# The Value of Slack: Capacity Utilization in Multi-Product Firms*

Nuno Clara Sampreet Singh Goraya S. Lakshmi Naaraayanan

August 2026

**Abstract**

Why do firms hold slack capacity? Using product-level data on installed capacity and physical output for manufacturing firms, we show that slack is concentrated in the product lines facing the most volatile demand. Firms also commit less capacity to their more volatile product lines, and this allocation responds to product-specific risk but not to firm-wide risk. We develop a dynamic model in which a multi-product firm both accumulates capacity and allocates it across product lines under demand uncertainty. Shifting capacity toward lines whose demand is currently less volatile serves as internal insurance against demand risk. In a series of counterfactuals, we show that (a) absent this internal insurance, firms are substantially smaller and less profitable; (b) because slack is chosen rather than imposed, within-firm dispersion in capacity utilization and in returns to capital can reflect demand-risk management rather than misallocation; and (c) that the gap between average and marginal $q$ narrows, rather than widens, as slack increases.

*Keywords*: Slack, utilization, Tobin's $q$, multi-product firms
*JEL Classification*: D25, G31, G11, G50

*Clara: Duke University (Email: nuno.clara@duke.edu); Goraya: Department of Economics, Stockholm School of Economics (Email: sampreet.goraya@hhs.se ); Naaraayanan: London Business School and CEPR (Email: lnaaraayanan@london.edu). For helpful comments and suggestions, we thank seminar participants. We are grateful to London Business School's Research and Materials Development Grant and Wheeler Institute for Business and Development for supporting this research.

# 1 Introduction

Firms in modern economies routinely operate well below full capacity. Among U.S. firms, average *firm-level* capacity utilization ranges between 68-78% (Walker et al., 2024; Grullon and Ikenberry, 2025). We document a similar pattern in India, where average *product-level* utilization is around 57%, with a typical range of 49–66% across products within a firm. More broadly, underutilization is pervasive across countries, raising a natural question: why do firms maintain so much unused capacity? Existing explanations typically interpret idle capital as a symptom of underlying frictions, viewing it as evidence of weak demand, rigidities that prevent capacity from being redeployed, or misallocation of resources across producers. This paper argues that unused capacity can instead reflect firms' optimal choices. When firms must commit capacity to a product line before demand is realized, they have an incentive to build in slack— capacity that remains unused in most states of the world but becomes valuable when demand is high.

Product-level data allow us to test a sharp prediction of this mechanism. If slack reflects weak demand or stranded capital, it should be concentrated in products with low expected demand. If it reflects optimal choice to counter demand uncertainty, it should instead be concentrated in products with greater demand risk. We study this mechanism using novel product-level data on installed capacity and physical output from manufacturing firms, together with a structural model of a multi-product firm that allocates capacity before demand is realized. In both the data and the model, firms hold slack where demand risk is greatest and reallocate capacity toward less volatile product lines. Firms thus manage demand risk through the internal allocation of capacity, even in the absence of financing or agency frictions. Therefore, viewing slack as an optimal choice shifts how we interpret standard measures of dispersion in revenue productivity: within-firm differences in capacity utilization and returns to capital need not indicate allocative inefficiency.

The main empirical challenge is that utilization is rarely observed at the level at which firms make capacity decisions. Most datasets report aggregate utilization or survey-based firm-level measures, even though capacity is typically chosen product by product. We overcome this limitation using detailed product-level data on Indian firms, made available through disclosure requirements, that account for more than 70% of the country's industrial output (Goldberg et al., 2010; Bau and Matray, 2023). Firms report installed capacity and physical production separately for each product they manufacture, allowing us to measure utilization—physical output relative to installed capacity—at the firm-product-year level.[1]

[1]This definition is consistent with the existing literature, which has relied on survey-based measures and defines firm-level capacity utilization as the ratio of actual output to sustainable maximum output

We measure demand volatility by estimating product-level demand curves using marginal cost as a supply-side instrument for price, following Foster et al. (2008). We then decompose the resulting demand shocks into a firm-wide component, which moves all of a firm's products together, and a product-specific component, which does not. This distinction is central to the paper: common and idiosyncratic demand risk have sharply different implications for how firms allocate capacity.

The empirical analysis yields four facts. First, dispersion in utilization is largely a within-firm phenomenon: more than half of the variation occurs across product lines within the same firm. Second, the within-firm slack is concentrated in more volatile products. Utilization falls with product-level demand volatility, and the relationship is unchanged when identified entirely from comparisons across products within the same firm and year. Third, capacity allocation itself responds to volatility: holding average demand fixed, firms allocate less capacity to more volatile product lines.

The fourth fact identifies the margin along which this reallocation operates. Utilization falls with exposure to both firm-wide and product-specific demand volatility, but capacity shares fall only with product-specific volatility. Firm-wide shocks move all of a firm's products together and therefore cannot be absorbed by shifting capacity across them while product-specific shocks can. Firms allocate capacity precisely along this insurable margin, providing direct evidence that internal capacity reallocation is used to manage demand volatility.

A model in which firms commit capacity before demand is realized rationalizes these empirical patterns and provides a disciplined framework for interpreting their quantitative implications. A firm accumulates capital subject to time-to-build and adjustment costs; installed capital determines total capacity, which the firm allocates across product lines *before* demand is realized and utilizes *after* demand is observed. Because demand is downward sloping and utilization costs are upward sloping, profit is concave in demand, so volatility reduces expected utilization. Slack is therefore the optimal response to demand volatility.[2]

The allocation rule delivers the fourth fact, that utilization declines with both firm-wide and product-specific demand volatility, while capacity shares decline only with product-specific volatility. This is because a firm-wide shock raises the expected profitability of every product line in the same proportion and therefore cancels out in relative capacity shares. As a result, these shares respond only to product-specific

(capacity) (see, e.g., Corrado and Mattey, 1997; Kim and Winston, 1974; Walker et al., 2024).

[2]An alternative interpretation is that product-level capacity reflects the allocation of shared production assets rather than dedicated production lines. Such a mechanism could generate persistent differences in measured utilization across products. Our framework is consistent with this interpretation in the sense that it treats capacity as an internal resource that can be allocated across products. The model's contribution is to show how this allocation responds systematically to differences in demand volatility and to quantify the value of such operational flexibility.

volatility. Reallocation acts as the firm's *internal insurance*. By redirecting committed capacity each period toward lines with stronger realized demand, the firm mitigates the impact of volatility on utilization and cash flows. A central implication is that firm-level volatility is itself an equilibrium outcome of internal capacity allocation rather than a primitive characteristic of the firm. Two firms with identical aggregate demand volatility may organize production very differently depending on how demand volatility is distributed across their product lines. Product-level data are therefore essential for distinguishing firms that face similar aggregate uncertainty but very different internal allocation problems.

We estimate the model by simulated method of moments. The utilization cost curve is disciplined by the level and within-firm dispersion of utilization, and the demand process by the variances, cross-product covariance, and persistence of product output. Crucially, no volatility response is targeted: how utilization and the capacity allocation react to demand volatility is left out of the estimation, and the model reproduces both untargeted patterns—utilization falling with either component of demand shocks, and capacity shares responding to idiosyncratic shocks but not at all to firm-wide shocks.

The calibrated model answers three questions. First, what is the value of the reallocation margin? Freezing the capacity plan at the time-invariant allocation rule reduces long-run profits by 3.1% and the capital stock by 2.0%. A firm with less flexibility to allocate capacity is not only less profitable but also smaller, because lower expected profits per unit of capacity reduce the returns to investment.

Second, does the composition of risk matter, holding its total fixed? We construct firms with identical firm-level demand volatility that differ only in how that volatility is split between common and product-specific components. Firms whose risk is mostly product-specific actively rotate capacity, run their lines at very different rates, and have smoother profit profiles; firms whose risk is mostly firm-wide have nothing to rotate against, so the shock passes directly through to profits. Firm-level data cannot distinguish these two types of firms by looking only at total demand volatility. Yet the firm whose risk is mostly product-specific exhibits 2.41 times more within-firm utilization dispersion, reallocates capacity 3.05 times more actively, and has 8.9% lower firm-level profit volatility.

In the model, a product line's revenue productivity is pinned down by its utilization rate. Consequently, within-firm dispersion in revenue productivity is driven by the firm's active management of demand risk through capacity reallocation. Viewed through this lens, observed within-firm dispersion in utilization need not be interpreted as evidence of misallocation. Instead, it may partly reflect an efficient response to heterogeneous demand risk, as the firm deliberately reallocates capacity across

product lines to insure against demand fluctuations.

Third, what does chosen slack imply for measurement of Tobin's $q$? Our simulated data reproduce the empirical finding that asset utilization improves the investment–$q$ relationship, even though the model contains none of the rigidities that motivate such corrections. The regression cannot distinguish slack that binds from slack that insures. The model offers a diagnostic that can. When slack is chosen, the gap between average and marginal $q$ *narrows* as volatility rises and utilization falls, the opposite of the widening predicted when idle capacity reflects rationed demand. It also delivers an aggregation result. At the optimal utilization policy, firm-level asset utilization is a near-sufficient statistic for investment opportunities.

**Related Literature.** This paper contributes to the literature on internal resource allocation within firms. A large literature studies how firms allocate resources across divisions, plants, and business units, and how these allocation decisions affect investment, productivity, and firm value (Stein, 1997; Scharfstein and Stein, 2000; Maksimovic and Phillips, 2002; Matvos and Seru, 2014; Giroud and Mueller, 2015). While much of this literature focuses on the allocation of financial capital, we study the allocation of productive capacity across product lines. We show that demand uncertainty is a fundamental determinant of this allocation: firms optimally commit relatively less capacity to products facing more volatile demand, implying lower equilibrium utilization. Unlike the internal capital markets literature, our mechanism does not rely on financing frictions or agency problems. Instead, heterogeneous demand risk alone generates systematic differences in the optimal allocation of productive capacity across products.[3]

More broadly, our findings speak to why firms organize multiple product lines under common ownership. A central insight of the theory of the firm is that common ownership allows the coordination and reallocation of resources that market transactions may not achieve (Coase, 1937; Grossman and Hart, 1986; Hart and Moore, 1990). The literature on multi-product firms, in turn, studies how firms choose and adjust the set of products they produce (Bernard et al., 2010a; Eckel and Neary, 2010; Goldberg et al., 2010; Mayer et al., 2014; Hottman et al., 2016; De Loecker et al., 2016; Argente et al., 2024). Our results identify productive capacity as another internally allocated

[3]In doing so, our results also contribute to the literature on corporate risk management. Existing work explains firms' hedging decisions through costly external finance, cash-flow volatility, and the interaction between financial and operational hedging (Froot et al., 1993; Almeida et al., 2017). In our framework firms are risk neutral, yet demand uncertainty still affects capacity allocation because expected profits are concave in demand. Firms therefore respond to product-level demand risk by adjusting how much capacity they commit to each product line. This differs from traditional financial hedging and from coinsurance-based theories of corporate diversification (Lewellen, 1971; Hann et al., 2013), which emphasize the covariance of cash flows across businesses; our mechanism operates through the allocation of productive resources within the firm.

resource. By showing that firms systematically allocate capacity according to product-level demand uncertainty, we highlight an additional benefit of organizing multiple product lines within a single firm: internal capacity allocation provides operational flexibility that allows firms to accommodate heterogeneous demand fluctuations.

Our results also qualify the interpretation of dispersion in the revenue product of capital. A large literature interprets such dispersion as evidence of capital misallocation (Hsieh and Klenow, 2009), though endogenous utilization complicates that inference at the firm level (Kabir et al., 2024). Kehrig and Vincent (2026) move inside the firm and trace differences in the returns to capital across plants to financing constraints. We show that substantial dispersion in capital productivity also arises within firms, across their products, and that it can reflect an optimal response to product-level demand risk rather than a friction in capital allocation. Therefore, observed differences in capital productivity need not identify misallocation when firms endogenously vary capacity and utilization across products.

The paper further contributes to the literature evaluating Tobin's $q$ as a measure of investment opportunities. Since Hayashi (1982), empirical applications of $q$-theory have relied on average $q$ as a proxy for the marginal $q$ that governs investment. Subsequent work identifies several reasons why this proxy may fail: measurement error (Erickson and Whited, 2000), the present-value construction of marginal $q$ (Abel and Blanchard, 1986), omitted intangible capital (Peters and Taylor, 2017), information contained in bond markets (Philippon, 2009), rents (Crouzet and Eberly, 2023), and excess capacity (Grullon and Ikenberry, 2025), which builds on demand-rationing models of investment (Precious, 1985; Licandro, 1992). We ask how the relationship between average and marginal $q$ changes when excess capacity is an optimal choice rather than a distortion. Our model delivers an aggregation result for multi-product firms: firm-level asset utilization and average $q$ together provide near-sufficient statistics for investment.

Finally, we contribute to the literature on capacity utilization under demand uncertainty. The idea that installed capacity provides flexibility—firms choose it before demand is realized but adjust utilization afterward—has a long history (Stigler, 1939; Hart, 1940). Variable utilization is also a central margin in macroeconomics (Berndt and Morrison, 1981; Greenwood et al., 1988; Basu and Kimball, 1996). A related literature, beginning with Arrow et al. (1951), studies capacity choice in newsvendor-style environments and has recently brought this framework to firm-level data (Butters, 2019, 2020; Amberg et al., 2026). Relative to the uncertainty–investment literature (Bloom, 2009), we focus not on how uncertainty affects the accumulation of capacity, but on how firms allocate a given capacity across products. We construct, to our knowledge, the first product-level measures of capacity utilization for multi-product

firms and show that slack is concentrated in the product lines that bear the most demand risk.

This evidence also bears on the sources of slack in developing economies. Existing work attributes slack among small firms to input indivisibilities, thin local demand, and missing complementary inputs such as skilled labor (Walker et al., 2024; Bassi et al., 2024). We instead study large manufacturing firms that account for the bulk of India's industrial output. Slack remains prevalent among these firms, but it is systematically concentrated in products facing greater demand risk—a pattern these alternative mechanisms do not readily explain.

**Outline.** The rest of the article is organized as follows. Section 2 describes the data and measurement. Section 3 documents the empirical facts. Section 4 develops the theoretical framework, lays out the estimation strategy, and presents the counterfactual results. Section 5 concludes.

## 2 Data

The following paragraphs detail the source and construction of each dataset.

**Firm Financials.** We obtain firm- and product-level data from Prowess, a database compiled by the Centre for Monitoring the Indian Economy (CMIE).[4] We use the most recent vintage of the database, which addresses the systematic survivorship bias present in earlier versions, as discussed by Siegel and Choudhury (2012).

CMIE collects information from balance sheets and income statements for roughly 37,000 publicly listed and private firms. Our sample spans the fiscal years 1992 to 2016; throughout, "year" refers to the fiscal year, which in India runs from April 1 to March 31. Firms in Prowess account for more than 70% of industrial output, 75% of corporate taxes, and over 95% of excise taxes collected by the Government of India, and primarily represent large and medium-sized firms in the economy (Bau and Matray, 2023).

**Product-level Capacity Utilization.** To examine within-firm capacity utilization, we use detailed product-level information that firms disclose under the reporting requirements of the Companies Act of 1956 and the Companies Act of 2013. The data report

[4]Prowess has been widely used in studies of Indian firms, including Bertrand et al. (2002); Gopalan et al. (2007); Lilienfeld-Toal et al. (2012); Gopalan et al. (2016); Naaraayanan and Nielsen (2021); and Naaraayanan and Wolfenzon (2024).

total sales and total quantity sold at the firm-product level, which allows us to construct unit prices and quantities. Importantly, the data also contain information on installed capacity, actual production, and sales reported in company annual reports; see Goldberg et al. (2010) and Bau and Matray (2023) for further details.[5]

Our final dataset contains approximately 2,300 distinct manufactured products, where product definitions are held fixed over time. We use Prowess's internal product classification, which is based on the National Industrial Classification (NIC). With these definitions, we construct a panel of product-level output and prices, defining the unit price of each product as total sales divided by total quantity sold.[6]

For each product, we define *Utilization* as the ratio of actual output to installed capacity at the firm-product-year level. Both quantities come from firms' own disclosures in a standardized format rather than from a survey instrument. This matters because survey-based utilization measures have long been criticized on measurement grounds. Respondents self-define "practical capacity," and the aggregate series are heavily smoothed, so measured utilization is close to output relative to trend (Shapiro, 1989; Corrado and Mattey, 1997). Our measure differs on both counts, as installed capacity is a disclosed physical quantity for each product and utilization is realized physical production relative to that quantity, product by product. It remains a *capacity*-utilization concept, output relative to installed capability, rather than a *capital*-utilization concept such as machine hours (Gorodnichenko and Shapiro, 2011). To the extent that reported installed capacity embeds a firm's normal-operations assumptions, levels are measured relative to that norm. Our empirical findings rely on comparisons across products within the same firm, which difference out reporting conventions applied uniformly across a firm's products. From now onward, with a slight abuse of terminology, we use capacity and capacity utilization interchangeably.

Table 1 reports descriptive statistics for the final estimation sample. The data reveal substantial heterogeneity across firms. Firm size, either measured by assets or sales, is highly right-skewed with the sample containing a small number of very large firms. The average firm has a leverage ratio of 0.24 and produces about three distinct products. Export participation, however, is highly concentrated. The median firm earns only 0.70% of revenues from exports, while firms at the 90th percentile earn more than 42% of revenues from foreign sales.

[5]The Companies Act does not require firms to report products in standardized units. As a result, units may vary across firms and within firms over time. We therefore standardize units within and across firms and drop observations when the available information is insufficient to reconcile changes in unit types within a firm-product over time.

[6]Figure IA.A.1 and Figure IA.A.2 report the distribution of products and the share of total manufacturing output accounted for by each two-digit manufacturing sector. Basic metals, vehicles, and chemicals account for the largest shares of value, while furniture and other wood products account for the smallest. These patterns are consistent with the product distributions documented for samples of multi-product firms in the United States (Bernard et al., 2010b) and India (De Loecker et al., 2016).

Panel B reports firm-product-level characteristics. Capacity utilization averages 0.57, but dispersion is large, with a standard deviation of 0.37. At the 10th percentile, utilization is only 0.07, indicating substantial slack in a nontrivial share of product lines. Revenue shares are also highly concentrated within firms. The median product accounts for 58% of firm revenues, while the 75th percentile product accounts for 99%. By the 90th percentile, the product effectively accounts for all firm revenues. Thus, although multi-product firms are common in the data, firm performance is often dominated by a single core product, with remaining product lines accounting for relatively small revenue shares.

TABLE 1: Summary statistics

| Variable | N | Mean | SD | p10 | p25 | p50 | p75 | p90 |
|---|---|---|---|---|---|---|---|---|
| *Panel A. Firm characteristics* | | | | | | | | |
| Assets | 34,856 | 3,675.68 | 24,053.50 | 68.60 | 149.50 | 421.70 | 1,463.70 | 5,173.60 |
| Sales | 23,482 | 2,703.29 | 12,776.06 | 50.20 | 148.10 | 447.55 | 1,496.90 | 4,852.40 |
| Leverage | 31,448 | 0.24 | 0.38 | 0.04 | 0.10 | 0.19 | 0.31 | 0.45 |
| Exporting intensity | 34,853 | 11.85 | 22.84 | 0.00 | 0.00 | 0.70 | 11.66 | 42.37 |
| Number of product lines | 34,860 | 3.08 | 2.53 | 1.00 | 1.00 | 2.00 | 4.00 | 6.00 |
| Investment/assets | 20,855 | -0.10 | 0.18 | -0.23 | -0.13 | -0.05 | -0.02 | -0.01 |
| Market-to-book | 5,510 | 0.80 | 1.15 | 0.08 | 0.17 | 0.39 | 0.94 | 2.05 |
| *Panel B. Firm-product characteristics* | | | | | | | | |
| Capacity utilization | 66,931 | 0.57 | 0.37 | 0.07 | 0.26 | 0.56 | 0.83 | 1.05 |
| Product-level sales | 60,706 | 1,313.25 | 11,387.42 | 7.20 | 40.80 | 180.30 | 642.90 | 2,020.50 |
| Proportion of firm sales | 60,637 | 0.54 | 0.40 | 0.01 | 0.10 | 0.58 | 0.99 | 1.00 |

Notes: This table summarizes the sample used for the analysis. The sample keeps product-year observations with nonmissing capacity utilization, product-level sales, proportion of firm sales, and the firm characteristics listed in the table. The time period spans the fiscal years between 1995 to 2016. In our analysis, "year" refers to the fiscal year as opposed to the calendar year because the financial year in India runs from April 1 to March 31. Product characteristics are measured at the product-year level, so they have the same number of observations. Firm characteristics are measured after reducing this same product-year subsample to one observation per firm-year. Capacity utilization is winsorized at the 5th and 95th percentiles, consistent with the main analysis. Data source: CMIE Prowess.

**Validation.** As a validation exercise, we compare the annual capacity-utilization series in our data with two external benchmarks for India. In panel A of Figure IA.A.3, the first external benchmark is the Reserve Bank of India OBICUS series, which reports quarterly capacity utilization for Indian manufacturing firms; we aggregate that quarterly series to the annual level. In panel B of Figure IA.A.3, the second external benchmark is the OECD capacity-utilization tendency series for India, available through FRED as 'BSCUJG02INQ160S'; we likewise convert the quarterly observations to annual averages. To place the three series on a common scale, each series is standardized over the common sample period before plotting. The figure shows that the movements in our data line up closely with the external aggregate series, which supports the use of the product-based utilization measure as a proxy for economy-wide capacity utilization.

**Quantity Productivity, Markups, and Marginal Costs.** For our analysis, we estimate quantity-based productivity following De Loecker, Goldberg, Khandelwal, and Pavcnik (2016), which allows us to separate physical productivity from markups, marginal costs, and quality differences. The estimation uses single-product firms to recover product-specific production function parameters and then applies these estimates to multi-product firms under standard assumptions about common product technologies and selection into product scope. We account for unobserved quality differences by using input price variation, modeled as a function of output prices, product market shares, product fixed effects, location fixed effects, and export status. This procedure yields firm-product-year measures of quantity productivity, markups, and marginal costs that are adjusted for product quality and selection into multi-product production. See Appendix IA.B for further details.

**Capacity Allocation.** We do not observe capacity allocation and therefore proxy the capacity share with the material cost share of each product. Material cost shares are recovered from the production function estimates. Specifically, we use the estimated revenue function to infer the share of product revenue attributable to material inputs, yielding product-year material cost shares used throughout the analysis. Appendix IA.B provides the derivation and estimation details.

**Product-level Demand Shocks.** We construct product-level demand shocks using an instrumental variables approach that addresses the endogeneity of prices. Our identification strategy exploits variation in marginal costs as an instrument for prices, following Foster, Haltiwanger, and Krizan (2008). Throughout, $i$ indexes products and $j$ indexes firms, consistent with the model in Section 4. Specifically, we estimate the following demand equation using two-stage least squares:

$$\ln Q_{ijt} = \eta \ln P_{ijt} + \alpha_{ij} + \delta_{it} + \varepsilon^{D}_{ijt}, \tag{1}$$

where $Q_{ijt}$ is the quantity produced of product $i$ by firm $j$ in year $t$, $P_{ijt}$ is the product price, $\alpha_{ij}$ are firm-product fixed effects, and $\delta_{it}$ are product-year fixed effects. We instrument for $\ln P_{ijt}$ using log marginal cost $\ln MC_{ijt}$, which shifts supply but is uncorrelated with demand shocks. The demand residual is the estimated shock in equation (1),

$$\hat{d}_{ijt} = \ln Q_{ijt} - \hat{\eta} \ln P_{ijt} - \hat{\alpha}_{ij} - \hat{\delta}_{it}, \tag{2}$$

where $\hat{\eta}$ is the IV estimate of the price elasticity. Subtracting the estimated fixed effects from the demand shocks matters for the volatility measures below because the product-year effects $\delta_{it}$ co-move across a firm's products and would otherwise be at-

tributed to firm-wide risk; it also aligns the risk measures with our regressions, which absorb product-year effects on the outcome side.

We measure each product's total demand volatility as the time-series standard deviation of the residual,

$$\hat{s}_{D,ij} = \mathrm{sd}_t\left(\hat{d}_{ijt}\right), \tag{3}$$

computed for any product with at least two years of data. This captures the unpredictable component of demand facing each product line, and it is the volatility measure used in Facts 2 and 3.

To separate firm-wide from product-specific demand risk, we use second moments rather than realizations. Because the product-specific component is independent across products within a firm, the time-series covariance between two products of the same firm isolates the variance of the firm-wide component,

$$\hat{s}^2_{f,j} = \frac{1}{|\mathcal{P}_j|} \sum_{(i,k)\in\mathcal{P}_j} \widehat{\mathrm{Cov}}_t\left(\hat{d}_{ijt}, \hat{d}_{kjt}\right), \tag{4}$$

where $\mathcal{P}_j$ is the set of product pairs of firm $j$ with at least five overlapping years of data. The product-specific variance is the remainder,

$$\hat{s}^2_{\varepsilon,ij} = \widehat{\mathrm{Var}}_t\left(\hat{d}_{ijt}\right) - \hat{s}^2_{f,j}, \tag{5}$$

and the volatilities $\hat{s}_{f,j}$ and $\hat{s}_{\varepsilon,ij}$ are the square roots, with negative variance estimates set to zero. These are the empirical counterparts of $s_f$ and $s_{\varepsilon,i}$ in the model, and total risk decomposes as $\hat{s}^2_{D,ij} = \hat{s}^2_{f,j} + \hat{s}^2_{\varepsilon,ij}$ on the subsample where the decomposition is estimable, mirroring the model's $s^2_{D,i} = s^2_f + s^2_{\varepsilon,i}$. The construction has two properties worth noting. Unlike a decomposition based on within-firm-year means of realizations, the pairwise covariance is unbiased for any number of products. Appendix IA.C details the identifying assumption and reports robustness to the overlap screen.

# 3 Empirical Facts

We organize the descriptive evidence around four empirical facts. The unit of observation is a firm-product-year, except where we explicitly aggregate to the firm-year level.

**Fact 1.** *Product-level utilization averages 56.9%, with within-firm range 48.8–66.2%. Firm fixed effects absorb approximately 38.3% of total dispersion; the remaining approximately 61.7% is within-firm across products.*

Table 2 shows that product-level capacity utilization displays substantial dispersion. This dispersion is not simply a comparison across large and small firms. When we absorb firm fixed effects, a meaningful share of the remaining variation continues to arise within firms, across product lines. In other words, firms do not operate all products at similar rates of capacity utilization. Instead, the same firm often combines highly utilized product lines with product lines that have substantial slack. This fact is useful because it motivates a product-level analysis of capacity. It shows that cross-sectional utilization differences are not primarily driven by firm-specific characteristics such as unobserved productivity or management practices.

Why would firms choose to run some product lines at full capacity while others are not? A simple reason could be that firms have fixed capacity for each product line, and demand for that product is volatile and uncorrelated across products, so utilization rates may differ. But this intuition relies on the unrealistic assumption that firms cannot shift capacity across product lines. If that is the case, the puzzle remains: why do firms not reallocate capacity toward high-demand or more profitable product lines? We will delve into these questions in the next three facts.

TABLE 2: Product-Level Capacity Utilization Dispersion

| | |
|---|---|
| *Panel A: Summary Statistics* | |
| Mean product-level utilization | 56.9% |
| Mean within-firm minimum | 48.8% |
| Mean within-firm maximum | 66.2% |
| | |
| *Panel B: Variance Decomposition* | |
| Share explained by firm fixed effects | 38.3% |
| Share explained by firm-year fixed effects | 49.9% |
| Within firm-year share (across products) | 50.1% |

*Notes:* The estimating equation is $u_{ijt} = \alpha_{jt} + \varepsilon_{ijt}$ where $u_{ijt}$ is capacity utilization for product $i$ in firm $j$ at time $t$, and $\alpha_{jt}$ is a firm-year fixed effect. The variance decomposition reports the $R^2$ from regressing utilization on firm fixed effects. Sample includes firm-product-years with nonmissing capacity utilization.

**Fact 2.** *Product-level utilization declines with product-level demand volatility. A one-unit increase in volatility is associated with a 7.8 percentage-point decline in utilization, equivalent to about 14% of mean utilization. Aggregating the same observations to the firm level yields a nearly identical decline of 7.5 percentage points.*

In Table 3, we relate product utilization to demand volatility. We estimate regressions of product-level utilization on product-level demand volatility, with year fixed effects and standard errors clustered at the firm level. The coefficient on demand volatility is negative and statistically significant: a one-unit increase in product-level

volatility is associated with a 7.8 percentage point decline in utilization. Products that face more volatile demand are operated less intensively. The results remain quantitatively similar if we use firm×year fixed effects, so that identification comes solely from comparisons across product lines within the same firm and year (column 2). The coefficient remains negative and significant ($\hat{\beta} = -0.083$), indicating that the pattern is not driven by differences across firms in scale, financial condition, management, or firm-level demand conditions.

This finding is consistent with the long-standing literature on optimal capacity and inventory choice under demand uncertainty, Arrow et al. (1951). In this class of models, the capacity (or inventory) commitment is made before demand is realized, so the firm trades off foregone sales when demand turns out high against the cost of idle capacity when it turns out low. In expectations, the firms maintain low utilization rates. We will get back to this prediction the theory Section 4.

The empirical counterpart of this literature has so far operated at the firm or market level, reflecting the absence of product-level capacity and utilization data. To compare our estimates with firm-level evidence, column (3) aggregates the same observations to the firm-year level, weighting product-level utilization and volatility by sales shares. The firm-level relationship remains negative and of similar magnitude. The two levels of aggregation answer different questions, however, because the firm-level regression cannot reveal *which* product lines hold the slack, nor separate volatile firms from volatile products. The within-firm estimate in column (2) shows that firms hold slack in the product lines where demand risk resides while running others intensively.

So far in the literature, demand volatility has been linked to capacity only indirectly, through its effect on utilization. Direct evidence on how firms allocate capacity across product lines remains scarce. We turn to this in the next fact.

**Fact 3.** *Firms' product sales share declines with product demand volatility (-0.078). Firms also allocate smaller capacity shares to more volatile product lines (-0.123).*

In Table 4, we examine how firms allocate capacity and sales across product lines with different demand characteristics. Although we observe installed capacity at the product level, capacity *shares* cannot be computed directly because capacity is reported in heterogeneous physical units across products. For the allocation analysis in this Fact and in the next Fact, we therefore proxy the capacity share with the material cost share of each product line, recovered from the production function estimates in Appendix IA.B; utilization, by contrast, is always measured directly as output over installed capacity.

The specification regresses a product's sales share (column 1) or capacity share (column 2) on product demand volatility, absorbing firm and product×year fixed ef-

TABLE 3: Demand Volatility and Capacity Utilization

| | Product-level (1) | Product-level (2) | Firm-level (3) |
|---|---|---|---|
| Demand volatility ($\hat{s}_D$) | -0.0781*** | -0.0826*** | |
| | (0.0069) | (0.0097) | |
| Demand volatility ($\hat{s}_D$) | | | -0.0753*** |
| | | | (0.0087) |
| $R^2$ | 0.03 | 0.52 | 0.03 |
| Year fixed effects | Yes | No | Yes |
| Firm × year fixed effects | No | Yes | No |

*Notes:* The estimating equation is $u_{ijt} = \beta \hat{s}_{D,ij} + \delta_t + \varepsilon_{ijt}$ where $u_{ijt}$ is capacity utilization and $\hat{s}_{D,ij}$ is product demand volatility, the time-series standard deviation of the demand residual in equation (2). Column (1) uses product-level observations with year fixed effects $\delta_t$. Column (2) replaces the year fixed effects with firm × year fixed effects, so identification comes from comparisons across product lines within the same firm and year. Column (3) aggregates utilization and volatility to the firm-year level using sales-weighted averages, with year fixed effects. Standard errors clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

fects. Identification therefore comes from comparing product lines within the same firm, holding fixed the firm's scale, financial condition, and common demand environment, as well as aggregate conditions in each product market. Both coefficients are negative: a one-unit increase in demand volatility is associated with a 7.8 percentage point lower sales share and a 12.3 percentage point lower capacity share. Columns (3) and (4) tighten identification further by absorbing firm×year fixed effects, so the comparison is across products of the same firm in the same year; both coefficients remain negative and significant ($-0.072$ and $-0.114$). Firms tilt both capacity and sales toward their more stable product lines.

These results highlight a risk-exposure channel in the internal allocation of capacity. Firms systematically tilt capacity away from product lines with volatile demand and toward stable ones, and hold slack where risk is concentrated. This margin of adjustment — the within-firm allocation of productive capacity in response to the second moment of product-level demand — has not, to our knowledge, been documented before, and it is invisible in firm-level data. It suggests that multi-product firms treat the capacity allocation itself as an instrument for managing their exposure to demand risk, a real-side counterpart to the financial mechanisms studied in the literature on internal capital markets and corporate risk management.

**Fact 4.** *Decomposing demand volatility into firm-wide and product-specific components, utilization falls with both (-0.106 and -0.057). By contrast, the capacity share falls with product-specific volatility (-0.104) but does not respond to firm-wide volatility.*

Previous facts establish that capacity allocation and utilization respond to demand

TABLE 4: Capacity Allocation and Demand Volatility

| | Sales Share (1) | Capacity Share (2) | Sales Share (3) | Capacity Share (4) |
|---|---|---|---|---|
| Demand volatility ($\hat{s}_D$) | -0.0775*** | -0.1230*** | -0.0721*** | -0.1144*** |
| | (0.0143) | (0.0252) | (0.0191) | (0.0335) |
| $R^2$ | 0.84 | 0.80 | 0.71 | 0.83 |
| Firm fixed effects | Yes | Yes | No | No |
| Firm × year fixed effects | No | No | Yes | Yes |
| Product × year fixed effects | Yes | Yes | Yes | Yes |

*Notes:* The estimating equation is $y_{ijt} = \gamma\, \hat{s}_{D,ij} + \alpha_j + \delta_{it} + \varepsilon_{ijt}$ 1where $y_{ijt}$ is either the sales share or the capacity share of product $i$ in firm $j$. Sales share is the proportion of firm sales from product $i$. Capacity share is the material cost share, recovered from the production function estimates in Appendix IA.B. $\hat{s}_{D,ij}$ is product demand volatility from equation (3), $\alpha_j$ are firm fixed effects, and $\delta_{it}$ are product-by-year fixed effects. Columns (3) and (4) replace the firm fixed effects with firm × year fixed effects, so identification comes only from comparing products of the same firm in the same year. Standard errors clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

risk, but conflate the sources of that risk, because the volatility measure aggregates all fluctuations a product faces, whatever their origin. From the firm's perspective, however, total volatility conflates two components with very different sources. Risk that is common to all of the firm's product lines moves the demand jointly, so no internal reallocation can escape it, and it should bear on the scale at which the firm operates. This is akin to the *aggregate risk* in macro-finance models. Risk that is idiosyncratic to a product line is precisely what the internal allocation margin can manage, by tilting capacity toward whichever lines are more stable. Next, we document how the capacity allocation and utilization respond to aggregate (firm-specific) vs idiosyncratic (product-specific) shocks.

In Table 5, we regress product utilization and capacity share on the two volatility components from equations (4) and (5), absorbing product-year fixed effects. Because $\hat{s}_{f,j}$ is constant within a firm, columns (1) and (3) identify both volatilities from comparisons across firms within a product-year, while columns (2) and (4) add firm fixed effects, so the product-specific slope is identified purely from differences across products within the same firm. The decomposition requires at least two products with five overlapping years per firm, so the sample is the multi-product subsample; Appendix IA.C shows the estimates are insensitive to this screen.

Both components are negatively related to product-level utilization. Utilization declines with firm-wide volatility ($-0.106$) and with product-specific volatility ($-0.057$), indicating that firms operate product lines less intensively when demand risk is larger, whatever its origin.

Furthermore, the capacity-share regression reveals a sharper distinction. The capacity share declines with product-specific volatility ($-0.104$) but does not respond to

firm-wide volatility (a coefficient of 0.045 with a standard error of 0.078). This asymmetry is exactly what risk management through reallocation implies, i.e., shifting capacity between product lines can absorb only the risk that is specific to a line, so the allocation prices product-specific volatility and ignores the common component, which moves all lines together and leaves relative allocations unchanged.

TABLE 5: Firm vs. Idiosyncratic Demand Volatility

| | Utilization (1) | Utilization (2) | Capacity Share (3) | Capacity Share (4) |
|---|---|---|---|---|
| Firm volatility ($\hat{s}_f$) | -0.1061*** | | 0.0445 | |
| | (0.0338) | | (0.0779) | |
| Idiosyncratic volatility ($\hat{s}_\varepsilon$) | -0.0572*** | -0.0627*** | -0.1041*** | -0.1169*** |
| | (0.0152) | (0.0166) | (0.0292) | (0.0356) |
| $R^2$ | 0.34 | 0.61 | 0.48 | 0.79 |
| Firm fixed effects | No | Yes | No | Yes |
| Product × year fixed effects | Yes | Yes | Yes | Yes |

*Notes:* The estimating equation is $y_{ijt} = \beta_\varepsilon \hat{s}_{\varepsilon,ij} + \beta_f \hat{s}_{f,j} + \delta_{it} + \varepsilon_{ijt}$ where $y_{ijt}$ is either utilization or the capacity share, and $\hat{s}_{f,j}$ and $\hat{s}_{\varepsilon,ij}$ are the firm-wide and product-specific demand-risk volatilities constructed from pairwise cross-product covariances in equations (4) and (5). $\delta_{it}$ are product-by-year fixed effects. Columns (2) and (4) add firm fixed effects, which absorb $\hat{s}_{f,j}$; the product-specific slope is then identified from within-firm comparisons alone. Standard errors clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

**Taking Stock.** Taken together, the four facts document how multi-product firms manage demand risk. Capacity utilization varies substantially within firms across product lines. Product lines facing more volatile demand are operated with more slack and receive smaller shares of the firm's capacity and sales. The decomposition in Fact 4 isolates the mechanism. Utilization falls with both common and idiosyncratic demand risk, but the internal allocation of capacity responds only to the idiosyncratic component, precisely the part of demand risk that reallocation across product lines can manage. Common firm-wide shocks move all product lines together and leave relative allocations unchanged. These patterns indicate that firms treat the internal allocation of capacity as an instrument for managing demand risk. They hold slack where risk is concentrated and tilt capacity toward stable lines. They also caution against purely firm-level analyses of capacity utilization, since the margin along which firms manage demand risk operates within the firm and is invisible in aggregated data. The next section develops a model that rationalizes these facts and quantifies the value of this *internal insurance margin*.

# 4 Theoretical Framework

We develop a dynamic model of a multi-product firm that chooses investment, capacity allocation across product lines, and utilization in response to stochastic product-level demand. The firm accumulates capital subject to a one-period time-to-build and a convex adjustment cost; installed capital pins down total capacity, which the firm divides across product lines before demand is realized and utilizes after demand is observed. The model delivers three results that map directly to the empirical facts: utilization falls with both common and idiosyncratic demand volatility; the capacity allocation responds only to the idiosyncratic component; and the firm's reallocation dampens the effect of volatility on utilization. Section 4.5 uses the calibrated model to quantify the value of this internal insurance margin.

## 4.1 Environment

A representative consumer has CES preferences over differentiated firm-product varieties. There is a unit mass of atomistic firms indexed by $j \in [0,1]$, and each firm produces $\mathcal{I}$ distinct products indexed by $i \in \{1,\ldots,\mathcal{I}\}$. Thus, each pair $(i,j)$ identifies a distinct variety. Aggregate consumption is

$$C_t = \left( \int_0^1 \sum_{i=1}^{\mathcal{I}} D_{ijt}^{1/\eta} c_{ijt}^{(\eta-1)/\eta} \, dj \right)^{\eta/(\eta-1)}, \qquad \eta > 1, \tag{6}$$

where $c_{ijt}$ is consumption of product $i$ supplied by firm $j$, $D_{ijt}$ is its stochastic demand shifter, and $\eta$ is the elasticity of substitution across varieties. The corresponding aggregate price index is

$$P_t = \left( \int_0^1 \sum_{i=1}^{\mathcal{I}} D_{ijt} p_{ijt}^{1-\eta} \, dj \right)^{1/(1-\eta)}. \tag{7}$$

Utility maximization implies the demand curve

$$c_{ijt} = D_{ijt} p_{ijt}^{-\eta} C_t P_t^{\eta}. \tag{8}$$

We study the firm's problem in partial equilibrium. Because firms are atomistic, each firm takes the aggregate demand index $C_t P_t^{\eta}$ as given; we normalize it to one. Market clearing for each variety requires $c_{ijt} = y_{ijt}$, where production is

$$y_{ijt} = u_{ijt} K_{jt} \omega_{ijt}. \tag{9}$$

Here $K_{jt}$ is firm $j$'s installed capacity, $\omega_{ijt} \in [0,1]$ is the share allocated to product $i$, with $\sum_i \omega_{ijt} = 1$, and $u_{ijt} \in [0,1]$ is the utilization rate of that product's capacity.

Inverting the demand curve, the price at which the market absorbs $y_{ijt}$ units is

$$p_{ijt} = \left( \frac{D_{ijt}}{y_{ijt}} \right)^{1/\eta}. \tag{10}$$

Product-level revenue is therefore

$$R_{ijt} = p_{ijt} y_{ijt} = D_{ijt}^{1/\eta} \left( u_{ijt} K_{jt} \omega_{ijt} \right)^{1-1/\eta}. \tag{11}$$

Utilization is costly:

$$C_{ijt} = \beta \, K_{jt} \, \omega_{ijt} \, u_{ijt}^{\gamma}, \qquad \beta > 0, \ \gamma > 1. \tag{12}$$

The convexity parameter $\gamma > 1$ captures the increasing marginal cost of pushing utilization toward capacity, reflecting overtime wages (Basu and Kimball, 1996), accelerated depreciation (Greenwood et al., 1988), and other margins through which higher utilization becomes increasingly costly. Operating profit on product $i$ is

$$\Pi_{ijt} = R_{ijt} - C_{ijt} = D_{ijt}^{1/\eta} (u_{ijt} K_{jt} \omega_{ijt})^{1-1/\eta} - \beta K_{jt} \omega_{ijt} u_{ijt}^{\gamma}. \tag{13}$$

Log demand for product $i$ at firm $j$ is the sum of a firm-wide component, a product-specific mean, and a product-idiosyncratic component:

$$\ln D_{ijt} = f_{jt} + \mu_{ij} + \nu_{ijt}. \tag{14}$$

The firm shock $f_{jt}$ is common to all of firm $j$'s product lines in year $t$; $\mu_{ij}$ is a product-specific mean capturing time-invariant differences in average demand across product lines; and $\nu_{ijt}$ is a product-firm-year idiosyncratic shock. The two stochastic components follow independent AR(1) processes:

$$f_{jt} = \rho_f \, f_{j,t-1} + \sigma_f \, \xi_{jt}, \qquad \xi_{jt} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \tag{15}$$

$$\nu_{ijt} = \rho_{\varepsilon,i} \, \nu_{ij,t-1} + \sigma_{\varepsilon,i} \, \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \tag{16}$$

with $\xi_{jt} \perp \varepsilon_{kjt}$ for all $k, t$ and $\varepsilon_{ijt} \perp \varepsilon_{kjt}$ for $i \neq k$. The persistence parameters satisfy $\rho_f, \rho_{\varepsilon,i} \in (-1,1)$. The unconditional variances are

$$s_f^2 \equiv \frac{\sigma_f^2}{1-\rho_f^2}, \qquad s_{\varepsilon,i}^2 \equiv \frac{\sigma_{\varepsilon,i}^2}{1-\rho_{\varepsilon,i}^2}, \qquad s_{D,i}^2 \equiv s_f^2 + s_{\varepsilon,i}^2, \tag{17}$$

where $s^2_{D,i}$ is the total unconditional log-demand variance for product $i$. The decomposition (14) is the structural counterpart to a one-factor model: a firm-wide common factor that moves all of a firm's products together, plus product-specific noise.

Finally, the firm accumulates capital according to

$$K_{jt+1} = (1-\delta)K_{jt} + i_{jt}, \qquad \delta \in (0,1), \tag{18}$$

where $\delta$ is the depreciation rate and $i_{jt}$ is investment. Investment incurs a quadratic adjustment cost $\Phi(i_{jt}, K_{jt}) = \frac{\theta}{2}(i_{jt}/K_{jt})^2 K_{jt}$ with $\theta > 0$.

## 4.2 Timing

Capital and capacity shares are committed before demand is realized, while utilization is chosen after: ex-ante choices made under uncertainty must balance the volatility of one product line against another, while the ex-post utilization decision absorbs whatever demand realization actually occurs. We summarize the timing and information revelation in Figure 1. Let

$$s_{jt} \equiv (f_{jt}, \nu_{1jt}, \ldots, \nu_{\mathcal{I}jt}) \tag{19}$$

denote the firm's state at the end of period $t$—the realized values of the firm-common shock and the product-idiosyncratic shocks. Within period $t$, decisions and information unfold in the following order. The firm enters period $t$ with predetermined capital $K_{jt}$ and a predetermined capacity allocation $\{\omega_{ijt}\}_{i=1}^{\mathcal{I}}$. The capacity allocation across product lines is set by the rule $\omega_{ijt} = \Omega_i(s_{j,t-1})$, derived in Section 4.3 below. The shocks $\xi_{jt}$ and $\{\varepsilon_{ijt}\}_{i=1}^{\mathcal{I}}$ then realize; the firm observes the period-$t$ state $s_{jt}$ and, through (14), the realized product-level demands $D_{ijt}$. Given $\{D_{ijt}\}_{i=1}^{\mathcal{I}}$ and the predetermined $(K_{jt}, \{\omega_{ijt}\}_{i=1}^{\mathcal{I}})$, the firm chooses utilization $u_{ijt}$ for each product line, generating output, revenues, and operating profits. At the end of period $t$, the firm chooses investment $i_{jt}$. Dividends are then paid,

$$d_{jt} = \pi_{jt} - i_{jt} - \Phi(i_{jt}, K_{jt}), \tag{20}$$

where $\pi_{jt} = \sum_i \Pi_{ijt}$ is total operating profit.

## 4.3 Firm Problem

**Utilization.** After demand is realized (Step 3 in Figure 1), the firm chooses utilization $u_{ijt}$ given $(D_{ijt}, K_{jt}, \omega_{ijt})$. This is a static product-by-product problem. Given $\omega_{ijt}$ and

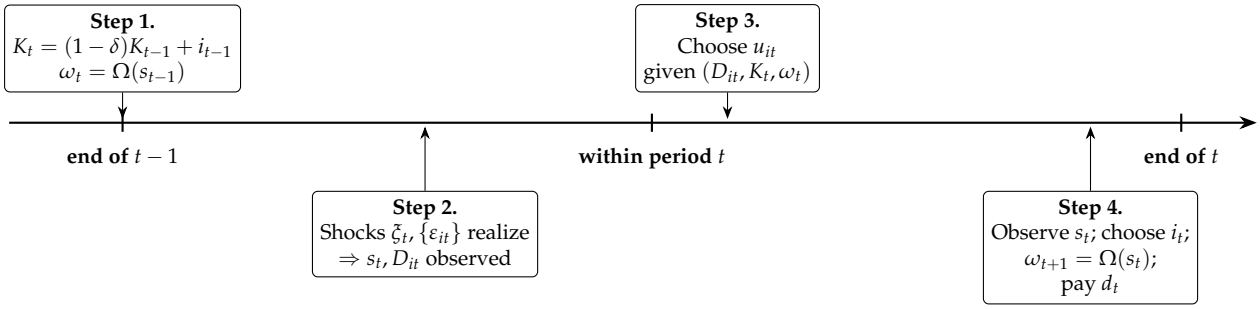



FIGURE 1: Timing in the model.

realized $D_{ijt}$, the optimal utilization (derived in Appendix IA.D.1) is

$$u^*_{ijt}(D_{ijt}, \omega_{ijt}) = \left(\frac{\eta - 1}{\gamma\beta\eta}\right)^{\eta\psi} \left(\frac{D_{ijt}}{K_{jt}\,\omega_{ijt}}\right)^{\psi}, \tag{21}$$

where $\psi \equiv \frac{1}{\eta(\gamma-1)+1}$ is the elasticity of utilization to demand shocks. Two distinct economic forces pin down $\psi$. The *cost channel* operates through $\gamma$: more convex utilization costs make it expensive to push capacity hard, so the firm smooths its response to demand realizations and $\psi$ falls. The *markup channel* operates through $\eta$: with more elastic demand (lower markup), the firm has weaker incentive to push output because additional sales come at a sharper price decline, again lowering $\psi$.

**Firm value function.** The value of the firm is

$$\begin{aligned} V_t(K_{jt}, \{\omega_{ijt}\}_{i=1}^{\mathcal{I}}, s_{jt}) = \pi^*_{jt}(K_{jt}, \{\omega_{ijt}\}_{i=1}^{\mathcal{I}}, s_{jt}) \\ + \max_{K_{jt+1},\, \{\omega_{ijt+1}\}_{i=1}^{\mathcal{I}}} \Big\{ -i_{jt} - \Phi(i_{jt}, K_{jt}) \\ + \frac{1}{1+r}\,\mathbb{E}_t\Big[V_{t+1}(K_{jt+1}, \{\omega_{ijt+1}\}_{i=1}^{\mathcal{I}}, s_{jt+1})\Big]\Big\}, \end{aligned} \tag{22}$$

where $\pi^*_{jt}$ is operating profit evaluated at the optimal utilization rule (21) and the expectation is taken over next-period shocks $(\xi_{j,t+1}, \{\varepsilon_{ij,t+1}\}_i)$ conditional on $s_{jt}$. The firm chooses next-period capital $K_{j,t+1}$ and the next-period capacity plan $\{\omega_{ij,t+1}\}_i$ after observing the current state $s_{jt}$, but before next-period demand is realized; utilization is chosen only after that realization. Marginal $q$ is defined as

$$q_{jt} \equiv \frac{\partial V_t(K_{jt}, \{\omega_{ijt}\}_{i=1}^{\mathcal{I}}, s_{jt})}{\partial K_{jt}},$$

the shadow value of an additional unit of installed capital. We return to $q$ when we examine the relation between average $q$, marginal $q$, and excess capacity in Section 4.5.

**Capacity allocation.** At the end of period $t-1$, the firm chooses the allocation $\omega_{ijt}$ to be deployed in period $t$. This choice conditions on the observed state $s_{t-1}$ and the anticipated within-period utilization response.

**Proposition 1.** *The capacity choice is*

$$\omega_{ijt} = \Omega_i(s_{jt-1}) \equiv \frac{M_{ijt|t-1}^{1/\phi}}{\sum_k M_{kjt|t-1}^{1/\phi}}, \qquad M_{ijt|t-1} \equiv \mathbb{E}_{t-1}[D_{ijt}^{\phi}], \tag{23}$$

*where $\phi = \gamma\psi$ is the elasticity of operating profit to product-level demand.*

**Proof.** See Appendix IA.D.2. ■

The optimal capacity share has a transparent economic structure. In our environment, $\phi \equiv \gamma\psi < 1$ for $\gamma, \eta > 1$, so operating profit is concave in demand. Two forces shape the allocation $\Omega_i$. The *level channel* tilts capacity toward products with higher expected demand: under lognormal $D_i$ with $\mathbb{E}[D_i] = \bar{D}_i$ held fixed, $M_i^{1/\phi} \propto \bar{D}_i \cdot \exp(-\frac{1}{2}(1-\phi)\, s_{D,i}^2)$ (see Appendix IA.D.2 for the derivation), so the leading term is proportional to mean demand. The *volatility channel* works through the exponent on $s_{D,i}^2$: because profit is concave in demand, a mean-preserving spread reduces expected profit, and the firm tilts capacity away from volatile product lines. The strength of this penalty is governed by $(1-\phi)$: more concave profit (larger $1-\phi$, i.e. lower demand elasticity $\eta$ or more convex utilization costs $\gamma$) produces a sharper allocation response to volatility. Notably, because $M_i$ depends only on the marginal distribution of $D_i$, the cross-product correlation structure of demand does not enter the allocation: two firms with identical marginal distributions but very different cross-product correlations face identical allocation problems.

Further, the capacity allocation does not depend on firm size $K_{jt}$; it is fully pinned down by expected product-level demand given the information set $s_{j,t-1}$. This separation reduces the dimensionality of the firm problem—the state space collapses to $(K_{jt}, s_{jt})$, with the allocation rule $\Omega(\cdot)$ solved analytically.

**Steady state comparison.** Here, we examine how product-level utilization rates are related to demand volatility in steady state, and uncover the structural parameters that drive this relationship. Throughout this discussion we suppress the firm index $j$ to lighten notation.

**Proposition 2** (Expected utilization)**.** *Under the shock structure* $\ln D_{it} = f_t + \mu_i + \nu_{it}$ *with unconditional variances* $s_f^2$ *and* $s_{\varepsilon,i}^2$*, expected steady-state utilization is*

$$\mathbb{E}[u_i^{*,ss}] = \tilde{\Lambda} \cdot \exp\left(-\tfrac{1}{2}(\gamma-1)\psi^2\,(s_f^2 + s_{\varepsilon,i}^2)\right), \tag{24}$$

*where* $\tilde{\Lambda} \equiv A \cdot [\alpha_1(1-\phi)]^{-1/\gamma} \left[(1+\theta\delta)(r+\delta) - \frac{\theta}{2}\delta^2\right]^{1/\gamma}$ *is a constant.*

**Proof.** See Appendix IA.D.3. ■

In the stationary benchmark, expected utilization is *independent of* $\mu_i$, because at $\omega^*$ the capacity-reallocation channel exactly offsets the direct effect of mean demand on utilization. Expected utilization declines with *both* firm-level volatility $s_f$ and product-idiosyncratic volatility $s_{\varepsilon,i}$, entering symmetrically through the total variance $s_f^2 + s_{\varepsilon,i}^2$ with coefficient $(\gamma-1)\psi^2 > 0$.

The coefficient $(\gamma-1)\psi^2$ measures how strongly demand volatility lowers expected utilization. Its source is the shape of the utilization rule (21): because $\psi < 1$, optimal utilization is concave in realized demand, so a mean-preserving spread in demand lowers utilization on average—high-demand realizations raise $u$ by less than low-demand realizations reduce it. The strength of this effect depends on two parameters that pull in *opposite directions* as $\gamma$ changes: more convex costs raise the penalty on utilization variation, captured by $(\gamma-1)$, but they also flatten the utilization response itself, lowering $\psi$. The coefficient is the product of these opposing forces and vanishes at both extremes. With linear costs ($\gamma \to 1$), utilization moves one-for-one with demand ($\psi \to 1$) and there is no curvature to generate a loss; with extremely convex costs ($\gamma \to \infty$), utilization barely responds to demand ($\psi \to 0$), so there is no variation to penalize. The demand elasticity $\eta$ operates through $\psi$ alone: more elastic demand weakens the firm's incentive to chase demand with utilization, lowers $\psi$, and so mutes the effect of volatility on expected utilization.

**Lemma 1** (Steady-state capacity allocation)**.** *In the steady-state benchmark with lognormal demand and unconditional variances* $s_f^2$, $s_{\varepsilon,i}^2$, *the optimal capacity share satisfies*

$$\omega_i^* \;\propto\; \bar{D}_i \cdot \exp\Big(-\tfrac{1}{2}(1-\phi)\, s_{\varepsilon,i}^2\Big).$$

**Proof.** See Appendix IA.D.2. ■

Two consequences follow from this Lemma. The firm-wide variance $s_f^2$ does not affect capacity allocation, because a common shock that scales every product's demand by the same factor creates no incentive to shift capacity across lines. Second, holding mean demand fixed, the share is strictly decreasing in the product-idiosyncratic variance with elasticity $-(1-\phi)/2 < 0$. More volatile lines receive less capacity; the strength of the tilt is governed by $(1-\phi)$, the concavity of profit in demand.

**Lemma 2** (Joint optimum vs. fixed allocation)**.** *The coefficient on* $s_{D,i}^2$ *that determines how strongly demand volatility lowers expected utilization differs between two benchmarks. At fixed* $\Omega$ *(not optimized), a mean-preserving-spread calculation gives a coefficient of* $\eta(\gamma -$

$1)\psi^2$. *At the joint optimum* $\Omega^*$*, the coefficient is* $(\gamma-1)\psi^2$*—smaller by a factor of* $\eta$*. The reduction comes from the reallocation channel: volatile products receive less capacity, and because* $u^* \propto \omega^{-\psi}$ *a smaller share mechanically raises utilization on those lines. As a result, expected utilization is less sensitive to demand volatility at the joint optimum than under any fixed allocation.*

**Proof.** See Appendix IA.D.5. ■

Taken together, these results rationalize the empirical facts. Expected utilization falls with both common and idiosyncratic demand volatility (Proposition 2), matching the utilization patterns in Facts 2 and 4; the capacity allocation tilts away from volatile product lines and responds only to the idiosyncratic component (Lemma 1), matching Facts 3 and 4. Lemma 2 highlights why this margin matters: by reallocating capacity toward stable product lines, the firm blunts the impact of demand volatility on its utilization, profits, and value. The next section calibrates the model and quantifies the value of this internal insurance margin.

## 4.4 Estimation Strategy

We estimate the model on the Prowess product-firm-year panel by the method of simulated moments. We implement the model with two product lines and map them in the data to each firm's two largest products by total sales over the sample period (product 1 the larger line, without loss of generality) using Prowess's product classification, whose product codes are consistent across years, so we follow each product line over time; the descriptive facts use all of a firm's products, whereas the structural estimation uses the leading two to match the two-product line model. On average these two products account for 96% of a firm's sales, so the two-product specification captures the bulk of its activity.

Our estimation strategy consists of two parts. We first fix a subset of parameters externally based on standard values and direct data counterparts. We then take these external parameters as given and use a simulated method of moments (SMM) procedure to estimate the remaining parameters by matching a targeted set of moments. We describe each step in detail.

**Externally set Parameters.** We fix the demand elasticity fixed at $\eta = 5$ (a standard value implying a steady-state markup $\eta/(\eta-1) = 1.25$). The depreciation rate $\delta = 0.08$ is set to the average depreciation rate in the data, and the adjustment cost $\theta = 1$ and interest rate $r = 0.064$ to standard values.

**Simulated Method of Moments.** We choose the utilization-technology parameters $(\gamma, \beta)$, the firm-specific and product-specific AR(1) innovation variance and persistence parameters $(\sigma_f^2, \rho_f; \sigma_{\varepsilon,i}^2, \rho_{\varepsilon,i})$, and the product demand-mean gap $\mu_1 - \mu_2$ to minimize the weighted distance between a vector of model-simulated moments and their data counterparts. The targeted moments are computed by the same code on the simulated and the Prowess panels: the mean and within-firm cross-product dispersion of capacity utilization; the variances, cross-product covariance, and first-order autocorrelations of demeaned log product output; and the average sales share of the leading product.[7] Each parameter is disciplined primarily by one moment: the cost scale $\beta$ by the mean utilization level; the curvature $\gamma$ by the volatility and within-firm dispersion of utilization; the firm-common variance $\sigma_f^2$ by the cross-product covariance of output; the idiosyncratic variances $\sigma_{\varepsilon,i}^2$ by the own output variances; the persistences by the corresponding autocorrelations; and the demand-mean gap by the sales share.

Table 6 reports the estimates and Table 7 the moment fit, which is close across all ten targets. The estimated demand is highly volatile (unconditional idiosyncratic volatilities $s_{\varepsilon,i} = \{1.32, 1.61\}$ and firm-common $s_f = 1.34$), reflecting the high volatility of product-level output and utilization in the data; product 1 is the larger, more persistent ($\rho_{\varepsilon,1} = 0.62$), and somewhat less volatile line.

TABLE 6: Calibrated model parameters

| Parameter | Symbol | Value | Source / Discipline |
|---|---|---|---|
| Demand elasticity | $\eta$ | 5.000 | Set externally |
| Utilization curvature | $\gamma$ | 1.165 | Estimated (SMM) |
| Utilization cost scale | $\beta$ | 1.057 | Estimated (SMM) |
| Depreciation rate | $\delta$ | 0.080 | Matched to data |
| Interest rate | $r$ | 0.064 | Set externally |
| Adjustment cost parameter | $\theta$ | 1.000 | Set externally |
| Firm shock persistence | $\rho_f$ | 0.445 | Estimated (SMM) |
| Firm shock innovation variance | $\sigma_f^2$ | 1.436 | Estimated (SMM) |
| Product demand log-mean | $\{\mu_1,\ \mu_2\}$ | $\{0.926,\ -0.926\}$ | Estimated (SMM) |
| Product idiosyncratic persistence | $\{\rho_{\varepsilon,1},\ \rho_{\varepsilon,2}\}$ | $\{0.623,\ 0.299\}$ | Estimated (SMM) |
| Product idiosyncratic innovation variance | $\{\sigma_{\varepsilon,1}^2,\ \sigma_{\varepsilon,2}^2\}$ | $\{1.068,\ 2.356\}$ | Estimated (SMM) |
| Product idiosyncratic volatility (uncond.) | $\{s_{\varepsilon,1},\ s_{\varepsilon,2}\}$ | $\{1.321,\ 1.609\}$ | Derived |

[7] Because physical product output has extreme, lumpy tails, the log-output moments use observations winsorized at the 2.5% level

Table 7 reports the moments of the calibrated baseline. The firm operates well below installed capacity, with mean firm-level utilization of 0.53, and runs its two lines at materially different rates, the within-firm utilization dispersion in Table 7 being the model counterpart of Fact 1's within-firm spread. Allocation concentrates on the larger line: product 1 takes a sales share of 0.78 versus 0.22 for product 2, matching the data; the tilt away from the more volatile line highlights the underlying capacity allocation, i.e., firms allocate capacity away from highly volatile product lines.

TABLE 7: Targeted Moments

| Moment | Data | Model |
|---|---|---|
| *Targeted moments* | | |
| Mean utilization | 0.601 | 0.516 |
| Std. dev. of log utilization | 0.995 | 0.879 |
| Within-firm utilization dispersion | 0.323 | 0.254 |
| Mean sales share (leading product) | 0.776 | 0.777 |
| Variance of log output, product 1 | 0.857 | 0.851 |
| Variance of log output, product 2 | 0.959 | 1.047 |
| Cross-product covariance of log output | 0.187 | 0.193 |
| Autocorrelation of log output, product 1 | 0.558 | 0.619 |
| Autocorrelation of log output, product 2 | 0.470 | 0.480 |
| Autocorrelation of firm-mean log output | 0.527 | 0.470 |

*Notes:* The targeted block reports the ten auxiliary moments matched by SMM, computed by the same code on the Prowess product-firm-year panel and on the simulated model; the output (co)variance and autocorrelation moments use demeaned log output winsorized at 2.5%.

### 4.4.1 Untargeted Moments

How utilization and the capacity allocation respond to demand volatility is not targeted in estimation, so these responses provide an out-of-sample test of the model; we set them against empirical Facts 2 and 4. We run two controlled experiments. In each of these experiments, we hold a product line's mean demand fixed, so that only the volatility of its demand changes, and we trace the line's mean utilization and its capacity share $\omega_i$. The first experiment raises the line's *idiosyncratic* volatility $s_\varepsilon$, holding the firm-common component fixed; the second experiment raises the *firm-common* volatility $s_f$ shared by both lines. Table 8 reports the slopes and Figure 2 the curves.

The model reproduces the pattern of responses in the data. First, utilization falls with volatility from either source (Fact 2): the firm runs more volatile lines with more slack, a precautionary response to demand risk. Second, the capacity allocation responds to *idiosyncratic* volatility but not to firm-common volatility. The model slope of the capacity share on idiosyncratic volatility is $-0.129$, against $-0.104$ in the data, while its slope on firm-common volatility is exactly zero—matching the null in Fact 4,

where the data coefficient of 0.045 is statistically indistinguishable from zero. This follows directly from the optimal allocation rule (Lemma 1). The variance of the firm-common component enters identically for both product lines and the capacity shares allocation cancels exactly. Thus only the idiosyncratic impacts allocation of capacity across product lines. Because only the idiosyncratic component reallocates capacity, the share is weakly related to *total* demand volatility (Fact 3) even as it responds to the idiosyncratic part (Fact 4). The model's slopes are uniformly steeper than their empirical counterparts, as expected if measurement error in the estimated volatility components attenuates the data coefficients.

TABLE 8: Untargeted Moments

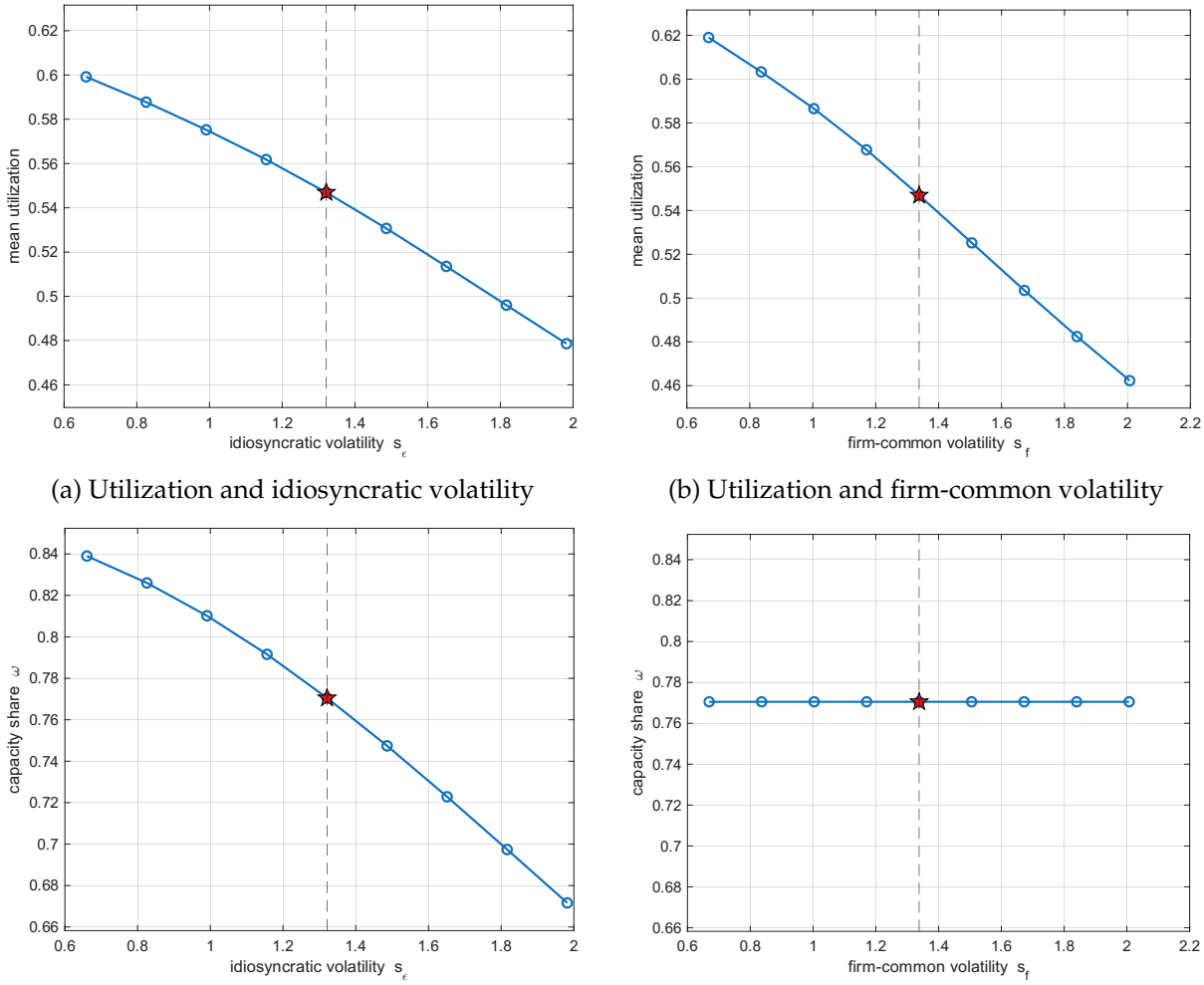
| | Slope on demand volatility | |
|---|---|---|
| Response | Data | Model |
| Utilization, idiosyncratic vol. $s_\varepsilon$ | -0.057*** | -0.092 |
| Utilization, firm-common vol. $s_f$ | -0.106*** | -0.119 |
| Capacity share, idiosyncratic vol. $s_\varepsilon$ | -0.104*** | -0.129 |
| Capacity share, firm-common vol. $s_f$ | 0.045 | 0.000 |

*Notes:* Slopes of product utilization and of the product capacity share on idiosyncratic and on firm-common demand volatility (the standard deviation of the corresponding log-demand component). The model column reports the slope from a controlled experiment that raises one volatility component while holding the model's mean demand fixed. The data column reports the coefficients on $\hat{s}_{\varepsilon,ij}$ and $\hat{s}_{f,j}$ from columns (1) and (3) of Table 5, in which both volatilities enter jointly. Standard errors clustered at the firm level. $^{***}$ $p < 0.01$, $^{**}$ $p < 0.05$, $^{*}$ $p < 0.10$.

## 4.5 Results

Having shown that the model reproduces the empirical facts, we now use it as a laboratory to answer three questions that the data alone cannot. The first question we tackle is how much is the internal reallocation margin worth? To do so, we shut down the firm's ability to adjust capacity shares in response to demand, i.e. we force the firm to optimize on every margin keeping the capacity shares at their unconditional average. Second we ask how does the composition of risk impact firms decision? We vary the level of idiosyncratic and common risk within a firm, while keeping the total risk the firm faces unchanged. We show that firms facing different risk compositions endogenously differ sharply in utilization dispersion, reallocation activity, and profit volatility. Finally, we ask what does optimal underutilization, or in other words the fact that firms optimally leave some production slack, imply for how we measure investment opportunities? We revisit the relation between average $q$, marginal $q$, and excess capacity, and show that the wedge between the two, depend on whether excess capacity reflect optimal slack or a binding constraint.

FIGURE 2: Untargeted moments.




(a) Utilization and idiosyncratic volatility

(b) Utilization and firm-common volatility

(c) Capacity share and idiosyncratic volatility

(d) Capacity share and firm-common volatility

*Notes:* Mean utilization and capacity share $\omega$ of a product line as its idiosyncratic demand volatility $s_\varepsilon$ or the firm-common volatility $s_f$ rises, holding mean demand fixed. Utilization falls with both volatility components; the capacity share falls with idiosyncratic volatility but is invariant to firm-common volatility. The marker denotes the benchmark calibration.

### 4.5.1 The value of within-firm capacity reallocation

We start by studying the value of internal reallocation margins for a firm. To understand the value of adjusting capacity, we compare the a firm that optimally allocates capacity across product lines to a counterfactual firm that cannot reallocate capacity. In the benchmark model, firms can reset their capacity plan each period to the conditionally optimal rule $\omega_t = \Omega(s_{t-1})$, so capacity shares optimally respond to the lagged state space. In our counterfactual we do not allow firms to adjust capacity. Instead, firms have their capacity plans frozen at $\omega_i^* = M_i^{1/\phi}/S$ (Proposition 1), the allocation that maximizes unconditional expected profit among all time-invariant plans. This comparison therefore allows us to isolate the option value of changing commit-

ted capacity across product lines, in particular, the value to shift capacity to lines with higher expected demand. Because the frozen-plan firms in the counterfactual commit their capacity to their best possible fixed allocation, the difference between the two scenarios are a lower bound on the value of allocation flexibility. To trace the value of capacity reallocation, we scale the idiosyncratic component of demand volatility, symmetrically for both product lines, holding the firm-common volatility $s_f$ fixed.

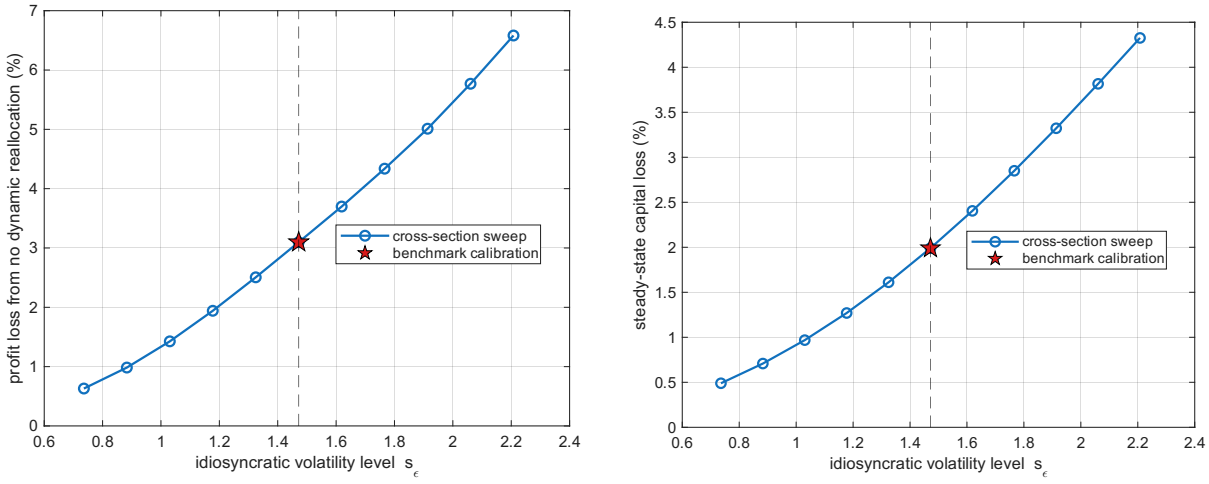
Panel (a) and (b) from Figure 3 plots the profit and capital losses, respectively, from this experiment. In particular, it plots the change in profits/capital from freezing capacity allocation on the $y$-axis as we vary the level of idiosyncratic volatility $s_\varepsilon$ on the $x$-axis. Both losses rise monotonically as we increase volatility. The increasing pattern reflects jointly the two ingredients that make reallocation valuable. On one hand, idiosyncratic shocks make the demand of each product line diverge, and the persistence of this shocks, make the divergence forecastable. If our product demand shocks where i.i.d., both the optimal conditional capacity plan and the unconditional capacity plan would coincided, and thus the value of capacity reallocation would be zero. Similarly, the firm-common shock moves both lines together and creates no reallocation motive, so it contributes nothing to this margin.

At our estimated level of idiosyncratic volatility (marked with a red start on the plots), freezing the capacity plan reduces long-run average profits by 3.1% and the capital stock by 2.0%. If firms could not adjust capacity in reaction to changes in demand across their product lines, optimal capital stock is endgenously lower. This is intuitive, since a frozen capital stock that cannot be reallocated across product lines, lowers expected profit per unit of capacity, and with it the return to installed capital. Thus a firm that cannot adjust its allocated capital, is not only less profitable, it is also smaller, with lower assets. This internal reallocation of capital is therefore an insurance margin that firms can use to smooth product specific shocks, and quantitatively an important driver of firm value and capital allocation. This margin is more valuable for firms with more volatile product lines, more persistent product lines and for firms whose volatility of demand is mainly driven by product specific volatility rather than common component.

#### 4.5.2 Same firm-level volatility, different demand-risk management

Next, we study whether the composition of demand risk matter, holding total risk fixed. In particular, we construct firms that are observationally identical in overall firm-level volatility but differ on whether most of the demand risk they face comes from the common firm level component or the idiosyncratic component. We hold each product's marginal demand distribution fixed, meaning that firms have the same log-mean $\mu_i$ and the same total log-demand variance $s^2_{D,i} = s^2_f + s^2_{\varepsilon,i}$, but we change

FIGURE 3: Firm Size and Demand Volatility




(a) Profit loss

(b) Capital loss

*Notes:* Steady-state profit loss and capital loss from freezing the capacity plan at $\omega^*$ instead of reallocating, plotted against the level of idiosyncratic demand volatility $s_\varepsilon$ (both product lines scaled together, holding the firm shock $s_f$ and the within-firm balance fixed). The dashed line and marker indicate the benchmark calibration.

the split between the variance of the firm-common component $s_f^2$, that is shared by both product lines, and the product-specific components $s_{\varepsilon,i}^2$. More specifically, we change the ratio of the two types of volatility, $s_f/s_D$, by increasing $s_f^2$ and lowering each $s_{\varepsilon,i}^2$ by the same amount, while holding the persistence fixed. Because each product's marginal distribution is unchanged, the firm-level demand volatility an econometrician would measure is identical across these firms. Only the cross correlation of demand across product lines changes, since that correlation is driven entirely by the shared volatility component. For each value of $s_f/s_D$ we simulate the model and three key within-firm statistics: the dispersion of utilization across product lines, the intensity of capacity reallocation (the time-series standard deviation of the capacity share $\omega_1$), and firm-level profit volatility.

Figure 4 plots the three objects against the common share $s_f/s_D$: panel (a) shows the dispersion in utilization, panel (b) the reallocation of capacity and panel (c) the volatility of firm level products. As we increase the share of volatility coming from the firm-level component, the dispersion in utilization across product lines falls, there is less capacity reallocation and the volatility of firm profits rises. This patterns can be easily rationalized, through the lenses of the allocation rule (Lemma 1) and are economically intuitive: the capacity plan responds only to the idiosyncratic component of demand risk, so a firm whose risk is mostly product-specific is more likely to reallocate capacity across product lines, while a form whose risk is mainly driven by a firm level component, cannot use the capacity margin, since all demand moves synced across product lines, making reallocation useless, and therefore profits more volatile.

The differences are economically large. Comparing a firm whose demand variance is predominantly idiosyncratic to one dominated by the common component, the former exhibits 2.41 times more within-firm utilization dispersion, reallocates capacity 3.05 times more actively, and has 8.9% lower firm-level profit volatility, despite the fact that total firm-level volatility was kept unchanged in this experiment. The key implication, is that two observationally equivalent firms in terms of how volatile their demand is, can differ systematically in their internal operations, their capacity to self-insure against demand shocks and ultimately how stable their profits are. Product scope is therefore a margin through which firms can alleviate volatility of their cash-flows, if they have a product mix that is uncorrelated.

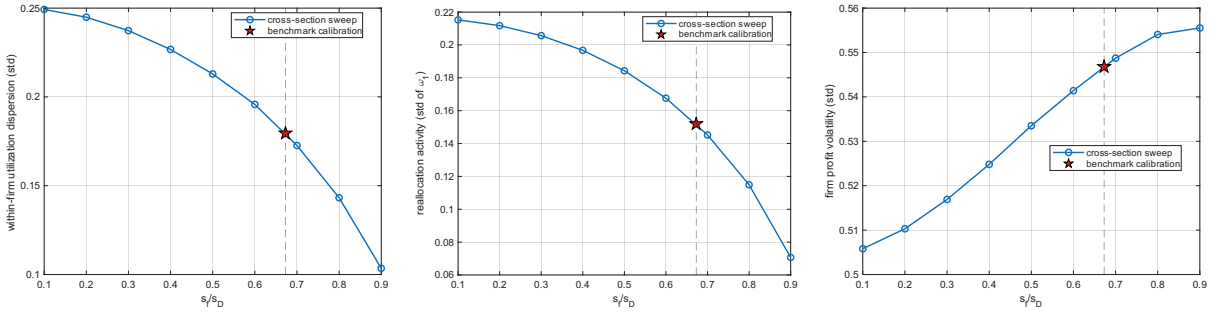
Dispersion in utilization connects our results to the misallocation literature. At the optimal utilization rule, the product-level average revenue product of capacity is log-linear in utilization, $\ln ARPK_{ijt} = \gamma \ln u_{ijt} + \text{const}$ (Appendix IA.D.4), so within-firm dispersion in $ARPK$ across product lines is utilization dispersion scaled by $\gamma^2$. A large literature reads dispersion in revenue products as evidence of capital misallocation (Hsieh and Klenow, 2009). Endogenous utilization can make this inference non-trivial. If utilization responds optimally to shocks, a higher $ARPK$ dispersion can mean a *better* allocative efficiency (Kabir et al., 2024). Our results underscore the same economic mechanism, but whitin a firm. The within-firm, across-product dispersion in revenue products that the model generates is the outcome of endogenous optimal demand-risk management, and not of frictions in capital allocation. For example, Kehrig and Vincent (2026) documents that within-firm dispersion in plant-level revenue products is "good dispersion" arising from internal capital markets under costly external finance. Our mechanism is complementary but different: dispersion in utilization arises from internal insurance rather than internal financing. Capacity and utilization respond optimally to the composition of demand risk. Panel A of Figure 4 shows exactly how this optimal dispersion varies with the composition of whithin firm risk. When the largest component of demand is is idiosyncratic, the possibility of moving capital across product lines, is very valuable, insuring firms. If on the contrary, the common component of firm level demand dominates, in the limit there is nothing left to insure against.

Our results also show that realized firm-level profit, an equilibrium outcome, in our model, is not sufficient to measure firm-level risk. Two firms facing identical product-level demand risk can display very different profit volatility depending on the correlation of that risk across their product lines and on the internal reallocation it generates. The decomposition of demand risk into common and idiosyncratic components, which requires product-level data, is important to fully grasp how risky a firm is.

We should note that our model abstracts from financial frictions. Thus the decline

of profit volatility highlight in panel (c) of the figure, is a by-product of the firm's operating decision. The firms in our model place no value in smoothing cashflows, to avoid having to raise external equity. If on the other hand firms faced costly external financing, and if the stochastic discount factor, would place a high price on states of the world where the firm would have to raise equity, and additional benefit of smoother cashflows would affect firm value ((Froot et al., 1993)), making capacity reallocation a way of firms to possibly substitute financial hedging.[8]

FIGURE 4: Optimal Slack: Firm vs Product-level Demand Volatility




(a) Utilization dispersion (b) Capacity reallocation (c) Firm-level profit volatility

*Notes:* Same firm-level volatility, different internal economics. Each panel plots one within-firm object against the firm common share $s_f/s_D$, holding every product's marginal demand distribution fixed; the marker denotes the benchmark calibration. Utilization dispersion is the cross-product standard deviation of realized utilization within a firm-year, averaged over the simulated panel. Capacity reallocation is the within-firm time-series standard deviation of the planned capacity share $\omega_1$, averaged across firms. Firm-level profit volatility is the within-firm time-series standard deviation of total operating profit, averaged across firms.

### 4.5.3 Slack as a choice: implications for measuring $q$

Since Hayashi (1982), the empirical content of $q$-theory has rested on average $q$ proxying for the marginal $q$ that governs investment. Grullon and Ikenberry (2025) document that excess capacity undermines this proxy—when firms expect new capital to sit partly idle, average $q$ decouples from marginal $q$, and investment regressions improve substantially once asset utilization is controlled for. We ask what the proxy problem looks like when excess capacity is a choice rather than a friction. Because slack arises optimally in our model and the model delivers the true marginal $q$, we can answer two questions that firm-level data cannot. First, how much information about investment opportunities is lost by measuring utilization at the firm rather than the product level? Second, how should the average–marginal $q$ gap be interpreted when slack is chosen?

We first verify that the model reproduces the empirical pattern in Grullon and Ikenberry (2025). On the simulated panel, a regression of the investment rate $I/K$ on aver-

[8]For example, Almeida et al. (2017), shows that firms replace supply-contract hedging with financial derivatives when the latter become available.

age $q$ yields an $R^2$ of 0.77; adding firm-level asset utilization, $AU = \sum_i \text{sales}_i / K$, lifts it to 0.84 with a positive coefficient (Table 9), mirroring the improvement they document in Compustat. Our model, however, contains no demand rationing. In their framework the firm cannot sell more than realized demand; in ours the firm can always sell more by moving down its demand curve, and slack is a choice rather than a constraint. The same regression evidence is therefore consistent with two different worlds, one in which slack binds and one in which it insures, and the investment regression alone cannot separate them. Separating them requires evidence beyond firm-level investment regressions. The within-firm facts of Section 3 provide it. No rationing account predicts that slack concentrates in the volatile product lines, or that the capacity allocation responds to product-specific risk while ignoring firm-wide risk.

TABLE 9: Optimal Investment and Slack

| | (1) $I/K$ | (2) $I/K$ | (3) $I/K$ |
|---|---|---|---|
| Average $q$ | 0.106 | 0.078 | 0.079 |
| Asset utilization $AU = \sum_i \text{sales}_i / K$ | | 0.046 | |
| product 1, $\text{sales}_1 / K$ | | | 0.051 |
| product 2, $\text{sales}_2 / K$ | | | 0.009 |
| $R^2$ | 0.769 | 0.843 | 0.851 |

*Notes:* Investment $I/K$ is regressed on average $q$ (col. 1), then adding firm asset utilization $AU = \sum_i \text{sales}_i / K$ (col. 2, GI's proxy for expected excess capacity), then replacing $AU$ with its product-level components $\{\text{sales}_i / K\}$ (col. 3). Adding $AU$ raises $R^2$ and enters positively, reproducing GI's excess-capacity control; the product-level components add almost nothing over firm $AU$, so firm-level asset utilization is a near-sufficient statistic for investment even with multiple products. $\Delta R^2$ from adding firm $AU$ (col. 2 vs. 1): $+0.074$; from the product-level components over firm $AU$ (col. 3 vs. 2): $+0.008$. Simulated firm-year panel; coefficients shown ($t$-statistics uninformative at this sample size).

Product-level data might be expected to improve these regressions. They do not for the following reason. Replacing firm-level asset utilization with its product-level components $\{\text{sales}_i / K\}$, entered separately, raises the $R^2$ by only 0.008 (Table 9). Firm-level asset utilization is thus a near-sufficient statistic for investment even with multiple product lines. At the optimal utilization rule, each product's operating profit is proportional to its revenue with a common factor of proportionality, so $MPK_t \propto \sum_i \text{sales}_{it} / K_t$ (Appendix IA.D.4). The firm's own optimization compresses the within-firm composition of demand into the firm-level sales-to-capital ratio, a multi-product, Hayashi-type aggregation result. Two implications follow. Firm-level asset-utilization corrections are robust to multi-product firms, because the within-firm composition of demand does not enter the mapping from utilization to the marginal product of capital.

Figure 5 turns to the second question. As we scale up the volatility of demand, holding each product's mean fixed, marginal $q$ stays flat, pinned by the investment first-order condition at $(1 + r)(1 + \theta\delta)$, while average $q$ falls because volatility de-

presses expected profits and hence firm value. The average–marginal gap therefore *narrows* as slack rises (Panel B). The implication is a reversal of the usual reading. When slack is chosen, average $q$ is a better proxy for marginal $q$ precisely in the high-slack firm-years where the rationing view says it is worst. Utilization-based corrections to $q$ therefore carry opposite interpretations in the two worlds, and applying a rationing-style correction to firms whose slack is insurance adjusts $q$ in the wrong direction. For the firms we study—where the placement of slack identifies it as insurance—the appropriate reading of a low-utilization firm-year is not that average $q$ overstates investment opportunities, but that the firm has bought protection against the volatility that also depresses its value.

TABLE 10: The Average−Marginal $q$ Wedge, Utilization, and Demand Volatility

| Object | Value |
|---|---|
| Mean firm utilization, low volatility | 0.630 |
| Mean firm utilization, high volatility | 0.338 |
| Marginal $q$, low volatility | 1.155 |
| Marginal $q$, high volatility | 1.148 |
| Wedge (avg − marginal $q$), low volatility | 3.473 |
| Wedge (avg − marginal $q$), high volatility | 2.180 |
| Corr. $(q^{avg} - q^{marg}, u)$, baseline panel | 0.358 |
| Corr. $(q^{avg}, u)$, baseline panel | 0.575 |

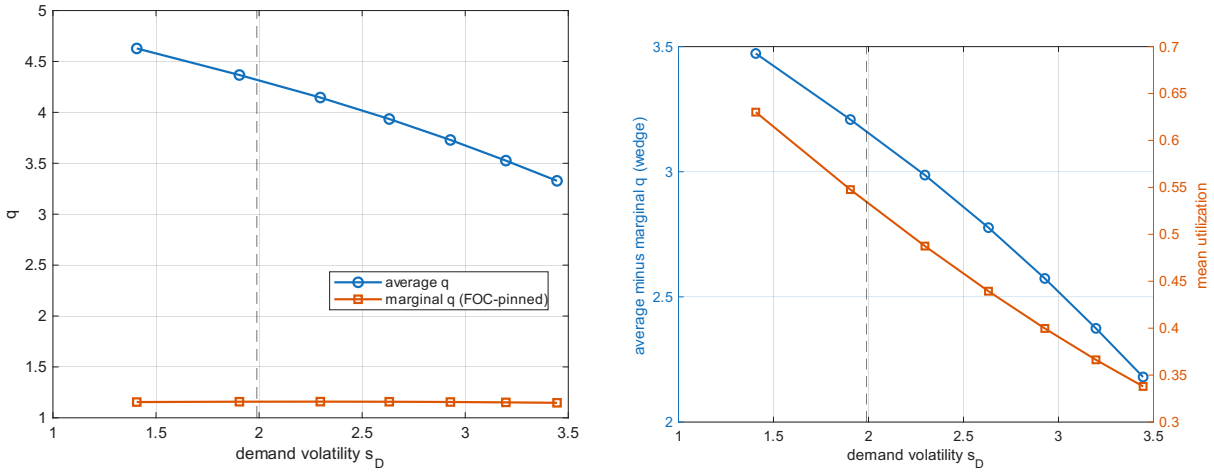
*Notes:* The first six rows are cross-steady-state means from the demand-volatility sweep. Marginal $q$ is pinned by the investment first-order condition while average $q$ falls with volatility, so the wedge *narrows* as excess capacity rises. The final two rows are raw pooled Pearson correlations across the baseline simulated firm-year panel, where firm utilization is $u = \sum_i \omega_i u_i$.

TABLE 11: The Average−Marginal $q$ Wedge, Utilization, and Demand Volatility

| Object | Value |
|---|---|
| Mean firm utilization, low volatility | 0.630 |
| Mean firm utilization, high volatility | 0.338 |
| Marginal $q$, low volatility | 1.155 |
| Marginal $q$, high volatility | 1.148 |
| Wedge (avg − marginal $q$), low volatility | 3.473 |
| Wedge (avg − marginal $q$), high volatility | 2.180 |
| Corr. $(q^{avg} - q^{marg}, u)$, baseline panel | 0.358 |
| Corr. $(q^{avg}, u)$, baseline panel | 0.575 |

*Notes:* The first six rows are cross-steady-state means from the demand-volatility sweep. Marginal $q$ is pinned by the investment first-order condition while average $q$ falls with volatility, so the wedge *narrows* as excess capacity rises. The final two rows are raw pooled Pearson correlations across the baseline simulated firm-year panel, where firm utilization is $u = \sum_i \omega_i u_i$.

FIGURE 5: Average−Marginal $q$ wedge




(a) Average and marginal $q$

(b) Wedge and utilization

*Notes:* The average−marginal $q$ wedge as demand volatility $s_D$ rises (both the firm-common and product-specific components scaled up; each product's mean held fixed). Average $q$ falls with volatility while marginal $q$ stays flat, pinned by the investment first-order condition. The wedge narrows even as mean utilization falls, i.e. as excess capacity rises. These panels show cross-steady-state means; Table 11 separately reports the pooled within-panel correlations. The dashed line marks the benchmark calibration.

# 5 Conclusion

Idle capacity is typically viewed as a symptom of underlying distortions such as weak demand, rigidities, or misallocated resources. Our results suggest that an important component of idle capacity instead reflects the optimal allocation of productive capacity under demand uncertainty. In the model, slack serves as operational insurance by allowing firms to accommodate product-specific fluctuations in demand. Using data that observe installed capacity and physical output for each product line, we show that slack is concentrated where demand risk resides. Within the same firm, more volatile product lines operate with lower utilization and receive smaller capacity shares, while the firm's internal allocation responds to product-specific risk but not to firm-wide risk—the component that cannot be insured through reallocation. A model in which the firm commits capacity across product lines before demand is realized rationalizes these patterns and quantifies their value. At our estimates, reallocating committed capacity toward currently strong product lines raises long-run profits by 3.1%, equivalent to 2.0% of the capital stock.

Viewing slack as an optimal choice rather than a distortion changes how familiar empirical diagnostics should be interpreted. Within-firm dispersion in utilization and in the revenue product of capital is, in part, a consequence of demand-risk management rather than evidence of inefficient capital allocation. Further, firm-level volatility is an equilibrium outcome rather than a primitive. Firms facing identical ag-

gregate demand risk can exhibit markedly different volatility depending on whether that demand risk is common across products or product-specific–a distinction that can only be identified with product-level data. Likewise, the gap between average and marginal *q narrows* as optimal slack increases, the opposite prediction of demand-rationing models of idle capacity. This provides a simple sign-based diagnostic for distinguishing slack that reflects operational insurance from slack that reflects binding demand constraints.

Our framework deliberately abstracts from financing frictions, so the insurance value we estimate should be interpreted as a lower bound for firms that value smoother cash flows in their own right. Integrating operational flexibility with financial hedging is therefore a natural next step. The analysis also treats the firm's product portfolio and organizational boundary as given: capacity is fungible across product lines within the firm but not across firms. Our analysis takes the firm's product portfolio as given and studies how capacity is allocated across existing product lines. Understanding how firms optimally choose the scope of their product portfolio when internal capacity reallocation itself creates value is a natural direction for future research.

# Bibliography


Abel, A. B. and O. J. Blanchard (1986). The present value of profits and cyclical movements in investment. *Econometrica 54*(2), 249–273.

Almeida, H., K. W. Hankins, and R. Williams (2017). Risk management with supply contracts. *The Review of Financial Studies 30*(12), 4179–4215.

Amberg, N., R. Friberg, and C. Syverson (2026). Have we got news for you: Theory and evidence on firms' optimal expected capacity utilization. *Working Paper*.

Argente, D., M. Lee, and S. Moreira (2024). The life cycle of products: Evidence and implications. *Journal of Political Economy 132*(2), 337–390.

Arrow, K. J., T. Harris, and J. Marschak (1951). Optimal inventory policy. *Econometrica 19*(3), 250–272.

Bassi, V., I. Rasul, O. A. Veroux, and A. Vitali (2024). Capital underutilization and the provision of firm training: Experimental evidence from uganda. Technical report, Working Paper.

Basu, S. and M. S. Kimball (1996). Procyclical productivity: Increasing returns or cyclical utilization? *The Quarterly Journal of Economics 111*(3), 719–751.

Bau, N. and A. Matray (2023). Misallocation and capital market integration: Evidence from India. *Econometrica 91*(1), 67–106.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010a). Multiple-product firms and product switching. *American Economic Review 100*(1), 70–97.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010b). Multiple-product firms and product switching. *American economic review 100*(1), 70–97.

Berndt, E. R. and C. J. Morrison (1981). Capacity utilization measures: Underlying economic theory and an alternative approach. *American Economic Review 71*(2), 48–52.

Bertrand, M., P. Mehta, and S. Mullainathan (2002). Ferreting out Tunneling: An Application to Indian Business Groups. *Quarterly Journal of Economics 117*(1), 121–148.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Butters, R. A. (2019). On demand uncertainty in the newsvendor model. *Economics Letters 185*, 108746.

Butters, R. A. (2020). Demand volatility, adjustment costs, and productivity: An examination of capacity utilization in hotels and airlines. *American Economic Journal: Microeconomics 12*(4), 1–44.

Coase, R. H. (1937). The nature of the firm. *Economica 4*(16), 386–405.

Corrado, C. and J. Mattey (1997). Capacity utilization. *Journal of Economic Perspectives 11*(1), 151–167.

Crouzet, N. and J. Eberly (2023). Rents and intangible capital: A q+ framework. *The Journal of Finance 78*(4), 1873–1916.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016). Prices, markups, and trade reform. *Econometrica 84*(2), 445–510.

Eckel, C. and J. P. Neary (2010). Multi-product firms and flexible manufacturing in the global economy. *Review of Economic Studies 77*(1), 188–217.

Erickson, T. and T. M. Whited (2000). Measurement error and the relationship between investment and q. *Journal of Political Economy 108*(5), 1027–1057.

Foster, L., J. Haltiwanger, and C. Krizan (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review 98*(1), 394–425.

Froot, K. A., D. S. Scharfstein, and J. C. Stein (1993). Risk management: Coordinating corporate investment and financing policies. *The Journal of Finance 48*(5), 1629–1658.

Giroud, X. and H. M. Mueller (2015). Capital and labor reallocation within firms. *The Journal of Finance 70*(4), 1767–1804.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010). Multiproduct firms and product turnover in the developing world: Evidence from India. *Review of Economics and Statistics 92*(4), 1042–1049.

Gopalan, R., A. Mukherjee, and M. Singh (2016). Do debt contract enforcement costs affect financing and asset structure? *Review of Financial Studies 29*(10), 2774–2813.

Gopalan, R., V. Nanda, and A. Seru (2007). Affiliated firms and financial support: Evidence from Indian business groups. *Journal of Financial Economics 86*(3), 759–795.

Gorodnichenko, Y. and M. D. Shapiro (2011). Using the survey of plant capacity to measure capital utilization. *US Census Bureau Center for Economic Studies Paper No. CES-WP-11-19*.

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

Grossman, S. J. and O. D. Hart (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy 94*(4), 691–719.

Grullon, G. and D. L. Ikenberry (2025). Excess capacity, marginal q, and corporate investment. *Journal of Finance 80*(3), 1533–1592.

Hann, R. N., M. Ogneva, and O. Ozbas (2013). Corporate diversification and the cost of capital. *The Journal of Finance 68*(5), 1961–1999.

Hart, A. G. (1940). *Anticipations, Uncertainty, and Dynamic Planning*. Chicago: University of Chicago Press.

Hart, O. and J. Moore (1990). Property rights and the nature of the firm. *Journal of Political Economy 98*(6), 1119–1158.

Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica 50*(1), 213–224.

Hottman, C. J., S. J. Redding, and D. E. Weinstein (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics 131*(3), 1291–1364.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics 124*(4), 1403–1448.

Kabir, P., E. Tan, and I. Vardishvili (2024). Quantifying the allocative efficiency of capital: The role of capital utilization. Working paper.

Kehrig, M. and N. Vincent (2026). Good dispersion, bad dispersion. *Review of Economic Studies 93*(2), 1103–1136.

Kim, Y. C. and G. C. Winston (1974). The optimal utilization of capital stock and the level of economic development. *Economica 41*(164), 377–386.

Lewellen, W. G. (1971). A pure financial rationale for the conglomerate merger. *The Journal of Finance 26*(2), 521–537.

Licandro, O. (1992). Investment dynamics and capacity utilization under monopolistic competition. *Annales d'Économie et de Statistique* (27), 91–113.

Lilienfeld-Toal, U. v., D. Mookherjee, and S. Visaria (2012). The distributive impact of reforms in credit enforcement: Evidence from Indian debt recovery tribunals. *Econometrica 80*(2), 497–558.

Maksimovic, V. and G. Phillips (2002). Do conglomerate firms allocate resources inefficiently across industries? theory and evidence. *The Journal of Finance 57*(2), 721–767.

Matvos, G. and A. Seru (2014). Resource allocation within firms and financial market dislocation: Evidence from diversified conglomerates. *The Review of Financial Studies 27*(4), 1143–1189.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2014). Market size, competition, and the product mix of exporters. *American Economic Review 104*(2), 495–536.

Naaraayanan, S. L. and K. M. Nielsen (2021). Does personal liability deter individuals from serving as independent directors? *Journal of Financial Economics 140*(2), 621–643.

Naaraayanan, S. L. and D. Wolfenzon (2024). Business group spillovers. *The Review of Financial Studies 37*(1), 231–264.

Peters, R. H. and L. A. Taylor (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics 123*(2), 251–272.

Philippon, T. (2009). The bond market's q. *Quarterly Journal of Economics 124*(3), 1011–1056.

Precious, M. (1985). Demand constraints, rational expectations and investment theory. *Oxford Economic Papers 37*(4), 576–605.

Scharfstein, D. S. and J. C. Stein (2000). The dark side of internal capital markets: Divisional rent-seeking and inefficient investment. *The Journal of Finance 55*(6), 2537–2564.

Shapiro, M. D. (1989). Assessing the federal reserve's measures of capacity and utilization. *Brookings Papers on Economic Activity 1989*(1), 181–241.

Siegel, J. and P. Choudhury (2012). A reexamination of tunneling and business groups: New data and new methods. *Review of Financial Studies 25*(6), 1763–1798.

Stein, J. C. (1997). Internal capital markets and the competition for corporate resources. *The Journal of Finance 52*(1), 111–133.

Stigler, G. J. (1939). Production and distribution in the short run. *Journal of Political Economy 47*(3), 305–327.

Walker, M. W., N. Shah, E. Miguel, D. Egger, F. Samy Soliman, and T. Graff (2024). Slack and economic development. Working Paper 33055, National Bureau of Economic Research.

# Internet Appendix

*The Value of Slack: Capacity Utilization in Multi-Product Firms*

Nuno Clara[1], Sampreet Singh Goraya[2], S. Lakshmi Naaraayanan[3]

[1]Duke University
[2]Stockholm School of Economics
[3]London Business School and CEPR

# Appendices

## IA.A Additional figures and tables

This section collects supplementary evidence on the representativeness of the sample and the behavior of the estimated price-cost objects. Figures IA.A.1 and IA.A.2 describe the industry composition of the product-level data, reporting each two-digit manufacturing sector's share of unique products and of total output. The product dimension is not concentrated in a handful of industries but spans the manufacturing cross-section, so the facts in the main text are not driven by a single sector. Figure IA.A.3 compares average capacity utilization in our sample with the Reserve Bank of India's survey, which covers a broader set of manufacturing firms; the two series display similar levels and cyclical movements.

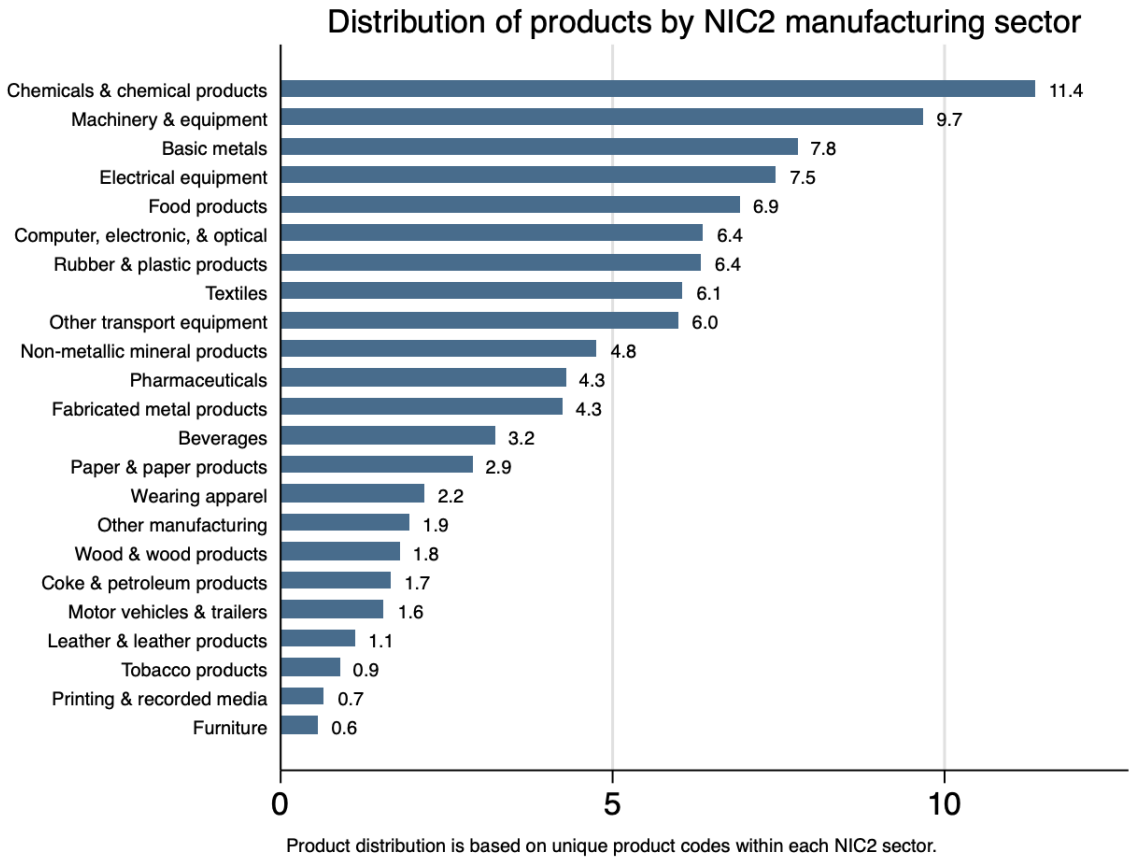



FIGURE IA.A.1: Industry Distribution: Product Coverage

*Notes:* This figure reports the distribution of products across NIC2 manufacturing sectors in the sample. For each sector, it plots the share of unique product codes assigned to that sector out of all unique products observed in the manufacturing sample. This figure shows how broadly product coverage is distributed across industries and helps document whether the product-level data are concentrated in a few sectors or span a wide cross-section of manufacturing activity.

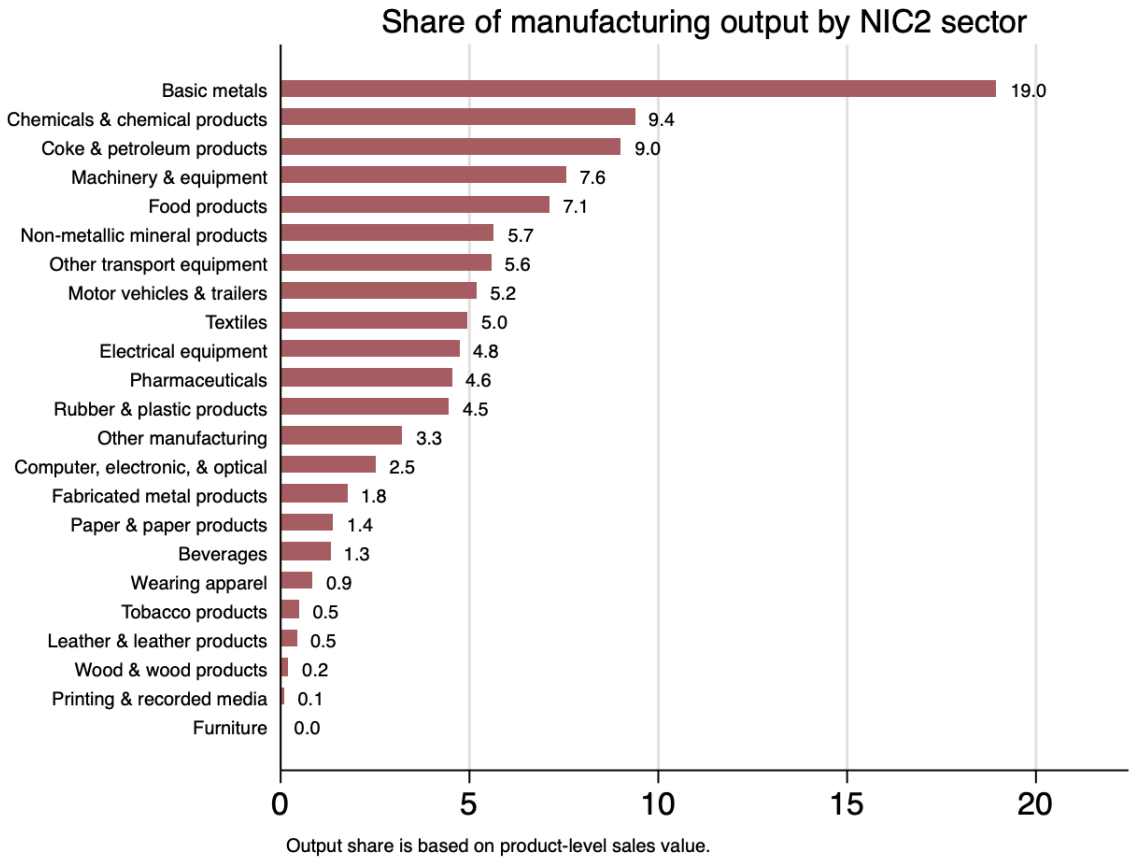



FIGURE IA.A.2: Industry Distribution: Output Share

*Notes:* This figure reports each NIC2 manufacturing sector's share of total manufacturing output in the sample. Output is measured using product-level sales value, aggregated within each NIC2 sector and divided by total sales value across all manufacturing sectors. This figure complements the product-count distribution by showing the economic weight of each sector, rather than simply the number of products observed.

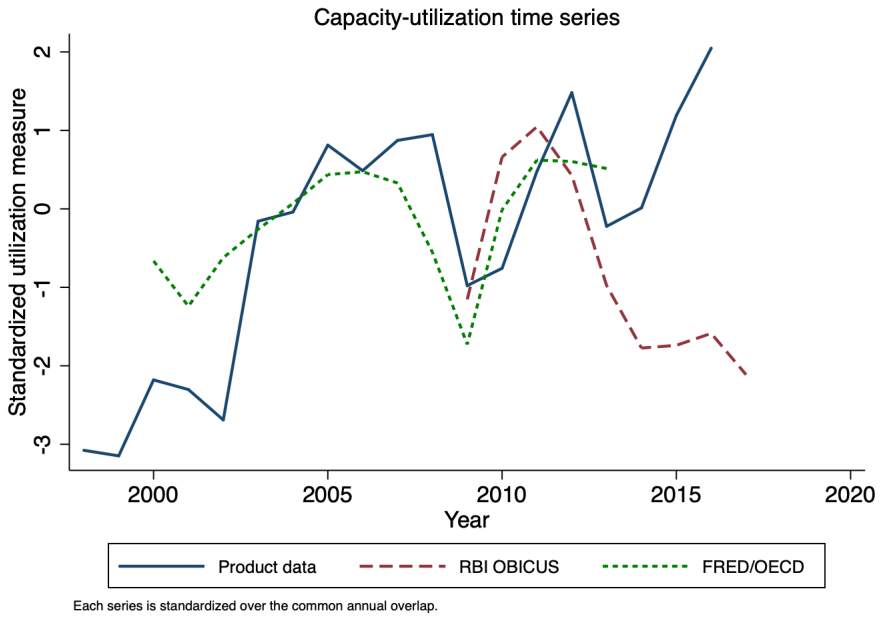



FIGURE IA.A.3: Capacity Utilization

*Notes:* The solid line reports the annual mean of firm-product capacity utilization in the product-level data, defined as as the ratio of actual output to installed capacity at the firm-product-year level. Dashed and short-dashed lines report annual averages from the RBI OBICUS survey and the FRED/OECD aggregate capacity-utilization series, respectively. Each series is standardized using the mean and standard deviation computed over the common annual overlap. The close comovement across series supports the use of the product-level measure as an indicator of capacity utilization.

## IA.B Estimating Quantity-Based Productivity

Estimating productivity is difficult in environments where firms face frictions in adjusting inputs. Capital constraints, unreliable electricity, infrastructure bottlenecks, and labor regulations can all weaken the link between observed input choices and underlying productivity. Standard proxy-variable approaches, such as Levinsohn and Petrin (2003), rely on the idea that firms can adjust intermediate inputs in response to productivity shocks. These methods typically recover revenue-based productivity, TFPR. While TFPR is informative, it combines physical efficiency with other forces that affect revenues, including markups, product quality, and product mix.

At the same time, revenue-based measures are not necessarily undesirable. If higher-quality products command higher prices, then prices and revenues may capture economically meaningful differences in firms' ability to produce quality, as emphasized by Atkin, Khandelwal, and Osman (2019). For this reason, we follow the methodology of De Loecker, Goldberg, Khandelwal, and Pavcnik (2016), which provides a way to estimate quantity-based productivity while allowing for systematic differences in product quality. The approach is flexible enough to accommodate a broad class of demand systems and does not require imposing a particular degree of pass-through from input prices to output prices. It also yields estimates of markups and marginal costs at the firm-product-year level.

The key idea is to use variation in input prices to proxy for differences in product quality. We model output quality as an increasing function of output prices, product market shares, and product indicators. This is motivated by the fact that producers selling higher-priced goods tend to use higher-priced inputs, a pattern documented, for example, by Kugler and Verhoogen (2012). Following De Loecker et al. (2016), we also assume that quality is correlated across inputs. Intuitively, producing a high-quality good generally requires not only better materials, but also complementary labor and capital. This lets us summarize input-quality differences using a single firm-product-level quality index.[9]

### IA.B.1 Assumptions

The estimation procedure relies on a set of simplifying assumptions, following De Loecker et al. (2016). These assumptions allow us to recover quantity productivity, marginal costs, and markups using the available data.

1. Producers of a given product share the same production technology, although their productivity may differ.
2. A firm has the same productivity level across all of the products it manufactures.
3. In the short run, firms adjust output through materials, while labor and capital are fixed or slow-moving.

[9] Because input prices are strongly related to output prices and other measures of product quality, we use output quality to control for input price variation across firms. The maintained assumption is that input quality is positively correlated across factors of production, so that a single product-quality index can capture systematic variation in the prices of materials, labor, and capital used to produce a given product.

4. Firms minimize contemporaneous short-run costs, taking current output and input prices as given.

5. Production function parameters are stable over the sample period.

6. More productive firms are more likely to operate as multi-product producers.

The second assumption is clearly restrictive, but it is commonly used because it makes it possible to estimate markups for firms that produce multiple goods. The third assumption is needed because labor and capital are observed only at the firm-year level, not separately by product. This assumption does not rule out longer-run substitution between capital and labor; it only implies that, within the short-run markup calculation, materials are treated as the adjustable input.

Under the assumptions that materials are the variable input and that firms minimize costs, markups can be written as the ratio of the output elasticity of materials to the materials share of revenue. The cost-minimization assumption also implies that input prices are not affected by the quantity of inputs purchased. This rules out static input-market power, such as monopsony power. As a result, the level of estimated markups may be biased downward. The estimates are therefore most useful for studying changes in markups rather than their absolute level. Finally, the stability of production coefficients is imposed because the data do not support estimating separate production functions for different subperiods.

### IA.B.2 Sources of bias and empirical corrections

Two empirical challenges are central for estimating TFPQ in our setting. First, for firms that make multiple products, we do not observe how inputs are allocated across products. Second, product quality is unobserved.

We address the first issue by estimating the production function using only single-product firms. This avoids the need to allocate inputs across products when recovering the production technology. However, single-product status is itself endogenous. Firms choose whether to produce one or several products, and this decision is likely related to productivity. To correct for this selection problem, we use the assumption that product scope increases with productivity. We estimate the probability that a firm remains a single-product producer as a function of its expected productivity and state variables, including product scope, materials, location, and other exogenous characteristics. The assumption that single- and multi-product firms share the same product-specific technology then allows us to apply the production function estimates from single-product firms to multi-product firms.

The second issue is that product quality is not directly observed. If higher-quality firms produce goods that differ systematically from lower-quality firms within the same product category, then quantity-based productivity estimates can be biased downward when quality is omitted.[10] We handle this by using input prices as a proxy for input quality, and by linking input quality to output quality. Because product-level input use is not observed for multi-product firms, the production function is again

[10]TFPR partly captures cross-sectional quality differences because prices enter revenues. However, prices also reflect markups and demand-side conditions, so TFPR can confound productivity, quality, and market power. The sign and size of the resulting bias depend on the setting.

estimated on the single-product subsample. We then use the estimated relationship between quality, prices, market shares, and product characteristics to control for input price differences across firms.[11]

The implementation proceeds in four steps:

1. We estimate product-specific production functions using single-product firms. The production technology is specified as translog, and the estimates are used to recover the output elasticity of materials.

2. We correct for selection into single-product production. Specifically, we estimate the productivity cutoff above which firms expand into multiple products, and control for the probability that a firm remains below that cutoff in a given year as a function of predicted productivity and state variables.

3. For multi-product firms, we impute product-level materials shares using the production function estimates from single-product firms. The allocation of materials across products is modeled as a function of firm-product-year output prices, product market shares, product and location fixed effects, and export status.

4. We compute markups and marginal costs at the firm-product-year level. A product's markup is the ratio of the output elasticity of materials to the materials share of revenue. Marginal cost is then obtained by dividing the product price by the estimated markup.

This procedure gives us quality-adjusted estimates of TFPQ, markups, and marginal costs while allowing for flexible differences in input and output quality across firms.

### IA.B.3 Validation exercises

We conduct several diagnostic checks to assess whether the estimates line up with basic economic intuition and with patterns documented in prior work. Figure IA.B.1 plots the relationship between output quantities and our estimated markups and marginal costs after removing product-year fixed effects.

The left panel shows that firms producing larger quantities tend to have higher markups. This is consistent with the idea that high-output firms have lower marginal costs and can sustain larger markups. The right panel relates marginal costs to quantities. Our estimated production elasticities indicate that many firms operate under increasing returns to scale, a feature also documented by De Loecker et al. (2016). Consistent with this pattern, marginal costs decline with the quantity produced.

We also evaluate whether applying production function estimates from single-product firms to multi-product firms produces plausible within-firm patterns. Figure IA.B.2 examines how markups and marginal costs vary across products within the same firm. To make products comparable within firms, we residualize markups and marginal costs using both product-year and firm-year fixed effects. We then relate these residualized outcomes to each product's share of the firm's sales.

[11] In the data, input prices increase with product quality. This supports the use of output-quality proxies, including output prices, product market shares, and product indicators, to account for input price differences across producers. This approach is consistent with evidence that firms producing more expensive goods tend to use more expensive inputs, as in Kugler and Verhoogen (2012).

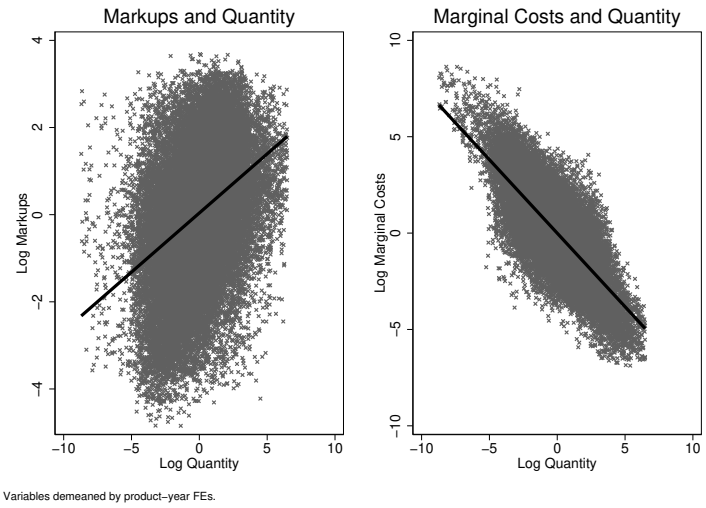



FIGURE IA.B.1: Marginal Costs, Markups, and Quantities

*Notes:* The left panel plots the relationship between the natural logarithm of product markups and output quantities. The right panel plots the relationship between the natural logarithm of product marginal costs and output quantities. Observations are at the firm-product-year level for 1992–2016. Markups, marginal costs, and quantities are residualized using product-year fixed effects so that comparisons are made across firms within the same product-year.

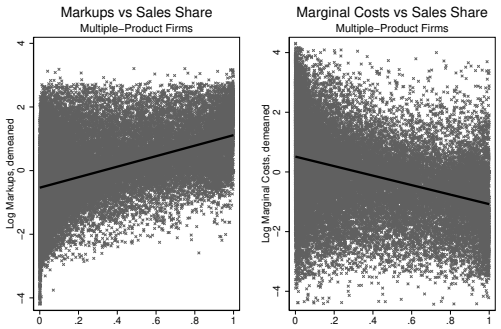



FIGURE IA.B.2: Markups, Costs and Product Sales Share

*Notes:* The left panel presents the correlation between the natural log of product markup and sales share. The right panel is between the natural log of product marginal cost and sales share. Data are at the firm-product-year level for the period 2005 to 2015. Data are winsorized at the 3rd and 97th percentiles. Markups, marginal costs, and quantities are demeaned by product-year and firm-year fixed effects to make them comparable across firms.

The patterns are consistent with the idea that multi-product firms have core products. Products that account for a larger share of firm sales tend to have higher markups and lower marginal costs. Conversely, products that are farther from the firm's main area of strength have lower markups and higher marginal costs. Although the estimation procedure does not impose a particular demand system or market structure, these correlations are in line with models of multi-product firms, including Eckel and Neary (2010), Mayer, Melitz, and Ottaviano (2014), and Melitz and Ottaviano (2008), as well as with the empirical evidence in De Loecker et al. (2016) for Indian manufacturing.

**Production Function Framework.** We estimate a gross output production function for each product line following the control function approach (Levinsohn and Petrin,

2003; Ackerberg et al., 2015). The production function takes the form:

$$y_{ijt} = \beta_l l_{ijt} + \beta_k k_{ijt} + \beta_m m_{ijt} + \omega_{ijt} + \varepsilon_{ijt} \tag{IA.B.1}$$

where $y_{ijt}$ is log output, $l_{ijt}$ is log labor, $k_{ijt}$ is log capital, $m_{ijt}$ is log materials, and $\omega_{ijt}$ is productivity. We use the proxy variable method to address the simultaneity of input choices and unobserved productivity.

**Recovering Material Shares.** From the production function estimation, we obtain estimates of the revenue function $\rho_{ijt}$, which captures the relationship between inputs and revenue. The material cost share for product $j$ in firm $i$ is then computed as:

$$\omega_{ijt}^{M} = \frac{P_{ijt}Q_{ijt}}{\exp(\hat{\rho}_{ijt})} \tag{IA.B.2}$$

where $P_{ijt}Q_{ijt}$ is product revenue (sales value) and $\hat{\rho}_{ijt}$ is the estimated revenue function from the production function estimation.

**Interpretation as Capacity Allocation.** The material share $\omega_{ijt}^{M}$ reflects the intensity of variable input usage for each product line. Under the assumption that material inputs scale with production intensity, products receiving a larger share of material inputs represent a greater allocation of the firm's productive capacity.

We winsorize the material shares at the 5th and 95th percentiles to limit the influence of outliers, yielding our final capacity share measure $\omega_{ijt}$.

**Remark.** An alternative interpretation relates to the markup estimation literature (Loecker and Warzynski, 2012). Under cost minimization, the material cost share relates to the output elasticity of materials and the markup:

$$\omega_{ijt}^{M} = \frac{\theta_{ijt}^{M}}{\mu_{ijt}} \tag{IA.B.3}$$

where $\theta_{ijt}^{M}$ is the output elasticity of materials and $\mu_{ijt}$ is the markup. Our capacity share measure thus also captures variation in production intensity relative to market power across product lines.

# IA.C Measurement of Demand Risk

This appendix details the construction of the demand-risk volatilities in Section 2, states its identifying assumption, and reports robustness of the decomposition.

**Identifying assumption.** The covariance identity behind equation (4) rests on the assumption that the product-specific components $\nu_{ijt}$ are uncorrelated across products within the firm. If they are correlated, the pair covariance identifies $s^2_{f,j} + \mathrm{Cov}(\nu_i, \nu_k)$ rather than $s^2_{f,j}$. Negative correlation biases the firm component downward and can produce negative estimates even when a common component exists. Two considerations support proceeding under the assumption. First, the model's allocation rule depends only on each product's marginal volatility, so cross-product correlation does not affect the prediction being tested. Second, the economic content of the decomposition is the separation of risk that internal reallocation cannot absorb from risk that it can; negatively correlated product risk belongs in the second category, so an estimator that assigns it to the product-specific component classifies it correctly for this purpose. A truncated zero should accordingly be read as the absence of positively comoving common risk, not necessarily the absence of any firm-wide shock.

**Measurement error.** Product-specific measurement error—misspecified demand elasticity, unit-value price error, and noise in physical quantities—enters each product's own variance one for one but never enters the cross-product covariance. The firm-component estimate is therefore robust to it, while the product-specific variance is overstated in level, and as a noisy regressor its coefficients are attenuated toward zero. Both forces work against the patterns documented in Facts 2 through 4, so the estimates are conservative in this respect. Both volatilities are generated regressors, so the clustered standard errors do not account for the estimation error in the volatilities themselves.

**Robustness of the decomposition.** The decomposition in Fact 4 requires a minimum number of overlapping years per product pair; the baseline uses five. Table IA.C.1 rebuilds the volatilities requiring five and ten years. Signs and significance are unchanged, and the share of firms with a truncated firm-component variance is essentially flat across different time frames, which rules out short panels as the main source of the truncation. Dropping the truncated firms instead of keeping them leaves the point estimates close to the baseline.

# IA.D Derivations

## IA.D.1 Optimal Utilization Rule (Equation 21)

The first-order condition for $\Pi_{ijt} = D_{ijt}^{1/\eta}(u_{ijt}K_{jt}\omega_{ijt})^{1-1/\eta} - \beta K_{jt}\omega_{ijt}u_{ijt}^{\gamma}$ with respect to $u_{ijt}$ is

$$(1 - 1/\eta)\, D_{ijt}^{1/\eta}\, (K_{jt}\omega_{ijt})^{1-1/\eta}\, u_{ijt}^{-1/\eta} = \gamma\beta\, u_{ijt}^{\gamma-1}\, K_{jt}\omega_{ijt}.$$

TABLE IA.C.1: Fact 4 Sensitivity to the Minimum-Years Screen

| | Utilization | | Capacity Share | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Firm volatility ($\hat{s}_f$) | -0.1062*** | -0.0783* | 0.0528 | 0.1089 |
| | (0.0343) | (0.0441) | (0.0784) | (0.1065) |
| Idiosyncratic volatility ($\hat{s}_\varepsilon$) | -0.0559*** | -0.0447** | -0.1014*** | -0.1194** |
| | (0.0155) | (0.0191) | (0.0290) | (0.0502) |
| Minimum years | 5 | 10 | 5 | 10 |
| $R^2$ | 0.34 | 0.37 | 0.49 | 0.54 |

*Notes:* Each column rebuilds the covariance-based volatilities requiring the indicated minimum number of overlapping years per product pair and per product, then reruns the specification with both volatilities and product-year fixed effects. Columns (1) to (3) use utilization as the outcome and columns (4) to (6) the capacity share. Standard errors clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Rearranging:

$$u_{ijt}^{\gamma-1+1/\eta} = \frac{\eta-1}{\gamma\beta\eta} D_{ijt}^{1/\eta} (K_{jt}\omega_{ijt})^{-1/\eta}.$$

Note $\gamma - 1 + 1/\eta = [\eta(\gamma-1)+1]/\eta = 1/(\eta\psi)$. Raising both sides to the power $\eta\psi$:

$$u_{ijt}^{*} = \left(\frac{\eta-1}{\gamma\beta\eta}\right)^{\eta\psi} \left(\frac{D_{ijt}}{K_{jt}\omega_{ijt}}\right)^{\psi} \equiv A \left(\frac{D_{ijt}}{K_{jt}\omega_{ijt}}\right)^{\psi},$$

where $A \equiv \left(\frac{\eta-1}{\gamma\beta\eta}\right)^{\eta\psi}$. ■

## IA.D.2 Proof of Proposition 1

The capacity plan for period $t$ is chosen at the end of period $t-1$, conditional on the state $s_{j,t-1}$. Conditional expected operating profit is separable across products: $\mathbb{E}_{t-1}[\pi_{jt}] = \sum_i \mathbb{E}_{t-1}[\Pi_{ijt}]$, each evaluated at the optimal utilization rule of Appendix IA.D.1. At that rule, the utilization first-order condition implies $(1-1/\eta)R_{ijt} = \gamma C_{ijt}$ state by state, hence $(1-1/\eta)\mathbb{E}_{t-1}[R_{ijt}] = \gamma\,\mathbb{E}_{t-1}[C_{ijt}]$.

By the envelope theorem,

$$\frac{\partial\,\mathbb{E}_{t-1}[\Pi_{ijt}]}{\partial\omega_{ijt}} = \frac{(\gamma-1)}{\omega_{ijt}}\,\mathbb{E}_{t-1}[C_{ijt}].$$

Equalizing marginal contributions across products subject to $\sum_i \omega_{ijt} = 1$ gives $\mathbb{E}_{t-1}[C_{ijt}]/\omega_{ijt} = \mathbb{E}_{t-1}[C_{kjt}]/\omega_{kjt}$ for all $i, k$. Since $\mathbb{E}_{t-1}[C_{ijt}] = \beta K_{jt}\omega_{ijt}\,\mathbb{E}_{t-1}[u_{ijt}^{*\gamma}]$ and $\mathbb{E}_{t-1}[u_{ijt}^{*\gamma}] \propto \omega_{ijt}^{-\phi} M_{ijt|t-1}$ with $M_{ijt|t-1} \equiv \mathbb{E}_{t-1}[D_{ijt}^{\phi}]$ (using $\gamma\psi = \phi$), the condition becomes $M_{ijt|t-1}/\omega_{ijt}^{\phi} = M_{kjt|t-1}/\omega_{kjt}^{\phi}$. Combining with the adding-up constraint:

$$\omega_{ijt} = \Omega_i(s_{j,t-1}) = \frac{M_{ijt|t-1}^{1/\phi}}{\sum_k M_{kjt|t-1}^{1/\phi}}.$$ ■

**Lognormal case (Lemma 1).** In the steady-state benchmark, replace the conditional moment with its unconditional counterpart $M_i = \mathbb{E}[D_i^{\phi}]$. Under the shock structure (14), $\ln D_{ijt}$ is Gaussian with mean $\mu_{ij}$ and variance $s_{D,i}^2 = s_f^2 + s_{\varepsilon,i}^2$, so

$$M_i = \exp\left(\phi\mu_{ij} + \tfrac{1}{2}\phi^2 s_{D,i}^2\right), \qquad M_i^{1/\phi} = \exp\left(\mu_{ij} + \tfrac{1}{2}\phi\, s_{D,i}^2\right).$$

In the share ratio, the common factor $\exp(\frac{1}{2}\phi\, s_f^2)$ cancels across products, leaving

$$\omega_i^* \propto \exp\left(\mu_{ij} + \tfrac{1}{2}\phi\, s_{\varepsilon,i}^2\right).$$

Holding mean demand $\bar{D}_i = \mathbb{E}[D_i] = \exp(\mu_{ij} + \frac{1}{2}s_{D,i}^2)$ fixed, substitute $\mu_{ij} = \ln \bar{D}_i - \frac{1}{2}s_{D,i}^2$:

$$\omega_i^* \propto \bar{D}_i \cdot \exp\left(-\tfrac{1}{2}(1-\phi)\, s_{\varepsilon,i}^2\right),$$

where terms involving only $s_f^2$ again cancel in the ratio. ■

## IA.D.3 Proof of Proposition 2

The proof proceeds in four steps: (1) express operating profit in closed form; (2) solve for steady-state capital; (3) derive steady-state utilization at the optimal allocation; (4) take expectations under lognormal demand.

**Step 1: Profit representation.** At the optimal utilization rule, $(1-1/\eta)R_{ijt} = \gamma C_{ijt}$, so $\Pi_{ijt} = R_{ijt}\left(1 - \frac{1-1/\eta}{\gamma}\right)$. Substituting $u_{ijt}^*$ into $R_{ijt} = D_{ijt}^{1/\eta}(u_{ijt}K_{jt}\omega_{ijt})^{1-1/\eta}$, the exponent on $D_{ijt}$ is $1/\eta + \psi(\eta-1)/\eta = \phi$ and the exponent on $K_{jt}\omega_{ijt}$ is $(1-1/\eta)(1-\psi) = 1-\phi$. Total operating profit is therefore

$$\pi_{jt} = \alpha_1\, K_{jt}^{1-\phi}\, \mathcal{D}_{jt}, \qquad \mathcal{D}_{jt} \equiv \sum_i D_{ijt}^{\phi}\, \omega_{ijt}^{1-\phi},$$

where $\alpha_1 \equiv \left(1 - \frac{\eta-1}{\eta\gamma}\right)\left(\frac{\eta-1}{\gamma\beta\eta}\right)^{(\eta-1)/[\eta(\gamma-1)+1]}$. The marginal product of capital is $MPK_{jt} = \alpha_1(1-\phi)K_{jt}^{-\phi}\mathcal{D}_{jt}$.

**Step 2: Steady-state capital.** Differentiating the value function with respect to $K_{jt+1}$ gives the investment first-order condition $1 + \theta(i_{jt}/K_{jt}) = \frac{1}{1+r}\mathbb{E}_t[q_{jt+1}]$, and the envelope condition gives $q_{jt} = MPK_{jt} + (1-\delta)[1+\theta(i_{jt}/K_{jt})] + \frac{\theta}{2}(i_{jt}/K_{jt})^2$. In steady state $i_{ss}/K_{ss} = \delta$, so

$$q_{ss} = (1+r)(1+\theta\delta), \qquad MPK_{ss} = (1+\theta\delta)(r+\delta) - \tfrac{\theta}{2}\delta^2.$$

Inverting $MPK_{ss} = \alpha_1(1-\phi)K_{ss}^{-\phi}\mathcal{D}_{ss}$:

$$K_{ss} = \left[\frac{\alpha_1(1-\phi)\,\mathcal{D}_{ss}}{(1+\theta\delta)(r+\delta) - \frac{\theta}{2}\delta^2}\right]^{1/\phi}.$$

**Step 3: Steady-state utilization at the optimal allocation.** Substituting $K_{ss}$ into the utilization rule of Appendix IA.D.1, and using $\psi/\phi = 1/\gamma$:

$$u_i^{*,ss}(D_i \mid \omega) = \underbrace{A\,[\alpha_1(1-\phi)]^{-1/\gamma}\,[(1+\theta\delta)(r+\delta) - \tfrac{\theta}{2}\delta^2]^{1/\gamma}}_{\equiv\,\tilde{\Lambda}}\; D_i^{\psi}\,\omega_i^{-\psi}\,\mathcal{D}_{ss}^{-1/\gamma}. \tag{IA.D.1}$$

Imposing the optimal shares $\omega_i^* = M_i^{1/\phi}/S$ with $S \equiv \sum_k M_k^{1/\phi}$: first, $(\omega_i^*)^{-\psi} = S^{\psi} M_i^{-1/\gamma}$; second, evaluating the demand index at the optimal shares, $\mathcal{D}_{ss} = \sum_k M_k(\omega_k^*)^{1-\phi} = S^{-(1-\phi)}\sum_k M_k^{1/\phi} = S^{\phi}$, so $\mathcal{D}_{ss}^{-1/\gamma} = S^{-\psi}$. The $S$ terms cancel:

$$u_i^{*,ss}(D_i) = \tilde{\Lambda}\, D_i^{\psi}\, M_i^{-1/\gamma}.$$

**Step 4: Expectation under lognormal demand.** Under the shock structure (14), $\ln D_i$ is Gaussian with mean $\mu_i$ and variance $s_{D,i}^2 = s_f^2 + s_{\varepsilon,i}^2$, so

$$\mathbb{E}[D_i^{\psi}] = \exp\!\left(\psi\mu_i + \tfrac{1}{2}\psi^2 s_{D,i}^2\right), \qquad M_i = \exp\!\left(\phi\mu_i + \tfrac{1}{2}\phi^2 s_{D,i}^2\right).$$

Therefore

$$\mathbb{E}[u_i^{*,ss}] = \tilde{\Lambda}\, M_i^{-1/\gamma}\,\mathbb{E}[D_i^{\psi}] = \tilde{\Lambda}\exp\!\left[\left(\psi - \frac{\phi}{\gamma}\right)\mu_i + \tfrac{1}{2}\left(\psi^2 - \frac{\phi^2}{\gamma}\right)s_{D,i}^2\right].$$

Using $\phi = \gamma\psi$: the coefficient on $\mu_i$ is $\psi - \psi = 0$, so mean demand drops out; the coefficient on $s_{D,i}^2$ is $\psi^2 - \gamma\psi^2 = -(\gamma-1)\psi^2$. Hence

$$\mathbb{E}[u_i^{*,ss}] = \tilde{\Lambda}\cdot\exp\!\left(-\tfrac{1}{2}(\gamma-1)\psi^2\,(s_f^2 + s_{\varepsilon,i}^2)\right).$$

■

## IA.D.4 Asset Utilization and the Marginal Product of Capital

At the optimal utilization rule, each product's operating profit is proportional to its revenue with a common factor: $(1-1/\eta)R_{ijt} = \gamma C_{ijt}$ implies

$$\Pi_{ijt} = \left(1 - \tfrac{\eta-1}{\eta\gamma}\right)R_{ijt},$$

where the factor of proportionality depends only on $(\eta,\gamma)$ and is common across product lines. Summing across products, $\pi_{jt} = \left(1 - \frac{\eta-1}{\eta\gamma}\right)\sum_i R_{ijt}$. Moreover, for *any* capacity shares $\{\omega_{ijt}\}$ with $\sum_i \omega_{ijt} = 1$, substituting the utilization rule gives

$\pi_{jt} = \alpha_1 K_{jt}^{1-\phi} \mathcal{D}_{jt}$, so profit is homogeneous of degree $1-\phi$ in $K_{jt}$ and

$$MPK_{jt} = (1-\phi)\,\frac{\pi_{jt}}{K_{jt}} = (1-\phi)\left(1 - \tfrac{\eta-1}{\eta\gamma}\right)\frac{\sum_i \text{sales}_{ijt}}{K_{jt}}.$$

The contemporaneous marginal product of capital is therefore exactly proportional to the firm-level sales-to-capital ratio. Two features of this result bear emphasis. First, it does *not* require the capacity allocation to be optimal: the derivation holds for arbitrary shares, so even a firm with a distorted allocation has $MPK$ exactly proportional to its asset utilization. Second, it *does* require optimal utilization: the common proportionality between profit and revenue comes from the utilization first-order condition, so product-specific distortions on the operating margin break the result.

**Product-level revenue products and utilization.** The same first-order condition delivers a product-level counterpart. Dividing $(1-1/\eta)R_{ijt} = \gamma\beta\, k_{ijt}\, u_{ijt}^{\gamma}$ by product-level capacity $k_{ijt} \equiv K_{jt}\omega_{ijt}$ gives the average revenue product of capacity,

$$ARPK_{ijt} \equiv \frac{R_{ijt}}{k_{ijt}} = \frac{\gamma\beta}{1-1/\eta}\, u_{ijt}^{\gamma},$$

and, since revenue is concave in output under CES demand, the marginal revenue product is proportional to it: $MRPK_{ijt} \equiv \partial R_{ijt}/\partial k_{ijt} = (1-1/\eta)\, ARPK_{ijt} = \gamma\beta\, u_{ijt}^{\gamma}$. The marginal profit product satisfies the same power relation, $\partial\Pi_{ijt}/\partial k_{ijt} = (\gamma - 1)\beta\, u_{ijt}^{\gamma}$. All three objects are therefore log-linear in utilization with a common slope,

$$\ln ARPK_{ijt} = \gamma \ln u_{ijt} + \text{const},$$

where the constant depends only on $(\eta, \gamma, \beta)$ and differs across the three notions only in level. Because empirical measures of the revenue product of capital are typically constructed as sales over capital, the measured object is $ARPK$; the proportionality means the results below apply to it directly. Three consequences follow. First, a regression of product-level $\ln ARPK$ on $\ln u$ identifies $\gamma$ directly, providing an overidentification check on the structural estimate. Second, within a firm, $\text{Var}_i(\ln ARPK) = \gamma^2 \text{Var}_i(\ln u)$ (and identically for $MRPK$, since the two differ by a constant): measured revenue-product dispersion across product lines reflects the optimal utilization response to heterogeneous demand rather than allocative wedges. Third, like the asset-utilization result above, the relation requires only interior optimal utilization—it holds for arbitrary capacity shares and breaks only when the utilization margin itself is distorted.

## IA.D.5 Proof of Lemma 2

At a fixed allocation $\omega$ (not re-optimized as volatility changes), expected utilization is proportional to $\mathbb{E}[D_i^{\psi}]$ by the utilization rule of Appendix IA.D.1. Under lognormal demand with mean $\bar{D}_i$ held fixed, $\mu_i = \ln \bar{D}_i - \frac{1}{2}s_{D,i}^2$, so

$$\mathbb{E}[D_i^{\psi}] = \exp\left(\psi\mu_i + \tfrac{1}{2}\psi^2 s_{D,i}^2\right) = \bar{D}_i^{\psi}\, \exp\left(-\tfrac{1}{2}\psi(1-\psi)\, s_{D,i}^2\right).$$

Using $1 - \psi = \eta(\gamma - 1)\psi$, the coefficient is $\psi(1 - \psi) = \eta(\gamma - 1)\psi^2$: at fixed $\omega$, a mean-preserving spread lowers expected utilization at rate $\frac{1}{2}\eta(\gamma - 1)\psi^2$ per unit of $s^2_{D,i}$.

At the joint optimum, Proposition 2 gives a coefficient of $(\gamma - 1)\psi^2$—smaller by the factor $\eta$. The difference is the reallocation term: the optimal share falls with $s^2_{D,i}$ (Lemma 1), and since $u^* \propto \omega^{-\psi}$, the smaller share mechanically raises utilization, offsetting part of the direct effect. ■

# Bibliography


Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Atkin, D., A. K. Khandelwal, and A. Osman (2019). Measuring productivity: Lessons from tailored surveys and productivity benchmarking. In *AEA Papers and Proceedings*, Volume 109, pp. 444–449. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016). Prices, markups, and trade reform. *Econometrica 84*(2), 445–510.

Eckel, C. and J. P. Neary (2010). Multi-product firms and flexible manufacturing in the global economy. *Review of Economic Studies 77*(1), 188–217.

Kugler, M. and E. Verhoogen (2012). Prices, plant size, and product quality. *Review of Economic Studies 79*(1), 307–339.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies 70*(2), 317–341.

Loecker, J. D. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–2471.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2014). Market size, competition, and the product mix of exporters. *American Economic Review 104*(2), 495–536.

Melitz, M. J. and G. I. Ottaviano (2008). Market size, trade, and productivity. *Review of Economic Studies 75*(1), 295–316.