NBER WORKING PAPER SERIES

HAVE WE GOT NEWS FOR YOU:
THEORY AND EVIDENCE ON FIRMS' OPTIMAL EXPECTED CAPACITY UTILIZATION

Niklas Amberg
Richard Friberg
Chad Syverson

Working Paper 33400
http://www.nber.org/papers/w33400

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2025, Revised May 2026

For helpful comments and suggestions, we thank Christoph Boehm, Aaron Flaaen, and Hannes Malmberg, as well as seminar and conference participants at the Booth School of Business (University of Chicago), the Econometric Society World Congress 2025 (Seoul), EARIE 2025 (Valencia), the EEA 2025 Congress (Bordeaux), the Federal Reserve Bank of St. Louis, MACCI 2026, Norio 2025, Sveriges Riksbank, the Wharton School (University of Pennsylvania), and the World Bank. The opinions expressed in this article are the sole responsibility of the authors and should not be interpreted as reflecting the views of Sveriges Riksbank or the National Bureau of Economic Research. The authors have no financial interests relevant to this paper.

Have We Got News For You: Theory and Evidence on Firms' Optimal Expected Capacity Utilization
Niklas Amberg, Richard Friberg, and Chad Syverson
NBER Working Paper No. 33400
January 2025, Revised May 2026
JEL No. D20, L11, L23, L60

ABSTRACT

Using quarterly micro data on capacity utilization among Swedish manufacturing firms, we show that idiosyncratic factors are more important than aggregate influences in explaining variation in capacity utilization across firms and over time. Idiosyncratic does not mean unpredictable, however. A newsvendor model of optimal capacity predicts that higher demand uncertainty lowers expected capacity utilization, especially for high-markup firms. We find support for these predictions in data containing firm-specific, forward-looking measures of uncertainty: firms facing high uncertainty on average have seven percentage points lower capacity utilization than firms facing low uncertainty; among high-markup firms, the difference is nine percentage points.

Niklas Amberg
Sveriges Riksbank
niklas.amberg@riksbank.se

Richard Friberg
Norwegian School of Economics and
Stockholm School of Economics
nerf@hhs.se

Chad Syverson
The University of Chicago
Booth School of Business
and NBER
chad.syverson@chicagobooth.edu

A line of research in macroeconomics has explored how aggregate fluctuations both shape and are shaped by patterns in average capacity utilization in the economy (see, e.g., Greenwood, Hercowitz and Huffman, 1988, Michaillat and Saez, 2015, and Kuhn and George, 2019). However, much less work has explored how capacity utilization is determined at the microeconomic level, where capacity decisions are actually made. This knowledge deficit is sharpened by the fact that cross-sectional variation at the firm level far outweighs time-series variation in capacity utilization. We show below that firm fixed effects explain almost half the variation in capacity utilization, whereas industry-time fixed effects account for less than 10 percent.

The purpose of this paper is to shed light on capacity-utilization decisions at the firm level. We demonstrate that the cross-sectional variation in utilization is not simply unobserved heterogeneity or a reflection of noisy shocks realized after capacity decisions have been made. Firms deliberately hold idle capacity on average, and for measurable microeconomic reasons. More specifically, we show that firms' capacity-utilization levels are negatively related to the amount of demand uncertainty they face, and the more so the higher their markups are.

We establish these findings in three steps. We begin by using quarterly micro data from the largest business survey in Sweden—covering Swedish manufacturing firms' capacity utilization over the period 2010Q1–2024Q1—to document the following three stylized facts about capacity utilization at the firm level: (i) a majority of firms routinely operate below full capacity, (ii) the main reason for operating below full capacity is "insufficient demand," and (iii) the variation in capacity utilization across firms and over time is predominantly idiosyncratic.

We then formulate a newsvendor-style model (Arrow, Harris and Marschak, 1951) to generate testable predictions about this heterogeneity in capacity utilization. The newsvendor model is traditionally formulated in terms of inventory decisions, but it can easily be translated into a model of optimal capacity. The key element in this class of models is that firms need to make capacity (or inventory) decisions before demand is realized. They therefore face a trade-off: if the chosen capacity turns out to be too low relative to realized demand, an opportunity cost arises ex post due to foregone profits from demand the firm cannot satisfy; if the chosen capacity turns out to be too high, on the other hand, profits are hurt by the cost of maintaining unused capacity. Rather than focusing on capacity per se, we focus on *expected capacity utilization*, the equivalent of the expected share of inventories sold.

We derive two key testable predictions from the model. First, firms facing higher demand uncertainty operate with lower capacity utilization in expectation.[1] This result has been demon-

[1]A large theoretical literature shows that idle capacity also can be motivated by strategic considerations to optimally accommodate or deter entry (see Lieberman, 1987, for an early empirical application). This is, however,

strated previously by Butters (2019) using a newsvendor model with fixed markups. We generalize Butters's result by showing that it also holds in a setting with optimal pricing. Second, the effect of demand uncertainty on expected capacity utilization is stronger for firms with higher markups, as the opportunity cost of failing to satisfy demand increases in the markup.[2] We are not aware of any paper that has derived or tested this prediction before.

The third and final step of the analysis is to take the model predictions to the data. We do so using the aforementioned micro data, which comprise not only information about firms' capacity utilization, but also a firm-level measure of perceived, forward-looking demand uncertainty, which is critical when testing the predictions of the newsvendor model. The uncertainty measure is based on firms' answers to the survey question "Predicting the future development of our business situation is currently...," to which there are four possible answers: easy, fairly easy, fairly difficult, and difficult.

Our results show that the classic newsvendor framework explains very well patterns in capacity utilization at the firm level, both in the cross section and within firms over time. Even while controlling for contemporaneous demand shocks and other relevant firm conditions along with firm and time fixed effects, firms' capacity utilization levels when they find it difficult to predict the future are on average seven percentage points lower than when they find prediction easy. This difference corresponds to almost 40 percent of mean idle capacity in the sample. We also find that the negative effect of demand uncertainty on capacity utilization is significantly stronger for firms with higher markups, whether markups are estimated using the production approach of De Loecker and Warzynski (2012) or as EBIT margins.

Related literature. Beyond work already discussed, our study mainly relates to three strands of literature. The first is the literature on capacity utilization, which has largely focused on the consequences of capacity constraints and idle capacity for macroeconomic dynamics. A seminal contribution here is Fagnart, Licandro and Portier (1999), who develop a dynamic general equilibrium model where firms make forward-looking decisions that imply less than full capacity utilization in response to demand uncertainty. Their work forms the conceptual framework for later empirical research, for example by Boehm and Pandalai-Nayar (2022), who tie capacity utilization to industry supply curves in US data, and Balleer and Noeller (2024), who use firm-level data from Germany to examine how price responses to monetary policy shocks depend on

unlikely to be quantitatively important in our empirical context because Swedish manufacturing firms face intense international competition.

[2]The prediction holds under quite general conditions, but not universally. Our model pins down the necessary conditions precisely.

firms' capacity utilization. These and other papers in the empirical literature thus focus on how the response of prices and quantities to various shocks depends on capacity utilization—taking the latter as given—whereas our contribution is to provide empirical evidence on how firms make capacity decisions in the first place.[3]

Second, we add to the literature on newsvendor models by showing that this class of models is useful not only for studying inventory decisions, but also for understanding how manufacturing firms make capacity decisions. Moreover, while newsvendor models are central in operations management, that literature overwhelmingly consists of theoretical contributions (see, e.g., Qin et al., 2011, and DeYong, 2020). By documenting that higher demand uncertainty is associated with lower capacity utilization, and especially so for high-markup firms, we provide support for several key implications of newsvendor models that are not well established in the literature.

Finally, our findings relate to a rich literature examining the effect of uncertainty on the amount and timing of investment through mechanisms such as irreversibility, adjustment costs, risk aversion, and mode of competition (see Hartman, 1972; Abel, 1983; Caballero, 1991; and Dixit and Pindyck, 1994, for early influential work). We instead focus on the idea that uncertainty affects the kind of investments firms undertake through its interaction with capacity constraints.

1 Stylized Facts About Capacity Utilization at the Firm Level

The stylized facts presented in this section are based on micro data from *Konjunkturbarometern*, the most important business and household survey in Sweden (Konjunkturinstitutet, Various years). We describe *Konjunkturbarometern* in detail in section 3.1 below.

Fact 1 *A majority of firms routinely operate below full capacity, even outside of recessions*

The average capacity utilization of Swedish manufacturing firms hovers between 80 and 85 percent in normal times, while fewer than 15 percent operate at full capacity in an average period (Figure 1, Panels A and B). Hence, most firms routinely operate with meaningful idle capacity, even outside recessions. Both the average capacity utilization and the share of firms operating at full capacity are markedly procyclical; for example, the share of firms operating at

[3] See also Sun (2025), who calibrates a macro model with inattentive consumers to show capacity competition can lead to chronic excess capacity, which creates an important role for demand shocks in business cycles. Also related is Butters (2020), who examines capacity utilization in hotels and airlines. He finds higher demand variability is associated with lower average capacity utilization among hotels, where adjusting capacity involves substantial costs, but not in the airline industry, where capacity adjustment is less costly.

full capacity was below seven percent during the recession associated with the European debt crisis, but above 25 percent during the post-pandemic recovery period.

Fact 2 *"Insufficient demand" is the main reason for operating below capacity*

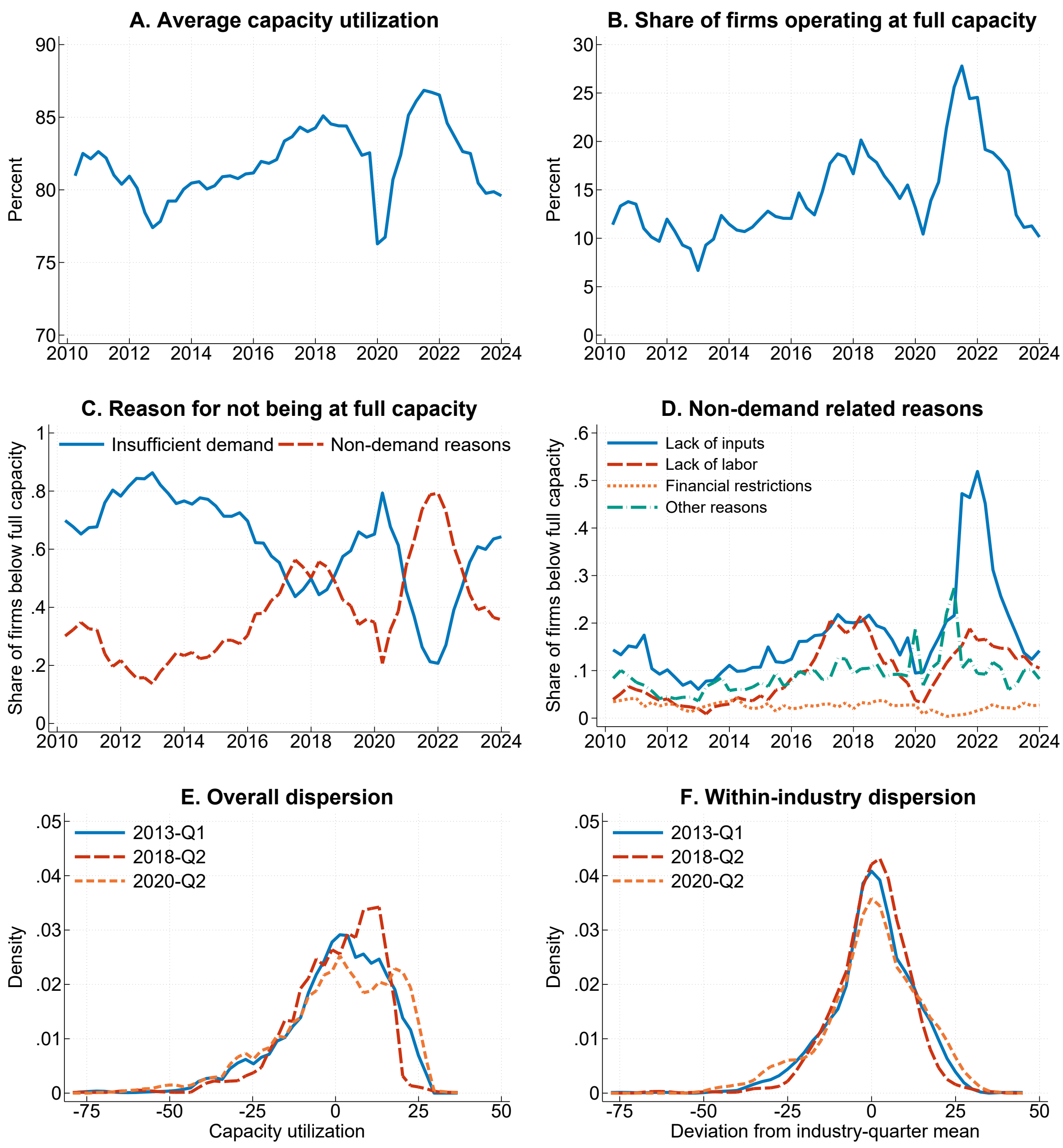
The most common reason for not operating at full capacity is "insufficient demand," which typically accounts for almost two-thirds of the cases in which firms operate below capacity (Figure 1, Panel C). The second most common reason is shortages of intermediate inputs (Panel D), which on average accounts for around 15 percent of cases. Insufficient demand thus outstrips other reasons for operating below full capacity by a wide margin. During the COVID-19 pandemic, however, half of the firms operating below full capacity cited a lack of intermediate inputs (compared to 20 percent citing insufficient demand), reflecting the importance of supply-side shocks in the decline in manufacturing output during the pandemic.

Fact 3 *The variation in capacity utilization across firms and over time is mainly idiosyncratic*

The cross-sectional dispersion in capacity utilization at any given point in time is large. Panel E of Figure 1 shows kernel density plots of firm-level capacity utilization in three quarters: one in which GDP growth was positive (2018Q2) and two in which GDP growth was negative (2013Q1 and 2020Q2). Panel F shows the corresponding density plot for industry-quarter demeaned capacity utilization, where industries are defined at the five-digit level. Most of the dispersion across firms remains after the demeaning, implying capacity utilization varies substantially even within narrowly defined industries. Indeed, the standard deviation of quarter-industry demeaned capacity utilization is 0.133, compared to 0.151 for the quarterly demeaned series.

To better understand the sources of variation in capacity utilization across firms and over time, we use regressions to assess to what extent various groups of factors explain the observed variation in capacity utilization. The results are reported in Table 1. Columns (1) and (2) show time fixed effects and interacted industry-time fixed effects explain 4.5 and 17 percent, respectively, of the variation in capacity utilization. Hence, macro shocks and industry-level shocks only account for a small part of the variation in firm-level capacity utilization. On the other hand, the R^2 from the regression of capacity utilization on firm fixed effects is 0.46 (column 3). Time-invariant firm characteristics thus account for almost half of the variation in capacity utilization. Finally, industry-time and firm fixed effects jointly account for 55 percent of the variation; the remaining 45 percent is thus due to some combination of time-varying firm characteristics and other time-varying factors at group level, such as location-specific shocks.

Figure 1: Capacity utilization over time and across firms



This figure reports descriptive statistics on capacity utilization at quarterly frequency among Swedish manufacturing firms. Panel A plots the average capacity utilization and Panel B the share of firms that operate below full capacity in each quarter. Panel C plots the shares of firms that operate below full capacity for demand and non-demand related reasons, respectively, while Panel D provides a further breakdown of the non-demand reasons. The series shown in Panels A–D are deseasonalized. The bottom two panels are kernel density plots of capacity utilization (Panel E) and industry-quarter demeaned capacity utilization (Panel F) in three quarters: one in which GDP growth was positive (2018-Q2) and two in which GDP growth was negative (2013-Q1 and 2020-Q2).

Table 1: Accounting for the variation in firm-level capacity utilization

	Dependent variable: Capacity utilization ($CU_{i,t}$)			
	(1) Time FE	(2) Industry-time FE	(3) Firm FE	(4) Industry-time and firm FE
R^2	0.045	0.172	0.455	0.546
Number of obs.	9,324	9,324	9,324	9,324

This table shows the R^2s obtained when regressing firm-level capacity utilization ($CU_{i,t}$) on different sets of fixed effects. Industries are defined at the level of five-digit SNI/NACE codes. Firms and industry-quarters with fewer than 12 observations are excluded from the estimations.

Taken together, our stylized facts demonstrate most firms deliberately operate with idle capacity, and the amount of idle capacity they maintain is primarily determined by idiosyncratic factors.[4] In what follows, we develop and test the hypothesis that firms choose their expected capacity utilization based on the demand uncertainty they face.

2 A Newsvendor Model of Expected Capacity Utilization

To formalize the link between capacity choice and uncertainty we build a newsvendor-style model (Arrow, Harris and Marschak, 1951) in which the firm chooses both price and capacity before demand is realized. Relative to the fixed-markup benchmark in Butters (2019), we allow for an ex-ante price choice (Raz and Porteus, 2006), which lets the model speak directly to markup heterogeneity. The model delivers two main predictions that map directly into the empirical analysis: higher demand uncertainty lowers expected capacity utilization, and that effect is stronger for higher-markup firms.

2.1 Setup and the capacity rule

Consider a risk-neutral firm that sells a single product over one period. At the start of the period, it chooses a posted price p and effective capacity K before demand is realized. After the demand

[4]Our Facts 1 and 2 have been noted previously by, for example, Boehm and Pandalai-Nayar (2022) and Sun (2025), but we are not aware of any prior work documenting Fact 3.

D is drawn, sales are

$$y = \min\{D, K\}. \tag{1}$$

Here K is effective capacity over the planning horizon: the maximum output the firm can deliver given its equipment, staffing, organization, and operating plan. Installing capacity costs $c > 0$ per unit. In our baseline model there is no additional variable production cost.[5]

Let $F_D(d \mid p, \theta)$ denote the conditional cdf of demand, where θ indexes the uncertainty state. Expected profits are

$$\Pi(p, K; \theta) = p\,\mathbb{E}[\min\{D, K\} \mid p, \theta] - cK. \tag{2}$$

For a given posted price,

$$\mathbb{E}[\min\{D, K\} \mid p, \theta] = \int_0^K [1 - F_D(x \mid p, \theta)]\; dx, \tag{3}$$

so the first-order condition for capacity choice is

$$p\,[1 - F_D(K^*(p, \theta) \mid p, \theta)] = c. \tag{4}$$

Define the Lerner-index form for the markup

$$\alpha(p) \equiv 1 - \frac{c}{p} = \frac{p - c}{p}, \qquad p > c, \tag{5}$$

which is also the well-known critical fractile as in Arrow, Harris and Marschak (1951).

Proposition 1 (Capacity is a demand quantile) *Under the regularity conditions in Appendix D1, for any fixed $p > c$ the capacity problem is strictly concave and the optimal capacity satisfies*

$$K^*(p, \theta) = Q_D(\alpha(p) \mid p, \theta)\,, \tag{6}$$

where $Q_D(u \mid p, \theta) \equiv F_D^{-1}(u \mid p, \theta)$ is the conditional demand quantile function.

This is the standard newsvendor logic: the firm sets capacity at a demand quantile by trading off the cost of carrying one more unit of capacity against the opportunity cost of missing out on one more unit of demand in high-demand states. Panel A of Figure 2 illustrates the cutoff rule.

[5] Appendix D7 adds a constant per-unit production cost m. The main structure is unchanged, except the relevant markup index becomes the contribution-margin fractile $\alpha_m(p) = 1 - c/(p - m)$.

2.2 Multiplicative demand and expected utilization

To keep the joint price-capacity problem tractable, assume demand takes the multiplicative form

$$D = d(p)\,\varepsilon, \tag{7}$$

where $d(p)$ is the deterministic demand scale with $d'(p) < 0$, and $\varepsilon > 0$ is a demand shifter. Let $Q(\cdot;\theta)$ denote the quantile function of ε. This specification is useful because price scales the whole demand distribution, while θ governs the distribution of the shock. Proposition 1 then implies

$$K^*(p,\theta) = d(p)\,Q(\alpha(p);\theta)\,. \tag{8}$$

Define expected capacity utilization at the optimum as

$$U(p,\theta) \equiv \frac{\mathbb{E}[\min\{D, K^*(p,\theta)\} \mid p,\theta]}{K^*(p,\theta)}. \tag{9}$$

Using the quantile identity $\mathbb{E}[\min\{\varepsilon, Q(\alpha;\theta)\} \mid \theta] = \int_0^\alpha Q(u;\theta)\,du + (1-\alpha)Q(\alpha;\theta)$, expected utilization can be written as

$$U(p,\theta) = U(\alpha(p),\theta) = \frac{\int_0^{\alpha(p)} Q(u;\theta)\,du}{Q(\alpha(p);\theta)} + [1-\alpha(p)]\,. \tag{10}$$

It is convenient to set

$$R(\alpha(p),\theta) \equiv \frac{\int_0^{\alpha(p)} Q(u;\theta)\,du}{Q(\alpha(p);\theta)}, \qquad U(\alpha(p),\theta) = R(\alpha(p),\theta) + (1-\alpha(p)). \tag{11}$$

The term $1-\alpha(p)$ is the probability that capacity binds. The term $R(\alpha(p),\theta)$ is the contribution of nonbinding states to expected utilization, scaled by installed capacity. The key tractability gain from our multiplicative demand assumption is that the scale term $d(p)$ cancels from the utilization ratio. Once capacity is chosen optimally, price matters for utilization only through the markup state $\alpha(p)$.

The firm still chooses price and capacity jointly. Substituting the capacity rule into profits, therefore, reduces the joint choice problem to a price choice:

$$\max_{p>c}\; p d(p) \int_0^{1-c/p} Q(u;\theta)\,du. \tag{12}$$

Appendix D1 gives conditions guaranteeing existence and a convenient sufficient condition for uniqueness. For our empirical work, the important point is straightforward: the endogenous price choice enters utilization only through the markup index $\alpha(p)$. By the same reason, in what follows we write α for the induced markup state whenever doing so creates no confusion.

2.3 Comparative statics: uncertainty, markup, and their interaction

Three comparative statics deliver the model's empirical predictions.

Proposition 2 (Higher uncertainty lowers expected utilization) *Fix a markup state $\alpha \in (0,1)$. Suppose the move from θ_L to θ_H is a mean- and median-preserving dispersive increase in the distribution of ε.*[6] *Then*

$$U(\alpha, \theta_H) \leq U(\alpha, \theta_L). \tag{13}$$

The proof, in Appendix D2, nests the fixed-markup utilization result in Butters (2019). Intuitively, higher uncertainty raises the value of insuring against high-demand realizations, capacity rises relative to expected sales, so expected utilization falls.

Proposition 3 (Higher markup lowers expected utilization) *On any region where $Q(\alpha;\theta) > 0$ and the relevant derivatives exist,*

$$U_\alpha(\alpha, \theta) = -R(\alpha, \theta)\, \partial_\alpha \log Q(\alpha; \theta) < 0. \tag{14}$$

Higher-markup firms thus choose deeper buffers against high demand realizations and operate with more slack capacity on average.

The empirical examination further examines the interaction between markup and uncertainty: the negative uncertainty effect (Proposition 2) is stronger for high-markup firms under a local condition.

Proposition 4 (Local uncertainty-markup interaction) *The local cross effect of markup and demand uncertainty on expected capacity utilization is negative, i.e., $U_{\alpha\theta}(\alpha, \theta) \leq 0$ if and only if an increase in uncertainty moves the upper-tail quantile that pins down capacity by more than it moves the lower quantiles that pin down expected sales. Formally,*

$$U_{\alpha\theta}(\alpha, \theta) \leq 0 \tag{15}$$

[6] Formally, the two distributions have the same mean and median, and their quantile spreads satisfy

$$Q(\beta; \theta_H) - Q(\alpha; \theta_H) \geq Q(\beta; \theta_L) - Q(\alpha; \theta_L)$$

for all $0 < \alpha < \beta < 1$.

if and only if

$$\underbrace{\partial_{\alpha\theta} \log Q(\alpha;\theta)}_{\textit{endpoint response}} \geq \underbrace{\partial_{\alpha} \log Q(\alpha;\theta)}_{\textit{local tail slope}} \left[\underbrace{\partial_{\theta} \log Q(\alpha;\theta) - \frac{\int_0^{\alpha} Q(u;\theta)\, \partial_{\theta} \log Q(u;\theta)\, du}{\int_0^{\alpha} Q(u;\theta)\, du}}_{\textit{endpoint relative to lower quantiles}} \right]. \tag{16}$$

See Appendix D3 for the formal proof.[7] This result says the interaction between uncertainty and markups is negative when uncertainty disproportionately thickens the upper tail of demand that is relevant for high-markup firms' capacity choices.

The interpretation is easiest to see by comparing two firms. Consider first a firm that chooses a relatively central quantile of demand, so a dispersive shift in the demand distribution ε barely changes its capacity target. Now compare to a higher-markup firm that chooses a cutoff farther out in the right tail, where the same shift moves the relevant quantile much more. Capacity then rises faster than expected sales, so utilization falls by more.

Our data capture discrete variation in forward-looking demand uncertainty with cross-firm heterogeneity in markups. The natural empirical counterpart to the model is therefore a finite-difference comparison across uncertainty and markup states. In Appendix D6, we establish that the utilization gap between high- and low-uncertainty firms is more negative among high-markup firms if Proposition 4 holds.

2.4 A lognormal illustration

The two-dimension comparative-statics logic stated in Proposition 4 can be worked out explicitly for a broad class of positive-support demand-shock specifications. In particular, when $\log \varepsilon$ belongs to a location-scale family, such as the normal, logistic, or Gumbel, the model yields analogous expressions for expected utilization and the uncertainty-markup interaction. We focus on the lognormal here as a clean visual illustration.

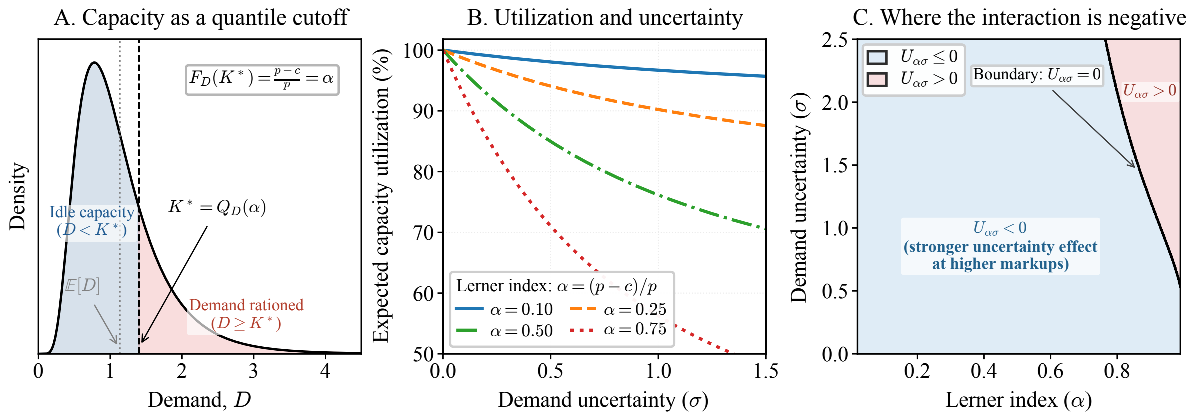
When ε is lognormal with uncertainty parameter σ, its quantile function is

$$Q(\alpha;\sigma) = \exp\{\nu(\sigma) + \sigma z_{\alpha}\}, \qquad z_{\alpha} \equiv \Phi^{-1}(\alpha),$$

where the location term $\nu(\sigma)$ drops out of utilization. In that case, utilization has the closed

[7] Appendix D8 gives an example of when the cross-derivative is positive (due to lower uncertainty shocks for higher-markup firms) and establishes that differences across uncertainty levels attenuate in the extreme upper tail.

Figure 2: Capacity choice, uncertainty, and expected utilization



Panel A illustrates the quantile cutoff rule: the firm sets optimal capacity K^* at the α-quantile of demand, where $\alpha = (p - c)/p$. Panel B plots expected utilization in the lognormal case using (17) for four markup states. Utilization declines with uncertainty, and the decline is steeper at higher markups. Panel C maps the region in (α, σ)-space where the local interaction in (18) is negative. The shaded area labeled $U_{\alpha\sigma} < 0$ is the region in which the uncertainty effect is stronger at higher markups; the wedge labeled $U_{\alpha\sigma} > 0$ shows where the interaction becomes positive.

form

$$U(\alpha, \sigma) = e^{\sigma^2/2 - \sigma z_\alpha} \Phi(z_\alpha - \sigma) + (1 - \alpha), \tag{17}$$

and the interaction condition becomes

$$U_{\alpha\sigma}(\alpha, \sigma) \leq 0 \iff 1 \geq \sigma \left[z_\alpha - \sigma + \frac{\phi(z_\alpha - \sigma)}{\Phi(z_\alpha - \sigma)} \right]. \tag{18}$$

Appendix D5 derives (17)–(18).

Figure 2 visualizes the model. Panel A shows the cutoff rule. Panel B makes the two core predictions visible: utilization falls with uncertainty, and the decline becomes steeper as markup rises. Panel C shows that, in the lognormal case, the negative-interaction region covers most of the relevant parameter space, with the positive region confined to combinations of very high markups and very high uncertainty.

3 Demand Uncertainty and Capacity Utilization in the Data

3.1 Data and sample

Data sources. We examine the effects of demand uncertainty on expected capacity utilization using data from two sources. The first is the aforementioned survey *Konjunkturbarometern*, which has been conducted by the National Institute for Economic Research (henceforth the NIER) since the 1950s. The NIER is a government agency operating independently under the Ministry of Finance; its principal responsibility is to support Swedish economic policymakers by producing forecasts and economic analyses. *Konjunkturbarometern* is a monthly survey, but several of the questions necessary for our analysis are only asked in the last month of each quarter. We therefore treat our data as quarterly and only use observations corresponding to the quarter's final month. The data we have access to spans the period 2010Q1–2024Q1.

The survey covers almost all sectors in the economy. Firms with at least 100 employees are always included, while smaller firms are selected through random sampling stratified by sector and size class.[8] A new sample is drawn each year; hence, selected smaller business units stay in the sample for at least four quarters, but may disappear after that. Around 5,800 firms are included in the survey annually, but they are not legally required to respond. The response rates on the questions we use are between 47 and 51 percent, a similar rate to comparable surveys in other countries (see, e.g., Ropele, Gorodnichenko and Coibion, 2024).

Our second data source is Serrano, an annual firm-level panel that contains the universe of incorporated firms in Sweden (Dun & Bradstreet, 1998–2021). Serrano contains detailed accounting data plus information about, among other things, a firm's industry, location, and age. We use Serrano to compute markups and to assign five-digit industry codes to our sample firms. We link the Serrano data to the NIER data by means of the unique identifier (*organisationsnummer*) belonging to every Swedish firm.

Sample composition and characteristics. Our analysis sample comprises all manufacturing firms that respond to the NIER survey and covers the period 2021Q2–2024Q1. We restrict the sample period because one question critical for our purposes—namely, how uncertain firms are about the future—was only introduced into the survey in 2021Q2. We drop observations with

[8] Strictly speaking, the unit of observation is the business unit (*verksamhetsenhet*), which comprises the production units within a firm that operate in the same industry. Only about one percent of the firms in the survey have multiple business units, however, so we aggregate the data to the firm level. The aggregation is done by taking the modal response to multiple-choice questions and the mean response to numeric questions.

a missing value for any variable in our empirical model to ensure a consistent sample across specifications. These restrictions leave us with 6,603 observations across 953 firms.[9]

Table A1 in Online Appendix A provides descriptive statistics for those of our sample firms that responded to the survey in 2021Q4 . The median firm is 36 years old and has 393 million SEK in annual sales, 251 million SEK in total assets, and 133 employees (the USD/SEK exchange rate is around 10). The corresponding means are substantially larger due to the strong right-skewedness of the firm size distribution; the average firm is 44 years old and has 1.6 billion SEK in sales, 2.0 billion SEK in total assets, and 297 employees.

3.2 Empirical model

We examine the relationship between capacity utilization and demand uncertainty using the following model:

$$CU_{i,t} = \alpha_i + \psi_{j,t} + \sum_{k=2}^{4} \beta_k \cdot \mathbb{1}\{DU_{i,t} = k\} + \mathbf{\Omega} \cdot \mathbf{X_{i,t}} + \varepsilon_{i,t}. \tag{19}$$

The dependent variable, $CU_{i,t}$, is the capacity utilization reported by firm i in the NIER survey in period t.[10] $DU_{i,t}$—our measure of firms' perceived, forward-looking demand uncertainty—is a categorical variable recording firm i's response to the following question in the NIER survey in period t: "The future development of our business situation is currently... easy to predict / fairly easy to predict / fairly difficult to predict / difficult to predict."[11] α_i and $\psi_{j,t}$ are firm and industry-period fixed effects, respectively. (We include different sets of fixed effects in different specifications below.) The coefficients of interest are the β_k, which capture the conditional average difference in capacity utilization between firms in uncertainty bins 1 and k, respectively, where bin 1 corresponds to the answer "easy to predict." Standard errors are clustered by firm.

The vector $\mathbf{X_{i,t}}$ includes two key control variables. The first is contemporaneous output

[9] To assuage concerns about external validity in the face of this short panel, we show in Online Appendix B that the results are qualitatively similar if we instead estimate the effects of demand uncertainty on capacity utilization using an alternative micro data set from Statistics Sweden, which covers a longer time period. The drawback of that data is we do not directly observe firms' subjective, forward-looking uncertainty. We must instead rely on a statistically derived industry-level proxy for demand uncertainty. This is why we use the NIER data in the main part of the paper.

[10] We exclude observations for which reported capacity utilization is equal to zero or strictly larger than 100 percent. These observations make up around one percent of the initial sample.

[11] One may argue for the use of *lagged* demand uncertainty as explanatory variable, as it takes some time for a firm to adjust capacity in response to shocks. A common assumption in the literature is that capacity adjustments take three months (Sun, 2025), suggesting a one-quarter lag. We nevertheless use contemporaneous demand uncertainty because our data is a rotating panel, so we lose over 20 percent of observations when lagging the explanatory variable. That said, our results are robust to using the one-quarter lag of demand uncertainty (see Table A2 in Online Appendix A).

growth, measured using firms' responses to the following question in the NIER survey: "Our production volume has in the past three months... increased / not changed / decreased." We include this control because we are interested in estimating the effect of demand uncertainty on firms' *expected* capacity utilization, but our dependent variable measures firms' *realized* capacity utilization.[12] By controlling for firms' contemporaneous output growth, we partial out the influence of realized demand and supply shocks on a firm's capacity utilization and can consequently interpret the β_k as the effect of demand uncertainty on a firm's choice of *expected* capacity utilization.

The second key control is expected output growth in the near future, measured using firms' responses to the question: "We expect our production volume in the coming three months to... increase / not change / decrease." We include this control to avoid confounding variation in uncertainty with variation in point estimates of future outcomes. In some specifications, we also include two inventory-related controls constructed on the basis of firms' responses to the questions: "Our inventory of raw materials is currently... too large / just right / too small" and "Our inventory of finished goods is currently... too large / just right / too small."

Figure A1 in Online Appendix A provides descriptive statistics on the main variables in the model: capacity utilization, demand uncertainty, contemporaneous output growth, and expected output growth. Responses to the question about how difficult it is to predict the firm's future business situation are on average distributed as follows: 2.3 percent respond that it is easy, 26.6 percent fairly easy, 53.3 percent fairly difficult, and 17.8 percent difficult.

3.3 The effect of demand uncertainty on capacity utilization

We report the estimation results in Table 2, beginning with a stripped-down version of the model that includes only demand uncertainty and industry-time fixed effects as independent variables.[13] The estimates show capacity utilization decreases monotonically with demand uncertainty. Firms that find their future business situation fairly difficult and difficult to predict, respectively, have on average 6.1 and 9.5 percentage points lower capacity utilization than firms that find it easy. The difference between firms that find it easy and fairly easy is 1.5 percentage points, but this gap is not statistically significant. These differences are economically meaningful. Average capacity utilization in the sample is 83.4 percent. Hence, the difference be-

[12]More precisely, $CU_{i,t}$ is the empirical analog of q/K^*, where $q = \bar{q} + \epsilon$, whereas we want to estimate the effect of demand uncertainty on $\bar{q}/K^*$.

[13]Industries are defined by five-digit SNI/NACE codes, and we assign the most recently observed industry code in Serrano to each firm.

tween firms in the easy and difficult categories amounts to 57 percent of mean idle capacity (0.095/0.166).

Next, we add contemporaneous output growth to the set of controls to ensure our results are not biased by realized demand or supply shocks. This leads, as shown in column (2), to slight declines in the magnitudes of the coefficient estimates for the answers fairly difficult and difficult, but the estimates remain statistically and economically significant. The conditional average difference in capacity utilization between firms in the easy and difficult categories is now 7.0 percentage points, or 42 percent of mean idle capacity in the sample. The estimates for the contemporaneous output growth question are informative in their own right. Capacity utilization among firms whose output has decreased in the past three months is 9.8 percentage points lower than among firms whose output did not change. Firms whose output has increased, meanwhile, have on average 2.9 percentage points higher capacity utilization than firms whose output did not change. These values are sensible and indicate our output-growth proxy does a good job of controlling for contemporaneous shocks.

The results reported in the following two columns are from estimations in which we further augment the model with controls for expectations about future output growth (column 3) and current inventory levels (column 4). The coefficients of interest are not meaningfully affected by including these control variables.

In column (5), we substitute firm and time fixed effects for the industry-time fixed effects. The firm fixed effects substantially increase R^2, from 0.38 to 0.70, but the coefficient estimates are hardly affected. The relationship between demand uncertainty and expected capacity utilization therefore also exists within firms over time. The estimates indicate that when firms find their future business situation fairly easy, fairly difficult, and difficult to predict, respectively, they have on average 1.8, 4.3, and 6.5 percentage points lower capacity utilization than when they find it easy to predict their future business situation.

The results in the final two columns address a potential concern with our measure of demand uncertainty. Namely, the survey question refers to firms' "business situation," a rather broad concept that could be interpreted as covering not just demand conditions but also supply-side factors. We address this by estimating our model on two subsamples of firms: one contains firms that either operate at full capacity or operate below capacity because of "insufficient demand" (column 6), while the other comprises firms that either operate at full capacity or operate below full capacity for non-demand related reasons (column 7).[14] If our findings above reflect demand-

[14]Firms can report multiple reasons for operating below full capacity. We assign a firm that operates below full capacity to the sample in column (6) if at least one provided reason is insufficient demand and otherwise to the

Table 2: The effect of demand uncertainty on capacity utilization

	Dependent variable: Capacity utilization ($CU_{i,t}$)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicting the future development of our business situation is currently... [Omitted category: Easy]							
–Fairly easy	–0.015	–0.018	–0.020	–0.018	–0.018	–0.019	0.008
	(0.016)	(0.015)	(0.015)	(0.015)	(0.011)	(0.019)	(0.012)
–Fairly difficult	–0.061***	–0.050***	–0.053***	–0.049***	–0.043***	–0.050**	–0.001
	(0.016)	(0.015)	(0.015)	(0.015)	(0.012)	(0.022)	(0.012)
–Difficult	–0.095***	–0.070***	–0.072***	–0.069***	–0.065***	–0.069***	–0.016
	(0.018)	(0.017)	(0.017)	(0.017)	(0.013)	(0.022)	(0.014)
Our production volume has over the past three months... [Omitted category: Not changed]							
–Increased		0.029***	0.034***	0.032***	0.027***	0.032***	0.018***
		(0.005)	(0.005)	(0.005)	(0.003)	(0.006)	(0.004)
–Decreased		–0.098***	–0.094***	–0.088***	–0.068***	–0.069***	–0.057***
		(0.007)	(0.007)	(0.007)	(0.004)	(0.006)	(0.008)
We expect our production volume over the next three months to... [Omitted category: Remain the same]							
–Increase			–0.021***	–0.021***	–0.010***	–0.018***	–0.007**
			(0.005)	(0.005)	(0.003)	(0.006)	(0.004)
–Decrease			–0.017**	–0.014**	–0.002	–0.001	–0.008
			(0.007)	(0.006)	(0.004)	(0.005)	(0.007)
Industry × time FE	Yes	Yes	Yes	Yes	No	No	No
Firm and time FE	No	No	No	No	Yes	Yes	Yes
Inventory controls	No	No	No	Yes	Yes	Yes	Yes
Number of obs.	6,603	6,603	6,603	6,603	6,603	3,387	3,332
Number of firms	953	953	953	953	953	695	710
R^2	0.294	0.360	0.363	0.375	0.701	0.777	0.733

This table reports estimation results for the regression specified in (19). The industry-time fixed effects are constructed based on five-digit SNI/NACE codes. The estimation sample in column (6) comprises firms that either operate at full capacity or operate below capacity because of "insufficient demand," while the sample in column (7) comprises firms that either operate at full capacity or operate below capacity for non-demand related reasons. Standard errors are clustered at the firm level in all regressions. *, **, and *** denote statistical significance at the ten, five, and one percent levels, respectively.

related uncertainty, we should only find a significant effect in the former group, because there is no reason to expect demand uncertainty to explain variation in capacity utilization driven by non-demand reasons. Reassuringly, the estimates in column (6) are statistically significant and close in magnitude to the prior estimates, while the estimates from the placebo test in column (7) are close to zero and statistically insignificant.

3.4 The role of markups

We now turn to the second main prediction of the newsvendor model, namely, that the effect of demand uncertainty on capacity utilization is stronger for firms with higher markups. We examine the role of markups using cross-sectional heterogeneity analysis, in which we estimate (19) separately for firms with high and low markups and then test whether the effect of demand uncertainty differs across the groups.

We estimate markups following the production function approach proposed by De Loecker and Warzynski (2012) (see Online Appendix C for details) and classify firms in the top tercile of the markup distribution as high-markup firms and firms in the bottom tercile as low-markup firms. We sort firms based on their average markups over the five-year period preceding our sample period (2016–2020) to ensure the sorting is not influenced by contemporaneous shocks that might also affect firms' capacity utilization.

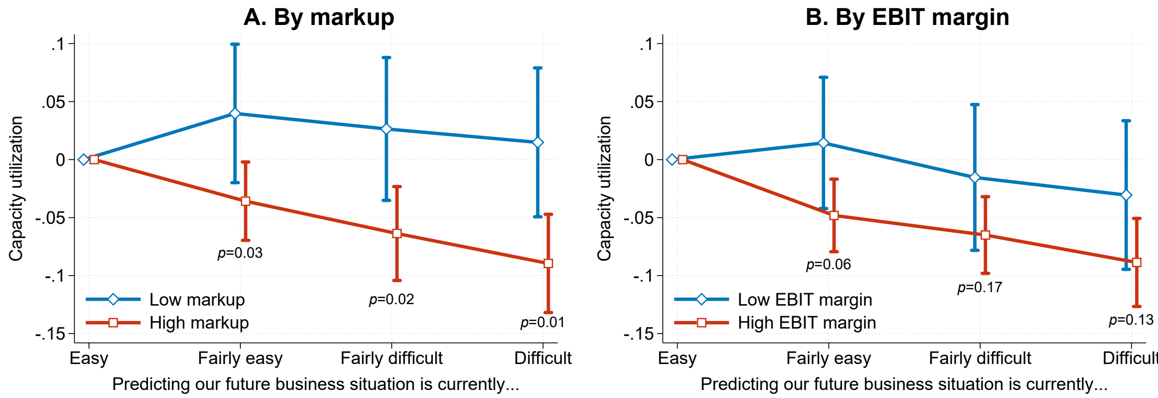
We plot the estimation results graphically in Panel A of Figure 3. The blue diamonds are the estimates of β_k obtained when estimating our specification on the sample of low-markup firms, while the red squares are the corresponding estimates for the high-markup firms (recall that the β_k capture the conditional average difference in capacity utilization between firms in uncertainty bins 1 and k). Capacity utilization declines monotonically and substantially with uncertainty among high-markup firms: difference between firms that find it easy and difficult to predict their future business situation is 8.9 percentage points. Among low-markup firms, on the other hand, there is no relationship between uncertainty and capacity utilization. The differences between the respective low- and high-markup subsample β_k estimates are statistically significant at the 5 percent level in all three cases.

There is some debate about the accuracy of markups measured using the production function approach.[15] For the sake of robustness, we redo the analysis as before, but sort firms based on their EBIT margins instead of their estimated markups. The EBIT margin has its own issues as a

sample in column (7).

[15] See Bond, Hashemi, Kaplan and Zoch (2021) and De Ridder, Grassi and Morzenti (2026), for example.

Figure 3: Heterogeneity in the effect of demand uncertainty on capacity utilization



This figure reports the estimates of β_k obtained when estimating equation (19) on samples comprising low-markup/low-EBIT margin firms (blue diamonds) and high-markup/high-EBIT margin firms (red squares), respectively; see the main text for details on how firms are sorted into the respective groups. All estimations use the baseline version of equation (19), which is the same specification as used in column (5) of Table 2. The vertical bars represent 95-percent confidence bands, computed based on standard errors clustered at the firm level. The p-values are from two-sided tests of the null hypothesis that a given β_k does not differ between the two groups.

measure (see, e.g., De Loecker, Eeckhout and Unger, 2020), but offer an independently derived markup metric. The results of this exercise, reported in Panel B of Figure 3, are qualitatively similar to the main results, though the differences between the two subsamples' respective β_k estimates are no longer statistically significant. The qualitative parallels nevertheless suggest our findings are not sensitive to the precise way in which we estimate markups.

The results reported in Figure 3 thus confirm the newsvendor model's prediction that demand uncertainty has a larger effect on capacity utilization for firms with higher markups. These results also speak to the concern that realized demand shocks could drive our main results. That is, had (unobserved) firm-level demand shocks caused the negative association between demand uncertainty and capacity utilization, we would have observed a negative relationship also among low-markup firms. The fact that we don't speaks against the concern that realized demand shocks may drive our findings.

4 Concluding Remarks

Firms' amount of idle capacity has a history of prominence in macroeconomic policy, where estimates of aggregate capacity utilization and the output gap are key inputs in the policy-making process (see, e.g., Corrado and Mattey, 1997). We instead emphasize that there is much cross-sectional variation in capacity utilization and its changes, and we document that much of this is a rational response to demand uncertainty. Loosely speaking, whereas the previous literature has focused on the macroeconomic consequences of capacity constraints and idle capacity, we focus on the microeconomic question of how firms make capacity decisions in the first place. We find that a straightforward newsvendor model does quite well in explaining the considerable amount of across-firm variation in expected capacity utilization.

While our focus is to explain the firm-level variation, we conclude with some reflections on what our results may imply for aggregate capacity utilization and suggest avenues for future research. One observation is that in our class of models, higher uncertainty leads to more excess capacity in expectation. This excess capacity uses resources but can be seen as a cost of insurance against surplus losses from stock-outs. Evaluating the welfare cost of this insurance is a complex task, as one should also take account of the lower risk of rationing for these firms' customers. Still, with an average slack of over 15 percent, there is clearly scope for quantitatively important effects. Furthermore, the fact that the relationship between demand uncertainty and capacity utilization is stronger for firms with high markups points to a possible additional distortion of market power when uncertainty is high: not only is there deadweight loss from higher prices, but also the costs of maintaining unused capacity.

Finally, some observers have pointed to a decline in capacity utilization in the U.S. over recent decades (see, e.g., Pierce and Wisniewski, 2018; Gahn, 2020; and Bohr, 2024). Our results suggest one possible explanation for this is increases in markups and uncertainty. Growing markups have been found by, among others, De Loecker, Eeckhout and Unger (2020). And when it comes to uncertainty, it seems plausible that rapid technological change and the intensified international competition facing manufacturing firms (see, e.g., Autor et al., 2020) have made predicting demand more difficult. The set of measures capturing the uncertainty faced by firms is rapidly expanding (see e.g. Baker, Bloom and Davis, 2016; and Altig et al., 2022) and can be used to further examine the drivers of changes in aggregate capacity utilization.

References

Abel, Andrew B. 1983. "Optimal Investment Under Uncertainty." *American Economic Review*, 73(1): 228–233.

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83(6): 2411–2451.

Altig, David, Jose Maria Barrero, Nicholas Bloom, Steven J Davis, Brent Meyer, and Nicholas Parker. 2022. "Surveying Business Uncertainty." *Journal of Econometrics*, 231(1): 282–303.

Arrow, Kenneth J, Theodore Harris, and Jacob Marschak. 1951. "Optimal Inventory Policy." *Econometrica*, 19(3): 250–272.

Autor, David, David Dorn, Gordon H Hanson, Gary Pisano, and Pian Shu. 2020. "Foreign Competition and Domestic Innovation: Evidence from US Patents." *American Economic Review: Insights*, 2(3): 357–374.

Baker, Scott R, Nicholas Bloom, and Steven J Davis. 2016. "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics*, 131(4): 1593–1636.

Balleer, Almut, and Marvin Noeller. 2024. "Monetary Policy in the Presence of Supply Constraints: Evidence from German Firm-Level Data." CEPR Discussion Paper No. 18732.

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry. 2018. "Really Uncertain Business Cycles." *Econometrica*, 86(3): 1031–1065.

Boehm, Christoph E, and Nitya Pandalai-Nayar. 2022. "Convex Supply Curves." *American Economic Review*, 112(12): 3941–3969.

Bohr, Clement E. 2024. "Capacity Buffers: Explaining the Retreat and Return of the Phillips Curve." Unpublished manuscript.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. 2021. "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data." *Journal of Monetary Economics*, 121: 1–14.

Butters, R Andrew. 2019. "On Demand Uncertainty in the Newsvendor Model." *Economics Letters*, 185: 108746.

Butters, R Andrew. 2020. "Demand Volatility, Adjustment Costs, and Productivity: An Examination of Capacity Utilization in Hotels and Airlines." *American Economic Journal: Microeconomics*, 12(4): 1–44.

Caballero, Ricardo J. 1991. "On the Sign of the Investment-Uncertainty Relationship." *American Economic Review*, 81(1): 279–288.

Corrado, Carol, and Joe Mattey. 1997. "Capacity Utilization." *Journal of Economic Perspectives*, 11(1): 151–167.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review*, 102(6): 2437–2471.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics*, 135(2): 561–644.

De Ridder, Maarten, Basile Grassi, and Giovanni Morzenti. 2026. "The Hitchhiker's Guide to Markup Estimation: Assessing Estimates from Financial Data." *Econometrica*, 94(1): 137–168.

DeYong, Gregory D. 2020. "The Price-Setting Newsvendor: Review and Extensions." *International Journal of Production Research*, 58(6): 1776–1804.

Dixit, Avinash K, and Robert S Pindyck. 1994. *Investment under Uncertainty.* Princeton University Press.

Dun & Bradstreet. 1998–2021. "The Serrano Database."

Fagnart, Jean-Franççois, Omar Licandro, and Franck Portier. 1999. "Firm Heterogeneity, Capacity Utilization, and the Business Cycle." *Review of Economic Dynamics*, 2(2): 433–455.

Gahn, Santiago. 2020. "Is There a Declining Trend in Capacity Utilization in the US Economy? A Technical Note." *Review of Political Economy*, 32(2): 283–296.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W Huffman. 1988. "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review*, 402–417.

Hartman, Richard. 1972. "The Effects of Price and Cost Uncertainty on Investment." *Journal of Economic Theory*, 5(2): 258–266.

Konjunkturinstitutet. Various years. "Konjunkturbarometern." Documentation available at: https://www.konj.se/media/njddjtc0/metodbok.pdf.

Kuhn, Florian, and Chacko George. 2019. "Business Cycle Implications of Capacity Constraints under Demand Shocks." *Review of Economic Dynamics*, 32: 94–121.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70(2): 317–341.

Lieberman, Marvin B. 1987. "Excess Capacity as a Barrier to Entry: An Empirical Appraisal." *Journal of Industrial Economics*, 35(4): 607–627.

Michaillat, Pascal, and Emmanuel Saez. 2015. "Aggregate Demand, Idle Time, and Unemployment." *Quarterly Journal of Economics*, 130(2): 507–569.

Pierce, Justin, and Emily Wisniewski. 2018. "Some Characteristics of the Decline in Manufacturing Capacity Utilization." FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Qin, Yan, Ruoxuan Wang, Asoo J Vakharia, Yuwen Chen, and Michelle MH Seref. 2011. "The Newsvendor Problem: Review and Directions for Future Research." *European Journal of Operational Research*, 213(2): 361–374.

Raz, Gal, and Evan L Porteus. 2006. "A Fractiles Perspective to the Joint Price/Quantity Newsvendor Model." *Management Science*, 52(11): 1764–1777.

Ropele, Tiziano, Yuriy Gorodnichenko, and Olivier Coibion. 2024. "Inflation Expectations and Misallocation of Resources: Evidence from Italy." *American Economic Review: Insights*, 6(2): 246–61.

Rovigatti, Gabriele. 2020. "MARKUPEST: Stata module for markup estimation via Micro and Macro approaches." *Statistical Software Components, Boston College Department of Economics*.

Statistics Sweden. 1998–2023. "Industrins kapacitetsutnyttjande." Documentation available at: https://www.scb.se/hitta-statistik/statistik-efter-amne/naringsverksamhet-och-utrikeshandel/foretagens-produktion-forsaljning-och-ekonomi–kortperiodisk-statistik/industrins-kapacitetsutnyttjande/.

Sun, Tiancheng. 2025. "Excess Capacity and Demand-Driven Business Cycles." *Review of Economic Studies*, 92(4): 2730–2764.

Online Appendix for "Have We Got News For You: Theory and Evidence on Firms' Optimal Expected Capacity Utilization"

Niklas Amberg, Richard Friberg, Xiang Liu, and Chad Syverson

Appendix A. Additional Tables and Figures

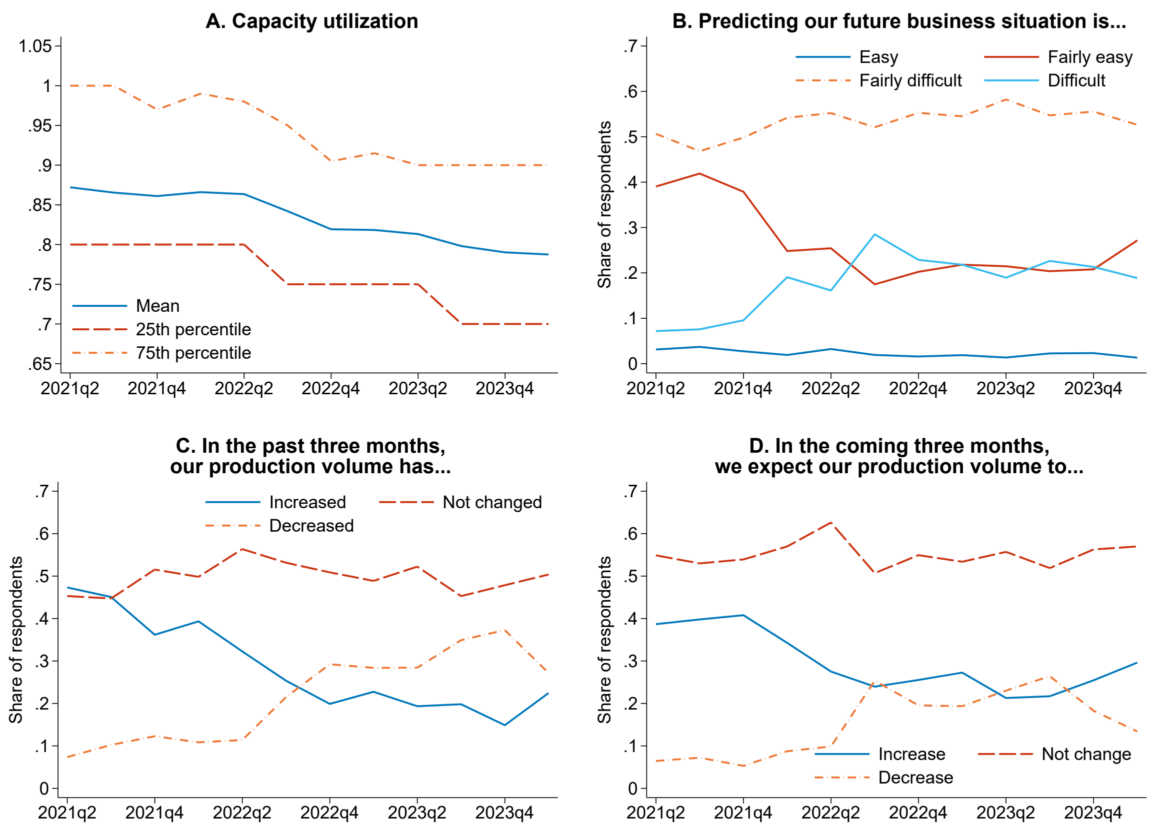
This appendix provides additional tables and figures referred to in the main text of the paper. Table A1 presents descriptive statistics for those of our sample firms that responded to the NIER survey in 2021Q4; Figure A1 descriptive statistics on the key questions in the NIER survey; and Table A2 estimation results for a version of equation (19) in which the main explanatory variable (demand uncertainty) is lagged by one quarter.

Table A1: Descriptive statistics for sample firms (2021Q4)

	Mean	Pct. 25	Median	Pct. 75	SD	*N*
Sales (MSEK)	1,604	149	393	983	6,371	579
Total assets (MSEK)	1,957	84	251	727	14,101	579
Number of employees (FTEs)	297	57	133	253	712	579
Age (years)	44	27	36	60	27	579
Cash holdings/Total assets	0.078	0.000	0.016	0.117	0.118	579
Liabilities/Total assets	0.643	0.521	0.659	0.783	0.188	579
EBIT margin (EBIT/Sales)	0.075	0.031	0.068	0.130	0.155	579
Markup (2016–20 average)	1.183	1.014	1.113	1.259	0.358	449

This table provides descriptive statistics for those of our sample firms that responded to the NIER survey in 2021Q4. The variables are constructed based on data from Serrano for the fiscal year 2021.

Figure A1: Descriptive statistics on key survey questions



Panel A plots the mean of the capacity-utilization variable in the NIER data, along with the 25th and 75 percentiles, in each quarter of the sample period. Panel B shows the proportion of respondents in each quarter that answered easy, fairly easy, difficult, and fairly difficult on the question about the difficulty of predicting the firm's future business situation. Panels C and D show the corresponding proportions for the questions about output growth over the past three months and expectations about output growth over the coming three months, respectively.

Table A2: The effect of lagged demand uncertainty on capacity utilization

	Dependent variable: Capacity utilization ($CU_{i,t}$)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
–Fairly easy	–0.018	–0.016	–0.015	–0.015	–0.021**	–0.038**	–0.004
	(0.019)	(0.018)	(0.017)	(0.017)	(0.011)	(0.015)	(0.014)
–Fairly difficult	–0.055***	–0.044**	–0.044**	–0.042**	–0.039***	–0.059***	–0.017
	(0.019)	(0.018)	(0.017)	(0.017)	(0.012)	(0.017)	(0.015)
–Difficult	–0.087***	–0.065***	–0.064***	–0.062***	–0.066***	–0.080***	–0.032**
	(0.021)	(0.019)	(0.019)	(0.019)	(0.013)	(0.019)	(0.016)
Our production volume has over the past three months... [Omitted category: Not changed]							
–Increased		0.028***	0.033***	0.032***	0.029***	0.037***	0.022***
		(0.006)	(0.006)	(0.006)	(0.004)	(0.007)	(0.004)
–Decreased		–0.100***	–0.095***	–0.089***	–0.070***	–0.070***	–0.060***
		(0.008)	(0.008)	(0.007)	(0.005)	(0.006)	(0.009)
We expect our production volume over the next three months to... [Omitted category: Remain the same]							
–Increase			–0.022***	–0.022***	–0.008**	–0.014**	–0.008**
			(0.006)	(0.006)	(0.004)	(0.006)	(0.004)
–Decrease			–0.021***	–0.017**	–0.004	–0.003	–0.007
			(0.007)	(0.007)	(0.004)	(0.005)	(0.008)
Industry × time FE	Yes	Yes	Yes	Yes	No	No	No
Firm and time FE	No	No	No	No	Yes	Yes	Yes
Inventory controls	No	No	No	Yes	Yes	Yes	Yes
Number of obs.	5,134	5,134	5,134	5,134	5,118	2,740	2,668
Number of firms	915	915	915	915	810	587	594
R^2	0.294	0.364	0.367	0.381	0.717	0.785	0.753

This table reports estimation results for a modified version of the regression specified in (19), in which the main explanatory variable is lagged by one quarter instead of entering contemporaneously. The industry-time fixed effects are constructed based on five-digit SNI/NACE codes. The estimation sample in column (6) comprises firms that either operate at full capacity or operate below capacity because of "insufficient demand," while the sample in column (7) comprises firms that either operate at full capacity or operate below capacity for non-demand related reasons. Standard errors are clustered at the firm level in all regressions. *, **, and *** denote statistical significance at the ten, five, and one percent levels, respectively.

Appendix B. An Alternative Test of The Effect of Demand Uncertainty on Capacity Utilization

Our main test of the effect of demand uncertainty on capacity utilization relies on data from the National Institute for Economic Research's survey *Konjunkturbarometern*. The benefit of using this data is that we can measure subjective, forward-looking uncertainty at the firm level directly based on firms' responses to the survey question about how difficult it is to predict their future business situation. The drawback of the NIER data, however, is that the panel is rather short, which raises a question about the external validity of the baseline results.

In this appendix, we therefore estimate the effects of demand uncertainty on capacity utilization using an alternative data set from Statistics Sweden, which cover a much longer time period (1998Q1-2023Q1). The drawback of the Statistics Sweden data is that we do not directly observe firms' subjective, forward-looking uncertainty, and instead have to rely on a cruder industry-level proxy for demand uncertainty in the estimations; this is why we use the NIER data in the main part of the paper. The longer time period covered by the Statistics Sweden data nevertheless makes these data useful for probing the external validity of our baseline results.

B1 Data

The alternative data set is the micro data underlying the aggregate industrial capacity utilization series produced by Statistics Sweden, the official Swedish statistics agency (Statistics Sweden, 1998–2023). Statistics Sweden collects these data through a quarterly industrial capacity utilization survey (*Industrins kapacitetsutnyttjande*) conducted at the business-unit level. The unit of observation is the business unit, as in NIER's survey (see section **??**). The survey covers business units in the mining and manufacturing sectors with at least 50 employees.[B1] We aggregate the data to the firm level, since fewer than two percent of the firms in the data have multiple business units in a given period.

The capacity utilization data is an unbalanced panel at quarterly frequency spanning 1998Q1-2023Q1. There are 91,263 observations distributed across 2,987 firms, comprising two main variables. The first is capacity utilization, defined as the ratio between a firm's actual production level and full production capacity. Full production capacity, in turn, is defined

[B1]More specifically, the sampling procedure involves including the largest business units in an industry and then moving down the size distribution until the included business units jointly account for 95 percent of the industry's sales. Since 2009, however, only business units with at least 50 employees are asked about capacity utilization. To ensure that we have a consistently defined sample over time, we drop firms with fewer than 50 employees in the years before 2009.

as the maximum level of production that the firm can achieve with its current machinery and normal staffing under currently prevailing production methods. Respondents are instructed to disregard seasonal factors like vacations and to consider working hours that "may be deemed normal" when assessing their production capacity. This implies, for example, that firms should not revise estimated capacity downwards when furloughing workers, but should do so after laying off staff permanently. The survey also specifies reported utilization may exceed 100 percent if a firm employs overtime labor.

Second, firms reporting capacity utilization levels below 100 percent are asked why. They select one or several of the following pre-specified options: lack of high-skilled workers, lack of other workers, lack of intermediate inputs, insufficient demand, production disturbances, and other reasons.

B2 Empirical model

Our alternative test of the relationship between capacity utilization and demand uncertainty is based on the following model, which follows the baseline model closely:

$$CU_{i,t} = \beta \cdot DU_{j(i),t} + \gamma \cdot e_{i,t} + \mathbf{\Omega} \cdot \mathbf{X_{i,t-4}} + \alpha_i + \psi_t + \varepsilon_{i,t}. \tag{B1}$$

The dependent variable, $CU_{i,t}$, is the capacity utilization reported by firm i in Statistics Sweden's survey in period t (see Section **??**). $e_{i,t}$ is firm i's realized demand shock in period t, defined as the residual from the following first-order autoregressive model for sales growth:

$$\Delta y_{i,t} = \rho \cdot \Delta y_{i,t-1} + \mu_i + \lambda_{j,t} + e_{i,t}, \tag{B2}$$

where $\Delta y_{i,t}$ is firm i's sales growth between periods $t-1$ and t, μ_i is a firm fixed effect, and $\lambda_{j,t}$ is an industry-year fixed effect. Demand uncertainty, $DU_{j(i),t}$, is in turn measured as the inter-quartile range of $e_{i,t}$ in firm i's five-digit SNI/NACE industry j in year t, which is similar to how Bloom et al. (2018) measure productivity uncertainty.[B2] $\mathbf{X_{i,t-4}}$ is a vector comprising the following time-varying firm controls measured in period $t-4$: total assets in logs, the number of employees in logs, age (number of years since incorporation) in logs, and the ratios of cash holdings to total assets, liabilities to total assets, and EBIT to sales, respectively. α_i and ψ_t,

[B2] We choose the inter-quartile range as our dispersion measure because it corresponds closely to dispersive orderings, which our theoretical predictions are based on. We exclude observations belonging to industry-year cells with fewer than 12 firms from the estimations, since a dispersion-based measure like $DU_{j(i),t}$ requires a minimum number of observations to make sense.

finally, are firm and time fixed effects. Standard errors are two-way clustered at the firm and industry-year levels.

The coefficient of interest is β, which captures the percentage point change in capacity utilization following a one-unit increase in demand uncertainty. We assess the magnitude of the estimated effects as $-\beta \cdot \sigma^{DU} / \left(1 - \overline{CU}\right)$, which corresponds to the estimated increase in idle capacity relative to the sample mean following a one standard-deviation increase in demand uncertainty.

B3 Results

The results are presented in Table B1. Column (1), which reports the results when estimating (B1) without time-varying firm controls, shows that a one-unit increase in demand uncertainty is associated with a statistically significant decrease in capacity utilization of 3.5 percentage points; hence, a one-standard-deviation increase in demand uncertainty (0.122) leads idle capacity to increase by 0.4 percentage points, which corresponds 3.5 percent of mean idle capacity in the sample (12 percent). This estimate is robust to augmenting the model with the additional firm-level controls collected in $\mathbf{X_{i,t-4}}$, as shown in column (2).

In the final two columns of Table B1, we report results from estimations of (B1) on two subsamples of firms: one comprising firms that either operate at full capacity or operate below capacity because of "insufficient demand" (column 3), and one comprising firms that either operate at full capacity or operate below full capacity for non-demand related reasons (column 4). The motivation is the same as for the corresponding tests in the main part of the paper: if our baseline finding indeed is explained by demand uncertainty, we should only find a significant effect in the former group, because there is no reason to expect demand uncertainty to explain variation in capacity utilization driven by, say, production disruptions or insufficient access to intermediate inputs. The estimation reported in column (4) thus amounts to a placebo test. The results turn out in line with expectations. The estimate of β that we obtain when dropping firms that operate below capacity for non-demand reasons is statistically significant and similar in magnitude to the baseline estimate (column 4), whereas the estimate we obtain when dropping firms that operate below capacity for demand reasons is smaller and statistically insignificant (column 5).

Table B1: The effect of demand uncertainty on capacity utilization

	Dependent variable: Capacity utilization ($CU_{i,t}$)			
	(1)	(2)	(3)	(4)
$DU_{j(i),t}$	–0.035***	–0.033***	–0.030***	–0.012
	(0.011)	(0.010)	(0.011)	(0.008)
$e_{i,t}$	0.096***	0.115***	0.118***	0.041***
	(0.007)	(0.008)	(0.008)	(0.007)
Firm and time fixed effects	Yes	Yes	Yes	Yes
Time-varying firm controls	No	Yes	Yes	Yes
Adjusted R^2	0.474	0.483	0.560	0.419
Number of observations	83,929	82,511	64,189	45,556
Number of firms	2,587	2,575	2,485	2,176
$-\widehat{\beta} \cdot \sigma^{DU} / \left(1 - \overline{CU}\right)$	–0.035	–0.034	–0.030	–0.012

This table reports estimation results for the regression specified in (B1). The set of time-varying controls comprise one-year lags of total assets in logs, sales in logs, employment in logs, firm age in logs, and the ratios of cash to total assets, liabilities to total assets, and EBIT to sales, respectively. The estimation sample in column (3) comprises firms that either operate at full capacity or operate below capacity because of "insufficient demand," while the sample in column (4) comprises firms that either operate at full capacity or operate below capacity for non-demand related reasons. Observations belonging to industry-year cells with fewer than 12 firms are excluded from the estimations. The number in the bottom row is the estimated effect of a one-standard deviation increase in demand uncertainty on capacity utilization, expressed as a share of mean idle capacity in the sample. Standard errors are two-way clustered at the firm and industry-year level, respectively, in all regressions. *, **, and *** denote statistical significance at the ten, five, and one percent levels, respectively.

B4 Comparing the estimates

In closing, let us compare the estimates reported in Table B1 with the baseline estimates reported in Table 2 in the main part of the paper. The two sets of estimates are qualitatively similar: higher demand uncertainty is in both cases associated with lower capacity utilization. The effect estimates reported in Table B1 are, however, an order of magnitude smaller than the baseline estimates. The likely reason for this is that the latter are based on a quite precise measure of firms' subjective, forward-looking uncertainty, whereas the alternative estimates are based on an industry-level proxy.

More specifically, our NIER data indicates that variation in subjective uncertainty at the firm level to a large extent is idiosyncractic: when we regress dummies corresponding to each of the four possible responses to the question about how difficult it is to predict the firm's future business situation on industry-period fixed effects, we obtain R^2s between 0.20 and 0.24. Hence, industry-level factors explain only a small part of the variation in firm-level uncertainty. Our alternative estimates are thus likely to suffer from attenuation bias, which pulls the coefficient estimates towards zero. That the alternative estimates have the right sign and are statistically significant nevertheless suggests that our baseline results are not specific to the time period covered by the NIER data. This alleviates the concern that the external validity of our baseline results are limited because of the shortness of the sample period.

Appendix C. Measuring Markups

We estimate the markups used in Section 3.4 of the paper using the production-function approach proposed by De Loecker and Warzynski (2012). In what follows, we describe in more detail how we implement their method using our data.

We estimate markups using annual financial accounts data from Serrano, which span the years 1998-2021 (see Section 3.1). Serrano comprises the universe of incorporated Swedish firms, but we restrict the sample used for the markup estimation to firms above a size threshold. The reasons are that (i) Serrano includes a very large number of firms that are considerably smaller than the smallest firms in our sample, and (ii) production functions may differ across firms of different sizes. Estimating markups using all firms in Serrano would therefore risk making the estimates less precise. More specifically, we exclude firm-year observations for which total assets, sales, or employment fall below the first percentile of the respective distributions among the firms in our sample. We thus drop firm-years with sales below 25.0 million SEK, total assets below 14.2 million SEK, or number of employees below 18.

We compute the markups using Rovigatti's (2020) Stata module `markupest`. For the production-function estimation, we follow Levinsohn and Petrin (2003) and Ackerberg, Caves and Frazer (2015). We use value added as the measure of output, the book value of physical capital (buildings, land and machinery) as state variable, labor costs as free variable, and the value of intermediate input purchases as proxy variable. All variables are deflated using producer price indices defined at the level of two-digit SNI/NACE codes. We conduct the estimation at the level of five-digit SNI/NACE codes and only include observations belonging to industries with least 250 observations. The mean and median number of observations per industry in the estimations are 744 and 501, respectively, whereas the mean and median number of firms per industry are 77 and 47, respectively.

We report descriptive statistics on the resulting markup estimates in Table A1 in Online Appendix A.

Appendix D. Model Details

This appendix states the assumptions used in Section 2, proves the main-text results, and records the extensions and limits that matter for interpretation.

D1 Assumptions, the capacity rule, and the reduced price problem

For each posted price $p > c$ and uncertainty state θ, demand is nonnegative and integrable. The conditional cdf $F_D(\cdot \mid p, \theta)$ is continuous and strictly increasing on the support relevant for the optimum, and the corresponding quantile function $Q_D(\cdot \mid p, \theta)$ is continuously differentiable wherever derivatives are taken. Under multiplicative demand, $d(p) > 0$, $d'(p) < 0$, $d(\cdot)$ is continuous on $[c, \infty)$, and $pd(p) \to 0$ as $p \to \infty$.

A sufficient condition for an interior capacity choice is strictly positive support, $F_D(0 \mid p, \theta) = 0$. More generally, the interior quantile rule applies whenever

$$p\,[1 - F_D(0 \mid p, \theta)] > c.$$

If this fails, the boundary choice $K^*(p, \theta) = 0$ is optimal. The main text works on the interior region, which is the case relevant for the empirical interpretation.

A useful identity for expected sales follows from the fact that, for nonnegative demand,

$$\min\{D, K\} = \int_0^K \mathbf{1}\{D \geq x\}\, dx.$$

Taking expectations and applying Tonelli's theorem gives

$$\mathbb{E}[\min\{D, K\} \mid p, \theta] = \int_0^K \Pr(D \geq x \mid p, \theta)\, dx = \int_0^K [1 - F_D(x \mid p, \theta)]\; dx. \tag{D1}$$

Proof of Proposition 1. Substituting (D1) into profits gives

$$\Pi(p, K; \theta) = p \int_0^K [1 - F_D(x \mid p, \theta)]\; dx - cK. \tag{D2}$$

Differentiating with respect to K,

$$\Pi_K(p, K; \theta) = p\,[1 - F_D(K \mid p, \theta)] - c. \tag{D3}$$

Because $F_D(\cdot \mid p, \theta)$ is increasing, Π_K is strictly decreasing in K, so the capacity problem is strictly concave on the interior region. The first-order condition therefore yields

$$F_D(K^*(p,\theta) \mid p,\theta) = 1 - \frac{c}{p} \equiv \alpha(p),$$

and since $F_D(\cdot \mid p,\theta)$ is continuous and strictly increasing,

$$K^*(p,\theta) = Q_D(\alpha(p) \mid p,\theta)\,. \tag{D4}$$

That proves Proposition 1.

Under multiplicative demand, $D = d(p)\varepsilon$, so optimal capacity becomes

$$K^*(p,\theta) = d(p)\,Q(\alpha(p);\theta)\,, \tag{D5}$$

where $Q(\cdot;\theta)$ is the quantile function of ε. Expected sales at that optimum are

$$\mathbb{E}[\min\{D, K^*(p,\theta)\} \mid p,\theta] = d(p)\left[\int_0^{\alpha(p)} Q(u;\theta)\,du + (1-\alpha(p))\,Q(\alpha(p);\theta)\right]. \tag{D6}$$

Substituting (D5)–(D6) into profits and using $c = p(1-\alpha(p))$ yields the one-dimensional price problem

$$\max_{p>c}\; pd(p)\int_0^{1-c/p} Q(u;\theta)\,du. \tag{D7}$$

Equivalently, writing $p = c/(1-\alpha)$,

$$\Pi^r(\alpha,\theta) = \frac{c}{1-\alpha}\,d\!\left(\frac{c}{1-\alpha}\right)\int_0^{\alpha} Q(u;\theta)\,du. \tag{D8}$$

As $\alpha \downarrow 0$, the integral vanishes, so $\Pi^r(\alpha,\theta) \to 0$. As $\alpha \uparrow 1$, the integral is bounded above by $\mathbb{E}[\varepsilon \mid \theta] < \infty$, while the prefactor equals $pd(p) \to 0$ by assumption. Hence Π^r extends continuously to $[0,1]$ with zero endpoints and therefore attains a maximum. Because the objective is strictly positive at any interior α when demand has positive support, at least one maximizer is interior.

A convenient sufficient condition for uniqueness is that, for each θ, both

$$\alpha \mapsto \log\left[\frac{c}{1-\alpha}\,d\left(\frac{c}{1-\alpha}\right)\right] \quad \text{and} \quad \alpha \mapsto \log\left(\int_0^{\alpha} Q(u;\theta)\,du\right)$$

are concave on $(0,1)$, with at least one strictly concave. Then $\log \Pi^r(\alpha,\theta)$ is strictly concave, so the maximizer is unique.

Any interior price optimizer also satisfies

$$0 = (pd(p))' \int_0^{\alpha(p)} Q(u;\theta)\,du + pd(p)\,Q(\alpha(p);\theta)\,\alpha_p(p), \qquad \alpha_p(p) = \frac{c}{p^2}. \tag{D9}$$

We do not use (D9) in the empirical mapping, but it clarifies how a higher posted price trades off the usual direct revenue effect against the larger capacity buffer induced by a higher markup state.

D2 Utilization, uncertainty, and markup

A second useful identity is

$$\mathbb{E}[\min\{\varepsilon, Q(\alpha;\theta)\} \mid \theta] = \int_0^{\alpha} Q(u;\theta)\,du + (1-\alpha)Q(\alpha;\theta). \tag{D10}$$

To see this, write

$$\mathbb{E}[\min\{\varepsilon, Q(\alpha;\theta)\} \mid \theta] = \mathbb{E}[\varepsilon\,\mathbf{1}\{\varepsilon \le Q(\alpha;\theta)\} \mid \theta] + Q(\alpha;\theta)\,\Pr(\varepsilon > Q(\alpha;\theta) \mid \theta)\,,$$

and then change variables from the realization of ε to its cumulative-probability index $u = F_\varepsilon(\varepsilon;\theta)$ in the first term. Dividing (D10) by $Q(\alpha;\theta)$ gives

$$U(\alpha,\theta) = \frac{\int_0^{\alpha} Q(u;\theta)\,du}{Q(\alpha;\theta)} + (1-\alpha). \tag{D11}$$

Define

$$I(\alpha,\theta) \equiv \int_0^{\alpha} Q(u;\theta)\,du, \qquad R(\alpha,\theta) \equiv \frac{I(\alpha,\theta)}{Q(\alpha;\theta)}, \qquad U(\alpha,\theta) = R(\alpha,\theta) + (1-\alpha). \tag{D12}$$

Proof of Proposition 2. Fix $\alpha \in (0,1)$ and write

$$Q_L(u) \equiv Q(u;\theta_L), \qquad Q_H(u) \equiv Q(u;\theta_H), \qquad h(u) \equiv Q_L(u) - Q_H(u).$$

A mean- and median-preserving dispersive increase implies that h is decreasing on $(0,1)$, that $h(1/2) = 0$, and that

$$\int_0^1 h(u)\,du = \int_0^1 Q_L(u)\,du - \int_0^1 Q_H(u)\,du = \mathbb{E}[\varepsilon \mid \theta_L] - \mathbb{E}[\varepsilon \mid \theta_H] = 0.$$

Define

$$q_j \equiv Q_j(\alpha), \qquad x_j \equiv \int_0^\alpha Q_j(u)\,du + (1-\alpha)Q_j(\alpha), \qquad U_j \equiv \frac{x_j}{q_j}, \qquad j \in \{L, H\}.$$

We need to show $U_H \le U_L$.

Case 1: $\alpha \ge 1/2$. Since h is decreasing and $h(1/2) = 0$, one has $q_H \ge q_L$. Also,

$$x_L - x_H = \int_0^\alpha h(u)\,du + (1-\alpha)h(\alpha).$$

Because h is decreasing, $h(u) \le h(\alpha)$ for all $u \in [\alpha, 1]$. Hence

$$\int_\alpha^1 h(u)\,du \le (1-\alpha)h(\alpha).$$

Using $\int_0^1 h(u)\,du = 0$,

$$0 = \int_0^\alpha h(u)\,du + \int_\alpha^1 h(u)\,du \le \int_0^\alpha h(u)\,du + (1-\alpha)h(\alpha) = x_L - x_H.$$

So $x_H \le x_L$. Since also $q_H \ge q_L > 0$,

$$U_H = \frac{x_H}{q_H} \le \frac{x_L}{q_H} \le \frac{x_L}{q_L} = U_L.$$

Case 2: $\alpha < 1/2$. Now $q_H \le q_L$. Rewrite utilization as

$$1 - U_j = \frac{\int_0^\alpha \left[q_j - Q_j(u)\right]\,du}{q_j}.$$

Because h is decreasing and $u < \alpha$, we have $h(u) \ge h(\alpha)$. Therefore

$$q_H - Q_H(u) = q_L - Q_L(u) + h(u) - h(\alpha) \ge q_L - Q_L(u)$$

for every $u \in (0, \alpha)$. Integrating and using $q_H \le q_L$ gives

$$1 - U_H = \frac{\int_0^\alpha \left[q_H - Q_H(u)\right]\,du}{q_H} \ge \frac{\int_0^\alpha \left[q_L - Q_L(u)\right]\,du}{q_L} = 1 - U_L.$$

Hence $U_H \le U_L$.

The two cases together prove Proposition 2.

Proof of Proposition 3. Using (D12),

$$R_\alpha(\alpha,\theta) = \frac{Q(\alpha;\theta)^2 - I(\alpha,\theta)Q_\alpha(\alpha;\theta)}{Q(\alpha;\theta)^2} = 1 - R(\alpha,\theta)\frac{Q_\alpha(\alpha;\theta)}{Q(\alpha;\theta)}.$$

Subtracting one gives

$$U_\alpha(\alpha,\theta) = -R(\alpha,\theta)\,\partial_\alpha \log Q(\alpha;\theta). \tag{D13}$$

Quantiles are increasing in their probability index, so $Q_\alpha(\alpha;\theta) > 0$ and therefore $U_\alpha(\alpha,\theta) < 0$. That proves Proposition 3.

D3 The interaction term and the finite-difference mapping

Proof of Proposition 4. Differentiate (D12) with respect to θ:

$$R_\theta(\alpha,\theta) = \frac{\int_0^\alpha Q_\theta(u;\theta)\,du}{Q(\alpha;\theta)} - \frac{I(\alpha,\theta)Q_\theta(\alpha;\theta)}{Q(\alpha;\theta)^2}.$$

Using $Q_\theta = Q\,\partial_\theta \log Q$ and $R = I/Q$ gives

$$U_\theta(\alpha,\theta) = R(\alpha,\theta)\left[\frac{\int_0^\alpha Q(u;\theta)\,\partial_\theta \log Q(u;\theta)\,du}{\int_0^\alpha Q(u;\theta)\,du} - \partial_\theta \log Q(\alpha;\theta)\right]. \tag{D14}$$

This is the fixed-markup uncertainty effect written as a comparison between the uncertainty sensitivity of the endpoint quantile and the weighted-average sensitivity of the lower quantiles.

Now differentiate (D13) with respect to θ:

$$U_{\alpha\theta}(\alpha,\theta) = -R_\theta(\alpha,\theta)\,\partial_\alpha \log Q(\alpha;\theta) - R(\alpha,\theta)\,\partial_{\alpha\theta} \log Q(\alpha;\theta).$$

Substituting (D14) yields

$$\begin{aligned} U_{\alpha\theta}(\alpha,\theta) = -R(\alpha,\theta)\Bigg\{&\partial_{\alpha\theta}\log Q(\alpha;\theta) \\ &+ \partial_\alpha \log Q(\alpha;\theta)\left[\frac{\int_0^\alpha Q(u;\theta)\,\partial_\theta \log Q(u;\theta)\,du}{\int_0^\alpha Q(u;\theta)\,du} - \partial_\theta \log Q(\alpha;\theta)\right]\Bigg\}. \end{aligned} \tag{D15}$$

Because $R(\alpha,\theta) > 0$, rearranging (D15) gives the condition in (16). That proves Proposition 4.

Proof of Corollary 5. If $U_{\alpha\theta}(a,\vartheta)$ is integrable on $[\alpha_L,\alpha_H]\times[\theta_L,\theta_H]$, then iterated integration gives

$$\Big[U(\alpha_H,\theta_H) - U(\alpha_H,\theta_L)\Big] - \Big[U(\alpha_L,\theta_H) - U(\alpha_L,\theta_L)\Big] = \int_{\alpha_L}^{\alpha_H}\int_{\theta_L}^{\theta_H} U_{\alpha\theta}(a,\vartheta)\,d\vartheta\,da. \tag{D16}$$

If $U_{\alpha\theta}(a,\vartheta) \le 0$ throughout that rectangle, the double integral is nonpositive as well. That proves Corollary 5.

Nothing in this mapping requires solving for the total response of the optimized markup $\alpha^*(\theta)$. The empirical interaction corresponds to the finite difference in (D16), which is why Section 2 focuses on state-space comparisons across markup and uncertainty bins rather than on the derivative of one firm's optimizer.

D4 A broader positive-support class

It is useful to record a broader class of positive-support demand shocks for which the main objects in Section 2 take a similar form. Suppose the quantile function of ϵ can be written as

$$\log Q(\alpha;\theta) = \nu(\theta) + \theta s(\alpha), \qquad s'(\alpha) > 0.$$

Equivalently,

$$Q(\alpha;\theta) = \exp\{\nu(\theta) + \theta s(\alpha)\}.$$

This specification is automatically consistent with the multiplicative demand system $D = d(p)\epsilon$ with $\epsilon > 0$. It covers the lognormal case and, more generally, any case in which $\log\epsilon$ belongs to a location-scale family. For example,

$$s(\alpha) = \Phi^{-1}(\alpha) \quad \text{(normal)}, \qquad s(\alpha) = \log\left(\frac{\alpha}{1-\alpha}\right) \quad \text{(logistic)}, \qquad s(\alpha) = -\log(-\log\alpha) \quad \text{(Gumbel)}.$$

From the definition of $R(\alpha,\theta)$,

$$R(\alpha,\theta) = \frac{\int_0^\alpha Q(u;\theta)\,du}{Q(\alpha;\theta)} = \int_0^\alpha e^{\theta[s(u)-s(\alpha)]}\,du,$$

so expected utilization becomes

$$U(\alpha,\theta) = \int_0^\alpha e^{\theta[s(u)-s(\alpha)]}\,du + (1-\alpha).$$

Thus the level term $\nu(\theta)$ drops out of utilization, just as the scale term $d(p)$ drops out under multiplicative demand in the main text.

The derivatives entering Proposition 4 are

$$\partial_\alpha \log Q(\alpha;\theta) = \theta s'(\alpha), \qquad \partial_\theta \log Q(\alpha;\theta) = \nu'(\theta) + s(\alpha), \qquad \partial_{\alpha\theta} \log Q(\alpha;\theta) = s'(\alpha).$$

The weighted-average term in the local interaction condition becomes

$$\frac{\int_0^\alpha Q(u;\theta)\,\partial_\theta \log Q(u;\theta)\,du}{\int_0^\alpha Q(u;\theta)\,du} = \nu'(\theta) + \frac{\int_0^\alpha e^{\theta[s(u)-s(\alpha)]}s(u)\,du}{\int_0^\alpha e^{\theta[s(u)-s(\alpha)]}\,du}.$$

Substituting into the condition in Proposition 4, the common term $\nu'(\theta)$ cancels, so the local interaction condition reduces to

$$U_{\alpha\theta}(\alpha,\theta) \le 0 \iff 1 \ge \theta\left[s(\alpha) - \frac{\int_0^\alpha e^{\theta[s(u)-s(\alpha)]}s(u)\,du}{\int_0^\alpha e^{\theta[s(u)-s(\alpha)]}\,du}\right].$$

This shows that the sign condition in Proposition 4 is not specific to the lognormal case. What matters is how uncertainty stretches the upper-tail quantile relative to the lower quantiles that govern expected sales. The lognormal worked example in the next subsection is therefore a transparent special case rather than a knife-edge one.

D5 The lognormal illustration

Let

$$Q(\alpha;\sigma) = \exp\{\nu(\sigma) + \sigma z_\alpha\}, \qquad z_\alpha \equiv \Phi^{-1}(\alpha). \tag{D17}$$

Because the location term $\nu(\sigma)$ cancels from utilization, the normalization of mean demand is irrelevant for the formulas below. Then

$$\begin{aligned}
\int_0^\alpha Q(u;\sigma)\,du &= \int_0^\alpha \exp\{\nu(\sigma) + \sigma\Phi^{-1}(u)\}\,du \\
&= \exp\{\nu(\sigma)\}\int_{-\infty}^{z_\alpha} e^{\sigma z}\phi(z)\,dz \\
&= \exp\{\nu(\sigma) + \sigma^2/2\}\int_{-\infty}^{z_\alpha} \phi(z-\sigma)\,dz \\
&= \exp\{\nu(\sigma) + \sigma^2/2\}\Phi(z_\alpha - \sigma).
\end{aligned} \tag{D18}$$

Dividing (D18) by $Q(\alpha;\sigma)$ and using (D11) yields

$$U(\alpha,\sigma)=e^{\sigma^2/2-\sigma z_\alpha}\Phi(z_\alpha-\sigma)+(1-\alpha), \tag{D19}$$

which is equation (17) in the main text.

To obtain the interaction condition, specialize (16). Here

$$\partial_\alpha \log Q(\alpha;\sigma)=\sigma\,\partial_\alpha z_\alpha, \qquad \partial_\sigma \log Q(\alpha;\sigma)=\nu'(\sigma)+z_\alpha, \qquad \partial_{\alpha\sigma} \log Q(\alpha;\sigma)=\partial_\alpha z_\alpha.$$

The common scale term $\nu'(\sigma)$ cancels from the bracketed difference in (16). For the weighted-average term,

$$\frac{\int_0^\alpha Q(u;\sigma)\,\partial_\sigma \log Q(u;\sigma)\,du}{\int_0^\alpha Q(u;\sigma)\,du}=\nu'(\sigma)+\sigma-\frac{\phi(z_\alpha-\sigma)}{\Phi(z_\alpha-\sigma)}. \tag{D20}$$

Substituting (D20) into (16) and canceling the positive factor $\partial_\alpha z_\alpha$ gives

$$U_{\alpha\sigma}(\alpha,\sigma)\le 0 \iff 1\ge\sigma\left[z_\alpha-\sigma+\frac{\phi(z_\alpha-\sigma)}{\Phi(z_\alpha-\sigma)}\right], \tag{D21}$$

which is equation (18) in the main text.

D6 Mapping to the empirical interaction

Let $\theta_H>\theta_L$ denote two uncertainty states and $\alpha_H>\alpha_L$ two markup states. Then

$$\Big[U(\alpha_H,\theta_H)-U(\alpha_H,\theta_L)\Big]-\Big[U(\alpha_L,\theta_H)-U(\alpha_L,\theta_L)\Big]=\int_{\alpha_L}^{\alpha_H}\int_{\theta_L}^{\theta_H}U_{\alpha\theta}(a,\vartheta)\,d\vartheta\,da. \tag{D22}$$

Corollary 5 (Mapping to the empirical interaction) *If $U_{\alpha\theta}(a,\vartheta)\le 0$ for all $(a,\vartheta)\in[\alpha_L,\alpha_H]\times[\theta_L,\theta_H]$, then*

$$\Big[U(\alpha_H,\theta_H)-U(\alpha_H,\theta_L)\Big]-\Big[U(\alpha_L,\theta_H)-U(\alpha_L,\theta_L)\Big]\le 0. \tag{D23}$$

Corollary 5 is the model analogue of the uncertainty-by-markup interaction estimated in the data. It says that the utilization gap between high- and low-uncertainty firms should be more negative among high-markup firms. The mapping is ordinal rather than literal: the empirical markup measures need not coincide exactly with $\alpha(p)$. What matters is that they rank firms by the value of serving one more unit of demand in high-demand states. Firms with more to lose from stockouts respond to uncertainty by carrying more slack capacity.

D7 Extensions

A constant per-unit production cost. Suppose each unit sold also requires a constant variable production cost m, so expected profits become

$$\Pi_m(p, K; \theta) = (p - m)\,\mathbb{E}[\min\{D, K\} \mid p, \theta] - cK, \qquad p > m + c. \tag{D24}$$

The capacity first-order condition is now

$$(p - m)\,[1 - F_D(K \mid p, \theta)] = c,$$

so the relevant state variable becomes

$$\alpha_m(p) \equiv 1 - \frac{c}{p - m}. \tag{D25}$$

Hence the capacity rule remains quantile-based:

$$K_m^*(p, \theta) = Q_D(\alpha_m(p) \mid p, \theta)\,. \tag{D26}$$

Under multiplicative demand,

$$K_m^*(p, \theta) = d(p)\,Q(\alpha_m(p); \theta), \qquad U_m(p, \theta) = \frac{\int_0^{\alpha_m(p)} Q(u; \theta)\,du}{Q(\alpha_m(p); \theta)} + [1 - \alpha_m(p)]\,, \tag{D27}$$

and the reduced price problem becomes

$$\max_{p > m + c}\,(p - m) d(p) \int_0^{\alpha_m(p)} Q(u; \theta)\,du. \tag{D28}$$

So the mathematics of the baseline is unchanged; what changes is the interpretation of the markup state.

Additive demand. Now suppose demand is additive rather than multiplicative:

$$D = d(p) + \varepsilon, \qquad d'(p) < 0. \tag{D29}$$

Assume $d(p) + \varepsilon \geq 0$ almost surely and $K_A^*(p, \theta) > 0$ on the region of interest. Since $F_D(K \mid$

$p, \theta) = F_\varepsilon(K - d(p); \theta)$, the capacity rule becomes

$$K_A^*(p, \theta) = d(p) + Q_\varepsilon(\alpha(p); \theta). \tag{D30}$$

At that optimum,

$$U_A(p, \theta) = \frac{d(p) + \int_0^{\alpha(p)} Q_\varepsilon(u; \theta)\, du + [1 - \alpha(p)]\, Q_\varepsilon(\alpha(p); \theta)}{d(p) + Q_\varepsilon(\alpha(p); \theta)}. \tag{D31}$$

Holding $M = d(p)$ fixed and writing $Q(\alpha; \theta) = Q_\varepsilon(\alpha; \theta)$,

$$\partial_\alpha U_A(M, \alpha, \theta) = -\frac{\left[\alpha M + \int_0^\alpha Q(u; \theta)\, du\right] Q_\alpha(\alpha; \theta)}{[M + Q(\alpha; \theta)]^2} < 0. \tag{D32}$$

So the basic markup effect survives.

The fixed-markup uncertainty effect also survives on an economically relevant region. Define

$$U_\varepsilon(\alpha, \theta) \equiv \frac{\int_0^\alpha Q(u; \theta)\, du}{Q(\alpha; \theta)} + (1 - \alpha),$$

so that

$$U_A(M, \alpha, \theta) = \frac{M + U_\varepsilon(\alpha, \theta) Q(\alpha; \theta)}{M + Q(\alpha; \theta)}.$$

For fixed M, this expression is increasing in U_ε and decreasing in Q because

$$\frac{\partial U_A}{\partial U_\varepsilon} = \frac{Q}{M + Q} > 0, \qquad \frac{\partial U_A}{\partial Q} = \frac{M(U_\varepsilon - 1)}{(M + Q)^2} \leq 0.$$

By Proposition 2, $U_\varepsilon(\alpha, \theta_H) \leq U_\varepsilon(\alpha, \theta_L)$. And when $\alpha \geq 1/2$, dispersive order with common mean and median implies $Q(\alpha; \theta_H) \geq Q(\alpha; \theta_L)$. Therefore

$$U_A(M, \alpha, \theta_H) \leq U_A(M, \alpha, \theta_L) \qquad \text{for } \alpha \geq \frac{1}{2}. \tag{D33}$$

What is lost is the clean reduction to a two-dimensional state space indexed only by markup and uncertainty. Under additive demand, the level term $M = d(p)$ enters utilization directly, so the interaction becomes a property of (M, α, θ) rather than just (α, θ).

D8 Limits and scope

Why there is no global sign theorem. The local interaction need not be negative everywhere. A simple counterexample is

$$Q(\alpha, \theta) = \alpha + \theta\sqrt{\alpha}. \tag{D34}$$

Then

$$U(\alpha, \theta) = \frac{\alpha^2/2 + (2/3)\theta\alpha^{3/2}}{\alpha + \theta\sqrt{\alpha}} + (1 - \alpha),$$

so differentiating first with respect to θ and then with respect to α gives

$$U_{\alpha\theta}(\alpha, 0) = \frac{1}{12\sqrt{\alpha}} > 0. \tag{D35}$$

A blanket global sign claim would therefore be too strong.

Attenuation in the extreme upper tail. If $\mathbb{E}[\varepsilon \mid \theta] < \infty$ and $Q(\alpha; \theta) \to \infty$ as $\alpha \uparrow 1$, then

$$R(\alpha, \theta) = \frac{\int_0^\alpha Q(u; \theta)\, du}{Q(\alpha; \theta)} \to 0, \qquad U(\alpha, \theta) = R(\alpha, \theta) + (1 - \alpha) \to 0. \tag{D36}$$

The numerator in $R(\alpha, \theta)$ stays bounded by $\mathbb{E}[\varepsilon \mid \theta]$, while the denominator diverges. Hence for any two uncertainty levels $\theta_H > \theta_L$,

$$U(\alpha, \theta_H) - U(\alpha, \theta_L) \to 0 \qquad \text{as } \alpha \uparrow 1. \tag{D37}$$

Even when the interaction is negative over empirically relevant markup ranges, differences across uncertainty levels must therefore attenuate in the extreme upper tail.

Scope of the model. These limits do not weaken the role the model plays in the paper. The point is not to claim a universal sign theorem. It is to provide a disciplined link between forward-looking demand uncertainty, markup heterogeneity, and planned slack capacity. The model is a theory of ex ante price choice and ex ante capacity choice under demand uncertainty. It is not a model of within-period scarcity pricing after the demand shock is realized. That timing is the one that fits the empirical application.