how changes in market power are distributed across sectors.

We combine empirical evidence, a tractable menu-cost model, and a quantitative multi-sector model to evaluate this mechanism. The empirical analysis begins by documenting the evolution of market power in the United States. We estimate firm-level markups from Compustat using the production-based approach of Edmond et al. (2023), building on De Loecker and Warzynski (2012), and aggregate them to the sector level. Aggregate markups increase substantially between 1980 and 2016, but this increase is far from uniform across sectors. Some sectors experience large and persistent increases in market power, while markups in others change much less.

We next provide three pieces of empirical evidence consistent with the proposed mechanism. First, the real effects of monetary policy vary with the prevailing markup state. Using local projections and externally identified monetary policy shocks following Gertler and Karadi (2015), we find that contractionary monetary policy shocks are associated with larger declines in real GDP when aggregate markups are higher. Second, higher markups are associated with a lower frequency of price adjustment. Using the historical micro-price series of Nakamura et al. (2018), we find that this relationship remains strong after controlling for inflation and appears for both price increases and price decreases. The latter distinction is informative because inflation generates a markedly asymmetric pattern: higher inflation is strongly associated with more frequent price increases but has little relationship with the frequency of price decreases. Markups, by contrast, are negatively associated with adjustment frequencies in both directions, as predicted by the menu-cost mechanism.

Third, the frequency of price adjustment helps account for the markup dependence of monetary transmission. Conditional on both markups and inflation, contractionary monetary shocks have larger real effects when prices are adjusted less frequently. Adding the interaction between monetary shocks and the frequency of price adjustment also systematically attenuates the estimated interaction

between monetary shocks and markups. Because markups and pricing frequencies are endogenous equilibrium outcomes, we interpret these estimates as conditional correlations. They distinguish the proposed mechanism from the competing inflation channel and motivate the theoretical analysis.[1]

We formalize the mechanism in a tractable continuous-time menu-cost model. Firms face CES demand, so their desired markup is determined by the elasticity of substitution. A key object for the firm's pricing decision is the profit loss from allowing its posted price to deviate from the frictionless optimum. Under CES demand, this loss becomes locally flatter as the desired markup rises. Solving the firm's pricing problem yields an explicit mapping from market power to nominal rigidity: the width of the optimal (s, S) inaction region is increasing and concave in the desired markup, while the frequency of price adjustment is decreasing and convex. Equivalently, expected price duration is increasing and concave in the markup.

We then introduce a marginal monetary shock using the framework of Alvarez and Lippi (2022). The real response to the shock can be written directly as a function of the steady-state frequency of price adjustment. Combining this result with the pricing problem delivers a transparent markup–frequency–monetary-transmission channel: higher markups reduce adjustment frequency and increase the real effects of monetary shocks at every positive horizon. Cumulative monetary non-neutrality is increasing but concave in the desired markup. The concavity result is particularly useful for our quantitative exercise because it implies that the cross-sectoral distribution of markup changes matters even conditional on the aggregate markup.

To quantify the mechanism, we develop a multi-sector menu-cost model with heterogeneous markups and sector-specific price-setting parameters. We feed into the model the observed paths of sectoral markups and expenditure weights from 1980 to 2016. The remaining sector-specific parameters governing idiosyncratic shocks and price adjustment are calibrated to micro-price evidence from Nakamura and Steinsson (2008), matching the pricing moments over 1998–2005. These pricing parameters are then held fixed over time. Apart from changes in sectoral composition, the observed evolution of markups generates all time variation in endogenous price-adjustment behavior.

Between 1980 and 2016, model-implied monetary non-neutrality, measured by the cumulative real response to a nominal-spending shock, increases by approximately 42 percent. An exact decomposition shows that changes in sectoral markups account for 75.9 percent of this increase, making changes in market power the dominant force in the model. The model also generates a pronounced secular decline in the frequency of price adjustment that is broadly consistent with the decline observed in the micro-price data, even though this time-series pattern is not targeted in the calibration.

The unequal rise in market power matters separately from its aggregate increase. Because monetary non-neutrality is concave in the markup, concentrating markup growth in sectors that already experience large increases produces diminishing marginal effects. We illustrate this point with a

[1] More specifically, inflation may drive the relationship between markups and frequency of price adjustment. Lower inflation is associated with less frequent price adjustment (Ball et al., 1988) and has also been linked to higher markups (Bénabou, 1992; Banerjee and Russell, 2001).

counterfactual that preserves the aggregate markup in every year but equalizes markup changes across sectors. Under the observed heterogeneous evolution of markups, monetary non-neutrality increases by 42 percent. Under the equal-change counterfactual, it would instead increase by 58 percent. Thus, the uneven sectoral rise in market power has dampened the increase in monetary non-neutrality relative to an economy in which the same aggregate increase in market power were distributed more uniformly across sectors.

Finally, we quantify why higher markups increase monetary non-neutrality. In a Calvo counterpart in which sectoral adjustment frequencies are held fixed, monetary non-neutrality is nearly unchanged as markups rise. The 42 percent increase in the menu-cost model is therefore driven by the endogenous response of firms' pricing decisions. Consistent with this interpretation, applying the sufficient-statistic decomposition of Alvarez et al. (2022) shows that approximately 90 percent of the markup-induced increase in monetary non-neutrality operates through the decline in the frequency of price adjustment, with changes in the selection of firms into adjustment accounting for the remainder. Examining the underlying pricing policies further shows that the fall in adjustment frequency is driven primarily by the widening of firms' inaction regions rather than by large changes in the distribution of price gaps. An extension with intermediate inputs following Nakamura and Steinsson (2010) substantially increases monetary non-neutrality in levels but leaves the proportional increase associated with rising markups similar to that in the baseline model.

Related Literature. This paper contributes to three strands of the literature. First, it relates to a growing literature studying how market power shapes the transmission of monetary policy. Recent work has emphasized a *real-rigidity* channel, whereby strategic complementarities in price setting cause firms to adjust prices by less than they would in a frictionless benchmark; see, among others, Mongey (2021), Wang and Werning (2022), and Ueda (2023). We emphasize a distinct *nominal-rigidity* channel. In our setting, greater market power flattens firms' profit losses around their optimal prices, widens their inaction regions, and lowers the frequency of price adjustment. Market power therefore changes not only the size of price adjustments conditional on adjustment, but also firms' endogenous propensity to adjust prices in the first place.[2] Relatedly, Chiavari et al. (2021) study heterogeneous firm-level markup responses to monetary policy within sectors. Our focus instead is on how persistent differences in sectoral market power and their heterogeneous evolution affect aggregate monetary transmission. Meier and Reinelt (2024) emphasize the reverse direction of causality: firms facing more rigid prices optimally choose higher markups because of a precautionary price-setting motive. We instead study how higher desired, or steady-state, markups alter firms' incentives to adjust prices and thereby affect nominal rigidity.

Second, the paper contributes to the state-dependent pricing literature initiated by Barro (1972) and developed, among others, by Dotsey et al. (1999), Hobijn et al. (2006), Golosov and Lucas (2007),

[2] Mongey (2021) studies an environment with both nominal and real rigidities and emphasizes the role of strategic complementarities in amplifying monetary non-neutrality. Our contribution is to isolate and quantify how market power itself changes the endogenous degree of nominal rigidity through firms' adjustment frequencies.

Gertler and Leahy (2008), Midrigan (2011), Vavra (2014), Alvarez et al. (2016), Klenow and Willis (2016), Karadi and Reiff (2019), Karadi et al. (2022), Alvarez et al. (2022), Alvarez and Lippi (2022), Auclert et al. (2022), and Alvarez et al. (2023). This literature emphasizes adjustment costs, profit curvature, idiosyncratic volatility, and selection as key determinants of monetary non-neutrality. We highlight that one of these objects—profit curvature, and hence the endogenous frequency of price adjustment—varies systematically with market power. Our tractable model provides an explicit mapping from desired markups to profit curvature, inaction regions, adjustment frequencies, and monetary non-neutrality. This mechanism is absent from time-dependent models with an exogenous adjustment hazard: consistent with this distinction, our Calvo counterfactual shows that holding adjustment frequencies fixed largely shuts down the effect of rising markups on monetary non-neutrality. More broadly, our results imply that nominal rigidity need not be a fixed structural feature of the economy, but can evolve endogenously with market structure.

Finally, the paper contributes to the literature emphasizing sectoral heterogeneity in price rigidity as a determinant of aggregate monetary non-neutrality; see, for example, Carvalho (2006), Nakamura and Steinsson (2010), Carvalho and Schwartzman (2015), Gautier and Bihan (2018), Carvalho et al. (2020), Pasten et al. (2020), Alvarez et al. (2022), Lan et al. (2026) and Li (2026). Existing work typically takes sectoral differences in price rigidity as primitives or disciplines them directly using micro-price evidence. We show that heterogeneous nominal rigidities can instead arise endogenously from heterogeneity in market power. Moreover, because the relationship between market power and monetary non-neutrality is nonlinear, the cross-sectoral distribution of markup changes matters even conditional on the aggregate markup. This feature is central to our finding that the unequal rise in U.S. sectoral markups has dampened the increase in monetary non-neutrality relative to a counterfactual in which the same aggregate markup increase is distributed uniformly across sectors.

Outline. The remainder of the paper is organized as follows. Section 2 presents the data and empirical evidence. Section 3 develops the tractable menu-cost model and characterizes the link between markups, price adjustment, and monetary non-neutrality. Section 4 introduces the quantitative multi-sector model. Section 5 presents the calibration, quantitative results, counterfactuals, and mechanism exercises. The final section concludes.

2. Data and Motivating Facts

This section presents the data and empirical evidence that motivate the paper's theoretical and quantitative analysis. The analysis is organized around three predictions of the menu-cost mechanism: monetary transmission is stronger when markups are higher; higher markups are associated with less frequent price increases and decreases; and pricing frequency helps account for markup-dependent monetary transmission.

The empirical analysis documents state-dependent monetary transmission and examines whether

the observed patterns are consistent with these predictions. Because markups, inflation, and pricing frequencies are endogenous equilibrium outcomes, we interpret the estimated coefficients as conditional correlations. The tractable model in Section 3 formalizes the link between markups, firms' price-adjustment decisions, and monetary non-neutrality. The quantitative model in Section 4 then assesses the mechanism quantitatively and permits counterfactual policy analysis.

The remainder of this section proceeds as follows. Section 2.1 states the three empirical hypotheses. Section 2.2 describes the data and measurement, and Section 2.3 documents the evolution of aggregate and sectoral markups. Sections 2.4–2.6 test the three hypotheses in turn.

2.1 Empirical Hypotheses

The three hypotheses follow from a simple menu-cost mechanism. When price adjustment is costly, a higher desired markup makes profits less sensitive to deviations of the current price from its frictionless optimum. Firms consequently tolerate larger price gaps and adjust their prices less frequently, so a larger fraction of a monetary policy shock is transmitted to real activity rather than prices.

2.1.1 Hypothesis 1: Monetary transmission is stronger when markups are higher

Let $\epsilon_t^m > 0$ denote a contractionary monetary policy shock and define the response of real GDP at horizon h, conditional on the markup state μ, as

$$R_h(\mu) \equiv \frac{\partial y_{t+h}}{\partial \epsilon_t^m}.$$

Because a contractionary monetary policy shock reduces real activity, $R_h(\mu) < 0$. Hypothesis 1 predicts that the real effects of monetary policy become stronger as markups rise:

$$\frac{\partial R_h(\mu)}{\partial \mu} < 0.$$

We test this prediction using local projections that allow the output response to externally identified monetary policy shocks to vary with the aggregate markup prevailing at the time of the shock.

Inflation may generate a similar empirical relationship. Lower inflation is associated with less frequent price adjustment (Ball et al., 1988) and has also been linked to higher markups (Bénabou, 1992; Banerjee and Russell, 2001). Markup-dependent monetary transmission alone therefore does not isolate the proposed mechanism. Hypotheses 2 and 3 use additional predictions of the menu-cost mechanism to distinguish it from this competing inflation channel.

2.1.2 Hypothesis 2: Higher markups are associated with less frequent price increases and decreases

Hypothesis 2 concerns the price-adjustment margin directly. A higher desired markup widens the firm's inaction region on both sides of its optimal price and therefore reduces the frequency of both price increases and price decreases. This symmetric prediction distinguishes the markup mechanism from changes in price-setting behavior associated with inflation. Nakamura et al. (2018) show that lower inflation primarily reduces the frequency of price increases while having little effect on the frequency of price decreases. As discussed by Nakamura and Steinsson (2008), this asymmetry arises naturally in menu-cost models with positive inflation. Examining price increases and decreases separately therefore helps distinguish the markup mechanism from variation in inflation.

2.1.3 Hypothesis 3: Pricing frequency helps account for markup-dependent monetary transmission

Hypothesis 3 links the first two predictions. If endogenous price adjustment is the channel connecting markups to monetary transmission, then pricing frequency should help account for the stronger real effects of monetary policy in high-markup states. We evaluate this implication in two complementary ways. First, we ask whether controlling for the frequency of price adjustment attenuates the interaction between markups and monetary policy shocks. Second, we test whether, conditional on markups and inflation, lower pricing frequency is associated with stronger real responses to monetary policy shocks. Both specifications control for inflation to distinguish the nominal-rigidity mechanism from the competing inflation-based explanation.

2.2 Data and Measurement

Our empirical analysis combines three sources of data. Firm-level balance-sheet data provide the markup measures, CPI microdata provide the frequencies of price adjustment, and externally identified innovations measure monetary policy shocks. This subsection describes the construction and role of each variable. Appendix A provides additional details.

Markups. We estimate firm-level markups from quarterly Compustat data for publicly traded U.S. firms over 1980–2016. The estimation uses the production-based approach of Edmond et al. (2023), building on the framework of De Loecker and Warzynski (2012). We then aggregate firm-level markups within one-digit NAICS sectors using cost weights. Appendix A.1 describes the markup estimation, variable construction, sample selection, and aggregation procedure.

Smoothed Markup Measures. Our analysis focuses on persistent variation in market power rather than high-frequency fluctuations in measured markups. We therefore smooth transitory movements and measurement noise using centered seven-year moving averages. For the descriptive analysis of sectoral markup dynamics and the quantitative exercises, we define the smoothed markup of sector

k in year t, denoted by $\mu^{ss}_{k,t}$, as the centered seven-year moving average of the annual sectoral markup $\mu_{k,t}$. We apply the same procedure to the aggregate markup series used in the time-series analysis.[3]

Frequency of Price Adjustment. We use two complementary sets of measures from the Bureau of Labor Statistics Consumer Price Index Research Database. For the quantitative calibration in Section 4, we use the micro-pricing moments constructed by Nakamura and Steinsson (2008). These data cover 270 Entry Level Items in the non-shelter component of the CPI over 1998–2005 and report detailed moments of price adjustment across goods and services. We use the moments to discipline the sectoral pricing parameters in the menu-cost model.[4]

For the time-series analysis, we use the historical price-adjustment measures constructed by Nakamura et al. (2018) from the same CPI microdata. These series span 1978–2014 and measure how frequently consumer prices change over time. We denote the aggregate median frequency of price adjustment by FPA_t, and the corresponding frequencies of price increases and decreases by FPA^+_t and FPA^-_t, respectively. Separating these two margins allows us to distinguish the symmetric prediction of the markup mechanism from the asymmetric relationship between inflation and price adjustment discussed in Section 2.5.

Monetary Policy Shocks. Our baseline monetary policy shock series follows the proxy structural vector autoregression approach of Gertler and Karadi (2015), extended through 2015Q4. The external instrument is based on high-frequency changes in interest-rate futures around Federal Open Market Committee announcements. We normalize the shock so that a positive value corresponds to a contractionary monetary policy innovation.

High-frequency monetary surprises may contain information about the Federal Reserve's private forecasts (Nakamura and Steinsson, 2018; Jarociński and Karadi, 2020; Miranda-Agrippino and Ricco, 2021; Zhang, 2022). As a robustness check, we use the monetary policy shocks of Miranda-Agrippino and Ricco (2021), which are orthogonal to the Federal Reserve's internal forecasts.

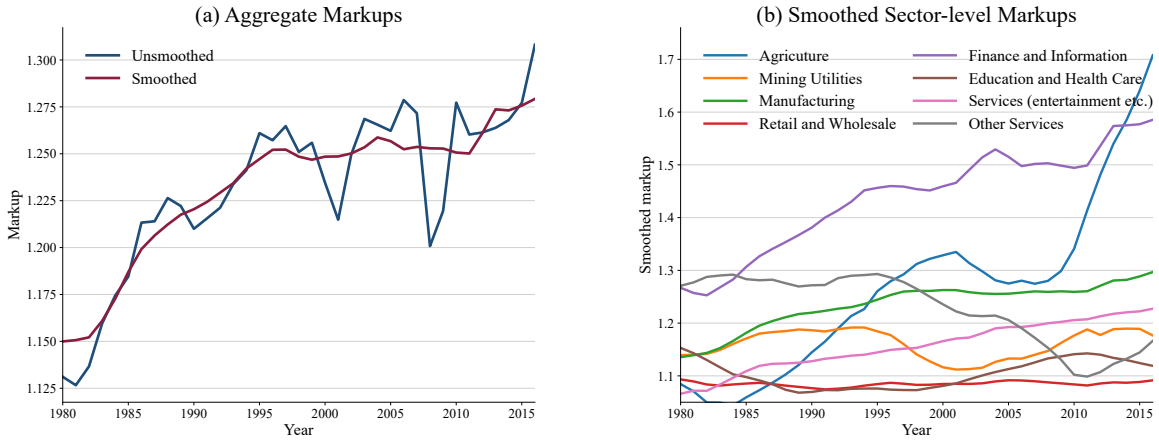
2.3 Evolution of Markups

Figure 1 documents two features of U.S. markups over 1980–2016. The aggregate markup rises substantially, while sectoral markups differ in both their levels and their evolution. The aggregate increase therefore does not reflect a uniform rise in market power across sectors.

[3]The results are stable across alternative smoothing windows.

[4]The CPI Research Database is a standard source for studies of micro price adjustment and has also been used by, among others, Bils and Klenow (2004), Bils et al. (2012), and Nakamura et al. (2018).

Figure 1: Evolution of Aggregate and Sectoral Markups in the United States



Notes: Panel (a) plots the smoothed aggregate markup, and Panel (b) plots smoothed markups for one-digit NAICS sectors. The markups are constructed from Compustat over 1980–2016 using the procedure described in Section 2.2. All smoothed series are centered seven-year moving averages. Appendix A.1 provides details on markup estimation and aggregation.

Fact 1: The Aggregate Markup Has Increased. Panel (a) of Figure 1 shows that the smoothed aggregate markup rises from approximately 1.15 in 1980 to about 1.30 in 2016, an increase of nearly 15 percent. This pattern is consistent with the evidence in De Loecker et al. (2020).

This secular increase motivates our analysis. If firms' desired markups affect their incentives to adjust prices, a persistent rise in market power can alter nominal rigidity and, in turn, the real effects of monetary policy. The aggregate series, however, masks substantial differences in the evolution of market power across sectors.

Fact 2: Markup Levels and Their Changes Differ Substantially across Sectors. Panel (b) of Figure 1 shows substantial and persistent differences in markup levels across one-digit NAICS sectors. In 2000, for example, sectoral markups range from approximately 1 in the lowest-markup sector to about 1.5 in the highest-markup sector.

Sectoral markups also follow different paths over time. The aggregate increase is concentrated in a subset of sectors: agriculture, finance and information, and services experience particularly pronounced increases in their smoothed markups. Markups in most remaining sectors are comparatively stable, while one sector experiences a decline.

This cross-sector heterogeneity is central to the quantitative analysis. If higher markups are associated with less frequent price adjustment, the uneven evolution of sectoral markups changes the distribution of nominal rigidity across the economy.

2.4 Hypothesis 1: Markup-Dependent Monetary Transmission

We test Hypothesis 1 by allowing the response of real GDP to an identified monetary policy shock to vary with the aggregate markup prevailing when the shock occurs. The hypothesis predicts a larger decline in real activity following a contractionary shock when the aggregate markup is higher.

Empirical Specification. We estimate quarterly state-dependent impulse responses using the local-projection method of Jordà (2005). For each horizon $h = 0, 1, \ldots, 24$, we estimate

$$y_{t+h} = \alpha_h + \beta_1^h \epsilon_t^m + \beta_2^h \epsilon_t^m \times \mu_t^{ss} + \mathbf{\Gamma}_h' \mathbf{X}_t + u_{t+h}, \tag{2.1}$$

where y_t denotes log real GDP, ϵ_t^m is the identified monetary policy shock, and μ_t^{ss} is the smoothed aggregate markup. A positive value of ϵ_t^m corresponds to a contractionary monetary policy innovation. The interaction term $\epsilon_t^m \times \mu_t^{ss}$ captures how the response to the same shock varies with the markup state. The vector $\mathbf{X}_t$ contains four lags of log real GDP, the log GDP deflator, the two-year Treasury yield, and a linear time trend.

Markup-Dependent Responses. Conditional on a markup state μ^{ss}, the response of log real GDP at horizon h to a unit monetary policy shock is

$$MN^h(\mu^{ss}) \equiv \frac{\partial y_{t+h}}{\partial \epsilon_t^m} = \beta_1^h + \beta_2^h \mu^{ss}.$$

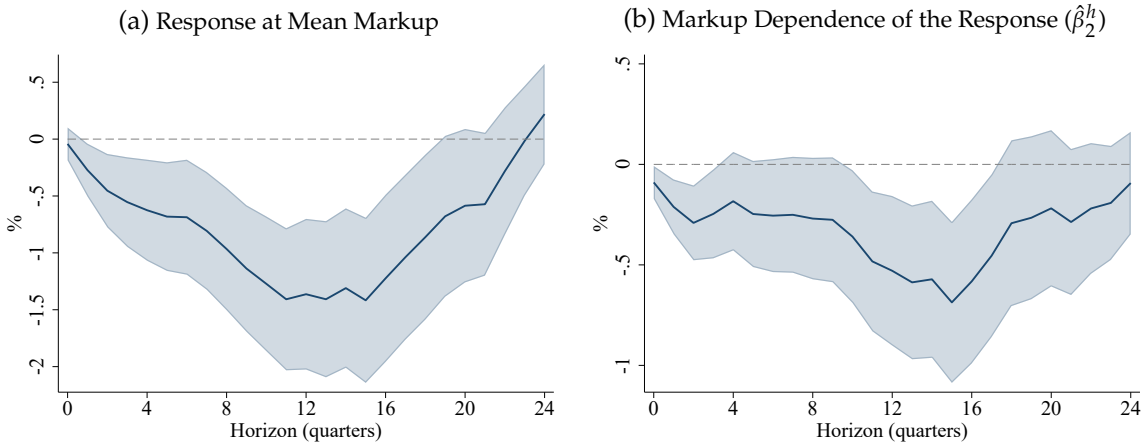
We refer to $MN^h(\mu^{ss})$ as monetary non-neutrality at horizon h. Because a positive monetary policy shock is contractionary, Hypothesis 1 corresponds to

$$\beta_2^h < 0.$$

A negative β_2^h means that the response becomes more negative as the markup rises.

Figure 2 reports the estimates of equation (2.1). Panel (a) plots the response evaluated at the sample mean of the smoothed markup, $\overline{\mu^{ss}}$: $\widehat{MN}^h(\overline{\mu^{ss}}) = \hat{\beta}_1^h + \hat{\beta}_2^h \overline{\mu^{ss}}$. Panel (b) plots $\hat{\beta}_2^h$, the change in the response associated with a 1% increase in the markup. The shaded areas report 90 percent confidence intervals based on Newey–West standard errors.

Figure 2: Markup-Dependent Effects of Contractionary Monetary Policy Shocks



Notes: This figure reports estimates of the markup-dependent response of real GDP to a contractionary monetary policy shock. Panel (a) plots $\hat{\beta}_1^h + \hat{\beta}_2^h \overline{\mu^{ss}}$, the estimated response evaluated at the sample mean of the smoothed markup. Panel (b) plots $\hat{\beta}_2^h$, which measures how the response varies with the markup state. The coefficients are estimated from equation (2.1). The monetary policy shock ϵ_t^m follows the identification strategy of Gertler and Karadi (2015), with positive values corresponding to contractionary innovations. The shaded areas indicate 90 percent confidence intervals constructed using Newey–West standard errors.

Panel (a) shows that contractionary monetary policy shocks reduce real GDP on average. Panel (b) shows that $\hat{\beta}_2^h$ is negative over the horizons at which this contractionary response is most pronounced. The same contractionary shock is therefore associated with a larger decline in real activity when the aggregate markup is higher.[5]

These results are consistent with Hypothesis 1: higher markups are associated with greater monetary non-neutrality. The next two subsections evaluate whether price-adjustment behavior is consistent with the proposed channel. They also account for inflation, which may covary with both markups and pricing frequency and generate a similar state-dependent relationship.[6]

2.5 Hypothesis 2: Markups, Inflation, and the Frequency of Price Adjustment

We test Hypothesis 2 by relating the aggregate markup to the frequencies of price increases and decreases. The directional frequencies distinguish the markup mechanism from the inflation channel discussed in Section 2.1.1: inflation predicts an asymmetric relationship with price adjustment, whereas higher markups predict lower adjustment frequencies in both directions.

[5] Appendix Figure D.1 shows that the results are similar using the alternative monetary policy shocks of Miranda-Agrippino and Ricco (2021).

[6] Appendix Figure D.2 plots the smoothed aggregate markup, inflation, and the aggregate frequency of price adjustment over 1978–2014.

Empirical Specification. To estimate the relationship between markups and pricing frequency conditional on inflation, we use

$$y_t = \alpha + \gamma_\mu \log\left(\mu_t^{ss}\right) + \gamma_\pi \pi_t + \mathbf{\Gamma}'\mathbf{X}_t + u_t, \tag{2.2}$$

where the dependent variable y_t is, in turn, the total frequency of price adjustment, FPA_t, the frequency of price increases, FPA_t^+, or the frequency of price decreases, FPA_t^-. The variable μ_t^{ss} is the smoothed aggregate markup, π_t is CPI inflation, and $\mathbf{X}_t$ contains macroeconomic controls.

The two mechanisms imply different coefficient patterns. Higher inflation causes posted prices to drift downward more rapidly relative to firms' desired prices, so the inflation channel predicts

$$\gamma_\pi^{FPA} > 0, \qquad \gamma_\pi^{FPA+} > 0, \qquad \gamma_\pi^{FPA-} \approx 0.$$

Higher inflation therefore raises total adjustment frequency primarily through more frequent price increases, with little effect on the frequency of price decreases. This asymmetry is consistent with the evidence in Nakamura et al. (2018) and with menu-cost models featuring positive trend inflation (Nakamura and Steinsson, 2008).

The markup mechanism instead operates through the width of firms' inaction regions. A higher desired markup widens the inaction region in both directions and predicts

$$\gamma_\mu^{FPA} < 0, \qquad \gamma_\mu^{FPA+} < 0, \qquad \gamma_\mu^{FPA-} < 0.$$

The behavior of FPA_t^- is particularly informative because inflation should have little effect on the frequency of price decreases, whereas higher markups should reduce it.

Table 1 reports the results. Columns (1), (3), and (5) omit inflation and relate each measure of pricing frequency to the smoothed aggregate markup and the macroeconomic controls. The markup coefficient is negative and statistically significant in all three specifications. Higher markups are associated with lower frequencies of total price adjustment, price increases, and price decreases.

Columns (2), (4), and (6) add CPI inflation. The coefficients match the asymmetric prediction of the inflation channel. Inflation is positively and significantly associated with total adjustment frequency and, especially, the frequency of price increases. Its association with the frequency of price decreases is economically small and statistically insignificant.

The markup coefficient remains negative and statistically significant for all three measures after inflation is included. For total FPA, the coefficient changes from -0.25 to -0.22. For price increases, it becomes less negative, changing from -0.73 to -0.53, while the coefficient for price decreases remains -0.20.

Interpretation. The directional results are consistent with Hypothesis 2. Inflation exhibits the predicted asymmetry, while higher markups are associated with less frequent adjustment in both directions. Equation (2.2) is therefore a mechanism test based on the joint behavior of markups,

Table 1: Markups, Inflation, and Frequency of Price Adjustment

	Dependent Variable					
	FPA_t		FPA_t^+		FPA_t^-	
	(1)	(2)	(3)	(4)	(5)	(6)
Markups	-0.25***	-0.22***	-0.73***	-0.53***	-0.20***	-0.20***
	(0.043)	(0.045)	(0.15)	(0.13)	(0.037)	(0.040)
Inflation		0.12***		0.62***		-0.0072
		(0.044)		(0.21)		(0.041)
Macro controls	✓	✓	✓	✓	✓	✓
Observations	112	112	112	112	112	112
Adjusted R^2	0.93	0.93	0.58	0.64	0.62	0.62

Notes: This table reports estimates of equation (2.2). The dependent variables are the total frequency of price adjustment, FPA_t, the frequency of price increases, FPA_t^+, and the frequency of price decreases, FPA_t^-. Columns (1), (3), and (5) omit inflation, while columns (2), (4), and (6) additionally control for CPI inflation. Markups are measured by the log of the smoothed aggregate markup. Newey–West standard errors are reported in parentheses. $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

inflation, and directional pricing frequencies. The next subsection asks whether pricing frequency also helps account for the markup dependence of monetary transmission documented in Section 2.4.

2.6 Hypothesis 3: The Relevance of the FPA Channel

We test Hypothesis 3 in two complementary ways. First, we ask whether monetary transmission varies with the frequency of price adjustment after conditioning on markups and inflation. Second, we examine whether adding the FPA interaction attenuates the markup dependence documented in Section 2.4.

Empirical Specification. We begin with a local-projection specification in which the response of real GDP depends on the markup state and lagged inflation:

$$y_{t+h} = \alpha_h + \beta_{0h}\epsilon_t^m + \beta_{\mu h}\epsilon_t^m \times \mu_t^{ss} + \beta_{\pi h}\epsilon_t^m \times \pi_{t-1} + \mathbf{\Gamma}_h'\mathbf{X}_t + u_{t+h}, \tag{2.3}$$

where ϵ_t^m denotes a contractionary monetary policy shock, μ_t^{ss} is the smoothed aggregate markup, π_{t-1} is lagged inflation, and $\mathbf{X}_t$ contains the baseline macroeconomic controls described in Section 2.4. We refer to equation (2.3) as the markup–inflation specification.

We then augment this specification by allowing monetary transmission to vary with the lagged frequency of price adjustment:

$$y_{t+h} = \alpha_h + \beta_{0h}\epsilon_t^m + \beta_{\mu h}\epsilon_t^m \times \mu_t^{ss} + \beta_{\pi h}\epsilon_t^m \times \pi_{t-1} + \beta_{Fh}\epsilon_t^m \times FPA_{t-1} + \mathbf{\Gamma}_h'\mathbf{X}_t + u_{t+h}. \tag{2.4}$$

We refer to equation (2.4) as the FPA-augmented specification. We estimate both specifications over the same sample with the same baseline controls and lag structure.

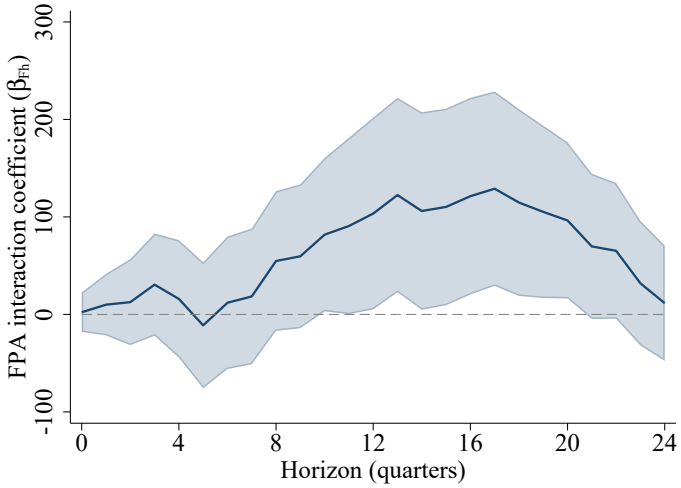
The menu-cost mechanism predicts

$$\beta_{Fh} > 0. \tag{2.5}$$

Because a positive ϵ_t^m denotes a contractionary shock, a positive β_{Fh} means that a higher FPA makes the output response less negative. Equivalently, monetary contractions have larger real effects when prices adjust less frequently.

FPA-Dependent Monetary Transmission. Figure 3 plots the estimated FPA interaction coefficient, $\widehat{\beta}_{Fh}$, together with its 90 percent confidence interval.

Figure 3: FPA-Dependent Effects of Contractionary Monetary Policy Shocks



Notes: This figure plots the estimated coefficient $\widehat{\beta}_{Fh}$ on the interaction between the contractionary monetary policy shock of Gertler and Karadi (2015) and lagged FPA in equation (2.4). The shaded area reports the pointwise 90 percent confidence interval based on Newey–West standard errors.

The estimate $\widehat{\beta}_{Fh}$ is positive at nearly all horizons and becomes statistically distinguishable from zero at the 10 percent level from quarters 10 through 20. This medium-horizon pattern is consistent with equation (2.5): conditional on markups and inflation, a higher frequency of price adjustment dampens the decline in output following a contractionary monetary policy shock. Equivalently, monetary contractions have larger real effects when prices adjust less frequently.

Appendix Section A.2 reports complementary partial-R^2 measures of the incremental explanatory power of the FPA interaction. The partial R^2 peaks at quarter 17, when adding the FPA interaction accounts for 4.34 percent of the residual variation under the markup–inflation specification. The adjusted partial R^2 is positive from quarters 8 through 22 and also peaks at quarter 17, at 3.10 percent.

Attenuation of Markup-Dependent Transmission. We next examine how conditioning on inflation and the FPA changes the estimated markup dependence of monetary transmission. If inflation

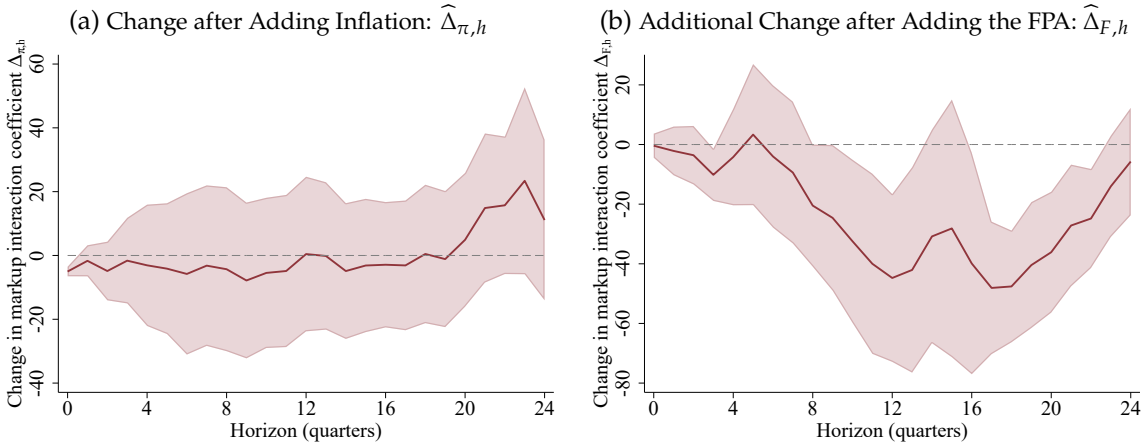
accounts for the original markup interaction, adding the inflation interaction should substantially attenuate the markup coefficient. If pricing frequency helps account for markup-dependent transmission, adding the FPA interaction should move the markup coefficient further toward zero.

To quantify these changes, we define

$$\Delta_{\pi,h} = \widehat{\beta}^{\mu}_{\mu h} - \widehat{\beta}^{\mu,\pi}_{\mu h},$$
$$\Delta_{F,h} = \widehat{\beta}^{\mu,\pi}_{\mu h} - \widehat{\beta}^{\mu,\pi,F}_{\mu h},$$

where the superscripts indicate the state variables included in each specification. Because the estimated markup interaction is negative, a negative value of either attenuation statistic means that the augmented specification yields a less negative markup coefficient. Appendix Section A.3 provides the estimation and inference details.

Figure 4: Attenuation of Markup-Dependent Monetary Transmission



Notes: Panel (a) plots $\widehat{\Delta}_{\pi,h} = \widehat{\beta}^{\mu}_{\mu h} - \widehat{\beta}^{\mu,\pi}_{\mu h}$, the change in the estimated markup interaction after adding the interaction between the monetary policy shock and lagged inflation. Panel (b) plots $\widehat{\Delta}_{F,h} = \widehat{\beta}^{\mu,\pi}_{\mu h} - \widehat{\beta}^{\mu,\pi,F}_{\mu h}$, the additional change after adding the interaction between the monetary policy shock and lagged FPA. Since the markup interaction is negative, a negative value indicates that the augmented specification produces a less negative markup interaction coefficient, moving it toward zero. The shaded areas are pointwise 90 percent confidence intervals constructed from a Bartlett heteroskedasticity-and-autocorrelation-consistent estimator applied to the difference in the two coefficient influence functions.

Adding inflation changes the estimated markup interaction relatively little. The 90 percent confidence interval for $\Delta_{\pi,h}$ contains zero at every horizon except on impact. Variation in inflation therefore does not appear to be the primary explanation for the markup dependence documented in Section 2.4.

Adding the FPA produces a systematic attenuation of the markup interaction. The estimate $\Delta_{F,h}$ is negative at nearly every horizon and statistically different from zero at quarter 3, quarters 9–13, and quarters 16–22. Together with the positive FPA interaction coefficient, this attenuation is consistent with Hypothesis 3: lower pricing frequency is associated with stronger monetary transmission and helps account for its markup dependence.

Taken together, the three empirical findings are consistent with the proposed markup–FPA–monetary-transmission channel. The next section develops a tractable menu-cost model that formalizes this channel and characterizes how higher markups reduce pricing frequency and amplify the real effects of monetary policy.

3. Illustrating the Mechanism: A Tractable Menu-Cost Model

This section develops a tractable menu-cost model that isolates the mechanism linking markups to monetary non-neutrality. The model provides an analytical characterization of the mechanism; Section 4 introduces the model used for quantification. Specifically, it delivers an analytical mapping from markups to firms' pricing incentives and, through the frequency of price adjustment, to the real effects of monetary shocks.

We illustrate the mechanism in two steps. First, a higher desired markup widens the inaction region and reduces the frequency of price adjustment. Second, following Alvarez and Lippi (2022), we study a marginal monetary shock that shifts all firms' desired prices by a common amount. In the canonical menu-cost model, the propagation of this shock can be expressed directly as a function of the steady-state frequency of price adjustment. Combining these results yields a transparent markup–FPA–monetary-transmission channel. Appendix B provides the derivations and proofs.

3.1 Environment

Time is continuous. Consider a sector populated by a continuum of monopolistically competitive firms. We suppress the sector index and compare otherwise identical sectors that differ only in their desired markup. Firm i faces CES demand,

$$Y_{it} = Y_t \left(\frac{P_{it}}{P_t} \right)^{-\theta}, \qquad \theta > 1,$$

where P_{it} is the firm's price, P_t is the sectoral price index, and Y_t is sectoral demand. The desired gross markup is

$$\mu = \frac{\theta}{\theta - 1} > 1.$$

Let W_{it} denote nominal marginal cost. Absent price-adjustment frictions, the firm's desired price is

$$P_{it}^* = \mu W_{it}.$$

Define $p_{it} \equiv \log P_{it}$, $p_{it}^* \equiv \log P_{it}^*$, and the log price gap

$$z_{it} \equiv p_{it} - p_{it}^*.$$

The desired log price follows a driftless Brownian motion,

$$dp^*_{it} = \sigma\, dB_{it}, \qquad \sigma > 0,$$

where B_{it} is a standard Brownian motion. Between adjustments, the posted price is fixed, so

$$dz_{it} = -\sigma\, dB_{it}.$$

A firm can reset its price by paying a fixed adjustment cost $\psi > 0$, measured in units of flow profit at the frictionless optimum. Upon adjustment, the firm resets z_{it} to zero. We hold ψ and σ fixed across sectors so that the comparative statics isolate the role of the desired markup.

The key object for the pricing decision is the profit loss generated by a nonzero price gap. Appendix B.1 shows that, around $z = 0$, the loss relative to frictionless flow profit is

$$\ell(z;\mu) = \frac{1}{2}\kappa(\mu)z^2 + o(z^2), \qquad \kappa(\mu) \equiv \frac{\mu}{(\mu-1)^2}. \tag{3.1}$$

The curvature of normalized profits satisfies

$$\kappa'(\mu) = -\frac{\mu+1}{(\mu-1)^3} < 0.$$

Thus, higher desired markups make normalized profits locally flatter. For a given deviation of the posted price from its frictionless optimum, the associated profit loss is smaller when the markup is higher. This curvature effect is the primitive force behind the results that follow.

3.2 Markups and the Frequency of Price Adjustment

We treat the quadratic approximation in equation (3.1) as the flow-loss function of an auxiliary menu-cost problem. The firm chooses when to pay ψ and reset its price gap to zero. The optimal policy is symmetric: the firm leaves its price unchanged while

$$z \in \left(-\overline{z}(\mu), \overline{z}(\mu)\right)$$

and adjusts when the price gap reaches either boundary.

Proposition 3.1 (Markups and price adjustment)**.** *The optimal adjustment threshold is*

$$\overline{z}(\mu) = \left[\frac{12\psi\sigma^2(\mu-1)^2}{\mu}\right]^{1/4}.$$

The width of the Ss inaction region, $S(\mu) \equiv 2\bar{z}(\mu)$, is increasing and concave in the desired markup:

$$S'(\mu) > 0, \qquad S''(\mu) < 0.$$

The long-run frequency of price adjustment is given by

$$N(\mu) = \frac{\sigma^2}{\bar{z}(\mu)^2} = \frac{\sigma}{\sqrt{12\psi}} \frac{\sqrt{\mu}}{\mu - 1}, \tag{3.2}$$

and is decreasing and convex in the desired markups:

$$N'(\mu) < 0, \qquad N''(\mu) > 0.$$

Equivalently, expected price duration $D(\mu) \equiv 1/N(\mu)$ is increasing and concave in the desired markup.

Appendix B.2 provides the proof.

Proposition 3.1 establishes the first step of the mechanism. Because higher markups flatten normalized profits, firms are willing to tolerate larger deviations of their current price from its frictionless optimum. The inaction region therefore widens and prices adjust less frequently. The effect also exhibits diminishing returns: both the width of the inaction region and expected price duration are concave in the desired markup.

Relation to Barro (1972) and generality of the result. The comparative statics in Proposition 3.1 build on the central insight of Barro (1972) that the gains from price adjustment depend on the curvature of profits around the frictionless optimum.[7] Our result makes explicit how this curvature changes with the desired markup under CES demand.

The concavity result requires more structure than the qualitative prediction that flatter profits reduce adjustment frequency. In particular, it depends on how the curvature of the profit-loss function varies with the markup. Appendices B.3 and B.4 make these conditions explicit for a more general profit-loss function and relate the CES benchmark to the linear-demand environment considered by Barro (1972). This distinction clarifies which parts of Proposition 3.1 reflect the general menu-cost mechanism and which depend on the particular mapping from market power to profit curvature.

3.3 A Marginal Monetary Shock

We next connect the pricing mechanism to monetary transmission using the analytical framework of Alvarez and Lippi (2022). Suppose the economy is initially in the stationary distribution of price gaps. At $t = 0$, an unexpected expansionary monetary shock of size $\delta > 0$ raises all firms' desired log

[7] Gopinath and Itskhoki (2010) similarly emphasize that flatter profits reduce the gains from price adjustment and thereby lower adjustment frequency. Alvarez et al. (2016) characterize price-setting behavior in terms of adjustment costs, profit curvature, and the volatility of the frictionless price.

prices by the same amount:

$$p_{i0}^{*,+} = p_{i0}^{*,-} + \delta.$$

Because posted prices are predetermined on impact, every firm's price gap shifts to the left:

$$z_{i0}^{+} = z_{i0}^{-} - \delta.$$

The stationary distribution of price gaps in the symmetric menu-cost model is triangular. Let $\overline{g}(z;\mu)$ denote its density. Immediately after the shock,

$$g(z, 0; \delta, \mu) = \overline{g}(z + \delta; \mu).$$

Because the stationary density vanishes at the adjustment boundaries, the fraction of firms pushed across a boundary on impact is of order $O(\delta^2)$. To first order in δ, the shock therefore displaces the distribution of price gaps without generating a discrete mass of immediate price adjustments. This is the marginal monetary experiment studied by Alvarez and Lippi (2022).

Write the output response to an expansionary shock as

$$\Delta y(t; \delta, \mu) = \delta \mathcal{R}(t; \mu) + o(\delta),$$

where $\mathcal{R}(t;\mu)$ is the real response per unit of the marginal monetary shock. We adopt the normalization in Alvarez and Lippi (2022), under which an individual firm's first-order contribution to the real response is proportional to the negative of its price gap. This normalization allows us to focus on the propagation of the disturbance rather than on its units.

The eigenfunction representation in Alvarez and Lippi (2022) then gives the following closed-form impulse response.

Proposition 3.2 (Markups and monetary transmission)**.** *Following a marginal expansionary monetary shock, the normalized real output response is*

$$\mathcal{R}(t;\mu) = \sum_{j=0}^{\infty} \frac{8}{\pi^2(2j+1)^2} \exp\left[-\frac{\pi^2}{2}(2j+1)^2 N(\mu) t\right], \qquad t \geq 0, \tag{3.3}$$

where $N(\mu)$ is the steady-state frequency of price adjustment in equation (3.2). The impact response is independent of the markup, $\mathcal{R}(0;\mu) = 1$, while, at every positive horizon,

$$\frac{\partial \mathcal{R}(t;\mu)}{\partial \mu} > 0, \qquad t > 0. \tag{3.4}$$

Thus, higher markups increase the real effect of a marginal monetary shock at every positive horizon.

Appendix B.6 provides the derivation.

For an expansionary shock, equations (3.4) means that output remains further above its steady-state level when the desired markup is higher. For a contractionary shock, symmetry reverses the sign of the response, so output falls by more in absolute value at every positive horizon.

The intuition follows directly from Proposition 3.1. A higher markup widens the inaction region and lowers $N(\mu)$. Firms therefore take longer to adjust their prices and eliminate the price gaps created by the monetary disturbance. More of the nominal shock is consequently absorbed by real quantities for a longer period.

3.4 The FPA Channel

Equation (3.3) makes the role of the frequency of price adjustment particularly transparent. Define

$$\mathscr{R}(t, N) \equiv \sum_{j=0}^{\infty} \frac{8}{\pi^2(2j+1)^2} \exp\left[-\frac{\pi^2}{2}(2j+1)^2 N t\right]. \tag{3.5}$$

Then

$$\mathcal{R}(t;\mu) = \mathscr{R}\left(t, N(\mu)\right),$$

and

$$\frac{\partial \mathscr{R}(t,N)}{\partial N} < 0, \qquad t > 0.$$

Within this tractable environment, the desired markup therefore affects the propagation of a marginal monetary shock entirely through its effect on steady-state price-adjustment frequency. Conditional on N, the markup does not enter equation (3.5). In this precise sense, FPA is a sufficient statistic for markup-dependent monetary propagation in the canonical model.

The same result can be summarized using the cumulative real effect,

$$C(\mu) \equiv \int_0^{\infty} \mathcal{R}(t;\mu)\, dt.$$

Integrating equation (3.3) gives

$$C(\mu) = \frac{1}{6N(\mu)} = \frac{\sqrt{\psi}}{\sqrt{3}\sigma} \frac{\mu - 1}{\sqrt{\mu}}. \tag{3.6}$$

It follows immediately that

$$C'(\mu) > 0, \qquad C''(\mu) < 0.$$

Hence, higher markups increase the cumulative real effect of monetary policy, but at a diminishing rate. Appendix Section B.7 provides derivation details.

Equation (3.6) is also the canonical menu-cost special case of the sufficient-statistic result in Alvarez et al. (2016). In the present model, every nonzero price change has absolute size $\overline{z}(\mu)$, so the kurtosis of the nonzero price-change distribution equals one. Their sufficient statistic—the kurtosis of nonzero price changes divided by six times their frequency—therefore reduces to $1/[6N(\mu)]$.

The analytical model thus delivers the mechanism underlying the empirical analysis. Higher desired markups flatten normalized profits around the frictionless optimum. Firms consequently tolerate larger price gaps and adjust prices less frequently. Following a monetary shock, those price gaps persist for longer, increasing the fraction of the nominal disturbance absorbed by real quantities. In the canonical model, the markup matters for monetary propagation precisely through this endogenous variation in the frequency of price adjustment. Moreover, the cumulative degree of monetary non-neutrality is increasing and concave in the desired markup.

Generality of the mechanism. The mechanism does not rely on monopolistic competition per se. Its essential ingredient is that greater market power reduces the sensitivity of profits to deviations of the firm's price from its frictionless optimum. The CES monopolistic-competition environment provides a particularly transparent mapping from the desired markup to profit curvature, but the same economic force can arise under alternative market structures. Appendix B.8, for example, shows that the relevant profit-curvature mechanism also operates under oligopolistic competition. More generally, whenever weaker competition or greater market concentration flattens profits with respect to price gaps, firms have weaker incentives to pay a fixed cost to adjust prices, generating lower endogenous FPA and stronger monetary non-neutrality.

4. A Multi-sector Model with Heterogeneous Markups

In this section, we develop a quantitative multi-sector menu-cost model. The model incorporates standard ingredients from second-generation menu-cost models, including random menu costs and leptokurtic idiosyncratic productivity shocks, which allow it to reproduce salient features of the empirical distribution of price changes; see, for example, Vavra (2014), Karadi and Reiff (2019), and Alvarez et al. (2021).

Motivated by the empirical evidence in Section 2, our main departure from the standard framework is to allow market power to vary systematically across sectors. Specifically, sectors differ in their within-sector elasticities of substitution and therefore in their desired markups.

4.1 Household

The economy consists of a representative household and a continuum of monopolistically competitive firms operating in K sectors. Firms are indexed by (k, i), where $k = 1, 2, \ldots, K$ denotes the sector and $i \in [0, 1]$ denotes a firm within the sector.

The representative household has preferences

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\gamma}}{1-\gamma} - \kappa \frac{N_t^{1+\lambda}}{1+\lambda} \right], \tag{4.1}$$

where β is the discount factor, γ governs the intertemporal elasticity of substitution, λ is the inverse

of the Frisch elasticity of labor supply, and κ scales the disutility of labor. Aggregate consumption is denoted by C_t, and aggregate labor supply by N_t.

Aggregate consumption is a CES composite of sectoral consumption bundles:

$$C_t = \left(\sum_{k=1}^{K} \omega_k^{\frac{1}{\eta}} C_{k,t}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}, \tag{4.2}$$

where η is the elasticity of substitution across sectors and ω_k denotes the expenditure weight of sector k, with $\sum_{k=1}^{K} \omega_k = 1$.

Within each sector, the sectoral consumption bundle is

$$C_{k,t} = \left(\int_0^1 c_{k,i,t}^{\frac{\theta_k - 1}{\theta_k}} di \right)^{\frac{\theta_k}{\theta_k - 1}}, \tag{4.3}$$

where $\theta_k > 1$ is the elasticity of substitution across differentiated goods within sector k.

The household faces the nominal budget constraint

$$\sum_{k=1}^{K} \int_0^1 p_{k,i,t} c_{k,i,t} di + Q_{t+1} B_{t+1} \leq B_t + W_t N_t + \Pi_t, \tag{4.4}$$

where B_t denotes the payoff from nominal bonds purchased in the previous period, Q_{t+1} is the period-t price of a nominal claim paying one unit in period $t+1$, $p_{k,i,t}$ is the nominal price charged by firm i in sector k, W_t is the nominal wage, and Π_t denotes aggregate firm profits rebated to the household.

Household optimality conditions. The representative household chooses differentiated consumption $\{c_{k,i,t}\}$, aggregate labor supply N_t, and bond holdings B_{t+1} to maximize (4.1) subject to (4.4). The resulting demand for an individual good is

$$c_{k,i,t} = \omega_k \left(\frac{p_{k,i,t}}{P_{k,t}} \right)^{-\theta_k} \left(\frac{P_{k,t}}{P_t} \right)^{-\eta} C_t, \tag{4.5}$$

where the sectoral and aggregate price indices are

$$P_{k,t} = \left(\int_0^1 p_{k,i,t}^{1-\theta_k} di \right)^{\frac{1}{1-\theta_k}}, \tag{4.6}$$

$$P_t = \left(\sum_{k=1}^{K} \omega_k P_{k,t}^{1-\eta} \right)^{\frac{1}{1-\eta}}. \tag{4.7}$$

The household's intertemporal and intratemporal optimality conditions are

$$\frac{P_{t+1}Q_{t+1}}{P_t} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma},$$
$$\frac{W_t}{P_t} = \kappa C_t^{\gamma} N_t^{\lambda}. \tag{4.8}$$

A key feature of the model is that θ_k varies across sectors. Under CES demand and flexible prices, a firm in sector k charges the desired gross markup

$$\mu_k \equiv \frac{\theta_k}{\theta_k - 1}.$$

Thus, a lower θ_k corresponds to greater market power and a higher desired markup. As $\theta_k \to \infty$, the desired markup converges to one and the sector approaches perfect competition.

This mapping provides the link between the empirical and quantitative analyses. We interpret the smoothed sectoral markups measured in the data as empirical counterparts of the sector-specific steady-state markups $\{\mu_k\}_{k=1}^K$ in the model. Changes in the empirical cross-sectional distribution of markups are therefore implemented through changes in the corresponding sectoral demand elasticities $\{\theta_k\}_{k=1}^K$.

4.2 Firms

Firm i in sector k produces differentiated good $y_{k,i,t}$ using labor according to the linear production technology

$$y_{k,i,t} = a_{k,i,t} n_{k,i,t},$$

where $n_{k,i,t}$ denotes labor input and $a_{k,i,t}$ is idiosyncratic productivity.

Firm-level productivity follows a "Poisson-normal" process:

$$\log a_{k,i,t} = \begin{cases} \rho_k^a \log a_{k,i,t-1} + \sigma_k^a \epsilon_{k,i,t}^a, & \text{with probability } \alpha_k, \\ \log a_{k,i,t-1}, & \text{with probability } 1 - \alpha_k, \end{cases} \tag{4.9}$$

where $\epsilon_{k,i,t}^a \sim N(0,1)$ is independent across firms. This process generates infrequent but potentially large innovations to desired prices and therefore helps the model reproduce the leptokurtic distribution of observed price changes. We denote the associated transition probability by

$$\Pr_k (a'|a).$$

Firms face random menu costs. In each period, a firm receives an opportunity to adjust its price at zero cost with probability ϕ_k. Otherwise, changing its nominal price requires paying a fixed cost

$\overline{f}_k$, denominated in units of labor. Hence,

$$f_{k,i,t} = \begin{cases} 0, & \text{with probability } \phi_k, \\ \overline{f}_k, & \text{with probability } 1-\phi_k. \end{cases} \tag{4.10}$$

The specification nests several commonly used models of price adjustment. For example, sufficiently large $\overline{f}_k$ generates price adjustment primarily through the free-adjustment opportunities, approaching a Calvo-type environment (Calvo (1983)), whereas $\phi_k = 0$ yields a standard menu-cost model (Golosov and Lucas (2007)).

Firm optimality conditions.. Firm i in sector k chooses its nominal price to maximize the expected present value of real profits:

$$\max_{\{p_{k,i,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} q_{0,t} \pi_{k,i,t},$$

where $q_{0,t} \equiv q_{0,1} q_{1,2} \cdots q_{t-1,t}$ is the cumulative real stochastic discount factor and $q_{t,t+1} = Q_{t+1} P_{t+1} / P_t$.

Let $\Gamma_{k,t}(p_{-1}, a, f)$ denote the distribution of firms in sector k over their beginning-of-period idiosyncratic states, consisting of the previous nominal price p_{-1}, current productivity a, and current menu cost f. The firm's problem can be written recursively as

$$V_{k,t}(p_{k,i,t-1}, a_{k,i,t}, f_{k,i,t}) = \max_{p_{k,i,t}} \left\{ u_{k,i,t} + \mathbb{E}_t \left[q_{t,t+1} V_{k,t+1}(p_{k,i,t}, a_{k,i,t+1}, f_{k,i,t+1}) \right] \right\}, \tag{4.11}$$

where current real profits net of adjustment costs are

$$u_{k,i,t} \equiv \left(\frac{p_{k,i,t}}{P_t} - \frac{W_t}{a_{k,i,t} P_t} \right) c_{k,i,t} - \mathbb{1}_{\{p_{k,i,t} \neq p_{k,i,t-1}\}} f_{k,i,t} \frac{W_t}{P_t},$$

and demand $c_{k,i,t}$ is given by (4.5).

Let $\psi_{k,t}(p_{k,i,t-1}, a_{k,i,t}, f_{k,i,t})$ denote the optimal pricing policy. Given this policy and the distribution of firm states, the sectoral and aggregate price index satisfy

$$P_{k,t} = \left(\int \left[\psi_{k,t}(p_{-1}, a, f) \right]^{1-\theta_k} d\Gamma_{k,t}(p_{-1}, a, f) \right)^{\frac{1}{1-\theta_k}},$$

$$P_t = \left(\sum_{k=1}^{K} \omega_k P_{k,t}^{1-\eta} \right)^{\frac{1}{1-\eta}}.$$

The distribution of firm states evolves according to firms' pricing decisions and the exogenous productivity and menu-cost processes. For any measurable set $\mathcal{B} \subseteq \mathbb{R}$,

$$\Gamma'_{k,t+1}(\mathcal{B}, a', f') = \left[\phi_k \mathbb{1}_{\{f'=0\}} + (1-\phi_k) \mathbb{1}_{\{f'=\overline{f}_k\}} \right] \times \int_{\{(p_{-1},a,f):\psi_{k,t}(p_{-1},a,f)\in\mathcal{B}\}} \Pr_k(a'|a) d\Gamma_{k,t}(p_{-1}, a, f). \tag{4.12}$$

Because we study the transition following a one-time unexpected aggregate monetary shock, the

pricing rules $\{\psi_{k,t}\}$ and firm distributions $\{\Gamma_{k,t}\}$ generally vary over time.[8]

4.3 Monetary Policy

Following the menu-cost literature, we specify monetary policy in terms of aggregate nominal expenditure. Let $S_t \equiv P_t C_t$ denote nominal spending. Its logarithm follows

$$\log S_t = \mu + \log S_{t-1} + \epsilon_t^S,$$

where ϵ_t^S denotes a monetary policy innovation. Our quantitative experiments consider a one-time unexpected innovation to nominal spending.

Throughout the quantitative analysis, monetary non-neutrality refers to the real response to this nominal expenditure shock. Following the menu-cost literature, we summarize monetary non-neutrality by the cumulative response of log real consumption, and hence real output, to the shock:

$$\mathcal{M} \equiv \sum_{h=0}^{\infty} \frac{\partial c_{t+h}}{\partial \epsilon_t^S},$$

where lowercase letters denote logarithms.

The nominal-spending formulation makes transparent the connection between price adjustment and monetary non-neutrality. Taking logs of the identity $S_t = P_t C_t$ gives

$$s_t = p_t + c_t.$$

For a given nominal-spending innovation, a smaller response of the aggregate price level necessarily implies a larger response of real activity.

Differences in firms' price-adjustment behavior translate directly into differences in monetary non-neutrality.

In the quantitative model, heterogeneity in $\{\theta_k\}$, and hence in desired markups $\{\mu_k\}$, generates corresponding heterogeneity in price-adjustment behavior across sectors. Changes in the distribution of sectoral markups can therefore affect aggregate monetary transmission.

4.4 The Equilibrium

Given the laws of motion for the exogenous shocks (4.9), (4.10) and the path of monetary policy shocks, an equilibrium of the economy consists of:

[8] We focus on one-time unexpected monetary shocks rather than systematic monetary policy shocks, as in Nakamura and Steinsson (2010). With state-dependent pricing, solving for equilibrium under aggregate uncertainty requires tracking the distribution of firms over previous prices and idiosyncratic states as an infinite-dimensional aggregate state variable, which substantially increases the computational burden.

1. the household's demand for differentiated goods $\{c_{k,i,t}\}$ given by condition (4.5) of households' decision problem,
2. firms' value functions $\left\{\{V_{k,t}\}_{k=1}^{K}\right\}_{t=0}^{\infty}$ and policy functions $\left\{\{\psi_{k,t}\}_{k=1}^{K}\right\}_{t=0}^{\infty}$, that solve firms' price-setting problem (4.11),
3. aggregate output C_t and sectoral outputs $\{C_{k,t}\}_{k=1}^{I}$ that satisfy equations (4.2) and (4.3),
4. aggregate price functions W_t, Q_t, $\{P_{k,t}\}_{k=1}^{K}$ and P_t that are determined by equations (4.8), (4.6) and (4.7),
5. the distributions on firms' states $\left\{\{\Gamma_{k,t}(p_{-1}, a, f)\}_{k=1}^{K}\right\}_{t=0}^{\infty}$ that evolve according to (4.12),
6. and the aggregate nominal spending $S_t = P_t C_t$.

Computing the Transition Path. In Section 5, we compute the deterministic transition of the economy following a one-time unexpected monetary policy shock. We briefly describe the computational procedure. We first assume that the economy starts in the steady state and returns to it after 200 periods. We next conjecture the entire path of the sectoral prices over the transition path. Given these prices, we solve firms' pricing problems backward. We then simulate firms in the economy forward using a non-stochastic simulation algorithm similar to Young (2010) to get the distribution over individual state variables at every period. We compute the sectoral prices using this distribution and the firms' policy rules. If these prices differ from the conjectured ones, we update the guessed prices until the equilibrium converges. In doing so, we obtain the aggregate and sectoral impulse response functions to the monetary policy shock.

5. Quantification

This section quantifies how the observed evolution of markups affects monetary transmission in the multi-sector menu-cost model. We first describe the calibration and the monetary experiment. We then use the calibrated model to trace the evolution of monetary non-neutrality from 1980 to 2016. As an external validation of the mechanism, we compare the model-implied evolution of price-adjustment frequency and size with their empirical counterparts. We next decompose the increase in monetary non-neutrality into the contributions of sectoral markups and sectoral expenditure weights and examine the role of heterogeneous markup growth across sectors. Finally, we quantify the mechanism linking markups to monetary non-neutrality using a constant-frequency Calvo benchmark, a sufficient-statistic decomposition, and the model-implied pricing policies.

5.1 Calibration and Model Fit

The calibration combines time-varying sectoral markups and expenditure weights with sector-specific parameters disciplined by micro evidence on price adjustment. The model contains $K = 8$ sectors.

The first set of time-varying parameters is the collection of within-sector elasticities of substitution, $\left\{\left\{\theta_{k,t}\right\}_{t=1980}^{2016}\right\}_{k=1}^{K}$. For each sector and year, we set $\theta_{k,t} = \mu_{k,t}^{ss}/(\mu_{k,t}^{ss} - 1)$, where $\mu_{k,t}^{ss}$ is the smoothed sectoral markup constructed in Section 2.2.

The second set of time-varying parameters is the vector of sectoral expenditure weights, $\{\omega_{k,t}\}_{k=1}^{K}$, which we construct from annual industry GDP data from the Bureau of Economic Analysis.

We set the remaining common parameters to conventional values from the menu-cost literature. The model is calibrated at a monthly frequency, with $\beta = 0.96^{1/12}$. We set $\gamma = 1$ and $\lambda = 0$, implying logarithmic utility over consumption and linear disutility from labor. We choose κ so that labor supply in the flexible-price steady state equals 1/3. The mean growth rate of nominal spending is set to 0.0018, matching the average growth rate of nominal GDP relative to real GDP over 1998–2005.[9] Following Nakamura and Steinsson (2010), we set the persistence of idiosyncratic productivity to $\rho_z = 0.7$. Table 2 summarizes these parameters.

Table 2: Externally Calibrated Parameters

Parameter	Description	Value	Target/Source
β	Discount factor	$0.96^{1/12}$	Convention
γ	Consumption curvature	1	Nakamura and Steinsson (2010)
λ	Inverse Frisch elasticity	0	Nakamura and Steinsson (2010)
κ	Labor coefficient	2.25	Steady-state labor supply
μ	Mean growth rate of S_t	0.0018	Nominal GDP growth minus real GDP growth
ρ_z	Idiosyncratic productivity persistence	0.7	Nakamura and Steinsson (2010)
$\{\omega_{k,t}\}$	Sectoral weights		Bureau of Economic Analysis

Four parameters remain to be calibrated for each sector: the standard deviation of idiosyncratic productivity innovations, σ_k^z; the probability of receiving an idiosyncratic productivity shock, α_k; the probability of drawing a zero menu cost, ϕ_k; and the positive menu cost, $\overline{f}_k$. These parameters determine the frequency and size distribution of price changes.

We discipline them using the micro-pricing statistics constructed from the data of Nakamura and Steinsson (2008). These data cover the manufacturing and service sectors during 1998–2005 and report price-adjustment moments for 270 Entry Level Items (ELIs) in the non-shelter component of the CPI.[10]

To combine the ELI-level pricing evidence with the NAICS-based markup measures, we construct a crosswalk between the two classifications. We match ELI categories to six-digit NAICS industries using product descriptions and individual item descriptions in the CPI micro data, allowing for many-to-many matches where necessary. This procedure delivers a common eight-sector classification for markups and price-adjustment moments. Appendix Table D.1 reports the sectoral pricing moments.

[9] The micro price-adjustment moments used in the calibration of Nakamura and Steinsson (2010) are measured over 1998–2005.

[10] The underlying CPI Research Database contains the micro price observations used to construct the CPI and has been widely used in the empirical price-setting literature; see, among others, Bils and Klenow (2004), Nakamura and Steinsson (2008), Bils et al. (2012), and Nakamura et al. (2018).

For each sector, we choose $\left(\sigma_k^z, \alpha_k, \phi_k, \overline{f}_k\right)$ to match four moments: the median frequency of price adjustment, the 25th percentile of the absolute size of price changes, the median absolute price change, and the 75th percentile of the absolute price-change distribution. The calibration uses each sector's average markup over 1998–2005. The resulting sector-specific parameters are reported in Table 3. We hold these parameters fixed when studying the evolution of the economy from 1980 to 2016; only sectoral markups and expenditure weights vary over time. Appendix Table D.1 shows that the calibrated model also performs well for untargeted pricing moments.

Table 3: Internally Calibrated Parameters

Sector	Markup	σ^z	α	ϕ	$\overline{f}$
Agriculture	1.32	0.258	0.161	0.044	0.031
Mining and Utilities	1.11	0.131	0.108	0.059	0.017
Manufacturing	1.26	0.125	0.084	0.048	0.023
Retail and Wholesale	1.09	0.065	0.073	0.016	0.052
Services (Information and Finance)	1.48	0.079	0.054	0.054	0.033
Education and Health Care	1.09	0.097	0.052	0.015	0.065
Services (Entertainment, etc.)	1.17	0.056	0.086	0.033	0.041
Other Services	1.22	0.049	0.140	0.013	0.075

Monetary experiment. To measure monetary non-neutrality, we introduce a one-time innovation to nominal spending. The economy receives an unexpected shock at $t = 1$, $\epsilon_1^S = \overline{\epsilon}$, with $\epsilon_t^S = 0$ for $t > 1$. We compute the transition following the shock under perfect foresight and normalize $\overline{\epsilon}$ so that nominal spending rises by 1 percent on impact.

Following the menu-cost literature, our measure of monetary non-neutrality is the cumulative response of log real GDP,

$$MN \equiv \sum_{h=0}^{\infty} \frac{\partial y_{t+h}}{\partial \epsilon_t^S} \overline{\epsilon}. \tag{5.1}$$

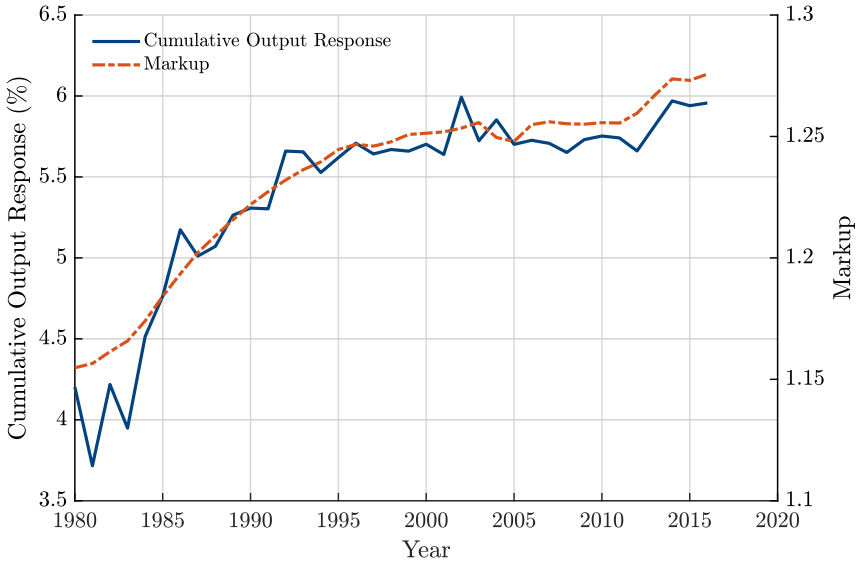
This measure captures both the magnitude and persistence of the real response to a nominal disturbance.

5.2 The Evolution of Monetary Non-neutrality

We first ask how the observed evolution of markups changes monetary transmission in the model. For each year from 1980 to 2016, we solve the model using the sectoral markups and expenditure weights observed in that year and hold these objects fixed over the transition following a one-time unexpected monetary shock. The sector-specific parameters governing idiosyncratic shocks and menu costs remain at their calibrated values. This exercise therefore maps the observed evolution of the sectoral composition and market power into a sequence of model economies.

Figure 5 reports the results. The dashed line plots the aggregate smoothed markup, while the solid line plots model-implied monetary non-neutrality. Monetary non-neutrality rises substantially over the sample and closely tracks the secular increase in markups. The same qualitative relationship holds within the individual sectors; Appendix Figure D.4 reports the sector-level results.

Figure 5: Aggregate Monetary Non-neutrality over Time



Note: The figure plots the evolution of the aggregate markup in the United States (dashed line) and model-implied aggregate monetary non-neutrality (solid line). Monetary non-neutrality is the cumulative response of log real GDP to the nominal-spending shock defined in equation (5.1). The shock is normalized to increase nominal spending by 1 percent on impact.

Between 1980 and 2016, the aggregate markup rises by approximately 15 percent, while model-implied monetary non-neutrality rises by approximately 42 percent. The average arc elasticity,

$$\frac{\Delta \log MN}{\Delta \log \mu},$$

is therefore close to 3 over the sample. The model implies that changes in market power of the magnitude observed in the data can materially alter the real effects of nominal disturbances.

Intermediate inputs. As a robustness exercise, Appendix C.1 extends the model to incorporate intermediate inputs through a parsimonious roundabout-production structure, following Nakamura and Steinsson (2010). Intermediate inputs substantially increase monetary non-neutrality in levels: across the sample, the cumulative real response is approximately three times as large as in the baseline model. This amplification reflects strategic complementarity in price setting, because sluggish input-price adjustment makes marginal costs respond more slowly to an aggregate nominal disturbance.

Despite this large level effect, the proportional increase in monetary non-neutrality associated with rising markups is similar to that in the baseline model. Roundabout production strengthens

the endogenous-price-rigidity mechanism: when higher markups reduce price adjustment in one sector, sluggish input prices further dampen the incentives of firms in other sectors to adjust. At the same time, the concavity of monetary non-neutrality with respect to markups limits the additional proportional amplification. Intermediate inputs therefore raise the level of monetary non-neutrality substantially without materially changing our conclusion about how its evolution is shaped by rising markups.

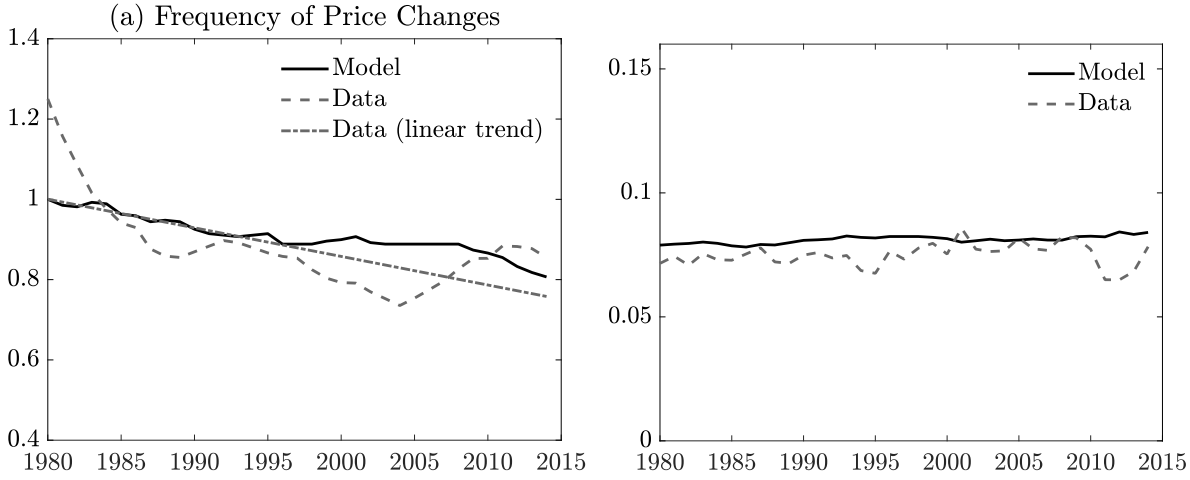
Interpretation of the historical exercise.. Our historical exercise should be interpreted as a sequence of comparative-static monetary experiments. For each year, we solve the model using that year's sectoral markups and expenditure weights and hold these objects fixed over the transition following a one-time unexpected monetary shock. The exercise therefore asks how the same nominal disturbance would propagate in economies characterized by different market structures, rather than tracing a historical transition in which firms anticipate future changes in markups or sectoral composition. This approximation is less restrictive when the real effects of the shock are concentrated at relatively short horizons. Appendix Figure D.3 shows that, using the 2000 calibration, the aggregate output response has a half-life of about six months and largely dissipates within two years. Hence, the cumulative monetary non-neutrality measure is driven primarily by short- and medium-run responses.

5.3 Price Adjustment over Time: Model Validation

The preceding subsection shows that the observed evolution of sectoral markups generates a substantial increase in model-implied monetary non-neutrality. The mechanism in Section 3 implies a corresponding prediction for price setting: as markups rise, firms should adjust prices less frequently. We therefore ask whether the same sequence of yearly calibrations also reproduces the observed evolution of price-adjustment behavior.

Figure 6 compares the evolution of two pricing moments in the model and in the data. Panel (a) reports the frequency of price adjustment (FPA). The dashed and dash-dotted lines show, respectively, a moving average and a linear trend constructed from the micro-price data of Nakamura et al. (2018); both indicate a substantial secular decline in the frequency of price changes. The solid line reports the corresponding model-implied FPA. Because Nakamura et al. (2018) aggregate price-adjustment frequencies at the major-group level, whereas our model aggregates across NAICS one-digit sectors, we focus on relative changes rather than levels. Specifically, we normalize both the model and data trend series to one in 1980 and plot FPA_t/FPA_{1980}. The model generates a pronounced decline in aggregate FPA that is broadly comparable in magnitude to the decline observed in the data. Panel (b) shows that the model also broadly reproduces the evolution of the average absolute size of price changes.

Figure 6: Frequency and Absolute Size of Price Changes



Note: Panel (a) compares the evolution of FPA in the data and the model. The dashed and dash-dotted lines report, respectively, the moving-average series and linear trend in the data from Nakamura et al. (2018); the solid line reports model-implied aggregate FPA. We normalize the data trend and model-implied frequency to one in 1980 and plot FPA_t/FPA_{1980}. Panel (b) compares the evolution of the absolute size of price changes in the data and the model. Model-implied moments are computed from the same yearly calibrations used in Figure 5.

This comparison provides an untargeted validation of the model. The sector-specific pricing parameters are calibrated to cross-sectional price-adjustment moments over 1998–2005, but the time-series evolution of FPA from 1980 to 2016 is not targeted. Instead, the decline in model-implied FPA emerges from the observed evolution of sectoral markups and expenditure weights. Thus, the same changes in the economic environment that generate stronger monetary non-neutrality also generate an increase in endogenous nominal rigidity consistent with the micro-price evidence.

Having established these aggregate time-series implications, we next ask how much of the increase in monetary non-neutrality is attributable to changing sectoral markups, as opposed to changes in the sectoral composition of the economy.

5.4 Markup and Weight Decomposition

The preceding exercise allows both sectoral markups and sectoral expenditure weights to evolve over time. We next quantify the contribution of each.

With the Cobb–Douglas aggregation across sectors,

$$\log C_t = \sum_{k=1}^{K} \omega_{k,t} \log C_{k,t},$$

aggregate monetary non-neutrality can be written as

$$MN_t = \sum_{k=1}^{K} \omega_{k,t} MN_{k,t},$$

where $MN_{k,t}$ denotes the cumulative real response in sector k in the model calibrated to year t.

The change relative to 1980 admits the exact decomposition

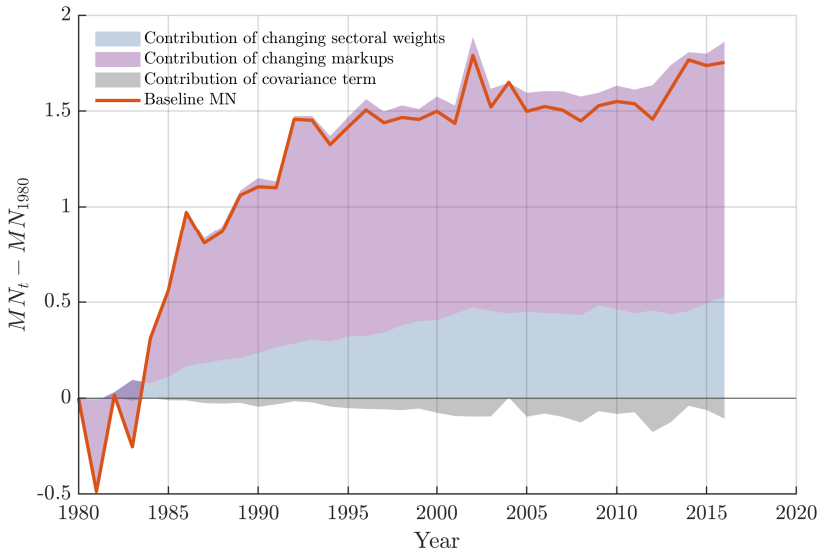
$$\begin{aligned} MN_t - MN_{1980} = & \underbrace{\sum_k \left(\omega_{k,t} - \omega_{k,1980}\right) MN_{k,1980}}_{\text{sectoral-weight changes}} \\ & + \underbrace{\sum_k \omega_{k,1980} \left(MN_{k,t} - MN_{k,1980}\right)}_{\text{markup-induced changes}} \\ & + \underbrace{\sum_k \left(\omega_{k,t} - \omega_{k,1980}\right) \left(MN_{k,t} - MN_{k,1980}\right)}_{\text{interaction term}}. \end{aligned} \tag{5.2}$$

The first component captures the effect of changes in sectoral expenditure shares, holding sector-level monetary non-neutrality at its 1980 value. The second captures changes in sector-level monetary non-neutrality induced by changing markups, holding expenditure weights at their 1980 values. The final term captures their interaction: it is positive when sectors whose monetary non-neutrality rises more also gain expenditure share.

Figure 7 reports the decomposition. Between 1980 and 2016, changing sectoral markups account for 75.9 percent of the increase in aggregate monetary non-neutrality. Changes in sectoral weights account for 30.2 percent, while the interaction term contributes −6.1 percent. Thus, the rise in monetary non-neutrality is driven primarily by changes within sectors rather than by a reallocation of economic activity across sectors.

Changes in sectoral composition nevertheless reinforce this effect. The positive weight component indicates that sectors with relatively high monetary non-neutrality in 1980 gained economic weight, on net, over the sample. The negative interaction term partially offsets this contribution: sectors experiencing larger subsequent increases in monetary non-neutrality tended, on average, to gain less weight or lose weight relative to other sectors. Quantitatively, however, this offset is modest. Overall, the decomposition points to the evolution of sectoral markups, rather than structural change in sectoral composition, as the main force behind the increase in aggregate monetary non-neutrality.

Figure 7: Decomposition of the Evolution of Monetary Non-neutrality



Note: The solid line plots model-implied aggregate monetary non-neutrality. The shaded areas report the contributions of changes in sectoral markups, changes in sectoral expenditure weights, and their interaction, as defined in equation (5.2).

5.5 Equal-Change Counterfactual

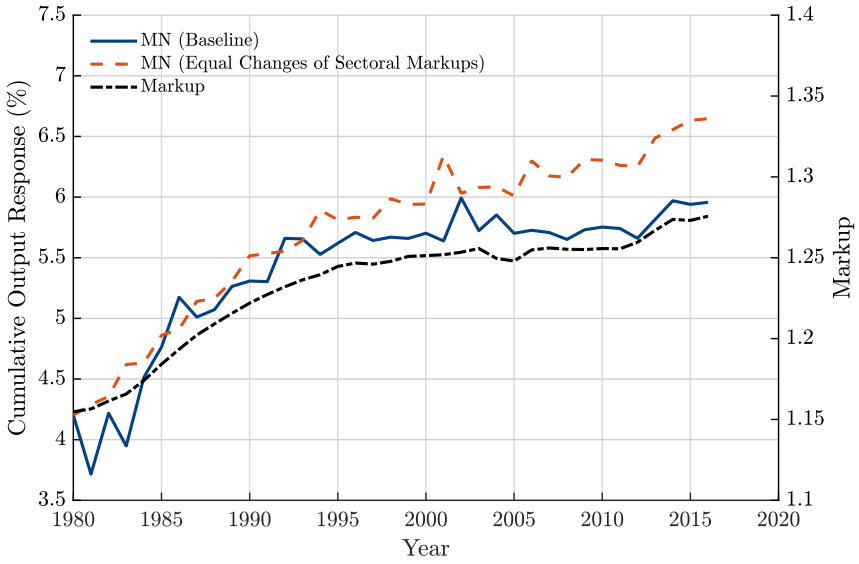
The aggregate increase in markups masks substantial heterogeneity across sectors. Figure 1 shows that some sectors experience much larger increases than others. Because the relationship between markups and monetary non-neutrality is concave, this dispersion in markup growth can matter for the aggregate response even holding fixed the path of the aggregate markup.

To quantify this effect, we construct an equal-change counterfactual. In each year, we replace the observed sectoral markup changes since 1980 with common sectoral changes chosen to preserve the aggregate markup. We then resolve the model using these counterfactual sectoral markups, while leaving all other calibrated objects unchanged. Thus, the exercise changes the cross-sectoral distribution of markup growth without changing its aggregate path.

Figure 8 compares the resulting monetary non-neutrality with the baseline. Under the observed heterogeneous evolution of sectoral markups, monetary non-neutrality rises by approximately 42 percent between 1980 and 2016. When markup increases are instead equalized across sectors, it rises by 58 percent. The corresponding markup elasticity increases from approximately 3 in the baseline to 3.7 in the equal-change counterfactual.

The result reflects the concavity results discussed in Section 3. Concentrating a given aggregate increase in markups in a subset of sectors yields a smaller increase in aggregate monetary non-neutrality than distributing the same markup increase more evenly across sectors. Heterogeneous sectoral markup growth therefore dampens the aggregate increase in monetary non-neutrality.

Figure 8: Monetary Non-neutrality over Time: An Equal-Change Counterfactual



Note: The solid line plots model-implied monetary non-neutrality under the observed evolution of sectoral markups. The dashed line plots monetary non-neutrality under a counterfactual that preserves the aggregate markup in each year but equalizes markup changes across sectors. The remaining line reports the aggregate markup.

5.6 The Frequency Mechanism: Quantification

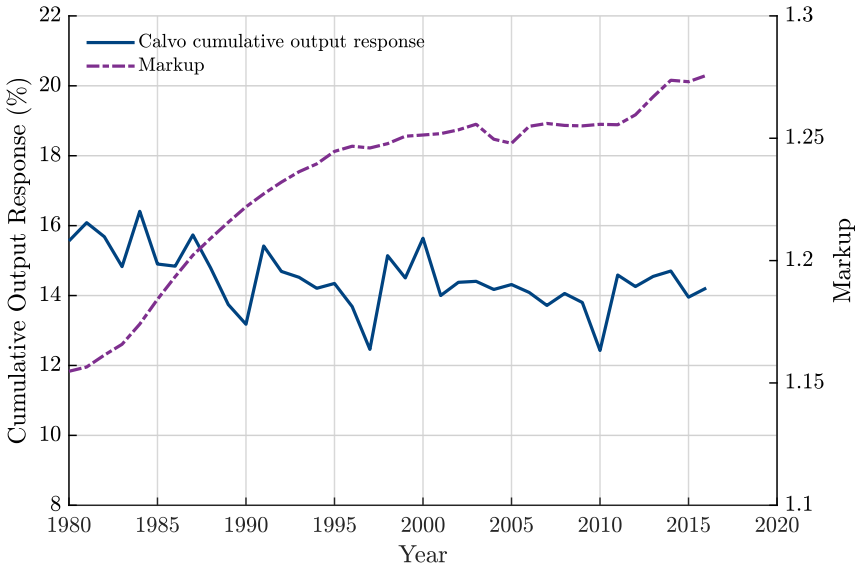
The preceding results show that the model generates both a substantial increase in monetary non-neutrality and a decline in the frequency of price adjustment consistent with the data. We now ask whether the latter is quantitatively responsible for the former.

We proceed in three steps. First, we shut down the endogenous response of price-adjustment frequency using a Calvo benchmark. Second, we use a sufficient-statistic decomposition to separate the frequency and selection margins. Third, we examine the pricing policies underlying the decline in adjustment frequency by studying the inaction region and the stationary distribution of price gaps.

5.6.1 A Constant-Frequency Benchmark: The Calvo Model

We first construct a Calvo counterpart of the quantitative model. For each sector, we choose a fixed probability of price adjustment to match its benchmark price-adjustment frequency and hold this probability constant as markups evolve over time. The remaining structure of the model, including the evolution of sectoral markups and expenditure weights, is unchanged. The exercise therefore removes the endogenous response of adjustment frequency to market power while preserving other channels through which markups may affect equilibrium.

Figure 9: Aggregate Monetary Non-neutrality over Time: Calvo Model



Note: The figure plots the aggregate markup (dashed line) and model-implied monetary non-neutrality (solid line) in the Calvo counterpart of the model. Sectoral price-adjustment probabilities are held fixed over time. Monetary non-neutrality is defined in equation (5.1), and the nominal-spending shock is normalized to increase nominal spending by 1 percent on impact.

Figure 9 reports the resulting evolution of monetary non-neutrality. In sharp contrast to the menu-cost model, monetary non-neutrality is nearly flat over the sample and, if anything, declines slightly despite the substantial increase in the aggregate markup. Thus, a higher markup by itself does not mechanically imply stronger monetary non-neutrality. The pronounced increase in the baseline model emerges because the frequency of price adjustment responds endogenously to changes in market power.

This comparison also illustrates a distinction between state-dependent and time-dependent pricing. In the Calvo model, the degree of nominal rigidity is imposed directly through an exogenous adjustment hazard. Changing firms' desired markups therefore leaves this source of rigidity unchanged. In the menu-cost model, by contrast, the adjustment hazard is an equilibrium outcome of firms' pricing incentives. The same change in markups alters the profitability of adjustment and hence the effective degree of nominal rigidity.

5.6.2 A Sufficient-Statistic Decomposition

We next quantify the relative importance of the frequency and selection margins using the sufficient-statistic approach of Alvarez et al. (2022). For a small log nominal-spending shock Δs, the cumulative real response can be approximated by

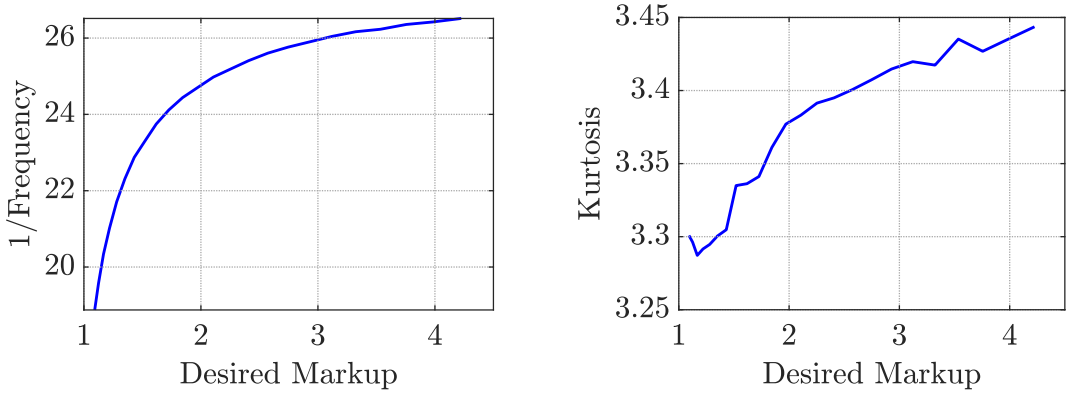
$$\mathcal{M}(\Delta s; \mu_{ss}) = \frac{1}{6} \frac{Kurt(\mu_{ss})}{Freq(\mu_{ss})} \Delta s + o(\Delta s), \tag{5.3}$$

where $Freq(\mu_{ss})$ denotes the steady-state frequency of price changes and $Kurt(\mu_{ss})$ the kurtosis of the nonzero price-change distribution.

Equation (5.3) separates two forces. The *frequency effect* is captured by $1/Freq(\mu_{ss})$: less frequent adjustment makes a nominal shock more persistent in real terms. The *selection effect* is summarized by $Kurt(\mu_{ss})$: stronger selection of firms into adjustment accelerates the aggregate price response and reduces monetary non-neutrality.

Panels (a) and (b) of Figure 10 show how these two sufficient statistics change with the desired markup in the calibrated model. Both $1/Freq$ and $Kurt$ rise with the markup over the relevant range. The frequency margin is quantitatively dominant: approximately 90 percent of the increase in monetary non-neutrality associated with higher markups is attributable to the decline in adjustment frequency. The selection margin contributes the remainder.

Figure 10: Monetary Non-neutrality: Frequency and Selection



(a) Inverse Frequency (b) Kurtosis

Note: The figure varies the desired markup in the calibrated model while holding other structural parameters fixed. Panel (a) plots the inverse FPA, which captures the frequency component of the sufficient statistic in equation (5.3). Panel (b) plots the kurtosis of the nonzero price-change distribution, which summarizes the selection component.

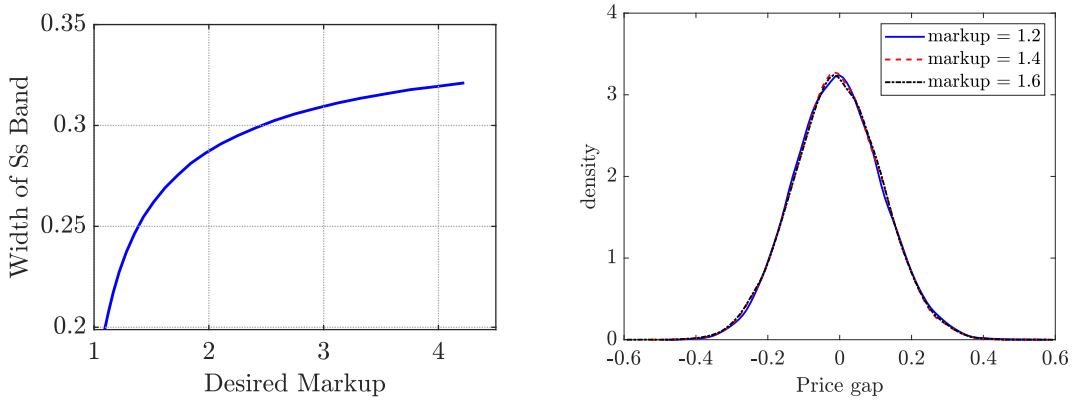
5.6.3 Why Does the Frequency of Price Adjustment Fall?

The dominance of the frequency margin raises a more primitive question: why does the frequency of price adjustment respond so strongly to markups in the quantitative model?

The frequency with which firms adjust depends on both the location of the adjustment boundaries and the distribution of firms within those boundaries. A firm adjusts when its price gap, which is the difference between its current price and its frictionless optimal price, leaves the (s, S) inaction region. Higher adjustment frequency can therefore arise either because the inaction region narrows or because more probability mass moves toward its boundaries.

Panel (a) of Figure 11 shows that the first force is central. As the desired markup rises, the width of the inaction region increases substantially. Moreover, the relationship is concave: increases in the markup have progressively smaller effects on the adjustment boundaries when markups are already

Figure 11: Width of Inaction Region and Price-Gap Distribution



(a) Width of the Inaction Region

(b) Distribution of Price Gaps

Note: The figure varies the desired markup in the calibrated model while holding other structural parameters fixed. Panel (a) plots the width of the (s, S) band. Panel (b) plots the distribution of price gaps when markups are set at 1.2, 1.4 and 1.6.

high. Panel (b) of Figure 11, by contrast, shows that the stationary distribution of price gaps changes relatively little as the markup varies. The decline in FPA is therefore driven primarily by firms' greater willingness to tolerate a given price misalignment rather than by large shifts in the distribution of price gaps.

This result mirrors the analytical mechanism in Section 3. Higher desired markups make normalized profits less sensitive to deviations from the frictionless optimal price. Firms consequently require a larger price gap before paying the menu cost, which widens the inaction region and reduces adjustment frequency. Because the widening of the inaction region itself exhibits diminishing returns, the resulting increase in monetary non-neutrality is also concave.

Summary. Taken together, the Calvo counterfactual, the sufficient-statistic decomposition, and the behavior of the inaction region provide complementary evidence for the same mechanism. Rising markups matter for monetary transmission primarily because they endogenously increase nominal rigidity. This frequency channel accounts for most of the model-implied increase in monetary non-neutrality between 1980 and 2016.

6. Concluding Remarks

This paper shows that rising market power can strengthen monetary non-neutrality by lowering firms' incentives to adjust prices. Consistent with this mechanism, higher markups are associated with lower FPA and stronger monetary transmission. Quantitatively, monetary non-neutrality rises by about 42 percent between 1980 and 2016, with changing sectoral markups accounting for 75.9 percent of the increase and the frequency margin playing the dominant role. More broadly, the results imply

that nominal rigidity evolves endogenously with market structure and that monetary transmission depends on both the level and distribution of market power.

References

Alvarez, F. E., A. Ferrara, E. Gautier, H. Le Bihan, and F. Lippi (2021): "Empirical investigation of a sufficient statistic for monetary shocks,"Technical report, National Bureau of Economic Research.

Alvarez, F., H. Le Bihan, and F. Lippi (2016): "The real effects of monetary shocks in sticky price models: a sufficient statistic approach," *American Economic Review*, 106, 2817–51.

Alvarez, F., and F. Lippi (2022): "The analytic theory of a monetary shock," *Econometrica*, 90, 1655–1680.

Alvarez, F., F. Lippi, and A. Oskolkov (2022): "The macroeconomics of sticky prices with generalized hazard functions," *The Quarterly Journal of Economics*, 137, 989–1038.

Alvarez, F., F. Lippi, and P. Souganidis (2023): "Price setting with strategic complementarities as a mean field game," *Econometrica*, 91, 2005–2039.

Atkeson, A., and A. Burstein (2008): "Pricing-to-market, trade costs, and international relative prices," *American Economic Review*, 98, 1998–2031.

Auclert, A., R. D. Rigato, M. Rognlie, and L. Straub (2022): "New Pricing Models, Same Old Phillips Curves?"Technical report, National Bureau of Economic Research.

Ball, L., N. G. Mankiw, D. Romer, G. A. Akerlof, A. Rose, J. Yellen, and C. A. Sims (1988): "The new Keynesian economics and the output-inflation trade-off," *Brookings Papers on Economic Activity*, 1988, 1–82.

Banerjee, A., and B. Russell (2001): "The relationship between the markup and inflation in the G7 economies and Australia," *Review of Economics and Statistics*, 83, 377–384.

Baqaee, D. R. (2018): "Cascading failures in production networks," *Econometrica*, 86, 1819–1838.

Baqaee, D. R., and E. Farhi (2020): "Productivity and misallocation in general equilibrium," *The Quarterly Journal of Economics*, 135, 105–163.

Barro, R. J. (1972): "A theory of monopolistic price adjustment," *The Review of Economic Studies*, 39, 17–26.

Bénabou, R. (1992): "Inflation and markups: theories and evidence from the retail trade sector," *European Economic Review*, 36, 566–574.

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

Bils, M., P. J. Klenow, and B. A. Malin (2012): "Reset price inflation and the impact of monetary policy shocks," *American Economic Review*, 102, 2798–2825.

Calvo, G. A. (1983): "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, 12, 383–398.

Carvalho, C. (2006): "Heterogeneity in price stickiness and the real effects of monetary shocks," *Frontiers in Macroeconomics*, 2.

Carvalho, C., N. A. Dam, and J. W. Lee (2020): "The cross-sectional distribution of price stickiness implied by aggregate data," *Review of Economics and Statistics*, 102, 162–179.

Carvalho, C., and F. Schwartzman (2015): "Selection and monetary non-neutrality in time-dependent pricing models," *Journal of Monetary Economics*, 76, 141–156.

Chiavari, A., D. Smirnov, and M. Morazzoni (2021): "Heterogeneous Markups Cyclicality and Monetary Policy," *Working paper*.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The rise of market power and the macroeconomic implications," *The Quarterly Journal of Economics*, 135, 561–644.

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

Dotsey, M., R. G. King, and A. L. Wolman (1999): "State-dependent pricing and the general equilibrium dynamics of money and output," *The Quarterly Journal of Economics*, 114, 655–690.

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

Gautier, E., and H. L. Bihan (2018): "Shocks vs menu costs: Patterns of price rigidity in an estimated multi-sector menu-cost model," *The Review of Economics and Statistics*, 1–45.

Gertler, M., and P. Karadi (2015): "Monetary policy surprises, credit costs, and economic activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

Gertler, M., and J. Leahy (2008): "A Phillips curve with an Ss foundation," *Journal of Political Economy*, 116, 533–572.

Golosov, M., and R. E. Lucas (2007): "Menu costs and phillips curves," *Journal of Political Economy*, 115, 171–199.

Gopinath, G., and O. Itskhoki (2010): "Frequency of price adjustment and pass-through," *The Quarterly Journal of Economics*, 125, 675–727.

Hobijn, B., F. Ravenna, and A. Tambalotti (2006): "Menu costs at work: restaurant prices and the introduction of the euro," *The Quarterly Journal of Economics*, 121, 1103–1131.

Jarociński, M., and P. Karadi (2020): "Deconstructing monetary policy surprises - the role of information shocks," *American Economic Journal: Macroeconomics*, 12, 1–43.

Jordà, Ò. (2005): "Estimation and inference of impulse responses by local projections," *American Economic Review*, 95, 161–182.

Karadi, P., and A. Reiff (2019): "Menu costs, aggregate fluctuations, and large shocks," *American Economic Journal: Macroeconomics*, 11, 111–46.

Karadi, P., R. Schoenle, and J. Wursten (2022): "Price Selection in the Microdata."

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

Lan, T., L. Li, and M. Li (2026): "Cyclical inequality in the cost of living and implications for monetary policy," *International Economic Review*, 67, 839–874.

Li, M. (2026): "Anatomy of Countercyclical Price-Change Dispersion: HANK meets Non-homothetic Demand."

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

Midrigan, V. (2011): "Menu costs, multiproduct firms, and aggregate fluctuations," *Econometrica*, 79, 1139–1180.

Miranda-Agrippino, S., and G. Ricco (2021): "The transmission of monetary policy shocks," *American Economic Journal: Macroeconomics*, 13, 74–107.

Mongey, S. (2021): "Market structure and monetary non-neutrality,"Technical report, National Bureau of Economic Research.

Nakamura, E., and J. Steinsson (2008): "Five facts about prices: A reevaluation of menu cost models," *The Quarterly Journal of Economics*, 123, 1415–1464.

——— (2010): "Monetary non-neutrality in a multisector menu cost model," *The Quarterly Journal of Economics*, 125, 961–1013.

——— (2018): "High-frequency identification of monetary non-neutrality: the information effect," *The Quarterly Journal of Economics*, 133, 1283–1330.

Nakamura, E., J. Steinsson, P. Sun, and D. Villar (2018): "The elusive costs of inflation: Price dispersion during the US great inflation," *The Quarterly Journal of Economics*, 133, 1933–1980.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297, http://www.jstor.org/stable/2171831.

Pasten, E., R. Schoenle, and M. Weber (2020): "The propagation of monetary policy shocks in a heterogeneous production economy," *Journal of Monetary Economics*, 116, 1–22.

Ueda, K. (2023): "Duopolistic competition and monetary policy," *Journal of Monetary Economics*, 135, 70–85.

Vavra, J. (2014): "Inflation Dynamics and Time-Varying Volatility: New Evidence and an Ss Interpretation," *The Quarterly Journal of Economics*, 129, 215–258.

Wang, O., and I. Werning (2022): "Dynamic oligopoly and price stickiness," *American Economic Review*, 112, 2815–2849.

Young, E. R. (2010): "Solving the incomplete markets model with aggregate uncertainty using the Krusell–Smith algorithm and non-stochastic simulations," *Journal of Economic Dynamics and Control*, 34, 36–41.

Zhang, D. (2022): "Term Structure, Forecast Revision, and the Signaling Channel of Monetary Policy," *Journal of the European Economic Association*, 20, 1522–1553.

Supplementary Appendix

The (Unequal) Rise of Market Power, Nominal Rigidities, and Monetary Non-Neutrality

Christian Höynck Minghao Li Donghai Zhang

August 2026

A Data Appendix **A-2**
A.1 Markup Estimation and Variable Construction A-2
A.2 Incremental Explanatory Power of the FPA Interaction A-3
A.3 Inference for Changes in the Markup Interaction Coefficient A-4

B Appendix for the Tractable Menu-Cost Model **A-5**
B.1 Markup and the Local Profit Loss A-5
B.2 Optimal Inaction Region and Frequency of Price Adjustment A-6
B.3 General Conditions for the Curvature of the Inaction Region A-8
B.4 Understanding the Barro (1972) Example A-11
B.5 The Stationary Distribution and a Marginal Monetary Shock A-13
B.6 The Monetary-Shock Impulse Response A-14
B.7 Cumulative Real Effects and the FPA Channel A-15
B.8 Oligopolistic Competition A-17

C Appendix for Quantitative Results **A-19**
C.1 Intermediate Inputs and Roundabout Production A-19

D Figures and Tables **A-22**

A. Data Appendix

A.1 Markup Estimation and Variable Construction

This appendix describes the construction of our markup measures. We estimate firm-level markups following the method in Edmond et al. (2023), and then aggregate the estimates to the sector-year level used in the empirical analysis.

Firm-level markups. The estimation follows the framework developed by Edmond et al. (2023), which builds on the production-based approach of De Loecker and Warzynski (2012). We use quarterly Compustat data for publicly traded U.S. firms from 1980 to 2016. The data cover sales, employment, capital, and input expenditures across manufacturing and service sectors. We estimate firm-level markups using the single-input approach of De Loecker and Warzynski (2012). According to this approach, the markup $\mu_{i,t}$ of a firm i at time t can be computed from one flexible input, X_i, as the ratio of the output elasticity of the input, $\epsilon_{Q,X_i,t}$, to the revenue share of that input, $s_{R,X_i,t}$

$$\mu_t(i) = \frac{\epsilon_{Q,X_i,t}}{s_{R,X_i,t}}.$$

Compustat reports Cost of Goods Sold (COGS), a composite of intermediate and labor inputs, which we use as the partially flexible input, X_i. We use a variant of the technique introduced by Olley and Pakes (1996) and described in De Loecker and Warzynski (2012) to estimate a Cobb-Douglas function and obtain a time-varying estimate of output elasticity at the sector level. We calculate firm-level markups by dividing the sector-year output elasticity by the firm-year COGS revenue share. In terms of implementation, we follow the procedure described in De Loecker et al. (2020) with the adjustments described in Baqaee and Farhi (2020). In particular, we estimate time-varying output elasticities and deflate using gross output price indices from KLEMS sector-level data.[11]

When transforming the data, we drop all firms in the government sector or the sector of the economy composed of finance, insurance, and real estate that are heavily regulated. We retain positive observations, linearly interpolate variables missing for one period, and trim calculated markups at the 1 percent level.[12]

Aggregation to sector-level and aggregate markups. We follow Edmond et al. (2023) to aggregate firm-level markups using cost weights. Specifically, the markup of sector s in year t is constructed as

$$\mu_{s,t} = \sum_i \mu_t(i) \frac{c_t(i)}{C_{s,t}} di = \left(\sum_i \omega_t(i) \frac{1}{\mu_t(i)} \right)^{-1} \tag{A.1}$$

where $c_t(i)/C_{s,t}$ denotes the firm's share of total sector variable costs, measured consistently with the flexible-input markup estimate. This is equivalent to a sales-weighted harmonic average of firm-level

[11] KLEMS stands for capital (K), labor (L), energy (E), materials (M), and service (S).
[12] Our findings are robust to trimming at 5%.

markups, where $\omega_t(i) = \text{Sales}_{it} / \sum_{j \in s} \text{Sales}_{jt}$ denotes firm i's share of total sector sales, as shown in expression (A.1).

Smoothed markups. We finally define the smoothed markup of sector k in year t, denoted by $\mu^{ss}_{k,t}$, as the seven-year centered moving average of the annual sectoral markup $\mu_{k,t}$.

$$\mu^{ss}_{k,t} = \frac{1}{7} \sum_{j=-3}^{3} \mu_{k,t+j}.$$

We use the same smoothing procedure to construct the aggregate markup series when persistent markup variation is the relevant state variable.

A.2 Incremental Explanatory Power of the FPA Interaction

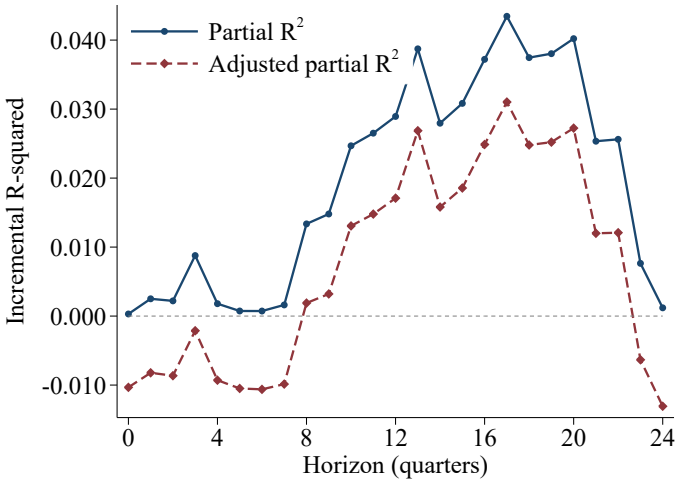
We assess whether the FPA interaction improves the fit of the markup–inflation specification. Because equation (2.4) nests equation (2.3), the conventional in-sample R^2 cannot fall when the FPA interaction is added. We therefore calculate the partial R^2 associated with the FPA interaction:

$$R^{2,\text{partial}}_{FPA,h} = \frac{SSR_{\mu,\pi,h} - SSR_{\mu,\pi,F,h}}{SSR_{\mu,\pi,h}} = 1 - \frac{SSR_{\mu,\pi,F,h}}{SSR_{\mu,\pi,h}}, \tag{A.2}$$

where $SSR_{\mu,\pi,h}$ is the sum of squared residuals from equation (2.3), and $SSR_{\mu,\pi,F,h}$ is the corresponding quantity from equation (2.4). The statistic measures the share of residual variation in y_{t+h} under the markup–inflation specification accounted for by adding the FPA interaction. We also report an adjusted partial R^2, which penalizes the FPA-augmented specification for the additional regressor.

The improvement in fit is concentrated at the horizons for which the FPA interaction is estimated most precisely. The partial R^2 peaks at quarter 17, when adding the FPA interaction accounts for 4.34 percent of the residual variation under the markup–inflation specification. After penalizing the augmented model for the additional regressor, the adjusted partial R^2 is positive from quarters 8 through 22 and also peaks at quarter 17, at 3.10 percent. The gains in fit are modest but persistent over these horizons, indicating that the FPA contains incremental information about the state dependence of monetary transmission beyond markups and inflation.

Figure A.1: Incremental Explanatory Power of the FPA Interaction



Notes: This figure reports the partial R^2 defined in equation (A.2) and the adjusted partial R^2, defined as $1 - \frac{SSR_{\mu,\pi,F,h}/df_{\mu,\pi,F,h}}{SSR_{\mu,\pi,h}/df_{\mu,\pi,h}}$. Both statistics measure the incremental explanatory power of the FPA interaction relative to the markup–inflation specification in equation (2.3).

A.3 Inference for Changes in the Markup Interaction Coefficient

To assess how the estimated markup dependence of monetary transmission changes when inflation and the frequency of price adjustment are included, we estimate the three local-projection specifications in equations (2.1), (2.3), and (2.4). At each horizon, the FPA-augmented specification determines the common estimation sample, and the two restricted specifications are re-estimated on exactly the same observations. We calculate

$$\widehat{\Delta}_{\pi,h} = \widehat{\beta}^{\mu}_{\mu h} - \widehat{\beta}^{\mu,\pi}_{\mu h}, \qquad \widehat{\Delta}_{F,h} = \widehat{\beta}^{\mu,\pi}_{\mu h} - \widehat{\beta}^{\mu,\pi,F}_{\mu h}.$$

Because the coefficients entering each difference are estimated from the same observations, their sampling covariance must be taken into account. For each specification, we construct the observation-level influence function of the estimated markup interaction coefficient. If $X_{s,h}$ is the regressor matrix, $\widehat{u}_{s,t,h}$ is the residual, and e_μ selects the markup interaction coefficient in specification s, the contribution of observation t is

$$\widehat{\psi}_{s,t,h} = e_\mu'(X_{s,h}'X_{s,h})^{-1}x_{s,t,h}\widehat{u}_{s,t,h}.$$

We form the difference in influence functions for each comparison, $\widehat{q}_{\pi,t,h} = \widehat{\psi}_{\mu,t,h} - \widehat{\psi}_{\mu,\pi,t,h}$ and $\widehat{q}_{F,t,h} = \widehat{\psi}_{\mu,\pi,t,h} - \widehat{\psi}_{\mu,\pi,F,t,h}$, and estimate its long-run variance using a Bartlett heteroskedasticity-and-autocorrelation-consistent estimator. The bandwidth is selected automatically using the fuller specification in each pair: the markup-and-inflation specification for $\widehat{\Delta}_{\pi,h}$ and the FPA-augmented specification for $\widehat{\Delta}_{F,h}$. This procedure incorporates the cross-model covariance and the serial corre-

lation induced by overlapping local-projection outcomes. The reported pointwise 90% confidence interval is

$$\widehat{\Delta}_{j,h} \pm 1.645\,\widehat{\mathrm{se}}(\widehat{\Delta}_{j,h}), \qquad j \in \{\pi, F\}.$$

B. Appendix for the Tractable Menu-Cost Model

This appendix provides the derivations and proofs for the analytical results in Section 3. We first derive the local profit-loss function under CES demand and solve the resulting auxiliary menu-cost problem. We then characterize more general conditions under which markups widen the inaction region and increase price duration, and relate these conditions to the linear-demand example in Barro (1972). Finally, we derive the response to a marginal monetary shock and discuss the generality of the profit-curvature mechanism under oligopolistic competition.

B.1 Markup and the Local Profit Loss

Firm i's flow profit is

$$\Pi_i(P_i; P_i^*) = \left(P_i - \frac{P_i^*}{\mu}\right) Y \left(\frac{P_i}{P}\right)^{-\theta},$$

where we have used $P_i^* = \mu W_i$. Let $z = \log P_i - \log P_i^*$, so that $P_i = P_i^* e^z$.

The ratio of actual flow profit to frictionless flow profit is

$$\frac{\Pi_i(P_i; P_i^*)}{\Pi_i(P_i^*; P_i^*)} = \frac{P_i^* e^z - P_i^*/\mu}{P_i^* - P_i^*/\mu} e^{-\theta z} = \frac{\mu e^z - 1}{\mu - 1} e^{-\theta z}. \tag{B.1}$$

Define the normalized flow-profit loss as

$$\ell(z; \mu) \equiv 1 - \frac{\Pi_i(P_i; P_i^*)}{\Pi_i(P_i^*; P_i^*)}.$$

Using equation (B.1),

$$\ell(z; \mu) = 1 - \frac{\mu e^z - 1}{\mu - 1} e^{-\theta z}.$$

At the frictionless optimum,

$$\ell(0; \mu) = 0.$$

Moreover, using

$$\mu = \frac{\theta}{\theta - 1},$$

we obtain

$$\ell_z(0; \mu) = 0, \qquad \ell_{zz}(0; \mu) = \theta(\theta - 1).$$

A second-order expansion around $z = 0$ therefore gives

$$\ell(z;\mu) = \frac{1}{2}\theta(\theta-1)z^2 + o(z^2). \tag{B.2}$$

Since $\theta = \mu/\mu - 1$ and $\theta - 1 = 1/\mu - 1$, equation (B.2) becomes

$$\ell(z;\mu) = \frac{1}{2}\frac{\mu}{(\mu-1)^2}z^2 + o(z^2).$$

Thus, defining

$$\kappa(\mu) \equiv \frac{\mu}{(\mu-1)^2},$$

we have

$$\kappa'(\mu) = -\frac{\mu+1}{(\mu-1)^3} < 0.$$

Hence, higher desired markups make normalized profits locally flatter around the frictionless optimal price.

B.2 Optimal Inaction Region and Frequency of Price Adjustment

We treat the quadratic approximation

$$L(z;\mu) = \frac{1}{2}\kappa(\mu)z^2$$

as the exact flow-loss function of an auxiliary menu-cost problem. Let $v(z)$ denote the relative value function and ρ the long-run average cost. Within the inaction region, the average-cost Hamilton–Jacobi–Bellman equation is

$$\rho = \frac{1}{2}\kappa(\mu)z^2 + \frac{\sigma^2}{2}v''(z), \qquad |z| < \overline{z}. \tag{B.3}$$

By symmetry, it is sufficient to consider the upper boundary. Value matching and smooth pasting require

$$v(\overline{z}) = v(0) + \psi,$$
$$v'(\overline{z}) = 0,$$

together with the symmetry condition

$$v'(0) = 0. \tag{B.4}$$

Rearranging equation (B.3),

$$v''(z) = \frac{2\rho}{\sigma^2} - \frac{\kappa(\mu)}{\sigma^2}z^2.$$

Integrating and imposing equation (B.4) yields

$$v'(z) = \frac{2\rho}{\sigma^2} z - \frac{\kappa(\mu)}{3\sigma^2} z^3. \tag{B.5}$$

The smooth-pasting condition implies

$$\rho = \frac{\kappa(\mu)}{6} \overline{z}^2. \tag{B.6}$$

Integrating equation (B.5) from zero to z gives

$$v(z) - v(0) = \frac{\rho}{\sigma^2} z^2 - \frac{\kappa(\mu)}{12\sigma^2} z^4.$$

Evaluating this expression at $z = \overline{z}$, imposing value matching, and substituting equation (B.6) yield

$$\psi = \frac{\kappa(\mu)}{12\sigma^2} \overline{z}^4.$$

The optimal threshold is therefore

$$\overline{z}(\mu) = \left[\frac{12\psi\sigma^2}{\kappa(\mu)} \right]^{1/4} = \left[\frac{12\psi\sigma^2(\mu-1)^2}{\mu} \right]^{1/4}.$$

The total width of the inaction region is

$$S(\mu) = 2(12\psi\sigma^2)^{1/4}(\mu-1)^{1/2}\mu^{-1/4}.$$

Its first derivative is

$$S'(\mu) = S(\mu)\frac{\mu+1}{4\mu(\mu-1)} > 0,$$

while its second derivative is

$$S''(\mu) = -\frac{(12\psi\sigma^2)^{1/4}\left(3\mu^2+6\mu-5\right)}{8\mu^{9/4}(\mu-1)^{3/2}} < 0,$$

because $3\mu^2 + 6\mu - 5 > 0 \qquad$ for all $\mu > 1$.

To derive the frequency of price adjustment, let $\tau = \inf\{t \geq 0 : |z_t| = \overline{z}\}$ denote the first-exit time following a reset to $z = 0$, and define $T(z) = \mathbb{E}[\tau \mid z_0 = z]$. The expected exit time satisfies

$$\frac{\sigma^2}{2} T''(z) = -1, \qquad z \in (-\overline{z}, \overline{z}),$$

with boundary conditions $T(-\overline{z}) = T(\overline{z}) = 0$. The solution is

$$T(z) = \frac{\overline{z}^2 - z^2}{\sigma^2}.$$

Starting from the reset point,

$$\mathbb{E}_0[\tau] = \frac{\overline{z}^2}{\sigma^2}.$$

The long-run frequency of price adjustment is therefore

$$N(\mu) = \frac{1}{\mathbb{E}_0[\tau]} = \frac{\sigma^2}{\overline{z}(\mu)^2} = \frac{\sigma}{\sqrt{12\psi}} \frac{\sqrt{\mu}}{\mu - 1}. \tag{B.7}$$

Its first derivative is

$$N'(\mu) = -\frac{\sigma}{\sqrt{12\psi}} \frac{\mu + 1}{2\sqrt{\mu}(\mu - 1)^2} < 0,$$

and its second derivative is

$$N''(\mu) = \frac{\sigma}{\sqrt{12\psi}} \frac{3\mu^2 + 6\mu - 1}{4\mu^{3/2}(\mu - 1)^3} > 0.$$

Expected price duration is

$$D(\mu) = \frac{1}{N(\mu)} = \frac{\sqrt{12\psi}}{\sigma} \frac{\mu - 1}{\sqrt{\mu}}.$$

It follows that

$$D'(\mu) = \frac{\sqrt{12\psi}}{\sigma} \frac{\mu + 1}{2\mu^{3/2}} > 0$$

and

$$D''(\mu) = -\frac{\sqrt{12\psi}}{\sigma} \frac{\mu + 3}{4\mu^{5/2}} < 0.$$

These results establish Proposition 3.1.

B.3 General Conditions for the Curvature of the Inaction Region

The CES benchmark provides a particularly simple mapping from the desired markup to profit curvature. This subsection characterizes the more general conditions under which the inaction region and expected price duration are increasing and concave in the markup. The exercise also clarifies the role of the primitives held fixed in the comparative static.

Consider a generic local quadratic loss

$$L(z; \mu) = \frac{1}{2}\kappa(\mu)z^2, \qquad \kappa(\mu) > 0,$$

where $\kappa(\mu)$ is the curvature of the profit loss normalized by frictionless flow profit. Allow both the adjustment cost relative to frictionless profit, $\psi(\mu)$, and the volatility of the desired log price, $\sigma(\mu)$, to depend on the markup.

Repeating the argument in Appendix B.2, the optimal adjustment threshold is

$$\overline{z}(\mu) = \left[\frac{12\psi(\mu)\sigma^2(\mu)}{\kappa(\mu)}\right]^{1/4}. \tag{B.8}$$

Define

$$H(\mu) \equiv \frac{\psi(\mu)\sigma^2(\mu)}{\kappa(\mu)}.$$

The width of the inaction region can then be written as

$$S(\mu) = 2[12H(\mu)]^{1/4}.$$

Differentiation gives

$$S'(\mu) > 0 \quad \Longleftrightarrow \quad H'(\mu) > 0, \tag{B.9}$$

and

$$S''(\mu) = \frac{2(12)^{1/4}}{16H(\mu)^{7/4}}\left[4H(\mu)H''(\mu) - 3H'(\mu)^2\right].$$

Hence,

$$S''(\mu) \leq 0 \quad \Longleftrightarrow \quad 4H(\mu)H''(\mu) \leq 3H'(\mu)^2. \tag{B.10}$$

Equations (B.9) and (B.10) are necessary and sufficient conditions for the inaction-band width to be increasing and concave in the desired markup. In general, the result depends jointly on profit curvature, the adjustment cost relative to profits, and desired-price volatility.

Fixed adjustment cost and desired-price volatility. The benchmark exercise in the main text holds ψ and σ fixed. In this case,

$$H(\mu) \propto \kappa(\mu)^{-1},$$

so

$$S'(\mu) > 0 \quad \Longleftrightarrow \quad \kappa'(\mu) < 0.$$

Thus, the inaction region widens with the markup precisely when normalized profits become locally flatter as the markup rises.

The concavity condition becomes

$$S''(\mu) \leq 0 \quad \Longleftrightarrow \quad 4\kappa(\mu)\kappa''(\mu) \geq 5\kappa'(\mu)^2.$$

Flatter profits are therefore sufficient for the monotonicity result but not, by themselves, for concavity. Concavity additionally restricts how rapidly profit curvature declines with the markup.

Expected price duration depends on a different combination of the same primitives. Since

$$D(\mu) = \frac{\overline{z}(\mu)^2}{\sigma^2(\mu)},$$

equation (B.8) implies

$$D(\mu) = [12J(\mu)]^{1/2}, \qquad J(\mu) \equiv \frac{\psi(\mu)}{\kappa(\mu)\sigma^2(\mu)}.$$

Therefore,

$$D'(\mu) > 0 \quad \Longleftrightarrow \quad J'(\mu) > 0,$$

and

$$D''(\mu) \leq 0 \quad \Longleftrightarrow \quad 2J(\mu)J''(\mu) \leq J'(\mu)^2.$$

When ψ and σ are fixed, these conditions reduce to

$$D'(\mu) > 0 \quad \Longleftrightarrow \quad \kappa'(\mu) < 0$$

and

$$D''(\mu) \leq 0 \quad \Longleftrightarrow \quad 2\kappa(\mu)\kappa''(\mu) \geq 3\kappa'(\mu)^2.$$

The conditions for the curvature of $S(\mu)$ and $D(\mu)$ are therefore distinct. Even with constant σ,

$$D(\mu) = \frac{S(\mu)^2}{4\sigma^2},$$

so concavity of the inaction-band width does not by itself imply concavity of price duration. If desired-price volatility also varies with the markup, its contribution to duration must be taken into account separately, which is the case in Barro (1972).

CES demand. For the CES specification in the main text,

$$\kappa(\mu) = \frac{\mu}{(\mu-1)^2},$$

with

$$\kappa'(\mu) = -\frac{\mu+1}{(\mu-1)^3} < 0$$

and

$$\kappa''(\mu) = \frac{2(\mu+2)}{(\mu-1)^4} > 0.$$

For the width of the inaction region,

$$4\kappa(\mu)\kappa''(\mu) - 5\kappa'(\mu)^2 = \frac{3\mu^2+6\mu-5}{(\mu-1)^6}0, \qquad \mu > 1.$$

Thus, the inaction region is increasing and concave in the desired markup.

For expected price duration,

$$2\kappa(\mu)\kappa''(\mu) - 3\kappa'(\mu)^2 = \frac{\mu + 3}{(\mu - 1)^5} > 0, \qquad \mu > 1.$$

Expected price duration is therefore also increasing and concave. The concavity results in Proposition 3.1 arise from the particular way in which normalized CES profit curvature declines with the desired markup.

These conditions also clarify the comparison with Barro (1972). Varying the slope of Barro's linear demand curve while holding fixed the variance of the additive demand disturbance and the nominal menu cost does not hold fixed either desired-price volatility or the adjustment cost relative to frictionless profits. The resulting comparative-static exercise is therefore different from the one in Proposition 3.1. We make this distinction explicit next.

B.4 Understanding the Barro (1972) Example

The general results in Appendix B.3 show that the relationship between markups and price duration depends on more than the curvature of profits. In particular,

$$D(\mu) = \left[\frac{12\psi(\mu)}{\kappa(\mu)\sigma^2(\mu)}\right]^{1/2}, \tag{B.11}$$

where $\kappa(\mu)$ is normalized profit curvature, $\psi(\mu)$ is the adjustment cost relative to frictionless profits, and $\sigma(\mu)$ is the volatility of the desired log price. A higher markup can therefore lengthen or shorten price duration depending on how these three objects vary with the markup.

The linear-demand example in Barro (1972) provides a useful illustration. Consider

$$Q = \alpha - \beta P + u, \qquad \beta > 0,$$

and normalize the frictionless optimum at $u = 0$ to $P^* = Q^* = 1$. With constant marginal cost, the desired gross markup is

$$\mu = \frac{\beta}{\beta - 1}, \qquad \beta = \frac{\mu}{\mu - 1}.$$

Thus, varying the slope of demand generates variation in the desired markup.

As in our CES benchmark, higher markups make normalized profits flatter. In particular, normalized profit curvature with respect to the log price gap is

$$\kappa_B(\mu) = 2\left(\frac{\mu}{\mu - 1}\right)^2, \tag{B.12}$$

which is decreasing in μ. Taken in isolation, this curvature effect makes firms more willing to tolerate

price gaps and therefore tends to increase price duration.

However, varying β while holding fixed the nominal menu cost and the volatility of the additive demand disturbance changes other primitives of the pricing problem at the same time. Let γ denote the nominal menu cost and suppose $du = \sigma_u \, dB$. Under the normalization above, frictionless operating profit is $\Pi^* = 1/\beta$, so the menu cost relative to frictionless profit is

$$\psi_B(\beta) = \frac{\gamma}{\Pi^*} = \gamma\beta.$$

Moreover, the frictionless optimal price responds to the demand disturbance according to

$$\frac{\partial P^*}{\partial u} = \frac{1}{2\beta}.$$

Since $P^* = 1$ at the normalization point, desired log-price volatility is

$$\sigma_{p^*}(\beta) = \frac{\sigma_u}{2\beta}.$$

Consequently, an increase in the markup generated by a decline in β not only flattens normalized profits, but also changes the adjustment cost relative to profits and raises desired-price volatility. Substituting these objects into equation (B.11) gives

$$D_B(\mu) \propto \left(\frac{\mu}{\mu - 1}\right)^{1/2},$$

so that

$$D_B'(\mu) < 0, \qquad D_B''(\mu) > 0.$$

In this experiment, the increase in desired-price volatility more than offsets the effect of flatter profits, so higher markups are associated with shorter rather than longer price duration.

This comparison illustrates why the objects held fixed in a markup comparative static matter. If instead we use the same linear-demand environment but hold fixed the normalized adjustment cost ψ and desired-price volatility σ_{p^*}, the only remaining effect of the markup operates through profit curvature. Equations (B.12) and (B.11) then imply

$$D_B(\mu) \propto \kappa_B(\mu)^{-1/2} \propto \frac{\mu - 1}{\mu},$$

which is increasing and concave in the markup. Likewise, the width of the inaction region satisfies

$$S_B(\mu) \propto \kappa_B(\mu)^{-1/4} \propto \left(\frac{\mu - 1}{\mu}\right)^{1/2},$$

and is also increasing and concave.

The Barro example therefore provides a useful interpretation of the general conditions in Appendix B.3. The effect of market power on price adjustment is not determined by the markup alone, but by how the economic variation generating the markup changes profit curvature, normalized adjustment costs, and desired-price volatility. Our benchmark comparative static isolates the profit-curvature channel by holding the latter two objects fixed.

B.5 The Stationary Distribution and a Marginal Monetary Shock

Let $a \equiv \overline{z}(\mu)$. Between adjustments, $dz_t = -\sigma\, dB_t$, and the firm resets z to zero whenever the state reaches either $-a$ or a. Away from the reset point, the stationary density $\overline{g}(z)$ satisfies the Kolmogorov forward equation

$$0 = \frac{\sigma^2}{2}\overline{g}''(z).$$

Together with the absorbing-boundary conditions $\overline{g}(-a) = \overline{g}(a) = 0$, symmetry, continuity at zero, and $\int_{-a}^{a} \overline{g}(z)\, dz = 1$, the stationary density is triangular:

$$\overline{g}(z) = \frac{1}{a} - \frac{|z|}{a^2}, \qquad z \in [-a, a].$$

Its derivative away from the reset point is

$$\overline{g}'(z) = \begin{cases} \dfrac{1}{a^2}, & z \in (-a, 0), \\ -\dfrac{1}{a^2}, & z \in (0, a). \end{cases} \tag{B.13}$$

Consider an expansionary monetary shock of size $\delta > 0$ that raises every firm's desired log price by δ. Since $z = p - p^*$, the shock lowers each firm's price gap by δ. Abstracting momentarily from firms that cross an adjustment boundary on impact, the post-shock density is

$$g(z, 0; \delta) = \overline{g}(z + \delta).$$

For a marginal shock,

$$g(z, 0; \delta) - \overline{g}(z) = \delta \overline{g}'(z) + o(\delta). \tag{B.14}$$

It remains to show that boundary crossings are second order in the size of the shock. Following an expansionary shock, the affected firms have pre-shock price gaps

$$z \in [-a, -a + \delta).$$

Their mass is

$$\int_{-a}^{-a+\delta} \overline{g}(z)\, dz = \int_{-a}^{-a+\delta} \frac{a + z}{a^2}\, dz$$

$$= \frac{\delta^2}{2a^2} = O(\delta^2).$$

Immediate adjustments therefore do not contribute to the first-order response. This is the marginal-shock experiment studied by Alvarez and Lippi (2022).

B.6 The Monetary-Shock Impulse Response

We next derive the marginal impulse response using the eigenfunction representation in Alvarez and Lippi (2022). Let

$$q(z) = -z$$

denote the normalized contribution of a firm with price gap z to the first-order real output response. Because the pricing problem and stationary distribution are symmetric, the response can be obtained by following firms until their first price adjustment after the shock.

Consider Brownian motion on $[-a, a]$ with absorbing boundaries. The backward operator is

$$\mathcal{L} = \frac{\sigma^2}{2}\frac{\partial^2}{\partial z^2}.$$

Its orthonormal eigenfunctions are

$$\phi_j(z) = \frac{1}{\sqrt{a}} \sin\left[\frac{j\pi(z+a)}{2a}\right], \qquad j = 1, 2, \ldots,$$

with eigenvalues

$$\lambda_j = -\frac{\sigma^2}{2}\left(\frac{j\pi}{2a}\right)^2 = -\frac{\sigma^2}{a^2}\frac{j^2\pi^2}{8} = -N\frac{j^2\pi^2}{8},$$

where $N = \sigma^2/a^2$ is the steady-state frequency of price adjustment.

For any function $h(z)$, define its projection on the j-th eigenfunction as

$$b_j[h] = \int_{-a}^{a} h(z)\phi_j(z)\, dz.$$

For $q(z) = -z$, direct integration gives

$$b_j[q] = \frac{4a^{3/2}}{j\pi}, \qquad j = 2, 4, 6, \ldots,$$

and $b_j[q] = 0$ for odd j, up to an immaterial normalization of the signs of individual eigenfunctions.

From equation (B.14), the initial perturbation per unit of the monetary shock is $\overline{g}'(z)$. Using

equation (B.13),

$$b_j[\overline{g}'] = \frac{8}{j\pi a^{3/2}} \quad \text{for} \quad j = 2 + 4k, \qquad k = 0, 1, 2, \ldots,$$

and the relevant projection is zero otherwise. Hence,

$$b_j[q]b_j[\overline{g}'] = \frac{32}{(j\pi)^2}$$

for $j = 2 + 4k$.

The marginal output response is therefore

$$\mathcal{R}(t) = \sum_{k=0}^{\infty} \frac{8}{\pi^2(2k+1)^2} \exp\left[-\frac{\pi^2}{2}(2k+1)^2 N t\right]. \tag{B.15}$$

Substituting $N = N(\mu)$ from equation (B.7) gives equation (3.3) in the main text.

At $t = 0$,

$$\mathcal{R}(0) = \frac{8}{\pi^2} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} = \frac{8}{\pi^2}\frac{\pi^2}{8} = 1.$$

The normalized impact response is therefore independent of both N and the desired markup.

For every $t > 0$, equation (B.15) is strictly decreasing in the adjustment frequency:

$$\frac{\partial \mathcal{R}(t; N)}{\partial N} < 0.$$

Since

$$N'(\mu) < 0,$$

the chain rule gives

$$\frac{\partial \mathcal{R}(t; \mu)}{\partial \mu} = \frac{\partial \mathcal{R}(t; N)}{\partial N} N'(\mu) > 0, \qquad t > 0.$$

Thus, higher markups increase the real response at every positive horizon by lowering the frequency of price adjustment. This establishes Proposition 3.2.

B.7 Cumulative Real Effects and the FPA Channel

The role of price-adjustment frequency is particularly transparent when the response is integrated over time. Using equation (B.15),

$$C \equiv \int_0^{\infty} \mathcal{R}(t)\, dt = \sum_{k=0}^{\infty} \frac{8}{\pi^2(2k+1)^2} \frac{2}{\pi^2(2k+1)^2 N} = \frac{16}{\pi^4 N} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^4}.$$

Using

$$\sum_{k=0}^{\infty} \frac{1}{(2k+1)^4} = \frac{\pi^4}{96},$$

we obtain

$$C = \frac{1}{6N}. \tag{B.16}$$

Substituting equation (B.7),

$$C(\mu) = \frac{1}{6} \frac{\sqrt{12\psi}}{\sigma} \frac{\mu - 1}{\sqrt{\mu}} = \frac{\sqrt{\psi}}{\sqrt{3}\sigma} \frac{\mu - 1}{\sqrt{\mu}}.$$

Its first derivative is

$$C'(\mu) = \frac{\sqrt{\psi}}{\sqrt{3}\sigma} \frac{\mu + 1}{2\mu^{3/2}} > 0,$$

and its second derivative is

$$C''(\mu) = -\frac{\sqrt{\psi}}{\sqrt{3}\sigma} \frac{\mu + 3}{4\mu^{5/2}} < 0.$$

Thus, the cumulative real effect of the marginal monetary shock is increasing and concave in the desired markup.

Equations (B.15) and (B.16) also make the FPA channel explicit. Conditional on N, the desired markup enters neither the marginal impulse response nor its cumulative effect:

$$\mathcal{R}(t; \mu) = \mathscr{R}\left(t, N(\mu)\right)$$

and

$$C(\mu) = \frac{1}{6N(\mu)}.$$

Within the canonical model, the effect of the desired markup on marginal monetary transmission therefore operates entirely through its effect on the steady-state frequency of price adjustment.

Finally, equation (B.16) coincides with the sufficient-statistic expression in Alvarez et al. (2016). A firm hitting the lower boundary changes its log price by $+\bar{z}$, while a firm hitting the upper boundary changes its log price by $-\bar{z}$. Conditional on adjustment,

$$\Delta p \in \{-\bar{z}, +\bar{z}\}$$

with equal probability. The kurtosis of the distribution of nonzero price changes is therefore

$$\text{Kurt}(\Delta p) = \frac{\mathbb{E}[(\Delta p)^4]}{\mathbb{E}[(\Delta p)^2]^2} = 1.$$

The sufficient statistic in Alvarez et al. (2016) consequently reduces to

$$\frac{\text{Kurt}(\Delta p)}{6N} = \frac{1}{6N},$$

which is exactly equation (B.16).

B.8 Oligopolistic Competition

The profit-curvature mechanism emphasized in the main text is not specific to monopolistic competition. This subsection shows that the same force arises in a standard oligopolistic environment: greater market power raises firms' desired markups and simultaneously makes normalized profits less sensitive to deviations from the frictionless optimal price.

We follow the nested-CES oligopolistic environment of Atkeson and Burstein (2008). There is a continuum of sectors, each containing a finite number of firms that compete in prices. The elasticity of substitution across sectors is η, while the elasticity of substitution across goods within a sector is θ, with $\theta > \eta > 1$. Let s_i denote firm i's expenditure share within its sector. The elasticity of demand perceived by the firm is

$$\epsilon_i = \theta - (\theta - \eta)s_i. \tag{B.17}$$

The frictionless gross markup is therefore

$$\mu_i = \frac{\epsilon_i}{\epsilon_i - 1}. \tag{B.18}$$

In a symmetric sector with N firms, $s_i = 1/N$, so

$$\mu(N) = 1 + \frac{1}{\theta - (\theta - \eta)/N - 1}.$$

A smaller number of firms raises each firm's market share and hence its desired markup.

We next characterize how market share affects local profit curvature. Let $x_i \equiv \log P_i$, and write firm profits as $\Pi_i = (P_i - W_i)Y_i$. At the frictionless optimum, define normalized profit-loss curvature by

$$\kappa_O(s_i) \equiv -\frac{\partial^2 \Pi_i / \partial x_i^2}{\Pi_i^*}.$$

Using

$$\frac{\partial \log Y_i}{\partial x_i} = -\epsilon_i$$

and the firm's first-order condition, straightforward differentiation yields

$$\kappa_O(s_i) = \epsilon_i(\epsilon_i - 1) + \frac{\partial \epsilon_i}{\partial x_i}.$$

Within-sector expenditure shares satisfy

$$\frac{\partial s_i}{\partial x_i} = -(\theta - 1)s_i(1 - s_i).$$

Combining this expression with equation (B.17),

$$\frac{\partial \epsilon_i}{\partial x_i} = (\theta - \eta)(\theta - 1)s_i(1 - s_i).$$

Hence,

$$\kappa_O(s) = \left[\theta - (\theta - \eta)s\right]\left[\theta - (\theta - \eta)s - 1\right] + (\theta - \eta)(\theta - 1)s(1 - s).$$

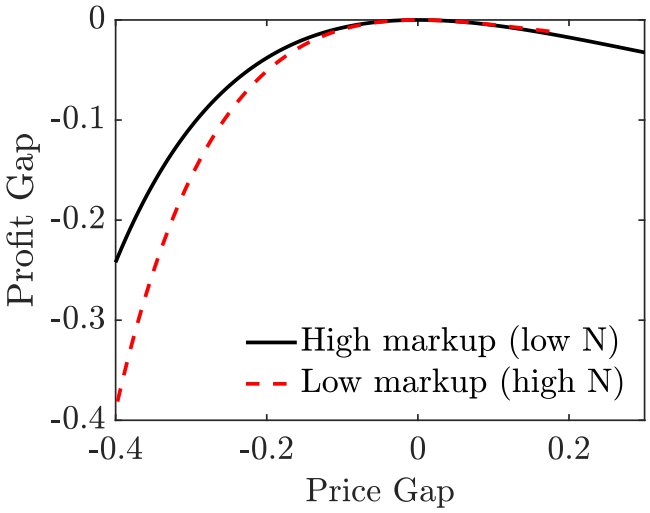
Differentiating with respect to the firm's market share gives

$$\frac{d\kappa_O(s)}{ds} = -(\theta - \eta)\left[\theta + 2(\eta - 1)s\right] < 0.$$

Thus, a larger market share lowers the local curvature of normalized profits. Since a larger market share also raises the desired markup through equations (B.17) and (B.18), greater market power is associated with flatter normalized profits, exactly as in the monopolistic-competition benchmark.

Figure B.1 illustrates this result for symmetric sectors with $N = 3$ and $N = 6$. The sector with fewer firms has the higher desired markup and the flatter profit function around the frictionless optimum.

Figure B.1: Market Power and Profit Curvature



Note: The figure plots firm profits as a function of the price gap under oligopolistic competition for a high-market-power case ($N = 3$) and a low-market-power case ($N = 6$). All other parameters are held fixed across the two cases. The price gap is the deviation of the firm's log price from its frictionless optimal log price. The sector with fewer firms has a higher desired markup and a flatter profit function around the optimum.

The oligopoly example shows that the mechanism in Section 3 does not depend on monopolistic competition itself. What matters is the relationship between market power and the curvature of

normalized profits. Whenever greater market power makes profits less sensitive to price gaps, a fixed cost of price adjustment gives firms a stronger incentive to tolerate price misalignment, lowering endogenous adjustment frequency and strengthening the propagation of nominal disturbances.

C. Appendix for Quantitative Results

C.1 Intermediate Inputs and Roundabout Production

This appendix extends the baseline model to incorporate intermediate inputs through roundabout production. As emphasized by Nakamura and Steinsson (2010), intermediate inputs generate strategic complementarity in price setting because firms' marginal costs depend on prices set by other firms. In addition, when differentiated goods are themselves used as inputs, markups become embedded in production costs and accumulate along the production chain, as in the cascading-markup mechanism of Baqaee (2018).

Production with intermediate inputs. We replace the labor-only technology in the baseline model with

$$y_{ikt} = A_{ikt} \left(\frac{\ell_{ikt}}{1-\alpha} \right)^{1-\alpha} \left(\frac{m_{ikt}}{\alpha} \right)^{\alpha},$$

where m_{ikt} is a composite intermediate input and $\alpha \in [0, 1)$ is its cost share. We assume that intermediate inputs are aggregated using the same economy-wide composite as final consumption, so that

$$P_t^M = P_t.$$

This parsimonious roundabout-production structure captures the main strategic-complementarity mechanism without introducing a full input-output network.

Cost minimization implies

$$MC_{ikt} = A_{ikt}^{-1} W_t^{1-\alpha} P_t^{\alpha},$$

or, in logs,

$$mc_{ikt} = (1-\alpha) w_t + \alpha p_t - a_{ikt}.$$

The frictionless desired price is therefore

$$p_{ikt}^{*} = \log \mu_{kt} + (1-\alpha) w_t + \alpha p_t - a_{ikt}.$$

The term αp_t creates strategic complementarity: when other firms' prices adjust sluggishly, marginal costs and desired prices also respond more slowly, weakening firms' incentives to adjust their own prices.

Roundabout production also amplifies markup distortions. In a symmetric flexible-price economy,

$$P = \mu A^{-1} W^{1-\alpha} P^{\alpha},$$

which implies

$$\frac{P}{W} = \left(\frac{\mu}{A}\right)^{\frac{1}{1-\alpha}} \qquad \text{and} \qquad \frac{\partial \log(P/W)}{\partial \log \mu} = \frac{1}{1-\alpha} > 1.$$

Thus, markups are effectively compounded through successive rounds of production.

Calibration. Following Nakamura and Steinsson (2010), we recalibrate the pricing parameters after introducing intermediate inputs. Holding the baseline pricing parameters fixed would generally change both the frequency and size of price adjustments and would therefore confound the propagation effect of intermediate inputs with a change in the model's fit to micro pricing behavior.

For each sector k, we choose the menu-cost parameter and the volatility of idiosyncratic productivity,

$$\psi_k^I \qquad \text{and} \qquad \sigma_{A,k}^I,$$

to match the same sectoral frequency of price adjustment and average absolute size of price changes as in the baseline calibration. All other structural parameters and empirical inputs, including sectoral markups, expenditure weights, preferences, and the aggregate shock process, are unchanged. The recalibration is conducted at the baseline calibration point, after which these parameters are held fixed in the historical experiments from 1980 to 2016.

This procedure keeps the two models comparable along the microeconomic dimensions that discipline nominal rigidity and isolates the additional propagation generated by roundabout production.

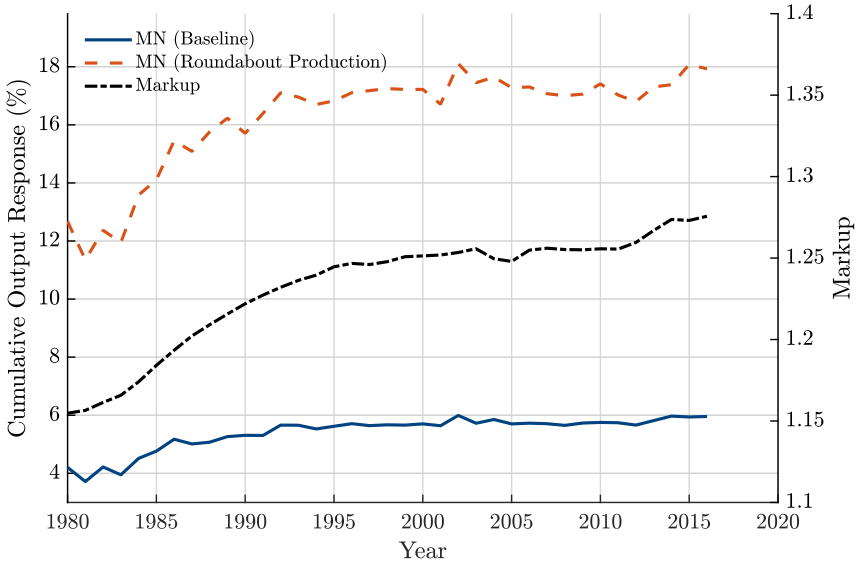
Quantitative results. Figure C.1 shows that intermediate inputs substantially increase the *level* of monetary non-neutrality. Across the sample, the cumulative real response is approximately three times as large as in the baseline model. The amplification reflects strategic complementarity: sluggish input prices make marginal costs and desired prices respond more slowly to a monetary shock, reducing firms' incentives to adjust and increasing the persistence of the real response.

Despite this large level effect, intermediate inputs have little effect on the proportional increase in monetary non-neutrality associated with rising markups. In particular,

$$\frac{MN_{2016}^I}{MN_{1980}^I} \approx \frac{MN_{2016}^B}{MN_{1980}^B},$$

where I and B denote the intermediate-input and baseline economies. Since both economies face the same historical evolution of sectoral markups, their implied markup elasticities of monetary non-neutrality are also very similar. Thus, roundabout production mainly shifts monetary non-neutrality upward rather than materially changing its growth over the sample.

Figure C.1: Monetary Non-neutrality over Time: Adding Intermediate Inputs



Note: The solid line plots model-implied monetary non-neutrality in the baseline model. The dashed line reports the corresponding series in the model with roundabout production. The remaining line reports the aggregate markup.

Two offsetting forces help explain this result. First, intermediate inputs strengthen the endogenous-frequency channel. If a higher markup lowers price-adjustment frequency in one sector, its price becomes more sluggish. Because this price enters the intermediate-input bundle, marginal costs in other sectors also become less responsive, further reducing their incentives to adjust. Roundabout production therefore propagates markup-induced nominal rigidity across sectors.

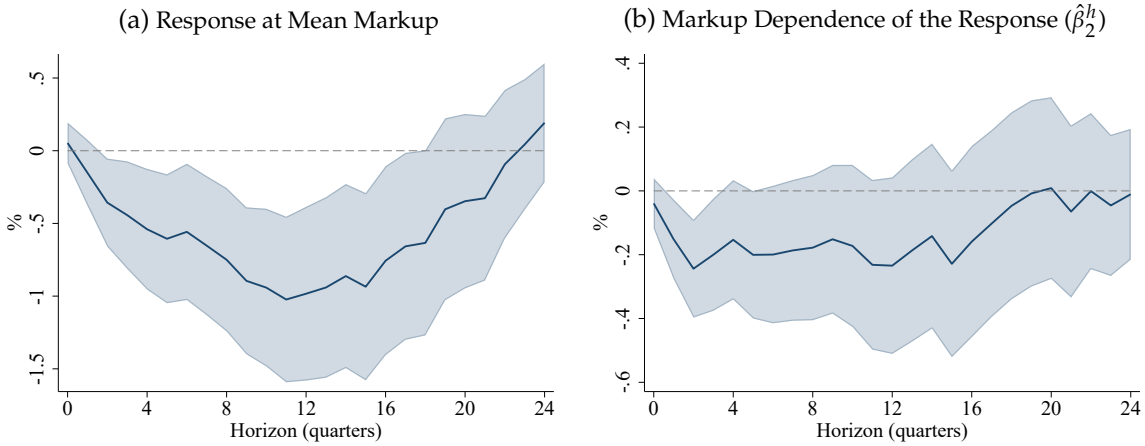
Second, roundabout production raises the effective markup distortion at both the beginning and the end of the sample. Because monetary non-neutrality is concave in market power in the quantitative model, moving both economies into a region with a larger effective markup wedge reduces the marginal effect of further markup increases. This concavity offsets the stronger cross-sector propagation generated by intermediate inputs.

The recalibration further removes any mechanical change in nominal rigidity caused solely by changing the production technology. Quantitatively, these forces imply a large amplification of monetary non-neutrality in levels but little change in its elasticity with respect to the historical rise in markups.

The main quantitative conclusion is unchanged with intermediate inputs. Roundabout production substantially increases the overall degree of monetary non-neutrality, but it does not materially alter the contribution of rising markups to the evolution of monetary transmission.

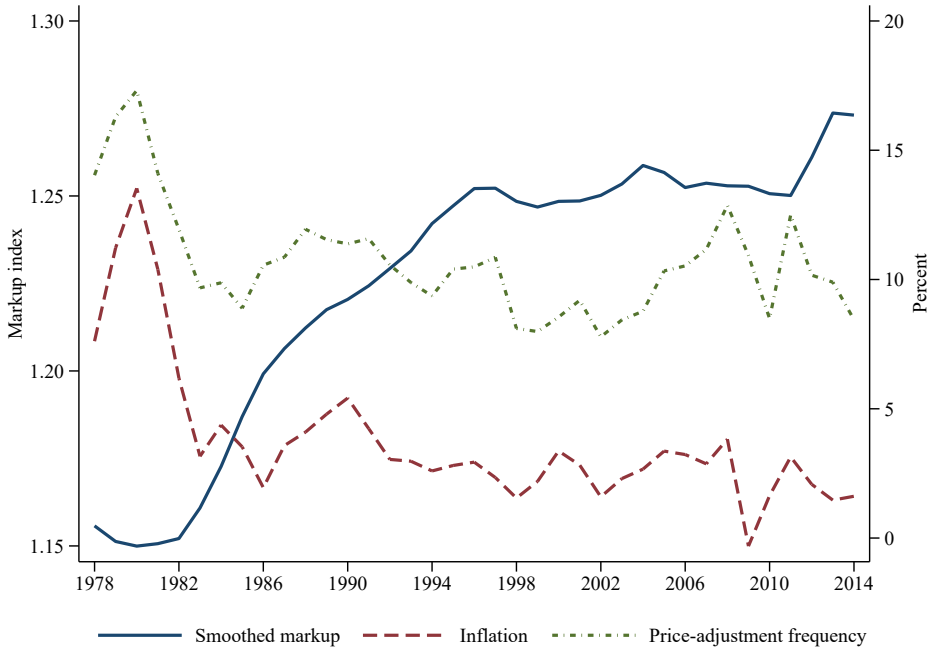
D. Figures and Tables

Figure D.1: Markup-Dependent Effects of Contractionary Monetary Policy Shocks (MR shock)



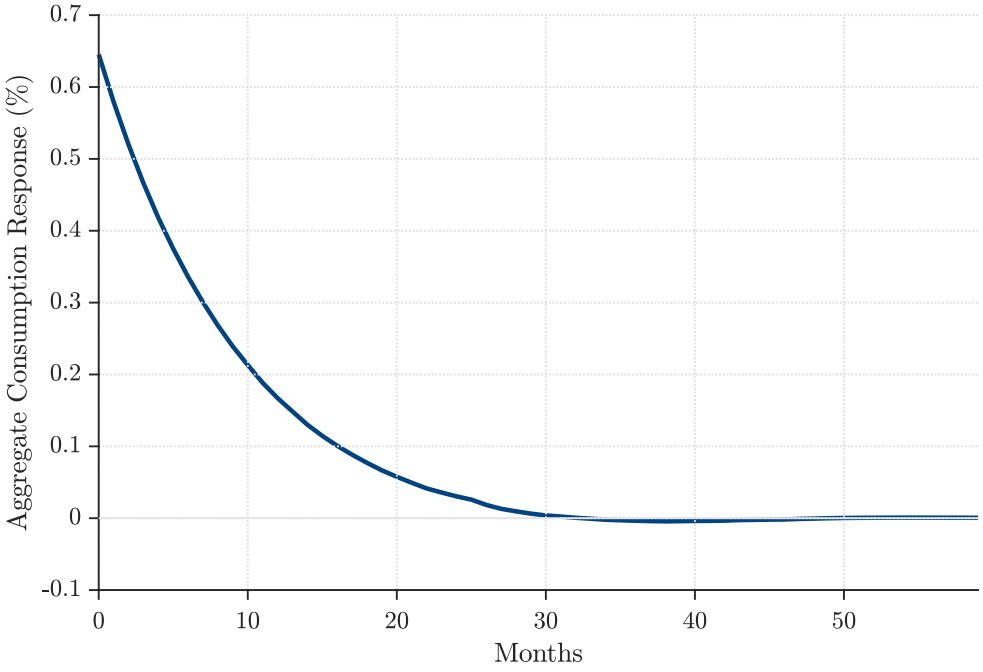
Notes: This figure reports estimates of the markup-dependent response of real GDP to a contractionary monetary policy shock. Panel (a) plots $\hat{\beta}_1^h + \hat{\beta}_2^h \overline{\mu^{ss}}$, the estimated response evaluated at the sample mean of the smoothed markup. Panel (b) plots $\hat{\beta}_2^h$, which measures how the response varies with the markup state. The coefficients are estimated from equation (2.1). The monetary policy shock ϵ_t^m follows the identification strategy of Miranda-Agrippino and Ricco (2021), with positive values corresponding to contractionary innovations. The shaded areas indicate 90 percent confidence intervals constructed using Newey–West standard errors.

Figure D.2: Markups, Inflation and Frequency of Price Adjustment



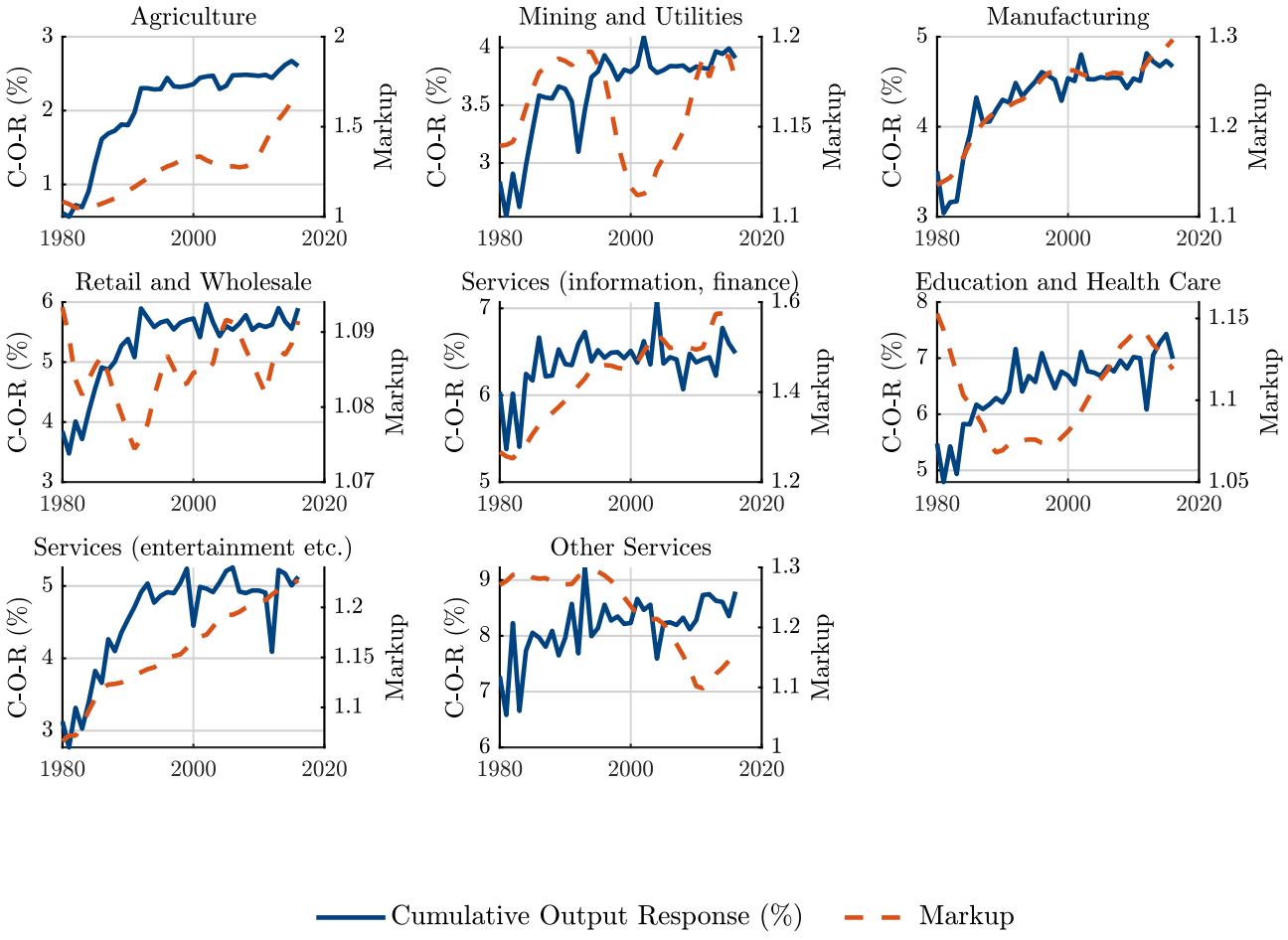
Notes: The left axis shows the smoothed aggregate markup index. The right axis shows annual inflation and the median frequency of price adjustment, in percent. Inflation is the calendar-year mean of quarterly CPI inflation observations. Sample: 1978–2014.

Figure D.3: The IRF of Aggregate Consumption to Monetary Policy Shocks



Notes: This figure shows the impulse response of aggregate consumption to a nominal-spending shock that raises nominal spending by 1 percent on impact.

Figure D.4: Sectoral Cumulative Output Response Over Time (eight sectors)



Note: This figure plots the evolution of U.S. sectoral markups (dashed red lines) and the corresponding sectoral monetary non-neutrality implied by the eight-sector model (solid blue lines).

Table D.1: Data and Model Moments

Moment	Agriculture	Mining and Utilities	Manufacturing	Retail and Wholesale	Services (information, finance)	Education and Health Care	Services (entertainment etc.)	Other services
Data Moments								
Moments Targeted								
Frequency	0.099	0.132	0.079	0.035	0.063	0.027	0.049	0.025
Absolute size (median)	0.154	0.068	0.081	0.046	0.041	0.076	0.056	0.069
25th percentile size	0.095	0.014	0.039	0.030	0.010	0.035	0.029	0.054
75th percentile size	0.382	0.125	0.157	0.098	0.091	0.116	0.096	0.128
Moments Not Targeted								
Size of price increase	0.236	0.062	0.076	0.062	0.052	0.104	0.069	0.105
Size of price decrease	0.272	0.064	0.097	0.075	0.062	0.133	0.107	0.142
Model Moments								
Moments Targeted								
Frequency	0.119	0.132	0.082	0.041	0.062	0.035	0.048	0.022
Absolute size (median)	0.201	0.066	0.081	0.050	0.042	0.038	0.059	0.082
25th percentile size	0.083	0.014	0.026	0.024	0.012	0.083	0.023	0.049
75th percentile size	0.383	0.151	0.155	0.076	0.091	0.104	0.099	0.125
Moments Not Targeted								
Size of price increase	0.214	0.076	0.087	0.053	0.052	0.076	0.065	0.090
Size of price decrease	0.296	0.156	0.133	0.027	0.068	0.075	0.062	0.021