Price of Scarcity:

Search Theory of Inventories and Markups*

Lukasz A. Drozd

Federal Reserve Bank of Philadelphia

lukaszadrozd@gmail.com

Marina M. Tavares

International Monetary Fund

marinamendestavares@gmail.com

November 2026

Abstract

Markups are difficult to observe, but the inventory-to-sales ratio is not, and it rises in recessions and falls in expansions. Standard theory reads the ratio as evidence that markups are countercyclical: inventories help firms sell, so a firm with a high markup holds more of them. After the pandemic, gross margins rose sharply while the ratio fell, and across firms margins rose most where the ratio fell most. We propose a search theory of markups that reconciles these observations. In a steady state, margins cover the user cost of selling capacity, and sellers with more market power hold more stock relative to sales. Over the cycle, the relation may reverse: because replacement takes time, a sale uses up the seller's ability to serve the buyers who arrive before the replacement, so when stock is scarce, existing capacity earns rents that raise the price relative to production cost. Uneven demand shocks across sectors can steepen the Phillips curve because markups are convex in demand. Our numerical examples relate gross margins after COVID to the model's predictions.

Keywords: Phillips curve; markups; search; inventories; capacity utilization; supply chains.

JEL: D83, E22, E31, E32.

*Previous versions of the manuscript circulated under the titles "Monetary Policy Shocks: A Case Against Countercyclical Markups" and "Rethinking Inventories and Markups over the Business Cycle." We thank Ben Lester, Urban Jermann, Joachim Hubmer, João Gomes, Rasmus Lentz, Christian Moser, Roc Armenter, Diego Känzig, Mathias Trabandt, Kim Ruhl, Jonas Arias, Simon Mongey, Isaac Baley, Lukas Nord, Nikolai Roussanov, Oleksiy Kryvtsov, Krzysztof Makarski, Piotr Żoch, Jan Werner, and François Gourio; an anonymous referee at the *Journal of Political Economy*; and participants at the 2025 Barcelona School of Economics Summer Forum workshop on Macro Fluctuations with Micro Frictions, the 2024 Warsaw Summer Workshop on Macroeconomics and Finance, the 2024 meeting of the Society for Economic Dynamics, the Spring 2024 Midwest Macroeconomics Meetings, the Wharton School brown bag, and the Federal Reserve Bank of Philadelphia lunch seminar for comments. The views expressed herein are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, the International Monetary Fund, its Executive Board, or IMF management.

1 Introduction

Markups are difficult to observe in the data, but the inventory-to-sales ratio is not, and it has a strong cyclical pattern: it rises in recessions and falls in expansions. Standard theory implies that the ratio reveals markups, and hence that markups are countercyclical (Bils and Kahn, 2000; Kryvtsov and Midrigan, 2013).[1] Inventories help firms sell. A firm that earns a high margin on each sale wants more of them, and because holding them is costly, the ratio rises with the markup. The logic applies to a wide class of macroeconomic models.

The post-pandemic inflation surge data run counter to this logic. Gross margins rose sharply in 2021 while the inventory-to-sales ratio fell, and they fell back as inventories were rebuilt. Across firms, margins rose most where the ratio fell most (Figure 1). In recessions, margins fall as inventories pile up relative to sales. How can these observations be reconciled?

We propose a theory in which both forces operate. Across steady states that differ in market power, sellers that earn a larger rent per sale due to higher market power build more selling capacity and hold more stock relative to sales—as the standard logic predicts. This also applies to lower-frequency and persistent shocks. In a steady state, a sale costs the production cost plus the user cost of the selling capacity that holds the unit (its operating, financing, and depreciation costs over the restocking time), neither of which depends on market power; the margin is fixed by the elasticity of demand and these costs, and the firm chooses how much stock to hold given that margin. Causality in this case runs from market-power-derived markups to inventories. Over the cycle, however, irreversibility keeps production and stock-holding capacity largely fixed, and transient demand surges have the opposite effect. When replacement takes time, a sale uses up the seller's ability to serve buyers until the replacement with fixed capacity arrives, and each of those buyers would have paid a rent over the cost of the good. The more buyers arrive per unit of stock, the larger the opportunity cost and the higher the markup. Low persistence of the shock does not justify a semi-permanent capacity buildout that outlasts the shock. Causality then runs the other way: scarce stock driven by surging demand raises the measured markup

[1] Kryvtsov and Midrigan (2013) develop this argument by evaluating several prominent models of inventory dynamics in the New Keynesian framework in response to monetary policy shocks. The focus on monetary policy shocks is essential: Khan and Thomas (2007) show that an equilibrium model with (S, s) inventory policies, perfectly competitive markets, and productivity shocks yields a countercyclical inventory-to-sales ratio.

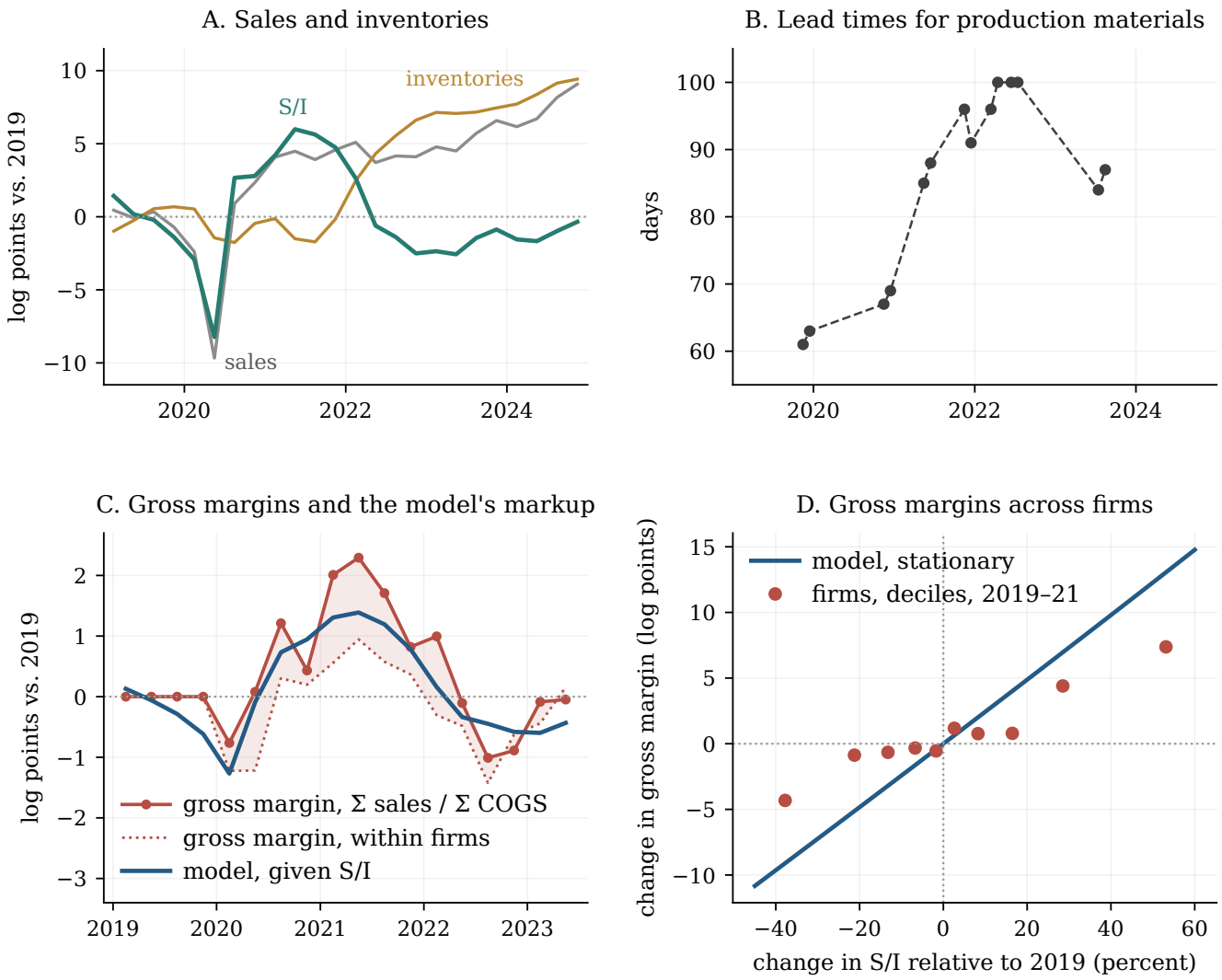

Figure 1: Sales, inventories, and markups after COVID: data and model

Notes: Panel A: real sales and inventories of manufacturing and trade industries (Bureau of Economic Analysis) and the sales-to-inventory ratio S/I, quarterly averages of 100 log relative to the 2019 average. Panel B: average lead time for production materials reported by purchasing managers (Institute for Supply Management, 2019–2023), in the months for which we have the report. Panel C: gross margins of U.S. manufacturing and trade firms in Compustat, with the exclusions of Section 2.6, relative to the same quarter of 2019, in two measures, with the range between them shaded: 100 log of total sales over total cost of goods sold (COGS) of the firms present in every quarter from 2019Q1 to 2023Q2 (solid), and the within-firm change in 100 log(sales/COGS), mean weighted by 2019 COGS (dotted). Model: the markup that the calibrated model predicts given the monthly path of the sales-to-inventory ratio, whose quarterly averages panel A shows, with 2019 as the steady state, as total revenue over the production cost of the goods sold in each quarter (Section 4.2). Panel D: change in the gross margin between 2019 and 2021 against the percent change in the sales-to-inventory ratio, for firms that report all four quarters of both years, with annual sales, cost of goods sold, and average inventories. Dots: deciles of the change in the ratio, COGS-weighted mean of the change in the margin within each decile. Line: stationary markup of the calibrated model with the stock of capacity fixed (Figure 2A).

relative to production cost. Because in our model demand moves sales faster than stock can follow, this effect dominates over the business cycle.

We build a model around this logic and use it to quantify the post-pandemic surge. Its key prediction is that the markup is close to linear in the sales-to-inventory ratio, the inverse of the inventory-to-sales ratio, with a slope equal to the rent per sale times the time it takes to restock. We feed the observed monthly path of the sales-to-inventory ratio into the model's pricing equation, with gross margins entering only through their 2019 level, a calibration target. The predicted markup tracks gross margins in Compustat, measured as total sales over total cost of goods sold, with a correlation of 0.91 from 2020 through mid-2023, and its mid-2021 rise, 1.39 log points, lies between the rise of this ratio, 2.29, and the mean rise of firms' own margins, 0.95 (Figure 1C). Across firms, margins rose close to linearly with the sales-to-inventory ratio, by 0.10 log points per percent increase in a firm's ratio between 2019 and 2021 and by 0.19 across three-digit manufacturing industries, against 0.24 in the model (Figure 1D). Sales enter both the ratio and the measured margin, but inventories have explanatory power of their own: among firms with the same sales growth, those that accumulated more inventory raised margins less. Across manufacturing industries, producer prices rose with sales per unit of inventory after conditioning on plant utilization (Section 4.1).

In each of the three recessions before the pandemic, margins fell by 0.3 to 4.1 log points at the trough of the sales-to-inventory ratio (1.6 to 3.1 within firms), and the model predicts falls of 0.5 to 2.3 (Section 5.2). After a monetary tightening, by contrast, the sales-to-inventory ratio moves little: its estimated fall implies a markup decline of at most 0.20 log points, and firms' gross margins show no robust response.

In the model, producers sell through units of selling capacity that we call *outposts*. An outpost is installed capacity to fill orders at a certain rate: a space on a dealer's showroom floor, a shelf in a department store, a bay in a warehouse, a production line, a table at a restaurant, or a chair at a barber shop. Such capacity is costly to install, hard to remove, and costs the same whether it is used or idle. Once the car on display is sold, the space remains, but the dealer cannot use it to make another sale until a replacement arrives; a restaurant table serves its next party only after the last one leaves. An outpost that holds stock is ready to serve its next buyer, and after a sale it restocks at a pace the seller cannot hurry. Goods on display attract buyers, so stock helps sell, as in the standard theory. Selling capacity extends well beyond the goods that firms count as inventories, but inventories record the part of it that holds goods ready for sale, so the ratio of sales to inventories serves as a litmus test of the theory.

Distributors, which sell to final buyers at the retail price P, send shoppers to draw price quotes from the outposts that hold stock. Shoppers value each match differently, so an outpost that raises its price loses some buyers but retains those with particularly good matches, and it earns a search rent. Let v be the unit cost of production and $1/\tau$ the mean time to restock. With exponentially distributed match values of mean η_0, the search rent per sale is $\eta_0 P$. Buyers arrive at each stocked outpost at the rate Λ, which equals the ratio of sales to inventories because, as a normalization, each outpost has unit stock-holding capacity. With a given stock of outposts, and when discounting, depreciation, and inventory carrying costs are small relative to the restocking rate, the price an outpost quotes, the wholesale price, satisfies

$$\tilde{p} \;\approx\; v + \eta_0 P\left(1 + \frac{\Lambda}{\tau}\right) \;\approx\; v + \frac{\eta_0 P}{1-u},$$

where utilization u is sales as a share of the largest flow that the stock of outposts can sustain.

The first expression shows why the relation is linear. Its last term is the opportunity cost of a sale: buyers arrive at the rate Λ, each worth the search rent $\eta_0 P$, over the restocking time $1/\tau$. At a given retail price, the markup is therefore linear in the ratio, with a slope that is the same whether inventories are ample or scarce. Allowing the retail price to adjust leaves this relation close to linear in the calibrated model. The second expression shows why the supply curve, the markup as a function of demand at a given stock of capacity, is convex. The share of outposts ready to sell is approximately $1 - u$. As sales approach capacity, each additional sale leaves fewer outposts ready to serve the buyers who follow, so the ratio, and with it the markup, rises more and more steeply.

The convexity bears on the shape of the Phillips curve. For two decades before the pandemic, prices responded little to slack, and the Phillips curve looked flat and stable (Del Negro et al., 2020; Hazell et al., 2022). In 2021 prices rose sharply, and firms report raising prices more after increases in demand than they cut them after decreases (Bunn et al., 2026). In our theory, the slope of the Phillips curve depends on inventories: the curve is flat and stable while inventories are ample relative to the time it takes to restock and steep when they are thin. The sales-to-inventory ratio measures the tightness of product markets, as the ratio of vacancies to job seekers measures the tightness of the labor market.

Selling capacity is costly to build and cheap to keep. After a fall in demand, firms stop opening outposts but retain the ones they have, inventories pile up relative to sales, and markups stay below their normal level until depreciation absorbs the excess capacity. In a surge, building outposts quickly is costly, so demand pushes sellers toward the steep part of the supply curve until new capacity arrives. When a surge ends, the capacity built in it sits idle, and markups fall below their normal level before they recover. In our calibration to U.S. manufacturing and trade, inventories cover 1.5 months of sales and replenishment takes 2 months on average. With capacity held fixed, a 20 percent increase in demand raises the markup by 15.6 log points, whereas an equal decrease lowers it by 8.9 log points (Table 2). Inventory coverage and the time to restock pin down utilization and therefore how convex the response is; given them, market power sets only its scale.

The theory gives a quantitative account of margins after COVID. The pandemic destroyed part of firms' organizational capital, the firm-specific knowledge embodied in their workforces (Drozd and Tavares, 2026). If selling capacity is part of that capital, the recovery was a surge relative to the capacity that remained. We feed the calibrated model the observed path of sales and a loss of 3.7 percent of outposts in April 2020, measured by the jobs that manufacturing and trade had not recovered a year later, and the model produces both the inventory-to-sales ratio and the markup (Figure 7). Stock is scarce from August 2020 to March 2022, and the markup peaks in 2021Q1 at 1.37 log points, between the two measures of gross margins, which peak in 2021Q2 at 2.29 and 0.95. From 2020Q3 to 2022Q4, the correlation between the model's markup and total sales over total cost of goods sold is 0.81; from June 2020 to December 2022, the correlation between inventories in the model and in the data is 0.51. The loss of capacity drives the fit: with the same path of sales and no loss of outposts, the correlation of inventories falls to -0.19. A destructive recession followed by a sharp recovery raises markups through the shortage of stock it leaves behind.

Because the supply curve is convex, where demand goes matters as well as how much of it there is. In the calibrated model, a 10 percent increase in total demand raises the average markup by 6.6 log points when spread evenly across markets and by 12.8 log points when concentrated in a quarter of them (Figure 6). Total sales and inventories are identical in the two experiments. In the second, goods held in markets where demand does not rise do nothing to relieve the scarcity where it does, so the aggregate ratio

understates the pressure on prices. In companion work (Drozd and Tavares, 2026), we document that industry employment gains were more dispersed after COVID than after the Great Recession and show how uneven workforce losses raised the cost of rebuilding. Here uneven spending affects prices through the stocks sellers have available when demand returns.

Related literature. Our model builds on search theories in which buyers draw price quotes and differ in their valuations of a match (Wolinsky, 1986; Anderson and Renault, 1999; Menzio, 2024), and on models in which inventories help firms sell or produce (Kahn, 1987; Bils and Kahn, 2000; Khan and Thomas, 2007; Kryvtsov and Midrigan, 2013). Search frictions in product markets are large: differences in sales across firms come mainly from the number of customers, and firms' investment in winning customers rivals their investment in physical capital (Afrouzi et al., 2023; Hubmer and Nord, 2026). These papers study how product-market frictions shape the level and dispersion of markups and the allocation of customers across firms; we study how they shape the cyclical behavior of markups when selling requires stock. In our theory, markups rise with sales per unit of selling capacity, as across firms they rise with sales per customer but not with the number of customers (Afrouzi et al., 2023). Limited capacity plays a role in search models of pricing and matching (Burdett et al., 2001), Michaillat and Saez (2015) study tightness in the product market, and investment in an outpost resembles investment in customer capital (Drozd and Nosal, 2012; Gourio and Rudanko, 2014; Gilchrist et al., 2017). Our departure is to tie a seller's ability to serve its next buyer to the stock available for sale.

The closest work on convex supply curves relates the price response to how intensively firms use productive capacity (Boehm and Pandalai-Nayar, 2022; Bohr, 2024). In our theory the relevant capacity is the ability to sell, and inventories and replenishment times jointly measure its utilization. This mechanism complements explanations for a nonlinear Phillips curve based on menu costs and trend inflation (Bunn et al., 2026), strategic complementarities in pricing (Harding et al., 2023), and labor-market tightness (Benigno and Eggertsson, 2023). The form of the constraint matters. Diminishing returns under Calvo pricing can flatten the Phillips curve (Holm et al., 2024). Here flexible prices respond to the opportunity cost of a sale, a component of effective marginal cost whose elasticity with respect to output rises with

utilization. This distinction bears on the finding of Gagliardone et al. (2025) that before the pandemic prices responded strongly to marginal costs even though marginal costs responded little to output.

Several accounts of post-pandemic inflation emphasize binding capacity constraints (Comin et al., 2023), shipping delays and stockouts (Alessandria et al., 2023; Cavallo and Kryvtsov, 2023), or congestion along supply chains (Bai et al., 2024). Changes in the composition of demand can also raise inflation when wages are rigid downward (Guerrieri et al., 2021) or labor is costly to move across sectors (Ferrante et al., 2023). Our mechanism gives sectoral shifts a further consequence through the markups of the sellers who receive or lose the business. It links the rise and reversal of margins in the goods sector to the scarcity of inventories, adding to evidence on markups, profits, and wages during the recovery (Glover et al., 2023b,a; Afrouzi et al., 2024).

Section 2 presents the model, the measurement of markups, and the calibration. Section 3 derives the supply curve and studies how capacity adjusts after a change in demand. Section 4 applies the theory to COVID-era inflation, including the loss of capacity in the pandemic, and Section 5 compares the model with monetary tightenings in normal times and with recessions. Section 6 concludes.

2 Model

The model describes a wholesale market for intermediate goods. Time is continuous, $t \in [0, \infty)$. The economy starts in a steady state and is hit at $t = 0$ by an unexpected, perfectly foreseen change in the paths of aggregate demand and of the discount rate; in two experiments, the path is revised again at later dates, each time unexpectedly, and values can jump when news arrives. Producing a unit of the good uses one unit of a variable input, which stands for materials and production labor and is supplied at a fixed price v; v serves as the numéraire, and we keep the symbol to make units explicit. A producer's sales are therefore its gross output, and v is its cost of goods sold per unit. Holding v fixed is a short-run assumption: in the model, all movements of prices relative to v are movements of markups.

2.1 Environment

There are three types of agents. *Producers* make intermediate goods and sell them through atomless selling units, called *outposts*. *Distributors* send *shoppers* to buy intermediate goods from outposts, turn them into final goods, and sell the final goods in a competitive retail market at price P_t. Retail demand is $D_t(P) = D_{0t}P^{-\varepsilon}$, with elasticity ε and an exogenous level D_{0t}. Agents of each type are identical, act as price takers with respect to aggregate variables, and play symmetric pure strategies.

2.2 Distributors and search

A shopper who buys a unit of an intermediate good at the quoted price $\tilde{p}$, the wholesale price, brings the distributor $1+\eta$ units of the final good, where η is a match-specific taste shock, exponentially distributed with mean η_0, distribution G, and density g. The distributor's surplus from the match is $P_t(1+\eta) - \tilde{p} = P_t - p$, where

$$p := \tilde{p} - \eta P_t \tag{1}$$

is the *effective price*. Because the effective price summarizes the value of a match, the distributor instructs its shoppers to accept any quote whose effective price does not exceed a reservation value $\bar{p}_t$. Lemma 3 in Appendix A.1 shows that this reservation policy is optimal. When all outposts quote the same price $\tilde{p}_t^*$, the policy can be summarized by the share of quotes a shopper rejects, $\pi_t := \Pr(p > \bar{p}_t) = G\big((\tilde{p}_t^* - \bar{p}_t)/P_t\big)$, which we call *search precision*.

Search technology. A shopper draws quotes at random from all outposts that hold stock. Each draw costs $c_0 v$ and takes a time of the order of $(dt)^2$. A shopper can therefore draw as many quotes as it needs within an instant of length dt, and, as long as some quotes are acceptable, the probability that it returns without a good vanishes as $dt \to 0$. The number of draws until acceptance is geometric with success probability $1 - \pi_t$. Rejecting more quotes raises both the expected search cost and the expected surplus of the good a shopper brings back. The formal result is below.

Lemma 1. *In an equilibrium with a single quoted price $\tilde{p}_t^*$ and precision $\pi_t \in [0,1)$, the expected search cost per good is $c(\pi_t)v = c_0 v/(1-\pi_t)$, the expected surplus per good is*

$$s_t(\pi_t) = P_t(1+\eta_0) - \tilde{p}_t^* - \eta_0 P_t \log(1-\pi_t), \tag{2}$$

and the reservation effective price is $\bar{p}_t = \tilde{p}_t^* + \eta_0 P_t \log(1-\pi_t)$.

The distributor chooses precision to maximize the expected net surplus per good, $s_t(\pi) - c(\pi)v$, whose maximum is at $\pi_t = \max\{0, 1 - c_0 v/(\eta_0 P_t)\}$. Distributors are *selective* when $\eta_0 P_t > c_0 v$: shoppers then reject some quotes, and optimal precision is

$$\pi_t = 1 - \frac{c_0 v}{\eta_0 P_t}. \tag{3}$$

Distribution is competitive. Operating a shopper costs χv per good, and free entry drives net surplus to that cost:

$$P_t - \tilde{p}_t^* - \eta_0 P_t \log\left(\frac{c_0 v}{\eta_0 P_t}\right) = \chi v. \tag{4}$$

The retail market clears, $Q_t\big(1 + \mathbb{E}_{\pi_t}[\eta]\big) = D_{0t} P_t^{-\varepsilon}$, where Q_t is the flow of intermediate goods sold and $\mathbb{E}_\pi[\eta] = \eta_0(1 - \log(1-\pi))$ is the mean taste of accepted matches. We specify the path of sales Q_t, which we call aggregate demand, and let D_{0t} adjust to deliver it.

2.3 Outposts and their demand

An outpost can sell only if it holds stock. Let M_t be the mass of outposts that hold a unit of stock. Every shopper buys, and shoppers sample stocked outposts uniformly, so each stocked outpost sells at the rate

$$\Lambda_t = \frac{Q_t}{M_t}, \tag{5}$$

and shoppers draw its quote at the rate $\Lambda_t/(1-\pi_t)$. We call Λ_t demand per outpost. Because each stocked outpost holds one unit, it is also the sales-to-inventory ratio, the inverse of the inventory-to-sales ratio M_t/Q_t.

An outpost that quotes $\tilde{p} \neq \tilde{p}_t^*$ while all others quote $\tilde{p}_t^*$ sells to a shopper who draws its quote whenever the taste shock satisfies $\tilde{p} - \eta P_t \leq \bar{p}_t$. With exponential tastes, the deviating outpost sells at the rate

$$\lambda_t(\tilde{p}) = \Lambda_t \min\left\{ \exp\left(-\frac{\tilde{p} - \tilde{p}_t^*}{\eta_0 P_t} \right), \frac{1}{1-\pi_t} \right\}, \tag{6}$$

where the second term applies at prices below the reservation effective price $\bar{p}_t$, at which every shopper who draws the quote buys. Above $\bar{p}_t$, each outpost faces a demand with elasticity $\tilde{p}/(\eta_0 P_t)$. Market power comes from the dispersion of tastes, and its strength is governed by η_0.

2.4 Producers

An outpost is in one of two states. In the *marketing* state it holds one unit of stock and waits for a buyer. In the *production* state it holds no stock, cannot sell, and restocks at Poisson rate $\hat{\tau}_t \leq \tau$, paying v per unit. Let N_t be the mass of outposts in the production state, and let $K_t := M_t + N_t$ be the stock of selling capacity. Outposts depreciate at rate δ, and stock held by a depreciating outpost is lost. Keeping an outpost costs $\zeta_0 v$ per unit of time, and holding stock costs an additional ζv.

Let V_{1t} and V_{0t} be the values of an outpost in the marketing and production states, and let

$$X_t := V_{1t} - V_{0t}$$

be the value of the unit of stock an outpost holds. The values of outposts that are kept satisfy

$$\rho_t V_{1t} = -(\zeta_0 + \zeta)v + \max_{\tilde{p}} \left\{ \lambda_t(\tilde{p})\,(\tilde{p} - X_t) \right\} - \delta V_{1t} + \dot{V}_{1t}, \tag{7}$$

$$\rho_t V_{0t} = -\zeta_0 v + \max_{0 \leq \hat{\tau} \leq \tau} \left\{ \hat{\tau}\,(X_t - v) \right\} - \delta V_{0t} + \dot{V}_{0t}. \tag{8}$$

A sale earns the price but turns a stocked outpost into an empty one, at a capital loss X_t; restocking costs v and yields the gain X_t. The outpost restocks at the maximal rate whenever $X_t \geq v$. Given the demand (6), the price that maximizes (7) is $\max\{\bar{p}_t, X_t + \eta_0 P_t\}$, so in a symmetric selective equilibrium it is

$$\tilde{p}_t = X_t + \eta_0 P_t, \tag{9}$$

which lies above $\bar{p}_t$ because shoppers reject some quotes.

Creating and destroying capacity. Producers create new outposts at rate a_t per unit of capacity; new outposts enter in the production state. Each new outpost requires a sunk investment ϕv, and creating outposts faster than they depreciate costs $\kappa_a \max\{a_t - \delta, 0\}^2 K_t$ in addition, where K_t is aggregate capacity, which producers take as given. Producers can liquidate outposts in the production state at rate d_t, at a cost $\kappa_d d_t^2 N_t$ with $\kappa_d \to 0$. Maximizing the value of their outposts, producers choose

$$a_t = \begin{cases} 0 & \text{if } V_{0t} < \phi v, \\ \in [0, \delta] & \text{if } V_{0t} = \phi v, \\ \delta + (V_{0t} - \phi v)/(2\kappa_a) & \text{if } V_{0t} > \phi v, \end{cases} \qquad d_t > 0 \text{ only if } V_{0t} \leq 0. \tag{10}$$

In the limit $\kappa_d \to 0$, an empty outpost can be liquidated at zero value, so $V_{0t} \geq 0$, and outposts are liquidated only when $V_{0t} = 0$. Its value then satisfies

$$\min \left\{ V_{0t},\ (\rho_t + \delta) V_{0t} + \zeta_0 v - \max_{0 \leq \hat{\tau} \leq \tau} \left\{ \hat{\tau} \left(X_t - v \right) \right\} - \dot{V}_{0t} \right\} = 0. \tag{11}$$

While empty outposts are kept, the second term is zero, which is (8); where liquidating them is strictly better, $V_{0t} = 0$ and the second term is positive. Capacity follows

$$\dot{M}_t = \hat{\tau}_t N_t - Q_t - \delta M_t, \qquad \dot{N}_t = Q_t - (\hat{\tau}_t + \delta + d_t) N_t + a_t K_t. \tag{12}$$

In the limits $\kappa_a \to 0$ and $\kappa_d \to 0$, creation and liquidation can occur in atoms, and liquidation can be a flow while $N_t = 0$. Let $\mathcal{C}_t$ and $\mathcal{L}_t$ be the cumulative masses of outposts created and liquidated, nondecreasing and possibly with jumps; the second equation then reads $dN_t = \left[Q_t - (\hat{\tau}_t + \delta) N_t \right] dt + d\mathcal{C}_t - d\mathcal{L}_t$, with $d\mathcal{C}_t = a_t K_t \, dt$ and $d\mathcal{L}_t = d_t N_t \, dt$ when the rates are finite. In these limits, outposts are created only where $V_{0t} = \phi v$, with $V_{0t} \leq \phi v$, and liquidated only where $V_{0t} = 0$. When $0 < V_{0t} < \phi v$, no outposts are created or destroyed. Because creating an outpost is costly and keeping one is cheap, this *inaction region* is nonempty: after a fall in demand, producers keep the outposts they have.

Holding a single unit. As a normalization, an outpost has room for one unit and restocks one unit at a time; producers add room for stock only by creating outposts. The normalization keeps the two features the model rests on: a constraint on the stock that selling capacity can hold, and a fixed cost of adding to that capacity. It also keeps the state of an outpost binary, and selling capacity is the binding friction: each outpost can serve at most one buyer per restocking cycle.[2]

2.5 Equilibrium

Throughout, v, c_0, η_0, τ, δ, κ_a, and ρ_t are positive and ϕ, ζ_0, ζ, and χ nonnegative, with $\zeta_0 + (\rho+\delta)\phi > 0$; the limit $\kappa_a \to 0$ and the case $\phi = 0$ are named where they are used.

Definition 1. Given paths $\{\rho_t, D_{0t}\}$ and initial capacity (M_0, N_0), a selective equilibrium consists of prices $\{\tilde{p}_t^*, P_t\}$, distributors' policies $\{\pi_t, Q_t\}$, capacity $\{M_t, N_t\}$, bounded values $\{V_{0t}, V_{1t}\}$, and producers' policies $\{\hat{\tau}_t, a_t, d_t\}$ such that $\eta_0 P_t > c_0 v$, distributors choose precision optimally (3) and earn zero profits (4), the retail market clears, outposts value and price optimally, (7), (11), with empty outposts kept only where its second term is zero, and (9), with $\tilde{p}_t = \tilde{p}_t^*$ and demand (5)–(6), producers create and destroy capacity optimally (10), and capacity evolves according to (12), in cumulative form in the limits. We call these equilibria.

A steady state has constant demand, discount rate, and capacity.

Assumption A1. $(1-\eta_0)\, c_0/\eta_0 < \tau^{-1}\big(\phi(\rho+\delta) + \zeta_0 + \tau\big) + \chi$.

Lemma 4 in Appendix A.1 shows that a selective steady state exists, and is unique, if and only if Assumption A1 holds. The assumption puts the steady-state retail price above $c_0 v/\eta_0$, where rejecting some quotes raises distributors' net surplus. In the steady state, new outposts just replace depreciating ones ($a = \delta$), so the value of an empty outpost equals the cost of creating it, $V_0^{ss} = \phi v$, and (8) gives the value of stock

$$X^{ss} = v + \frac{\zeta_0 + \phi(\rho+\delta)}{\tau}\, v. \tag{13}$$

[2] Appendix A.1 checks the normalization: a stocked outpost that could rent a slot for a second unit at 7 times its operating cost would leave the slot unused on every path we solve.

The steady-state price exceeds the production cost v by two terms: the user cost of an outpost, spread over the expected restocking time $1/\tau$, and the search rent $\eta_0 P$.

2.6 Measurement and calibration

We measure markups with gross margins. Let firm i produce $Y_i = \mathscr{Y}_i(V_i, F_{1i}, F_{2i}, \ldots)$ with a variable input V_i and inputs $F_{1i}, F_{2i}, \ldots$ that are fixed within the period, and let $c_i(Y_i)$ be the minimal cost of the variable input needed to produce Y_i, where the input's price $v_i(V_i)$ may rise with the quantity bought. Define the markup $\mu_i := \log\left(p_i / c_i'(Y_i)\right)$, the gross margin $\bar{\mu}_i := \log\left(p_i Y_i / c_i(Y_i)\right)$, the output elasticity $\alpha_i := \log\left(\partial_{V_i} \mathscr{Y}_i V_i / Y_i\right)$, and the input-price elasticity $\omega_i := \log\left(1 + v_i'(V_i) V_i / v_i(V_i)\right)$. Cost minimization implies

$$\mu_i = \bar{\mu}_i - \omega_i + \alpha_i. \tag{14}$$

Under a Cobb–Douglas technology with input prices taken as given, ω_i and α_i are constant, and markups and margins move one for one. More generally, they move in the same direction unless the elasticities vary enough over the cycle to reverse the comovement, which would also make marginal cost curves steep (Nekarda and Ramey, 2020). We measure $c_i(Y_i)$ by the cost of goods sold (COGS), which records the direct costs of production, and $p_i Y_i$ by sales. In the model, the log gross margin of a producer is $\log(\tilde{p}/v)$, which equals its markup.

The distinction matters for interpreting the pandemic. In companion work, we find that much of the initial increase in industry prices lay outside measured unit labor and material costs (Drozd and Tavares, 2026). That accounting residual can include unmeasured costs of expanding production. Section 4 asks whether the gross margins recorded by firms moved with an observable cost of selling a unit from stock.

Sales and inventories are the BEA's monthly real sales and real inventories of manufacturing and trade industries; inventories are stocks at the end of each month. We measure gross margins for U.S. manufacturing and trade firms in Compustat, whether or not they report inventories, relative to the same quarter of 2019, which removes seasonality without filtering. We exclude petroleum refining, fuel dealers, and gasoline stations, whose margins move with energy prices, together with the auto dealers that share an industry code with gasoline stations. The main measure is 100 log of total sales over total COGS of the

Table 1: Parameters

Parameter		Value	Target or source
Discount rate (monthly)	ρ^{ss}	7.94×10^{-3}	Cost of capital, 10% a year
Restocking rate	τ	0.50	Delivery lag, 60 days
Mean taste shock	η_0	0.107	Jointly matched (see notes)
Search cost per quote	c_0	0.0736	Jointly matched
Cost of a shopper	χ	0.809	Jointly matched
Sunk cost of an outpost	ϕ	6.33	Jointly matched
Operating cost of an outpost	ζ_0	0.0727	Jointly matched
Depreciation of outposts	δ	1.85×10^{-3}	Warehouses (BEA), 2.22% a year
Cost of building fast	κ_a	20.0	Post-pandemic recovery (text)
Cost of holding stock	ζ	0	Normalization

Notes: The unit cost of production is the numéraire, $v = 1$. The five jointly matched parameters target a distribution margin of 38 percent, a gross margin of 49 percent, an inventory-to-sales ratio of 1.5 months, spending on selling capacity of 22.8 percent of producers' expenses, and three quotes per purchase. Implied steady state: value of stock $X^{ss} = 1.269$, retail price $P^{ss} = 2.056$, inventory-to-sales ratio 1.50 months, mean accepted taste shock $\mathbb{E}_\pi[\eta] = 0.23$, utilization of selling capacity 57 percent.

1,383 firms present in every quarter from 2019Q1 to 2023Q2, which account for 92 percent of the sample's COGS in 2019. Like the BEA sales-to-inventory ratio and the 2019 gross margin that the calibration targets, it is a ratio of sums, so it does not depend on how a given set of sales and costs is divided among firms. The second measure is the change in each firm's log gross margin, averaged across firms with 2019 COGS weights, excluding firm-quarters with negative margins or changes of more than 50 log points.

We calibrate the model to U.S. manufacturing and trade and use it throughout to quantify the results. A model period is a month, and Table 1 lists the parameters. The steady-state discount rate matches a cost of capital of 10 percent a year, and the restocking rate $\tau = 0.5$ matches a delivery lag of 60 days in U.S. manufacturing (Deloitte Research Center for Energy & Industrials, 2024). We hold the variable cost of holding stock at $\zeta = 0$.

Outposts depreciate at the rate of commercial warehouses in the national accounts, 2.22 percent a year (Bureau of Economic Analysis, n.d.), so that $\delta = 1.85 \times 10^{-3}$ a month. Five parameters—the mean taste shock η_0, the search cost c_0, the cost of a shopper χ, the sunk cost of an outpost ϕ, and its operating cost ζ_0—are set jointly to match five targets. Two come from these Compustat manufacturing and trade firms in 2019, the steady state of the COVID exercises: a gross margin $(\tilde{p} - v)/v$ of 49 percent, their sales less the cost of goods sold relative to the cost of goods sold; and producers' spending on selling capacity, $(\phi\delta + \zeta_0)vK$, of 22.8 percent of their expenses, their selling, general, and administrative expenses relative

to these expenses plus the cost of goods sold. The other three are a distribution margin $(P - \tilde{p})/\tilde{p}$ of 38 percent, computed from the 1997 input-output tables, an inventory-to-sales ratio of 1.5 months, and three quotes per purchase. The parameters match the five targets closely; Appendix A.2 reports the model's values. In the steady state, a little more than half of the outposts are restocking at any time: the utilization of selling capacity, defined in Section 3.2, is 57 percent.

The depreciation rate governs how long a downturn lasts (Section 3.3): in the calibrated model, no outposts are created for 7 months after a permanent 1 percent fall in demand and for 30 months after a 5 percent fall. The cost of building fast, κ_a, governs how long a surge lasts (Section 3.4). Outside Sections 3.3, 3.4, and 4.3 and Appendix A.2.3, the quantitative exercises either hold the stock of capacity fixed or take the sales-to-inventory ratio from the data, so their results hold whatever the length of downturns and the value of κ_a.

We set the size of the rise and κ_a by the post-pandemic recovery, treated as a permanent rise in demand. In the monthly BEA data behind Figure 1A, the sales-to-inventory ratio peaked 6.8 log points above its 2019 level in April 2021 and fell to half of that peak 9.4 months later, interpolating between months. With the model's inventories measured as BEA measures them, at the end of each month, a permanent rise in sales of 3.9 percent with $\kappa_a = 20.0$ reproduces both numbers (Appendix A.2). The implied rise in sales is close to the 4.4 percent by which real sales exceeded their 2019 level in 2021–22, an untargeted moment.

3 Supply curve of selling capacity

We derive the supply curve of selling capacity and what it implies for the response of prices to demand. Prices are set off the user cost of selling capacity (Section 3.1). With a given stock of capacity, the markup is convex in demand and, for a given retail price, linear in the sales-to-inventory ratio, while across steady states that differ in market power the inventory-to-sales ratio rises with the markup (Section 3.2). Because capacity is costly to create and is kept when demand falls, downturns and surges move the economy along different parts of this curve (Sections 3.3, 3.4, and 3.5). Market power sets the scale of the curve and the utilization of selling capacity its curvature (Section 3.6), and because the curve is convex, uneven demand raises the average markup (Section 3.7).

3.1 User-cost pricing

Subtracting (8) from (7) and using the pricing rule (9), the value of stock satisfies, whenever empty outposts are kept and restock at the maximal rate,

$$(\rho_t + \tau + \delta)\, X_t = (\tau - \zeta)v + \Lambda_t\, \eta_0 P_t + \dot{X}_t. \tag{15}$$

Solving forward, when (15) holds at all future dates,

$$X_t = \int_t^\infty e^{-\int_t^s (\rho_l + \tau + \delta) dl} \left[(\tau - \zeta)v + \Lambda_s \eta_0 P_s \right] ds. \tag{16}$$

A unit of stock is worth the search rents it is expected to earn, $\eta_0 P_s$ per sale at the rate Λ_s, plus the restocking cost it saves. The advantage of a stocked outpost over an empty one lasts only until the empty outpost restocks, so these flows are discounted at the rate $\rho + \tau + \delta$.

The price (9) decomposes into the production cost, the user cost of capacity, and the search rent:

$$\tilde{p}_t = v + (X_t - v) + \eta_0 P_t, \qquad \tau(X_t - v) = (\rho_t + \delta) V_{0t} + \zeta_0 v - \dot{V}_{0t}. \tag{17}$$

The second equation, which is (8) for an empty outpost that is kept, says that the user cost of capacity, $X_t - v$, is the flow cost of an empty outpost—its operating cost and the financing and depreciation of its value—spread over the restocking time $1/\tau$. A sale leaves the outpost out of stock until it restocks, so the price must cover what the outpost forgoes while it restocks; the user cost of capacity is this opportunity cost of a sale. When selling capacity is scarce, an outpost is valuable and the price is high; when capacity is idle, the reverse holds. We call this *user-cost pricing*.

The retail price depends on the wholesale price through distributors' zero-profit condition (4). Substituting (9) into (4) gives an equation in P_t and X_t alone. Its solution rises with the value of stock, but by less, so the wholesale price and the markup rise with the value of stock. The formal result is below.

Lemma 2 (Price block). *For every $X > \underline{X} := (1-\eta_0)c_0 v/\eta_0 - \chi v$, the retail price is the unique solution $P = P(X) > c_0 v/\eta_0$ of*

$$P\left[1 - \eta_0 + \eta_0 \log\left(\frac{\eta_0 P}{c_0 v}\right)\right] = X + \chi v,$$

with $P'(X) = 1/\left[1+\eta_0 \log(\eta_0 P/c_0 v)\right] \in (0,1)$, and P is strictly concave. The wholesale price $\tilde{p} = X + \eta_0 P(X)$ and the markup $\tilde{p}/v$ are increasing in X.

Markups therefore move only through the value of stock. Equation (15) shows how demand per outpost moves it. Writing $\mathbf{T}_{0t} := 1/(\rho_t + \tau + \delta)$ for the effective restocking time, (15) and (9) give

$$\tilde{p}_t = (\tau - \zeta)\mathbf{T}_{0t}\, v + \eta_0 P_t \left(1 + \Lambda_t \mathbf{T}_{0t}\right) + \mathbf{T}_{0t}\dot{X}_t. \tag{18}$$

In a stationary state, in which the stocks of outposts and the value of stock are constant ($\dot{X}_t = 0$), the markup over the discounted restocking cost is the search rent scaled up by $1+\Lambda_t \mathbf{T}_{0t}$, the ratio of the length of a selling cycle—waiting for a buyer, $1/\Lambda_t$, plus restocking, $\mathbf{T}_{0t}$—to the waiting time alone. An outpost maximizes its profit per unit of time. A higher price lengthens the wait for a buyer, and the cost of a longer wait depends on how much of the cycle is spent waiting. When buyers are scarce, waiting dominates the cycle, a higher price is costly, and the markup is low. When buyers are plentiful, the outpost spends most of the cycle restocking, waiting is cheap, and the markup is high. With instantaneous restocking ($\tau \to \infty$, so that $\mathbf{T}_{0t} \to 0$), the markup is the constant search rent, as in standard models of search with taste shocks.

3.2 Supply curve

Define the *utilization of selling capacity* as

$$u_t := \frac{\Lambda_t}{\Lambda_t + \tau + \delta} \in (0,1), \qquad \text{so that} \qquad (\tau + \delta)\frac{M_t}{Q_t} = \frac{1-u_t}{u_t}. \tag{19}$$

Utilization rises with the sales-to-inventory ratio Λ_t relative to the restocking rate. When capacity is maintained by replacement and sales are constant, (12) implies $M = (\tau K - Q)/(\tau+\delta)$, so that $u = Q/(\tau K)$:

sales relative to the maximal flow that the stock of capacity can sustain. It is also, approximately, the share of outposts that are restocking, so that $1-u$ is the share that holds stock and is ready to sell.

When the stock of capacity is given, the price exceeds the production cost by the search rent scaled up by the odds that an outpost is restocking. This scaling makes the supply curve convex in utilization, and utilization summarizes how inventories and restocking times affect the markup. The formal result is below.

Proposition 1 (Supply curve of selling capacity)**.** *Let ρ, τ, the stock of capacity K, and sales Q be constant, with capacity held at K by replacement (producers create outposts at the rate $a=\delta$ and liquidate none), and consider a stationary state given K and Q in which distributors are selective and outposts restock at the maximal rate. Then:*

(i) $u = Q/(\tau K)$ *and* $M/Q = (1-u)/[(\tau+\delta)u]$;

(ii) the price satisfies

$$\tilde{p} - v = \eta_0 P(X)\left[1 + \theta\,\frac{u}{1-u}\right] - \theta' v, \qquad \theta := \frac{\tau+\delta}{\rho+\tau+\delta}, \quad \theta' := \frac{\rho+\delta+\zeta}{\rho+\tau+\delta}; \tag{20}$$

(iii) if $\zeta \le \tau$, the markup $\tilde{p}/v$ and its logarithm are increasing and strictly convex in u on the stationary states that satisfy these conditions, both for a given retail price P and with $P = P(X)$;

(iv) as $(\rho+\delta+\zeta)/\tau \to 0$, $\tilde{p} - v \to \eta_0 P(X)/(1-u)$, and the restocking rate and the inventory-to-sales ratio affect the markup only through u.

The proof is in Appendix A.1. A stationary state takes the stock of capacity as given; the steady state is the stationary state in which producers also choose that stock. The restriction $\zeta \le \tau$, which the calibration satisfies with $\zeta = 0$, keeps the value of stock without search rents, $(\tau-\zeta)v/(\rho+\tau+\delta)$, nonnegative. The price in part (ii) grows without bound as utilization approaches one; near that limit, outposts would want room for a second unit. With a slot for a second unit at $\sigma = 7$ times the operating cost of an outpost, they would leave it unused in stationary states up to a utilization of 80.1 percent, beyond the range of Figure 2, and along each path we solve (Appendix A.1). Part (iv) gives the approximation stated in the introduction: the price is the production cost plus the search rent divided by the share of capacity that

holds stock. In the calibration of Section 2.6, $\theta = 0.984$ and $\theta' = 0.019$, and the approximation overstates $\tilde{p} - v$ by 5 percent at the steady state. The proposition has three implications.

First, the supply curve is convex. When much of the capacity holds stock, a sale empties an outpost that other stocked outposts easily stand in for, and the price is close to the production cost plus the search rent. When little of it does, each sale empties scarce capacity, and the price rises steeply. The convexity comes from the user cost of capacity, which rises with $u/(1-u)$, approximately the odds that an outpost is restocking.

Second, utilization is, approximately, a sufficient statistic for the markup (part (iv)). Long restocking times and thin inventories raise markups through the same channel, the sales-to-inventory ratio relative to the restocking rate. Utilization of selling capacity plays the role that capacity utilization plays for the supply elasticity in Boehm and Pandalai-Nayar (2022), with the difference that it is observable through inventories.

Third, the markup is a forward-looking object. Outside a stationary state, (16) shows that the markup depends on the expected path of demand per outpost, and the supply curve (20) applies to a demand change that is expected to last well beyond the time it takes to restock. With a given stock of capacity, a surge expected to end sooner moves the markup less. When capacity responds, the markup rises less the faster producers are expected to build, so a surge known to be temporary can raise the markup more than a permanent one (Appendix A.2.3).

Proposition 1 takes a simple form in terms of the sales-to-inventory ratio. The odds $u/(1-u)$ equal $\Lambda/(\tau+\delta)$, so the factor that scales up the search rent is linear in sales per unit of inventory. For a given retail price, the price is therefore an affine function of the sales-to-inventory ratio, and the convexity of the supply curve comes entirely from the way the ratio responds to demand. With a given stock of capacity, each additional sale draws down inventories that are replaced only with a delay, so the ratio rises more and more steeply as sales approach the largest flow that the stock of capacity can sustain. The formal result is below.

Corollary 1 (The sales-to-inventory ratio)**.** *Under the conditions of Proposition 1, let* $\Lambda = Q/M$ *be the sales-to-inventory ratio and* $\mathbf{T}_0 = 1/(\rho+\tau+\delta)$*. Then:*

(i) $\tilde{p} - v = \eta_0 P(X)\left(1 + \Lambda \mathbf{T}_0\right) - \theta' v$, *so that, for a given retail price, the price is an increasing affine function of* Λ*;*

(ii) $\Lambda = (\tau + \delta)\, u/(1-u) = (\tau + \delta)\, Q/(\tau K - Q)$, *which is increasing and convex in utilization and, for a given stock of capacity, in sales, and unbounded as* $Q \to \tau K$.

The proof is in Appendix A.1. The slope of the line, $\eta_0 P\, \mathbf{T}_0$, is the search rent times the effective restocking time: an outpost that sells forgoes, while it restocks, the rents of the buyers who arrive at the rate Λ. This is the model's key prediction. The intercept and the slope are fixed by market power and restocking times, the ratio is observable, and the line is the same whether capacity is idle or scarce. With the retail price responding to the value of stock, the relation stays close to linear. In the calibrated model, over a fivefold range of the sales-to-inventory ratio, from half to two and a half times its steady-state value, the markup deviates from a straight line by at most 1.0 percent of its range (Figure 2A). Over this range of the ratio, utilization rises from 40 to 77 percent, and in terms of utilization the markup is convex (Figure 2B). Along this curve, a 1 percent increase in demand raises the markup by 0.6 log points, a 20 percent increase by 15.6 log points, and a 40 percent increase by 51.1 log points, while a 30 percent fall lowers it by 12.0 log points. The sales-to-inventory ratio is therefore the measure of slack in which the supply curve is linear: a given change in the ratio moves the markup by the same amount whether capacity is idle or scarce, while a given change in demand moves the ratio, and with it the markup, by more when capacity is scarce.

Across steady states that differ in market power, markups and inventories are related with the opposite sign. In a steady state, producers create outposts until an empty one is worth what it costs to create, so the value of stock equals the production cost plus the user cost of capacity evaluated at the cost of creating an outpost (13), independently of the parameters that govern search. By (16), the value of stock rises with the search rents that a unit earns at the rate Λ, so with a larger rent per sale, the same value of stock goes with fewer buyers per unit of stock. A seller with more market power therefore holds more inventory relative to sales and charges a higher markup, as in the standard reading of inventories (Bils and Kahn, 2000). The formal result is below.

Corollary 2 (Inventories and markups across steady states)**.** *Consider steady states in which distributors are selective and that differ only in the parameters of search,* η_0, c_0, *and* χ. *Then:*

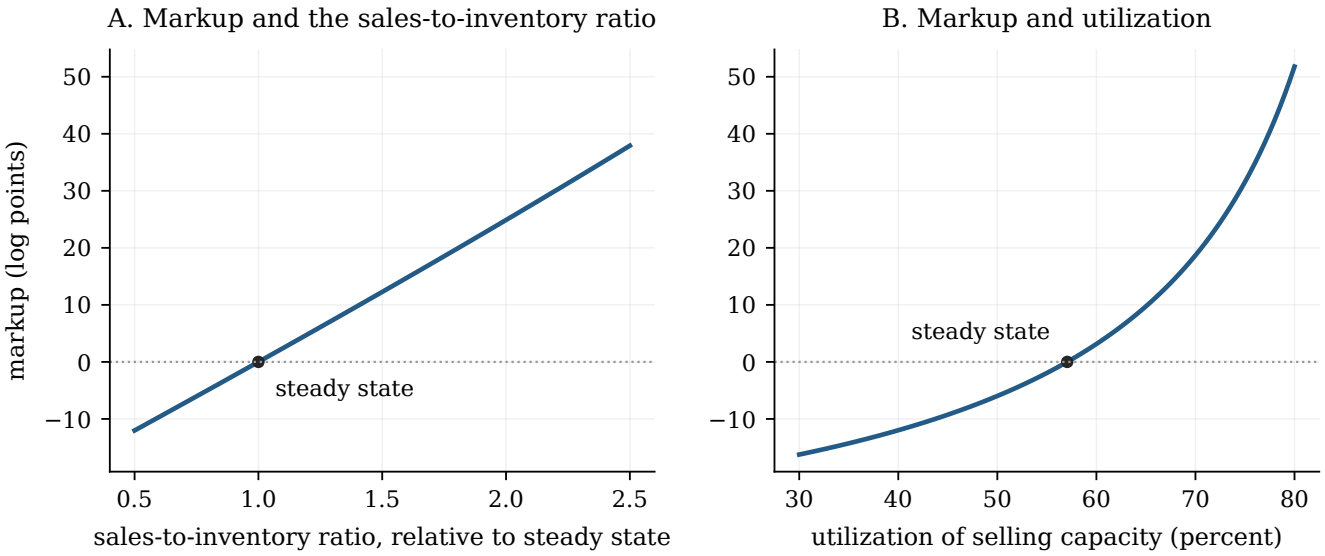

Figure 2: Supply curve of selling capacity

Notes: Stationary states of the calibrated model with the stock of capacity fixed and maintained by replacement (Proposition 1). Markup: $100 \log(\tilde{p}/v)$ relative to the steady state. Panel A: against the sales-to-inventory ratio $\Lambda = Q/M$, relative to its steady-state value. Panel B: against the utilization of selling capacity $u = \Lambda/(\Lambda + \tau + \delta) = Q/(\tau K)$, up to 80 percent.

(i) the value of stock X^{ss} is the same in all of them, and the inventory-to-sales ratio is an increasing affine function of the price,

$$\frac{M^{ss}}{Q^{ss}} = \frac{\tilde{p}^{ss} - X^{ss}}{(\rho + \tau + \delta)X^{ss} - (\tau - \zeta)v};$$

(ii) the search rent $\eta_0 P^{ss}$ is increasing in η_0, so more market power raises both the markup $\tilde{p}^{ss}/v$ and the inventory-to-sales ratio.

The proof is in Appendix A.1. In the calibrated model, raising η_0 by 20 percent raises the markup by 2.5 log points and the inventory-to-sales ratio from 1.50 to 1.76 months. In a steady state, the margin over production cost covers the user cost of capacity and the search rent. With a given stock of capacity, demand moves the value of stock away from this user cost (Corollary 1): when sales per unit of inventory rise, existing capacity earns rents above its user cost, and because the cost of goods sold records only the production cost v, these rents raise measured margins. The creation of outposts, which fixes the value of stock in the corollary, also sets the inventory-to-sales ratio to which capacity returns after a change in demand (Section 3.4).

3.3 Downturns

A fall in demand moves the economy into the inaction region, where capacity is kept, inventories pile up relative to sales, and markups fall slightly.

When demand falls, the value of an empty outpost falls below the cost of creating one, and producers stop creating outposts. They keep the ones they have, because keeping an outpost is cheap and they will need it again. Proposition 2 characterizes this response for a permanent fall in demand at constant interest rates, in the limit in which creating capacity faster than depreciation is free ($\kappa_a \to 0$). The limit isolates the role of the sunk cost ϕv. The formal result is below.

Proposition 2 (Downturns). *Let $\phi > 0$, $\zeta \leq \tau$, $\kappa_a \to 0$, ρ constant, and Assumption A1 holds. At $t = 0$, demand falls permanently and unexpectedly: retail demand D_{0t} follows the path that keeps sales at $\bar{Q} < Q^{ss}$ for all $t \geq 0$, with $\bar{Q}$ close enough to Q^{ss} that no outposts are liquidated, outposts keep restocking at the maximal rate, and distributors stay selective. Then there is $T \in (0, \infty)$ such that:*

(i) for $t \in [0, T)$, no outposts are created or destroyed, and capacity declines at the rate of depreciation;

(ii) for $t \in [0, T)$, utilization is below, the inventory-to-sales ratio above, and the markup below their steady-state values;

(iii) for $t \geq T$, utilization, the inventory-to-sales ratio, and the markup are at their steady-state values.

In the economy with $\phi = 0$ and $\zeta_0 > 0$, in which producers do not create and liquidate outposts at the same date, starting from that economy's own steady state, with $\bar{Q}$ close enough to Q^{ss} that distributors stay selective, empty outposts are liquidated as they appear, and utilization, the inventory-to-sales ratio, and the markup return to their steady-state values at $t^* = \delta^{-1} \log\left[(\bar{Q} + \delta M^{ss})/(\bar{Q} + \delta M^{ss}\bar{Q}/Q^{ss})\right]$, which equals $(1 - \bar{Q}/Q^{ss})M^{ss}/\bar{Q}$, a fraction $Q^{ss}/\bar{Q} - 1$ of the steady-state inventory-to-sales ratio, up to terms of order δ, and stay there.

The proof, in Appendix A.1, constructs the equilibrium in four steps. First, during inaction the equations for the two stocks of outposts are linear and independent of prices: with no creation or destruction of outposts and restocking at the maximal rate, (12) becomes

$$\dot{M}_t = \tau N_t - \bar{Q} - \delta M_t, \qquad \dot{N}_t = \bar{Q} - (\tau + \delta) N_t, \tag{21}$$

and capacity K_t declines at rate δ. Demand per outpost $\Lambda_t = \bar{Q}/M_t$ starts below Λ^{ss} and reaches it in finite time T, because capacity keeps declining. Second, from T on, producers create outposts so as to keep demand per outpost at its steady-state value, and the value of stock is at its steady-state value. Third, a comparison argument applied to (15) shows that $X_t < X^{ss}$ for all $t < T$: the value of stock is the discounted sum of rents that are lower than in the steady state until T and equal to them afterwards. By Lemma 2, the markup is below its steady-state value. Fourth, the value of an empty outpost is the discounted value of the gains from restocking, $\tau(X_t - v) - \zeta_0 v$, which are below their steady-state value before T, so $V_{0t} < \phi v$, and producers indeed create no outposts.

Figure 3 shows the downturn in the calibrated model after a permanent fall in demand of 5 percent. In panel A, each point is the economy's selling capacity, measured against the new, permanently lower sales $\bar{Q}$: horizontally, the outposts holding stock, which is the inventory-to-sales ratio, and vertically, the outposts that are restocking. During the downturn, both stocks change only as outposts sell, restock, and depreciate, as in (21). Inventories are constant on the line $\tau N = \bar{Q} + \delta M$ and rise above it; restocking outposts are constant on the line $(\tau + \delta)N = \bar{Q}$ and fall above it. The shock moves the economy to A: the old capacity, measured against the new sales. A lies above both lines, because the restocking outposts were set up for the old, higher sales. In a first phase, of about 4 months, restocking outposts finish their work faster than new ones join them, and meanwhile they deliver more than sales take, so inventories rise by 4.7 log points and the path moves right and down to the peak. In a second phase, inventories fall, but only as fast as depreciation removes outposts: the path creeps left along the line on which restocking outposts are constant, and capacity declines at the rate of depreciation (part (i)). At B, 30 months after the shock, the ratio is back at its steady-state value, and creating outposts pays again (part (iii)). New outposts start empty, so they add to the restocking outposts, and the economy moves up from B to C, the new steady state.

Panel C shows what happens to the markup. Its horizontal axis is the same inventory-to-sales ratio, and the curve is the supply curve of Figure 2: at each ratio, the markup at which the value of stock is constant, $\dot{X} = 0$ in (15). When the path lies above the curve, the value of stock exceeds its stationary value at that ratio, and by (15) it rises; below the curve it falls (Appendix A.1, proof of Proposition 1). By Lemma 2, the markup moves with it. On impact, sales fall, the ratio jumps to A, and the markup falls

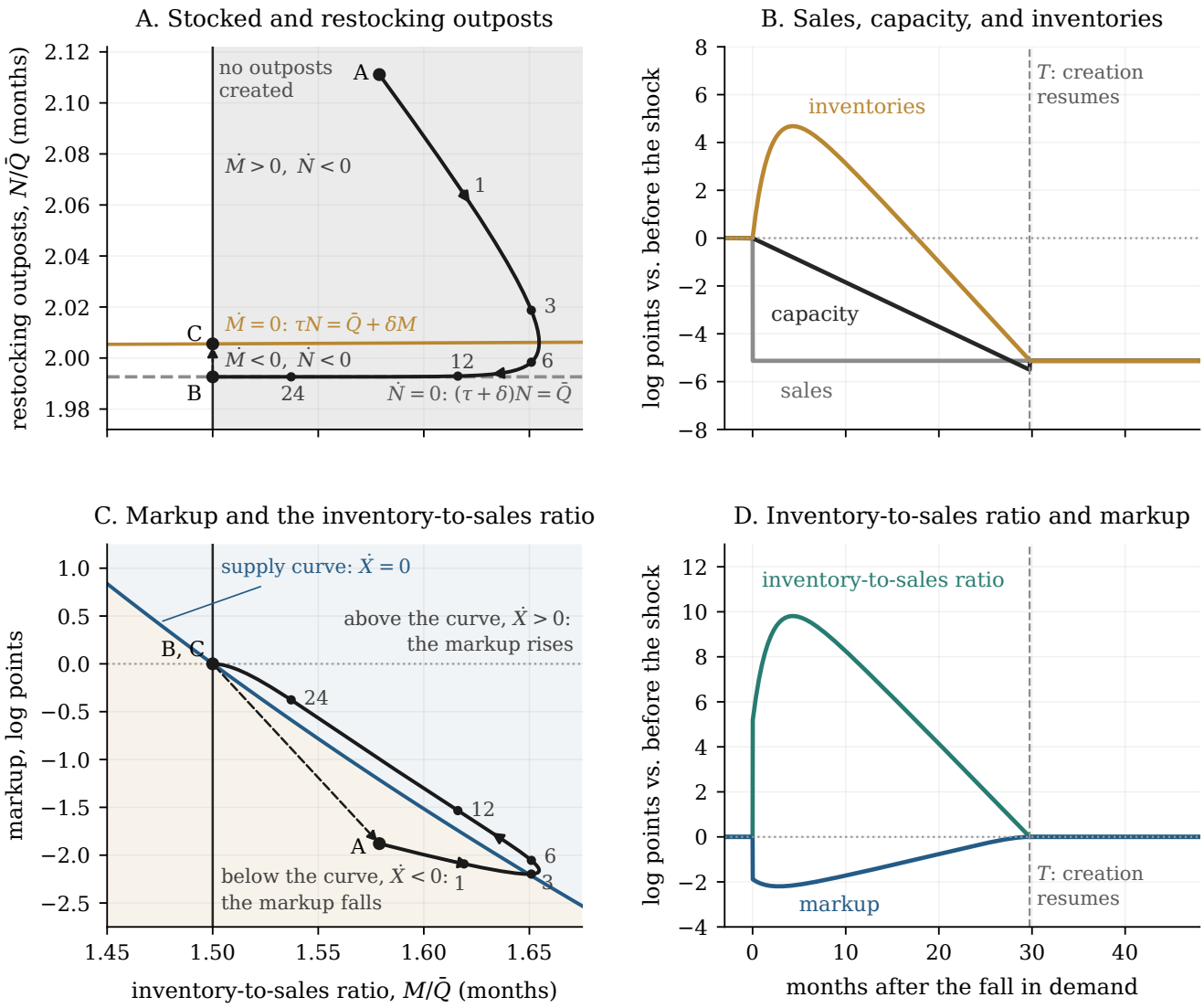

Figure 3: Permanent fall in demand

Notes: Calibrated model. At month 0, demand falls unexpectedly and permanently by 5 percent; producers create outposts only when demand per outpost is back at its steady-state value ($\kappa_a \to 0$, Proposition 2). Top row: quantities; bottom row: prices. Panel A: stocked outposts M against restocking outposts N, both per unit of post-shock sales $\bar{Q}$, so that the horizontal axis is the inventory-to-sales ratio. The two lines are where each stock is constant in (21), inventories on $\dot{M} = 0$ and restocking outposts on $\dot{N} = 0$, and each region shows the signs of $\dot{M}$ and $\dot{N}$. Shaded: the ratio is above its steady-state value $1/\Lambda^{ss}$; along the path, no outposts are created there. A: the economy at the shock; B: the date T at which producers resume creating outposts; C: the new steady state. Panel C: the markup, $100 \log(\tilde{p}/v)$ relative to the steady state, against the same ratio. The curve is the supply curve of Figure 2, the markup at which the value of stock is constant ($\dot{X} = 0$ in (15)); above it the markup rises, and below it the markup falls. B and C coincide with the point before the shock. Dots in panels A and C: 1, 3, 6, 12, and 24 months after the shock. Panels B and D: the same paths over time, in 100 log deviations from their values before the shock; capacity is $K = M + N$. Dashed line: the date T; the outposts created at T start empty, so capacity jumps slightly.

by 1.9 log points: by (16), a unit of stock is worth the search rents it is expected to earn, and demand per outpost stays low until T. The markup lands below the curve, so it keeps falling. It reaches its low point, 2.2 log points below its steady-state value, after about 3 months, where the path crosses the curve,

while inventories are still rising: the value of stock looks ahead, so the markup turns before the ratio does. From then on the path lies above the curve, and the markup rises back to its steady-state value, which it reaches at T (part (ii)). Panels B and D show the same paths over time; the inventory-to-sales ratio peaks 9.8 log points above its steady-state value.

Two features of the downturn matter for the evidence. First, inventories reach their trough after sales, as in the data (Appendix A.3), and the inventory-to-sales ratio rises, as in the recessions of Section 5.2. Second, the fall in the markup is small but lasts as long as the excess capacity, 30 months. Without the sunk cost, producers would close outposts as they empty, and the economy would be back at its steady state in about 4 days, the time it takes to sell the excess stock of an economy without the sunk cost, whose steady state holds 2.5 months of sales in inventories.

Throughout the downturn, utilization is low and the economy is on the flat part of the supply curve. By Corollary 1 and Lemma 2, at the steady state the elasticity of the markup with respect to the sales-to-inventory ratio is $(1+\Theta)\,\eta_0 P\,\Lambda \mathbf{T}_0 \big/ \left[\tilde{p}\,(1-\Theta\Lambda\mathbf{T}_0)\right]$, where $\Theta := \eta_0 P'(X^{ss})$: the share of the search rent in the price, times the ratio of restocking time to waiting time, amplified by the response of the retail price. It is 0.243 in the calibrated model. Along the fixed-capacity supply curve, the elasticity with respect to sales is larger by the factor $1/(1-u)$. This factor rises with utilization because an additional sale directs subsequent buyers to a smaller stock of available goods. A downturn that lowers the sales-to-inventory ratio by a few percent lowers markups by a fraction of a percent.

With capacity given, higher interest rates reinforce the decline in markups, because they lower the value of stock in (16): carrying stock is more expensive. This sign is the opposite of that in customer-market models, in which higher discount rates raise markups as firms harvest their customer base (Gilchrist et al., 2017).

3.4 Surges

A rise in demand moves the economy toward the steep part of the supply curve: capacity is scarce while producers build it, inventories are thin relative to sales, and markups are high. The cost of building fast sets how long this lasts.

When demand rises, demand per outpost rises, and so do the value of stock and the value of an empty outpost. Once an empty outpost is worth more than its cost, producers create outposts faster than they depreciate, at the rate (10), which rises with the gap between the value of an empty outpost and its cost and falls with the cost of building fast, κ_a. Capacity therefore grows gradually, and until it catches up with the new sales, utilization is above its steady-state value. Proposition 3 characterizes this response for a permanent rise in demand at constant interest rates, to first order in the size of the rise. The formal result is below.

Proposition 3 (Surges). *Let $\phi > 0$, $\zeta \leq \tau$, $\underline{\kappa} < \kappa_a < \infty$, ρ constant, and Assumption A1 holds, where $\underline{\kappa}$ is the bound in Appendix A.1. At $t = 0$, demand rises permanently and unexpectedly: retail demand D_{0t} follows the path that keeps sales at $\bar{Q} > Q^{ss}$ for all $t \geq 0$. Then, to first order in $\bar{Q}/Q^{ss} - 1$:*

(i) at every date, producers create outposts faster than they depreciate, and capacity rises monotonically toward its new steady-state level, $K^{ss}\bar{Q}/Q^{ss}$;

(ii) at every date, the inventory-to-sales ratio is below, and the markup above, their steady-state values;

(iii) the deviations from the new steady state are the sum of two decaying exponentials: a fast one, whose rate tends to $\tau + \delta$ as $\kappa_a \to \infty$, and a slow one, whose rate falls as κ_a rises and tends to zero as $\kappa_a \to \infty$.

The proof, in Appendix A.1, linearizes the model around the new steady state, which has the same value of stock and of an empty outpost as the old one (13) and stocks of outposts scaled up by $\bar{Q}/Q^{ss}$. Capacity and stock are given when demand rises, while the values of stock and of an empty outpost jump. When $\kappa_a > \underline{\kappa}$, the linear system has two stable and two unstable roots, all real, so, from the stocks of outposts at the time of the rise, a unique path of the linear system converges to the new steady state. Along it, stock falls short of its new level at every date. By (16), the value of stock, the discounted sum of the search rents it earns, is then above its steady-state value; the value of an empty outpost, the discounted sum of the gains from restocking (17), exceeds the cost of an outpost; and producers create outposts faster than they depreciate. Because building fast is costly, capacity falls short of what the new sales need at every date. When the slow rate is small relative to the rates $\tau + \delta$ and r_X of the linearized system (Appendix A.1), it is close to $\left[\sqrt{(\rho+\delta)^2 + 2\beta/\kappa_a} - (\rho+\delta)\right]/2$, where β is the fall in the gain from restocking, $\tau(X-v) - \zeta_0 v$, per unit of log capacity along the supply curve: a 1 percent rise in capacity lowers the gain by $\beta/100$.

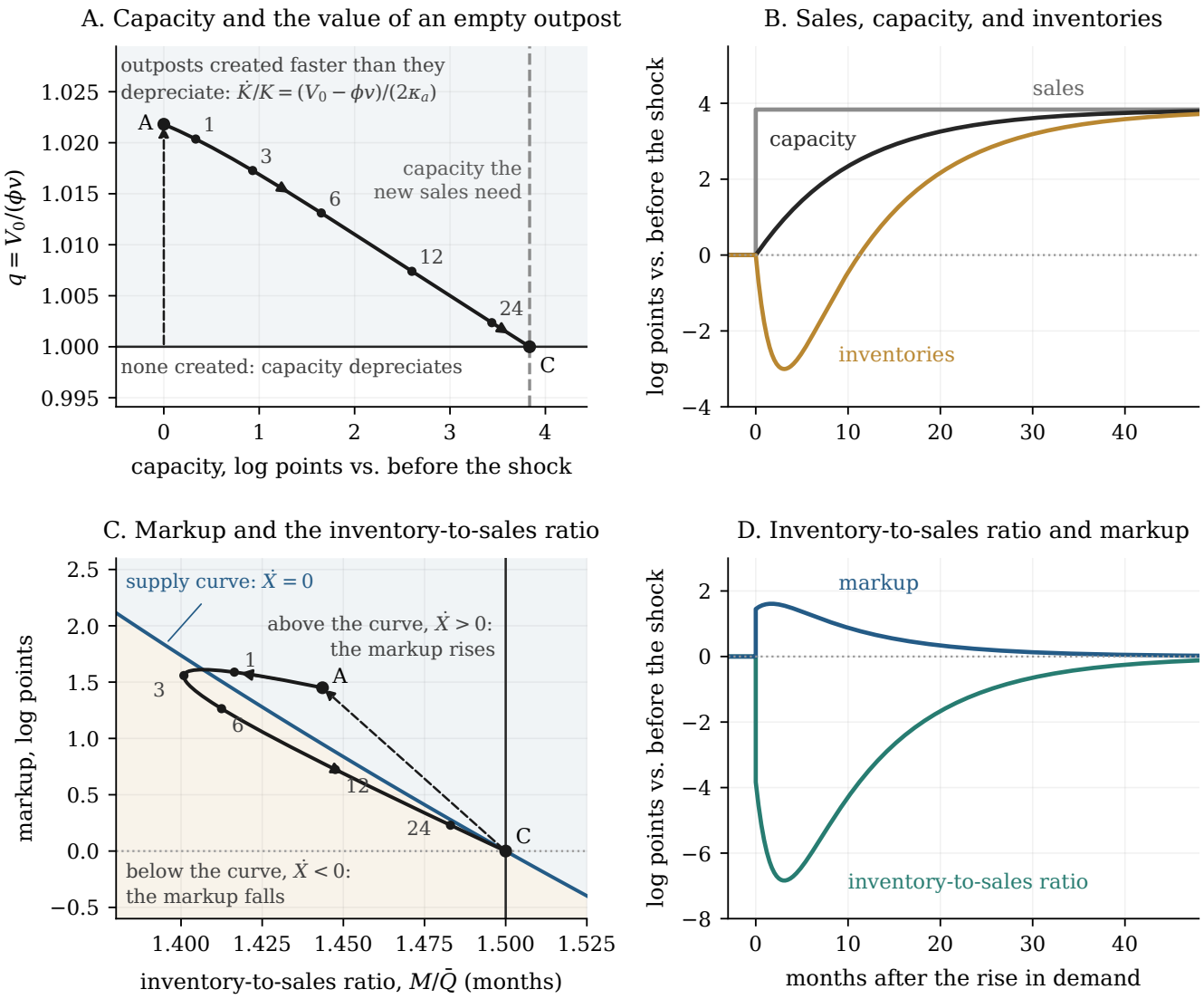

Figure 4: Permanent surge in demand

Notes: Calibrated model, with the size of the surge and κ_a set by the post-pandemic recovery (Section 2.6). At month 0, sales rise unexpectedly and permanently by 3.9 percent; producers create outposts at the rate (10). Panel A: the value of an empty outpost relative to its cost, $q = V_0/(\phi v)$, against capacity, 100 log K relative to its value before the shock. Above the line $q = 1$, outposts are created faster than they depreciate, and capacity grows at the rate $(V_0 - \phi v)/(2\kappa_a)$; below it, none are created. Dashed line: the capacity that the new sales need. A: the economy at the shock; C: the new steady state. Panel C: the markup, $100 \log(\tilde{p}/v)$ relative to the steady state, against the inventory-to-sales ratio $M/\bar{Q}$, with the supply curve of Figure 2 ($\dot{X} = 0$ in (15)); above the curve the markup rises, and below it the markup falls. Dots in panels A and C: 1, 3, 6, 12, and 24 months after the shock. Panels B and D: the same paths over time, in 100 log deviations from their values before the shock.

This is the speed of investment in Tobin's q theory, in which capacity grows in proportion to $q-1$, where $q = V_0/(\phi v)$ is the value of an empty outpost relative to its cost.

At the calibrated values (Section 2.6), κ_a is well above $\underline{\kappa} = 4.2$, and parts (i) and (ii) of Proposition 3 hold along the exact, nonlinear paths for permanent rises in sales of up to 20 percent. Figure 4 shows the calibrated surge. In panel A, each point is the economy's capacity, horizontally, and q, the value of

an empty outpost relative to its cost, vertically. Above the line $q = 1$, the path moves right at a speed proportional to its height above the line. When demand rises, q jumps to 2.2 percent above 1 at A, and the economy moves down toward C as capacity catches up with the new sales. Half of the gap closes in 7 months, and after a year capacity is 2.6 log points higher, against the 3.8 log points that the new sales need.

Panel C shows the markup, as in Figure 3. On impact, sales rise, the ratio falls to A, and the markup rises by 1.5 log points. The path lies above the curve, so the markup keeps rising. Meanwhile, sales draw down stock faster than the restocking outposts replace it, and inventories fall for about 3 months, by 3.0 log points (panel B). The markup peaks at 1.6 log points after about 2 months, where the path crosses the curve, while inventories are still falling: as in the downturn, the markup turns before the ratio does. From then on, the path lies below the curve, and the markup falls as capacity grows; after a year, it is 0.7 log points above its steady-state value.

3.5 Size and sign of demand shocks

The response of prices to demand is convex in the size of the shock and asymmetric in its sign (Proposition 1), and both properties are strongest in the short run. Large surges move the economy into the steep part of the supply curve, where it stays until producers have built the capacity the new sales need; large declines move it into the flat part, where it stays until depreciation removes the excess capacity.

We quantify the nonlinearity. We hit the calibrated economy with temporary demand shocks that peak after 12 months and fade over the following years, of sizes from 5 to 30 percent and of both signs, and hold capacity at its steady-state level over the horizon of the shock. Holding capacity fixed isolates the shape of the supply curve; Section 3.4 shows how capacity creation dampens the response to positive shocks over time. In a downturn, capacity is kept: in the calibration, no outposts are liquidated after a permanent fall in demand of up to 36.5 percent (Proposition 2 and Appendix A.1), more than the largest fall in Figure 5.

Figure 5 shows the results. A 5 percent increase in demand raises the markup by 2.9 log points at its peak, and a 5 percent fall lowers it by 2.6 log points: for small shocks, the response is nearly symmetric, with a slope of about 0.55. A 30 percent increase raises the markup by 27.4 log points, 0.91 per percentage point

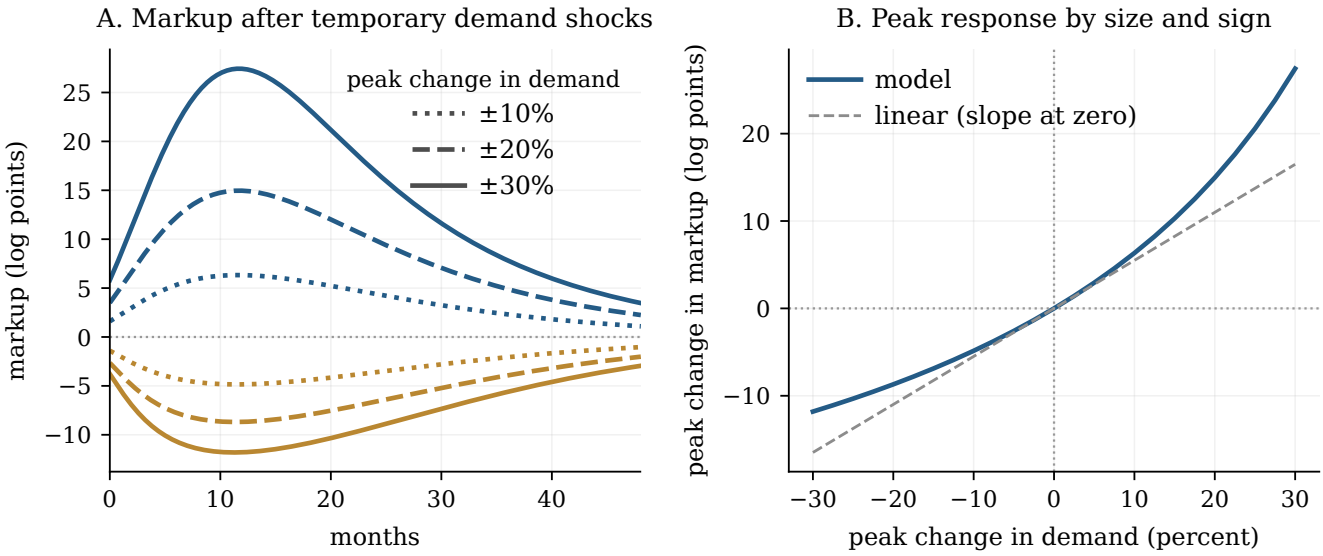

Figure 5: Response of markups to demand shocks of different sizes and signs

Notes: Calibrated model. Demand follows $Q_t = Q^{ss}[1 + s\,(t/12)e^{1-t/12}]$, which peaks at $1 + s$ after 12 months; capacity is held at its steady-state level with replacement; the value of stock solves (15) under perfect foresight. Panel A: markup, $100 \log(\tilde{p}/v)$ relative to the steady state, after increases (blue) and decreases (gold) in demand. Panel B: peak response of the markup as a function of s; the dashed line extrapolates the slope for small shocks.

of demand, while a 30 percent fall lowers it by 11.8 log points, 0.39 per percentage point. The retail price rises by 17.0 log points after the large surge and falls by 6.8 log points after the large decline. With unit production cost fixed in the short run, the markup contributes directly to the change in the wholesale price level. This part of the price response is nearly linear for moderate demand shocks, grows sharply in large surges, and is muted in large downturns.

3.6 Scale and curvature

We ask what makes the supply curve more or less convex. Two features of the economy shape it: market power, through the search rent, and the utilization of selling capacity in normal times. The first sets the scale of the curve and the second its shape.

With the stock of capacity given, utilization moves in proportion to demand, and, for a given retail price, the price moves with $u/(1-u)$, scaled by the search rent (Proposition 1(ii)). A change in demand therefore moves the price by the search rent times a function of utilization alone, and how much more prices rise in a surge than they fall in a downturn of the same size depends only on how much of the capacity is restocking before the shock. The formal result is below.

Corollary 3 (Scale and shape)**.** *Under the conditions of Proposition 1, hold the retail price fixed, and let demand change permanently from Q to $(1+x)Q$ with the stock of capacity given, where $x > -1$ and $(1+x)u < 1$. Then:*

(i) the price changes by

$$\Delta\tilde{p} = \theta\,\eta_0 P\,\frac{xu}{(1-u)\big(1-(1+x)u\big)};$$

(ii) for $0 < x < 1$, the ratio of the responses to an increase and to a decrease in demand of the same size x is

$$\frac{1-(1-x)u}{1-(1+x)u},$$

which depends only on utilization, $u = 1/\big[1+(\tau+\delta)M/Q\big]$, and increases with it.

The proof is in Appendix A.1. Utilization depends on two observables, the inventory-to-sales ratio and the restocking time, so the calibration pins down the curvature of the supply curve. Table 2 quantifies the corollary in the calibrated model, with the retail price responding to the value of stock and the other parameters recalibrated in each row. Lowering the inventory-to-sales ratio from 1.5 to 1 month raises utilization from 57 to 67 percent. A 20 percent increase in demand then raises the markup by 29.6 log points instead of 15.6, while a 20 percent fall lowers it by 12.1 instead of 8.9, so the response to the surge rises from 1.8 to 2.4 times the response to the fall. A restocking time of three months instead of two has the same effect, because it raises utilization by the same amount (Proposition 1(iv)). Lowering the gross margin from 49 to 37 percent lowers every response by about 20 percent and leaves their ratio unchanged. The closed form of part (ii) is within 0.13 of the ratio in the calibrated model in every row.

The corollary separates how strongly prices respond to demand from how convex the response is. Market power governs the first, and the slack in selling capacity in normal times governs the second. An economy with thin inventories relative to the time it takes to restock has a steeper and more convex goods supply curve. With ample inventory coverage, the markup responds less and more symmetrically to demand. Anything that lengthens restocking times when demand surges, such as congestion upstream or shipping delays, raises utilization at given sales and inventories and moves the economy further up the curve. Section 4.2 shows how large this effect would have been in 2021 if purchasing managers' lead times measured restocking times.

Table 2: What sets the scale and the curvature of the supply curve

	Utilization	Change in demand			Surge-to-fall ratio	
Calibration	(percent)	+1%	+20%	−20%	Model	Closed form
Baseline	57	0.57	15.6	−8.9	1.76	1.72
Inventories of 1 month	67	0.88	29.6	−12.1	2.45	2.33
Inventories of 2 months	50	0.42	10.6	−7.0	1.52	1.50
Restocking time of 1 month	40	0.27	6.3	−4.8	1.31	1.31
Restocking time of 3 months	67	0.89	30.1	−12.3	2.44	2.32
Gross margin of 37 percent	57	0.46	12.5	−7.1	1.76	1.72

Notes: Stationary states of the calibrated model with the stock of capacity fixed and maintained by replacement; the retail price responds to the value of stock. Change in demand: change in the markup, $100 \log(\tilde{p}/v)$, in log points, after a permanent change in demand of the size shown. Baseline: the calibration of Section 2.6, with inventories of 1.5 months of sales, a restocking time of 2 months, and a gross margin of 49 percent. Each other row changes one of these targets, with inventories in months of sales, and sets η_0, c_0, χ, ϕ, and ζ_0 to match the remaining four. Surge-to-fall ratio: the change after a 20 percent increase in demand over the absolute change after a 20 percent fall; closed form: $(1 - 0.8u)/(1 - 1.2u)$, from Corollary 3(ii) with the retail price fixed. With a slot for a second unit at $\sigma = 7$ (Appendix A.1), outposts would leave it unused in every state.

3.7 Uneven demand

Consider two recoveries with the same total sales and the same stock of selling capacity in each market. How much can we learn about prices from their aggregate inventories? With capacity maintained by replacement, inventories in market i satisfy $M_i = (\tau K - Q_i)/(\tau + \delta)$. Summing across markets shows that once total sales are given, total inventories are the same in the two recoveries. Yet shifting buyers from one market to another raises sales per unit of inventory by more in the market that gains buyers than it lowers them in the market that loses buyers. Since it is the local sales-to-inventory ratio that enters the seller's pricing rule, average markups rise. So does aggregate revenue over the production cost of the goods sold, which weights each market's price by its sales.

Corollary 4 (Uneven demand). *Consider a continuum of markets $i \in [0, 1]$, each as in Proposition 1, with the same stock of capacity K and sales $Q_i = Q(1 + x_i)$, where $x_i > -1$, $\int_0^1 x_i \, di = 0$, and $\sup_i Q_i < \tau K$. Let $\Lambda_i = Q_i/M_i$. Then:*

(i) the aggregate sales-to-inventory ratio, $\int_0^1 Q_i \, di \big/ \int_0^1 M_i \, di = (\tau + \delta)Q/(\tau K - Q)$, depends only on average sales;

(ii) the average of the markets' ratios, $\int_0^1 \Lambda_i \, di$, exceeds the aggregate ratio unless $x_i = 0$ for almost all i, and, for a given retail price, the average price $\int_0^1 \tilde{p}_i \, di$ is an increasing affine function of it;

(iii) if $\zeta \leq \tau$, the log markup $\log(\tilde{p}_i/v)$ is strictly convex in utilization u_i, both for a given retail price and when each market's retail price responds to its own value of stock, so the average log markup exceeds its value under even demand ($x_i = 0$ for all i), and any nontrivial mean-preserving spread of the distribution of x_i that keeps every market under the conditions of Proposition 1 raises it;

(iv) if $\zeta \leq \tau$, the same holds for aggregate revenue over the production cost of the goods sold, $\int_0^1 Q_i \tilde{p}_i \, di \big/ \left(v \int_0^1 Q_i \, di\right)$, the counterpart of total sales over total cost of goods sold.

The proof is in Appendix A.1. When markets share the same capacity and demand is uneven, the aggregate sales-to-inventory ratio understates the pressure on selling capacity, because the average price depends on the average of the markets' ratios. The prediction from the aggregate ratio in Section 4.2 abstracts from this dispersion, while the cross-sectional evidence relates each firm's margin to its own ratio and so measures the supply curve market by market.

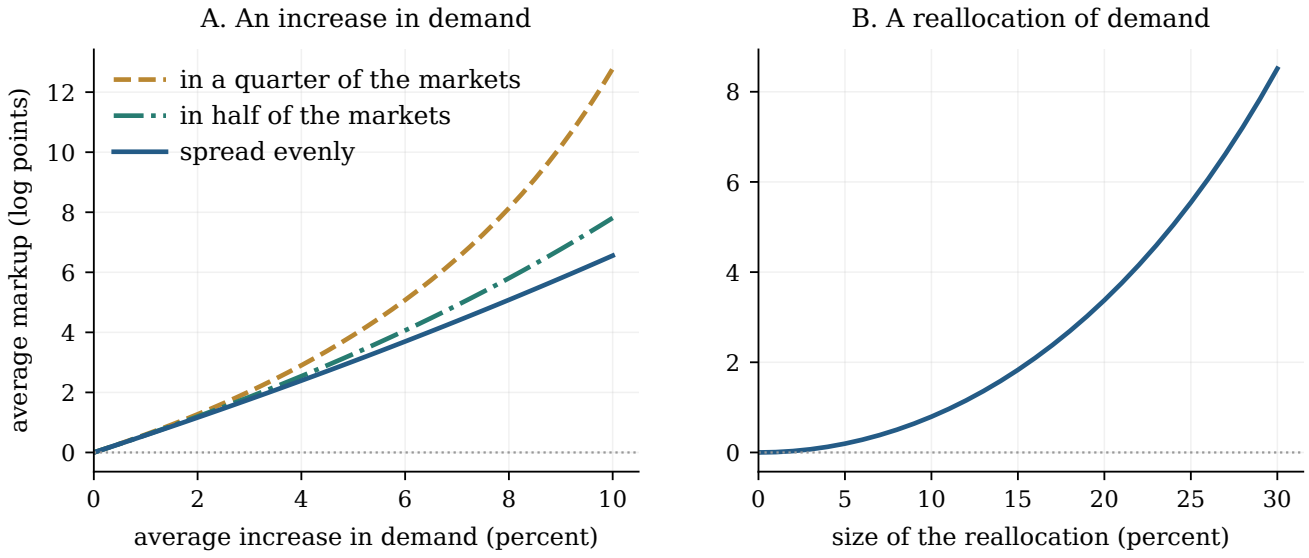

Figure 6: Uneven demand and the average markup

Notes: Calibrated model. Identical markets in stationary states with the same stock of capacity, fixed and maintained by replacement (Corollary 4); each market's retail price responds to its own value of stock. Average markup: mean over markets of $100 \log(\tilde{p}_i/v)$, relative to the steady state. Panel A: an increase in demand spread evenly, or concentrated in half or in a quarter of the markets with no change in the others; the aggregate sales-to-inventory ratio is the same in the three cases, and utilization stays below 80 percent in every market. Panel B: demand rises by the percentage on the horizontal axis in half of the markets and falls by the same percentage in the other half; total demand and the aggregate sales-to-inventory ratio do not change.

Figure 6 quantifies the difference for markets with the same initial capacity, allowing each market's retail price to respond to its own value of stock. In panel A, a 10 percent increase in total demand raises the

average markup by 6.6 log points if demand grows evenly, by 7.8 if the increase is concentrated in half of the markets, and by 12.8 if it is concentrated in a quarter. Aggregate revenue over the production cost of the goods sold, the counterpart of the main measure of gross margins, rises by 8.8 and 19.3 log points in the two concentrated cases (Corollary 4(iv)), more than the average markup, because the markets whose prices rise most also sell more. A statistician observing only total sales and inventories would record exactly the same increase in tightness in all three cases. In panel B, demand rises by 20 percent in half of the markets and falls by 20 percent in the others. Both aggregate sales and aggregate inventories stay at their initial levels, yet the average markup rises by 3.4 log points. These are short-run comparisons at fixed capacity. A wider dispersion of demand can raise the level of goods prices at a given level of aggregate activity. A rise in the markup contributes to inflation while the price level adjusts. As inventories recover, the same mechanism can ease price pressure.

4 Application to COVID-era inflation

The post-pandemic recovery lets us ask where selling capacity was scarce and what happened to prices when that scarcity eased. We first document the behavior of sales, inventories, and gross margins in U.S. manufacturing and trade. We compare sales per unit of inventory with plant utilization as a measure of tightness, then use the observed sales-to-inventory ratio to predict the markup over the recovery. Finally, we feed the model the path of sales and the loss of selling capacity in the pandemic.

4.1 Three facts

Fact 1: sales surged while inventories lagged. In 2020 and 2021, demand shifted toward goods, and producers and distributors could not expand the capacity to deliver them quickly. Real sales of manufacturing and trade fell 9.7 log points below their 2019 average in the second quarter of 2020 and were 4.5 log points above it a year later. In quarterly averages, real inventories stayed below their 2019 level until 2022 and were then rebuilt, to 7.1 log points above it by mid-2023. The sales-to-inventory ratio therefore rose 6.0 log points above its 2019 average by 2021Q2 and returned to it by mid-2022 (Figure 1A). Lead times lengthened (Figure 1B): the average lead time for production materials reported by purchasing

managers rose from 63 days in December 2019 to 88 days in June 2021 and 100 days in mid-2022, and was still 84 to 87 days in the summer of 2023 (Institute for Supply Management, 2019–2023).

Fact 2: aggregate gross margins rose and fell with the sales-to-inventory ratio. On the main measure of Section 2.6, gross margins fell in the first quarter of 2020 (within firms, also in the second), rose by 2.29 log points by the second quarter of 2021 (0.95 within firms), and fell back in 2022, to 1.01 log points below their 2019 level in the third quarter (1.42 within firms; Figure 1C). At the 2021 peak, the gap between the two measures comes mainly from shifts of sales toward firms with higher margins. Margins were highest when inventories were thinnest relative to sales and fell back as inventories were rebuilt. Over 2020Q1–2023Q2, the correlation between the margin and the sales-to-inventory ratio is 0.79 (0.83 within firms), and the margin rises by 0.20 log points (0.15 within firms) per percent increase in the ratio (Newey–West standard error 0.06).

Fact 3: across firms, margins rose most where the sales-to-inventory ratio rose most. Figure 1D sorts firms into deciles by the change in their sales-to-inventory ratio between 2019 and 2021, measured with annual sales and average inventories. The decile of firms whose ratio rose most, by 53 percent at the median, raised their margins by 7.4 log points on average; the decile whose ratio fell most, by 38 percent, saw margins fall by 4.3 log points. Across 1,247 firms, the margin rises by 0.10 log points (standard error 0.03) per percent increase in the ratio with COGS weights and by 0.08 (0.01) with equal weights; across 99 three-digit industries, it rises by 0.14 (0.05), and across the 69 manufacturing industries among them by 0.19 (0.06). The relation is close to linear in the level of the ratio. Adding the square of the percent change gives a coefficient of −0.01 (0.03) with COGS weights and −0.03 (0.01) with equal weights, so the relation is close to a straight line, with a slight concavity when firms are weighted equally. Both sales and inventories contribute: given sales growth, firms that accumulated more inventory raised their margins less (Appendix Table A2). Published industry data show the same relation for prices: across 35 manufacturing industries in the Census survey of manufacturers' shipments and inventories, the producer price index rose by 0.37 log points (standard error 0.17) per percent increase in the sales-to-inventory ratio between 2019 and 2021 (Appendix Table A3).

Four features of accounting data bear on these facts. When input prices rise quickly, firms that value inventories first-in, first-out charge older, cheaper units to the cost of goods sold, which raises measured margins while input prices rise and lowers them when input prices fall. Input prices rose quickly through 2021 and the first half of 2022 and fell in the second half of 2022, so this bias may account for part of the aggregate pattern. The cross-sectional relation is robust to it: the bias is larger for firms that hold more inventories relative to their costs, which works against a positive relation between margins and the sales-to-inventory ratio, and controlling for the 2019 level of the ratio barely changes the estimates. Two others bias the cross-section toward the relation: sales are measured in current dollars, so a firm's own price increase raises both its margin and its ratio, and fixed costs included in the cost of goods sold make margins rise with sales. The coefficient on inventories in column (3) of Appendix Table A2 is free of both, and it has the sign the theory predicts. Given sales, absorption costing moves fixed overhead out of the cost of goods sold when inventories build up and raises measured margins, which works against the relation.

Plant utilization offers a direct comparison. Across 85 manufacturing industry groups outside petroleum, we regress producer-price growth from 2019 to 2021 on both the change in plant utilization and the percent change in sales per unit of inventory. With 2019 sales weights, the coefficient on sales per unit of inventory is 0.43 (standard error 0.10), while the coefficient on plant utilization is −0.13 (0.10). Industries where goods sold more quickly relative to stocks thus had larger price increases even conditional on measured factory utilization (Appendix Table A4). The measures capture different margins of adjustment. The survey asks how much a plant could produce, whereas selling pressure depends on the stock it can deliver before replenishment. These coefficients describe comovement. Industry sales are nominal, and producer prices reflect both input costs and markups.

4.2 Predicted markups

The model's pricing equation is a structural relation between the markup and the path of the sales-to-inventory ratio, which equals demand per outpost: given the path, (15) delivers the value of stock and Lemma 2 the markup. We use it to predict markups from the monthly real sales-to-inventory ratio of manufacturing and trade from 2017 to 2026 (Figure 1A). We treat the ratio's 2019 average as the steady

state, hold the discount rate and the restocking rate at their steady-state values, and let agents foresee the realized path. No parameter is re-estimated, and margins enter the prediction only through their 2019 level, a calibration target.

Lead times are the other determinant of utilization. If the purchasing managers' lead time for production materials measured the restocking time of outposts, the restocking rate would have fallen by more than a third by mid-2022, utilization would have reached 68 percent in November 2021, and the model's markup would have risen by 12.4 log points at its peak in 2021Q3, about 5 times the measured rise. We read this as evidence that these commitment lead times, the time between placing an order and its delivery, overstate the increase in the time needed to replenish stock.

The predicted markup depends on the expected change in the value of stock as well as on the current ratio, through the term $\mathbf{T}_{0t}\dot{X}_t$ in (18). The value of stock discounts rents at $\rho + \tau + \delta = 0.510$ a month, so the markup looks ahead about two months, and except around abrupt reversals the prediction is close to the stationary markup at the current ratio, read off the supply curve of Figure 2A. The exception is the spring of 2020: anticipating the rebound, the predicted markup falls by 0.09 log points in the second quarter, against 1.74 log points in the stationary reading.

Figure 1C and Table 3 compare the predicted markup with measured gross margins. The predicted markup falls in the first half of 2020, rises by 1.39 log points by the second quarter of 2021, between the rise of total sales over total COGS and the mean rise within firms (Fact 2), and is below its 2019 level from mid-2022. Over the 14 quarters from 2020Q1 to 2023Q2, the correlation between the predicted markup and total sales over total COGS is 0.91 (0.84 within firms), and a regression of that margin on the prediction has a slope of 1.14 (Newey–West standard error 0.12). Measured margins fall further than the predicted markup in the second half of 2022, by 1.01 log points in the third quarter (1.42 within firms) against 0.45 (Section 4.1).

The model also accounts for the shape of the relation. By Corollary 1 and Figure 2A, its markup is close to linear in the sales-to-inventory ratio, with a slope of 0.24 log points per percent change in the ratio (Figure 1D). In the data, the relation is close to linear as well (Facts 2 and 3); its slope is close to the model's for aggregate margins over time and across manufacturing industries, and about two-fifths of it across firms. Firms' book inventories include materials and work in process and so measure the stock

Table 3: Sales-to-inventory ratio, predicted markup, and gross margins after COVID

	Data	Model	Data: gross margin		
	Sales-to-inventory	Predicted		Within firms	
Quarter	ratio	markup	Σ sales / Σ COGS	Mean	Median
2020Q2	−8.2	−0.09	+0.08	−1.22	−0.24
2020Q4	+2.8	+0.95	+0.43	+0.20	+0.60
2021Q2	+6.0	+1.39	+2.29	+0.95	+0.88
2021Q4	+4.7	+0.78	+0.82	+0.37	−0.77
2022Q3	−1.4	−0.45	−1.01	−1.42	−1.65
2023Q2	−2.6	−0.43	−0.05	+0.16	+0.13

Notes: Changes relative to 2019, in log points. Data, sales-to-inventory ratio: BEA real sales over real inventories of manufacturing and trade, quarterly mean, relative to the 2019 average. Model, predicted markup: the markup that the calibrated model predicts given the monthly path of the ratio, as total revenue over the production cost of the goods sold in the quarter (Appendix A.2). Data, gross margin: U.S. manufacturing and trade firms in Compustat, whether or not they report inventories, with the exclusions of Section 2.6, relative to the same quarter of 2019: 100 log of total sales over total cost of goods sold (COGS) of the firms present in every quarter from 2019Q1 to 2023Q2 and the within-firm change in 100 log(sales/COGS), mean weighted by 2019 COGS, as in Figure 1C; and the median of the within-firm change across firms in each quarter.

of goods ready for sale with error; averaging within industries removes the part of this error specific to each firm (Appendix A.3.4). The episode tests this linear relation: aggregate utilization of selling capacity rose from 57 to 58.5 percent, a range over which the supply curve is close to linear in demand as well.

The reversal helps interpret the contribution of markups to inflation. As inventories caught up with sales, a stocked seller could expect fewer buyers to arrive while it waited for a replacement. This reduced the incentive to keep prices elevated relative to production cost. Thus rebuilding inventories can contribute to disinflation even while sales remain above their pre-pandemic level. The calculation conditions on the realized path of sales and inventories, which can reflect both changes in demand and disruptions to supply.

The recovery from COVID offers a reason to take the distribution across markets seriously. In Drozd and Tavares (2026), the dispersion of industry employment losses was similar in the COVID and Great Recession contractions, but the variance of employment growth during the recovery was about two and a half times as large after COVID. In that paper's production-network model, equalizing inherited workforce capacity relative to steady-state employment between hit and safe producers within each chain and tier reduces the rebuilding contribution to cumulative above-trend inflation through 2022 by 57 percent, from 4.76 to 2.05 percentage points. The experiment holds group capacity totals and the

inferred demand and productivity paths fixed. Here we hold the initial distribution of capacity fixed and vary where buyers arrive. In the companion model, concentrating losses makes recruitment and training more costly at the producers that must rebuild most. In our model, concentrating buyers raises the rents a seller forgoes while restocking where demand has surged. Shifts in spending can also affect inflation through downward wage rigidity (Guerrieri et al., 2021) or costly movements of labor across sectors (Ferrante et al., 2023).

How much might the aggregate ratio miss? We apply the fixed-capacity supply curves in Figure 6 to relative changes in real shipments across 35 manufacturing industries, removing aggregate shipments growth and weighting industries by their initial shipments (Appendix A.3.6). If these relative changes are treated as shifts in demand across initially identical markets, the 2019–21 distribution raises the model's average markup by 0.61 log points relative to an even allocation. The corresponding increase is 0.37 log points for 2017–19 and 1.03 for 2013–15. Uneven demand can matter at unchanged aggregate sales, but this manufacturing comparison also shows that its effect need not be unusually large in the pandemic window. These are separate fixed-capacity experiments, but they suggest that uneven demand could account for part of the gap between the conditional prediction in Table 3 and the rise of total sales over total COGS in 2021.

4.3 Loss of capacity and recovery

Section 4.2 took the sales-to-inventory ratio from the data; here the model produces it. The pandemic was a destructive recession followed by a sharp recovery. When demand returned, firms had lost part of their capacity to sell, so the recovery was a surge relative to the capacity that remained: the reverse of the transient surge in Appendix A.2.3, which is a surge followed by a downturn. Fed with the path of sales and this loss of capacity, the calibrated model tracks the shortage of stock in 2021 and the timing of the rise and reversal of margins.

Drozd and Tavares (2026) measure the organizational capital that firms lost in the pandemic. In the model, a fall in demand alone leaves capacity in place: producers keep their outposts after a permanent fall of up to 36.5 percent (Proposition 2 and Appendix A.1). We therefore model the pandemic as an unexpected loss of outposts in April 2020, stocked and restocking alike, and feed the model the observed

path of sales relative to its pre-pandemic trend from April 2020 to December 2022.[3] Treating outposts as part of organizational capital, we measure their loss by the jobs that manufacturing and trade had not recovered a year later, 3.7 percent of their employment.[4] The loss is close to the one that the calibration of Section 2.6 implies: because only sales relative to capacity matter in the model, the permanent rise in sales of 3.9 percent in Figure 4 is the same as a loss of 3.8 percent of outposts with sales unchanged.[5]

Figure 7 shows two versions. In the first, producers foresee the path of sales from April 2020 on. In the second, they expect the sales of each month of the lockdown, April and May 2020, to last for good, and the rebound in June comes as a surprise.

In the lockdown, sales fall 19 log points below trend in April and 11 in May (panel B). Producers keep the outposts they have, inventories pile up relative to sales (panel D), and the markup falls (panel C). If the rebound comes as a surprise, producers stop creating outposts, as q falls to 0.77 on average in April (panel A); if they foresee it, they keep building through the lockdown. When sales return, capacity falls short of what they need. From August 2020 to March 2022, stock is scarce: the inventory-to-sales ratio is below normal, its steady-state value in the model and its 2019 average in the data, and the markup is above normal. The ratio reaches its low in April 2021, 7.7 log points below normal, against 6.8 in the data in the same month, the low that the calibration of Section 2.6 targets; the markup peaks in 2021Q1 at 1.37 log points, between the two measures of gross margins in panel C, which peak in 2021Q2 (Fact 2). Early in 2022, sales fall below trend again, to about where they were before the pandemic; from late February 2022 to mid-May 2022, q is just below 1, producers create no outposts, and capacity shrinks with depreciation. From May 2022, the ratio is above normal, in the model as in the data. This glut comes

[3]Sales are BEA real sales of manufacturing and trade, in log deviations from a trend that grows at their least-squares growth rate over 2015–2019, 1.8 percent a year, and passes through their 2019 average. They change at the start of each month and return linearly to trend over the 12 months after December 2022. In January–February 2020, sales were 2.0 log points below this trend, and the inventory-to-sales ratio was 1.5 log points above its 2019 average.

[4]The loss is the shortfall of private employment in manufacturing, wholesale, and retail trade in the Quarterly Census of Employment and Wages (QCEW), with the dating and weights of Drozd and Tavares (2026): the change from January–February 2020 to January–February 2021 in each 6-digit NAICS industry, applied to its employment in 2019, summed within each of the 237 BEA industries of these sectors, and summed over the BEA industries that lost employment. Netting gains against losses across the three sectors gives 3.4 percent, and subtracting the 20 percent of the shortfall that Drozd and Tavares (2026) let firms recover through recalls gives 2.9 percent. With these two losses, the correlations of the model's inventories and markup with the data reported below are 0.48 and 0.82, and 0.38 and 0.82; with the rebound a surprise, they are 0.59 and 0.80.

[5]The measured loss is likely a lower bound on the loss of capacity relative to demand. Gains offset losses within each BEA industry, and demand also shifted across finer markets, so capacity sat idle in some markets while others faced scarcity (Section 3.7). Section 4.1 documents such shifts across firms.

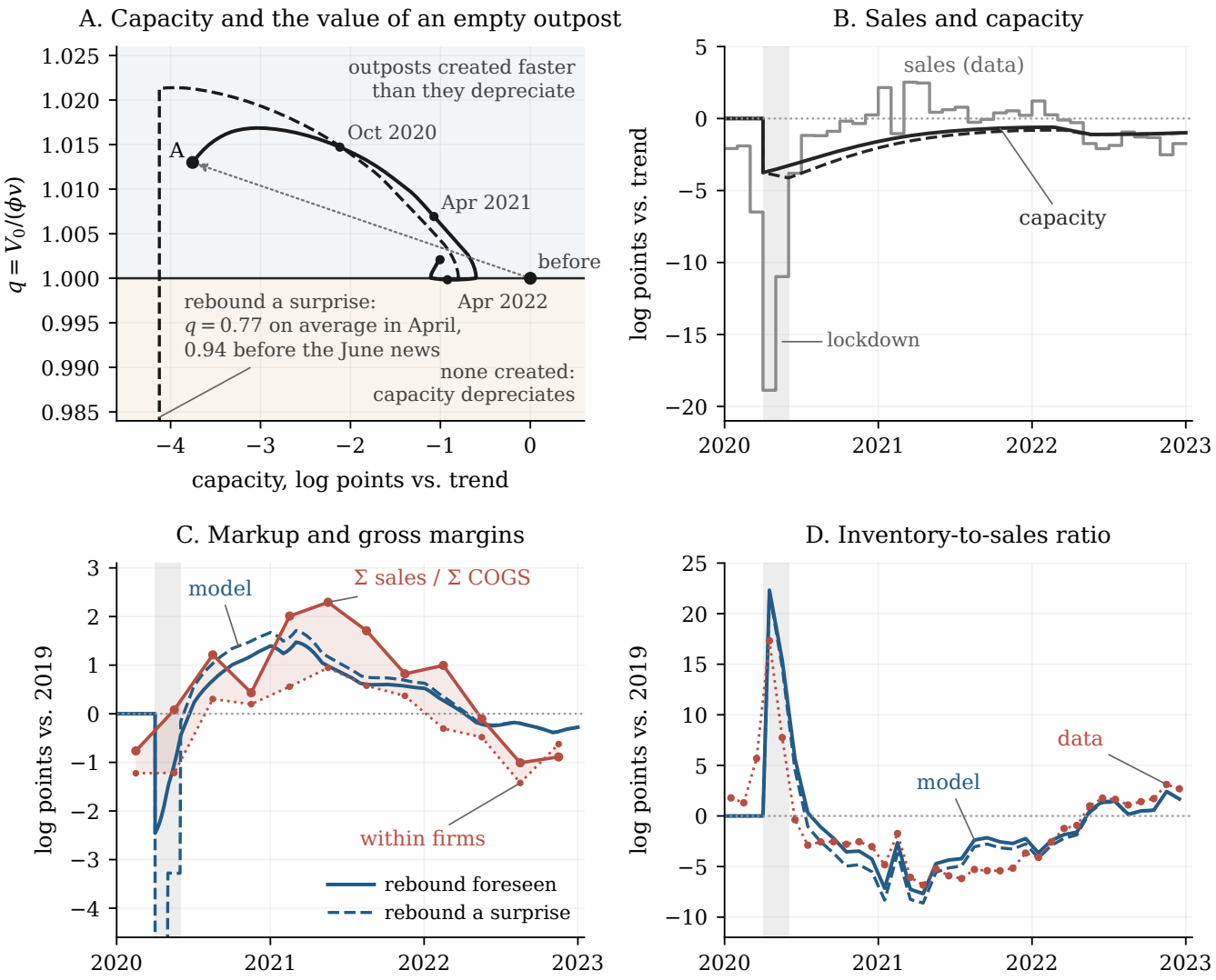

Figure 7: Destructive shock with a sharp recovery: COVID era

Notes: Calibrated model, with κ_a as in Figure 4. In April 2020, 3.7 percent of outposts, stocked and restocking alike, are lost, and sales follow BEA real sales of manufacturing and trade relative to trend (text). Solid lines: producers foresee the path of sales from April 2020 on. Dashed lines: in April and in May 2020, they expect that month's sales to last for good, and the rebound in June comes as a surprise. Before April 2020, the model is in its steady state. Panel A: the paths of $q = V_0/(\phi v)$ and capacity, $100 \log K$ relative to trend; above the line $q = 1$, outposts are created faster than they depreciate, and below it none are created. The dot on the line at zero capacity is the economy before the pandemic; the loss of outposts and the news move it at once to A (dotted arrow). On the dashed path, the news in May and in June 2020 moves q at the capacity inherited then, and the path is below the panel until the June news (values shown). The other dots mark the middle of the months shown and, at the end of the paths, December 2022. Panels B–D: shaded, the lockdown, April and May 2020; the data start in January 2020. Panel B: sales (data, gray) and capacity, in log points relative to trend. Panel C: the markup (blue), $100 \log(\tilde{p}/v)$ relative to the steady state, and gross margins (red, with dots at quarters) in the two measures of Figure 1C, total sales over total cost of goods sold (solid) and the mean change within firms (dotted), with the range between them shaded. The dashed markup falls to −5.4 log points in April 2020, below the panel. Panel D: the inventory-to-sales ratio, inventories at the end of each month over the month's sales, in the model (blue) and in the BEA data (red, dotted, with dots), in 100 log deviations from 2019.

from the path of sales, which the model takes from the data: with sales back at trend after the lockdown, the ratio would still be below normal at the end of 2022.

Because the model's ratio shares its denominator with the data's, the test is inventories and the markup. We measure the model's inventories as BEA measures them, at the end of each month. Because the calibration of κ_a also uses the speed of the ratio's recovery, the comparison rests on the rest of the path of inventories and on the markup. From June 2020 to December 2022, the correlation between inventories in the model and in the data, both relative to the trend of sales at the 2019 ratio, is 0.51, and the root-mean-square gap between the model's ratio and the data's is 1.9 log points, against 3.4 if inventories had stayed on trend. With the same sales and no loss of outposts, the gap is 4.2 log points and the correlation of inventories −0.19. Over 2020Q3–2022Q4, the correlation between the model's markup, measured as total revenue over the production cost of the goods sold in each quarter, and total sales over total cost of goods sold (COGS) is 0.81 (0.83 within firms; Section 4.1).

The model misses three features of the data. In the lockdown, restocking continues and inventories rise in the model, while in the data they fell: the model's ratio rises 22.3 log points in April 2020, against 17.3 in the data, and stays above the data's through September 2020. The model then rebuilds stock earlier and by less: inventories reach their low in April 2021 in the model and in August 2021 in the data, and from August 2022 they are above trend in the data, while in the model they are back at trend only from November 2022. Margins fell further in the second half of 2022, on both measures (Section 4.1).

5 Normal times and recessions

The pre-pandemic response to a monetary tightening and the three recessions before the pandemic provide comparisons with ordinary fluctuations in demand and with downturns. They test the same pricing relation where the sales-to-inventory ratio barely moves, after monetary tightenings, and where it falls, in recessions.

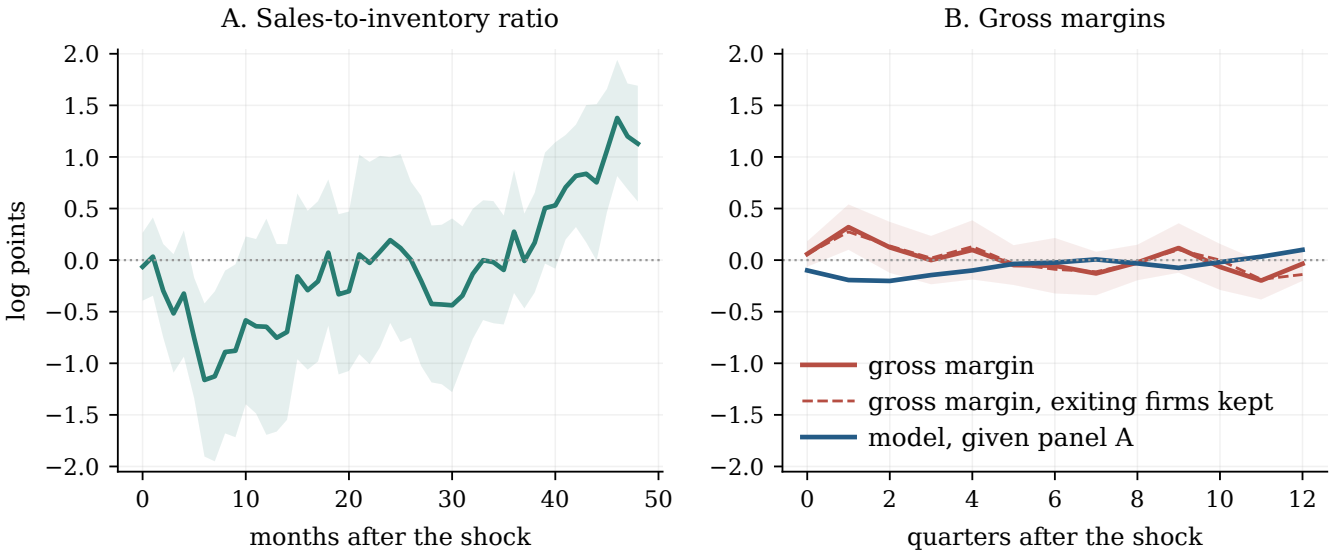

Figure 8: Sales-to-inventory ratio and gross margins after a monetary tightening

Notes: Panel A: local projection of 100 log of the BEA real sales-to-inventory ratio of manufacturing and trade; the one-year yield is instrumented by the FF4 surprise of Gertler and Karadi (2015), 1991:1–2012:6; controls: 12 lags of the one-year yield, the excess bond premium, the log consumer price index, and log real sales and inventories; 90 percent Newey–West intervals. Panel B: firm-level local projections of the change in 100 log(sales/COGS), manufacturing and trade firms in Compustat North America, with the same surprise summed within the quarter, 1991Q1–2012Q2; COGS weights and firm fixed effects; 90 percent intervals (Driscoll and Kraay, 1998); dashed: firms that stop reporting kept at their last margin. Model: the markup that the calibrated model predicts given the response of panel A, months 0–48, with the discount rate at its steady-state value; quarterly means. Responses are scaled to a shock that raises the one-year yield by 0.195 percentage points on impact.

5.1 Normal times

In normal times, demand moves less, and so does the sales-to-inventory ratio. We estimate the response of the ratio to a monetary tightening with local projections (Jordà, 2005). The one-year Treasury yield is instrumented by the surprise in three-month-ahead federal funds futures around policy announcements (FF4 in Gertler and Karadi, 2015), over 1991:1–2012:6, with 12 lags of the one-year yield, the excess bond premium of Gilchrist and Zakrajšek (2012), the log consumer price index, and log real sales and inventories as controls (Montiel Olea and Plagborg-Møller, 2021). After a tightening that raises the one-year yield by about 20 basis points, sales fall, inventories follow with a lag, and the ratio falls by 1.2 log points after 6 months (standard error 0.45; Figure 8A).

We estimate the response of gross margins at the firm level and at quarterly frequency, which respects the timing of fiscal-quarter reports. The change in a firm's log gross margin between the quarter before the shock and h quarters after it is regressed on the one-year yield, averaged over the quarter and

instrumented by the sum of the surprises within the quarter. The controls are four lags of the macroeconomic variables, the firm's lagged margin change, its size, quarter-of-year dummies, and firm fixed effects; observations are weighted by the firm's share of the cost of goods sold, and standard errors allow for correlation across firms and over time (Driscoll and Kraay, 1998). Margins show no robust response (Figure 8B): over 13 quarters, they stay within 0.32 log points of zero and are significant at the 10 percent level at 2 horizons, and keeping firms that stop reporting at their last margin changes little. Given the estimated response of the ratio, with the discount rate held at its steady-state value, the model predicts a fall in the markup of at most 0.20 log points, and the prediction lies within the 90 percent intervals of the estimated margins at 9 of the 13 horizons. The rise in interest rates that comes with a tightening would lower the model's markup further (Section 3.3). Appendix A.3 reports the responses to the shocks of Bauer and Swanson (2023) and Jarociński and Karadi (2020), which give the same picture.

5.2 Recessions

In a recession, inventories pile up relative to sales, and the theory predicts that markups fall along the flat part of the supply curve. Appendix Table A5 reports the three recessions before the pandemic at the quarter in which the sales-to-inventory ratio of manufacturing and trade reached its trough. Gross margins are measured relative to the same quarter of the last full year of the preceding expansion, in the two measures of Figure 1C. In each recession both fell: total sales over total COGS of the firms present in every quarter of the five-year window by 0.3 to 4.1 log points, and the COGS-weighted mean within firms by 1.6 to 3.1 log points; so did the median margin. Given the path of the ratio, the calibrated model predicts falls of 0.5 to 2.3 log points: smaller than the mean fall within firms in 1990–91 and 2001, and larger in 2007–09. The recoveries fit less well: in 1990–91 and 2001, margins within firms were still below their pre-recession level a year after the ratio had returned to its own, and in 2007–09 they recovered before the ratio. Fixed costs included in the cost of goods sold make measured margins fall with sales (Section 4.1), so part of the measured falls is mechanical. Under the reading in which inventories help firms sell, the rise in inventories relative to sales would instead signal higher markups.

6 Conclusion

The opportunity cost of a sale depends on how soon the seller can replace the unit it sells. When buyers arrive more quickly than a seller can replenish its stock, selling today means giving up the chance to earn a rent from another buyer. This gives the firm a reason to raise its price relative to production cost even when the underlying degree of market power is unchanged. Sales per unit of inventory measure this opportunity cost, and because the ratio rises more steeply as sales approach capacity, a high ratio signals that goods prices will respond strongly to further increases in demand. Over the cycle, inventories therefore signal markups with the sign opposite to the standard reading, which holds across steady states that differ in market power. In a downturn, firms retain their selling capacity, inventories pile up relative to sales, and markups fall along the flat part of the supply curve. During the post-pandemic recovery, sales outran inventories, and gross margins rose most at firms where stocks became scarce relative to sales. As stocks were rebuilt, the opportunity cost of a sale fell, helping account for the reversal in goods margins.

A shift in spending raises markups where buyers arrive faster than sellers can replenish their stocks. Because those markups rise along an increasingly steep supply curve, their rise can outweigh the decline in markets that lose buyers. The distribution of demand can therefore put pressure on goods prices even when aggregate sales and inventories change little. This gives an aggregate Phillips curve a reason to depend on where spending returns. The response of inflation to employment also depends on wage setting, the response of production costs, and monetary policy.

References

Hassan Afrouzi, Andres Drenik, and Ryan Kim. Concentration, market power, and misallocation: The role of endogenous customer acquisition. Working Paper 31415, National Bureau of Economic Research, 2023.

Hassan Afrouzi, Andres Blanco, Andres Drenik, and Erik Hurst. A theory of how workers keep up with inflation. Working Paper 33233, National Bureau of Economic Research, 2024.

George A Alessandria, Shafaat Y Khan, Armen Khederlarian, Carter B Mix, and Kim J Ruhl. The aggregate effects of global and local supply chain disruptions: 2020–2022. Working Paper 30849, National Bureau of Economic Research, January 2023. URL http://www.nber.org/papers/w30849.

Simon P Anderson and Regis Renault. Pricing, product diversity, and search costs: A bertrand-chamberlin-diamond model. *RAND Journal of Economics*, pages 719–735, 1999.

Xiwen Bai, Jesús Fernández-Villaverde, Yiliang Li, and Francesco Zanetti. The causal effects of global supply chain disruptions on macroeconomic outcomes: Evidence and theory. Working Paper 32098, National Bureau of Economic Research, 2024.

Michael D. Bauer and Eric T. Swanson. A reassessment of monetary policy surprises and high-frequency identification. *NBER Macroeconomics Annual*, 37:87–155, 2023.

Pierpaolo Benigno and Gauti B. Eggertsson. It's baaack: The surge in inflation in the 2020s and the return of the non-linear Phillips curve. Working Paper 31197, National Bureau of Economic Research, 2023.

Mark Bils and James A. Kahn. What inventory behavior tells us about business cycles. *American Economic Review*, 90(3):458–481, June 2000. doi: 10.1257/aer.90.3.458. URL https://www.aeaweb.org/articles?id=10.1257/aer.90.3.458.

Christoph E. Boehm and Nitya Pandalai-Nayar. Convex supply curves. *American Economic Review*, 112 (12):3941–69, December 2022. doi: 10.1257/aer.20210811. URL https://www.aeaweb.org/articles?id=10.1257/aer.20210811.

Clement E. Bohr. Capacity buffers: Explaining the retreat and return of the Phillips curve. Working paper, Northwestern University, January 2024.

Philip Bunn, Lena Anayi, Emily Barnes, Nicholas Bloom, Aaron Jalca, Paul Mizen, Gregory Thwaites, and Ivan Yotzov. How curvy is the Phillips curve? Working paper, Bank of England, Stanford University and Federal Reserve Bank of Atlanta, April 2026.

Kenneth Burdett, Shouyong Shi, and Randall Wright. Pricing and matching with frictions. *Journal of Political Economy*, 109(5):1060–1085, 2001.

Bureau of Economic Analysis. BEA rates of depreciation, service lives, declining-balance rates, and Hulten–Wykoff categories. https://apps.bea.gov/national/pdf/BEA_depreciation_rates.pdf, n.d.

Alberto Cavallo and Oleksiy Kryvtsov. What can stockouts tell us about inflation? evidence from online micro data. *Journal of International Economics*, 146:103769, 2023. doi: https://doi.org/10.1016/j.jinteco.2023.103769. URL https://www.sciencedirect.com/science/article/pii/S0022199623000557.

Diego A. Comin, Robert C. Johnson, and Callum J. Jones. Supply chain constraints and inflation. Working Paper 31179, National Bureau of Economic Research, 2023.

Marco Del Negro, Michele Lenza, Giorgio E Primiceri, and Andrea Tambalotti. What's up with the phillips curve? Technical report, National Bureau of Economic Research, 2020.

Deloitte Research Center for Energy & Industrials. 2024 manufacturing industry outlook. Industry report, Deloitte Research Center for Energy & Industrials, October 2024.

John C. Driscoll and Aart C. Kraay. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, 80(4):549–560, 1998.

Lukasz A. Drozd and Jaromir B. Nosal. Understanding international prices: Customers as capital. *American Economic Review*, 102(1):364–95, February 2012. doi: 10.1257/aer.102.1.364. URL https://www.aeaweb.org/articles?id=10.1257/aer.102.1.364.

Lukasz A. Drozd and Marina M. Tavares. Organizational capital destruction and post-COVID inflation. Working paper, Federal Reserve Bank of Philadelphia and International Monetary Fund, October 2026.

Francesco Ferrante, Sebastian Graves, and Matteo Iacoviello. The inflationary effects of sectoral reallocation. *Journal of Monetary Economics*, 140(S):S64–S81, 2023.

Luca Gagliardone, Mark Gertler, Simone Lenzu, and Joris Tielens. Anatomy of the Phillips curve: Micro evidence and macro implications. Working Paper 31382, National Bureau of Economic Research, 2025. Revised May 2025.

Mark Gertler and Peter Karadi. Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76, January 2015. doi: 10.1257/mac.20130329. URL https://www.aeaweb.org/articles?id=10.1257/mac.20130329.

Simon Gilchrist and Egon Zakrajšek. Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720, June 2012. doi: 10.1257/aer.102.4.1692. URL https://www.aeaweb.org/articles?id=10.1257/aer.102.4.1692.

Simon Gilchrist, Raphael Schoenle, Jae Sim, and Egon Zakrajšek. Inflation dynamics during the financial crisis. *American Economic Review*, 107(3):785–823, 2017.

Andrew Glover, José Mustre-del Río, and Jalen Nichols. Corporate profits contributed a lot to inflation in 2021 but little in 2022—a pattern seen in past economic recoveries. *Federal Reserve Bank of Kansas City Economic Bulletin*, May 2023a.

Andrew Glover, José Mustre-del Río, and Alice von Ende-Becker. How much have record corporate profits contributed to recent inflation? *Federal Reserve Bank of Kansas City Economic Review*, 108(1): 1–13, 2023b.

Francois Gourio and Leena Rudanko. Customer capital. *The Review of Economic Studies*, 81(3):1102–1136, 2014. URL https://EconPapers.repec.org/RePEc:oup:restud:v:81:y:2014:i:3:p:1102-1136.

Veronica Guerrieri, Guido Lorenzoni, Ludwig Straub, and Iván Werning. Monetary policy in times of structural reallocation. In *Macroeconomic Policy in an Uneven Economy: Proceedings of the Economic Policy Symposium, Jackson Hole*, Kansas City, MO, 2021. Federal Reserve Bank of Kansas City.

Martin Harding, Jesper Lindé, and Mathias Trabandt. Understanding post-COVID inflation dynamics. *Journal of Monetary Economics*, 140(Supplement):S101–S118, 2023.

Jonathon Hazell, Juan Herreno, Emi Nakamura, and Jon Steinsson. The slope of the phillips curve: evidence from us states. *The Quarterly Journal of Economics*, 137(3):1299–1344, 2022.

Martin B. Holm, Lars O. Lerdalen, and David Vines. The effect of capacity constraints on the slope of the Phillips curve. *Oxford Economic Papers*, 76(4):1226–1238, 2024.

Joachim Hubmer and Lukas Nord. Investment in demand and dynamic competition for customers. PIER Working Paper 26-006, Penn Institute for Economic Research, 2026.

Institute for Supply Management. Manufacturing ISM report on business. Monthly reports, December 2019, December 2020, June 2021, December 2021, April 2022, July 2022, August 2023, 2019–2023.

Marek Jarociński and Peter Karadi. Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics*, 12(2):1–43, 2020.

Òscar Jordà. Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182, March 2005. doi: 10.1257/0002828053828518. URL https://www.aeaweb.org/articles?id=10.1257/0002828053828518.

James A. Kahn. Inventories and the volatility of production. *The American Economic Review*, 77(4): 667–679, 1987. ISSN 00028282. URL http://www.jstor.org/stable/1814538.

Aubhik Khan and Julia K. Thomas. Inventories and the business cycle: An equilibrium analysis of (s, s) policies. *American Economic Review*, 97(4):1165–1188, September 2007. doi: 10.1257/aer.97.4.1165. URL https://www.aeaweb.org/articles?id=10.1257/aer.97.4.1165.

Oleksiy Kryvtsov and Virgiliu Midrigan. Inventories, Markups, and Real Rigidities in Menu Cost Models. *The Review of Economic Studies*, 80(1):249–276, 07 2013. ISSN 0034-6527. doi: 10.1093/restud/rds028. URL https://doi.org/10.1093/restud/rds028.

Guido Menzio. Markups: A search-theoretic perspective. Working Paper 32888, National Bureau of Economic Research, 2024.

Pascal Michaillat and Emmanuel Saez. Aggregate demand, idle time, and unemployment. *Quarterly Journal of Economics*, 130(2):507–569, 2015.

José Luis Montiel Olea and Mikkel Plagborg-Møller. Local projection inference is simpler and more robust than you think. *Econometrica*, 89(4):1789–1823, 2021.

Christopher J. Nekarda and Valerie A. Ramey. The cyclical behavior of the price-cost markup. *Journal of Money, Credit and Banking*, 52(S2):319–353, 2020. doi: https://doi.org/10.1111/jmcb.12755. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jmcb.12755.

Asher Wolinsky. True monopolistic competition as a result of imperfect information. *Quarterly Journal of Economics*, 101(3):493–511, 1986. ISSN 00335533, 15314650. URL http://www.jstor.org/stable/1885694.

Online Appendix

Price of Scarcity: Search Theory of Inventories and Markups

Lukasz A. Drozd Marina M. Tavares

Contents

A.1 Proofs **A-2**

A.2 Calibration and solution **A-13**

A.3 Data and robustness **A-18**

A.4 Replication package **A-25**

A.1 Proofs

Time subscripts are omitted when no confusion arises. G and g are the distribution and density of the taste shock η, exponential with mean η_0.

A.1.1 Distributors and search

Lemma 3 (Reservation policy)**.** *Among stationary acceptance rules, sets of effective prices accepted with positive probability, a distributor that maximizes the expected surplus net of search costs uses a reservation effective price: shoppers accept a quote if and only if its effective price is at most $\bar{p}$.*

Proof. Because the distributor's surplus from a match depends on the quote and the taste shock only through the effective price $p = \tilde{p} - \eta P$, we can restrict attention to rules defined on effective prices. With a single quoted price $\tilde{p}^*$, effective prices lie in $(-\infty, \tilde{p}^*]$ and have a continuous distribution with full support there. Let $\mathcal{A}$ be any measurable acceptance set with $\Pr(\mathcal{A}) = 1 - \pi > 0$, and let $I_0 = (-\infty, \bar{p}_0]$ with $\Pr(I_0) = \Pr(\mathcal{A})$; $\bar{p}_0$ exists and is unique because the distribution of effective prices is continuous and strictly increasing on its support. Each draw is accepted with the same probability under both rules, so the number of draws, and hence the expected search cost, is the same. Let $E_0 := \mathcal{A} \setminus I_0$ and $E_2 := I_0 \setminus \mathcal{A}$. Then $\Pr(E_0) = \Pr(E_2)$, every price in E_0 exceeds $\bar{p}_0$, and every price in E_2 is at most $\bar{p}_0$. If $\Pr(E_0) > 0$,

$$\int_{\mathcal{A}} p \, d\Pr(p) - \int_{I_0} p \, d\Pr(p) = \int_{E_0} p \, d\Pr(p) - \int_{E_2} p \, d\Pr(p) > 0,$$

so the expected effective price conditional on acceptance is strictly lower under I_0, and the expected surplus is strictly higher. If $\Pr(E_0) = 0$, the two rules coincide up to a set of measure zero. □

Proof of Lemma 1. Search cost. Each draw is accepted with probability $1-\pi$, independently, so the number of draws until acceptance is geometric with mean $1/(1-\pi)$. All draws occur within an instant, so discounting within the instant vanishes as $dt \to 0$, and the expected cost is $c_0 v/(1-\pi)$.

Surplus. Let $\bar{\eta} := (\tilde{p}^* - \bar{p})/P$ be the taste shock that makes a quote just acceptable, so that $\pi = G(\bar{\eta})$ and $\bar{\eta} = -\eta_0 \log(1-\pi)$. A quote is accepted if and only if $\eta \geq \bar{\eta}$. Changing variables from p to η over the acceptance region $(-\infty, \bar{p}]$,

$$s(\pi) = \mathbb{E}\left[P - p \,\middle|\, p \leq \bar{p}\right] = P - \tilde{p}^* + P\,\mathbb{E}[\eta \,|\, \eta \geq \bar{\eta}] = P - \tilde{p}^* + P(\bar{\eta} + \eta_0),$$

where the last equality uses the memorylessness of the exponential distribution. Substituting $\bar{\eta}$ gives (2), and $\bar{p} = \tilde{p}^* - P\bar{\eta} = \tilde{p}^* + \eta_0 P \log(1-\pi)$. The mean accepted taste is $\mathbb{E}_\pi[\eta] = \bar{\eta} + \eta_0 = \eta_0(1 - \log(1-\pi))$.

Distributor's choice. The net surplus $s(\pi) - c_0 v/(1-\pi)$ has derivative $\eta_0 P/(1-\pi) - c_0 v/(1-\pi)^2$, which is positive for $1 - \pi > c_0 v/(\eta_0 P)$ and negative below, so the optimum is $\pi = \max\{0, 1 - c_0 v/(\eta_0 P)\}$, which is (3) when $\eta_0 P > c_0 v$, that is, when distributors are selective. At the selective optimum, the search cost equals $\eta_0 P$, and the net surplus is $P - \tilde{p}^* - \eta_0 P \log\left(c_0 v/(\eta_0 P)\right)$, which gives (4). □

A.1.2 Steady state and price block

Proof of Lemma 2. Substituting $\tilde{p}^* = X + \eta_0 P$ into (4) gives $f(P) = X + \chi v$ with $f(P) := P\left[1 - \eta_0 + \eta_0 \log(\eta_0 P/c_0 v)\right]$. Optimal precision (3) must lie in $[0, 1)$, so $P \geq c_0 v/\eta_0$. On this domain, $f'(P) = 1 + \eta_0 \log(\eta_0 P/c_0 v) \geq 1$, $f(c_0 v/\eta_0) = (1-\eta_0) c_0 v/\eta_0$, and $f(P) \to \infty$ as $P \to \infty$. Hence, for every $X > \underline{X}$, there is a unique solution $P(X)$, $P'(X) = 1/f'(P(X)) \in (0,1)$, and $P''(X) = -\eta_0 P'(X)^3/P(X) < 0$. Assumption A1 is equivalent to $\underline{X} < X^{ss}$; in the calibration, $\underline{X} < 0$. Then $\tilde{p} = X + \eta_0 P(X)$ has derivative $1 + \eta_0 P'(X) > 0$. □

Lemma 4 (Steady state)**.** *A selective steady state, with $\pi^{ss} \in (0,1)$, exists if and only if Assumption A1 holds, and it is unique.*

Proof. In a steady state, $\dot{M} = \dot{N} = 0$ in (12) gives $aK = \delta K + dN$. If $d > 0$, then $V_0 \leq 0 < \phi v$, and (10) gives $a = 0$, which contradicts $aK = \delta K + dN > 0$; with $\phi = 0$, we rule out creating and liquidating

outposts at the same time. Hence $d = 0$ and $a = \delta > 0$, which requires $V_0 = \phi v$ by (10). Empty outposts are kept, so (11) reduces to (8), which with $\dot{V}_0 = 0$ gives $\max_{\hat{\tau} \le \tau} \hat{\tau}(X - v) = [\zeta_0 + (\rho + \delta)\phi]v > 0$, so $X > v$, $\hat{\tau} = \tau$, and $X = X^{ss}$ in (13). By Lemma 2, the retail price solves $f(P) = X^{ss} + \chi v$ on $P \ge c_0 v/\eta_0$. Since f is strictly increasing with $f(c_0 v/\eta_0) = (1 - \eta_0)c_0 v/\eta_0$, a solution with $P > c_0 v/\eta_0$, that is, with $\pi \in (0, 1)$, exists if and only if $(1 - \eta_0)c_0 v/\eta_0 < X^{ss} + \chi v$, which is Assumption A1, and it is unique. Equation (15) with $\dot{X} = 0$ then gives

$$\Lambda^{ss} = \frac{(\rho + \tau + \delta)X^{ss} - (\tau - \zeta)v}{\eta_0 P^{ss}} = \frac{(\rho + \delta)X^{ss} + \tau(X^{ss} - v) + \zeta v}{\eta_0 P^{ss}} > 0.$$

Given P^{ss} and π^{ss}, retail market clearing determines Q^{ss}, and $M^{ss} = Q^{ss}/\Lambda^{ss}$ and $N^{ss} = (Q^{ss} + \delta M^{ss})/\tau$ follow. The pricing rule, $V_1 = V_0 + X$, and (7) are then satisfied. □

A slot for a second unit. This paragraph checks the one-unit normalization of Section 2.4. Suppose that an outpost holding one unit could rent a slot to restock a second unit at rate τ, at a cost $\sigma\zeta_0 v$ per unit of time while the slot is in use, hold the second unit at an additional cost ζv, and stop using the slot after a sale. Let V_{1t}^+ be the value of a stocked outpost that may use the slot from t on; $V_{1t}^+ \ge V_{1t}$, because the outpost can leave the slot unused. Let $H_t(\omega) := \max_{\tilde{p}} \lambda_t(\tilde{p})(\tilde{p} - \omega)$ be the payoff from selling a unit whose sale costs the outpost ω. By (6), profit rises with the price below $\bar{p}_t$, and above it its derivative has the sign of $\omega + \eta_0 P_t - \tilde{p}$, so the best price is $\max\{\bar{p}_t, \omega + \eta_0 P_t\}$ and

$$H_t(\omega) = \begin{cases} \Lambda_t \eta_0 P_t \exp\big((X_t - \omega)/(\eta_0 P_t)\big) & \text{if } \omega + \eta_0 P_t \ge \bar{p}_t, \\ \Lambda_t(\bar{p}_t - \omega)/(1 - \pi_t) & \text{otherwise;} \end{cases}$$

at $\omega = X_t$, the best price is (9) and $H_t(X_t) = \Lambda_t \eta_0 P_t$. Let V_{2t} be the value of an outpost holding two units, which sells one of them and then stops using the slot, $\rho_t V_{2t} = -\big((1+\sigma)\zeta_0 + 2\zeta\big)v + H_t(V_{2t} - V_{1t}) - \delta V_{2t} + \dot{V}_{2t}$. If outposts holding one unit do not use the slot, subtracting (7) gives the value of the second unit, $B_t := V_{2t} - V_{1t}$:

$$\dot{B}_t = (\rho_t + \delta)B_t + (\sigma\zeta_0 + \zeta)v - H_t(B_t) + \Lambda_t \eta_0 P_t.$$

Using the slot for an instant gains $\Gamma_t := \tau \max\{B_t - v, 0\} - \sigma\zeta_0 v$ per unit of time. If $\Gamma_t \leq 0$ at every date, V_{1t} and V_{2t} solve the Bellman equations of an outpost that may use the slot, so $V_{1t}^+ = V_{1t}$; if $\Gamma_t > 0$ at some date, using the slot for a short time from that date raises the value of a stocked outpost. Hence $V_{1t}^+ = V_{1t}$ at every date if and only if $\Gamma_t \leq 0$ at every date. In a stationary state with $\zeta_0 > 0$, B falls with σ, because $(\rho+\delta)B - H(B)$ rises with B. Let $(\rho+\tau+\delta)B^* = (\tau-\zeta)v + H(B^*) - \Lambda\eta_0 P$. If $B^* \leq v$, then $B \leq v$ at $\sigma = 0$, the slot is unprofitable for every $\sigma \geq 0$, and $\sigma^* := 0$. Otherwise, $\Gamma = 0$ at $\sigma^* := \tau(B^* - v)/(\zeta_0 v) > 0$, and $\Gamma \leq 0$ for $\sigma \geq \sigma^*$. In the calibration with $\zeta = 0$, $\sigma^* = 1.114$ at the steady state, 1.23 at the post-COVID peak utilization of 58.5 percent, and 6.9 at 80 percent, the highest utilization we study (Figure 2B); at each of these thresholds, the best price for the second unit is above $\tilde{p}$. Along a path, B_t solves the equation above backward from its stationary value at the end of the path. With $\sigma = 7$, $\sigma^* \leq \sigma$ in stationary states up to a utilization of 80.1 percent. Along the paths of Figures 3–5, 7, and A1 and of the predicted markups of Sections 4 and 5, on each path that producers expect at the time, $\Gamma_t \leq -0.27\, v$ at every date, so outposts would leave the slot unused.

A.1.3 Proof of Proposition 1

(i) With capacity maintained by replacement and constant sales, $\dot{M} = 0$ in (12) gives $\tau N = Q + \delta M$. Substituting $N = K - M$, $M = (\tau K - Q)/(\tau+\delta)$, so that $\Lambda = Q/M = (\tau+\delta)u/(1-u)$ with $u = Q/(\tau K)$. Then $\Lambda/(\Lambda+\tau+\delta) = u$, which agrees with (19), and $M/Q = 1/\Lambda = (1-u)/[(\tau+\delta)u]$.

(ii) In a stationary state, $\dot{X} = 0$ in (15), so $X - v = \big[\Lambda\eta_0 P - (\rho+\delta+\zeta)v\big]/(\rho+\tau+\delta)$. Since $\Lambda/(\rho+\tau+\delta) = \theta u/(1-u)$, adding $\eta_0 P$ to both sides and using (9) gives (20).

(iii) Let $R(u) := \theta u/(1-u)$, which is increasing and strictly convex on $(0,1)$ and unbounded as $u \to 1$, and let $\hat{X} := (\tau-\zeta)v/(\rho+\tau+\delta)$ be the value of stock without search rents, which is nonnegative because $\zeta \leq \tau$. By (ii) and (9), the stationary value of stock and the price are

$$X = \hat{X} + \eta_0 P\, R(u), \qquad \tilde{p} = \hat{X} + \eta_0 P\big(1 + R(u)\big).$$

For a given P, $\tilde{p}$ is an increasing affine function of $R(u)$, so it is increasing and strictly convex in u.

With $P = P(X)$, substituting X into the price block of Lemma 2 and dividing by $\eta_0 P$ gives

$$R = \log\frac{P}{P_m} + \frac{1}{\eta_0} - 1 - \frac{\hat{A}}{\eta_0 P}, \qquad P_m := \frac{c_0 v}{\eta_0}, \quad \hat{A} := \hat{X} + \chi v,$$

where $\hat{A} \geq 0$ because $\hat{X} \geq 0$ and $\chi \geq 0$. The right side is continuous and strictly increasing in P on (P_m, ∞) and unbounded above. Each R therefore corresponds to at most one retail price with selective distributors, $P(R)$, which rises without bound with R, and, because $P(X)$ is strictly increasing, to at most one stationary value of stock. With $y := \hat{A}/(\eta_0 P) \geq 0$, implicit differentiation gives

$$P_R = \frac{P}{1+y} > 0, \qquad P_{RR} = \frac{P(1+2y)}{(1+y)^3} > 0, \qquad (\log P)_{RR} = \frac{y}{(1+y)^3} \geq 0,$$

so $\tilde{p}_R = \eta_0[P + (1+R)P_R] > 0$ and $\tilde{p}_{RR} = \eta_0[2P_R + (1+R)P_{RR}] > 0$. The price is therefore increasing and strictly convex in R, hence in u, and $\log P$ is increasing and convex in u.

For the logarithm, write $\tilde{p} = \hat{X} + e^{\Psi(u)}$, with $\Psi(u) := \log\eta_0 + \log P + \log\big(1 + R(u)\big)$ and P fixed or $P = P(R(u))$. Since

$$\frac{d}{du}\log\big(1 + R(u)\big) = \frac{\theta}{(1-u)\big(1-(1-\theta)u\big)}$$

is positive and strictly increasing for $0 < \theta < 1$, $\Psi' > 0$ and $\Psi'' > 0$. With $\psi := e^{\Psi}/\tilde{p} \in (0, 1]$, because $\hat{X} \geq 0$,

$$\frac{d^2 \log\tilde{p}}{du^2} = \psi\,\Psi'' + \psi(1-\psi)\,(\Psi')^2 > 0,$$

so the log markup is increasing and strictly convex in u.

Finally, given Λ, let $h(X) := X - \hat{X} - \eta_0 P(X) R$ with $R = \Lambda/(\rho + \tau + \delta)$, so that (15) reads $\dot{X} = F(X, \Lambda)$, with $F(X, \Lambda) := (\rho+\tau+\delta)X - (\tau - \zeta)v - \Lambda\eta_0 P(X) = (\rho+\tau+\delta)h(X)$. Since P is concave (Lemma 2), h is convex on $[\underline{X}, \infty)$. At the lower end, where $P = P_m$, $h(\underline{X}) = \eta_0 P_m(R_m - R)$, where R_m is the right side of the equation for R at $P = P_m$; it is negative when the stationary state has selective distributors. At the stationary value of stock, $h = 0$ and, using $P' = 1/[1 + \eta_0 \log(P/P_m)]$ and the equation for R,

$$h'(X) = 1 - \eta_0 P'(X) R = \frac{\eta_0 + \hat{A}/P}{1 + \eta_0 \log(P/P_m)} > 0.$$

A convex function with a single root and a positive slope there is negative below the root and positive above it. Given Λ, $F(\cdot, \Lambda)$ is therefore negative from $\underline{X}$ up to the stationary value of stock and positive above it: the value of stock rises above the supply curve and falls below it.

(iv) As $(\rho + \delta + \zeta)/\tau \to 0$, $\theta \to 1$, $\theta' \to 0$, and $\hat{X} \to v$, so, at each u, X solves $X = v + \eta_0 P(X) u/(1-u)$, which depends on τ and M/Q only through u, and $\tilde{p} - v = \eta_0 P(X)/(1-u)$. □

A.1.4 Proof of Corollary 1

(i) By Proposition 1(i), $u/(1-u) = \Lambda/(\tau+\delta)$, so $\theta\, u/(1-u) = \Lambda/(\rho+\tau+\delta) = \Lambda \mathbf{T}_0$. Substituting into (20) gives $\tilde{p} - v = \eta_0 P(X)(1 + \Lambda \mathbf{T}_0) - \theta' v$, which, for a given P, is affine in Λ with slope $\eta_0 P\, \mathbf{T}_0 > 0$.

(ii) By Proposition 1(i), $M = (\tau K - Q)/(\tau + \delta)$ and $u = Q/(\tau K)$, so $\Lambda = Q/M = (\tau + \delta)Q/(\tau K - Q) = (\tau + \delta)u/(1-u)$. The function $u/(1-u)$ is increasing and convex on $(0,1)$ and unbounded as $u \to 1$; for a given K, u is proportional to Q, so Λ is increasing and convex in Q and unbounded as $Q \to \tau K$. □

A.1.5 Proof of Corollary 2

(i) By Lemma 4, every steady state has $V_0 = \phi v$, and (8) gives X^{ss} in (13), which depends on v, ζ_0, ϕ, ρ, δ, and τ but not on η_0, c_0, or χ. By (9), $\tilde{p}^{ss} - X^{ss} = \eta_0 P^{ss}$, and (15) with $\dot{X} = 0$ and $\Lambda = Q/M$ gives $(\rho+\tau+\delta)X^{ss} - (\tau-\zeta)v = (Q^{ss}/M^{ss})\, \eta_0 P^{ss}$, which is positive because $X^{ss} > v$. Dividing the first equation by the second gives the result, with a positive slope.

(ii) By Lemma 2, P^{ss} solves $f(P) = X^{ss} + \chi v$. In terms of the search rent $\mathcal{R} := \eta_0 P$, this condition reads $\ell(\mathcal{R}, \eta_0) := \mathcal{R}\left[1/\eta_0 - 1 + \log\left(\mathcal{R}/(c_0 v)\right)\right] - X^{ss} - \chi v = 0$. Distributors are selective, so $\pi^{ss} = 1 - c_0 v/\mathcal{R} \in (0,1)$ by (3), and $\mathcal{R} > c_0 v$. Hence $\partial \ell/\partial \mathcal{R} = 1/\eta_0 + \log\left(\mathcal{R}/(c_0 v)\right) > 0$ and $\partial \ell/\partial \eta_0 = -\mathcal{R}/\eta_0^2 < 0$, and the implicit function theorem gives

$$\frac{d\mathcal{R}}{d\eta_0} = \frac{\mathcal{R}}{\eta_0\left[1 + \eta_0 \log\left(\mathcal{R}/(c_0 v)\right)\right]} > 0.$$

Since X^{ss} does not depend on η_0, $\tilde{p}^{ss} = X^{ss} + \mathcal{R}$ increases with η_0, and so does M^{ss}/Q^{ss} by (i). □

A.1.6 Proof of Corollary 3

(i) By Proposition 1(i), $u = Q/(\tau K)$, so with K given, utilization moves from u to $(1+x)u$. By (20), for a given P,

$$\Delta\tilde{p} = \theta\,\eta_0 P\Big[\frac{(1+x)u}{1-(1+x)u} - \frac{u}{1-u}\Big] = \theta\,\eta_0 P\,\frac{xu}{(1-u)\big(1-(1+x)u\big)}.$$

(ii) By (i), the ratio of $\Delta\tilde{p}$ at x to $|\Delta\tilde{p}|$ at $-x$ is $\big(1-(1-x)u\big)/\big(1-(1+x)u\big)$, which depends only on u; its derivative with respect to u is $2x/\big(1-(1+x)u\big)^2 > 0$. By Proposition 1(i), $M/Q = (1-u)/[(\tau+\delta)u]$, so $u = 1/\big[1+(\tau+\delta)M/Q\big]$. □

A.1.7 Proof of Corollary 4

By Proposition 1(i), market i has $u_i = Q_i/(\tau K)$ and $M_i = (\tau K - Q_i)/(\tau+\delta)$, so that $u_i = u(1+x_i)$, where $u := Q/(\tau K)$ is utilization under even demand.

(i) Since $\int_0^1 Q_i\,di = Q$, $\int_0^1 M_i\,di = (\tau K - Q)/(\tau+\delta)$, and the ratio follows.

(ii) By Corollary 1(ii), $\Lambda_i = (\tau+\delta)Q_i/(\tau K - Q_i)$, a strictly convex function of Q_i on $[0, \tau K)$, and by (i) the aggregate ratio is the same function evaluated at Q. Since $\sup_i Q_i < \tau K$, $\int_0^1 \Lambda_i\,di$ is finite, and Jensen's inequality gives $\int_0^1 \Lambda_i\,di \geq (\tau+\delta)Q/(\tau K - Q)$, with equality only if $Q_i = Q$ for almost all i. By Corollary 1(i), for a given P, $\tilde{p}_i = (1-\theta')v + \eta_0 P(1+\mathbf{T}_0\Lambda_i)$, so $\int_0^1 \tilde{p}_i\,di = (1-\theta')v + \eta_0 P\big(1+\mathbf{T}_0\int_0^1 \Lambda_i\,di\big)$.

(iii) By Proposition 1(iii), the log markup of market i is the same strictly convex function of its utilization $u_i = u(1+x_i)$ in every market, both for a given retail price and when each market's retail price responds to its own value of stock. Since u_i is affine in x_i, $\log(\tilde{p}_i/v)$ is strictly convex in x_i, and Jensen's inequality gives an average log markup above its value under even demand, unless $x_i = 0$ for almost all i. The utilizations at which distributors are selective and outposts restock at the maximal rate form an interval, because the stationary value of stock rises with u, and the log markup is strictly convex on it. A nontrivial mean-preserving spread that keeps every market in this interval raises the average of any function that is strictly convex on it, which gives the last claim.

(iv) Since $Q_i = \tau K u_i$ by Proposition 1(i), aggregate revenue over the production cost of the goods sold is $\int_0^1 u_i\,\tilde{p}(u_i)\,di\big/(vu)$, where $\tilde{p}(\cdot)$ is the stationary price as a function of utilization, the same in every

market, and $u = \int_0^1 u_i \, di$ does not change under a mean-preserving spread. By Proposition 1(iii), $\tilde{p}' > 0$ and $\tilde{p}'' > 0$ on the interval of (iii), so $\big(s\tilde{p}(s)\big)'' = 2\tilde{p}'(s) + s\tilde{p}''(s) > 0$, and the argument of (iii) applies to $s\tilde{p}(s)$. □

A.1.8 Proof of Proposition 2

The level of retail demand D_{0t} adjusts so that sales are $Q_t = \bar{Q} < Q^{ss}$ for $t \geq 0$, the discount rate is constant, and capacity starts at (M^{ss}, N^{ss}), so $\Lambda_0 = \bar{Q}/M^{ss} < \Lambda^{ss}$. We study the limiting economy $\kappa_a \to 0$, in which creation of outposts can be a measure (an atom of new outposts at a date): any value of an empty outpost above its cost would trigger unbounded creation, so $V_{0t} \leq \phi v$ at all $t > 0$, and outposts are created only when $V_{0t} = \phi v$. The shock is small enough that the solution V_{0t} of (8) stays nonnegative (no liquidation), $X_t \geq v$ (restocking at the maximal rate), and $X_t > \underline{X}$ (selective distributors, Lemma 2) along the path. In the calibration, $\underline{X} < v$, so the second condition implies the third, and the first two hold for permanent falls in sales of up to 36.5 percent, on a grid of half a percentage point; $V_{0t} \geq 0$ fails at 37.0 percent, and $X_t \geq v$ is slack. We construct the equilibrium and verify it.

Step 1: quantities during inaction. Conjecture that no outposts are created or destroyed on $[0, T)$. Then the quantity block is (21), whose solution is

$$K_t = K_0 e^{-\delta t}, \qquad N_t = \frac{\bar{Q}}{\tau + \delta} + \left(N_0 - \frac{\bar{Q}}{\tau + \delta}\right) e^{-(\tau+\delta)t}, \qquad M_t = K_t - N_t.$$

Let $M^* := \bar{Q}/\Lambda^{ss}$, the stock at which demand per outpost is at its steady-state value. Since $M_0 = M^{ss} > M^*$, and M_t falls below M^* in finite time because $K_t \to 0$ while $N_t \to \bar{Q}/(\tau + \delta) > 0$, the date $T := \inf\{t \geq 0 : M_t \leq M^*\}$ is in $(0, \infty)$, $\Lambda_t = \bar{Q}/M_t < \Lambda^{ss}$ on $[0, T)$, and $\Lambda_T = \Lambda^{ss}$ with $\dot{M}_T \leq 0$, so that $\tau N_{T^-} \leq \bar{Q} + \delta M^*$.

Step 2: after T. From T on, producers create outposts so as to keep $M_t = M^*$: an atom of new outposts at T raises N from N_{T^-} to $N^* := (\bar{Q} + \delta M^*)/\tau$, and replacement creation $a_t = \delta$ follows. Demand per outpost is Λ^{ss}, and with a constant discount rate the unique bounded solution of (15) is $X_t = X^{ss}$:

writing (15) as $\dot{X} = F(X, \Lambda)$ with

$$F(X, \Lambda) := (\rho + \tau + \delta)X - (\tau - \zeta)v - \Lambda\eta_0 P(X),$$

the proof of Proposition 1(iii) shows that, because $\zeta \leq \tau$ and $\chi \geq 0$, $F(\cdot, \Lambda^{ss})$ is negative from $\underline{X}$ up to X^{ss}, its only root, and positive above it. A solution that starts above X^{ss} therefore rises without bound, and one that starts below falls at a rate bounded away from zero until it leaves the branch on which distributors are selective. The unique bounded solution on that branch is $X_t = X^{ss}$. Then (8) gives $V_{0t} = \phi v$ for $t \geq T$, which is consistent with creating outposts.

Step 3: the value of stock before T. The atom of new outposts at T enters the restocking state, so M_t and Λ_t are continuous, and by (16) so is X_t; hence $X_T = X^{ss}$. Let $\Delta_t := X^{ss} - X_t$. Since $F(X^{ss}, \Lambda^{ss}) = 0$,

$$\dot{\Delta}_t = -F(X_t, \Lambda_t) = \underbrace{\left[F(X^{ss}, \Lambda^{ss}) - F(X_t, \Lambda^{ss})\right]}_{=A_t\Delta_t} + \underbrace{\left[F(X_t, \Lambda^{ss}) - F(X_t, \Lambda_t)\right]}_{=:S_t},$$

where $A_t := \int_0^1 \partial_X F\left(X_t + s\Delta_t, \Lambda^{ss}\right) ds$ is bounded and $S_t = -(\Lambda^{ss} - \Lambda_t)\eta_0 P(X_t) < 0$ on $[0, T)$. Solving backward from $\Delta_T = 0$, and using that the kernel $e^{-\int A}$ is positive,

$$\Delta_t = -\int_t^T e^{-\int_t^s A_l\, dl}\, S_s\, ds > 0 \qquad \text{for } t \in [0, T).$$

Hence $X_t < X^{ss}$ on $[0, T)$, and by Lemma 2 the markup is below its steady-state value. By Step 1 and (19), the inventory-to-sales ratio $1/\Lambda_t$ is above, and utilization is below, their steady-state values on $[0, T)$.

Step 4: no outposts are created before T. With $\hat{\tau} = \tau$, (8) gives, for $t < T$,

$$V_{0t} = \int_t^T e^{-(\rho+\delta)(s-t)}\left[\tau(X_s - v) - \zeta_0 v\right] ds + e^{-(\rho+\delta)(T-t)}\phi v.$$

The steady-state identity $(\rho + \delta)\phi v = \tau(X^{ss} - v) - \zeta_0 v$ implies that the same expression with $X_s = X^{ss}$ equals ϕv. Subtracting,

$$V_{0t} - \phi v = \tau \int_t^T e^{-(\rho+\delta)(s-t)}\left(X_s - X^{ss}\right) ds < 0,$$

so creating no outposts is optimal on $[0, T)$, which verifies the conjecture of Step 1. Liquidating none is optimal because this $V_{0t} \geq 0$, so (11) holds with its second term zero.

The case $\phi = 0$. With no sunk cost and $\kappa_a, \kappa_d \to 0$, creating and liquidating empty outposts is free, so $V_{0t} = 0$ at all t, and outposts are created or liquidated only when $V_{0t} = 0$ (Section 2.4); by hypothesis, producers do not create and liquidate outposts at the same date, and $\zeta_0 > 0$. Start from the steady state of this economy, in which empty outposts are kept and (8) gives $X^{ss} = v(1 + \zeta_0/\tau)$, and let $M^* := \bar{Q}/\Lambda^{ss} < M^{ss}$. Conjecture that liquidation has an atom of size N^{ss} at $t = 0$ and then the density $d\mathcal{L}_t = \bar{Q}\, dt$, so that outposts are liquidated as they empty, until $M_t = M^*$, at a date t^*, and that at t^* an atom of new outposts raises N to $(\bar{Q} + \delta M^*)/\tau$, with replacement creation afterwards. Stocked outposts cannot be liquidated. Until t^*, $N_t = 0$ and $\dot{M}_t = -\bar{Q} - \delta M_t$, so

$$t^* = \frac{1}{\delta} \log \frac{\bar{Q} + \delta M^{ss}}{\bar{Q} + \delta M^*} = \frac{(1 - \bar{Q}/Q^{ss})M^{ss}}{\bar{Q}} + O(\delta),$$

and from t^* on the economy is at its steady state, as in Step 2. Before t^*, $V_0 = 0$ and $X_t = V_{1t}$, so (7) gives

$$\dot{X}_t = (\rho + \delta)X_t + (\zeta_0 + \zeta)v - \Lambda_t \eta_0 P(X_t), \qquad X_{t^*} = X^{ss}.$$

The right side is zero at (X^{ss}, Λ^{ss}), by (7) in the steady state, and it falls with Λ, so the comparison argument of Step 3, with $\Lambda_t = \bar{Q}/M_t < \Lambda^{ss}$ on $[0, t^*)$, gives $X_t < X^{ss}$ there. An empty outpost kept and restocked would then earn $\tau \max\{X_t - v, 0\} - \zeta_0 v < \tau(X^{ss} - v) - \zeta_0 v = 0$ per unit of time, so liquidating it is optimal. In (11), the second term is then positive while $V_{0t} = 0$, so the condition holds on $[0, t^*)$, and from t^* on empty outposts are kept and (8) holds, which verifies the conjecture. Utilization, the inventory-to-sales ratio, and the markup are at their steady-state values from t^* on. □

A.1.9 Proof of Proposition 3

Let $s := \bar{Q}/Q^{ss} - 1$. By (13), the new steady state has the same value of stock X^{ss}, the same value of an empty outpost ϕv, and the same demand per outpost Λ^{ss} as the old one, and stocks of outposts scaled up by $1 + s$. Near it, outposts restock at the maximal rate ($X^{ss} > v$), none are liquidated ($V_0^{ss} = \phi v > 0$), and creation follows the branch $a_t = \delta + (V_{0t} - \phi v)/(2\kappa_a)$ of (10), which we verify below. Let k, m, ξ, and ν

be the deviations of K, M, X, and V_0 from the new steady state. To first order, (12), (15), and (8) give

$$\dot{k} = \gamma\, v, \qquad \dot{m} = \tau k - (\tau+\delta)m, \qquad \dot{\xi} = b\, m + r_X\, \xi, \qquad \dot{v} = -\tau\xi + r_V\, v, \tag{A.1}$$

with $\gamma := K/(2\kappa_a)$, $b := \Lambda\eta_0 P/M$, $r_X := \rho+\tau+\delta-\Lambda\eta_0 P'(X)$, and $r_V := \rho+\delta$, evaluated at the new steady state. Here $b>0$ and $r_V>0$, and $r_X = (\rho+\tau+\delta)h'(X^{ss})>0$, with h the function in the proof of Proposition 1(iii), because $\zeta \le \tau$ and $\chi \ge 0$. The capacity k and the stock m are given at $t=0$; ξ and v are forward-looking.

Roots. The characteristic polynomial of (A.1) is $C(z) = C_0(z) + \gamma b \tau^2$, with $C_0(z) := z(z+\tau+\delta)(z-r_X)(z-r_V)$. The function C_0 vanishes at $-(\tau+\delta)$, 0, r_V, and r_X, and it is negative on $(-(\tau+\delta),0)$ and between r_V and r_X. Let $\bar{C}_1$ and $\bar{C}_2$ be the largest values of $-C_0$ on these two intervals, and $\underline{\kappa} := Kb\tau^2/(2\min\{\bar{C}_1,\bar{C}_2\})$, with $\underline{\kappa}=\infty$ if $r_X = r_V$, so that $\kappa_a > \underline{\kappa}$ implies $r_X \neq r_V$. For $\kappa_a > \underline{\kappa}$, $\gamma b\tau^2 < \min\{\bar{C}_1,\bar{C}_2\}$, so C has two roots $z_f < z_s$ in $(-(\tau+\delta),0)$ and two positive roots between r_V and r_X. The bounded solutions of (A.1) are the combinations of the eigenvectors of z_f and z_s. For a root z, the eigenvector in the order (k,m,ξ,v) is $\big(1, w(z), (r_V-z)z/(\gamma\tau), z/\gamma\big)$, with $w(z) := \tau/(z+\tau+\delta)$.

Initial conditions. To first order, $k_0 = -sK$ and $m_0 = -sM$, so $m_0 = (M/K)\,k_0$, with $M/K = \tau/(\tau+\delta+\Lambda)$. Since $-(\tau+\delta) < z_f < z_s < 0$, $w_f := w(z_f) > w_s := w(z_s) > w(0) = \tau/(\tau+\delta) > M/K$. Because $w_f \neq w_s$, the two eigenvectors project onto linearly independent vectors of the two given stocks, so the initial conditions pick a unique bounded solution. Writing $k_t = \alpha_s e^{z_s t} + \alpha_f e^{z_f t}$, $\alpha_s = k_0(w_f - M/K)/(w_f - w_s) < 0$ and $\alpha_f = k_0(M/K - w_s)/(w_f - w_s) > 0$.

Signs. First,

$$k_t = \frac{k_0}{w_f - w_s}\Big[(w_f - M/K)e^{z_s t} - (w_s - M/K)e^{z_f t}\Big] < 0,$$

because $w_f - M/K > w_s - M/K > 0$ and $e^{z_s t} \ge e^{z_f t}$. Second, $m_t = k_0\big[(w_f - M/K)w_s e^{z_s t} - (w_s - M/K)w_f e^{z_f t}\big]/(w_f - w_s)$, and the bracket is at least $e^{z_f t}\big[(w_f - M/K)w_s - (w_s - M/K)w_f\big] = e^{z_f t}(M/K)(w_f - w_s) > 0$, so $m_t < 0$: the inventory-to-sales ratio $M_t/\bar{Q}$ is below $1/\Lambda^{ss}$ at every date. Third, the bounded solutions of the last two equations of (A.1) are

$$\xi_t = -b\int_t^\infty e^{-r_X(s'-t)} m_{s'}\, ds' > 0, \qquad v_t = \tau\int_t^\infty e^{-r_V(s'-t)}\xi_{s'}\, ds' > 0.$$

By Lemma 2, the markup is above its steady-state value. Since $v_t > 0$, $V_{0t} > \phi v$, which verifies the branch of (10): producers create outposts faster than they depreciate, $\dot{k}_t = \gamma v_t > 0$, and capacity rises toward its new level, which it approaches from below. This proves (i) and (ii).

Rates. The deviations are combinations of $e^{z_f t}$ and $e^{z_s t}$. As $\kappa_a \to \infty$, $\gamma \to 0$, and the roots of C tend to those of C_0: $z_f \to -(\tau + \delta)$ and $z_s \to 0$. The slow root z_s is the largest root of C in $(-(\tau + \delta), 0)$, and $C(0) = \gamma b \tau^2 > 0$, so $C'(z_s) > 0$, and differentiating $C(z_s) = 0$ gives $dz_s/d\gamma = -b\tau^2/C'(z_s) < 0$. As κ_a rises, γ falls and z_s rises toward zero, which proves (iii).

The slow rate. When $|z_s|$ is small relative to $\tau + \delta$ and r_X, replacing $z + \tau + \delta$ by $\tau + \delta$ and $z - r_X$ by $-r_X$ in C_0 gives $(\tau + \delta) r_X z(z - r_V) = \gamma b \tau^2$, whose negative root is $\left[r_V - \sqrt{r_V^2 + 2\beta/\kappa_a}\right]/2$ with $\beta := K b \tau^2/[(\tau + \delta) r_X]$. On the supply curve, the stationary value of stock $\bar{X}(\Lambda)$ solves (15) with $\dot{X} = 0$, so $\bar{X}'(\Lambda) = \eta_0 P/r_X$, and with sales given, $d \log \Lambda / d \log K = -(\Lambda + \tau + \delta)/(\tau + \delta)$ by Proposition 1(i). Hence $\beta = -\tau \, d\bar{X}/d \log K$, the fall in the gain from restocking per unit of log capacity: a 1 percent rise in capacity lowers it by $\beta/100$. □

A.2 Calibration and solution

A.2.1 Steady-state targets

A model period is a month and $v = 1$. The discount rate is $\rho^{ss} = \log(1.10)/12$, a cost of capital of 10 percent a year, and the restocking rate is $\tau = 0.5$, a delivery lag of 60 days (Deloitte Research Center for Energy & Industrials, 2024). We set $\zeta = 0$. Outposts depreciate at the BEA rate for commercial warehouses, 2.22 percent a year, with a service life of 40 years (Bureau of Economic Analysis, n.d.): $\delta = 0.0222/12 = 1.85 \times 10^{-3}$.

The remaining steady-state parameters $(\eta_0, c_0, \chi, \phi, \zeta_0)$ are chosen to match five moments. The steady state is block recursive: given $(\rho, \tau, \delta, \zeta)$, the value of stock is (13), the retail price solves the price block of Lemma 2, precision is (3), and demand per outpost is given in the proof of Lemma 4. The targets and the model's values are:

1. Distribution margin $(P - \tilde{p})/\tilde{p}$: target 0.38, model 0.380. The target cumulates trade and transportation margins on manufactured goods along the production chain in the 1997 input-output tables,

using the share of manufactured goods sold to final users (53 percent) and the share of manufacturing's intermediate inputs that are manufactured goods (59 percent), which imply a recirculation multiplier of 1.38.

2. Gross margin $(\tilde{p} - v)/v$: target 0.49, model 0.490. The target is sales less the cost of goods sold relative to the cost of goods sold in 2019, summed over all the U.S. manufacturing and trade firms in Compustat in 2019, with the exclusions of Section 2.6. In the model, a producer's revenue is its gross output, v is its cost of goods sold per unit, and selling capacity is paid for out of the gross margin, as selling, general, and administrative expenses are in the data.
3. Inventory-to-sales ratio M/Q: target 1.5 months, model 1.500, from the BEA series of Section 2.6.
4. Spending on selling capacity, $(\phi\delta + \zeta_0)vK$, as a share of producers' expenses, $(\phi\delta + \zeta_0)vK + v\tau N$: target 0.228, model 0.228. The target is selling, general, and administrative expenses relative to these expenses plus the cost of goods sold, for the same firms in 2019. Relative to producers' production cost, the model's spending is 0.295. With both targets measured in the same firms, the model's operating margin, $\left[\tilde{p}Q - v\tau N - (\phi\delta + \zeta_0)vK\right]/(\tilde{p}Q)$, with the production expenses $v\tau N$ of the target, which include the goods lost to depreciation, $v\delta M$, is 0.128, against 0.131 for sales less the cost of goods sold and selling, general, and administrative expenses, relative to sales, in the data.
5. Quotes per purchase $1/(1 - \pi)$: target 3, model 3.000.

The implied mean accepted taste shock is $\mathbb{E}_\pi[\eta] = 0.225$, $\Theta = \eta_0 P'(X^{ss}) = 0.096$, and the weights in Proposition 1 are $\theta = 0.984$ and $\theta' = 0.019$. Steady-state utilization of selling capacity is 0.571. The search rent is 0.148 of the price and the user cost of capacity 0.181.

A slot for a second unit. The check of the one-unit normalization in Appendix A.1 sets the cost of the slot at $\sigma = 7$ times the operating cost of an outpost; at that cost, outposts would leave the slot unused on each path we solve.

A.2.2 Solution

Stationary supply curve. Given a sales-to-inventory ratio Λ, the stationary value of stock solves (15) with $\dot{X} = 0$ and the exact price block of Lemma 2, and the markup is $100 \log\big((X + \eta_0 P(X))/v\big)$. Figure 2 and the stationary readings in Section 4 use this solution, with utilization $u = \Lambda/(\Lambda + \tau + \delta)$.

Paths of the sales-to-inventory ratio. Given a path $\{\Lambda_t\}$ that returns to Λ^{ss}, we solve (15) with the exact price block backward from the steady state, using a second-order Runge–Kutta scheme with a step of 0.02 months, holding ρ at its steady-state value, and τ too except in the lead-time counterfactual of Section 4.2, in which τ_t follows the purchasing managers' lead times. In the exercise of Section 4.2, $\Lambda_t = \Lambda^{ss} \exp(s_t/100)$, where s_t is 100 log of the monthly BEA sales-to-inventory ratio relative to its 2019 average, observed through 2026:8; the ratio then returns linearly to its 2019 average over 12 months and stays there. The quarterly markup is total revenue over the production cost of the goods sold in the quarter: each month's revenue is its BEA real sales times the model's price at the month's observation of the ratio. In the exercise of Section 5.1, s_t is the estimated response of Figure 8A for months 0–48, after which it returns linearly to zero over 12 months. Along the computed paths, $X_t > v$, so outposts restock at the maximal rate.

Demand shocks of different sizes and signs. For the exercise of Figure 5, we solve the quantity block (12) forward given the path of demand, with the stock of capacity held at its steady-state level and maintained by replacement, set $\Lambda_t = Q_t/M_t$, and then solve (15) backward as above, with a step of 0.05 months.

Creation at a finite speed. For Sections 3.4 and 4.3, we solve for the paths of stocked and restocking outposts, the value of stock, and the value of an empty outpost, given the path of sales. The four equations are (12) with creation (10), (15), and (8), with the stocks of outposts given at $t = 0$ and the two values at their steady-state levels after 600 months (900 for the transient surge and for Section 4.3). Along these paths $V_0 > 0$, so (11) reduces to (8). We discretize the equations with the trapezoidal rule on a grid of 0.05 months and solve them at all dates at once by Newton's method, so the values at each date are the

discounted sums of rents along the computed path, as perfect foresight requires. When sales change at a point of the grid, each interval of the grid uses its own sales, so that the change takes effect at that point. In the transient surge with an anticipated end, producers know the path of sales from $t = 0$ on. When the end comes as a surprise, the path up to the end is that of the permanent surge, and a second solution starts from the stocks of outposts at that date. The creation rate jumps at $V_0 = \phi v$, which the value of an empty outpost crosses in the transient surge: at month 19 when the end is anticipated, and when the surge ends if the end is a surprise. In that exercise, we replace the jump by a logistic step of width $10^{-4}v$ in V_0 and set the two values after 900 months at the steady state of the smoothed rule. Halving or doubling the width changes the markup by at most 0.006 log points. The solver reproduces the first-order solution of Proposition 3 for small surges, the paths of Figure 5 with capacity held fixed, and the downturn of Figure 3 as $\kappa_a \to 0$. Halving the time step changes the markup by less than 10^{-3} log points in the permanent surge and by at most 0.001 log points in the transient surge. The exercise of Section 4.3 uses the same smoothing. When the rebound is a surprise, the first two solutions hold the sales of April and of May 2020 to the horizon and are used for one month each, and a third starts in June from the stocks of outposts at that date. In both versions, halving or doubling the width changes the markup by at most 0.003 log points; halving the time step changes it by at most 0.001 log points. The model's inventories are compared with BEA's at the end of each month, over the month's sales; with the mean stock over the month instead, the root-mean-square gap of the ratio is 2.0 log points. The model's quarterly markup is total revenue over the production cost of the goods sold in the quarter, $100 \log \left[\int_q \tilde{p}_t Q_t \, dt \big/ \left(v \int_q Q_t \, dt \right) \right]$, relative to its steady-state value: the counterpart of total sales over total COGS.

The size of the surge and κ_a. The two moments are computed from the monthly BEA ratio of real sales to real inventories in manufacturing and trade, in 100 log deviations from its 2019 average: its peak after 2020 (6.8 log points, April 2021) and the months from the peak to the date at which the ratio is half of it, interpolated linearly between months (9.4; the first month at or below half is 10 months after the peak). BEA inventories are stocks at the end of the month, so in the model the ratio in each month is the month's sales over the stock at the end of the month, and the same two statistics are computed from it, along the path of a permanent rise in sales of size $\bar{Q}/Q^{ss} - 1$; the two parameters are solved for

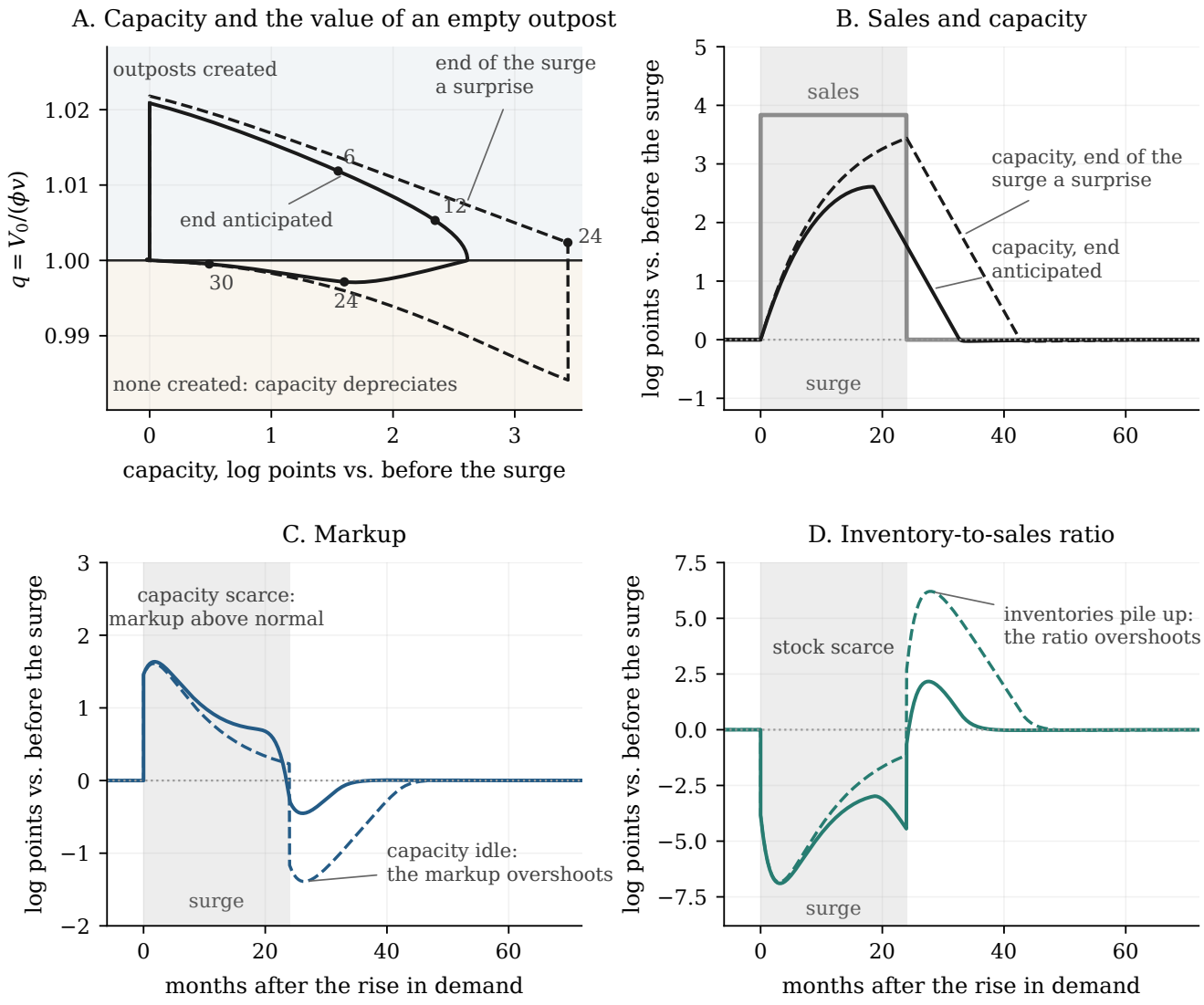

Figure A1: Transient surge in demand

Notes: Calibrated model, with κ_a as in Figure 4. Sales rise by 3.9 percent at month 0, stay there for 24 months (shaded), and then return to their old level for good. Solid lines: producers know from the start when the surge ends. Dashed lines: they expect the rise to be permanent, and the end comes as a surprise. Panel A: $q = V_0/(\phi v)$ against capacity, 100 log K relative to its value before the surge, as in Figure 4A; dots: months after the rise in demand. Panels B–D: sales and capacity, the markup, and the inventory-to-sales ratio, in 100 log deviations from their values before the surge.

jointly. The match gives $\bar{Q}/Q^{ss} - 1 = 3.9$ percent and $1/(2\kappa_a) = 0.0250$. Matching instead the model's continuous path, its peak and the date at which it is half of the peak, to the peak and the first month at or below half in the data gives a surge of 3.9 percent and $\kappa_a = 23.5$; matching the peak of the quarterly averages in Figure 1A, 6.0 log points, gives a surge of 3.4 percent and $\kappa_a = 19.8$.

A.2.3 Transient surge

A transient surge is the flip side of the exercise of Section 4.3: relative to capacity, a surge followed by a downturn instead of a downturn followed by a surge. It combines the cases of Propositions 2 and 3. Figure A1 shows a rise in sales of 3.9 percent that lasts 24 months, in two versions. In the first, producers know from the start when the surge will end. In the second, they expect it to last, and its end comes as a surprise: the path follows Figure 4 until month 24 and then a downturn that starts from the capacity built in the surge. During the surge, capacity is scarce and the markup is above its steady-state value, as in Proposition 3, except in the last 15 days of an anticipated surge, when the value of stock already reflects the idle capacity to come. Producers who see the end coming stop creating outposts at month 19, because outposts created late would soon be idle. They build less capacity, 2.6 log points at its peak against 3.4 when the end comes as a surprise, so stock is scarcer and the markup higher than in the permanent surge: 0.9 against 0.7 log points after a year. When the surge ends, the capacity built in it is idle: producers create no outposts, inventories pile up relative to sales, and the markup falls below its steady-state value until depreciation removes the excess capacity, as in Proposition 2. The markup overshoots, by 0.5 log points when the end is anticipated and by 1.4 log points when it is a surprise, and the inventory-to-sales ratio overshoots by 2.2 and 6.2 log points. The overshoot is smaller when the end is anticipated because less capacity is left to remove: 1.6 log points when the surge ends, against 3.4.

A.3 Data and robustness

A.3.1 Data

Macroeconomic series. Real sales and real inventories of manufacturing and trade industries (BEA, FRED series CMRMTSPL and INVCMRMTSPL, chained 2017 dollars, monthly), the one-year Treasury yield (GS1), and the consumer price index (CPIAUCSL), from FRED; the excess bond premium of Gilchrist and Zakrajšek (2012) as distributed with the replication files of Gertler and Karadi (2015). In the quarterly local projections, quarterly series are averages of monthly values, except inventories, which are end-of-quarter values; Figure 1 and Table 3 use quarterly averages of the monthly sales-to-inventory ratio.

Monetary policy shocks. The FF4 surprise of Gertler and Karadi (2015), 1991:1–2012:6; the orthogonalized surprises of Bauer and Swanson (2023), 1988:2–2023:12; and the pure monetary policy shocks of Jarociński and Karadi (2020), 1990:2–2024:1. The estimation samples with the latter two series end in 2019:12. At quarterly frequency, surprises are summed within the quarter.

Compustat. Compustat North America Quarterly Fundamentals: sales (SALEQ), cost of goods sold (COGSQ), and total inventories, for firms in SIC 20–39 and 50–59 that report quarterly. Each fiscal-quarter report is assigned to the calendar quarter given by DATACQTR, which Compustat sets to the calendar quarter containing most of the fiscal quarter's months. In our sample, 78 percent of firm-quarters end in the last month of a calendar quarter; the rest end one month earlier or later.

Lead times. Average commitment lead times for production materials from the Manufacturing ISM Report On Business (Institute for Supply Management, 2019–2023): 61 days (November 2019), 63 (December 2019), 67 (November 2020), 69 (December 2020), 85 (May 2021), 88 (June 2021), 96 (November 2021), 91 (December 2021), 96 (March 2022), 100 (April, June, and July 2022), 84 (July 2023), and 87 (August 2023); for the exercise of Section 4.2 that reads lead times as restocking times, we interpolate linearly between reports and assume a return to 63 days by the end of 2024.

A.3.2 Sales and inventories after monetary tightenings

Figure A2 reports monthly local projections of real sales, real inventories, and the sales-to-inventory ratio, in which the one-year yield is instrumented by each of the three shock series, with 12 lags of the one-year yield, the log consumer price index, and log real sales and inventories as controls (and of the excess bond premium for the Gertler–Karadi surprises) and Newey–West standard errors. The first-stage F statistics are 18.4 (Gertler–Karadi), 7.9 (Bauer–Swanson), and 6.6 (Jarociński–Karadi). With all three series, sales fall first and inventories reach their trough several months later, and the sales-to-inventory ratio falls within the first year: after six months, by 1.16 log points (standard error 0.45) with the Gertler–Karadi surprises and by 1.42 log points (0.84) with the Bauer–Swanson surprises.

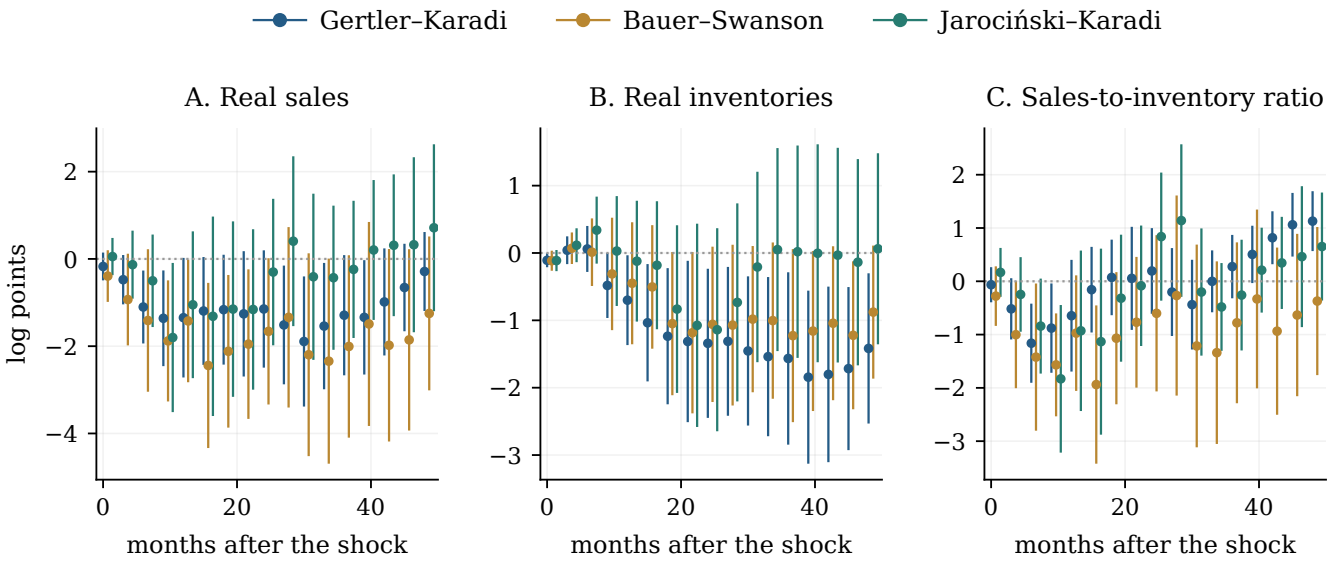

Figure A2: Local projections of sales and inventories

Notes: Local projections with the one-year yield instrumented by each shock series, at selected horizons, with 90 percent intervals; responses scaled to the shock in Gertler and Karadi (2015).

A.3.3 Gross margins after monetary tightenings

Table A1 reports the firm-level local projections of Section 5.1 at selected horizons for four specifications: the baseline (COGS weights, firm fixed effects, firms reporting at both dates), keeping firms that stop reporting at their last reported margin, equal weights (trimming the top and bottom percent of margin changes), and no firm fixed effects. No specification shows a robust response.

A.3.4 Cross-section of firms after COVID

Table A2 reports regressions of the change in the gross margin of each firm between 2019 and 2021 (annual sales over annual cost of goods sold, firms reporting four quarters in both years) on the percent change in its sales-to-inventory ratio (annual sales over average inventories). Column (2) adds the square of the percent change, column (3) replaces the ratio with the log changes in sales and inventories, and column (4) controls for the 2019 level of the inventory-to-sales ratio, which governs exposure to holding gains under first-in, first-out accounting. The relation is strong among manufacturers and weak among trade firms, whose margins are dominated by distribution services. Among manufacturers, the slope rises as firms are averaged within industries, from 0.12 across firms to 0.19 (0.06) across 69 three-digit

Table A1: Firm-level gross margins after a monetary tightening: robustness

Shock	Specification	$h = 0$	$h = 1$	$h = 2$	$h = 4$	$h = 8$	$h = 12$
Gertler–Karadi	Baseline	+0.06	+0.32*	+0.12	+0.10	−0.02	−0.04
		(0.08)	(0.13)	(0.15)	(0.17)	(0.10)	(0.10)
	Exiters kept	+0.06	+0.27*	+0.14	+0.13	−0.02	−0.14
		(0.07)	(0.12)	(0.14)	(0.17)	(0.09)	(0.10)
	Equal weights	−0.04	−0.04	+0.01	+0.01	+0.03	−0.04
		(0.06)	(0.08)	(0.06)	(0.10)	(0.09)	(0.08)
	No fixed effects	+0.05	+0.30*	+0.11	+0.07	−0.06	−0.09
		(0.08)	(0.13)	(0.15)	(0.17)	(0.10)	(0.10)
	First-stage F	19.2					
Bauer–Swanson	Baseline	−0.03	−0.04	+0.28	+0.09	−0.22	−0.26
		(0.08)	(0.15)	(0.19)	(0.21)	(0.17)	(0.29)
	Exiters kept	−0.03	−0.04	+0.25	+0.13	−0.25*	−0.37
		(0.08)	(0.14)	(0.18)	(0.19)	(0.15)	(0.26)
	Equal weights	−0.17*	−0.28*	−0.03	−0.18	−0.16	−0.24*
		(0.09)	(0.14)	(0.10)	(0.18)	(0.11)	(0.11)
	No fixed effects	−0.02	−0.03	+0.29	+0.08	−0.24	−0.25
		(0.08)	(0.16)	(0.18)	(0.20)	(0.17)	(0.29)
	First-stage F	9.7					
Jarociński–Karadi	Baseline	+0.09	+0.19	+0.14	+0.09	+0.01	+0.11
		(0.08)	(0.13)	(0.13)	(0.25)	(0.18)	(0.20)
	Exiters kept	+0.09	+0.18	+0.14	+0.13	+0.04	+0.02
		(0.08)	(0.12)	(0.13)	(0.25)	(0.17)	(0.15)
	Equal weights	−0.09	−0.05	+0.08	−0.04	+0.07	−0.15
		(0.10)	(0.18)	(0.09)	(0.14)	(0.16)	(0.19)
	No fixed effects	+0.07	+0.18	+0.13	+0.06	−0.02	+0.06
		(0.09)	(0.13)	(0.13)	(0.24)	(0.18)	(0.19)
	First-stage F	8.8					

Notes: Response of the gross margin, 100 log(sales/COGS), between quarter $t - 1$ and $t + h$ to a monetary tightening, scaled to the shock in Gertler and Karadi (2015); Driscoll–Kraay standard errors in parentheses; * significant at the 10 percent level.

industries and 0.25 (0.07) across 19 two-digit industries (COGS-weighted means within industries of at least three firms). Errors in measured inventories that are specific to firms bias the slope across firms toward zero and average out within industries.

A.3.5 Industry prices, margins, and sales-to-inventory ratios

The first two rows of Table A3 relate the change in producer prices between 2019 and 2021 or 2022 to the percent change in the sales-to-inventory ratio across industries. Sales and end-of-quarter inventories,

Table A2: Gross margins and inventories across firms, 2019–2021

		(1)	(2)		(3)		(4)
	N	$\Delta(S/I)$	$\Delta(S/I)$	$\Delta(S/I)^2/100$	$\Delta \log S$	$\Delta \log I$	$\Delta(S/I)$
All firms, COGS weights	1247	+0.101	+0.104	−0.010	+0.121	−0.085	+0.103
		(0.028)	(0.030)	(0.034)	(0.031)	(0.028)	(0.027)
All firms, equal weights	1247	+0.076	+0.099	−0.028	+0.123	−0.080	+0.079
		(0.013)	(0.019)	(0.012)	(0.020)	(0.016)	(0.013)
Manufacturing, COGS weights	1016	+0.119	+0.127	−0.034	+0.154	−0.093	+0.122
		(0.034)	(0.035)	(0.047)	(0.036)	(0.034)	(0.032)
Manufacturing, equal weights	1016	+0.082	+0.107	−0.034	+0.135	−0.082	+0.084
		(0.015)	(0.021)	(0.014)	(0.022)	(0.018)	(0.014)
Trade, COGS weights	231	+0.033	+0.020	+0.028	+0.022	−0.032	+0.033
		(0.025)	(0.033)	(0.031)	(0.040)	(0.028)	(0.026)
Trade, equal weights	231	+0.038	+0.019	+0.015	+0.012	−0.049	+0.037
		(0.016)	(0.027)	(0.013)	(0.031)	(0.022)	(0.015)

Notes: Dependent variable: change in the gross margin, 100 log(sales/COGS), between 2019 and 2021, in log points. $\Delta(S/I)$: percent change in the sales-to-inventory ratio; $\Delta \log S$, $\Delta \log I$: log points. U.S. manufacturing and trade firms in Compustat, with the exclusions of Section 2.6; firms with negative margins or changes of more than 50 log points in the margin or 150 log points in the ratio or sales excluded. Heteroskedasticity-robust standard errors in parentheses.

seasonally adjusted, come from the Census survey of manufacturers' shipments, inventories, and orders (M3) for 35 manufacturing industries and from the monthly wholesale trade survey for 16 merchant wholesale industries; prices are the matching producer price indexes of the Bureau of Labor Statistics. The ratio is monthly sales over end-of-quarter inventories, and both the ratio and prices are annual averages of quarterly values. Defense and nondefense industries that share a producer price index are merged; petroleum refineries and petroleum wholesalers are excluded, as in the firm-level sample, and wholesale totals are excluded to avoid overlap. Producer prices of manufacturers also move with costs. The wholesale indexes price the trade margin, the service of reselling, so they measure the margins of wholesalers directly.

The last row uses margins from the national accounts. Drozd and Tavares (2026) measure the margin of each of 341 detailed industries of the Bureau of Economic Analysis (BEA) as the ratio of nominal gross output to variable production costs. This measure captures the part of the output price not accounted for by observed payroll and intermediate-input costs; it is not a measure of net profits. Their inflation decomposition, equation (2) of that paper, states that an industry's price growth is the exact sum of its margin growth, the growth in its unit labor costs (wage growth net of labor productivity), and the cost-share-weighted inflation of its domestic and imported intermediate inputs; margin growth is the exact

residual required to balance the decomposition for each industry and year (their Online Appendix A.1). Input shares come from the 2017 benchmark input-output accounts of the BEA, after redefinitions, and are held at their 2017 values; gross output and its price come from the BEA GDP-by-industry accounts (September 2025 update), and employment and payroll from the Quarterly Census of Employment and Wages. We relate the change in this margin since 2019, the sum of its annual changes, to the percent change in the ratio of annual sales to average inventories (the end of the prior and the current year). Sales and inventories come from the Census Annual Integrated Economic Survey for six-digit NAICS manufacturing industries, which we sum to BEA industries with the BEA concordance, under which each six-digit NAICS code belongs to a single BEA industry. The 225 matched industries account for 21 percent of the 2019 gross output of the 341 industries. The sample is limited to manufacturing for two reasons: for 2019–2021, the Census publishes inventories by detailed industry only for manufacturing and trade, and the national accounts measure the output of trade industries by their margin on goods resold, so their margin is a margin on distribution services.

In every column the relation has the sign the theory predicts. It is significant at the 5 percent level for manufacturers weighted by sales and for the national-accounts margins in every column, and imprecise for wholesalers. The slope for national-accounts margins exceeds the slope across Compustat firms in Section 4, and labor productivity accounts for 0.21 of the slope of 0.43 with size weights and 0.27 of the slope of 0.35 with equal weights: where the ratio rose, real output rose relative to employment, and the decomposition counts the rise in output per worker as a fall in unit labor costs, which raises the margin at a given price.

A.3.6 Plant utilization and shipments

Table A4 compares the utilization of productive capacity with the sales-to-inventory ratio. We match the Census Quarterly Survey of Plant Capacity Utilization to annual manufacturing sales and year-end inventories from Census and to BLS producer price indexes. The survey's full-production utilization rate measures actual output relative to what plants could produce under normal operating conditions. Annual rates are averages of quarterly observations. Some utilization rates have low survey coverage,

Table A3: Prices, margins, and the sales-to-inventory ratio across industries

		2019 to 2021		2019 to 2022	
Industries	N	Size	Equal	Size	Equal
Manufacturing, prices	35	0.37	0.15	0.58	0.23
		(0.17)	(0.08)	(0.19)	(0.15)
Wholesale, trade margins	16	0.66	0.61	0.41	0.59
		(0.36)	(0.47)	(0.35)	(0.43)
Manufacturing (BEA), margins	225	0.43	0.35	0.41	0.33
		(0.05)	(0.05)	(0.09)	(0.09)

Notes: Coefficient of a regression of the change in 100 log of the price (first two rows) or of the margin (last row) on the percent change in the sales-to-inventory ratio across industries. N: number of industries. Size: weighted by 2019 sales (2019 gross output for the last row); Equal: unweighted. Heteroskedasticity-robust standard errors in parentheses. First two rows: Census M3 and monthly wholesale trade survey categories matched to BLS producer price indexes; for wholesale, the index of the trade margin. Last row: BEA manufacturing industries; the margin is the residual of the price decomposition of Drozd and Tavares (2026), summed over the years since 2019, and the ratio is annual sales over average inventories from the Census Annual Integrated Economic Survey.

which can weaken their estimated association with prices. For industry groups spanning several NAICS industries we sum sales and inventories and weight component price growth by 2019 sales.

Table A4: Producer-price growth, sales per unit of inventory, and plant utilization, 2019–21

Specification	Groups	Sales/inventories	Plant utilization
2019 sales weights	85	0.431	−0.132
		(0.099)	(0.097)
Equal weights	85	0.270	−0.180
		(0.108)	(0.103)
Excluding wood and primary metals	78	0.262	−0.083
		(0.097)	(0.093)

Notes: Each row reports a joint cross-sectional regression of the change in 100 log(PPI) on the percent change in nominal annual sales divided by year-end inventories and the percentage-point change in full-production plant utilization. Parentheses contain heteroskedasticity-robust standard errors. Sales weights use 2019 sales. All samples exclude petroleum and coal products (NAICS 324). The last row uses sales weights and also excludes wood products (321) and primary metals (331).

For the dispersion exercise in Section 4.2, we deflate annual shipments in each Census M3 manufacturing industry by its producer price index. We divide each industry's gross growth factor by the initial-shipment-weighted mean across industries. The weighted mean of the resulting net growth rates is zero. We apply each rate to the baseline sales of a market with the calibrated supply curve and compute the initial-shipment-weighted mean change in log markups. All markets stay within the utilization range of Figure 6 in the three windows reported.

A.3.7 Recessions

Table A5 reports the three recessions before the pandemic, at the quarter in which the sales-to-inventory ratio reached its trough. The reference years are 1989, 2000, and 2007, the last full years of the preceding expansions.

Table A5: Sales-to-inventory ratio, predicted markup, and gross margins in recessions

		Data	Model	Data: gross margin		
	Trough of	Sales-to-inventory	Predicted		Within firms	
Recession	the ratio	ratio	markup	Σ sales / Σ COGS	Mean	Median
1990–91	1991Q1	−5.8	−0.91	−0.34	−2.27	−0.60
2001	2001Q2	−2.9	−0.52	−4.10	−3.06	−1.35
2007–09	2009Q1	−10.8	−2.26	−0.48	−1.61	−0.90

Notes: Sales-to-inventory ratio: BEA real sales over real inventories of manufacturing and trade, quarterly mean of 100 log relative to its mean in the reference year. Predicted markup: the markup that the calibrated model predicts given the monthly path of the ratio, with the reference-year ratio as the steady state, as in Section 4.2. Gross margin: U.S. manufacturing and trade firms in Compustat, with the exclusions of Section 2.6, relative to the same quarter of the reference year: 100 log of total sales over total cost of goods sold (COGS) of the firms present in every quarter of the five-year window that starts with the reference year (70 to 89 percent of the reference year's COGS); and the within-firm change in 100 log(sales/COGS), mean weighted by the reference-year COGS, and median. All entries in log points.

A.4 Replication package

The replication package is a Python project that starts from raw data and produces every number, figure, and table in the paper. Estimates and model results quoted in the text are written by the code to `paper/tables/numbers*.tex` and read by LaTeX as macros. The lead times in Section 4.1 are typed from the ISM reports listed in `code/ism_lead_times.py`.

Running the code. The package's `README.md` lists the source series, the Compustat extracts and the environment variables that locate them, and the scripts in the order they run; `code/run_all.py` runs them and compiles both formats of the paper.

Data not distributed. Compustat is licensed; the package documents the variables and filters of the extract. The FRED series, the shock series, the Census and BLS industry data, and the BEA and QCEW data behind the national-accounts margins are public.