Inflation and Competition in an Old Keynesian Model

Yang Liu *
Boston College

November 11, 2026

[Incomplete and Preliminary]

Abstract

Standard competition models underlying the New Keynesian Phillips curve assume posted prices instantaneously clear all demand. Yet product unavailability is common in reality, and competitors can undercut only when their products are available. I formalize this insight in a shopping-time model with rationing. Nominal rigidities prevent firms from clearing excess demand and incentivize rationing. Rationing, particularly by low-price firms, raises effective search costs and households' reservation prices. When aggregate shocks push the economy into excess demand, increased rationing lowers residual demand elasticity and attenuates strategic complementarities. On the one hand, this weakens real rigidities and steepens the traditional Phillips curve, amplifying the transmission of marginal costs to inflation. On the other hand, desired markups rise endogenously, generating a novel form of excess-demand-driven inflation that operates independently of the marginal cost. In a quantitative model calibrated to U.S. data, excess demand generates substantial asymmetries in both the magnitude and persistence of inflation that are not confined to large shocks. Moreover, excess-demand-driven inflation implies that monetary policy effectiveness depends less on the slope of the Phillips curve, and that "looking through" transitory shocks overlooks the anti-competitive externality of excess demand.

JEL classification: E12, E31, D45
Keywords: Endogenous competition, Inflation, Old Keynesian, Phillips curve, Rationing

1 Introduction

In standard New Keynesian (NK) models, the effect of firm-level price stickiness on the aggregate price level relies on two key ingredients: imperfect price competition (Blanchard and Kiyotaki, 1987) and quasi-Walrasian market clearing (Calvo, 1979).[1] In this paper, I relax both assumptions. Firms may ration by restricting quantities, and imperfect competition is shaped by

*Email: yang.liu.7@bc.edu. Department of Economics, Boston College. I thank Susanto Basu, Ryan Chahrour, Joel Flynn, Tom Holden, Peter Ireland, Robert Ulbricht, Rosen Valchev, and participants at Boston College, the Fridays at the Boston Fed workshop, the Econometric Society NASM, the SED Meeting, the Inflation: Drivers and Dynamics Conference for helpful discussions. All errors are my own.

[1] Quasi-Walrasian market clearing assumes that, despite pricing frictions, firms must instantaneously satisfy all demand at their posted prices.

both price and availability dimensions. The interaction between rationing and imperfect competition turns out to play an important role in shock propagation. As rationing disrupts price competition, excess demand generates a steep Philips curve and persistently high inflation.

Non-clearing markets have important empirical relevance. During the COVID-19 pandemic, the U.S. experienced widespread and persistent shortages not seen for decades (Caldara et al., 2025), and a growing literature documents a strong time-series relationship between inflation and various measures of shortages (Lamont, 1995, Cavallo and Kryvtsov, 2023, Caldara et al., 2025, Bai et al., 2024). Shortages are also not limited to the COVID episode. Bils (2016) calculates an average stockout rate of 4.6% using the CPI micro data between 1988 and 2009. Cavallo and Kryvtsov (2023) find that online retailers had an average stockout rate of 10% even before COVID. According to Business Tendency Surveys from Denmark, 10.2% of firms in the service sector reported labor shortages between 2011 and 2019, while 6.3% of firms in the manufacturing sector reported shortages of labor, materials, or equipment between 1990 and 2019. Going forward, as rising geopolitical tensions continue to raise trade barriers and disrupt international trade, shortages may persist, particularly in sectors that rely on complex global supply chains.

New Keynesian models, however, are ill-suited to analyze these phenomena. As Barro (2025) points out, these models extend the excess supply case by imposing market clearing, and their theoretical coherence deteriorates when excess demand dominates. The strong non-rationing assumption, as I discuss below, is empirically counterfactual. Even if one sets aside the theoretical inconsistency and approximates rationing by reducing the number of firms, existing models remains inadequate. Monopolistic models take market structure as given, while oligopolistic ones tend to flatten the marginal-cost Phillips curve because higher concentration always increases strategic complementarities (Wang and Werning, 2022). By contrast, several recent studies find higher marginal cost pass-through during the COVID period (e.g., Amiti et al., 2022; Chin, 2023; Gagliardone et al., 2024), whereas none reports evidence of flattening. The difficulty of incorporating rationing into existing models highlights the need for a new theoretical framework that departs from standard NK models in its treatment of both market clearing and imperfect competition.

Theory. The theoretical analysis is motivated by the simple intuition that competitors can undercut only when they are selling. To formalize the interactions between rationing and competition, I depart from the New Keynesian benchmark in two respects. First, as in Old Keynesian models, I allow firms to ration and optimize over both prices and quantities. Second, I microfound product-market competition through search frictions to deliver a more realistic market structure. Such frictions are empirically relevant in a wide range of markets, from household consumption to firm-to-firm trade (Petrosky-Nadeau et al., 2016; Krolikowski and McCallum, 2021; Lenoir et al., 2023). In my model, each buyer pays a fixed cost to randomly meet a new seller and follows a reservation price rule à la McCall (1970). Meanwhile, sellers are not obligated to accept all buyers and may ration by randomly rejecting some of them. Thus, a successful match requires two conditions: (i) the seller's posted price is below the buyer's reservation price, and (ii) the seller accepts the buyer, i.e., the buyer is not rationed by the seller.[2]

[2]Once a match is formed, the seller sells any quantity demanded at the posted price. This is possible because each buyer is measure-zero in the model, so a single buyer's demand does not affect firm-level quantity choices. The absence of intensive-margin rationing greatly improves tractability, as it eliminates the need to model the combinatorial problem of purchasing from multiple sellers.

The resulting competition mechanism is analogous to the Salop (1979) model: sellers' market power arises from the possibility that buyers switch to other sellers. Rationing, particularly by low-price competitors, discourages buyers from continuing to search and increases the market power of all firms.

Formally, my model yields two analytical results. First, Proposition (1) shows that competitor rationing attenuates strategic complementarities: the effects of a competitor's posted price on other firms' residual demand and demand elasticity are proportional to the competitor's acceptance rate. The acceptance rate is defined as the ratio of quantity supplied to quantity demanded and is therefore one minus the rationing rate. Consistent with the economic intuition above, if a competitor rations completely (i.e., has a zero acceptance rate), changes in its posted price have precisely zero effects on other firms.

Second, I examine an analytically tractable special case with uniformly distributed search costs. In Proposition (2), I derive three sufficient statistics summarizing the extensive margin of demand, i.e., the number of buyers willing to purchase from the seller. The first statistic is the average acceptance rate (or, equivalently, the average non-rationing rate). The second is the *competitor price gap*, defined as the difference between the seller's own price and an acceptance-weighted competitor price index for its product market. The third is the *cross-market price gap*, defined as the difference between the acceptance-weighted competitor price index and the aggregate price level. Because the competitor price index is acceptance-weighted, it depends not only on competitors' average price changes but also on the covariance between acceptance rates and price changes, which captures which competitors tend to ration more. In the spirit of Amiti et al. (2019), I use these sufficient statistics to characterize the implications of rationing for endogenous demand elasticity, markups, and cost pass-through.

Neither result can be derived in a model with only rationing or only search frictions. In the class of monopolistic Old Keynesian models (e.g., Holden, 2024; Flynn et al., 2026), competitor interactions are absent by construction, and rationing by other firms is equivalent to an upward shift in aggregate demand. In models with only search frictions, the residual demand curve is endogenous but the availability dimension is absent. Although market tightness may be used to approximate rationing in DMP models (e.g., Bianchi et al., 2024), the matching function assumption again mutes competitor interactions, leaving competition exogenously determined.

In principle, my model features three distinct aggregate price indices. The first is the *consumption price index*, which measures the marginal monetary cost of the consumption bundle. The second is the *acceptance-weighted average posted price*. The third is the unweighted average of all firms' posted prices. The first two reflect the prices faced by buyers, whereas the third provides a direct measure of firms' price-setting behavior. I maintain a clear distinction among these three measures throughout the remainder of the paper.

Excess-demand-driven inflation. The theoretical framework above applies whenever rationing occurs, regardless of its underlying cause. In my general equilibrium model, I choose to generate rationing through sticky prices, a standard mechanism in the Old Keynesian literature. Specifically, I assume Calvo pricing. As Holden (2024) shows, rationing remains quantitatively important in more sophisticated menu-cost models.

To fix ideas, consider a negative productivity shock. According to the textbook New Keynesian Phillips curve, a decline in productivity raises real marginal costs and, consequently,

firms' reset prices. With greater price stickiness, fewer firms can reset their prices, mechanically dampening the inflation response to the shock. I refer to this traditional mechanism as "cost-driven inflation."

In my rationing model, price stickiness has a different implication. Firms that do not reset their prices increase rationing when their markups begin to turn negative, as implied by profit maximization. Greater price stickiness thus means more non-resetting firms and, consequently, more rationing in the economy. Moreover, the price-rationing covariance changes, as non-resetters are likely to have lower prices than resetters following an inflationary shock. In equilibrium, competition softens and residual demand elasticity falls, which serves as an endogenous markup shock in the traditional Phillips curve. As a result, reset prices increase with the degree of price stickiness, partially offsetting its dampening effect on inflation. I refer to this new mechanism as "excess-demand-driven inflation."

The novel "excess-demand-driven inflation" has important implications for monetary policy. First, because greater price stickiness dampens cost-driven inflation yet amplifies excess-demand-driven inflation, the slope of the traditional Phillips curve is a poor guide to inflation under excess demand. Consequently, the sacrifice ratio depends less on nominal rigidities under excess demand. By curbing excess demand, monetary tightening reduces rationing and restores price competition, lowering inflationary pressure directly. Second, "looking through" transitory shocks overlooks the anti-competitive externality of excess demand, as rationing by firms with outdated prices keeps competition weak and inflation elevated long after the shock has dissipated.

Quantitative results. In the third part of the paper, I examine the importance of rationing in a quantitative model. Following Aruoba et al. (2024), I calibrate idiosyncratic shocks to match the estimates by Foster et al. (2008). I then discipline the search cost distribution, assumed to be normally distributed and truncated at zero, using two moments. First, the empirical estimate of strategic complementarities in Gagliardone et al. (2023) identifies the dispersion of search costs. Second, the average steady-state markup pins down their mean. While there is little consensus on the average level of markups, I deliberately choose a relatively high target—a cost-weighted average of 29%—to emphasize the robustness of my results. High markups reduce the need for rationing. Consequently, in steady state firms accept 97.9% of customers, implying an average rationing rate of only 2.1%.

Despite its low steady-state level, rationing induces a pronounced asymmetry that is not confined to large shocks. For positive demand shocks, the consumption price inflation on impact is nearly 50% higher, compared to the deflation upon negative demand shocks. Internal persistence is likewise asymmetric: the autoregressive coefficient of inflation responses is over 0.1 higher following positive demand shocks than following negative ones, despite identical shock persistence. The intuition is that as the price level rises, firms with outdated prices continue to experience declining relative prices and are more likely to ration. While shocks are mean-reverting, price levels are not, generating a highly persistent feedback loop between rationing, relative price gaps, and inflation. Finally, the real wage stagnates after positive demand shocks even in a flexible wage model as labor demand is capped from above once firms decide to ration.

The general equilibrium model provides a qualitative account of several salient empirical

patterns observed during the surge in inflation between 2021 and 2023, which have been widely discussed in both academic and policy circles. First, the asymmetric nature of rationing gives rise to an asymmetric Phillips curve, consistent with the empirical observation of Benigno and Eggertsson (2023) and Ball et al. (2022). Second, post-COVID inflation (as well as inflation in the 1970s) is characterized not only by high levels but also by substantial persistence, consistent with my model's asymmetric persistence. Even though NK models with menu costs or capacity constraints can match the magnitude of responses, their shock propagation is typically less persistent. Third, as reviewed by Mongey (2025), the post-COVID economy is characterized by surprisingly low real wages. My model generates high inflation and depressed real wages under excess demand without relying on wage dynamics.

Empirical evidence. I use product-level data from the Indian Prowess database. Relative to other datasets, Prowess offers a unique panel comprising (i) material input prices and quantities, (ii) output prices and quantities, and (iii) product-level capacity utilization.

Before testing my main theory, I first examine whether capacity utilization serves as an appropriate proxy for rationing. To this end, I compare the responses of firms with different levels of capacity utilization to industry demand shocks. Following Giovanni et al. (2014), I proxy demand shocks using median sales growth at the 6-digit industry-year level. As expected, high-utilization (low-capacity) firms experience much smaller increases in output than low-utilization firms. However, they do not set higher prices relative to their low-utilization peers. In the absence of rationing, a relatively smaller increase in output must be accompanied by a relatively larger increase in price. The data clearly indicate otherwise, suggesting that capacity utilization is indeed a valid proxy for rationing in this dataset.[3]

Turning to strategic complementarities, I find strong empirical support for Proposition (1). In the data, firms' own price changes are significantly less correlated with price changes by low-capacity competitors, which are more likely to ration. Moreover, this finding is theoretically impossible in standard (i.e., non-rationing) Bertrand-Nash models, in which competitors' prices are sufficient statistics for competitive interactions. In these models, once controlling for competitors' price changes, the covariance between their capacity utilization and price changes should contain no additional information about firms' own price changes.

Lastly, I examine whether competitor rationing affects a firm's own cost pass-through. Theoretically, cost pass-through decreases in the own-price markup elasticity, defined as the elasticity of a firm's desired markup with respect to its own posted price. However, my theory does not yield an unambiguous prediction because the effect depends on the parameterization, making this ultimately an empirical question. In the data, I find that firms with more low-capacity competitors exhibit substantially higher cost pass-through without corresponding output losses. This finding suggests that competitor rationing weakens the own-price markup elasticity, thereby raising cost pass-through.

The implications of my competition test extend beyond capacity constraints. In practice,

[3]More generally, it is irrelevant for my test whether high-utilization firms raise prices or not, so long as prices increases fall short of what explains the quantity differences under a plausible range of residual demand elasticities. To overturn my interpretation, one would need to argue that high-utilization firms are subject to unobserved negative demand shocks or face more elastic demand. Both hypotheses are opposite of the common intuition that high-utilization firms are indeed more popular firms, and I find empirical support for neither.

firms may restrict quantities, voluntarily or otherwise, in response to a variety of factors, such as delivery delays, inventory stockouts, and other supply chain disruptions. I leave the investigation of these other factors for future research.

Literature. First, the paper contributes to the literature on rationing and non-clearing markets in macroeconomics, which dates back to Barro and Grossman (1971) and Keynes (1936) if not earlier.[4] Widespread stockouts and delivery delays during Covid renewed interest in rationing (Holden, 2024; Kharroubi and Smets, 2024; Flynn et al., 2026; Markauskas et al., 2026). Under monopolistic competition, Holden (2024) quantitatively analyzes how rationing affects the supply curve and consumer welfare, together with empirical tests for rationing at both micro and macro levels. Relatedly, Flynn et al. (2026) develop a theoretical model of rationing where both prices and quantities of monopolistic firms are pre-determined.

I extend their insights and show that rationing alters not only the supply curvature but also *the demand curvature*, an effect ruled out by construction in monopolistic settings.[5] For instance, monopolistic models are too stylized to explain why cost pass-through, one of the key determinants of the Phillips curve, increases when competitors ration. Second, in Holden (2024) and Flynn et al. (2026), households either forgo the rationed variety or passively accept the quantity constraint imposed by firms. In reality, households are more likely to visit multiple stores to circumvent such constraints, and search frictions appear much more realistic in this regard. Similarly for intermediate goods, search models are widely used to study firm-to-firm trade (Lenoir et al., 2023).

A closely related body of literature builds on aggregate matching frictions (e.g., Michaillat and Saez, 2015; Bianchi et al., 2024). Although aggregate matching functions can approximate aggregate rationing to some extent, they effectively assume monopolistic firms and abstract from the competitive interactions that are central to this paper. They also mute the covariance between prices and rationing, a channel that is no less important than the average level of rationing itself. Finally, compared to shopping time models that focus on households (e.g., Kaplan and Menzio, 2016), I demonstrate the importance of rationing and availability on the firm side.

Third, this paper contributes to the vast literature on imperfect competition and New Keynesian models (Rotemberg and Saloner, 1986; Rotemberg and Woodford, 1992; Kimball, 1995; Carlton, 1998). Following recent advances in international trade (Atkeson and Burstein, 2008; Amiti et al., 2019), there has been renewed interest in incorporating non-monopolistic competition into the NK framework (Wang and Werning, 2022; Alvarez et al., 2023; Gagliardone et al., 2023; Fujiwara and Matsuyama, 2026). My search model provides a unified framework where real rigidities endogenously evolves as the economy moves from excess supply to excess demand, a factor missing from both NK and OK models. Unsurprisingly, different micro-

[4] Many influential studies examined rationing prior to the New Neoclassical Synthesis. To name just a few, see Carlton (1977), Barro and Grossman (1974), Bénassy (1975), Svensson (1980), and Ohanian and Stockman (1994). While rationing largely disappears in modern New Keynesian models, it is, however, not unfamiliar in other fields such as finance (Jaffee and Modigliani, 1969) and industrial organization (Dana, 1999; Dana, 2001; Matsa, 2011; Conlone and Mortimer, 2013; Amberg et al., 2025; Li, 2026).

[5] Strictly speaking, residual demand shifts outward when competitors ration in Holden (2024) and Flynn et al. (2026). However, competition and the Phillips curve slope are determined not by the level but by the curvature of the residual demand curve.

foundations lead to different results and policy implications. In my model, lower competition due to rationing steepens the Phillips curve, amplifies inflationary shocks, but also reduces the sacrifice ratio for monetary policy. In oligopolistic NK models, however, lower competition always flattens the Phillips curve and raise the sacrifice ratio (Wang and Werning, 2022; Fujiwara and Matsuyama, 2026). Moreover, my model identifies a novel negative externality of excess-demand-driven inflation: the disruption of price competition (see Carlton, 1982 for an IO perspective). For central banks, this negative externality introduces a new cost of tolerating high inflation.

Empirically, to my best knowledge, my paper is the first in the macro literature to examine the competition effects of rationing at the micro level. Few macro papers have taken rationing seriously at the micro level. One exception is Holden (2024), who astutely tests the quantity response of retail products that are likely under-priced and finds evidence of rationing.[6] Together with my new theory, the unique availability of all data from inputs to outputs enables me to examine how rationing alters both strategic complementarities and cost pass-through.

The final contribution pertains to the non-linearity of the Phillips curve and inflation dynamics. A vast number of recent papers attempt to understand this non-linearity through the functional form of nominal rigidities (e.g., Gagliardone et al., 2024, Blanco et al., 2024a, Blanco et al., 2024b). Several others study the role of occasionally binding capacity constraints (e.g., Comin et al., 2023) and non-linear wage setting (e.g., Benigno and Eggertsson, 2023). Not all channels necessarily hold once rationing is allowed. In menu cost models, rationing attenuates the cost of underpricing because firms can stop selling when their prices lag far behind the aggregate price level.[7] Similarly, rationing reduces the need for higher prices to clear demand when capacity constraints bind or when the labor supply curve shifts considerably. By contrast, the endogenous competition channel in my model offers a new source of non-linearity that reconciles firm-level dampening and industry-level amplification.

The rest of the paper is organized as follows. Section 2 presents the core theory. Section 3 describes the general equilibrium model. Section 4 presents quantitative results, and Section 5 shows empirical evidence from micro data. Section 6 concludes.

2 Incorporating Rationing into Firms' Residual Demand

To formalize the interaction between rationing and competition, I begin by deriving the residual demand curve in a static partial equilibrium setting. I keep the supply side as given to focus on the demand side. The resulting insights are key drivers of the complete dynamic model developed in Section 3.

Regarding notation, uppercase and lowercase variables denote levels and logarithms, respectively, unless otherwise noted. Bold variables denote aggregate variables. All prices in this

[6] A separate strand of the literature examines capacity constraints. Both Balleer and Noeller (2023) and Ahlander et al. (2026) use firm-level data on capacity constraints, substantially extending earlier sector-level studies such as Boehm and Pandalai-Nayar (2022). Unfortunately, neither contains any test related to rationing, which limits the comparability between their empirical results and my theoretical predictions.

[7] Without rationing, the cost of underpricing is approximately a quadratic function of the price gap. With rationing, the cost becomes linear in the price gap, which substantially weakens the incentive to pay menu costs.

section are expressed in nominal terms, and **P** denotes the aggregate price index.

2.1 Search Frictions in a Single Product Market

Below I describe the micro-level structure of a single product market. A given product market is populated by a unit mass of firms (goods producers), indexed by s, and a unit mass of households, indexed by k.

Household demand. Each household wants to purchase Q units of goods in that market, which I endogenize in Section 3. To make the purchase, a household randomly meets firms until (i) she finds one whose price is below her reservation price (explained below), and (ii) the firm accepts new customers. Matches last one period.

Prices and quantities. Each firm s sells homogeneous goods and posts a unit nominal price P_s. I define $F(P)$ and $f(P)$ as the cumulative distribution function (CDF) and probability density function (PDF) of prices posted by firms. However, firms do not commit to supplying *all* households at their posted prices. Instead, they may endogenously impose production limits and ration supply for various reasons. As microfounded in Section 3.2 and widely analyzed in the Old Keynesian literature, nominal rigidities may prevent firms from clearing excess demand, thereby leading to rationing.

I assume that firms ration at the extensive margin by randomly rejecting households. Once matched, a household can purchase any quantity from the firm at the posted price. There is no need for intensive-margin rationing because each household is measure-zero in the model, so an individual buyer's demand does not affect firm-level variables.[8]

From a household's perspective, extensive-margin rationing by firms is equivalent to having a probabilistic acceptance rate. Let Ω_s denote the probability of acceptance by firm s. Equivalently, $1-\Omega_s$ is the probability of rationing by firm s. Then, the firm side is characterized by the joint distribution of two variables: (i) the posted price P and (ii) the probability of acceptance, Ω. The joint P-Ω distribution is known to all agents.

Search cost. Searching requires effort, which I model as a utility cost. Each household k is assigned a search cost $\nu_k \geq 0$ per unit of purchase in real terms. Household k can search an unlimited number of times using the same ν_k, and the cost of each round is equivalent to $\mathbf{P}\nu_k Q$ in nominal terms. Lastly, let $V(\cdot)$ and $v(\cdot)$ denote the cumulative distribution function (CDF) and probability density function (PDF) of ν, respectively.

2.2 Optimal Reservation Price

As in McCall (1970), household k sets an optimal reservation price $\bar{P}_k$, given ν_k and the joint (P_s, Ω_s) distribution. Intuitively, the optimal $\bar{P}_k$ balances the expected total cost against the possibility of finding a cheaper firm who is still selling. Mathematically, she minimizes the ex-

[8] Extensive-margin rationing is also important for tractability. If firms ration at the intensive margin, households may need to purchase from multiple firms, potentially giving rise to intractable combinatorial problems.

pected total cost for household k, EC_k:

$$EC_k = \min_{\bar{P}_k} \left[\mathbf{P}\nu_k Q + \int_{P_s \le \bar{P}_k} \Omega_s P_s Q ds + \left(\int_{P_s > \bar{P}_k} ds + \int_{P_s \le \bar{P}_k} (1-\Omega_s) ds \right) EC_k \right], \tag{2.1}$$

where $\int_{P_s > \bar{P}_k} ds$ and $\int_{P_s \le \bar{P}_k} (1-\Omega_s) ds$ are the probability that the match is rejected by the buyer and the firm, respectively.

Number of searches. If no firm rations, the probability of forming a match is simply $\int_{P_s \le \bar{P}_k} ds$ (i.e., $F(\bar{P}_k)$). When firms ration, the probability of matching with firm s is weighted by Ω_s to account for the possibility that the firm refuses to sell. The probability of forming a match is therefore $\int_{P_s \le \bar{P}_k} \Omega_s ds$, and its inverse is the expected number of searches:

$$N(\bar{P}_k) = \left[\int_{P_s \le \bar{P}_k} \Omega_s ds \right]^{-1} = \left[F(\bar{P}_k) \, \mathrm{E}(\Omega_s \mid P_s \le \bar{P}_k) \right]^{-1}. \tag{2.2}$$

Expected purchase price. Define $\check{P}(\bar{P}_k)$ as the Ω-weighted average posted price among firms whose prices are below the reservation price $\bar{P}_k$:

$$\check{P}(\bar{P}_k) = \frac{\int_{P_s \le \bar{P}_k} \Omega_s P_s ds}{\int_{P_s \le \bar{P}_k} \Omega_s ds} = \mathrm{E}(P_s \mid P_s \le \bar{P}_k) + \frac{\mathrm{Cov}(\Omega_s, P_s \mid P \le \bar{P}_k)}{\mathrm{E}(\Omega_s \mid P_s \le \bar{P}_k)}. \tag{2.3}$$

The second equality follows from expressing $\check{P}(\bar{P}_k)$ in terms of expectations and covariance. Economically, $\check{P}(\bar{P}_k)$ is the expected price at which k purchases, given the reservation price $\bar{P}_k$.

First-order condition. Given ν_k, the optimal reservation price $\bar{P}_k$ is implicitly defined below:

$$\bar{P}_k = \check{P}(\bar{P}_k) + \mathbf{P} N(\bar{P}_k) \nu_k. \tag{2.4}$$

Conversely, for a given $\bar{P}_k$, the implied search cost satisfies $\mathbf{P}\nu_k = \int_{P_s \le \bar{P}_k} \Omega_s (\bar{P}_k - P_s) ds$. Importantly, the implicit function is strictly increasing because $\frac{\partial \bar{P}_k}{\partial \nu_k} = \mathbf{P} N(\bar{P}_k) > 0$. Details are provided in Appendix B.1.

Fundamentally, rationing affects households' reservation prices by changing their effective search costs. To see this, note that, in the absence of rationing ($\Omega \equiv 1$), (2.4) can be rewritten as follows:

$$\bar{P}_k = \mathrm{E}(P_s \mid P_s \le \bar{P}_k) + \mathbf{P} F(\bar{P}_k)^{-1} \nu_k. \tag{2.5}$$

With rationing, the same expression applies after replacing the primitive search cost ν_k with an effective search cost ν_k^{eff},

$$\bar{P}_k = \mathrm{E}(P_s \mid P_s \le \bar{P}_k) + \mathbf{P} F(\bar{P}_k)^{-1} \nu_k^{\mathrm{eff}}. \tag{2.6}$$

where the effective search cost ν_k^{eff} is given by

$$\nu_k^{\text{eff}} = \underbrace{\frac{\nu_k}{\mathrm{E}(\Omega_s \mid P_s \le \bar{P}_k)}}_{\text{Average acceptance}} + \underbrace{\frac{\mathrm{Cov}(\Omega_s, P_s \mid P_s \le \bar{P}_k)}{\mathrm{E}(\Omega_s \mid P_s \le \bar{P}_k)}}_{\text{Price-acceptance covariance}} \frac{F(\bar{P}_k)}{\mathbf{P}}. \tag{2.7}$$

The first term captures how rationing, conditional on $P_s \le \bar{P}_k$, scales up the primitive search cost ν_k due to the difficulty of finding a non-rationing firm. The second term captures an additional effect arising from the covariance between posted prices and rationing. When low-price firms accept fewer buyers and ration more heavily, households have to raise their reservation prices even if the average acceptance rate stays unchanged.

2.3 Residual Demand at the Extensive Margin

After deriving the optimal reservation price, I characterize the residual demand curve and the associated endogenous demand elasticity. Despite the analytical complexity, I establish a general result on the relationship between rationing and price competition at the end of this section. I further establish analytical results for a special case in Section 2.4.

Residual demand. Consider firm s that sets price P_s. Let $\nu(P_s)$ denote the search cost of the marginal household whose reservation price equals P_s according to (2.4). The firm receives match requests from all households with costs above $\nu(P_s)$, where each household with ν is expected to visit $N(\bar{P}(\nu))$ times. The total number of match requests, $M(P_s)$, is therefore:

$$M(P_s) = \int_{\nu(P_s)} N(\bar{P}(\nu)) dV(\nu) = \int_{P_s} \frac{\upsilon(\nu(\bar{P}))}{\mathbf{P}} d\bar{P}. \tag{2.8}$$

Note that $\upsilon(\nu(\bar{P}))$ denotes the probability density evaluated at $\nu(\bar{P})$. The second equality follows from substituting $dV(\nu) = \upsilon(\nu(\bar{P}))\frac{\partial \nu}{\partial \bar{P}} d\bar{P}$, which cancels $N(\bar{P}(\nu))$. Relatedly, demand elasticity at the extensive margin, $\epsilon^M(P_s)$, is given by

$$\epsilon^M(P_s) = -\frac{d\ln[M(P_s)]}{dp_s} = \frac{\upsilon(\nu(P_s))}{M(P_s)} \frac{P_s}{\mathbf{P}}. \tag{2.9}$$

Lastly, the residual demand in this partial equilibrium example equals $M(P_s)Q$, which is the product of the extensive margin and the intensive margin.

Strategic complementarities in pricing. Implicitly, residual demand and its curvature depend on the prices and acceptance rates of all competitors. Although these interactions are generally difficult to characterize without specifying V, I establish below an analytical result on strategic complementarities in pricing that holds independently of the functional form of V. Consider a competitor j with $\{P_j, \Omega_j\}$. Differentiating the marginal buyer's search cost with respect to P_j gives

$$\frac{\partial \nu(P_s)}{\partial P_j} = -\frac{\mathbf{1}_{P_j < P_s}}{\mathbf{P}} \Omega_j. \tag{2.10}$$

Importantly, the effect of P_j is proportional to the acceptance rate, Ω_j. Thus, for a lower-price competitor (i.e. $P_j < P_s$), the more it rations, the less its posted price matters for firm s's marginal buyer. By (2.8) and (2.9), this multiplicative property also carries over to the effects of P_j on $M(P_s)$ and $\epsilon^M(P_s)$ (see Appendix B.2). While $\frac{\partial M(P_s)}{\partial P_j}$ and $\frac{\partial \epsilon^M(P_s)}{\partial P_j}$ do not admit general analytical solutions, their proportionality to Ω_j holds for any functional form of V. This yields the following proposition:

Proposition 1 (Rationing attenuates strategic complementarities)**.** *The effects of competitor j's posted price, P_j, on firm s's residual demand, $M(P_s)$, and its elasticity, $\epsilon^M(P_s)$, are proportional to its acceptance rate, Ω_j. In particular, these effects are exactly zero when competitor j rations completely (i.e., $\Omega_j = 0$). In this sense, rationing attenuates strategic complementarities.*

The proof is in Appendix B.2. In plain language, Proposition (1) formalizes the intuition in the Introduction: *competitors can undercut only when they are selling.*

2.4 A Special Case with Analytical Results

To obtain further insights beyond Proposition (1), I analytically examine a special case in which v is uniformly distributed over the interval $[0, v^{-1}]$. The full derivation is provided in Appendix B.3. Define the unconditional average acceptance rate, $\mathbf{\Omega}$, and the Ω-weighted average posted price, $\check{\mathbf{P}}$, as follows:

$$\mathbf{\Omega} = \int \Omega_s ds, \quad \check{\mathbf{P}} = \int \frac{\Omega_s}{\mathbf{\Omega}} P_s ds. \tag{2.11}$$

Suppose that no posted price is above the maximum reservation price, $\bar{P}(v^{-1})$. Equation (2.8) simplifies exactly to

$$M(P_s) = \frac{1}{\mathbf{\Omega}} + \frac{\check{\mathbf{P}} - P_s}{\mathbf{P}} v. \tag{2.12}$$

Let ω denote the log average acceptance rate, i.e., $\omega \equiv \ln \mathbf{\Omega}$. I prove in Appendix B.3:

Proposition 2 (Sufficient statistics for the extensive margin)**.** *Given (2.12), a first-order approximation around $P_s = \check{\mathbf{P}}$ yields*

$$d \ln [M(P_s)] \approx -d\omega - \epsilon^M(\check{\mathbf{P}})\, d(p_s - \check{\mathbf{p}}), \tag{2.13}$$

and

$$d \ln [\epsilon^M(P_s)] \approx d\omega + \left[1 + \epsilon^M(\check{\mathbf{P}})\right] d(p_s - \check{\mathbf{p}}) \; - d(\mathbf{p} - \check{\mathbf{p}}). \tag{2.14}$$

Thus, to a first-order approximation, residual demand and its elasticity are summarized by three sufficient statistics: changes in the average acceptance rate, the competitor price gap, $d(p_s - \check{\mathbf{p}})$, and the cross-market price gap, $d(\mathbf{p} - \check{\mathbf{p}})$.

Average acceptance rate (rationing). A decrease in the average acceptance rate, or equivalently an increase in the average degree of rationing, has a first-order effect on the extensive-

margin demand elasticity through $d\boldsymbol{\omega}$. As average rationing increases and $d\boldsymbol{\omega}$ becomes negative, demand not only increases but also becomes more inelastic. Proposition 2 also yields a sharp prediction on the magnitude: a change in the average degree of rationing locally translates one-for-one into a change in the extensive-margin demand elasticity.

Competitor price gap. I refer to the difference between p_s and $\check{\mathbf{p}}$ as the *competitor price gap*, which captures the difference between firm s's posted price and the Ω-weighted average posted price of its competitors. As expected, $\epsilon^M(P_s)$ increases in the competitor price gap, a result qualitatively similar to those in standard variable-elasticity models, such as the Kimball and nested CES aggregators.

What distinguishes (2.14) from standard models is the definition of $\check{\mathbf{p}}$. As Amiti et al. (2019) show, in a broad class of models, the index of competitor price changes can be approximated to first order by a market-share-weighted average of competitors' price changes. In my rationing model, however, $\check{\mathbf{P}}$ is an Ω-weighted average, which means that price changes alone are no longer a sufficient statistic. To a first-order approximation, $d\check{\mathbf{p}}$ can be decomposed as follows:

$$d\check{\mathbf{p}} \approx \underbrace{\int dp_j dj}_{\substack{\text{Non-rationing} \\ \text{benchmark}}} + \underbrace{\text{Cov}(\omega_j, dp_j)}_{dp_j \text{ reweighting}} + \underbrace{\text{Cov}(d\omega_j, p_j)}_{d\omega_j \text{ reweighting}}. \tag{2.15}$$

The first term is the unweighted average of price changes, analogous to AIK's index of price changes when competitors have equal market shares. The second term captures the covariance between ω_j and dp_j. It effectively assigns greater (lesser) weight to price changes among firms that ration less (more). The reweighting of dp_j is a direct application of Proposition 1 and provides the theoretical foundation for the empirical test in Section 5.3.1. Similarly, the third term captures the covariance between $d\omega_j$ and p_j.

Cross-market price gap. The last term in (2.14) is the change in the *cross-market price gap*, defined as the difference between $\check{\mathbf{p}}$ and the aggregate price index, $\mathbf{p}$. This term arises because search costs are constant in real terms. Its economic implication is consistent with the empirical analysis of market-power spillovers in Acharya et al. (2023), who conjecture that supply-driven inflation lowers demand elasticity even in sectors not directly affected by supply-chain shocks. I further characterize the cross-market price gap in general equilibrium in Section **??**.

By contrast, the classic channel of cross-market spillovers through relative prices changes the level of demand instead of its elasticity. In nested CES models, an increase in the aggregate price index raises demand for all firms within the same sector proportionally, leaving their market shares and thus residual demand elasticities unchanged. The same logic applies to multi-sector Kimball models.

2.5 From Residual Demand to Optimal Pricing

The extensive margin is an important component of firms' pricing decisions. To fix ideas, I assume that a firm's overall residual demand, $Y^d(P_s)$, and its elasticity, $\epsilon(P_s)$, take the following

form:

$$Y^d(P_s) = \underbrace{M(P_s)}_{\text{Extensive}} \times \underbrace{\left(\frac{P_s}{\mathbf{P}}\right)^{-\sigma}\mathbf{C}}_{\text{Intensive}}, \quad \epsilon(P_s) = \sigma + \epsilon^M(P_s), \tag{2.16}$$

where σ is the intensive-margin elasticity, $\epsilon^M(P_s)$ is defined in (2.9), and $\mathbf{C}$ is aggregate consumption. To isolate the role of the extensive margin, I keep the intensive margin as simple as possible. As I show in Section 3, this specification is consistent with my complete general equilibrium model. To obtain analytical results, I continue to work under the uniform distribution assumption, leaving the quantitative analysis to Section 4.

The elasticity zoo. I define the superelasticity as $\psi(P_s) = \frac{\partial \ln[\epsilon(P_s)]}{\partial p_s}$. Under the uniform distribution assumption, it simplifies to

$$\psi(P_s) = \frac{\epsilon^M(P_s)}{\epsilon^M(P_s)+\sigma}\left(\epsilon^M(P_s)+1\right). \tag{2.17}$$

Following AIK, I define $\Gamma(P_s)$ as the elasticity of the desired markup, $\frac{\epsilon(P_s)}{\epsilon(P_s)-1}$, with respect to P_s. It is given by

$$\Gamma(P_s) = \frac{\psi(P_s)}{\epsilon(P_s)-1}. \tag{2.18}$$

Optimal flexible price. Consider the pricing problem of a single flexible-price firm. Although its competitors may ration for various reasons, the firm itself does not, as it can always adjust its price to clear excess demand. As I show below in Section 3.2, the insights from this static problem are useful for understanding optimal pricing under Calvo pricing.

The flexible-price firm's optimal price, P^*, equals the desired markup, $\frac{\epsilon(P^*)}{\epsilon(P^*)-1}$, multiplied by nominal marginal cost, MC.

$$P^* = \frac{\epsilon(P^*)}{\epsilon(P^*)-1} MC. \tag{2.19}$$

Using Proposition 2 and log linearizing around $\check{\mathbf{P}}$, I obtain:

$$dp^* \approx \underbrace{\frac{1}{1+\Gamma(\check{\mathbf{P}})} dmc}_{\text{Cost-driven}} + \underbrace{\frac{\Gamma(\check{\mathbf{P}})}{1+\Gamma(\check{\mathbf{P}})}\left[d\check{\mathbf{p}} + \frac{-d\omega + d(\mathbf{p}-\check{\mathbf{p}})}{1+\epsilon^M(\check{\mathbf{P}})}\right]}_{\text{Endogenous elasticity}} \tag{2.20}$$

Together with (2.15), this expression provides a sharp contrast with the non-rationing bench-

mark in Amiti et al. (2019):

$$dp^* \approx \underbrace{\frac{1}{1+\Gamma(\check{\mathbf{P}})} dmc + \frac{\Gamma(\check{\mathbf{P}})}{1+\Gamma(\check{\mathbf{P}})} \int dp_j dj}_{\text{Non-rationing benchmark}} + \frac{\Gamma(\check{\mathbf{P}})}{1+\Gamma(\check{\mathbf{P}})} \left[\underbrace{\frac{-d\omega}{1+\epsilon^M(\check{\mathbf{P}})} + \text{Cov}(\omega_j, dp_j) + \text{Cov}(d\omega_j, p_j)}_{\text{Competitor rationing}} + \underbrace{\frac{d(\mathbf{p}-\check{\mathbf{p}})}{1+\epsilon^M(\check{\mathbf{P}})}}_{\substack{\text{Cross-market} \\ \text{price gap}}} \right] \tag{2.21}$$

Endogenous cost pass-through. By construction, the first-order analysis above abstracts from endogenous changes in Γ and, hence, in cost pass-through. Given (2.18), under the uniform distribution assumption,

$$\frac{\partial \Gamma(P_s)}{\partial \epsilon^M(P_s)} \begin{cases} > 0, & \text{if } \sigma > 1, \\ = 0, & \text{if } \sigma = 1, \\ < 0, & \text{if } \sigma < 1. \end{cases} \tag{2.22}$$

Beyond the special case, these effects are difficult to characterize analytically and ultimately constitute an empirical question. Section 5.3.2 suggests that Γ decreases when rationing is high. The quantitative model delivers a qualitatively similar result. Further exploring this relationship is left for future research.

3 Synthesizing the New and Old Keynesian Models

In this section, I incorporate the interaction between rationing and competition described above into an otherwise standard New Keynesian model.

3.1 Households

3.1.1 Consumption Bundle

In each period, household k first draws search costs $\{\nu_{k,i}\}$ for all product markets, indexed by i. Given the exogenous nominal budget X_k, she then chooses reservation prices $\{\bar{P}_{k,i}\}$ and starts the search process. After successfully matching with firms in all markets, she chooses purchase quantities $\{Q_{k,i}\}$ at the prices posted by the matched firms, $\{P_{k,i}\}$. Consumption is aggregated across products using a standard CES aggregator:

$$C_k = \left(\int Q_{k,i}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}, \quad k, i \in [0,1]. \tag{3.1}$$

Formally, household k maximizes U_k, defined as the expected consumption utility net of

search costs, subject to the budget constraint $X_k = \int P_i Q_{k,i} di$:

$$U_k = \max_{\{\bar{P}_{k,i}\}} \left[\mathrm{E}_{\{P_{k,i}\}} \left(\max_{\{Q_{k,i}\}} \frac{C_k^{1-\gamma}}{1-\gamma} \right) - \mathbf{C}^{1-\gamma} \int \frac{\mathbf{Q}_i}{\mathbf{C}} N_i(\bar{P}_{k,i}) \nu_{k,i} di \right], \tag{3.2}$$

Here, $\mathbf{C}$ and $\mathbf{Q}_i$ denote aggregate consumption bundle and aggregate consumption of good i, respectively. $N_i(\bar{P}_{k,i})$ denotes the expected number of searches in market i given reservation price $\bar{P}_{k,i}$, defined identically to $N(\cdot)$ in (2.2).

The first term in (3.2) takes an expectation over the joint distribution of $\{P_{k,i}\}$, since reservation prices are chosen before $\{P_{k,i}\}$ are realized. In the second term, $\mathbf{C}^{1-\gamma}$ ensures that search costs, which are modeled as utility costs, are proportional to consumption utility. $\frac{\mathbf{Q}_i}{\mathbf{C}}$ ensures that search costs are proportional to the amount of Q_i consumed. The proportionality assumption greatly simplifies the problem. As I show below, optimal reservation prices take the same simple form as in the single-product example in (2.4).

Ex-ante representative households. In Section 2.1, I assume that $\nu_{k,i}$ is distributed according to the same V across households k in every market i. However, when search costs are drawn i.i.d. across a continuum of markets, a given household's search costs across markets need not follow V (Judd, 1985). To ensure that both distributions coincide with V, I assign search costs deterministically via

$$\nu_{k,i} = V^{-1}\big((i+k) \bmod 1\big). \tag{3.3}$$

For instance, if $i = 0.8$ and $k = 0.6$, then $(i+k) \bmod 1 = 1.4 \bmod 1 = 0.4$. Since $(i+k) \bmod 1$ is uniformly distributed over i for any given k, and over k for any given i, each household faces the distribution V across markets, and each market faces the distribution V across households.

Together with symmetric product markets, (3.3) ensures that all households are ex ante representative: they take the same actions before the realization of $\{P_{k,i}\}$. Households differ only ex post, after drawing different realizations of $\{P_{k,i}\}$. Conditional on $\{P_{k,i}\}$, the ex-post consumption price index follows the standard CES formula $\left(\int P_{k,i}^{1-\sigma} di\right)^{\frac{1}{1-\sigma}}$.

Expected consumption price. Reservation prices depend on the *expected* consumption price because they are chosen before $\{P_{k,i}\}$ are realized. Hence I drop the k subscript. Let $\mathbf{P}$ denote the expected consumption price for the ex-ante representative household, which, to a first-order approximation, equals the ex-post average consumption price index. Under symmetric markets and to a first-order approximation, $\mathbf{P}$ is given by

$$\mathbf{P} = \left[\int \check{P}(\bar{P}(\nu))^{1-\sigma} dV(\nu) \right]^{\frac{1}{1-\sigma}}, \tag{3.4}$$

where $\bar{P}(\cdot)$ is the optimal reservation price as a function of ν defined below, and $\check{P}(\cdot)$ is the Ω-weighted average posted price defined in (2.3).

Optimal $\bar{P}(\cdot)$. At first order, the optimality condition regarding the reservation price is identical to (2.4).

Intertemporal. Lifetime utility function of the representative household

$$V_t(B_{t-1}) = \max\ U_t - \chi \frac{L^{1+\eta}}{1+\eta} + \frac{\beta}{\exp \xi_t} \mathrm{E}_t\, V_{t+1}(B_t). \tag{3.5}$$

subject to $\exp(-i_t)B_t = B_{t-1} + W_t L_t - \int Q_{i,t} P_{i,t}\, di + \Pi_t$, where B_t denotes bond holdings, Π_t denotes nominal firm profits, i_t is the log risk-free interest rate, and ξ_t is a reduced-form aggregate demand shock. The log-linearized Euler equation is

$$\hat{c}_t = \mathrm{E}_t\, \hat{c}_{t+1} - \frac{1}{\gamma}\Big(i_t - \mathrm{E}_t\, \pi_{t+1} + \ln \beta\Big) + \frac{1}{\gamma}\xi_t. \tag{3.6}$$

where π_t is the log growth of the consumption price index $\mathbf{P}_t$ defined in (3.4). Labor supply is given by $L^\eta = \chi^{-1} C_t^{-\gamma} \frac{W_t}{\mathbf{P}_t}$.

3.2 Firms

Firms face a dynamic problem due to nominal rigidities. I further impose idiosyncratic productivity and demand shocks on producers to ensure a non-degenerate price distribution in equilibrium. All variety markets are symmetric in equilibrium.

3.2.1 Setup

Production function. Let $C_{s,t}(x)$ denote the real cost of producing x units at time t for firm s. In the baseline model, each firm s produces output according to

$$Y_{s,t} = A_t A_{s,t} L_{s,t}^{1-\alpha}, \tag{3.7}$$

where A_t and $A_{s,t}$ denote aggregate and firm-level productivity, respectively. α is the returns to scale. Firms take the real wage W_t^r as given. The real cost function is given by:

$$C_{s,t}(Y_{s,t}) = W_t^r \left(\frac{Y_{s,t}}{A_{s,t}}\right)^{\frac{1}{1-\alpha}}. \tag{3.8}$$

Rationing. Since only a fraction $\Omega_{s,t}$ of match requests are accepted, total output of seller s is

$$Y_{s,t} = \Omega_{s,t} Y_{s,t}^d. \tag{3.9}$$

Idiosyncratic demand shock. Demand $Y_{s,t}^d$ in the complete model is an extension of (2.16). It is well known that in models with high superelasticity, implausibly large productivity shocks are required to generate reasonable price distributions (Klenow and Willis, 2016). In the context of the Kimball aggregator, Aruoba et al. (2024) show that introducing idiosyncratic demand shocks can generate realistic pricing dynamics. I adopt their approach and introduce an idiosyncratic demand shock. The idiosyncratic demand shock $\xi_{s,t}$ takes the following form. For seller s that

posts $P_{s,t}$, the quantity demanded by buyers, $Y^d_{s,t}$, is:

$$Y^d_{s,t} = M_t\left(\frac{P_{s,t}}{\exp(\xi_{s,t})}\right)Q_t(P_{s,t}). \tag{3.10}$$

where $Q(\cdot)$ is the intensive-margin demand and $M_t(\cdot)$ is the number of matches defined in (2.8). I choose the particular function form so that idiosyncratic demand $\xi_{s,t}$ changes demand elasticity in exactly the same way as in Aruoba et al. (2024).[9]

One can interpret $\xi_{s,t}$ as buyers' misperception about the price distribution. For example, some buyers overestimate how expensive things are, which raises both their reservation prices. For sellers matched with these buyers, it is as if they posted lower prices in the matching process, resulting in a positive $\xi_{s,t}$.

3.2.2 Calvo Pricing

Profit maximization. Prices are sticky à la Calvo, with θ_c denoting the probability of not resetting.[10] Let $x^{nr}_{s,t}$ denote the value of x conditional on *not* resetting the price, while $x_{s,t}$ refers to the resetting case. The seller's objective function in real terms is given by:

$$\begin{aligned} V_{s,t}(P_{s,t-1};A_{s,t},\xi_{s,t}) = \max_{\Omega_{s,t},P_{s,t},\Omega^{nr}_{s,t}} &(1-\theta_c)\Big[Y_{s,t}P_{s,t} - C_{s,t}\big(Y_{s,t}\big) + \mathrm{E}_t\,\beta V_{s,t+1}(P_{s,t};A_{s,t+1},\xi_{s,t+1})\Big] \\ &+\theta_c\Big[Y^{nr}_{s,t}P_{s,t-1} - C_{s,t}\big(Y^{nr}_{s,t}\big) + \mathrm{E}_t\,\beta V_{s,t+1}(P_{s,t-1};A_{s,t+1},\xi_{s,t+1})\Big], \end{aligned} \tag{3.11}$$

subject to $\Omega_{s,t} \le 1$, $\Omega^{nr}_{s,t} \le 1$, (3.8), and (3.10).

Optimal reset price. Let $\tilde{P}_{s,t} = \frac{P_{s,t}}{\mathbf{P}_t}$ denote the relative price posted by seller s. As shown in Appendix C.2.1, the seller's optimal relative reset price $\tilde{P}_{s,t}$ satisfies:

$$\tilde{P}_{s,t} = \frac{\epsilon_{s,t}}{\epsilon_{s,t}-1}C'_{s,t}(Y_{s,t}) + \beta\frac{\mathrm{E}_t\,V'_{s,t+1}(\tilde{P}_{s,t})}{Y_{s,t}(\epsilon_{s,t}-1)}, \tag{3.12}$$

where demand elasticity $\epsilon_{s,t}$ is given by $\epsilon_{s,t} = \epsilon_t\left(\frac{P_{s,t}}{\exp(\xi_{s,t})}\right)$.

Optimal rationing. $\Omega_{s,t}$ and $\Omega^{nr}_{s,t}$, are chosen such that.

$$\tilde{P}_{s,t-1} \ge C'_{s,t}(Y^{nr}_{s,t}) \quad \text{and} \quad \tilde{P}_{s,t} \ge C'_{s,t}(Y_{s,t}). \tag{3.13}$$

[9] In Aruoba et al. (2024), the derivative of demand elasticity with respect to idiosyncratic demand is the same as the derivative with respect to a firm's own price (see p.10). (3.10) also implies the same equality because $\frac{\partial}{\partial p_{s,t}}M_t\big(\frac{P_{s,t}}{\exp(\xi_{s,t})}\big) \equiv -\frac{\partial}{\partial \xi_{s,t}}M_t\big(\frac{P_{s,t}}{\exp(\xi_{s,t})}\big)$. Adding demand shocks on the intensive margin $Q_t\big(\frac{P_{s,t}}{\exp(\xi_{s,t})}\big)$ makes little quantitative difference as it does not change the desired markup.

[10] One would expect rationing to remain important in menu-cost models, since rationing weakens the need for faster price adjustments in the first place. In standard NK settings, firms are willing to pay the menu cost because the opportunity cost of holding a low price is approximately quadratic in the price gap. For OK firms, by contrast, the opportunity cost is approximately linear in the price gap, substantially reducing the incentive to pay the menu cost.

Description	Variable	Value	Note
a. Goods Producers			
Calvo pricing parameter	θ_c	0.7	Gagliardone et al. (2023)
S.D. of productivity innovations	σ_a	0.0224	Foster et al. (2008)
Persistence of productivity shocks	ρ_a	0.943	Foster et al. (2008)
S.D. of demand innovations	σ_ξ	0.0635	Foster et al. (2008)
Persistence of demand shocks	ρ_ξ	0.976	Foster et al. (2008)
Steady-state productivity	$\bar{a}$		Normalizing output to one
Returns to scale	$1-\alpha$	0.7	
b. Other Sectors			
Discount rate	β	0.99	
Frisch elasticity	η	2	
Labor disutility	χ		Normalizing wages to one
Risk aversion	γ	1.5	
Intensive-margin elasticity	σ	1.25	
Taylor rule (inflation)	φ_π	1.5	
Taylor rule (output)	φ_y	0.125	

Table 1: Externally Calibrated Parameters

That is, firms ration output whenever needed so that marginal cost never exceeds price.

Excess demand vs. excess supply. Introducing firm-level rationing $\Omega_{s,t}$ (and $\Omega^{nr}_{s,t}$) in (3.9) links the framework to Old Keynesian models described in Barro (2025). A goods producer faces excess supply when the constraint $\Omega_{s,t} \leq 1$ binds, i.e., when it would like to serve more customers than it actually has at the posted price. Conversely, it faces excess demand when the optimal $\Omega_{s,t}$ is strictly below one and the firm actively rations.

3.3 Other Elements

Monetary policy. I assume a standard Taylor rule $i_t = -\ln\beta + \varphi_\pi \pi_t + \varphi_y \hat{c}_t$, and π_t is the log change in the consumption price index $\mathbf{P}_t$ in (3.4). $\hat{c}_t$ is the log deviation from the steady-state value. i_t is the policy rate in log terms.

Numerical solution. I use the sequence-space Jacobian method to find the exact nonlinear dynamics under perfect foresight (Auclert et al., 2021). Compared to standard models, the industry equilibrium introduces an additional $P-\Omega$ fixed point, which requires more iterations. Fortunately, the firm's problem can be solved via the endogenous grid method, which greatly speeds up the iterations.

4 Inflation Dynamics in a Calibrated Model

4.1 Calibration

To discipline the model, I calibrate it using evidence on U.S. firm dynamics (Aruoba et al., 2024, Foster et al., 2008) and empirical estimates of strategic complementarities (Gagliardone et al., 2023, Amiti et al., 2019). Most microeconomic studies rely on data from the manufacturing sec-

Description	Target	Model	Note
a. Targeted Moments			
$\psi/(\epsilon-1)$	1.17	1.16	Gagliardone et al. (2023)
Average markup	30%	29.0%	Weighted by input costs
b. Model Parameters			
μ_v		0.147	Mean before truncation
σ_v		0.110	S.D. before truncation
c. Untargeted Steady-State Moments			
S.D. of prices		0.176	0.18 in Foster et al. (2008)
Average Ω		97.9%	
Average ϵ		4.59	
Average ψ		4.16	

Table 2: Internally Calibrated Parameters

tor, even though services account for the majority of output in advanced economies. Relative to manufacturing firms, service-sector firms adjust prices less frequently and exhibit greater price stickiness (Nakamura and Steinsson, 2008, Cotton and Garga, 2022). Because the incidence of rationing increases with price stickiness, this calibration strategy likely provides a lower bound on the quantitative importance of rationing and endogenous competition.

Externally calibrated parameters. The model is calibrated at a quarterly frequency with a discount factor of $\beta = 0.99$. On the household side, I assume a CRRA utility function with risk aversion $\gamma = 1.5$ and set the Frisch elasticity to $\eta = 2$. For the demand aggregator, I choose an intensive-margin elasticity of $\sigma = 1.25$.[11] A relatively low value of σ provides additional flexibility in targeting the average markup. Finally, monetary policy follows a Taylor rule with coefficients $\varphi_\pi = 1.5$ and $\varphi_y = 0.125$.

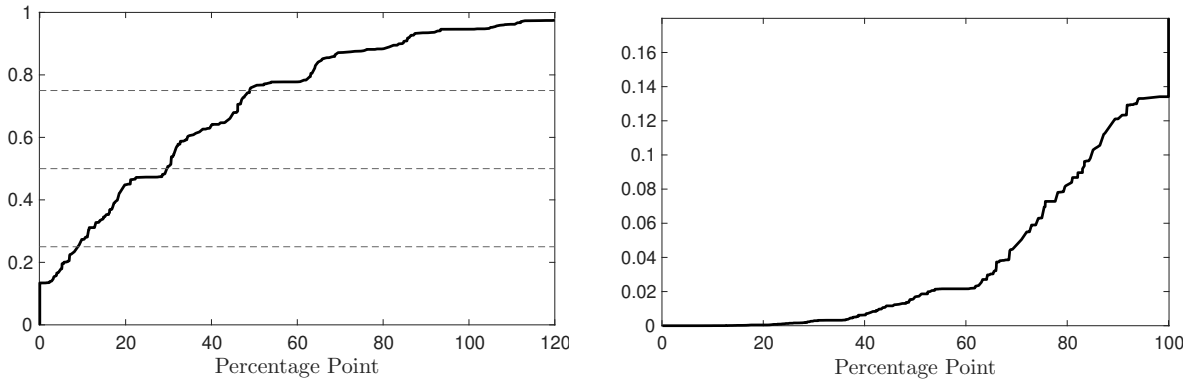
For idiosyncratic demand and productivity shocks, I use quarterly AR(1) processes to match the annual persistence and standard deviation reported by Foster et al. (2008). Idiosyncratic demand shocks, $\xi_{s,t}$, follow an AR(1) process with persistence $\rho_\xi = 0.976$ and innovation standard deviation $\sigma_\xi = 0.0635$. Idiosyncratic productivity shocks, $a_{s,t}$, follow an AR(1) process with persistence $\rho_a = 0.943$ and innovation standard deviation $\sigma_a = 0.0225$.

Finally, I set the Calvo parameter to $\theta_c = 0.7$ in line with Gagliardone et al. (2023). I further set $\alpha = 0.3$ for the production function in (3.7), consistent with the DSGE literature.[12] Labor disutility and aggregate productivity are chosen to normalize steady-state wages and output to one, respectively.

Internally calibrated search cost distribution. Search costs play a central role in the model because they determine demand elasticity and superelasticity. I assume that search costs follow

[11] The implied love for variety is $\frac{1}{1.25-1} = 4$.

[12] Empirical evidence suggests that long-run returns to scale in the manufacturing sector are close to unity (Lenzu et al., 2024, Gagliardone et al., 2023, Eslava et al., 2024). Short-run returns to scale, however, are much less well understood, as they depend critically on input flexibility and on the elasticity of substitution across inputs. Assuming constant short-run returns to scale implies infinite firm-level production capacity, which is at odds with observed capacity utilization. Besides, PMI delivery times exhibit substantial variation even prior to COVID-19, indicating that input adjustment is far from frictionless.



(a) Markup

(b) Acceptance Rate

Figure 1: Steady State Markups and Acceptance Rates

a truncated normal distribution with support on $\mathbb{R}+$, characterized by two parameters, μ_ν and σ_ν.

Two empirical moments pin down the two parameters. First, empirical estimates of strategic complementarities suggest that the ratio $\frac{\psi}{\epsilon-1}$ is close to unity. Following Gagliardone et al. (2023), I target a value of 1.17.[13] Because the model features heterogeneity in both elasticity and superelasticity, I compute this ratio using sales-weighted elasticity and superelasticity. This moment primarily identifies the ratio between μ_ν and σ_ν.

Second, I choose the level of μ_ν by targeting an average markup of 30 percentage points, similar to 29% in Aruoba et al. (2024). Markups are weighted by input costs as explained by Edmond et al. (2023). Even though many methodological questions remain (see Basu, 2019), a cost-weighted average of 30% certainly lies towards the higher end of the spectrum.[14] I deliberately choose a relatively high markup target to demonstrate the robustness of the results, as higher markups reduce the need for rationing.

Given these two moments, the calibrated search cost distribution implies $\mu_\nu = 0.147$ and $\sigma_\nu = 0.110$.[15] In steady state, the average markup is 29% when weighted by input costs and 33% when weighted by sales. The average elasticity and superelasticity are $\bar{\epsilon} = 4.59$ and $\bar{\psi} = 4.16$, implying $\frac{\bar{\psi}}{\bar{\epsilon}-1} = 1.16$.[16]

4.2 Untargeted Steady-State Moments

Price dispersion. The model generates a standard deviation of log price levels of 0.176, almost identical to the empirical estimate of 0.18 in Foster et al. (2008). Note that I solve the model under perfect foresight for technical reasons. The key difference between the deterministic and stochastic steady states is that price dispersion is wider in the latter, as it increases with the variance of aggregate shocks. Given the large variance of idiosyncratic shocks in the model, the additional contribution of aggregate shocks to price dispersion is arguably modest. If price dispersion were indeed substantially wider, rationing effects would be even stronger, as illustrated in panel (c) of Figure E6.

Markup distribution and the acceptance rate. In the steady state, the cost-weighted average markup is as high as 29% following my deliberately conservative calibration. High markups tend to undermine rationing and the competition channel. Panel (a) of Figure 1 plots the cumulative distribution function of markups. The 25th, 50th, and 75th percentiles are 8.5% and 29.6%, and 48.8%, respectively. Firms operating at a zero markup account for 9.9% of sales and 13.4% of costs. Zero-markup firms, however, do not necessarily ration aggressively. The average Ω conditional on a zero markup is 72.9%, while the unconditional average Ω is as high as 97.9%. A 2.1% rejection rate is therefore much lower than the average pre-COVID stockout rate of 12% reported by Cavallo and Kryvtsov (2023), which is unsurprising given the high markups.

4.3 Inflation Under Excess Demand

First, I impose aggregate demand shocks through ξ_t in (3.6). The aggregate shock hits at $t = 1$ and follows an AR(1) process with persistence $\rho = 0.5$. The initial shock size ranges from -8% to 8%, as shown on the x-axis of Figure 2. The first panel displays price changes upon impact as a function of the initial shock size, whereas the second panel shows cumulative price changes over the first four quarters and over the entire transitional path. All other panels report average responses of non-price variables over the first four quarters, while full transitional paths are shown in panel (a) of Figure A4.

I repeat the same exercise but with aggregate productivity shocks. Results are in Figure A5, which largely mirror responses to demand shocks.

4.3.1 Endogenous Erosion of Competition

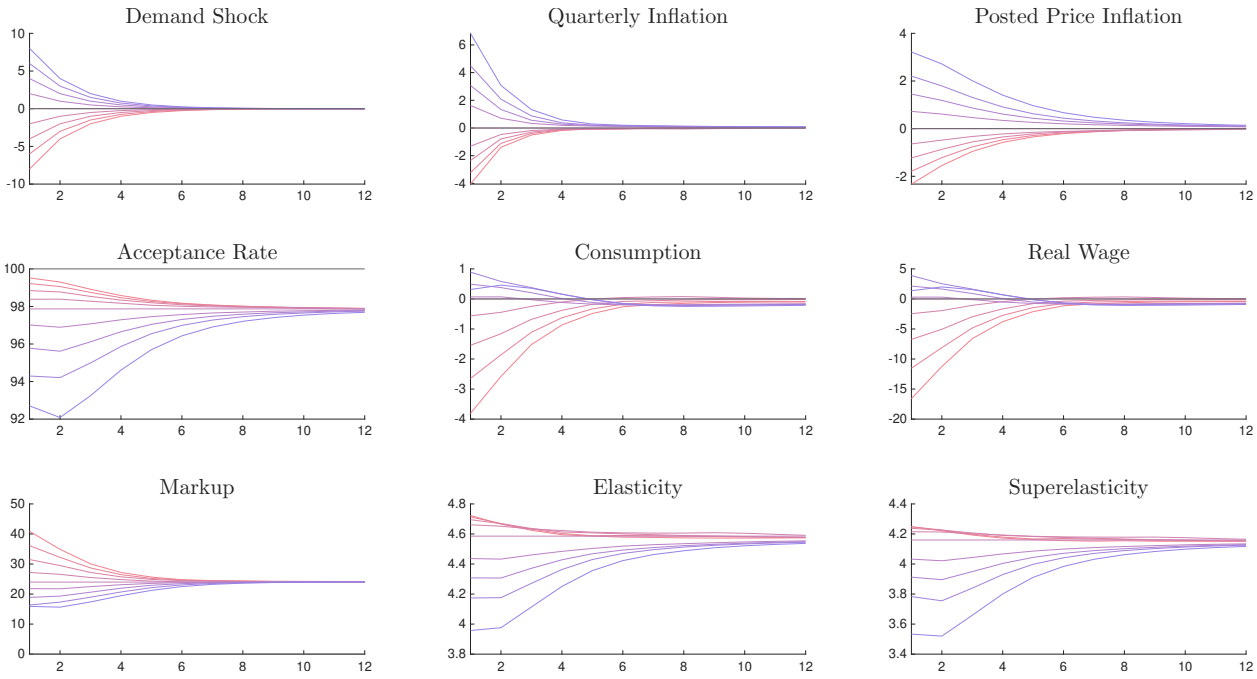
To be added. See figure 2. Markups and profits are both less countercyclical.

[13] In their notation, $1-\Omega = \left(1+\frac{\psi}{\epsilon-1}\right)^{-1}$. They estimate $\Omega \approx 0.54$, which implies $\frac{\psi}{\epsilon-1} \approx 1.17$.

[14] At probably the highest end, De Loecker et al. (2020) report a sales-weighted average markup of 61%. However, the corresponding cost-weighted average falls to roughly 35%. Moreover, such a high level appears highly implausible (e.g., Barkai, 2020, Foster et al., 2022, Benkard et al., 2025). Even in De Loecker et al. (2020), the cost-weighted average markup that accounts for overhead cost never exceeds 10%.

[15] As I show in Figure A3, high μ_v reduces elasticity, whereas high σ_v lowers superelasticity.

[16] In a NK model with no search frictions, one could compute the marginal-cost Phillips curve slope using $\frac{(1-\theta_c)(1-\beta\theta_c)}{\theta_c}\left(1+\frac{\psi}{\epsilon-1}\right)^{-1}$, which is approximately 0.060.



Notes: ...

Figure 2: Impulse Responses Following Demand Shocks

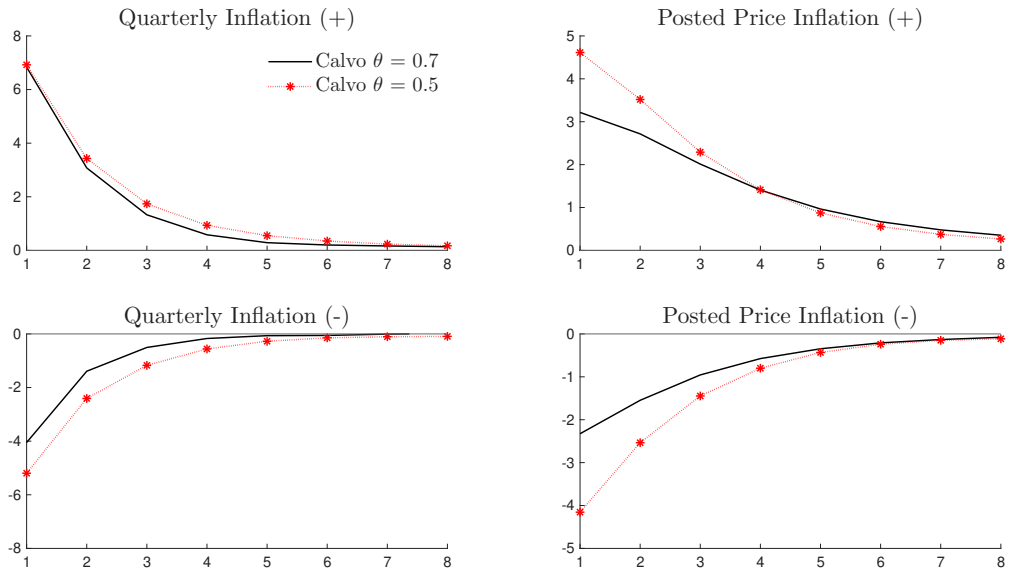
4.3.2 Asymmetry and Persistence

Asymmetric magnitude. Strong asymmetry emerges when demand shocks turn positive. An 8% positive shock raises the price level by over 6% upon impact and by 12% over four quarters, almost doubling the effect of its negative counterpart. This amplification is not limited to very large shocks. The dashed red line extrapolates impulse responses from negative shocks, and actual inflation diverges immediately as soon as the shock turns positive.

At the core of the asymmetry is the uneven response of rationing. In the model, the average acceptance rate falls from 98% in steady state to under 93% following an 8% shock, compared to 99% following a -8% shock. As rationing intensifies, households switch to more expensive sellers, and both elasticity and superelasticity decline. Not suprisingly, rationing dampens the countercyclicality of markups following positive shocks, as seen in the gap between the actual average markup and the linear extrapolation. The dampening effect is even more salient in the impulse response function in panel (a) of Figure A4. Markups remain highly countercyclical after negative shocks, as my model deviates from standard models only during positive shocks.

Asymmetric persistence. In addition to its magnitude, inflation displays a pronounced asymmetry in intrinsic persistence.[17] Both the posted price inflation and the consumption price inflation are substantially more persistent following positive demand shocks than negative demand shocks (also see the IRFs in Figure A4). The intuition for asymmetric persistence is straightforward. Rising price levels continue to lower relative prices for firms with old prices. Unlike mean-reverting shocks, price gaps persist until the next price adjustment. Even after the

[17] measure persistence by fitting an AR(1) model to the impulse response function.



Notes: ...

Figure 3: Inflation and Price Stickiness

shock has dissipated, firms with large price gaps continue to ration while waiting for the next reset, generating a highly persistent increase in rationing among low-price firms. Rationing reinforces inflation, while inflation creates more price gaps. This, in turn, strengthens the intrinsic persistence of aggregate inflation.

4.3.3 Paradox of Price Stickiness

To be added. Show that, as in Figure 3, demand shocks drive up inflation regardless of nominal rigidities due to rationing and endogenous competition. Higher price stickiness weakens cost-driven inflation. Yet it also leads to more rationing and weaker price competition, which strengthens excess-demand-driven inflation.

4.4 Monetary Policy Under Excess Demand

4.4.1 Effects of Monetary Policy Shocks

Interact MP shocks with positive & negative demand shocks, which is a DID to see MP shock effects under excess demand vs excess supply. Show that MP is better at controlling inflation under excess demand and stimulating output under excess supply.

4.4.2 Policy Effectiveness and Price Stickiness

To be added. Show that MP effectiveness depends much less on nominal rigidities when excess demand dominates. The conventional wisdom that greater nominal rigidities increases the sacrifice ratio is only true when excess supply dominates, but there isn't much inflation to deal with under excess supply anyways.

4.4.3 Risk of Persistently High Inflation

The quantitative exercise above reveals a novel negative externality of inflation: the disruption of price competition caused by rationing. When excess demand arises, either from supply constraints or from demand stimulus, this anti-competitive externality can cause inflation to persist even after the underlying shocks have dissipated. Consequently, monetary policy should respond more strongly to inflation driven by excess demand.

From a theoretical perspective, this novel externality echoes the earlier insight of Carlton (1982), who argues that inflation can disrupt the organization of markets. The mechanisms are difference, however. In my model, the externality arises because rationing weakens competitive pressure in product markets. By contrast, Carlton (1982) focuses on the difficulty of writing long-term fixed-price contracts when inflation creates uncertainty about future price levels.

5 Empirical Support from Micro Data

Testing for rationing is essentially testing whether firms stay on their demand curves. For example, Holden (2024) identifies retail products that are likely underpriced because of sticky prices. He then tests whether observed quantities increase sufficiently to clear the additional demand implied by the price gaps. My first test in Section 5.2 complements Holden (2024)'s approach by identifying ex ante quantity gaps using capacity information, and then I examine whether subsequent price adjustments are sufficient to clear the quantity gaps.

In Section 5.3 I test whether rationing softens competition. This requires a new empirical strategy, as rationing can affect competitive pressure without changing standard concentration measures such as the Herfindahl-Hirschman Index. I therefore extend Amiti et al. (2019) by examining whether industry-wide rationing, proxied by capacity constraints, changes firms' responses to their own cost shocks and competitors' prices. In Walrasian models, neither effect should differ because prices are sufficient statistics for competitive interactions. By contrast, my theory in Section 3 yields distinct predictions that directly speak to the empirical test.

5.1 Data

I use data on Indian manufacturing firms from the CMIE Prowess database. Relative to other datasets, Prowess uniquely combines information on input prices and quantities, output prices and quantities, and capacity utilization.[18] To the best of my knowledge, Prowess is the only dataset currently available that provides all three aspects at the product level, which is essential for my test in Section 5.3. On notation, products, firms, industries, and years are indexed by i, f, s, and t, respectively. Industries are defined at the 6-digit product classification level. Details on data cleaning are in Appendix F.

Output prices and quantities. I define a product as a unique combination of the product name reported by the producer, the product code, and the company identifier. The unit price is simply the product-wise sales divided by the product-wise quantity. Then, I take logs and

[18] The disclosure of product-level information was mandatory per the 1956 Companies Act until the general exemption in 2011 (Notification No S.O. 301(E) dated 08/02/2011). As a result, my sample ends in 2011.

calculated price changes $\Delta p_{i,t}$ and quantity changes $\Delta y_{i,t}$. Only manufacturing products are included.

Capacity utilization. The main variable of interest is called "product-wise installed capacity" in Prowess.[19] I calculate capacity utilization $\text{CU}^{\text{raw}}_{i,t}$ as quantity over lagged capacity in (5.1).

$$\text{CU}^{\text{raw}}_{i,t} = \frac{\text{Production Quantity}_{i,t}}{\text{Capacity}_{i,t-1}}. \tag{5.1}$$

To exclude outliers, I only include observations with $\text{CU}^{\text{raw}}_{i,t}$ between the 5th and 95th percentiles. Figure A1 shows the histogram of $\text{CU}^{\text{raw}}_{i,t}$. While one can observe a clear drop at 100% as expected, it is not unusual for firms to operate above 100% utilization.

Next, given the huge sectoral heterogeneity in capacity utilization, I residualize $\text{CU}^{\text{raw}}_{i,t}$ by removing the average capacity utilization $\text{CU}^{\text{ind.}}_{s,t}$ weighted by sales at the industry-year level.

$$\widehat{\text{CU}}_{i,t} = \text{CU}^{\text{raw}}_{i,t} - \text{CU}^{\text{ind.}}_{s,t}, \quad \text{where} \quad \text{CU}^{\text{ind.}}_{s,t} = \text{Weighted Avg. CU}^{\text{raw}}_{i,t}. \tag{5.2}$$

Importantly, the residualized $\widehat{\text{CU}}_{i,t}$ still has considerable product-level heterogeneity, as the unconditional correlation between $\widehat{\text{CU}}_{i,t}$ and $\widehat{\text{CU}}_{i,t-1}$ is as high as 0.76. Consequently, I always control for lagged $\widehat{\text{CU}}_{i,t}$ in the analysis below to separate the cyclical variation within products from the permanent variation between products.

Sales-to-inventory ratio. Given the importance of inventories (e.g., Bils and Kahn, 2000), I include the sales-to-inventory ratio as an extension. Inventories are only available at the firm level, hence I define the sales-to-inventory ratio as the firm-level sales of goods over lagged inventories.

$$\text{Inv}^{\text{raw}}_{f,t} = \frac{\text{Sales of Goods}_{f,t}}{\text{Inventories}_{f,t-1}}. \tag{5.3}$$

Like capacity utilization, a higher $\text{Inv}^{\text{raw}}_{f,t}$ generally indicates stronger demand relative to short-run supply. I remove outliers and residualize $\text{Inv}^{\text{raw}}_{f,t}$ as above. Industry-year averages are computed based on the product's industry rather than the firm's industry.

$$\widehat{\text{Inv}}_{i,t} = \text{Inv}^{\text{raw}}_{f,t} - \text{Inv}^{\text{ind.}}_{s,t}, \quad \text{where} \quad \text{Inv}^{\text{ind.}}_{s,t} = \text{Weighted Avg. Inv}^{\text{raw}}_{f,t}. \tag{5.4}$$

Firm-level input price shock ($\tau_{f,t}$). Prowess reports the price and quantity of each material input at the firm level. I closely follow Liu (2025) to construct the firm-level input price shock

[19] While there might not be standard definition of capacity, the self-reported capacity tends to be extremely stable. 84.8% of the raw data with non-missing capacity show exactly zero change in capacity, whereas 97.9% (99.8%) of the observations show absolute changes in capacity below 1% (5%). Changes in capacity are also lumpy. Conditional on having a positive change in capacity, the median increase is as much as 36.8%. Given these patterns, it is reasonable to think that the "installed capacity" variable, as its name suggests, reflects the constraint due to pre-determined capital inputs such as the size of the factory or the number of assembly lines. The data coverage is also high prior to 2011. In the raw data, 52.3% of product-year observations with non-missing sales contain non-missing capacity, covering 72.2% of total sales in the sample period.

from *major intermediate inputs*, i.e., the top 10% 6-digit product codes ranked by the number of unique buyers in Prowess. In principle, buyers are more likely to be price takers in these major intermediate input markets. Empirically, Liu (2025) shows that cost pass-through estimated using these major intermediate inputs appears highly consistent with results in the literature.

Define the firm-level price change of major inputs, $\Delta\rho^{\text{mj}}_{f,t}$, as the average price change of all major inputs weighted by their average purchase amounts between $t-1$ and t. Next, the firm-level exposure to $\Delta\rho^{\text{mj}}_{f,t}$ is the average expenditure share of major inputs in total variable costs, approximated by the cost of goods sold (COGS), between $t-1$ and t. Finally, the firm-level input price shock, $\tau_{f,t}$, is the product of $\Delta\rho^{\text{mj}}_{f,t}$ and its average expenditure share:

$$\tau_{f,t} = \text{Avg. Expenditure Share}_{f,t} \times \Delta\rho^{\text{mj}}_{f,t}. \tag{5.5}$$

Industry-level demand shock ($\delta_{s,t}$). Similar to Giovanni et al. (2014), I obtain a reduced-form shock, $\delta_{s,t}$, at the industry-year level by calculating the median sales growth (i.e., $\Delta p_{i,t} + \Delta y_{i,t}$) for each 6-digit industry in each year. The main advantage of this measure is that the treatment is homogeneous across industries, unlike alternative measures such as monetary policy shocks.[20] While I refer to $\delta_{s,t}$ as an industry-specific *demand* shock, in principle, it can be an arbitrary combination of synchronized demand, cost, and productivity shocks given its reduced-form nature. As I will show in Section 5.2, empirical results on prices and quantities are highly consistent with the demand shock interpretation. Furthermore, I directly observe firm-level material input costs, which allows for tighter control for cost shocks.

5.2 Product-Level Evidence of Rationing

5.2.1 Empirical Moments

I first plot moments of prices and quantities by capacity utilization. To hold the denominator fixed, I restrict the sample to observations for which capacity growth is less than 1% in absolute value. Detailed statistics are reported in Table A1.

First moment. Panel (a) of Figure 4 presents binned scatter plots of average $\Delta p_{i,t}$ and $\Delta y_{i,t}$ against the z-score of $\widehat{\text{CU}}_{i,t}$. Each of the 20 bins contains 5% of the sample. Quantity growth $\Delta y_{i,t}$ (red triangle) exhibits a strongly concave relationship with capacity utilization, which is hardly surprising. Products experiencing falling demand tend to have both low capacity utilization and declining sales, while products operating near full capacity cannot easily expand output further. Interestingly, the sharp kink occurs right in the middle of the sample, suggesting that almost half of the products are more or less affected by the residualized capacity measure.

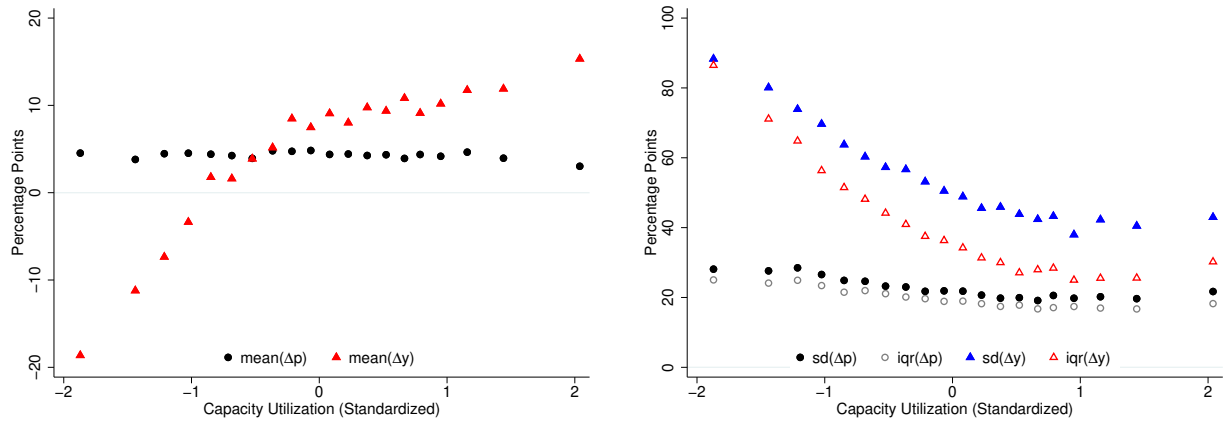
By sharp contrast, the average price growth (black circle) is almost completely unaffected by capacity utilization at the product level. If demand increases approximately linearly in capacity utilization, as suggested by quantity changes in the low-utilization region, then prices must in-

[20] Monetary policy shocks, while highly exogenous, are likely to affect industries differently. E.g., durable-goods industries are more sensitive to interest rates. Monetary policy shocks may also have supply-side effects on capital-intensive industries through the working-capital channel. If such heterogeneity is correlated with capacity utilization, the resulting estimates can be severely biased.

Variable	Definition	N	Mean	S.D.	Min	P10	P25	P50	P75	P90	Max
$\Delta p_{i,t}$	Price growth	53,033	0.036	0.229	-0.781	-0.198	-0.058	0.031	0.134	0.277	0.809
$\Delta y_{i,t}$	Output growth	53,033	-0.002	0.571	-2.508	-0.530	-0.157	0.035	0.217	0.474	2.565
$\delta_{i,t}$	Industry shock	53,033	0.090	0.101	-0.219	-0.021	0.018	0.086	0.158	0.215	0.390
$\tau_{i,t}$	Input shock	53,033	0.026	0.094	-0.262	-0.071	-0.017	0.017	0.065	0.137	0.363
$\Delta\text{Capacity}_{i,t}$	Capacity growth	53,033	0.028	0.088	-0.014	0.000	0.000	0.000	0.000	0.080	0.365
$\text{CU}^{\text{raw}}_{i,t}$	Raw capacity utilization	31,041	0.703	0.354	0.148	0.229	0.413	0.685	0.950	1.218	1.396
$\widehat{\text{CU}}_{i,t}$	Residualized capacity utilization	31,041	-0.135	0.352	-1.181	-0.594	-0.399	-0.145	0.100	0.354	1.057
$\widehat{\text{Inv}}^{\text{raw}}_{f,t}$	Raw inventory ratio	43,940	5.090	2.231	1.128	2.190	3.320	4.922	6.839	8.444	9.011
$\widehat{\text{Inv}}_{i,t}$	Residualized inventory ratio	43,940	-0.312	2.002	-7.831	-2.890	-1.640	-0.263	1.005	2.305	7.698
$\Delta p_{-i,t}$	Competitors' price	31,041	0.043	0.096	-0.221	-0.060	-0.012	0.037	0.095	0.160	0.343
$\Delta\hat{p}_{-i,t \mid \text{High CU}}$	See (5.8)	31,041	-0.002	0.025	-0.098	-0.026	-0.013	-0.002	0.006	0.021	0.102
$\text{Cov}\big(\Delta p_{-i,t}, \widehat{\text{CU}}^{\text{S.D.}}_{-i,t}\big)_{i,t}$	See (5.9)	31,041	-0.009	0.046	-0.226	-0.053	-0.025	-0.007	0.008	0.033	0.181

Notes: The sample corresponds to column (1) of Table 4.

Table 3: Summary Statistics



(a) Average Changes

(b) Dispersion

Notes: I restrict the sample to product-years in which the absolute change in capacity is below 1%. 20 bins for 82,482 observations. In panel (b), SD refers to the standard deviation and IQR to the inter-quartile range. The x-axis is the standardized $\widehat{\text{CU}}_{i,t}$.

Figure 4: Changes in Prices and Quantities

crease in the high-utilization region to clear any unsatisfied demand. That prices appear largely unchanged is a first piece of evidence consistent with rationing.

Second moment. Panel (b) of Figure 4 shows the standard deviations and inter-quartile ranges of $\Delta p_{i,t}$ and $\Delta y_{i,t}$ for each 5% partition. As capacity constraints become more binding, quantity adjustments should become less frequent, especially since I restrict the sample to observations with no change in capacity. Yet in the absence of rationing, smaller quantity fluctuations should imply larger price fluctuations, provided the stochastic properties of shocks

remain unchanged. Empirically, quantity fluctuations decline as capacity utilization increases. By contrast, the volatility of price changes, however measured, remains roughly constant. The absence of elevated price volatility, while difficult to explain under Walrasian market clearing, is a natural implication of rationing with pricing frictions.

5.2.2 Regression Specification

My formal test for rationing builds on the following double-interaction regression.

$$Y_{i,t} = \underbrace{\sum_{m=0}^{2}\sum_{n=0}^{2}\beta_{m,n}\widehat{\mathrm{CU}}^{*}_{i,t-m}\delta_{s,t-n} + \sum_{k=0}^{2}\beta_k\delta_{s,t-h}}_{\text{Industry demand shock}} + \underbrace{\sum_{m=0}^{2}\sum_{n=0}^{2}\gamma_{m,n}\widehat{\mathrm{CU}}^{*}_{i,t-m}\tau_{i,t-n} + \sum_{k=0}^{2}\gamma_k\tau_{i,t-k}}_{\text{Input price shock}}$$
$$+ \sum_{k=0}^{2}\eta_k\widehat{\mathrm{CU}}^{*}_{i,t-k} + \sum_{k=1}^{2}\eta'_k\Delta\text{Capacity}_{i,t-k} + \text{Fixed Effects} + \epsilon_{i,t}, \tag{5.6}$$

where $\widehat{\mathrm{CU}}^{*}_{i,t}$ is normalized $\widehat{\mathrm{CU}}_{i,t}$ defined below. Because rationing pertains to the joint behavior of prices and quantities, I separately estimate (5.6) using $\Delta p_{i,t}$ and $\Delta y_{i,t}$ as dependent variables.

Normalization of $\widehat{\mathrm{CU}}_{i,t}$. To facilitate interpretation, I standardize $\widehat{\mathrm{CU}}_{i,t}$ before estimating (5.6). The first measure is the z-score of $\mathrm{CU}_{i,t}$ within the regression sample. The second measure, which is more robust to outliers, is a centered decile rank of $\widehat{\mathrm{CU}}_{i,t}$. I transform $\widehat{\mathrm{CU}}_{i,t}$ into its decile rank within the sample and recenter the resulting values around zero. The ten deciles take values -4.5, -3.5, …, -0.5, 0.5, …, 3.5, 4.5. A one-unit increase corresponds to moving up one decile, and the median corresponds to zero. I then use this rank measure as a continuous variable in (5.6), while requires much fewer observations than a fully nonparametric categorical specification.

Lag controls. Given the potentially persistent heterogeneity in $\widehat{\mathrm{CU}}_{i,t}$, I include lags of all regressors up to $t-2$ to isolate the cyclical variation in $\widehat{\mathrm{CU}}_{i,t}$.[21] Coefficients are subscripted by the lag order. For example, $\beta_{0,0}$ denotes the interaction between current capacity utilization and the current industry shock $\delta_{i,t}$.

Industry demand shocks. $\beta_{0,0}$ plays a central role in the empirical test. Following a demand shock, both prices and quantities should increase. In the cross-section, low-capacity products should further exhibit a negative quantity gap relative to high-capacity products. To clear excess demand, however, prices of low-capacity products must rise by more to offset this gap. Demand elasticity then determines the magnitude of the required price adjustment. Therefore, the test boils down to whether (i) $\beta_{0,0} > 0$ in the price regression and $\beta_{0,0} < 0$ in the quantity regression, and (ii) their magnitudes are consistent with a plausible demand elasticity.

Input price shocks. The availability of input data in Prowess substantially sharpens my test. First, the double interaction with the input price shock controls for potential heterogeneity in firm-level input costs. Second, input price shocks provide a plausible range for the residual demand elasticity through γ_0 in the price and quantity regressions. Third, the double interaction

[21] In Table A2, I compare regressions with and without lag controls using a simplified specification. While both support the existence of rationing, omitting lag controls tends to overestimate the price response.

tests whether residual demand elasticity varies systematically across high- and low-capacity products through $\gamma_{0,0}$. Ideally, $\gamma_{0,0}$ should be close to zero once sufficient controls are included. If low-capacity products face more *inelastic* demand, for example because utilization is correlated with product appeal, this would only reinforce the evidence for rationing. Conversely, *more elastic* demand poses a threat to identification.

Confounding factors in $\widehat{\text{CU}}_{i,t}$. My identification strategy relies on variation in the tightness of capacity constraints driven by predetermined capacity levels. In the data, however, observed capacity utilization reflects substantial heterogeneity arising from differences in production technologies and product-specific trends. The residualization in (5.2) and the the lagged controls in (5.6) removes industry-year heterogeneity and persistent product-level heterogeneity, thereby improving the comparability of $\widehat{\text{CU}}_{i,t}$ among products.

The remaining confounding factors arise from transitory idiosyncratic shocks entering through the numerator of (5.1). Idiosyncratic demand shocks raise both prices and quantities, whereas rationing dampens both. Therefore, if high-utilization is systematically driven by high idiosyncratic demand, the resulting bias would only strengthen the evidence for rationing. Idiosyncratic supply shocks lower prices but raise quantities, and the opposite effects offset each other in the rationing test.

Fixed effects, clusters, and others. I control for lagged changes in capacity up to $t-2$ (i.e., $\Delta\text{Capacity}_{i,t-1}$ and $\Delta\text{Capacity}_{i,t-2}$). Although adding lagged $\Delta\text{Capacity}_{i,t}$ nearly doubles the R-squared of my quantity regressions, it has almost no effect on β's and γ's. I include fixed effects at the product, firm, 6-digit industry, and year levels. Standard errors are clustered by product, firm, and 6-digit industry. As explained in Appendix F, I also exclude industries with many outliers or poor data coverage.

5.2.3 Results

Effects of demand and supply shocks. Before estimating (5.6), I first regress $\Delta p_{i,t}$ and $\Delta y_{i,t}$ separately on the two shocks without $\widehat{\text{CU}}_{i,t}$ and its interactions. The results are reported in columns (1) and (2) of Table 4. Both prices and quantities rise substantially following positive industry shocks, indicating that $\delta_{s,t}$ mainly reflects common demand shocks. After a 1% positive input price shock, prices increase by 0.40% on average while output falls by 0.29%, implying a residual demand elasticity of 0.73. This estimate is in line with Fujiy et al. (2022), who obtain similar estimates using Indian firm-to-firm transaction data.[22]

Evidence of Rationing. I then estimate (5.6) and report the results results in columns (3)-(6). The estimated coefficient $\beta_{0,0}$ is indistinguishable from zero in the price regression but significantly negative in the quantity regression. Following a positive industry shock, low-capacity (high-utilization) products are no more expensive than other products, despite their substantially constrained supply.

[22] An estimate of 0.73 appears much lower than estimates using U.S. retailer data (e.g., DellaVigna and Gentzkow, 2019; Beck and Lein, 2020; Rosenthal-Kay et al., 2024). One reason is that I only include manufacturing firms and drop all retailers. Fujiy et al. (2022) find that factors such as institution quality and input specificity also affect the firm-level elasticity.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Tightness Measure			**Capacity Utilization**				**Sales-to-Inventory Ratio**			
CU / Inv Normalization			*Decile*		*S.D.*		*Decile*		*S.D.*	
	$\Delta p_{i,t}$	$\Delta y_{i,t}$	$\Delta p_{i,t}$	$\Delta y_{i,t}$	$\Delta p_{i,t}$	$\Delta y_{i,t}$	$\Delta p_{i,t}$	$\Delta y_{i,t}$	$\Delta p_{i,t}$	$\Delta y_{i,t}$
(a) Demand shock										
Industry Demand$_{s,t}$ (β_0)	0.46***	0.56***	0.48***	0.37***	0.49***	0.36***	0.50***	0.46***	0.50***	0.47***
	(0.05)	(0.07)	(0.05)	(0.08)	(0.05)	(0.08)	(0.07)	(0.10)	(0.07)	(0.10)
$\widehat{\text{CU}}^*_{i,t}$ × Demand$_{s,t}$ ($\beta^{cu}_{0,0}$)			0.00	-0.08***	0.01	-0.22***				
			(0.01)	(0.02)	(0.04)	(0.07)				
$\widehat{\text{Inv}}^*_{f,t}$ × Demand$_{s,t}$ ($\beta^{inv}_{0,0}$)							-0.00	-0.07***	-0.01	-0.21***
							(0.01)	(0.02)	(0.02)	(0.05)
(b) Cost shock										
Input Price$_{i,t}$ (γ_0)	0.40***	-0.29***	0.44***	-0.18***	0.44***	-0.17***	0.39***	-0.23***	0.39***	-0.23***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)
$\widehat{\text{CU}}^*_{i,t}$ × Input Price$_{i,t}$ ($\gamma^{cu}_{0,0}$)			-0.00	0.01	-0.00	0.06				
			(0.02)	(0.02)	(0.04)	(0.06)				
$\widehat{\text{Inv}}^*_{f,t}$ × Input Price$_{i,t}$ ($\gamma^{inv}_{0,0}$)							0.01	0.03	0.04	0.06
							(0.01)	(0.02)	(0.02)	(0.06)
(c) Other variables										
$\widehat{\text{CU}}^*_{i,t}$ (η^{cu}_0)			-0.01***	0.11***	-0.04***	0.34***				
			(0.00)	(0.00)	(0.01)	(0.01)				
$\widehat{\text{Inv}}^*_{f,t}$ (η^{inv}_0)							0.00**	0.04***	0.01***	0.11***
							(0.00)	(0.00)	(0.00)	(0.01)
Δ Capacity$_{i,t-1}$	-0.00	0.11***	-0.10***	1.02***	-0.13***	1.20***	-0.01	0.14***	-0.01	0.14***
	(0.01)	(0.04)	(0.02)	(0.05)	(0.02)	(0.05)	(0.01)	(0.04)	(0.01)	(0.04)
N	53,033	53,033	23,733	23,733	23,733	23,733	32,782	32,782	32,782	32,782
R^2	0.230	0.214	0.278	0.386	0.280	0.410	0.243	0.225	0.243	0.226
Within R^2	0.064	0.008	0.098	0.233	0.101	0.262	0.064	0.019	0.064	0.020
Products	8,867	8,867	3,927	3,927	3,927	3,927	5,973	5,973	5,973	5,973
Firms	3,523	3,523	2,097	2,097	2,097	2,097	2,351	2,351	2,351	2,351
6-digit Industries	132	132	75	75	75	75	93	93	93	93
(d) Low-slack products			*Top Decile* ($\beta_0 + 4.5\beta_{0,0}$)		*95th Percentile* ($\beta_0 + 1.663\beta_{0,0}$)		*Top Decile* ($\beta_0 + 4.5\beta_{0,0}$)		*95th Percentile* ($\beta_0 + 1.657\beta_{0,0}$)	
Demand$_{s,t}$			0.49***	0.01	0.51***	-0.01	0.49***	0.14	0.48***	0.12
			(0.08)	(0.13)	(0.09)	(0.15)	(0.08)	(0.11)	(0.07)	(0.11)

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Both $\widehat{\text{CU}}^*_{i,t}$ and $\widehat{\text{Inv}}^*_{f,t}$ are standardized, and their units are reported in the header. In all specifications, fixed effects are included at the product, firm, 6-digit industry, and year levels. Standard errors are clustered by product, firm, and 6-digit industry. In columns (3)-(4) and (7)-(8), I evaluate the tightness measure at its top decile ($\widehat{\text{CU}}^*_{i,t} = 4.5$, $\widehat{\text{Inv}}^*_{f,t} = 4.5$). In columns (5)-(6) and (9)-(10), I evaluate capacity utilization at its 95th percentile ($\widehat{\text{CU}}^*_{i,t} = 1.663$, $\widehat{\text{Inv}}^*_{f,t} = 1.657$).

Table 4: Capacity Utilization and Shock Responses

To assess the economic magnitude, I evaluate the response to demand shocks at the top decile as well as the 95th percentile of the capacity utilization distribution in panel (d) of Table 4. When capacity is in the top decile (columns 3-4), quantities increase by merely 0.01% in response to a 1% industry shock. Prices increase by 0.49%, almost identical to the average response of 0.48%. Results are similar under the z-score specification in columns (5) and (6), suggesting that outliers play a limited role. When capacity utilization is at the 95th percentile, prices increase by 0.51%, while quantities decrease by 0.01%.

What would be the "correct" price response in a standard non-rationing model? Based on estimates in column (3), the market-clearing price increase required for firms in the top utiliza-

tion decile to absorb excess demand would be:

$$0.48 + \frac{0.37 - 0.01}{\epsilon},$$

where ϵ is the residual demand elasticity. The market-clearing price increase would be as high as 0.97%, based on the reduced-form elasticity of 0.73 estimated in columns (1) and (2). The estimated price response of 0.49% from the data falls far short of this benchmark, and no plausible demand elasticity can reconcile such a price gap. I interpret this discrepancy as strong evidence of rationing.

Finally, $\gamma_{0,0}$ indicates whether high-utilization products face more *elastic* demand, which could potentially invalidate my test. The estimate of $\gamma_{0,0}$ is statistically insignificant in both the price and quantity regressions. Only in column (6) is the coefficient positive, suggesting potentially more *inelastic* demand, yet the effect is not statistically significant.

Results using the sales-to-inventory ratio. In columns (7)-(10), I repeat the analysis but using normalized $\widehat{\text{Inv}}_{f,t}$ instead of $\widehat{\text{CU}}_{i,t}$. Both $\widehat{\text{CU}}_{i,t}$ and $\widehat{\text{Inv}}_{f,t}$ are positively correlated with short-run quantity constraints, so the interpretation of the coefficients remains unchanged. Estimates are qualitatively similar: low-inventory firms (high $\widehat{\text{Inv}}_{f,t}$) do not increase prices by more after positive industry shocks, yet they reduce quantities significantly, consistent with rationing.

5.3 Effects of Rationing on Competition

5.3.1 Strategic Complementarities

Following Proposition 1, I examine whether strategic complementarities are affected by competitors' rationing decisions, proxied by their capacity utilization $\widehat{\text{CU}}_{j,t}$. My specification is given in (5.7).

$$Y_{i,t} = \sum_{k=0}^{2} \beta_k \Delta p_{-i,t-h} + \sum_{k=0}^{2} \tilde{\beta}_k X_{i,t-h} + \sum_{k=0}^{2} \gamma_k \tau_{i,t-k} + \sum_{k=1}^{2} \eta_k' \Delta \text{Capacity}_{i,t-k} + \text{FE} + \epsilon_{i,t} \tag{5.7}$$

Specification. The first term, $\Delta p_{-i,t}$, is a leave-one-out, sales-weighted average index of competitors' price changes. Amiti et al. (2019) show that such an index serves as a first-order sufficient statistic for competitors' prices in a broad class of Walrasian models. Competitors are defined as other products in the same 6-digit industry.

The second term, $X_{i,t}$, captures the interaction between rationing and competitors' price changes.[23] In the first version, I use $X_{i,t} = \Delta \hat{p}_{-i,t\,|\,\text{High CU}}$, which is the average deviation from $\Delta p_{-i,t}$ among competitors with above-median utilization $\widehat{\text{CU}}_{j,t}$. The average is weighted by sales, as defined in (5.8). These high-utilization (low-capacity) competitors are more likely to

[23] Qualitatively, specification (5.7) resembles a regression with the interaction term $\Delta p_{-i,t} \times \text{CU}^{\text{ind.}}_{s,t}$. Yet all variation in $\Delta p_{-i,t} \times \text{CU}^{\text{ind.}}_{s,t}$ is at the industry-time level, making it impossible to isolate unobserved industry trends. In contrast, $X_{i,t}$ constructed from prices and residualized utilization $\widehat{\text{CU}}_{j,t}$ exploits the cross-sectional distribution within industry-years.

	(1) $\Delta p_{i,t}$	(2) $\Delta y_{i,t}$	(3) $\Delta p_{i,t}$	(4) $\Delta y_{i,t}$	(5) $\Delta p_{i,t}$	(6) $\Delta y_{i,t}$
(a) Competitors' price						
$\Delta p_{-i,t}$	0.51***	-0.08**	0.58***	-0.10**	0.54***	-0.10**
	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)
$\Delta \hat{p}_{-i,t \mid \text{High CU}}$			-0.98***	0.22		
			(0.15)	(0.14)		
$\text{Cov}\left(\Delta p_{-i,t}, \widehat{\text{CU}}^{\text{S.D.}}_{-i,t}\right)_{i,t}$					-0.30***	0.14
					(0.09)	(0.09)
(b) Other variables						
Input Price$_{i,t}$	0.39***	-0.17***	0.37***	-0.16***	0.38***	-0.16***
	(0.03)	(0.05)	(0.03)	(0.05)	(0.03)	(0.05)
Δ Capacity$_{i,t-1}$	-0.00	0.12***	-0.00	0.12***	-0.00	0.12***
	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)	(0.04)
N	25,786	25,786	25,786	25,786	25,786	25,786
R^2	0.263	0.180	0.273	0.180	0.266	0.180
Within R^2	0.092	0.003	0.104	0.003	0.096	0.003
Products	4,128	4,128	4,128	4,128	4,128	4,128
Firms	2,182	2,182	2,182	2,182	2,182	2,182
6-digit Industries	78	78	78	78	78	78

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: In all specifications, fixed effects are included at the product, firm, 6-digit industry, and year levels. Standard errors are clustered by product, firm, and 6-digit industry. $\Delta p_{-i,t}$ denotes the leave-one-out average price change among competitors, weighted by sales. $\Delta p_{-i,t}$ (High CU) denotes the sales-weighted average price change among competitors with above-median utilization, $\widehat{\text{CU}}_{j,t}$. $\text{Cov}\left(\Delta p_{-i,t}, \widehat{\text{CU}}^{\text{S.D.}}_{-i,t}\right)_{i,t}$ denotes the covariance between competitors' price changes and the standardized (z-score) utilization rates of those competitors.

Table 5: Strategic Complementarities and Competitor Capacity Utilization

ration. If rationing matters for competition, the coefficient $\tilde{\beta}$ should be nonzero.

$$\Delta \hat{p}_{-i,t \mid \text{High CU}} = \sum_{j \neq i} \text{Sales Share}_{j,t} \times \left(\Delta p_{j,t} - \Delta p_{-i,t}\right) \times \mathbf{1}_{\widehat{\text{CU}}_{j,t} > \text{Median}}. \tag{5.8}$$

In the second version, I define $X_{i,t}$ as the sales-weighted covariance between competitors' price changes and their capacity utilization, $\widehat{\text{CU}}_{j,t}$, as in (5.9). To facilitate interpretation, capacity utilization is standardized using its z-score.

$$\text{Cov}\left(\Delta p_{-i,t}, \widehat{\text{CU}}^{\text{S.D.}}_{-i,t}\right)_{i,t} = \sum_{j \neq i} \text{Sales Share}_{j,t} \times \left(\Delta p_{j,t} - \Delta p_{-i,t}\right) \times \widehat{\text{CU}}^{\text{S.D.}}_{-i,t}. \tag{5.9}$$

Similarly, if rationing is irrelevant for competition, the covariance term should have no effect on outcome $Y_{i,t}$. Moreover, as I show later, the covariance specification directly relates to the theoretical mechanism in Section E.1.

Finally, I control for supply-side factors, namely the input price shock $\tau_{i,t}$ and lagged changes in capacity. Lags, fixed effects, and clustering follow (5.6).

Results. Table 5 reports the results. In columns (1) and (2), I first run the regressions without $X_{i,t}$. A 1% increase in competitors' prices is associated with a 0.51% increase in product i's price, consistent with findings in the literature (Amiti et al., 2019; Albagli et al., 2025). As expected, output losses are only 0.08%, far smaller compared to the price increase.

Next, I introduce $\Delta\hat{p}_{-i,t\,|\,\text{High CU}}$, the average deviation from $\Delta p_{-i,t}$ among high-utilization competitors. In the price regression in column (3), $\tilde{\beta} = -0.98$ and is statistically significant at the 1% level, indicating that prices are significantly less responsive to competitors that are more likely to ration. To better interpret the results, consider a simple example in which all competitors have equal weights. Suppose that competitors with below-median utilization do not change prices, whereas competitors with above-median utilization raise prices by 1%. Then, $\Delta p_{-i,t} = 0.5\%$ and $\Delta\hat{p}_{-i,t\,|\,\text{High CU}} = 0.25\%$. The expected price response of i is therefore $0.58 \times 0.5\% - 0.98 \times 0.25\% = 0.045\%$, which is almost completely irresponsive.

In column (5), I instead use the price-utilization covariance and find $\tilde{\beta} = -0.25$. The estimate is highly significant at the 1% level, likely because the covariance measure is more refined and less noisy than a simple binary split.

The empirical finding that firms react less to prices posted by rationing competitors is surprising yet intuitive. It is surprising because it challenges the standard Walrasian result that prices are sufficient statistics for competitive interactions. Yet it is intuitive in rationing models. Once competitors restrict quantities, they effectively withdraw from price competition, rendering their posted prices less relevant. The model in Section 3 provides a rigorous microfoundation for this mechanism.

5.3.2 Cost Pass-through

Next, I test whether industry-wide rationing affects firms' own cost pass-through. Intuitively, softer competition should make it easier to pass costs into prices. The industry-average capacity utilization ($\text{CU}^{\text{ind.}}_{s,t}$ in (5.2)) serves as a coarse proxy for industry-wide rationing. To remove industry-specific permanent heterogeneity in $\text{CU}^{\text{ind.}}_{s,t}$, I subtract the industry-level mean in the sample period $\overline{\text{CU}}^{\text{ind.}}_{s}$.

$$\widehat{\text{CU}}^{\text{ind.}}_{s,t} = \text{CU}^{\text{ind.}}_{s,t} - \overline{\text{CU}}^{\text{ind.}}_{s} \tag{5.10}$$

Specification. Interacting $\widehat{\text{CU}}^{\text{ind.}}_{s,t}$ with the input price shock $\tau_{i,t}$ yields specification (5.11).

$$\begin{aligned} Y_{i,t+h} = &\sum_{m=0}^{2}\sum_{n=0}^{2}\beta_{m,n}\widehat{\text{CU}}^{\text{ind.}*}_{s,t-m}\tau_{i,t-n} + \sum_{k=0}^{2}\beta_k\tau_{i,t-k} + \sum_{k=0}^{2}\gamma_k\widehat{\text{CU}}^{\text{ind.}*}_{s,t-k} \\ &+ \sum_{k=0}^{2}\eta_k\delta_{s,t-h} + \sum_{k=1}^{2}\eta_k'\Delta\text{Capacity}_{i,t-k} + \text{FE} + \epsilon_{i,t} \end{aligned} \tag{5.11}$$

As in Section 5.2.2, $\widehat{\text{CU}}^{\text{ind.}}_{s,t}$ is normalized using the centered decile rank. To study the dynamics, I place cumulative price and quantity changes up to two years on the left-hand side and use the same regression sample for all $h \in \{0,1,2\}$. I further control for the industry demand shock $\delta_{s,t}$

CU Normalization	(1)	(2)	(3)	(4)	(5)	(6)
	Decile					
	$x_{i,t} - x_{i,t-1}$		$x_{i,t+1} - x_{i,t-1}$		$x_{i,t+2} - x_{i,t-1}$	
	Δp	Δy	Δp	Δy	Δp	Δy
(a) Cost shock						
Input Price$_{i,t}$	0.42***	-0.27***	0.33***	-0.26***	0.22***	-0.30***
	(0.03)	(0.05)	(0.04)	(0.08)	(0.04)	(0.08)
Industry $\widehat{\text{CU}}^{*}_{s,t}$ × Input Price$_{s,t}$	0.01	0.01	0.05***	0.00	0.04***	0.04
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)
(b) Other variables						
Industry $\widehat{\text{CU}}^{*}_{s,t}$	-0.00***	0.01***	-0.00	0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Industry Demand$_{s,t}$	0.49***	0.45***	0.43***	0.36***	0.29***	0.42***
	(0.04)	(0.08)	(0.05)	(0.09)	(0.05)	(0.09)
Δ Capacity$_{i,t-1}$	-0.01	0.16***	-0.00	0.12**	-0.02	0.09*
	(0.01)	(0.04)	(0.02)	(0.05)	(0.02)	(0.05)
N	19,142	19,142	19,142	19,142	19,142	19,142
R^2	0.249	0.169	0.342	0.301	0.412	0.418
Within R^2	0.085	0.014	0.064	0.008	0.032	0.007
Products	3,070	3,070	3,070	3,070	3,070	3,070
Firms	1,650	1,650	1,650	1,650	1,650	1,650
6-digit Industries	69	69	69	69	69	69

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: In all specifications, fixed effects are included at the product, firm, 6-digit industry, and year levels. Standard errors are clustered by product, firm, and 6-digit industry. $\widehat{\text{CU}}^{*}_{s,t}$ is normalized as in Section 5.2.2.

Table 6: Cost Pass-Through in High-Utilization Industries

and changes in capacity. Lags, fixed effects, and clustering follow (5.6).

Results. As shown in Table 6, $\widehat{\text{CU}}^{\text{ind.}}_{s,t}$ significantly amplifies cost pass-through without reducing the output, consistent with the hypothesis that rationing softens competition. The effect emerges after one year, likely reflecting switching frictions, and remains significant two years later. Quantitatively, a one-decile increase in $\widehat{\text{CU}}^{\text{ind.}}_{s,t}$ amplifies the price response by 0.05. For a product in the top decile, the price response is expected to be higher by $0.05 \times 4.5 = 0.22$ after one year, corresponding to 66% of the average response. Finally, one potential concern is that $\widehat{\text{CU}}^{\text{ind.}}_{s,t}$ could be driven by strong and persistent industry demand rather than idiosyncrasies of granular competitors. However, this demand-based interpretation is inconsistent with the finding that $\widehat{\text{CU}}^{\text{ind.}}_{s,t}$ by itself has only negligible effects on outcomes.

In Table A3, I additionally interact the input price shock with $\overline{\text{CU}}^{\text{ind.}}_{s}$ and $\widehat{\text{CU}}_{i,t}$, thereby completing the full decomposition of raw capacity utilization $\text{CU}^{\text{raw}}_{i,t}$. Only $\widehat{\text{CU}}^{\text{ind.}}_{s,t}$, which captures cyclical fluctuations in industry-level capacity utilization, amplifies cost pass-through. Main estimates are virtually unchanged in Table A3.

6 Conclusions

This paper develops an Old Keynesian model that incorporates rationing and endogenous competition in a search framework. The model generates several results that depart sharply from standard New Keynesian predictions under excess demand. At the firm level, rationing attenuates price responses by giving firms an additional margin of adjustment. In industry equilibrium, rationing, particularly by low-price firms, softens competition, lowers elasticity and superelasticity, and raises both the consumption price index and posted prices. In general equilibrium, inflation becomes intrinsically more persistent due to the feedback loop between inflation and rationing. High inflation mechanically lowers firms' relative prices and leads to rationing; meanwhile, rationing reinforces inflation by raising households' reservation prices and softening competition. The resulting asymmetry in both magnitude and persistence is quantitatively important for inflation dynamics in a model calibrated to match firm dynamics data.

Empirical evidence at the product level supports both mechanisms. Products with high capacity utilization show no differential price response to demand shocks despite evident quantity constraints, while periods of high industry-level utilization amplify price responses without output losses. Together, these findings suggest that rationing and product-market frictions are quantitatively important for understanding inflation dynamics.

References

Acharya, V., M. Crosignani, T. Eisert, and C. Eufinger (Oct. 2023). *How Do Supply Shocks to Inflation Generalize? Evidence from the Pandemic Era in Europe*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.

Ahlander, E., N. Amberg, and M. Klein (2026). "Do Capacity Constraints Affect the Pass-Through of Monetary Policy to Prices?"

Albagli, E., F. Grigoli, E. Luttini, D. Quevedo, and M. Rojas (Aug. 2025). "Beyond Costs". *IMF Working Papers* 2025.164, p. 1. ISSN: 1018-5941.

Alvarez, F., F. Lippi, and P. Souganidis (2023). "Price Setting With Strategic Complementarities as a Mean Field Game". *Econometrica* 91.6. ISSN: 0012-9682.

Amberg, N., R. Friberg, and C. Syverson (Jan. 2025). *Have We Got News For You: Firm-Level Evidence on the Optimal Choice of Expected Capacity Utilization*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.

Amiti, M., S. Heise, F. Karahan, and A. Şahin (2022). *Pass-through of Wages and Import Prices Has Increased in the Post-COVID Period*. Tech. rep. Federal Reserve Bank of New York Liberty Street Economics.

Amiti, M., O. Itskhoki, and J. Konings (2019). "International Shocks, Variable Markups, and Domestic Prices". *Review of Economic Studies* 86.6. ISSN: 1467937X.

Aruoba, S. B., E. Oue, F. Saffie, and J. Willis (May 2024). *Reviving Micro Real Rigidities: The Importance of Demand Shocks*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.

Atkeson, A. and A. Burstein (2008). "Pricing-to-market, trade costs, and international relative prices". *American Economic Review* 98.5. ISSN: 00028282.

Auclert, A., B. Bardóczy, M. Rognlie, and L. Straub (2021). "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models". *Econometrica* 89.5. ISSN: 0012-9682.

Bai, X., J. Fernández-Villaverde, Y. Li, and F. Zanetti (Feb. 2024). *The Causal Effects of Global Supply Chain Disruptions on Macroeconomic Outcomes: Evidence and Theory*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.

Ball, L., D. Leigh, and P. Mishra (Oct. 2022). *Understanding U.S. Inflation During the COVID Era*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.

Balleer, A. and M. Noeller (2023). "Monetary Policy in the Presence of Supply Constraints: Evidence from German Firm-Level Data". *SSRN Electronic Journal*.

Barkai, S. (2020). "Declining Labor and Capital Shares". *Journal of Finance* 75.5. ISSN: 15406261.

Barro, R. (May 2025). *The Old Keynesian Model*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.

Barro, R. J. and H. I. Grossman (1971). "A General Disequilibrium Model of Income and Employment". *American Economic Review* 61.1. ISSN: 0002-8282.

Barro, R. J. and H. I. Grossman (1974). "Suppressed inflation and the supply multiplier". *Review of Economic Studies* 41.1. ISSN: 1467937X.

Basu, S. (2019). *Are price-cost markups rising in the United States? A discussion of the evidence*.

Beck, G. W. and S. M. Lein (Nov. 2020). "Price elasticities and demand-side real rigidities in micro data and in macro models". *Journal of Monetary Economics* 115, pp. 200–212. ISSN: 03043932.

Bénassy, J. P. (1975). "Neo-Keynesian Disequilibrium Theory in a Monetary Economy". *The Review of Economic Studies* 42.4. ISSN: 00346527.

Benigno, P. and G. B. Eggertsson (2023). "It's Baaack: The Surge in Inflation in the 2020s and the Return of the Non-Linear Phillips Curve".

Benkard, C. L., N. Miller, and A. Yurukoglu (Nov. 2025). *The Rise of Market Power and the Macroeconomic Implications: Comment*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.

Bianchi, J., A. McKay, and N. Mehrotra (Oct. 2024). *How Should Monetary Policy Respond to Housing Inflation?*

Bils, M. (2016). "Deducing markups from stockout behavior". *Research in Economics* 70.2. ISSN: 10909443.

Bils, M. and J. A. Kahn (2000). "What inventory behavior tells us about business cycles". *American Economic Review* 90.3. ISSN: 00028282.

Blanchard, O. J. and N. Kiyotaki (1987). "Monopolistic Competition and the Effects of Aggregate Demand". *American Economic Review* 77.4, pp. 647–666.

Blanco, A., C. Boar, C. Jones, and V. Midrigan (2024a). "Nonlinear Inflation Dynamics in Menu Cost Economies". *Finance and Economics Discussion Series* 2024-005. ISSN: 1936-2854.

Blanco, A., C. Boar, C. Jones, and V. Midrigan (May 2024b). *The Inflation Accelerator*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.

Boehm, C. E. and N. Pandalai-Nayar (2022). "Convex Supply Curves". *American Economic Review*. Vol. 112. 12.

Caldara, D., M. Iacoviello, and D. Yu (2025). "Measuring Shortages Since 1900".

Calvo, G. (1979). "Quasi-Walrasian theories of unemployment". *American Economic Review Papers and Proceedings* 69.2. ISSN: 00028282.

Carlton, D. W. (1977). "Uncertainty, Production Lags, and Pricing." *American Economic Review* 67.1. ISSN: 00028282.

Carlton, D. W. (1982). "The Disruptive Effect of Inflation on the Organization of Markets". *Inflation: Causes and Effects*. Ed. by R. E. Hall. University of Chicago Press. Chap. 6.
Carlton, D. W. (1998). "A Critical Assessment of the Role of Imperfect Competition in Macroeconomics". *Market Behaviour and Macroeconomic Modelling*.
Cavallo, A. and O. Kryvtsov (2023). "What can stockouts tell us about inflation? Evidence from online micro data". *Journal of International Economics* 146. ISSN: 18730353.
Chin, M. (2023). "The Pass-through of Wages to Consumer Prices in the COVID-19 Pandemic: Evidence from Sectoral Data in the U.S." *IMF Working Papers* 2023.233. ISSN: 1018-5941.
Comin, D., R. Johnson, and C. Jones (Apr. 2023). *Supply Chain Constraints and Inflation*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.
Conlone, C. T. and J. H. Mortimer (2013). "Demand estimation under incomplete product availability". *American Economic Journal: Microeconomics* 5.4. ISSN: 19457669.
Cotton, C. D. and V. Garga (Aug. 2022). *The Role of Industrial Composition in Driving the Frequency of Price Change*. Tech. rep. Boston, MA: Federal Reserve Bank of Boston.
Dana, J. D. (1999). "Equilibrium Price Dispersion under Demand Uncertainty: The Roles of Costly Capacity and Market Structure". *The RAND Journal of Economics* 30.4. ISSN: 07416261.
— (2001). "Competition in Price and Availability When Availability is Unobservable". *The RAND Journal of Economics* 32.3. ISSN: 07416261.
De Loecker, J., J. Eeckhout, and G. Unger (2020). *The rise of market power and the macroeconomic implications*.
DellaVigna, S. and M. Gentzkow (Nov. 2019). "Uniform Pricing in U.S. Retail Chains*". *The Quarterly Journal of Economics* 134.4, pp. 2011–2084. ISSN: 0033-5533.
Edmond, C., V. Midrigan, and D. Y. Xu (2023). "How Costly Are Markups?" *Journal of Political Economy* 131.7. ISSN: 1537534X.
Eslava, M., J. Haltiwanger, and N. Urdaneta (Jan. 2024). "The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand, and Wedges". *Review of Economic Studies* 91.1, pp. 259–300. ISSN: 0034-6527.
Flynn, J. P., G. Nikolakoudis, and K. Sastry (2026). *Pricing and Production Without The Invisible Hand*.
Foster, L., J. Haltiwanger, and C. Syverson (2008). "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?" *American Economic Review* 98.1. ISSN: 00028282.
Foster, L., J. Haltiwanger, and C. Tuttle (Sept. 2022). *Rising Markups or Changing Technology?* Tech. rep. Cambridge, MA: National Bureau of Economic Research.
Fujiwara, I. and K. Matsuyama (Apr. 2026). "Competition and the Phillips curve". *Journal of Monetary Economics* 159, p. 103920. ISSN: 03043932.
Fujiy, B. C., D. Ghose, and G. Khanna (2022). "Production Networks and Firm-level Elasticities of Substitution". *SSRN Electronic Journal*.
Gagliardone, L., M. Gertler, S. Lenzu, and J. Tielens (June 2023). *Anatomy of the Phillips Curve: Micro Evidence and Macro Implications*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.
— (2024). "Micro and Macro Cost-Price Dynamics across Inflation Regimes".
Giovanni, J. di, A. A. Levchenko, and I. Mejean (2014). "Firms, Destinations, and Aggregate Fluctuations". *Econometrica* 82.4. ISSN: 0012-9682.
Holden, T. (2024). *Rationing under Sticky Prices*.

Jaffee, D. M. and F. Modigliani (1969). “A Theory and Test of Credit Rationing”. *American Economic Review* 59.5.

Judd, K. L. (1985). “The law of large numbers with a continuum of IID random variables”. *Journal of Economic Theory* 35.1. ISSN: 10957235.

Kaplan, G. and G. Menzio (2016). “Shopping externalities and self-fulfilling unemployment fluctuations”. *Journal of Political Economy* 124.3. ISSN: 1537534X.

Keynes, J. M. (1936). *The general theory of employment, interest, and money*.

Kharroubi, E. and F. Smets (2024). “Monetary policy with profit-driven inflation”.

Kimball, M. S. (1995). “The Quantitative Analytics of the Basic Neomonetarist Model”. *Journal of Money, Credit and Banking* 27.4. ISSN: 00222879.

Klenow, P. J. and J. L. Willis (2016). “Real Rigidities and Nominal Price Changes”. *Economica* 83.331. ISSN: 14680335.

Krolikowski, P. M. and A. H. McCallum (2021). “Goods-market frictions and international trade”. *Journal of International Economics* 129. ISSN: 18730353.

Lamont, O. (1995). “Do "Shortages" Cause Inflation?” *National Bureau of Economic Research Working Paper Series* No. 5402.January.

Lenoir, C., J. Martin, and I. Mejean (2023). “SEARCH FRICTIONS IN INTERNATIONAL GOODS MARKETS”. *Journal of the European Economic Association* 21.1. ISSN: 15424774.

Lenzu, S., D. A. Rivers, J. Tielens, and S. Hu (2024). “Financial Shocks and Productivity: Pricing Response and the TFPR-TFPQ Bifurcation”. *SSRN Electronic Journal*. ISSN: 1556-5068.

Li, T. (2026). “Value of “Availability” during Demand Shocks: Evidence from U.S. Retailers during the COVID-19 Pandemic”.

Liu, Y. (2025). “Financial Constraints and Price Rigidities”.

Markauskas, J., M. O. Ravn, and J. V. Ríos-Rull (2026). “Sticky Prices and Constraints to Produce”.

Matsa, D. A. (2011). “Competition and product quality in the supermarket industry”. *Quarterly Journal of Economics* 126.3. ISSN: 15314650.

McCall, J. J. (1970). “Economics of Information and Job Search”. *The Quarterly Journal of Economics* 84.1. ISSN: 00335533.

Michaillat, P. and E. Saez (2015). “Aggregate demand, idle time, and unemployment”. *Quarterly Journal of Economics* 130.2. ISSN: 15314650.

Mongey, S. (2025). *“Hot” or not? Learning the right lessons from the labor market after COVID*.

Nakamura, E. and J. Steinsson (2008). “Five facts about prices: A reevaluation of menu cost models”. *Quarterly Journal of Economics* 123.4. ISSN: 00335533.

Ohanian, L. and A. Stockman (1994). “Short Run Effects of Money When Some Prices are Sticky”.

Petrosky-Nadeau, N., E. Wasmer, and S. Zeng (June 2016). “Shopping time”. *Economics Letters* 143, pp. 52–60. ISSN: 01651765.

Rosenthal-Kay, J., J. Traina, and U. Tran (2024). “Several Million Demand Elasticities”.

Rotemberg, J. J. and G. Saloner (1986). “A Supergame-Theoretic Model of Price Wars during Booms”. *The American Economic ReviewAmerican Economic Review*.

Rotemberg, J. J. and M. Woodford (1992). “Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity”. *Journal of Political Economy* 100.6. ISSN: 0022-3808.

Salop, S. C. (1979). “Monopolistic Competition with Outside Goods”. *The Bell Journal of Economics* 10.1. ISSN: 0361915X.

Svensson, L. E. O. (1980). “Effective Demand and Stochastic Rationing”. *The Review of Economic Studies* 47.2. ISSN: 00346527.

Wang, O. and I. Werning (2022). “Dynamic Oligopoly and Price Stickiness”. *American Economic Review* 112.8. ISSN: 19447981.

Appendix

A Tables and Figures

Capacity		Δ Price					Δ Quantity			
Decile	**Median**	**SD**	**IQR**	**Skewness**	**Kurtosis**	**Median**	**SD**	**IQR**	**Skewness**	**Kurtosis**
1	3.46	27.87	24.61	-0.01	4.67	-11.55	84.38	77.63	0.24	4.94
2	4.05	27.55	24.23	-0.05	4.81	-4.24	71.83	60.26	0.41	6.15
3	3.89	24.76	21.64	-0.03	5.40	-0.00	62.04	49.54	0.78	7.66
4	4.03	23.13	20.55	-0.06	5.82	0.42	56.97	42.84	0.92	9.08
5	4.06	21.83	19.22	0.05	6.32	2.97	51.85	36.78	1.14	10.75
6	3.97	21.25	18.66	-0.09	6.45	4.22	47.25	32.46	1.02	12.63
7	3.92	19.87	17.59	-0.07	6.87	5.06	44.88	28.44	1.13	14.90
8	3.82	19.87	16.94	-0.24	7.10	5.32	42.83	28.24	1.30	15.24
9	3.97	20.00	17.19	-0.09	7.25	6.88	40.19	25.21	0.98	17.20
10	3.21	20.71	17.41	-0.12	7.00	9.24	41.78	27.86	0.71	15.36

Notes: $N = 84,578$

Table A1: Empirical Moments Conditional On Capacity Utilization

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lags	N	N	Y	Y	N	N	Y	Y
CU Normalization			*Decile*				*S.D.*	
	$\Delta p_{i,t}$	$\Delta y_{i,t}$	$\Delta p_{i,t}$	$\Delta y_{i,t}$	$\Delta p_{i,t}$	$\Delta y_{i,t}$	$\Delta p_{i,t}$	$\Delta y_{i,t}$
Industry Demand$_{s,t}$	0.58***	0.43***	0.57***	0.41***	0.59***	0.33***	0.59***	0.23***
	(0.05)	(0.07)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)
$\text{CU}^{\text{raw}*}_{i,t} \times \text{Demand}_{s,t}$	0.01**	-0.06***	0.00	-0.08***	0.05***	-0.21***	0.03	-0.26***
	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)	(0.06)
$\text{CU}^{\text{raw}*}_{i,t}$	-0.01***	0.06***	-0.01***	0.10***	-0.02***	0.19***	-0.04***	0.31***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
N	23,733	23,733	23,733	23,733	23,733	23,733	23,733	23,733
R^2	0.242	0.258	0.250	0.342	0.244	0.269	0.254	0.365
Within R^2	0.053	0.073	0.063	0.178	0.055	0.087	0.067	0.207
Products	3,927	3,927	3,927	3,927	3,927	3,927	3,927	3,927
Firms	2,097	2,097	2,097	2,097	2,097	2,097	2,097	2,097
6-digit Industries	75	75	75	75	75	75	75	75
Low-slack products			*Top Decile* $(\beta_0 + 4.5\beta_{0,0})$				*95th Percentile* $(\beta_0 + 1.684\beta_{0,0})$	
Demand$_{s,t}$	0.63***	0.16**	0.58***	0.05	0.68***	-0.03	0.64***	-0.22**
	(0.05)	(0.07)	(0.05)	(0.11)	(0.06)	(0.06)	(0.06)	(0.11)

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes:

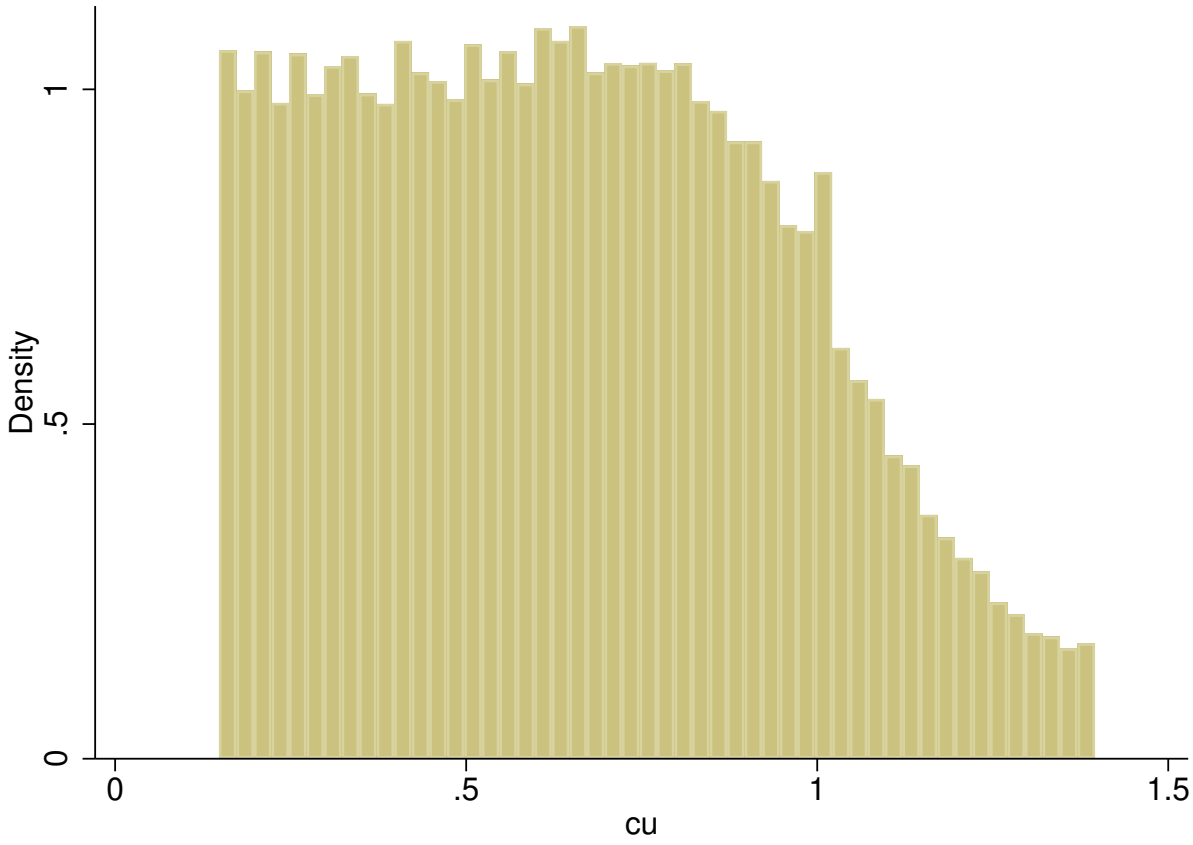
Table A2: Importance of Lagged Controls

CU Normalization	(1)	(2)	(3)	(4)	(5)	(6)
	Decile					
	$x_{i,t} - x_{i,t-1}$		$x_{i,t+1} - x_{i,t-1}$		$x_{i,t+2} - x_{i,t-1}$	
	Δp	Δy	Δp	Δy	Δp	Δy
(a) Cost shock						
Input Price$_{i,t}$	0.40***	-0.16***	0.31***	-0.16***	0.20***	-0.22***
	(0.03)	(0.04)	(0.04)	(0.06)	(0.03)	(0.07)
Industry $\widehat{\text{CU}}^*_{s,t}$ × Input Price$_{s,t}$	0.01	0.00	0.05***	-0.00	0.04***	0.03
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)
Industry $\overline{\text{CU}}^*_{s,t}$ × Input Price$_{s,t}$	-0.01	0.01	-0.02**	0.03*	-0.03**	0.04*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
$\widehat{\text{CU}}^*_{i,t}$ × Input Price$_{s,t}$	-0.01	0.00	-0.01	0.01	-0.02	0.01
	(0.01)	(0.03)	(0.01)	(0.04)	(0.01)	(0.04)
(b) Other variables						
Industry $\widehat{\text{CU}}^*_{s,t}$	-0.01***	0.02***	-0.00**	0.02***	-0.00	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\widehat{\text{CU}}^*_{i,t}$	-0.01***	0.11***	-0.01***	0.08***	-0.00*	0.07***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Industry Demand$_{s,t}$	0.51***	0.26***	0.44***	0.22***	0.29***	0.30***
	(0.04)	(0.06)	(0.05)	(0.08)	(0.05)	(0.08)
Δ Capacity$_{i,t-1}$	-0.15***	1.13***	-0.10***	0.97***	-0.09***	0.87***
	(0.02)	(0.04)	(0.02)	(0.05)	(0.03)	(0.06)
N	19,142	19,142	19,142	19,142	19,142	19,142
R^2	0.263	0.346	0.348	0.377	0.415	0.458
Within R^2	0.102	0.224	0.072	0.117	0.037	0.074
Products	3,070	3,070	3,070	3,070	3,070	3,070
Firms	1,650	1,650	1,650	1,650	1,650	1,650
6-digit Industries	69	69	69	69	69	69

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

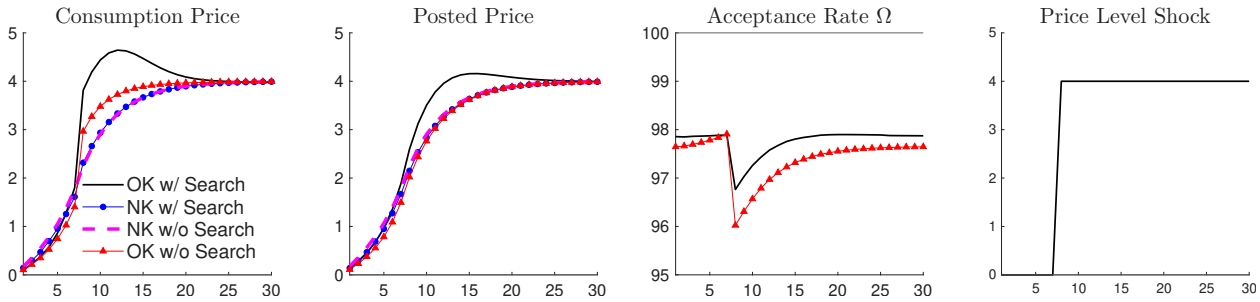
Notes: In all specifications, fixed effects are included at the product, firm, 6-digit industry, and year levels. Standard errors are clustered by product, firm, and 6-digit industry. $\widehat{\text{CU}}^*_{s,t}$ is normalized as in Section 5.2.2.

Table A3: Cost Pass-Through in High-Utilization Industries – Extended Version

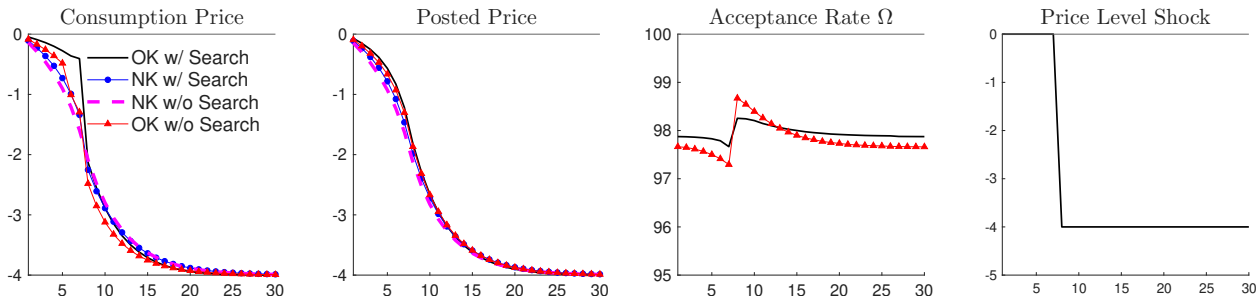


Notes: Prices are in levels.

Figure A1: Capacity Utilization Ratio



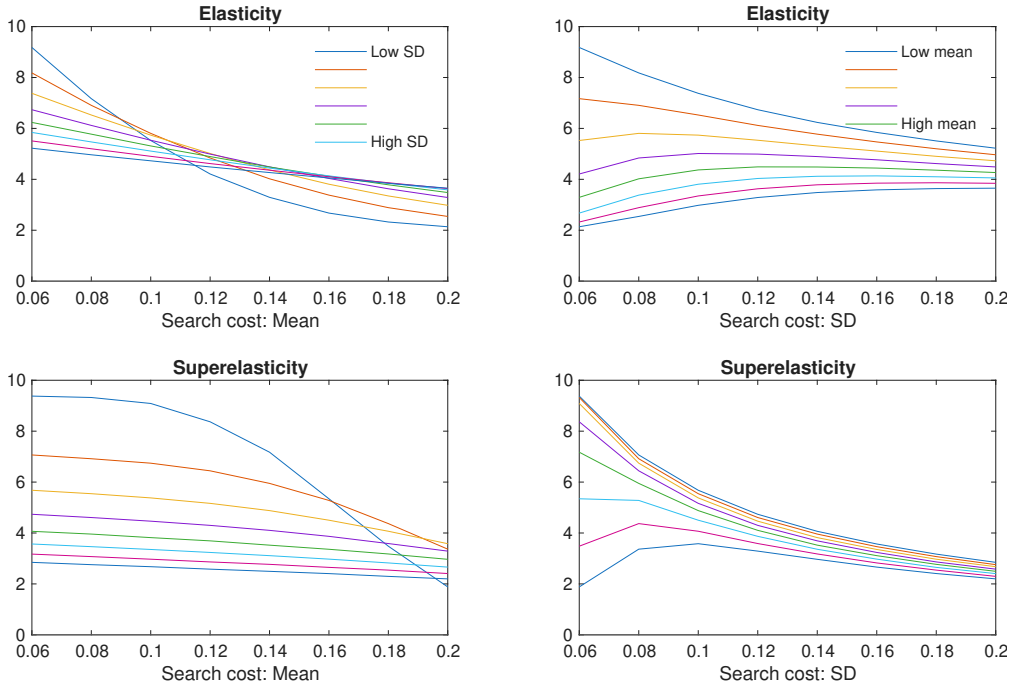
(a) Positive Inflation Shock



(b) Negative Inflation Shock

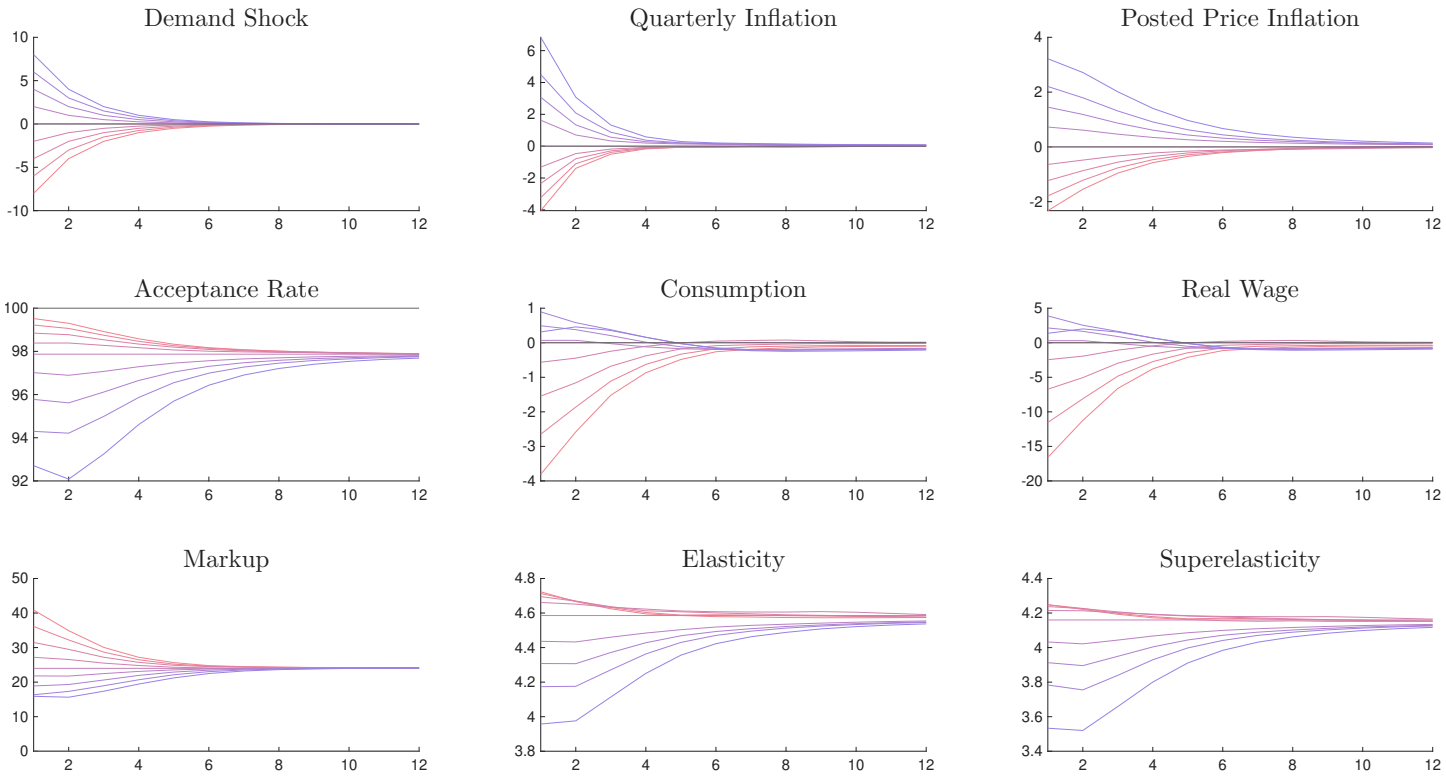
Notes: Prices are in levels.

Figure A2: Anticipated Inflation Shocks in Partial Equilibrium

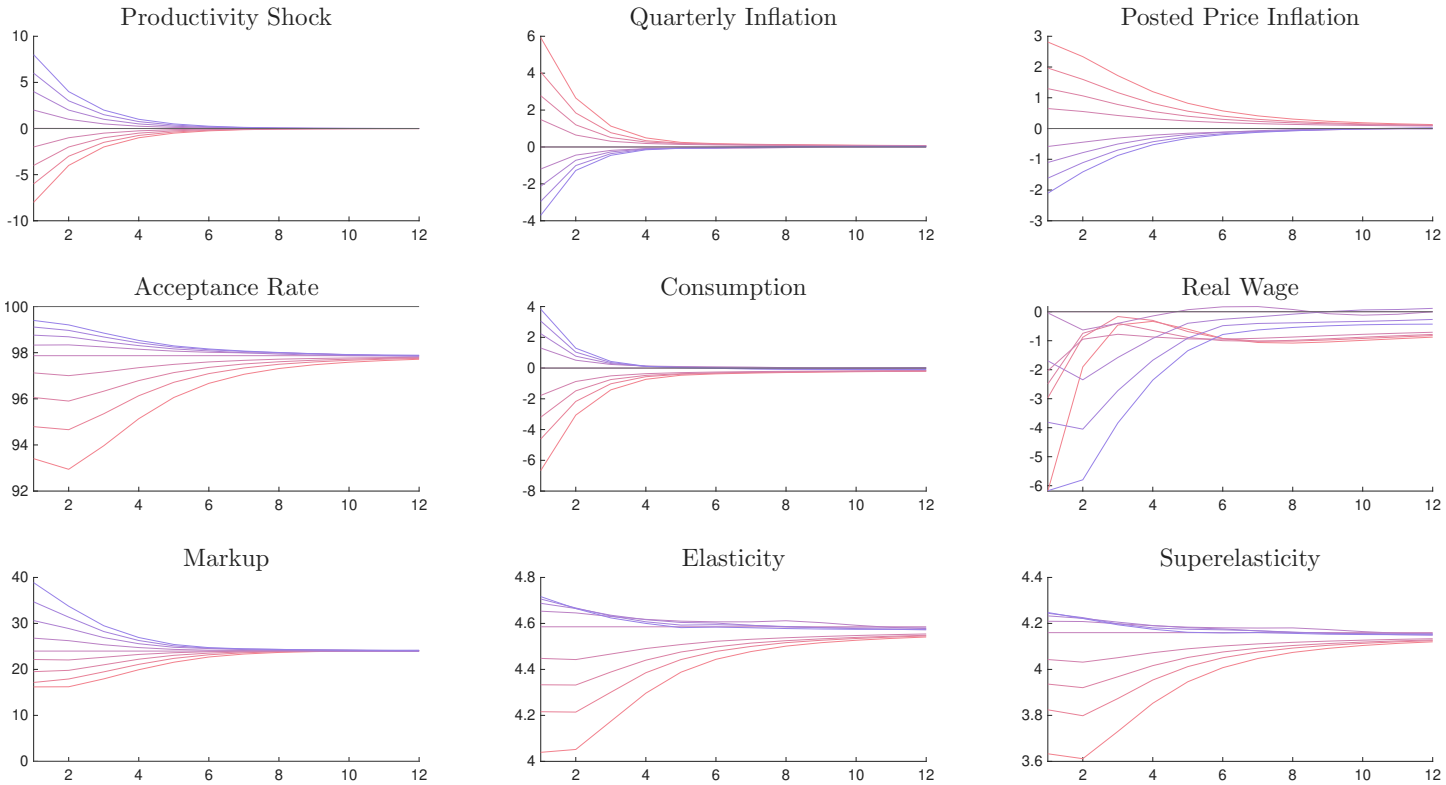


Notes: The figure is for illustration only. To generate the results efficiently, I assume that posted prices follow a normal distribution with $\sigma = 0.2$ and that no firm rations. In other words, it shows the partial equilibrium responses on the demand side while fixing the firm side constant.

Figure A3: Search Cost Distribution



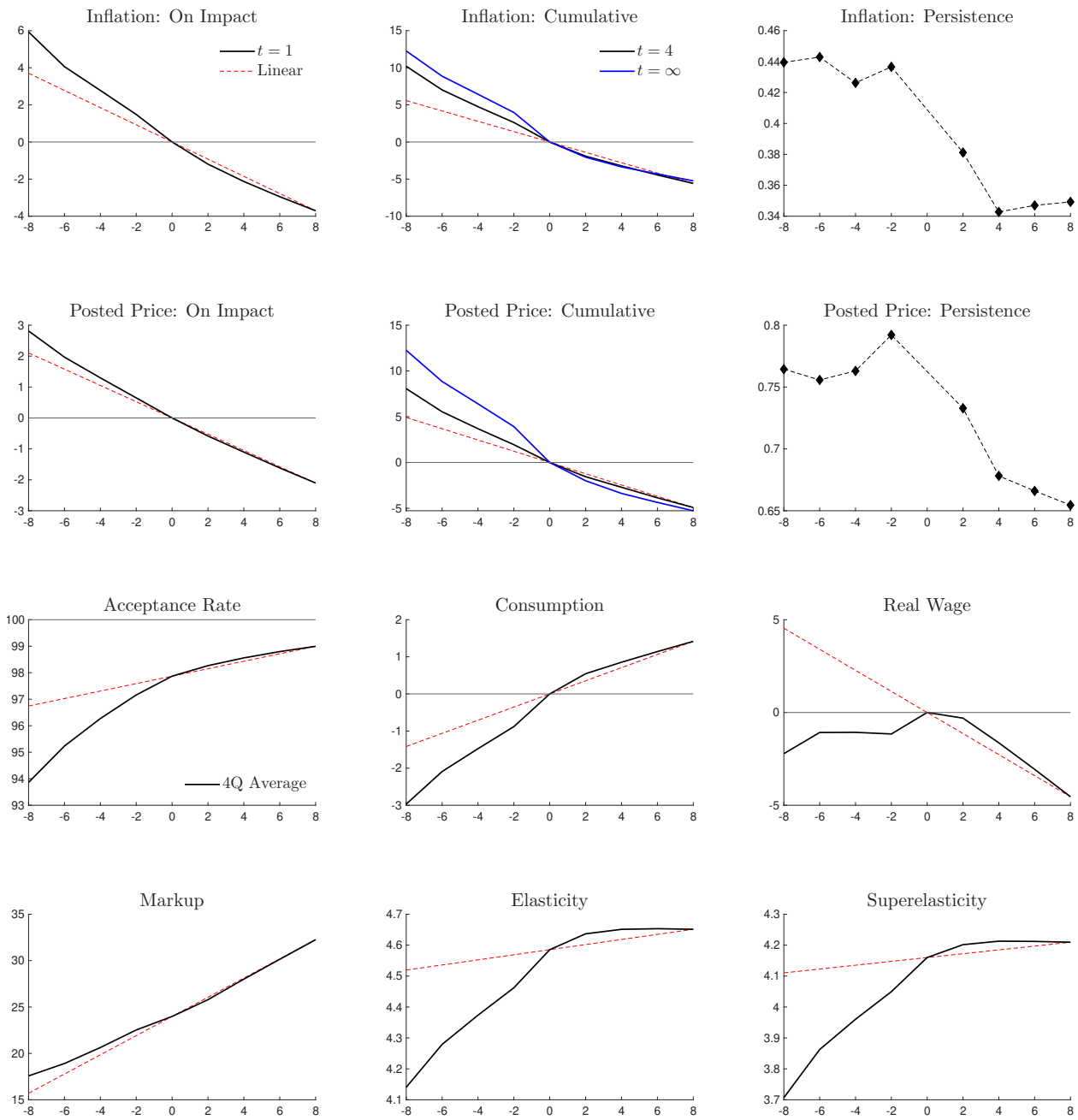
(a) Demand Shock



(b) Productivity Shock

Notes:

Figure A4: Transitional Dynamics After Shocks



Notes: Transitional paths are shown in Figure A4.

Figure A5: Impulse Responses Following Productivity Shocks

B Partial Equilibrium Model

B.1 Derivation

Given that firms posting the same price may choose different acceptance rates, let $\Omega(P)$ denote the *average* acceptance rate among firms posting price P:

$$\Omega(P) = \mathrm{E}(\Omega_s \mid P_s = P) = \frac{\int \Omega_s \mathbf{1}_{P_s=P} ds}{\int \mathbf{1}_{P_s=P} ds}. \tag{B.1}$$

and

$$\mathbf{P}\nu_k = \int^{\bar{P}_k} \Omega(P)(\bar{P}_k - P) dF(P). \tag{B.2}$$

and easy to see

$$\frac{\partial \nu(\bar{P}_k)}{\partial \bar{P}_k} = \frac{1}{\mathbf{P} N(\bar{P}_k)}. \tag{B.3}$$

B.2 Proof of Proposition 1

When υ is differentiable on $\mathbb{R}_+$,

$$\frac{\partial M(P_s)}{\partial P_j} = -\Big(\int_{\max(P_j,P_s)} \frac{\upsilon'(\nu(\bar{P}))}{\mathbf{P}^2} d\bar{P}\Big)\Omega_j \tag{B.4}$$

and

$$\frac{\partial \epsilon^M(P_s)}{\partial P_j} = \Big(\frac{\upsilon'(\nu(P_s))}{\upsilon(\nu(P_s))}\frac{\partial \nu(P_s)}{\partial P_j} - \frac{1}{M(P_s)}\frac{\partial M(P_s)}{\partial P_j}\Big)\epsilon^M(P_s) \tag{B.5}$$

When υ is not differentiable at certain points, as in the case of a uniform distribution, the multiplicative property can still be obtained by approximating υ with a differentiable function and taking the limit.

B.3 Uniformly distributed ν

Suppose that no firm sets a price above the maximum reservation price, $\bar{P}(\upsilon^{-1})$.

$$M(P_s) = \int_{\nu(P_s)}^{\frac{1}{\upsilon}} N(\bar{P}(\nu))\upsilon d\nu = \int_{P_s}^{\bar{P}(\upsilon^{-1})} N(\bar{P}(\nu))\upsilon \frac{\partial \nu(P)}{\partial P} dP = \frac{\upsilon}{\mathbf{P}}\big[\bar{P}(\upsilon^{-1}) - P_s\big]. \tag{B.6}$$

and

$$\bar{P}(v^{-1}) = \check{\mathbf{P}} + \frac{\mathbf{P}}{v}\frac{1}{\boldsymbol{\Omega}}, \tag{B.7}$$

$$M(P_s) = \frac{1}{\boldsymbol{\Omega}} + \frac{\check{\mathbf{P}} - P_s}{\mathbf{P}}v \tag{B.8}$$

$$= \frac{1}{\boldsymbol{\Omega}} + \frac{\check{\mathbf{P}} - P_s}{\check{\mathbf{P}}}\frac{\check{\mathbf{P}}}{\mathbf{P}}v \tag{B.9}$$

First-order approximation around $\check{\mathbf{P}} = P_s$:

$$\frac{1}{M(\check{\mathbf{P}})}dM(P_s) \approx -\frac{1}{\boldsymbol{\Omega}^2 M(\check{\mathbf{P}})}d\boldsymbol{\Omega} + (d\check{\mathbf{p}} - dp_s)\frac{v\check{\mathbf{P}}}{\mathbf{P}M(\check{\mathbf{P}})} \tag{B.10}$$

$$d\ln[M(P_s)] \approx -\frac{1}{\boldsymbol{\Omega}M(\check{\mathbf{P}})}d\boldsymbol{\omega} + (d\check{\mathbf{p}} - dp_s)\epsilon^M(\check{\mathbf{P}}) \tag{B.11}$$

Moreover,

$$M(\check{\mathbf{P}}) = \frac{1}{\boldsymbol{\Omega}} + \frac{\check{\mathbf{P}} - \check{\mathbf{P}}}{\mathbf{P}}v = \frac{1}{\boldsymbol{\Omega}} \tag{B.12}$$

$$d\ln[M(P_s)] \approx -d\boldsymbol{\omega} + (d\check{\mathbf{p}} - dp_s)\epsilon^M(\check{\mathbf{P}}) \tag{B.13}$$

B.3.1 Determinants of P

$$\mathbf{P}^{1-\sigma} = \int \check{P}(\bar{P}(v))^{1-\sigma}dV(v) \tag{B.14}$$

$$= \check{\mathbf{P}}^{1-\sigma}\int\left(1 - \frac{\check{\mathbf{P}} - \check{P}(\bar{P}(v))}{\check{\mathbf{P}}}\right)^{1-\sigma}dV(v) \tag{B.15}$$

$$\check{\mathbf{p}} - \mathbf{p} \approx \int \frac{\check{\mathbf{P}} - \check{P}(\bar{P}(v))}{\check{\mathbf{P}}}dV(v) \tag{B.16}$$

Note that

$$\int \check{P}(\bar{P}(v))dV(v) = \frac{v}{\mathbf{P}}\int_{P_{\min}}^{\bar{P}(v^{-1})} \check{P}(P)\frac{1}{N(P)}dP \tag{B.17}$$

$$= \frac{v}{\mathbf{P}}\int_{P_{\min}}^{\bar{P}(v^{-1})}\int_{P_{\min}}^{P}\Omega(P')P'dF(P')dP \tag{B.18}$$

$$= \frac{v}{\mathbf{P}}\int_{P_{\min}}^{\bar{P}(v^{-1})}\left(\int_{P'}^{\bar{P}(v^{-1})}dP\right)\Omega(P')P'dF(P') \tag{B.19}$$

$$= \frac{v}{\mathbf{P}}\int_{P_{\min}}^{\bar{P}(v^{-1})}\left(\bar{P}(v^{-1}) - P'\right)\Omega(P')P'dF(P') \tag{B.20}$$

$$= \frac{v}{\mathbf{P}}\bar{P}(v^{-1})\check{\mathbf{P}}\boldsymbol{\Omega} - \frac{v}{\mathbf{P}}\int_{P_{\min}}^{\bar{P}(v^{-1})}\Omega(P')(P')^2dF(P') \tag{B.21}$$

Observe that

$$\int \Omega(P)(P-\check{\mathbf{P}})^2 dF(P) = \int \Omega(P)P^2 dF(P) + \check{\mathbf{P}}^2 \int \Omega(P) dF(P) - 2\check{\mathbf{P}} \int \Omega(P) P dF(P) \tag{B.22}$$

$$= \int \Omega(P)P^2 dF(P) + \check{\mathbf{P}}^2 \mathbf{\Omega} - 2\check{\mathbf{P}}^2 \mathbf{\Omega} \tag{B.23}$$

$$= \int \Omega(P)P^2 dF(P) - \check{\mathbf{P}}^2 \mathbf{\Omega} \tag{B.24}$$

Define the Ω-weighted price dispersion $\mathrm{Var}_\Omega(P)$:

$$\mathrm{Var}_\Omega(P) = \int \frac{\Omega(P)}{\mathbf{\Omega}} (P-\check{\mathbf{P}})^2 dF(P) \tag{B.25}$$

and

$$\int \check{P}(\bar{P}(\nu)) dV(\nu) = \frac{\upsilon}{\mathbf{P}} \bar{P}(\upsilon^{-1}) \check{\mathbf{P}} \mathbf{\Omega} - \frac{\upsilon}{\mathbf{P}} [\check{\mathbf{P}}^2 \mathbf{\Omega} + \mathbf{\Omega} \mathrm{Var}_\Omega(P)] \tag{B.26}$$

$$= \check{\mathbf{P}} - \frac{\upsilon}{\mathbf{P}} \mathbf{\Omega} \mathrm{Var}_\Omega(P) \tag{B.27}$$

and

$$\check{\mathbf{p}} - \mathbf{p} = \upsilon \mathbf{\Omega} \frac{\mathrm{Var}_\Omega(P)}{\check{\mathbf{P}} \mathbf{P}} \tag{B.28}$$

C General Equilibrium Model

C.1 Households

C.1.1 Static Version

Objective function. I drop the k subscript in this section.

$$U = \max_{\{\bar{P}_i, L\}} \mathrm{E}_{\{P_i\}} \left\{ \max_{\{Q_i\}} \left(\frac{C^{1-\gamma}}{1-\gamma} \right) \right\} - \int \Psi_i \nu_i N_i(\bar{P}_i) di, \tag{C.1}$$

subject to

$$E = \int P_i Q_i di. \tag{C.2}$$

Define

$$G_i(P_i) = \left[N_i(P_i) \right]^{-1} = \int^{P_i} \Omega_i(P) dF_i(P) \tag{C.3}$$

Easy to see that given the reservation price $\bar{P}_i$, the conditional CDF and PDF of drawing $P_i \in (0, \bar{P}_i]$ are $\frac{G_i(P_i)}{G_i(\bar{P}_i)}$ and $\frac{\Omega_i(P_i) f_i(P_i)}{G_i(\bar{P}_i)}$, respectively.

Second stage. Conditional on $\{\bar{P}_i\}$ and $\{P_i\}$, maximizing $\frac{C^{1-\gamma}}{1-\gamma}$ subject to $E = \int P_i Q_i di$ gives the standard CES cost minization results:

$$\begin{aligned} Q_i &= \left(\frac{P_i}{P(\{P_i\})} \right)^{-\sigma} C(\{P_i\}), \quad P(\{P_i\}) = \left[\int P_i^{1-\sigma} di \right]^{\frac{1}{1-\sigma}}, \quad C(\{P_i\}) = \frac{E}{P(\{P_i\})}, \\ \frac{\partial C(\{P_i\})}{\partial P_i} &= -\frac{C(\{P_i\})}{P(\{P_i\})} \left(\frac{P_i}{P(\{P_i\})} \right)^{-\sigma}, \quad \frac{\partial}{\partial P_i} \frac{C(\{P_i\})^{1-\gamma}}{1-\gamma} = -C(\{P_i\})^{-\gamma} \frac{Q_i}{P(\{P_i\})}. \end{aligned} \tag{C.4}$$

Define $C(X; \{P_{-i}\})$, $P(X; \{P_{-i}\})$, $Q_i(X; \{P_{-i}\})$ as the functions with $P_i = X$ and all other P_j fixed.

First stage. Expand the first term:

$$\mathrm{E}_{\{P_i\}} \left\{ \max_{\{Q_i\}} \left(\frac{C^{1-\gamma}}{1-\gamma} \right) \right\} = \int^{\bar{P}_\infty} \cdots \int^{\bar{P}_i} \cdots \int^{\bar{P}_1} \left\{ \max_{\{Q_i\}} \left(\frac{C^{1-\gamma}}{1-\gamma} \right) \right\} \Pi_{j=1}^{\infty} \left(N_i(\bar{P}_j) \Omega_j(P_j) dF(P_j) \right). \tag{C.5}$$

Note that the last part the the product of the conditional PDF of all $P_i \le \bar{P}_i$ in all markets.

Differentiate $\mathrm{E}_{\{P_i\}} \left(\frac{C(\{P_i\})^{1-\gamma}}{1-\gamma} \right)$ with respect to $\bar{P}_i$ using the Leibniz rule:

$$\frac{\partial}{\partial \bar{P}_i} \mathrm{E}_{\{P_i\}} \left(\frac{C(\{P_i\})^{1-\gamma}}{1-\gamma} \right) = \mathrm{E}_{\{P_{-i}\}} N_i(\bar{P}_i) \Omega_i(\bar{P}_i) f_i(\bar{P}_i) \left[\frac{C(\bar{P}_i; \{P_{-i}\})^{1-\gamma}}{1-\gamma} - \underbrace{\mathrm{E}_{P_i} \left(\frac{C(P_i; \{P_{-i}\})^{1-\gamma}}{1-\gamma} \right)}_{\text{Average over } P_i, \text{ given } \{P_{-i}\}} \right]. \tag{C.6}$$

Meanwhile,

$$\frac{\partial}{\partial \bar{P}_i}\int \nu_j N_i(\bar{P}_j)dj = \nu_i \frac{\partial}{\partial \bar{P}_i} N_i(\bar{P}_i) = -\nu_i \Omega_i(\bar{P}_i) f_i(\bar{P}_i) N_i(\bar{P}_i)^2. \tag{C.7}$$

Approximation. At first order

$$\mathrm{E}_{P_i}\left(\frac{C(P_i;\{P_{-i}\})^{1-\gamma}}{1-\gamma}\right) \approx \left(\frac{C(\check{P}_i(\bar{P}_i);\{P_{-i}\})^{1-\gamma}}{1-\gamma}\right) \tag{C.8}$$

and more generally

$$\mathrm{E}_{\{P_i\}} X(\{P_i\}) \approx X(\{\check{P}_i(\bar{P}_i)\}). \tag{C.9}$$

Back to $\frac{\partial}{\partial \bar{P}_i}\mathrm{E}_{\{P_i\}}\left(\frac{C(\{P_i\})^{1-\gamma}}{1-\gamma}\right)$:

$$\frac{C(\bar{P}_i;\{P_{-i}\})^{1-\gamma}}{1-\gamma} - \mathrm{E}_{P_i}\left(\frac{C(P_i;\{P_{-i}\})^{1-\gamma}}{1-\gamma}\right) \tag{C.10}$$

$$\approx C(\check{P}_i(\bar{P}_i);\{P_{-i}\})^{1-\gamma}\left[\frac{\left(C(\bar{P}_i;\{P_{-i}\})/C(\check{P}_i(\bar{P}_i);\{P_{-i}\})\right)^{1-\gamma}-1}{1-\gamma}\right] \tag{C.11}$$

$$\approx C(\check{P}_i(\bar{P}_i);\{P_{-i}\})^{1-\gamma}\frac{\partial c(\check{P}_i(\bar{P}_i);\{P_{-i}\})}{\partial \check{P}_i(\bar{P}_i)}\left(\bar{P}_i - \check{P}_i(\bar{P}_i)\right) \tag{C.12}$$

$$= -\frac{C(\check{P}_i(\bar{P}_i);\{P_{-i}\})^{1-\gamma}\check{P}_i(\bar{P}_i)^{-\sigma}}{P(\check{P}_i(\bar{P}_i);\{P_{-i}\})^{1-\sigma}}\left(\bar{P}_i - \check{P}_i(\bar{P}_i)\right) \tag{C.13}$$

$$= -\frac{C(\check{P}_i(\bar{P}_i);\{P_{-i}\})^{-\gamma}Q_i(\check{P}_i(\bar{P}_i);\{P_{-i}\})}{P(\check{P}_i(\bar{P}_i);\{P_{-i}\})}\left(\bar{P}_i - \check{P}_i(\bar{P}_i)\right). \tag{C.14}$$

So

$$\frac{\partial}{\partial \bar{P}_i}\mathrm{E}_{\{P_i\}}\left(\frac{C(\{P_i\})^{1-\gamma}}{1-\gamma}\right) \approx -\mathrm{E}_{\{P_{-i}\}} N_i(\bar{P}_i)\Omega_i(\bar{P}_i) f_i(\bar{P}_i)\frac{C(\check{P}_i(\bar{P}_i);\{P_{-i}\})^{-\gamma}Q_i(\check{P}_i(\bar{P}_i);\{P_{-i}\})}{P(\check{P}_i(\bar{P}_i);\{P_{-i}\})}\left(\bar{P}_i - \check{P}_i(\bar{P}_i)\right) \tag{C.15}$$

$$\approx -N_i(\bar{P}_i)\Omega_i(\bar{P}_i) f_i(\bar{P}_i)\frac{C(\{\check{P}_i(\bar{P}_i)\})^{-\gamma}Q_i(\{\check{P}_i(\bar{P}_i)\})}{P(\{\check{P}_i(\bar{P}_i)\})}\left(\bar{P}_i - \check{P}_i(\bar{P}_i)\right) \tag{C.16}$$

And equating the marginal benefit and the marginal search cost:

$$\frac{\partial}{\partial \bar{P}_i}\mathrm{E}_{\{P_i\}}\left(\frac{C(\{P_i\})^{1-\gamma}}{1-\gamma}\right) = \frac{\partial}{\partial \bar{P}_i}\int \Psi_j N_i(\bar{P}_j)\nu_j dj \tag{C.17}$$

$$-N_i(\bar{P}_i)\Omega_i(\bar{P}_i) f_i(\bar{P}_i)\frac{C(\{\check{P}_i(\bar{P}_i)\})^{-\gamma}Q_i(\{\check{P}_i(\bar{P}_i)\})}{P(\{\check{P}_i(\bar{P}_i)\})}\left(\bar{P}_i - \check{P}_i(\bar{P}_i)\right) \approx -\Psi_i \nu_i \Omega_i(\bar{P}_i) f_i(\bar{P}_i) N_i(\bar{P}_i)^2. \tag{C.18}$$

Let Ψ_i be the expected amount of Q_i multiplied by the marginal consumption of utility

$$\Psi_i = \frac{C(\{\check{P}_i(\bar{P}_i)\})^{-\gamma} Q_i(\{\check{P}_i(\bar{P}_i)\})}{P(\{\check{P}_i(\bar{P}_i)\})} \tag{C.19}$$

At first order:

$$\bar{P}_i \approx \check{P}(\bar{P}_i) + P(\{\check{P}_i(\bar{P}_i)\}) N_i(\bar{P}_i) \nu_i \tag{C.20}$$

C.2 Firms

C.2.1 Optimal Pricing

Probability of not resetting prices: θ_c. X^* is the value conditional on resetting its price. The objective function:

$$\begin{aligned} V_{s,t}(\tilde{P}_{s,t-1}) = \max_{\Omega_{s,t}, \tilde{P}^*_{s,t}, \Omega^*_{s,t}} & (1-\theta_c)\Big[Y^*_{s,t} \tilde{P}^*_{s,t} - C_{s,t}\big(Y^*_{s,t}\big) + \mathrm{E}_t \beta V_{s,t+1}(\tilde{P}^*_{s,t}) \Big] \\ & + \theta_c \Big[Y_{s,t} \tilde{P}_{s,t-1} - C_{s,t}\big(Y_{s,t}\big) + \mathrm{E}_t \beta V_{s,t+1}(\tilde{P}_{s,t-1}) \Big], \end{aligned} \tag{C.21}$$

subject to $\omega_{s,t} = \ln \Omega_{s,t} \leq 0$. The Lagrangian as above with $\phi_{s,t}$ ($\phi^*_{s,t}$) for rationing.

FOCs. Differentiate w.r.t. $\tilde{P}^*_{s,t}$ and normalize by $Y^*_{s,t}$:

$$0 = (1-\epsilon^*_{s,t}) + C'_{s,t} \frac{1}{\tilde{P}^*_{s,t}} \epsilon^*_{s,t} + \beta \frac{\mathrm{E}_t V'_{s,t+1}(\tilde{P}^*_{s,t})}{Y^*_{s,t}} \tag{C.22}$$

Differentiate w.r.t. $\Omega_{s,t}$ ($\Omega^*_{s,t}$) as above.

$$\tilde{P}_{s,t-1} = C'_{s,t} + \frac{\phi_{s,t}}{Y_{s,t}}; \quad \tilde{P}^*_{s,t} = C'_{s,t} + \frac{\phi_{s,t}}{Y^*_{s,t}} \tag{C.23}$$

Differentiate w.r.t. $\tilde{P}_{s,t-1}$ and normalize by $Y_{s,t}$:

$$\frac{1}{\theta_c} \frac{V'_{s,t}(\tilde{P}_{s,t-1})}{Y_{s,t}} = (1-\epsilon_{s,t}) + C'_{s,t} \frac{1}{\tilde{P}_{s,t-1}} \epsilon_{s,t} + \beta \frac{\mathrm{E}_t V'_{s,t+1}(\tilde{P}_{s,t-1})}{Y_{s,t}} \tag{C.24}$$

Marginal cost

$$C_{s,t}(Y_{s,t}) = W^r_t \left(\frac{Y_{s,t}}{A_{s,t}} \right)^{\frac{1}{1-\alpha}}. \tag{C.25}$$

$$C'_{s,t}(Y_{s,t}) = \frac{1}{1-\alpha} W^r_t A_{s,t}^{-\frac{1}{1-\alpha}} Y_{s,t}^{\frac{\alpha}{1-\alpha}} \tag{C.26}$$

D More on First-Order Approximation

D.1 Industry Equilibrium Under Calvo-Style Pricing

Let $1-\theta$ be the share of resetters. Assume random selection (Calvo). The indicator is $z_s \in \{0,1\}$.

Assume that resetters never ration (their $d\Omega = 1-\Omega$). Around non-rationing pre-shock level ($\Omega = 1$).

Define $\omega_s = \ln \Omega_s$.

$$d\check{\mathbf{p}} \approx \int z_s dp_s^* ds \tag{D.1}$$

$$d\boldsymbol{\omega} \approx \int (1-z_s) d\omega_s ds \tag{D.2}$$

and

$$dp_s^* \approx \frac{1}{1+\Gamma} dmc_s + \frac{\Gamma}{1+\Gamma}\left[d\check{\mathbf{p}} + \frac{-d\boldsymbol{\omega} + d(\mathbf{p}-\check{\mathbf{p}})}{1+\epsilon^M}\right] \tag{D.3}$$

and

$$\check{\mathbf{p}} \approx \frac{1-\theta}{1-(1-\theta)\frac{\Gamma}{1+\Gamma}}\left(\frac{1}{1+\Gamma}\mathrm{E}(dmc) + \frac{\Gamma}{1+\Gamma}\frac{-d\boldsymbol{\omega} + d(\mathbf{p}-\check{\mathbf{p}})}{1+\epsilon^M}\right) \tag{D.4}$$

E More on the Mechanisms

E.1 Demand System

I first examine the demand system in partial equilibrium. The producer's problem (3.11) implies a strong link between prices and rationing: low-price firms (low P_s) have the strongest incentive to ration (low Ω_s). From the buyer's perspective, a high P-Ω covariance mechanically raise the average price $\check{P}(v)$ in (2.3), altering both the reservation price and the demand curve.

To illustrate this mechanism, I run a partial-equilibrium exercise that fixes the nominal posted price distribution as well as aggregate nominal spending. Varying the distribution of Ω_s changes buyers' reservation prices and input costs, leading to a different price of the consumption bundle. For the calibration, I adopt parameter values and the steady-state distribution of posted prices from Section 4.3.

Firm-specific rationing. In the first exercise, I construct the distribution of Ω_s using:

$$\Omega_s = \min\left(1,\ \bar{\mathscr{M}}\frac{P_s^{\frac{1-\alpha}{\alpha}}}{M(P_s)Q(P_s)}\right), \tag{E.1}$$

where $\alpha = 0.3$ and $P_s \sim N(0,\sigma_p)$. The functions $M(\cdot)$ and $Q(\cdot)$ are defined in (2.8) and (**??**). The free parameter $\bar{\mathscr{M}}$ pins down the average level of Ω_s in industry equilibrium.

Equation (E.1) provides an intuitive approximation to the optimal Ω_s implied by (3.8) and (3.11). I then lower $\bar{\mathscr{M}}$ to move the equilibrium from no rationing (average $\Omega_s = 100\%$) to high rationing (average $\Omega_s = 90\%$). The results are shown by the black lines in panel (a) of Figure E6.

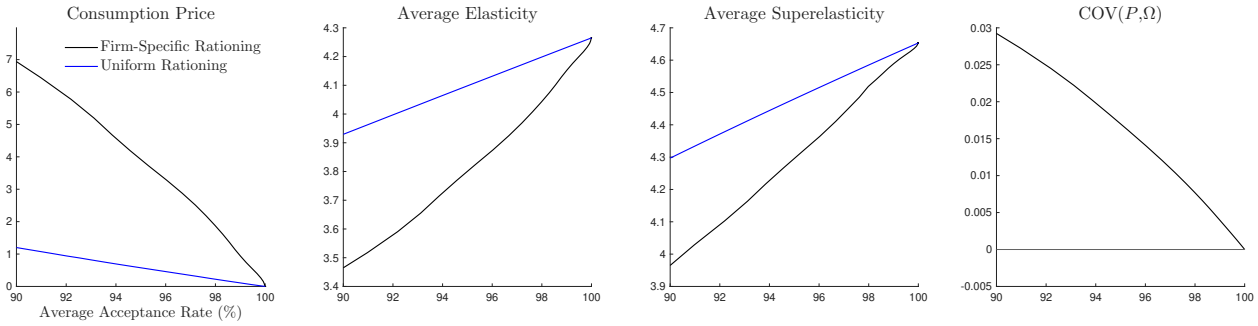
Uniform rationing. For comparison, I also report results (blue lines) obtained under a uniform rationing rule. Here, all firms choose the same Ω_s. I then compare the two cases while holding the average acceptance rate constant.

Importantly, the uniform rationing rule is closely related to the aggregate matching function commonly used in models with product market frictions, such as Michaillat and Saez (2015) and Kaplan and Menzio (2016).[24] Models may differ in details, yet reduced-form matching functions effectively impose uniform rationing as a function of aggregate market tightness. By construction, these functional assumptions cannot generate the price-rationing covariance implied by (E.1) in a micro-founded rationing model.

Prices and quantities. Under firm-specific rationing, the consumption price index rise sharply with rationing, reflecting higher purchase prices paid by households. As the average acceptance rate Ω falls from 100% to 90%, the consumption price rises by close to 6%, even though the posted price distribution remains completely fixed.

The effect is even more pronounced compared to the uniform rationing case. Under uniform rationing, the output price rises by only 1% even when $\Omega = 90\%$. The intuition is straightforward: with uniform rationing, buyers can still search and find low-price sellers (the second term in (2.3) is zero), albeit with more search efforts. In contrast, when low P_s also implies low

[24] While a growing literature studies product market frictions, to my knowledge, I am not aware of any search model that explicitly analyzes the Phillips curve or marginal cost pass-through at a level comparable to standard New Keynesian models.



Notes: For the posted price distribution, I use the steady-state distribution of the calibrated model in Section 4.3. The last panel shows the unconditional covariance between prices P and acceptance probabilities Ω.

Figure E6: Effects of Rationing on the Demand System

Ω_s, keeping searching is likely futile, so buyers end up accepting much higher input prices and passing costs fully into output prices.

Note that the extent of average rationing is only weakly related to quantities. In this exercise, nominal spending is fixed, so changes in quantities simply mirror changes in prices. For the same level of average rationing, buyers' willingness to pay can differ dramatically depending on the $P-\Omega$ joint distribution, resulting in large differences in quantities.

Elasticity and superelasticity. In the third and fourth columns of panel (a) of Figure E6, I evaluate demand elasticity and superelasticity weighted by sales. Under firm-specific rationing, demand elasticity falls from 3.4 to around 3.05 as average Ω_s declines to 90%. By comparison, under uniform rationing, elasticity declines more modestly to 3.23. That is, firm-specific rationing nearly doubles the decrease in elasticity. More importantly, firm-specific rationing also generates a larger decline in superelasticity. As average Ω_s falls to 90%, superelasticity drops from 7.6 to 6.6, compared to 7.05 under uniform rationing. Because real rigidities govern the slope of the Phillips curve, this reduction interacts with other macroeconomic shocks and amplifies inflation in addition to the one-time increase in desired markups.

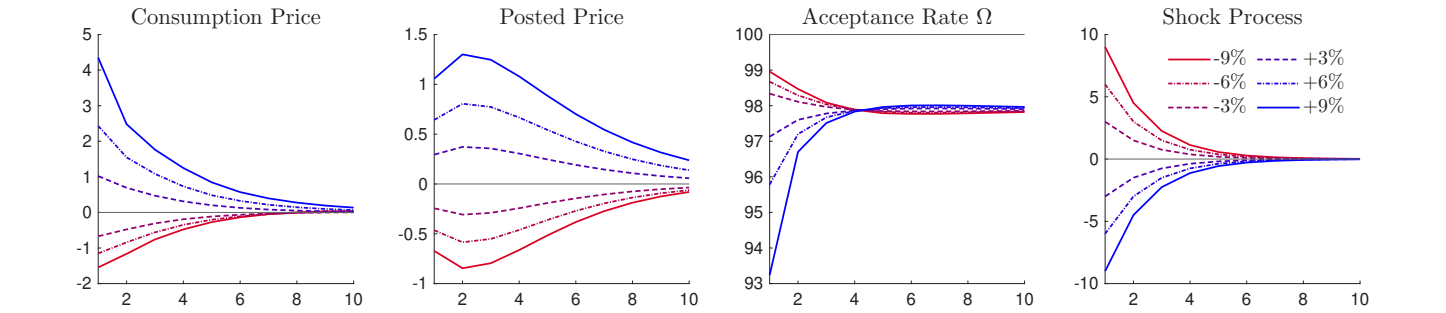
E.2 Phillips Curve in Different Models

E.2.1 Productivity Shocks

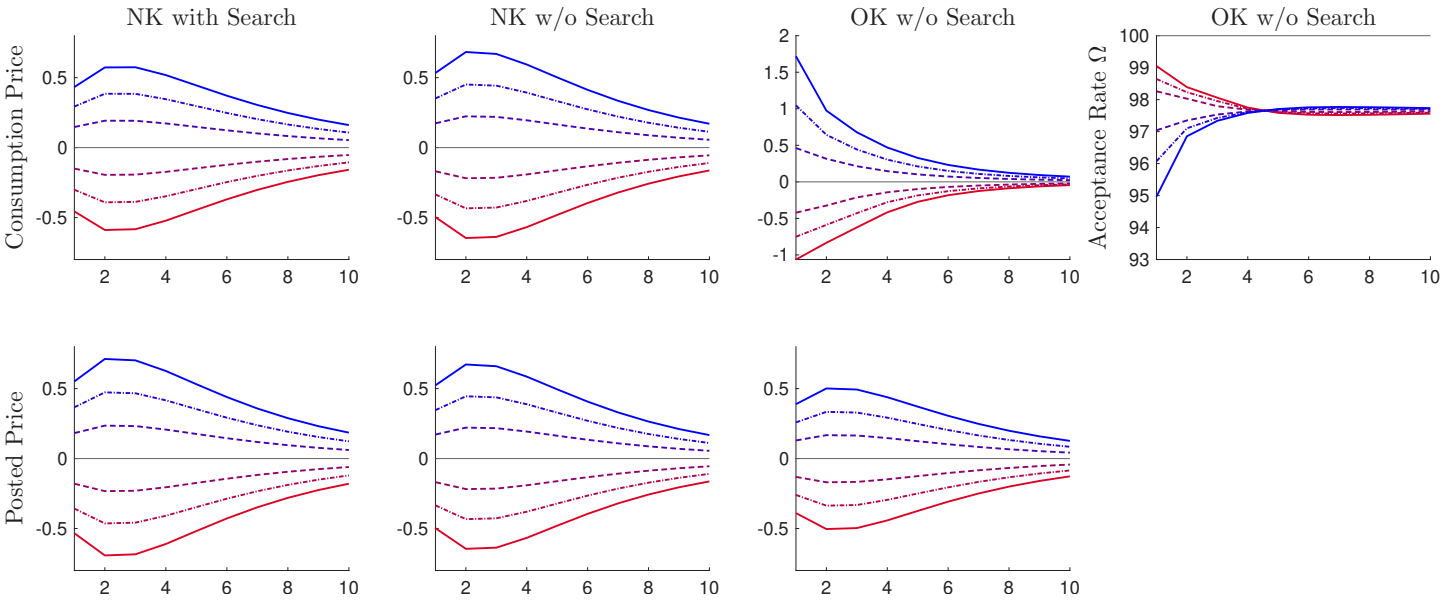
Next, I evaluate the Phillips curve numerically in partial equilibrium using (3.12) and the full calibration from Section 4.1. Following the sequence-space approach, any outcome of the firm block, denoted by $\mathbf{x}$, can be expressed as a function of real demand $\mathbf{c}$, the real wage $\mathbf{w}^r$, the aggregate price level $\mathbf{p}$, aggregate productivity shock $\mathbf{a}$, and aggregate demand shock ξ:

$$\mathbf{x} = F(\mathbf{c}, \mathbf{w}^r, \mathbf{p}, \mathbf{a}, \xi). \tag{E.2}$$

For this exercise, I fix $\{\mathbf{c}, \mathbf{w}^r, \mathbf{p}, \xi\}$ and impose different productivity shocks at $t = 1$. Each productivity shock follows an AR(1) process with an autoregressive coefficient of 0.5, and its initial magnitude ranges from -9% to 9%. The resulting shock paths are shown in the fourth column of



(a) Old Keynesian Model with Search Frictions



(b) Other Models

Notes: Price responses are in levels. The average posted price is the posted price integrated over the distribution of goods producers.

Figure E7: Productivity Shocks in Partial Equilibrium

panel (a) of Figure E7. The outcome variables of interest are the consumption price, the average posted price, and the average acceptance rate.[25]

Comparison between models. I label my model "OK with Search" in this section. To benchmark my model against other models in the literature, I construct three alternative models with comparable parameters. The first, labeled "NK with Search," eliminates rationing while preserving the search structure. The second, labeled "NK without Search," further shuts down search frictions and corresponds to the standard NK-Calvo model. Note that the second model has no real rigidities, so I increase its price stickiness θ_c from 0.7 to 0.79 to ensure comparability. The third model, labeled "OK without Search," extends the second by allowing for rationing

[25] Given the partial equilibrium nature, the consumption price shown in Figure E7 cannot coincide with the assumed **p**. Instead, I feed the equilibrium $P-\Omega$ distribution into the household's problem and calculate the expected unit cost of consumption.

but not search frictions. This model captures the core mechanism in monopolistic OK models: because households cannot search, they must reduce the number of varieties they consume when firms stop selling. The resulting welfare loss from reduced variety generates a wedge between the love-for-variety-adjusted consumption price and the average posted price. I show the responses in my Old Keynesian model in panel (a) of Figure E7 and those in the other three models in panel (b).

OK vs. NK. For the consumption price, my "OK with Search" model exhibits significant asymmetry and substantially greater price flexibility than the two NK models, despite sharing the same parameter values. Following a -9% productivity shock, the consumption price index jumps by over 4% in the OK model, compared with only 0.4-0.6% in the two NK models. Two forces drive this large increase in the consumption price index. First, as discussed above, the consumption price rises mechanically when households are willing to accept higher purchase prices, even if sellers keep posted prices completely fixed. Second, posted prices rise when negative productivity shocks soften price competition and induces greater cost pass-through.

To focus on the pricing behavior, I also plots changes in the average posted price. As expected, posted prices display asymmetry and greater pass-through following large negative shocks in my OK model relative to the two NK models. Nonetheless, posted prices in the OK model exhibit smaller initial responses but greater persistence than the consumption price, especially after negative shocks. Following a -9% shock, average posted prices initially rise by about 1%, accounting for only one-quarter of the increase in the consumption price. As the shock propagates, the consumption price falls quickly, whereas average posted prices remain elevated. Beginning at $t = 3$, changes in the average posted price explain the lion's share of the variation in the consumption price.

OK vs. OK. Panel (b) of Figure E7 also shows the responses in the "OK without search" model. Despite the absence of endogenous competition and the difference in magnitude, its LFV-adjusted consumption price shows remarkably similar asymmetric amplification in the first few quarters. However, several important differences distinguish the two OK models. First, posted price dynamics in "OK without Search" are nearly identical to those in the two NK models, exhibiting little amplification or asymmetry. In other words, rationing without search frictions primarily operates through the *unobserved* marginal utility of consumption on the demand side, rather than interacting with the *observed* posted prices on the supply side.

Second, the intrinsic persistence of price changes differs dramatically. In my "OK with Search" model, positive price responses are much more persistent than negative price responses, resulting in a long-lasting asymmetry, particularly in posted prices. In contrast, in the "OK without Search" model, positive and negative responses of the LFV-adjusted consumption price eventually become symmetric after five quarters.

E.2.2 Inflation Shocks

Next, I impose an inflation shock that permanently raises the price level without affecting real variables. In the notation of (E.2), the shock operates solely through $\mathbf{p}$, while $\{\mathbf{c}, \mathbf{w}^r, \mathbf{a}, \xi\}$ remain fixed. Compared with productivity shocks, inflation shocks imply the same cumulative price adjustment across models, allowing me to focus solely on differences in timing. Moreover, pro-

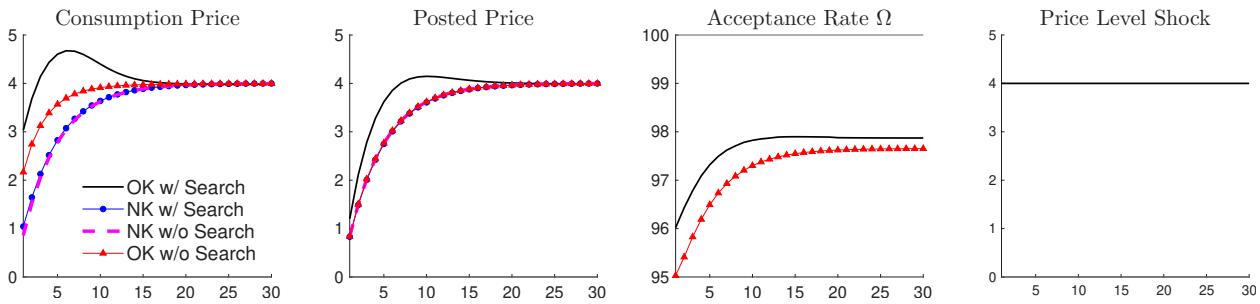
ductivity shocks are mean-reverting, whereas changes in the price level are permanent. Both are important to general equilibrium dynamics.

Contemporaneous shocks. Figure E8 shows the responses to contemporaneous inflation shocks. In panel (a), I introduce a 4% inflation shock. As sticky prices prevent firms from adjusting posted prices, the average acceptance rate Ω falls from nearly 98% to 96% in the "OK with Search" model. The resulting increase in rationing has a profound impact on both the consumption price and the average posted price (black line in the first two panels). As rationing raises households' reservation prices, they switch to more expensive sellers, leading to a rapid increase in the consumption price despite sticky posted prices. Quantitatively, rationing generates a nearly 2% consumption price gap between the "OK with Search" model and the two NK models. As the shock propagates, the peak response of the consumption price exceeds 4.5%, which is even higher than the 4% initial shock. The "OK without search" model lies between the "OK with Search" model and the two NK models in terms of the consumption price.

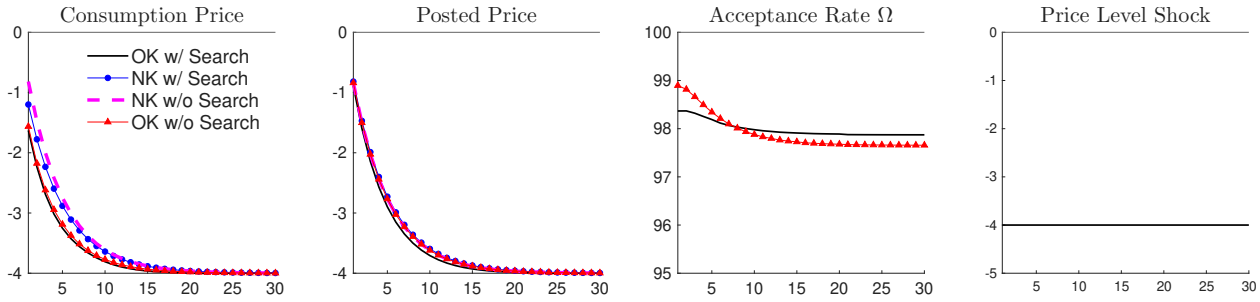
Second, posted prices rise more rapidly in the "OK with Search" model, albeit more slowly than the consumption price, as rationing softens competition and relaxes real rigidities. The peak response reaches around 4.2%, also exceeding the 4% initial shock. Relative to the NK models, rationing generates a posted price gap of 0.5–1% that persists for more than 10 quarters. By contrast, posted prices in the "OK without Search" model are virtually identical to those in the two NK models.

By sharp contrast, the three models differ only marginally following a negative inflation shock of -4% in panel (a). As the price level falls, firms initially experience higher relative prices and engage in less rationing. The resulting increase in availability lowers reservation prices and, consequently, the consumption price. However, the quantitative differences are very modest. For posted prices, the differences are almost negligible. The asymmetry after inflation shocks is even more profound than the asymmetry after productivity shocks in Figure E7.

News shocks. Finally, I examine the timing of responses to anticipated inflation shocks across models. In Figure A2, I introduce anticipated positive and negative inflation shocks that materialize at $t = 8$.



(a) Positive Inflation Shock



(b) Negative Inflation Shock

Notes: Price responses are in levels. The consumption price index is aggregated following (**??**). The average posted price is the posted price integrated over the distribution of goods producers.

Figure E8: Inflation Shocks in Partial Equilibrium

F Data

F.1 Data Cleaning

Imputation. In practice, firms occasionally report sales and quantities for a period other than 12 months, likely due to changes in the starting/ending month of their fiscal years. I adjust these observations to a 12-month basis only when no more than three months of data must be imputed. For example, suppose a firm reports sales from January 2000 to September 2000 and then from October 2000 to December 2001. I first impute sales for October-December 2000 proportionally using the latter reporting period. I then combine the observed sales from January-September 2000 with the imputed sales from October-December 2000 to recover annual sales for 2000. By contrast, if a firm reports sales from January-June 2000 and then from July 2000 to December 2001, I drop this observation because doing otherwise would require imputing six months of data.

My specifications require lags of up to two years, which can be potentially be demanding. To improve sample coverage, I impute missing lagged observations only when both the $t-1$ and $t+1$ values are available. Importantly, imputed observations are used solely to construct lagged controls and are never used as contemporaneous observations in period t. Consequently, all dependent variables and key regressors are based exclusively on reported data.

For $\Delta\text{Capacity}_{i,t}$, I set missing values to zero. This has negligible effects on regression results.

Decimal points. Occasionally, some decimal points also appear misplaced. I adjust them using the method in Liu (2025).

Capacity. Occasionally, the unit of production does not match the unit of capacity. I manually standardize the units and drop any unmatched observation.

COGS. Since COGS may also include the cost of services and others, I allocate the cost proportionally to sales, i.e., $\text{TVC}_{i,t} = \text{COGS}_{i,t} \times \text{Sales of Goods}_{i,t} / \text{Total Sales}_{i,t}$. To reduce measurement errors, I only include shocks when the average expenditure share is between 25% and 110%.[26] The median expenditure share in the regression sample is between 75% and 80% across specifications.

Outliers. I drop all observations with $|\Delta p_{i,t}| > \ln 3$ or $|\Delta y_{i,t}| > \ln 3$ following Amiti et al. (2019).

F.2 Sample Restrictions

I only use manufacturing output products in the following 2-digit sectors: [27, 57] and [63, 70].

In the capacity utilization regressions, I restrict the sample to observations for which both product-level and industry-level capacity utilization fall within a normal range.

- Product-level capacity utilization in period t must lie between the 5th and 95th percentiles of the full-sample distribution conditional on not having large changes in capacity ($|\Delta\text{Capacity}_{i,t}|$<1%). This step excludes products with extreme utilization ratios.

[26] As explained in Liu (2025), I allow the upper bound to be above 100% mainly because material expenditure data appear to have higher quality than COGS data and are unaffected by accounting methods such as FIFO.

- The 6-digit industry must contain at least 10 products with non-missing utilization rates. Industry-level utilization is computed using product-level utilization rates winsorized at the 5th and 95th percentiles. However, I drop industry-year observations for which more than 50% of products are winsorized.

I use the same procedure for inventory regressions.

Other sample restrictions include:

- Drop observations with extreme contemporaneous changes in capacity ($|\Delta\text{Capacity}_{i,t}| \geq 50\%$), as such changes deteriorate the quality of the utilization measure. Further tightening this restriction has little effect on the results.
- Drop industries with fewer than 50 observations.
- Drop observations if the input price shock $\tau_{i,t}$ covers less than 25% of input expenses.
- Drop observations if the industry shock $\delta_{s,t}$ covers fewer than 10 products.

Finally, $\Delta p_{i,t}$, $\Delta y_{i,t}$, $\delta_{s,t}$ and $\tau_{i,t}$ are winsorized at the 1st and 99th percentiles.