Dispersion Over the Business Cycle: Passthrough, Productivity, and Demand

Mikael Carlsson*, Alex Clymo†, Knut-Eric Joslin‡

January 15, 2023

Abstract

We characterize the cyclical dispersion of firm-level (physical) productivity and demand shocks using Swedish microdata that includes prices and a measure of utilization. We then analyze the consequences of dispersion for firms and the macroeconomy. We find that demand dispersion increases by more than productivity dispersion in recessions and that demand dispersion explains most of the variation in sales dispersion. Productivity shocks on the other hand pass through incompletely to prices and have only a limited effect on sales dispersion. In a heterogeneous-firm model matching the micro facts, demand dispersion has unambiguously negative effects on output via a "wait and see" channel. Productivity dispersion does not generate "wait and see" effects, but affects output negatively by inducing markup dispersion.

Keywords: Demand estimation, Productivity estimation, Variable markups, Business cycles, Dispersion, Uncertainty, Passthrough, Adjustment costs, Heterogeneous firm model.

JEL classification: D21, D22, D81, E32, L11.

*Uppsala University, UCLS and Sveriges Riksbank. Email: mikael.carlsson@nek.uu.se.

†University of Essex. Email: a.clymo@essex.ac.uk.

‡Kristiania University College. Email: knut-eric.joslin@kristiania.no.

We thank Christian Bayer, Matthias Kehrig, and Kjetil Storesletten for comments, as well as seminar participants at the 2019 SED meetings, 11th Nordic Conference on Register Data and Economic Modelling, 13th Nordic Summer Symposium in Macroeconomics, 1st Oslo Macro Conference, University of Kent, University of Manchester, Uppsala University, Sveriges Riksbank, Lancaster Macro Workshop 2019, Queen Mary Macro Workshop 2019, Royal Holloway SaM 2019, Econometric Society World Congress 2020, and EEA Congress 2020. The data are confidential, but the authors' access is not exclusive. We are grateful for financial support from the Ragnar Söderbergs Stiftelse (grant E52-13), the Handelsbankens Forskningstiftelse (grant P18-0082) and the Swedish Research Council (grant 2018-04581). 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.

1 Introduction

Recessions are times of increased dispersion across firms. Firms are worse off on average, but the range of outcomes also widens. Not only is across-firm dispersion of sales, employment, and prices countercyclical, but so is the underlying dispersion of firm-level shocks. Studies such as Bachmann and Bayer (2013), Kehrig (2015), and Bloom et al. (2018) have demonstrated that countercyclical dispersion of revenue productivity shocks (TFPR) is a feature of recessions common across countries and periods. This also holds for the Swedish recessions of 2001 and 2009.[1] Moreover, dispersion may itself play a role in propagating the business cycle. Pioneering work by Bloom (2009) has demonstrated how "uncertainty shocks" can have aggregate implications via a wait-and-see channel.[2]

In this paper, we delve deeper into the nature of cyclical dispersion using rich Swedish register data and then assess the role of dispersion for business cycles. We use firm-level price and utilization data to distinguish dispersion in firm-level demand from dispersion in firm-level physical productivity. We deliver novel insights into how firms react to each type of shock, and how dispersion of these shocks contributes to aggregate fluctuations.

The first question that we tackle is a fundamental one: Which shocks are becoming more dispersed in recessions? Earlier work has focused on dispersion in revenue productivity. But dispersion in revenue productivity is driven by shocks to both physical productivity (TFPQ) and demand. Our first contribution is to disentangle these distinct sources of cyclical dispersion. We exploit price and utilization data to measure TFPQ at the firm level, and then use TFPQ innovations to estimate firm-level demand, an approach inspired by Foster et al. (2008). We find that the dispersion of both TFPQ shocks and demand shocks rises during recessions, but that the increase is greater for demand shocks.

A second question that we pursue is how these shocks transmit to prices and sales. We estimate how TFPQ and demand shocks pass through to prices using an approach suggested by De Loecker et al. (2016). In contrast to what a simple pricing model would predict, we find that firms adjust their prices to only a limited degree in response to productivity shocks ("incomplete passthrough") and raise their prices in response to positive demand shocks: A 1% improvement in TFPQ lowers prices by less than 0.3%, while a 1% increase in demand causes firms to raise their prices between 0.2% to 0.3%. In our demand curve estimation,

[1] When we measure TFP in the same fashion as Bloom et al. (2018), we reproduce their dispersion results.

[2] Greater dispersion of shocks is systematically associated with higher uncertainty. As in other countries, the Swedish recessions in 2001 and 2009 were associated with increases in measures of uncertainty, such as the EPU index (Armelius et al., 2017).

we find statistically and economically significant deviations from the constant elasticity of substitution (CES) benchmark. This finding can help explain the passthrough results.

Using variance decompositions, we then investigate the relative importance of the two shocks for business cycle cyclicality. We find that demand shocks are the main driver of sales growth dispersion, both on average and over the business cycle. In particular, the increased dispersion of demand shocks during the Great Recession explains about 80% of the rise in sales dispersion during that period. The relative unimportance of TFPQ dispersion follows from the low passthrough from TFPQ shocks to prices which makes it difficult for changes in TFPQ dispersion to contribute to sales dispersion. In contrast, demand shocks directly affect the amount that firms can sell at a given price. Hence, increases in demand dispersion naturally translate into increases in sales dispersion.

Finally, we build a dynamic heterogeneous-firm model with non-convex adjustment costs in order to understand the aggregate implications of cyclical dispersion. Interpreting increases in dispersion as uncertainty shocks, the non-convex adjustment costs create "wait and see" behavior in response to increases in uncertainty: Firms may choose to pause investment when uncertainty is high rather than risk the costs associated with reversing an investment decision.[3] We also incorporate our estimated non-CES demand curves into this framework. Together, these features can rationalize our main empirical findings. The model predicts low passthrough from TFPQ to prices and non-zero passthrough from demand to prices. Moreover, the non-CES induced "real rigidity" causes firms to respond differently to demand and productivity uncertainty.

We find that uncertainty shocks produce large declines in aggregate output. However, demand uncertainty is more important than TFPQ uncertainty. Demand shocks directly shift the amount firms can sell and hence their desired investment. As a consequence, demand uncertainty is a powerful driver of wait-and-see behavior.[4]

In contrast, we find that TFPQ uncertainty has limited importance for firms. The reason is that firms do not adjust their production and investment much in response to TFPQ shocks when demand is sufficiently non-CES. Instead, TFPQ shocks cause firms to adjust their markup. TFPQ thus plays an aggregate role not via uncertainty but via the effect of realized dispersion on markups. Increases in TFPQ dispersion drive increases in markup

[3] Interpreting idiosyncratic dispersion as uncertainty is supported by the results of Bachmann et al. (2021), who show that managers' perceived uncertainty about the short-run future follows the conditional volatility of shocks measured from a GARCH model.

[4] Guiso and Parigi (1999) and Bloom et al. (2022) measure uncertainty based on a survey of managers who report a subjective probability distribution for their plant's future demand. Like us, they conclude that firm level demand uncertainty has a negative effect on to investment.

dispersion, causing misallocation which reduces aggregate output. In fact, this effect counteracts the "volatility overshoot" present in many standard models (Bloom, 2009). The volatility overshoot effect arises when firms' optimal sales are convex in their underlying productivity shocks, so that an increase in realised dispersion increases aggregate GDP. Given a non-CES demand curve, however, the positive effect is offset by increases in markup dispersion. For the demand curves that we estimate, the total effect is negative because endogenous markup changes mean that firms with negative TFPQ shocks lose more sales than firms with positive TFPQ shocks gain. Our analysis thus reveals distinct roles for demand and TFPQ dispersion: Demand dispersion hurts output via uncertainty, while productivity dispersion hurts output via realised dispersion.[5]

Related Literature Our paper relates to four broad strands in the literature. First, we provide new results on the cyclicality of dispersion in firm-level shocks and outcomes. We separately measure TFPQ and demand shocks and characterize the dispersion of each. We thus complement earlier work by distinguishing between two shocks that underlie TFPR. Our results strengthen the evidence that shocks become more dispersed in recessions: Productivity shocks become more dispersed in recessions—even after controlling for utilization—and so too do demand shocks.[6]

Second, we contribute to a literature that estimates how firms respond to idiosyncratic shocks. A number of recent papers have, like us, investigated the separate roles of demand and TFPQ shocks, often with a focus on prices (Carlsson et al., 2016; De Loecker et al., 2016; Foster et al., 2016; Hottman et al., 2016; Pozzi and Schivardi, 2016; Haltiwanger et al., 2018; Eslava and Haltiwanger, 2020; Kaas and Kimasa, 2021). Relative to this literature, we evaluate the role of non-CES demand and show that it can rationalise incomplete passthrough of productivity shocks. We also find support for key conclusions from this literature. For example, we find that demand plays a larger role in driving firm behavior than productivity, and that the responsiveness to shocks may be time-varying (see Berger and Vavra, 2019 and

[5]Bachmann and Bayer (2013) and and Mongey and Williams (2017) build models with non-convex adjustment costs on capital only, and not labor, and find smaller aggregate effects of uncertainty shocks, in contrast to Bloom (2009) and Bloom et al. (2018) who place adjustment costs on both capital and labor. Our finding that increased TFPQ dispersion leads to a fall in aggregate output via the realised volatility effect creates aggregate impacts of cyclical dispersion even in the absence of any adjustment costs. In the paper, we discuss robustness of our results to adjustment cost assumptions.

[6]For outcomes, Davis et al. (1996) show that employment growth dispersion is countercyclical, Bloom et al. (2018) show the same for sales growth, and Vavra (2014) does the same for prices. An exception to countercyclical dispersion is investment. For example, Bachmann and Bayer (2014) show that investment dispersion is procyclical.

Decker et al., 2020). Although not the focus of our analysis, we also find indirect evidence that shocks besides productivity and demand play a role for prices. One possibility is financial constraints which are subsumed into the residual term in our variance decomposition. For example, Gilchrist et al. (2017) use data similar to our own to show that firms who were financially constrained during the financial crisis raised prices.

Third, we contribute to the literature on non-CES demand curves. Since at least Ball and Romer (1990), it has been known that non-constant demand elasticities can generate incomplete passthrough from marginal costs to prices.[7] However, papers that directly estimate non-CES demand curves using firm-level price and quantity data are rare. Those that do estimate demand using microdata, find significant deviations from CES (Arkolakis et al., 2018; Haltiwanger et al., 2018). We confirm this finding and then show how it changes the aggregate effects of cyclical dispersion.

Finally, we relate to papers using or estimating dynamic heterogeneous-firm models. Our focus on cyclical dispersion and non-convex input adjustment costs places us closest to Bloom (2009), Bachmann and Bayer (2013), Mongey and Williams (2017), and Bloom et al. (2018). The novelty of our contribution is that we combine non-convex adjustment costs with an estimated non-CES demand curve and then use this model to study firm behavior and the distinct roles of cyclical demand and productivity shock dispersion.

The remainder of the paper is structured as follows. In Section 2, we describe our data and establish stylized facts. In Section 3, we present the measurement framework and the TFPQ and demand estimates. In Section 4, we document the cyclicality of TFPQ and demand dispersion. In Section 5, we estimate passthrough and perform our variance decomposition. In Section 6, we present the structural dynamic model and draw out the aggregate implications of dispersion and uncertainty. In Section 7, we conclude.

2 Data construction and summary

Our analyses are based on firm-level data at the annual frequency for the period between 1998 and 2013. During this period, Sweden experienced two recessions. The first was a comparatively minor slowdown between 2000 and 2002. The second was a sharp and deep contraction in 2009 associated with the Global Financial Crisis.

We construct our key variables using (1) bookkeeping data from financial statements, (2)

[7] Recent examples of papers that feature non-CES demand include Félix and Maggi (2019), Berger and Vavra (2019), Lindé and Trabandt (2018), and Arkolakis et al. (2018).

price data based on goods-level production data, and (3) capacity utilization data from managerial surveys. Our accounting data come from the *Företagens Ekonomi* (FEK) survey. In principle, this survey covers the universe of Swedish industrial firms, specifically firms classified by the European Union's NACE system as manufacturing (Section C). Our product-level price and quantity data are taken from Statistics Sweden's *Industrins Varuproduktion* (IVP) survey. The raw data from the IVP survey are reported at the 8-digit product level according to the Combined Nomenclature (CN). Our sample of the IVP includes observations from about 10,000 unique firms. Our utilization data are from the *Konjunkturstatistik för Industrin* (KFI) survey. These data are at the quarterly frequency and are reported by managers at the plant level based on a stratified sample of firms with at least 10 employees.

Our main sample is based on firm-year observations for which we have complete price, capacity utilization, bookkeeping and investment data. We define firms on the basis of a continuously operating set of plants and create a new firm identifier whenever the number of plants changes. This helps to ensure a consistent definition of firm-level prices. However, our results are robust to using the firm identifiers provided in the data, see Table A4 in the Appendix. In total, our unbalanced dataset comprises 3,181 unique firms and covers the period 1998-2013. An average firm is in the sample for about 5 years, and the total number of firm-year observations is 15,044. The median firm has 107 employees and 1.82 million Swedish kroner (SEK) of revenue per employee per year. The associated interquartile range for employees is 55 to 246 employees and for sales 1.26 to 2.76 million SEK per employee. However, the distribution of firm size is skewed. While the median firm has 107 employees, the average firm has 278 employees. Descriptive statistics are presented in Table A1 in the Appendix.

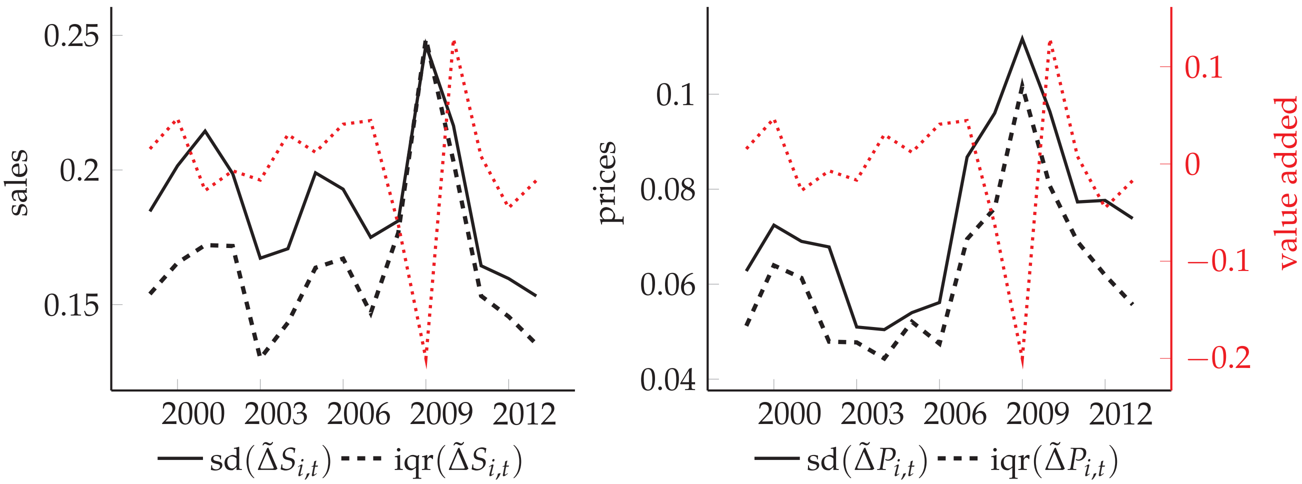
Our raw variables are sales, a firm-level price index, number of employees, intermediate goods, degree of factor utilization, and capital. In brief, sales is given by firm turnover; firm-level price is given by a price index computed based on product-level data; number of employees is measured in full-time equivalents; intermediate goods is given by the value of the stock of raw materials and consumables; factor utilization is based on managerial surveys; and capital is computed according to a perpetual inventory approach. We discuss our price and utilization data in conjunction with our discussion of TFPQ in the next section. Discussion of the data and other variables may be found in Appendix A.

The average level of within-sector dispersion is substantial. We illustrate the cyclicality of sales and price dispersion in Figure 1.[8] For both time series, we de-mean firm growth

[8]We provide an analogous figure that shows labor, intermediate goods, utilization, and investment in Fig-

(measured as log-changes) by the average growth in the relevant sector during the relevant year. We denote this transformation by $\tilde{\Delta}$.[9] We then plot the standard deviation (sd) and interquartile range (iqr) based on $\tilde{\Delta}S_{i,t}$ and $\tilde{\Delta}P_{i,t}$, where $S_{i,t}$ and $P_{i,t}$ are nominal sales and prices respectively. To emphasize cyclicality, each panel also includes the average firm-level growth of value-added. Dispersion is measured on the left axis (in black), while output growth is measured on the right axis (in red). What Figure 1 illustrates is that price and sale dispersion is counter-cyclical. Both exhibit higher dispersion in the recessionary periods.

Figure 1: Dispersion of sales and prices, 1999-2013



The left-hand plot presents sales dispersion, and the right-hand plot presents price disperison. Dispersion is measured across firms by the standard deviation (sd) and the interquartile range (iqr) of log changes computed within sector as $\text{sd}_t(\tilde{\Delta}x_{i,t})$ for sales $S_{i,t}$ and firm-level price index $P_{i,t}$. Sales $S_{i,t}$ is given by firm nominal turnover. Price $P_{i,t}$ is the firm-level price index. Each plot also includes the average growth rate of value added v deflated by a sectoral producer price index.

To quantify the degree of counter-cyclicality, we compare the level of dispersion in 2001 and 2009 to the average level of dispersion in all other years. In 2009, the iqr of $\tilde{\Delta}S_{i,t}$ is 58% greater than the average, and the iqr of $\tilde{\Delta}P_{i,t}$ is 79% higher than the average. In 2001, the same comparison shows increases of 8% for $\tilde{\Delta}S_{i,t}$ and 9% for $\tilde{\Delta}P_{i,t}$ relative to the average (for the standard deviation and other variables, see Table A3 in the Appendix). Investigating the sources and implications of this counter-cyclical dispersion is the main goal of the rest of this paper.

ure A1 in the Appendix. Most of these firm-level variables exhibit counter-cyclical dispersion. There is a dramatic rise in dispersion during the Great Recession and a smaller, though still meaningful, rise in dispersion around 2001. The only variable that deviates from this pattern is investment which is pro-cyclical.

[9] Specifically, for any (usually logged) variable $x_{i,t}$ we define $\tilde{\Delta}x_{i,t} = \Delta x_{i,t} - \text{mean}_{i \in I(j)}(\Delta x_{i,t})$, where $I(j)$ is the set of firms in industry j, and $\text{mean}_{i \in I(j)}(\Delta x_{i,t})$ gives the average value of $\Delta x_{i,t}$ in industry j at time t.

3 Production and demand framework

3.1 Productivity

We measure quantity productivity (TFPQ) at the firm level. Our theoretical starting point is a Cobb-Douglas production function based on capital $k_{i,t}$ and labor $l_{i,t}$, where i and t are firm and year subscripts respectively. We assume that factor elasticities depend on sector j, and that firm-level factor utilization $u_{i,t}$ can vary over time. Taking logarithms, our main specification has the following form:

$$\log v_{i,t} = z_{i,t} + \gamma_{K,j}(\log u_{i,t} + \log k_{i,t}) + \gamma_{L,j}(\log u_{i,t} + \log l_{i,t}), \quad (1)$$

where $\gamma_{K,j}$ and $\gamma_{L,j}$ denote sector specific factor elasticities of capital and labor. The left-hand side is a measure of physical production, real value added ($v_{i,t}$), defined below, while the right-hand side is usage of capital and labour ($u_{i,t}k_{i,t}$ and $u_{i,t}l_{i,t}$). In this specification, $z_{i,t}$ is the measure of physical productivity (TFPQ).

The interpretation of the production function is most clear for single product firms. However, most firms in our sample produce two or more products. Our general approach is therefore to deflate firm-level nominal value added $V_{i,t}$ by a firm price index $P_{i,t}$. This approach is similar to that suggested by Smeets and Warzynski (2013). Because of the possible issues associated with multiple product firms, however, we produce results for both our full sample and the sub-sample of single product firms.[10] The importance of the single product firm sample is coherence with our theoretical model; the importance of the full sample is representativeness.[11]

We compute value added $V_{i,t}$ based on the economic definition: Turnover ($S_{i,t}$) plus the change in the stock of partially finished goods ($D_{i,t}$), minus the use of raw materials and consumables ($M_{i,t}$): $V_{i,t} = S_{i,t} + D_{i,t} - M_{i,t}$. $P_{i,t}$ is constructed as a chained Laspeyres index using product-level price and sales data.[12] These price data come from a representative sample of industrial firms that are surveyed by Statistics Sweden to construct the producer price index (PPI). Real value added is then computed as $v_{i,t} \equiv V_{i,t}/P_{i,t}$. We maintain the convention that

[10] An alternative approach would be to estimate multi-product production functions as done in De Loecker et al. (2016) or using the methodology proposed by Dhyne et al. (2021).

[11] In preliminary analyses, we find that the average number of products per firm increased sharply during 2001. But we do not find evidence of cyclicality. Before and after 2001, the growth of number of products per firm is close to zero.

[12] Whether we use Laspeyres, Paasche, or Tornquist links does change our results. Additional information about our product-level price data and the construction of a firm-level price index can be found in Appendix A.1. We do not directly compare prices across firms, and instead all analysis is conducted using within-firm changes, reducing the need to quality-adjust prices across firms.

nominal variables are expressed in uppercase and real in lowercase letters, where possible.

We adjust for the degree of factor utilization ($u_{i,t}$) using a measure taken from the KFI. In the KFI, managers are asked to assess their degree of capacity utilization relative to intended production intensity, expressed as a percentage (note it was possible to report utilization in excess of "100%", see Appendix A.1 for details). Absent this utilization adjustment, demand shocks could be mis-classified as TFPQ shocks. For instance, a firm that experiences a negative demand shock may scale down production but also reduce utilization. This will look like a negative TFPQ shock unless the degree of utilization is accounted for. We substantiate this effect using a firm-level indicator of whether a firm was suffering from "insufficient demand" reported in the KFI data. Based on a regression that includes both firm and sector-year fixed effects, we find that firms that report "insufficient demand" have 15% lower utilisation on average and 26% lower utilization in the Great Recession. Hence, demand shocks are likely to create spurious movements in TFPQ if utilization is not corrected for.

3.1.1 TFPQ estimation

We estimate the input elasticities of the production function using a cost share approach implemented at the 2-digit sector level. This coincides with how other studies that focus on dispersion have estimated productivity, including Bloom et al. (2018). The cost share approach relies on the assumption that the production function is constant returns to scale (CRS) and that factor markets are competitive for capital and labour. To check the CRS assumption, we estimate the input elasticities based on control function approaches. Our control function estimates range from slightly below to slightly above $\gamma_{K,j} + \gamma_{L,j} = 1$. This is consistent with other studies, in particular those that attempt to account for factor utilization (Basu, 1996; Cette et al., 2015; Shapiro, 1993). CRS thus seems plausible and justifies the estimation of $\gamma_{K,j}$ and $\gamma_{L,j}$ based on each factor's share in total costs. Appendix B.1 presents further TFPQ estimation details.

We measure real labour costs $c^l_{i,t}$ as total payments to labour $C^L_{i,t}$ deflated by a sectoral price index. We measure capital by a user cost approach in which the cost of capital is given by $c^k_{i,t} = (r_t + \delta_j - i_{j,t})k_{i,t}$, where r_t is the yield on a 10-year Swedish government bond plus the spread between a 10-year treasury and Aaa bond, δ_j is a sector-specific depreciation rate taken from Melander (2009), and $i_{j,t}$ is the sector-specific change in the price of capital. Total costs are $c_{i,t} = c^l_{i,t} + c^k_{i,t}$, and for each sector j we estimate factor shares using the overall industry cost shares: $\gamma_{K,j} = \left(\sum_t \sum_{J(i)=j} c^k_{i,t}\right) / \left(\sum_t \sum_{J(i)=j} c_{i,t}\right)$ and $\gamma_{L,j} = 1 - \gamma_{K,j}$. Because the computation of cost shares does not rely on the utilization data, we compute the cost shares on a sample of about 8,000 firms and 50,000 observations for which price and capital (but not necessarily utilization) data are available. We get an average cost share for labour of 0.735, with some variation across industries.

Finally, we use equation (1) to back out our productivity measure $z_{i,t}$ from real value-added using the sector-specific elasticities together with measures of labor, capital, and factor utilization.[13] The labor variable $l_{i,t}$ is taken directly from our bookkeeping data which reports employment in terms of average full-time worker equivalents. For our capital variable $k_{i,t}$, we use a perpetual inventory method that relies on investment data. The perpetual inventory value tends to be preferable to the bookkeeping value because firms have incentives to write down assets in order to generate tax benefits and to inflate measures of return on capital. We therefore replace the book value of capital with the perpetual inventory measure whenever the latter is larger than the former. Capital and investment data are deflated based on sector specific changes in the price of gross fixed capital formation.

3.2 Demand

Our starting point is the constant elasticity of substitution (CES) model of demand. However, we find that deviations from CES are empirically and economically important, an issue we return to in the next section. We therefore model demand using a flexible specification that nests the CES model as a limiting case. We adapt the demand curve proposed by Gopinath et al. (2010)—henceforth GIR—to our setting and show how it can be estimated. This model allows the elasticity to differ along the demand curve in the manner of a Kimball (1995) aggregator.

The GIR model has form $q_{i,t} = (1 - \eta \log p_{i,t})^{\frac{\theta}{\eta}}$, where $q_{i,t}$ is real sales and $p_{i,t}$ is a firm's relative price. Here, $\theta > 0$ controls the average elasticity of demand and $\eta > 0$ controls how the elasticity of demand changes with the price. The key feature of the model is that the firm faces a non-constant elasticity of demand. For a given price, the elasticity is $\tilde{\theta}(p) \equiv -\frac{\partial \log q}{\partial \log p} = \frac{\theta}{1-\eta \log p}$. For $\eta > 0$, the firm's elasticity of demand rises as it increases its price. This captures the idea that it is hard for firms to gain new customers by lowering their price and easy for them to lose existing customers by raising their price. As a consequence, firms find it less appealing to change their price in response to productivity changes. This unresponsiveness is a type of "real rigidity" (Ball and Romer, 1990; Klenow and Willis, 2016).

[13]In an earlier version of this paper, we show that our results are robust to an alternative projection-based utilisation adjustment (Carlsson et al., 2022).

Like GIR, our model of demand is parameterized by θ and η, and thus allows for non-constant elasticity of demand:

$$\log q_{i,t} = \frac{\theta}{\eta} \log\left(1 - \eta \hat{p}_{i,t}\right) + \alpha_i + \mu_{j,t} + \epsilon_{i,t}. \tag{2}$$

However, we extend the model of GIR in two ways. First, we use demand shifters to allow firms to face different levels of demand from each other, and for this demand to be subject to shocks. α_i is a firm fixed effect capturing permanent differences in the level of demand across firms, and $\mu_{j,t}$ is a sector-time fixed effect capturing common changes in demand by year within a sector. The error term $\epsilon_{i,t}$ is thus an idiosyncratic demand shifter.

Second, we impose that all firms face the same elasticity of demand on average by normalizing each demand curve relative to the firm's average observed price. In the baseline GIR model in contrast, low price firms will always face a lower demand elasticity than high price firms. We abstract from permanent differences in elasticities because we are interested in how the elasticity changes within firm in response to price changes. $\hat{p}_{i,t}$ denotes the residual of a firm's log relative price after regressing on firm and sector-time fixed effects. Thus $\hat{p}_{i,t}$ has mean zero for all firms. This demeaning has no effect on the estimates in the CES case, since it is absorbed by the firm and sector-time fixed effects in the regression. However, for the general non-linear specification, this normalization ensures that all firms face elasticity θ on average.

The elasticity of demand in specification (2) is given by $\tilde{\theta}(\hat{p}) = \frac{\theta}{1-\eta\hat{p}}$. η captures how the elasticity of demand falls as a firm raises its price above its average price. The super-elasticity measures how the elasticity itself varies with the price, and is given by $\hat{\varepsilon}(\hat{p}) \equiv \partial \log \tilde{\theta}(\hat{p}) / \partial \log p = \frac{\eta}{1-\eta\hat{p}}$ (recall that $\hat{p} = \log p - c$ for some firm- and time-specific constant c). Since a firm's normalized relative price is mean zero by construction, all firms face elasticity θ on average, and average super-elasticity η (up to a Jensen's inequality correction).

The demand shock $\epsilon_{i,t}$ is the key object of interest. $\epsilon_{i,t}$ describes the idiosyncratic level of demand for firm i at time t when holding their price constant. Hence, a positive (negative) demand shock reflects the ability to sell more (less) in a given year without a corresponding reduction (increase) in price. Demand shocks reflect changes in the size of a firm's customer base or in customers' ability or willingness to pay.

We also work with a baseline CES model. This model is an important benchmark, and we rely on it in our variance decomposition exercises. Notice that the CES demand system is nested as the special case with $\eta \to 0$, in which case (2) reduces to

$$\log q_{i,t} = -\theta \log p_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,t}. \tag{3}$$

This model imposes that the elasticity of demand is constant at θ at all times.

3.2.1 Demand Estimation

The estimation of demand curves is a classic econometric challenge because prices and quantities are jointly determined in market equilibrium. To overcome this challenge, we employ the approach of Foster et al. (2008). Their identification strategy is based on the idea that the demand curve can be traced out by shifts in the marginal cost curve. Because the marginal cost curve falls as productivity improves, exogenous developments in productivity will lead to changes in price unrelated to shifts in demand. A natural way to estimate demand is thus to use productivity innovations—i.e. TFPQ—as an instrument for changes in price that are unrelated to shifts in demand. Indeed, we find that our productivity measures are strong instruments for price.

For the case of CES demand, we estimate θ using $\log z_{i,t}$ as an instrument for price. This is exactly the same procedure as used by Foster et al. (2008). For our general demand specification, we estimate θ and η based on a second-order approximation to the demand curve. A second order approximation of (2) around $\hat{p}_{i,t} = 0$ gives

$$\log q_{i,t} \simeq -\theta \hat{p}_{i,t} - \frac{\eta\theta}{2} \hat{p}^2_{i,t} + \alpha_i + \mu_{j,t} + \epsilon^*_{i,t}. \tag{4}$$

The coefficients in the regression $\log q_{i,t} = b_1 \hat{p}_{i,t} + b_2 \hat{p}^2_{i,t} + \hat{\alpha}_i + \hat{\mu}_{j,t} + \hat{\epsilon}^*_{i,t}$ enable us to identify the coefficients of the demand curve as $\theta = -b_1$ and $\eta = \frac{2b_2}{b_1}$. This is intuitive, since we are simply using a squared (log) price term to capture the nonlinearities in the model, relative to the (log) linear CES case. We estimate (4) using demeaned $\log z_{i,t}$ and its square as instruments for the relative price and relative price squared. This is similar to how Haltiwanger et al. (2018) estimate an approximation to a HARA demand curve. For clarity, we denote estimated shocks from the second-order approximation by $\epsilon^*_{i,t}$ to differentiate them those estimated from the CES-model which we denote $\epsilon_{i,t}$.

The utilization adjustment plays a potentially important role. If TFPQ is calculated without correcting for utilization (i.e. $u_{i,t} = 1$ in all periods), this biases the demand elasticities upwards because positive demand shocks raise utilization and hence measured TFPQ. This breaks the independence assumption in the IV estimation, since the instrument (TFPQ) becomes correlated with the error term (the demand shock).

The specifications presented in this section are pooled for the whole economy. This yields a single θ estimate which can be interpreted as the average demand elasticity across all sectors. Although we focus on the pooled estimate in the main text, sectoral estimates are close to the pooled estimate in most cases (see Appendix Table A5).

3.2.2 Demand Results

Demand estimates are presented in Table 1. In the first two columns, we present CES results. In the next two columns, we show results for the non-CES specification. Columns 1 and 3 are based on the sample of single product firms, while columns 2 and 4 are based on our main sample.

Results for the CES model yield a demand elasticity of about 4 in the sample of single product firms (column 1) and about 3 when estimated in our main sample (column 2). These estimates are in line with those from Foster et al. (2016) and with estimates based on Swedish data, in particular Carlsson et al. (2021) and Heyman et al. (2013).

Table 1: Demand estimation results

	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$
$\hat{p}_{i,t}$	-4.06***	-2.99***	-4.29***	-2.94***
	(0.56)	(0.20)	(0.54)	(0.20)
$\hat{p}_{i,t}^2$			-13.76*	-6.28***
			(5.41)	(1.60)
	Implied Structural Parameters			
θ	4.06***	2.99***	4.29***	2.94***
	(0.56)	(0.20)	(0.54)	(0.20)
η			4.82**	4.27***
			(1.79)	(1.04)
sample	single	main	single	main

Table 1 gives demand estimates based on our single product firm sample ($N = 3,350$) and our main sample ($N = 15,044$). The implied structural parameters θ and η are shown in the bottom panel. $q_{i,t}$ denotes firm i's real sales in year t and $\hat{p}_{i,t}$ denotes the log of firm i's relative price in year t de-meaned at the sector-year level. All specifications include firm and sector-year fixed effects. Standard errors are clustered at the firm level and given in parentheses. Level of significance at the 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively. The first two columns give results for CES demand, model 3. Columns three and four present results for the non-linear approximation, model 4. All estimates use $z_{i,t}$ (TFPQ) as an instrumental variable. The difference between each pair of regressions is the sample. Columns 1 and 3 and based on the sample of single product firms, whiles columns 2 and 4 are based on the main sample.

Results for our general demand specification are shown in columns three and four. The second order term is statistically significant at the 5% level in both samples, and the implied structural parameter, η, is statistically significant at the 1% level in both samples. The data thus prefers a model with non-constant elasticity of demand. Nevertheless, the first order coefficient is similar despite the addition of the second order term. The linear model may therefore be appropriate for modelling small changes.

How should we interpret the magnitude of the second order coefficient? In our main sample, the implied structural parameters are $\theta = 2.94$ and $\eta = 4.27$ (column 4). For these estimates, a 5% increase in a firm's price from $\hat{p} = 0$ to $\hat{p} = 0.05$ causes its demand elasticity to increase from 2.94 to $\frac{2.94}{1-4.27\times 0.05} = 3.74$. Similarly, a 5% reduction in price causes the elasticity to fall to 2.42.[14] As we show in our theoretical work, this is an economically meaningful result that can help explain incomplete passthrough. If a firm's elasticity rises when it raises its price, the firm gains little revenue from raising its price because it loses many customers. Conversely, if a firm's elasticity falls when it lowers its price, the firm gains little revenue from lowering its price because it gains few customers. In such a world, relatively small price adjustments will be optimal even when shocks have a large impact on marginal cost.

We conduct a number of robustness exercises. These include demand estimates based on (1) a piece-wise linear specification and (2) various samples, including a balanced panel and a sample that excludes the Great Recession. Results from these exercises reinforce the conclusions from the main text, see Appendix B.2. We also provide corroboration of our shocks. Our TFPQ measure is associated with process innovations and negative demand shocks correspond to reports of insufficient demand, see Appendix B.3.

4 The dispersion of productivity and demand shocks

4.1 Measuring cyclical dispersion

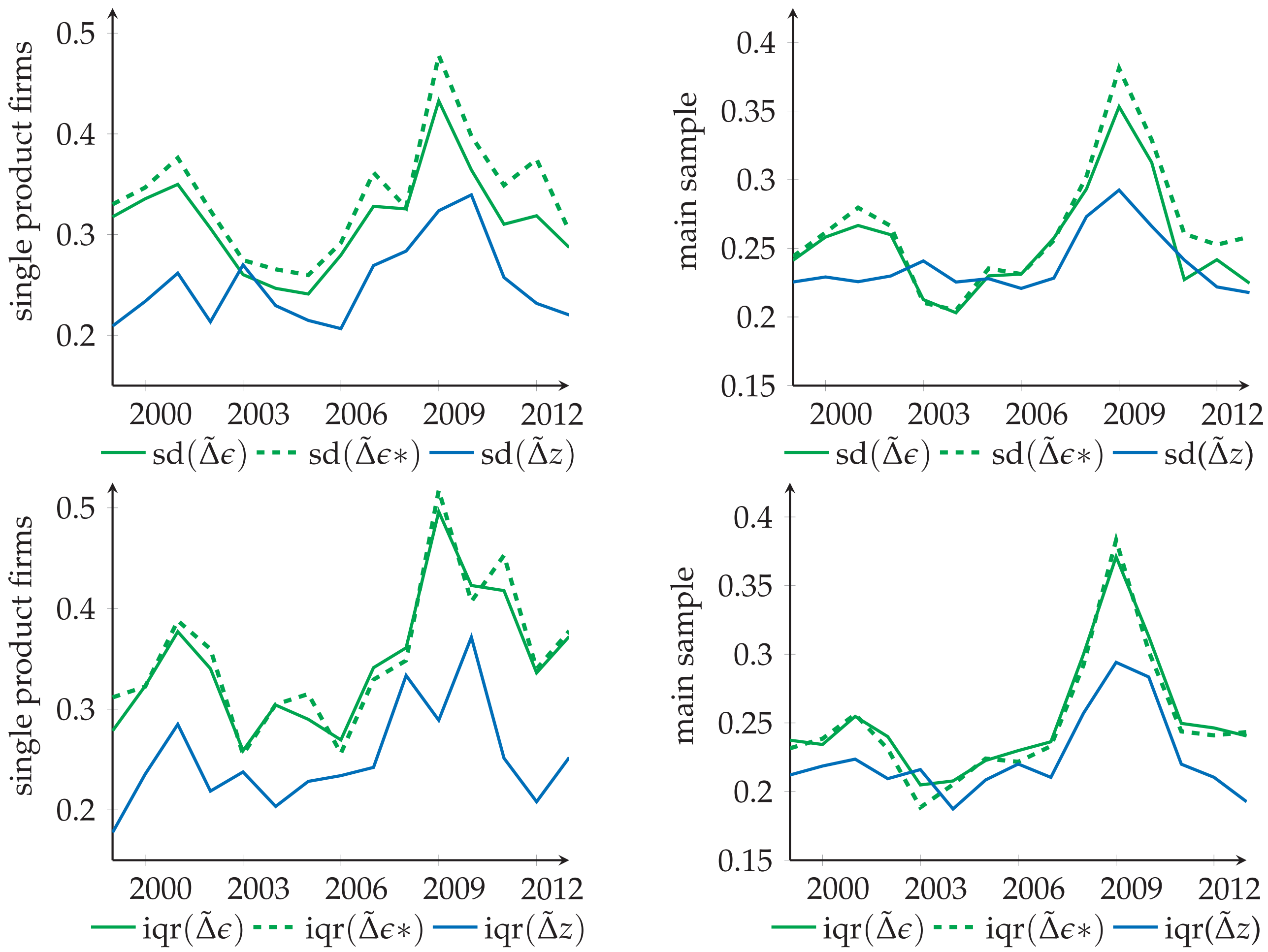
We measure shocks to TFPQ and demand as log changes relative to the previous year. To construct our dispersion measures, we first demean by the relevant average sector-year growth using our transformation $\tilde{\Delta}$. We then compute the standard deviation or interquartile range based on the de-meaned data. Dispersion in thus measured in terms of growth rates and reflects genuine heterogeneity across firms divorced from aggregate or sectoral volatility.

Both shocks exhibit countercyclical increases in dispersion. We illustrate this pattern in

[14] Comparing the deviation from CES estimated across papers is somewhat difficult due to the different functional forms used. If available, the super-elasticity serves as a useful metric. Relative to existing work, our estimated super-elasticity of $\eta = 4.27$ is around half of the value of 10 studied (but not estimated) in Klenow and Willis (2016), and larger than the value around 2 used in Berger and Vavra (2019).

Figure 2. This figure plots the within-sector dispersion of each shock over time. The left panels show results for single product firms. The right panels shows results for our main sample. The pattern of countercyclical volatility is present for both samples. We focus on results from the main sample in the remainder of the text, and report results from the single product firm sample in the appendix.

Figure 2: Dispersion of demand and TFPQ (1999-2013)



The left panels show times series for the across-firm dispersion of demand shocks ($\tilde{\Delta}\epsilon$ and $\tilde{\Delta}\epsilon^*$) and TFPQ shocks ($\tilde{\Delta}z$) for the sample of single product firms. The right panels present the same measures based on our main sample. The top row shows the standard deviation (sd), while the bottom row shows the interquartile range (iqr). For both demand and TFPQ, shocks are computed as log changes. The standard deviations are then computed across firms within sector as $\mathrm{sd}_t(\tilde{\Delta}x_{i,t})$ for each variable x as explained in the text.

4.2 Average level of productivity and demand shock dispersion

Both TFPQ and demand exhibit substantial firm-level volatility. Excluding the recession years 2001 and 2009, the standard deviation of TFPQ growth is 0.235 and the standard deviation of CES demand growth is 0.247. To put this in perspective, sectoral dispersion is an order of magnitude smaller: The standard deviation computed across sector-year averages is 0.071 for TFPQ growth and 0.055 for demand growth. Firm-level dispersion is, moreover, not driven by outliers. If we instead measure dispersion based on interquartile range, the results are similar: An IQR of 0.217 for TFPQ and 0.238 for demand.

Demand volatility is at least as large as that seen for productivity. This holds for both demand measures, ϵ and ϵ^*, and regardless of whether dispersion is measured using the standard deviation or the interquartile range. This highlights the potential importance of demand as an additional source of shocks that can drive dispersion in outcomes across firms.

4.3 Cyclicality of productivity and demand shock dispersion

Productivity and demand are characterized by counter-cyclical volatility. We quantify this pattern in Table 2 for our main sample. Each entry in this table presents a percentage change in dispersion for a given year relative to the average computed across all other years, excluding the 2001 and 2009 recessions. The left side presents these changes for TFPQ and demand shocks, while the right side presents changes for relative prices and real sales. For each variable, we present the change measured by the standard deviation (sd) and interquartile range (iqr). As is evident in this table, there was a dramatic increase in dispersion during the Great Recession. This holds for both shocks, sales, and prices. Table 2 also reveals an increase in dispersion around 2001, though the increase is smaller in comparison to the Great Recession. The 2001 recession is also somewhat ambiguous. The finding of increased dispersion is robust for demand and sales, but depends on the choice of dispersion measure for prices and TFPQ: For prices and TFPQ, we only measure countercyclicality during 2001 if we use the interquartile range as our measure of disperison.

Table 2 shows that demand dispersion is more cyclical than productivity dispersion. For example, during the Great Recession, the IQR of $\tilde{\Delta}\epsilon$ was 56% above average, whereas the IQR of $\tilde{\Delta}z$ was only 35.5% above average. The finding that demand dispersion is countercyclical, even more so than TFPQ, is new and potentially important for understanding the deeper sources and propagation of recessions.

Table 2: Cyclicality of dispersion, main sample

	shocks						outcomes			
	$\tilde{\Delta}z$		$\tilde{\Delta}\epsilon$		$\tilde{\Delta}\epsilon^*$		$\tilde{\Delta}s$		$\tilde{\Delta}p$	
	sd	iqr	sd	iqr	sd	iqr	sd	iqr	sd	iqr
2001	-3.9	3.0	8.0	7.2	9.5	9.8	17.0	9.2	-4.0	5.1
2009	24.6	35.5	43.1	56.2	49.3	64.1	34.5	57.8	51.5	83.2

This table presents percentage changes in dispersion measures for the 2001 and 2009 recessions relative to the average over all other years. The left side of the table (columns 1-6) shows the results for shocks (z, ε, and ε^*) while the right side of the table (columns 7-10) show changes in relative prices and real sales (p and s). For each variable, we show the change in the standard deviation (sd) and interquartile range (iqr) for 2001 and 2009. All measures have been de-meaned by sector-year. Additional statistics, including measures of skewness and kurtosis, can be found in the appendix.

5 The role of shocks for endogenous outcomes

To what extent do TFPQ and demand shocks drive sales and price dispersion over the business cycle? To evaluate this question empirically, we conduct two exercises. In the first exercise, we establish how shocks transmit to prices. We estimate a log-linear passthrough equation that describes how firms adjust their prices in response to TFPQ and demand innovations. Overall, we find a limited response of prices to TFPQ shocks, and a non-zero effect from demand shocks. In the second exercise, we describe how cyclical dispersion in sales and prices can be statistically attributed to dispersion in TFPQ and demand. We estimate a "semi-structural" variance decomposition that relies on our demand and passthrough results. The conclusion from this exercise is that demand plays a significant role in driving countercyclical volatility, while the role of TFPQ is limited.

5.1 Passthrough specification

The passthrough equation that we estimate is similar to that in De Loecker et al. (2016), but includes demand shocks as an additional driver of price setting:

$$\log p_{i,t} = \beta_z z_{i,t} + \beta_\epsilon \epsilon_{i,t} + \alpha_i + \mu_{j,t} + \tau_{i,t}. \tag{5}$$

Equation (5) specifies how firms set their prices in response to the shocks that they face. It can be interpreted as an estimated policy function. The parameters of interest are the "passthrough coefficients" β_z and β_ϵ. β_z measures the responsiveness of a firm's price to their level of TFPQ, and β_ϵ measures the responsiveness of their price to their demand shock.

We focus on within-firm variation and include firm and sector-time fixed effects α_i and $\mu_{j,t}$. The passthrough coefficients thus measure the average responsiveness of a firm's price to idiosyncratic TFPQ and demand shocks. As discussed in De Loecker et al. (2016), if $z_{i,t}$ and $\epsilon_{i,t}$ are exogenous shocks, (5) can be estimated by OLS because there is no endogeneity problem.

The passthrough equation is consistent with a benchmark static pricing model with CES demand in which prices are set as a constant markup over marginal cost: $p_{i,t} = (\theta/(\theta - 1))c_{j,t}/z_{i,t}$ for some (possibly sector specific) weighted input price $c_{j,t}$. In logs this gives $\log p_{i,t} = -\log z_{i,t} + \log c_{j,t} + \log(\theta/(\theta-1))$. Notice that in the benchmark model, TFPQ shocks directly affect prices via the impact on marginal cost—given a constant markup, passthrough is "complete"—while demand shocks only affect the quantity of sales, not the price. The benchmark model thus implies $\beta_z = -1$ and $\beta_\epsilon = 0$.

The error term $\tau_{i,t}$ is a "price wedge." This wedge captures changes in the prices that cannot be explained by the shocks. The price wedge is important as it provides a basis to evaluate how well the shocks account for firm behavior and the possible role of cyclical distortions.[15]

5.2 Passthrough results

Table 3 presents passthrough results based on our main sample. The first column shows estimates from an OLS fixed-effects estimation of equation (5). The second column shows results for the same model when using lagged values of TFPQ and demand as instruments. This is the favored approach of De Loecker et al. (2016) because instruments of this type can help correct for measurement error by relying on persistent changes. The third column presents equation (5) based on one year differences rather than levels, again using OLS. Additional passthrough exercises are commented on in Appendix C.1.

TFPQ passthrough is low across all specifications. Passthrough is -0.124 in the OLS specification and only -0.097 in the first difference specification (columns 1 and 3). This means that a firm only lowers its price by about 1% in response to a 10% reduction in costs. We find a higher passthrough when using the IV approach (column 2), but still much lower than complete passthrough. Since the IV approach focuses on persistent changes in TFPQ,

[15] For example, if firms face financial frictions which bind more at poorer firms, a recession could lead to larger declines in activity at poor firms than large firms in response to a negative aggregate *level* shock which is common to all firms . This would manifest as increased dispersion in sales across firms even if the dispersion of firm level shocks to TFPQ and demand had not increased. One interpretation of the price wedge is as something that shifts marginal cost, by replacing marginal cost with $mc_{i,t} = \tau_{i,t}c_{j,t}/z^u_{i,t}$. Through the lens of the constant markup model, this implies the wedge represents unmodelled changes in marginal cost, but one could equally think of the wedge as representing markup changes even if true marginal cost has not changed.

it is possible that the differences arise because firms are hesitant to change their prices in response to transitory (or perhaps mismeasured) changes in TFPQ.

With respect to demand, we find evidence of moderate passthrough. Our demand passthrough estimates range between 0.209 and 0.235 across specifications (columns 1 to 3). A demand shock of 10% thus leads to about 2% higher prices. In other words, firms increase their prices in response to an increased ability to sell at a given price.

Both the TFPQ and demand results are inconsistent with the simple static pricing model. Incomplete TFPQ passthrough ($\beta_z > -1$) suggests that firms allow their markup to rise rather than adjusting their price in proportion to a reduction in costs, while positive demand passthrough ($\beta_\epsilon > 0$) means that firms choose to raise their price and markup rather than simply selling more following a demand shock. These findings are important for understanding how demand and productivity shocks translate into firm behaviour.[16] A contribution of the theoretical model presented in the next section is an ability to rationalize these findings.

What can explain our passthrough results? Sticky prices probably play a role. Earlier work has documented price stickiness in our data—see for example Carlsson and Skans (2012); Carlsson (2017)—and sticky prices will lead to low TFPQ passthrough. Nevertheless, real rigidities appear to be important over-and-above sticky prices. To show this, we re-estimate passthrough based on the sample of firm-year observations that display flexible prices. We include in this sample only those observations for which all product level prices have been adjusted during a given year. Results are shown in column 4: TFPQ passthrough of about -0.1 and demand passthrough a bit larger than 0.2. It thus appears that some firms are reluctant to make large price adjustments because the environment limits the benefit of such adjustments, even in response to large TFPQ shocks. This exercise rules out Calvo or menu-cost style price stickiness as the sole driver of low passthrough from TFPQ to prices. To rule out Rotemberg price frictions—which lead to gradual price changes—we also ran the first-difference passthrough regression at longer horizons (two and three year differences) and found similar passthrough from TFPQ to prices.

[16] Our results are mostly consistent with the literature. Our TFPQ passthrough estimates are similar to those in De Loecker et al. (2016) and Pozzi and Schivardi (2016), although smaller than the average industry in Haltiwanger et al. (2018). Our finding of positive passthrough from demand shocks to prices is also comparable with estimates in Pozzi and Schivardi (2016) and Haltiwanger et al. (2018)—though we find higher demand passthrough overall.

Table 3: Passthrough estimates, main sample

	$\ln p_{i,t}$	$\ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$
$z_{i,t}$	-0.124***	-0.240***		
	(0.006)	(0.024)		
$\epsilon_{i,t}$	0.227***	0.235***		
	(0.005)	(0.009)		
$\Delta z^u_{i,t}$			-0.0965***	-0.106***
			(0.004)	(0.005)
$\Delta\epsilon_{i,t}$			0.209***	0.229***
			(0.005)	(0.005)
N	15042	10132	11108	5882
iv	no	L.z L.ϵ	no	no
sample	all	all	all	flexible prices

This tables presents passthrough results based on the main sample. $p_{i,t}$ denotes firm i's relative price in year t, while $z_{i,t}$ and $\epsilon_{i,t}$ denote firms i's TFPQ and demand in year t. First differences are indicated by Δ. All specifications include sector-year fixed effects. Specifications in levels include firm fixed effects. Standard errors are clustered at the firm level and given in parentheses. Three stars (***) indicates significance at the 0.1% level. The first and second columns show results for the estimation of the passthrough equation 5 in levels. The first column shows results based on OLS estimation while the second column shows results when using the lags of tfp and demand as instruments (L.z and L.ϵ). The third column presents model 5 estimated in first differences. Column 4 is the same model as column 3, but based solely on firms that exhibit flexible prices at the product level. The bottom panel presents information on the number of observations (N), the use of instrumental variables (iv), and the sample.

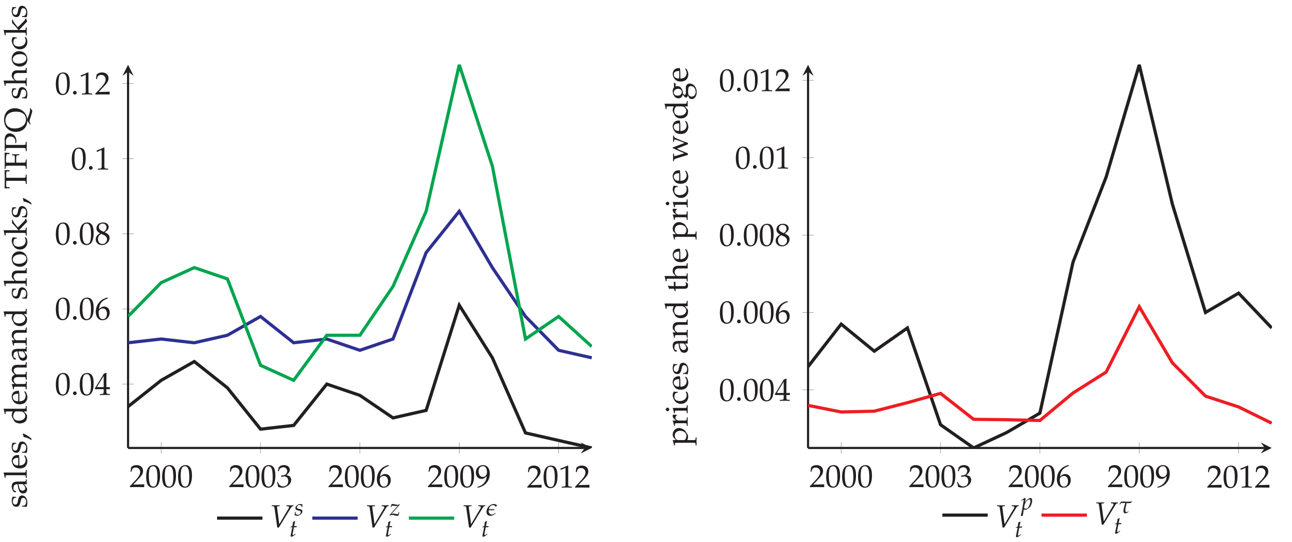
5.3 Variance decomposition of sales and price growth

To evaluate the economic importance of shocks for dispersion from an empirical perspective, we compute variance decompositions of prices and sales. These decompositions enable us to attribute the countercyclical dispersion of price and sales growth to the variation in underlying TFPQ and demand shocks, as mediated by the parameters of our estimated passthrough and demand relationships.[17]

Figure 3 illustrates the underlying dispersion of shocks that form part of our variance decomposition. We use the notation $V_t^x \equiv V_t(\tilde{\Delta} x_{i,t})$ to denote the variance of growth rates for variable x after removing sector-time variation. The left panel shows the variances of the TFPQ and demand shocks along with the variance of sales. Both demand and TFPQ disper-

[17] Eslava and Haltiwanger (2020) also perform variance decompositions across firms in a dataset featuring price information, and where, like us, they can estimate demand and TFPQ shocks. Our work is different from theirs in two main ways. We focus on first-differenced changes year to year and perform our variance decomposition over the business cycle, while they perform their decomposition over the firm lifecycle (i.e. by firm age). In addition, they measure their price wedge in a structural framework as the wedge relative to the statically optimal price, whereas we use the reduced-form coefficients β_z and β_ϵ to capture the deviation of passthrough from the CES benchmark. They additionally leverage firm-specific input price data, and explicitly model a multi-product firm.

Figure 3: Variances of sales, price, and underlying shocks



The left panel shows the variance over time of TFPQ shocks and demand shocks (V_t^z in blue and V_t^ϵ in green) alongside the variance of sales growth (V_t^s in black). The right panel shows the variance over time of price growth V_t^p (in black) and the variance of changes in the price wedge estimated from the passthrough equation V_t^τ (in red). For all variables, sector-year growth is removed before computing the variance. In the variance decomposition exercise, V_t^s and V_t^p are attributed to the V_t^z, V^ϵ and V_t^τ. The variances are presented in two panels because of differences in scale.

sion increase during the Great Recession (in green and blue, respectively), though demand increases by more. During 2001, only demand shocks show a clear relationship with sales dispersion (in black). The right panel shows the variance of firm prices p together with the variance of the price wedge τ (in black and red, respectively). Here we see a sharp spike in the dispersion of the price wedge in 2009 and little cyclicality otherwise.

5.3.1 Variance decomposition specification

The variance decomposition of prices is based on the log-linearity of the passthrough equation. Taking first differences of (5), subtracting sector-year means from both sides, and taking the variance across all firms yields an expression of the form:

$$V_t^p = V_t^{p,z} + V_t^{p,\epsilon} + V_t^{p,resid}. \tag{6}$$

This equation shows how the time-varying variance of price growth can be attributed to the variance of the shocks using the estimated parameters from the passthrough equation. There are three components: (1) $V_t^{p,z} \equiv \beta_z^2 V_t^z$ measures the contribution of TFPQ dispersion, (2) $V_t^{p,\epsilon} \equiv \beta_\epsilon^2 V_t^\epsilon$ measures the contribution of demand dispersion, and (3) $V_t^{p,resid}$ measures the residual variance. $V_t^{p,resid}$ contains the covariances between shocks as well as the variance of the price wedge: $V_t^{p,resid} \equiv V_t(\tilde{\Delta}\tau_{i,t}) + \beta_z\beta_\epsilon \text{cov}_t(\tilde{\Delta}z_{i,t}, \tilde{\Delta}\epsilon_{i,t}) + \beta_z \text{cov}_t(\tilde{\Delta}z_{i,t}, \tilde{\Delta}\tau_{i,t}) + \beta_\epsilon \text{cov}_t(\tilde{\Delta}\epsilon_{i,t}, \tilde{\Delta}\tau_{i,t})$.

The variance decomposition for sales has a similar structure. To get an equation that relates sales to a firm's shocks, we combine the passthrough equation with the log-linear approximation of demand given by the CES model. For our sales measure we use firm sales deflated by its sectoral price index: $s_{i,t} \equiv S_{i,t}/P^s_{i,t}$. This sales measure is similar to nominal sales since it is not deflated by the firm's own price. This sales measure can be exactly linked to a firm's shocks via the following procedure: Add $\log p_{i,t}$ to both sides of (3) to yield $\log s_{i,t} = (1-\theta)\log p_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,t}$. This relates a firm's sales to the price it chooses to set and its demand shock. Next, replace $\log p_{i,t}$ using the passthrough equation (5) to give

$$\log s_{i,t} = (1-\theta)\beta_z z^u_{i,t} + ((1-\theta)\beta_\epsilon + 1)\,\epsilon_{i,t} + (1-\theta)\tau_{i,t}. \tag{7}$$

We omit the firm and sector-year fixed effects as these drop out in the following steps. As in the price decomposition, we take the first difference of this equation over time, subtract the sector-year mean, and then take the variance across firms. Doing so yields a sales decomposition equation with structure:

$$V_t^s = V_t^{s,z} + V_t^{s,\epsilon} + V_t^{s,resid}. \tag{8}$$

There are three components of the decomposition: (1) $V_t^{s,z} \equiv (1-\theta)^2\beta_z^2 V_t^z$ measures the contribution of TFPQ dispersion, (2) $V_t^{s,\epsilon} = ((1-\theta)\beta_\epsilon + 1)^2 V_t^\epsilon$ measures the contribution of demand dispersion, and (3) $V_t^{s,resid}$ measures the residual variance. $V_t^{s,resid}$ is a residual variance term that contains the covariance terms and variance of the price wedge.[18]

5.3.2 Variance decomposition results

Variance decomposition results are shown in Figure 4. The left side shows the decomposition of sales, V_t^s, and the right side shows the decomposition of prices, V_t^p. Each plot includes the contribution from TFPQ (in blue), demand (in green), and the residual components. The residual components include one term that represents the contribution from the variance of the price wedge (in red) and three covariance terms between shocks and the price wedge. Here, we use the passthrough estimates β_z and β_ϵ from the first differences specification (column 3 in Table 3). We focus on the first difference estimates because the variance decom-

[18]Specifically, $V_t^{s,resid} \equiv (1-\theta)^2 V_t(\tilde{\Delta}\tau_{i,t}) + \beta_z(1-\theta)\,((1-\theta)\beta_\epsilon + 1)\,\text{cov}_t(\tilde{\Delta}z_{i,t}, \tilde{\Delta}\epsilon_{i,t}) + \beta_z(1-\theta)^2 \text{cov}_t(\tilde{\Delta}z_{i,t}, \tilde{\Delta}\tau_{i,t}) + (1-\theta)\,((1-\theta)\beta_\epsilon + 1)\,\text{cov}_t(\tilde{\Delta}\epsilon_{i,t}, \tilde{\Delta}\tau_{i,t})$.

position is conducted in first differences. However, results are similar if we instead use the IV estimates of passthrough (See appendix C.2 for additional discussion and quantification of the variance decomposition).

Figure 4: Variance Decompositions of Sales and Prices

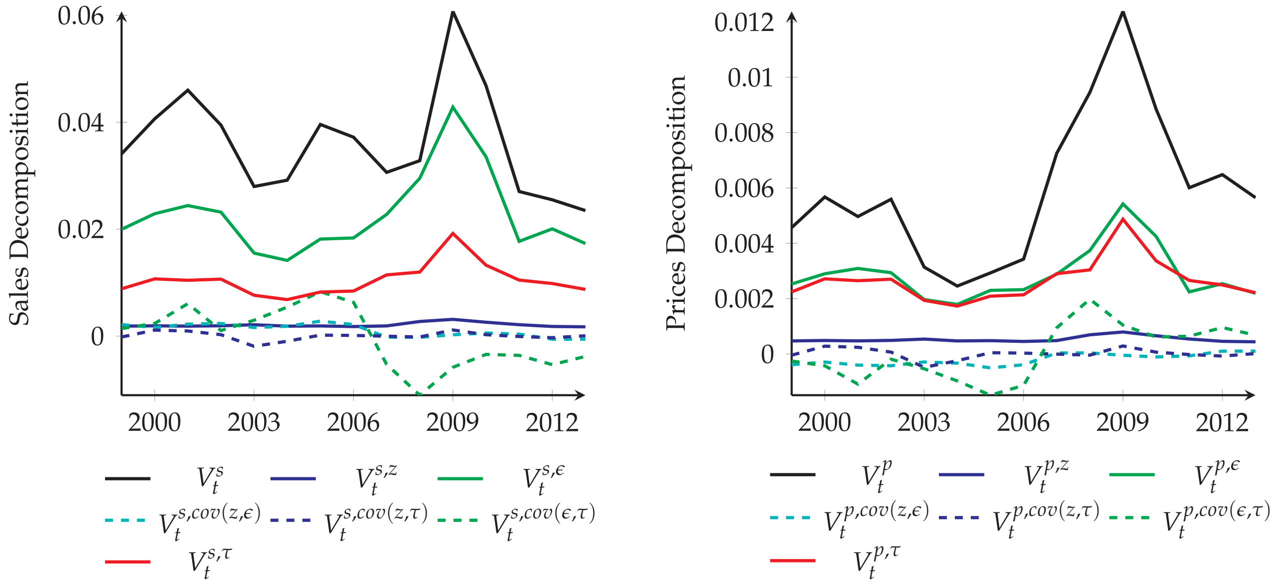


Figure 4 presents variance decompositions. The left panels present sales decompositions and the right panels price decompositions. The top row presents variance decompositions for which the passthrough equation coefficients is estimated in first differences. The bottom row presents variance decompositions based on the passthrough coefficients estimated in levels using the IV approach. The $V_t^{s,x}$ terms denote the portion of sales variance attributable to a variable or covariance between variables. The $V^{p,x}$ denote the same for the price dispersion.

As evident in the left panel, demand is the most important component of the sales decomposition. On average, demand accounts for 63% sales growth dispersion. Demand also drives the cyclicality of sales dispersion. The 82% increase in sales dispersion between 2008 and 2009 is mostly accounted for by demand. Comparing 2009 to non-recession years, we find increased demand dispersion explains about 80% of the increase in sales dispersion (See Table A6 for these calculations). Likewise, demand dispersion rises during 2000-2002 and tracks changes in sales dispersion during that period. In 2001, 27% of the rise in sales dispersion relative to non-recession years is attributable to increased demand dispersion. TFPQ dispersion in contrast makes only a marginal contribution to the dispersion of sales growth.

The picture is similar for price dispersion. As seen in the right panel, demand plays a more important role than TFPQ. Demand accounts for about 50% of price dispersion on average and 40% of the increase in price dispersion in 2009 relative to non-recession years. TFPQ explains about 10% of price dispersion on average and almost none of its movements over the cycle.

What explains these results? Why is demand the dominant driver of sales dispersion over the cycle, while TFPQ is largely irrelevant? The limited contribution of TFPQ to dispersion follows from the low TFPQ passthrough that we estimate. The contribution of TFPQ dispersion to sales dispersion is given by $V_t^{s,z} \equiv (1-\theta)^2 \beta_z^2 V_t^z$. Given our passthrough and demand curve coefficients ($\beta_z = -0.0965$ and $\theta = 2.99$), the variance of TFPQ is shrunk by a factor of about twenty five ($V_t^{s,z} \simeq 0.04 \times V_t^z$). Thus, the substantial levels of TFPQ dispersion that we observe are not transmitted into sales dispersion due to the structure of the relationships between TFPQ, prices, and sales. The $\beta_z^2 \simeq 0.01$ term in particular limits passthrough. The economic intuition for this result is that TFPQ does not affect sales directly, but only indirectly via the price. TFPQ shocks create price movements according to (5) which then affect sales because of movement along the demand curve (3). The low passthrough we observe from TFPQ to prices means that firms do not change their prices fully in response to TFPQ changes. This weakens the ability of TFPQ dispersion to affect sales growth dispersion, and as a consequence the inability of higher TFPQ dispersion in the Great Recession to explain the increase in sales dispersion.

Demand dispersion is, in contrast, the most important driver of sales growth dispersion because demand shocks have a direct impact on sales through the demand curve: In the absence of price adjustments, an increase in demand dispersion is transmitted one-for-one into an increase in sales dispersion. Although the effect on sales is dampened because firms raise their prices in response to demand shocks, the estimated demand passthrough still leaves a sizeable effect of demand shocks on sales. The combination of substantial demand dispersion on average and a large increase during the Great Recession explains both the average and cyclical importance of demand shocks for sales dispersion. In terms of our estimated parameters, the contribution of demand dispersion to sales dispersion is $V_t^{s,\epsilon} = ((1-\theta)\beta_\epsilon + 1)^2 V_t^\epsilon$. For the level of passthrough that we estimate ($\beta_\epsilon = 0.209$), $V_t^{s,\epsilon} \simeq 0.4 \times V_t^\epsilon$. Recalling that demand and TFPQ shocks have a similar variance on average, the relative importance of demand shocks reflects the ten-fold larger multiplier.

In the price decomposition, demand shocks are more important than TFPQ shocks for similar reasons. Passthrough from demand to prices is greater than passthrough from TFPQ to prices. In addition, demand dispersion increases by more than TFPQ dispersion during recessions.

5.3.3 Extension: Time-varying passthrough

Demand dispersion is more important than TFPQ dispersion in explaining the variance of firm level endogenous outcomes both on average and over the cycle. However, there remains significant unexplained dispersion, which we attribute to the price wedge.[19] For both sales and prices, the variance of the price wedge has significant explanatory power both on average and for cyclicality. For example, the price wedge contributes meaningfully to the rise in sales growth dispersion during the Great Recession, with an additional contribution from a rise in the contribution from the correlation between demand shocks and the price wedge. One set of explanations for the role of the price wedge relates to other, un-modelled, shocks or frictions. Another explanation is time-varying passthrough. We use passthrough coefficients for demand and TFPQ that are constant over time in our variance decompositions. But cyclical changes in passthrough are a real and important possibility. Recent work has emphasized the importance of changes in responsiveness to shocks (as opposed to changes in the dispersion of shocks) for driving countercyclical dispersion in endogenous variables (Berger and Vavra, 2019).

To investigate the possibility of time-varying passthrough, we estimate passthrough on a year by year basis. These estimates are presented in Figure A2 in the Appendix. We find evidence that passthrough varies systematically over the business cycle. TFPQ passthrough tends to increase during recessions (i.e. become more negative) while demand passthrough tends to fall. The decline in demand passthrough during recessions—when demand dispersion is high—is consistent with our firm-level evidence that passthrough is smaller in response to large idiosyncratic shocks.[20] Time-varying passthrough also explains why there is a large contribution from the correlation between demand shocks and the price wedge in 2009 in our variance decompositions (see Figure 4).

Taking the possibility of time-varying passthrough seriously, we re-compute our variance decompositions using using passthrough coefficients estimated period by period (see Figure A2 in the Appendix). We find that our main results are robust to this approach. In fact, the contribution of demand shocks to sales dispersion in the Great Recession is now even larger because firms change their prices relatively less in response to demand shocks during 2009.

[19]In related work, Eslava and Haltiwanger (2020) find that wedges play an important role in the firm lifecycle. They find that the *level* of their estimated wedge is correlated with the *level* of firms' demand and TFPQ shocks, and that over the lifecycle the fact that the growth in wedges is correlated with growth in demand and TFPQ reduces the variance in lifecycle sales growth by around 12%.

[20]We also find systematically higher demand passthrough overall in the latter half of our sample as compared with the first half.

6 Quantitative Model

In this section, we build a quantitative heterogeneous firm model informed by our empirical results. Our goals are twofold: The first is to investigate what features are needed to match our empirical findings regarding dispersion and passthrough. Our second purpose is to evaluate the economic importance of distinguishing between demand and productivity shocks when treating dispersion as inducing uncertainty. We show that it is crucial to distinguish between the two shocks because demand dispersion generates wait and see effects while productivity dispersion generates markup-induced misallocation.

6.1 Environment

The model is a continuous-time extension of Bloom (2009) and Bloom et al. (2018) that includes both demand and TFPQ shocks. The key extension we consider is a richer specification of demand, based on the finding that demand curves have non-constant elasticity. This invalidates the usual result—under CES—that TFPQ and demand shocks are isomorphic and can be studied as a single shock to TFPR. This extension also allows the model to generate passthrough from TFPQ shocks to prices in line with our empirical results.

In order to focus on our new features, we simplify our model in two ways relative to the existing literature. First, we consider only partial equilibrium results. This makes the model tractable even though the firm's problem has an additional state variable coming from the distinction between TFPQ and demand shocks. This is, moreover, a reasonable simplification given that general equilibrium effects are likely muted in the short-run, as discussed in Bloom et al. (2018). Second, we simplify the adjustment cost structure to reduce the dimensionality of the firm's problem.

Time is continuous and indexed by t. There is a unit mass of firms indexed by $i = [0, 1]$ who discount the future at rate r, and there is no entry or exit of firms. Aggregate prices are constant and taken as given by firms, with w denoting the real wage and P the aggregate price level. An aggregate state $s = \{1, 2\}$ denotes the level of uncertainty, which is common across firms, with $s = 1$ denoting low uncertainty and $s = 2$ denoting high uncertainty. Uncertainty switches to the other state according to a Poisson process with rate $\lambda^s(s)$.

6.2 Production and adjustment costs

Each firm produces output, q, from a Cobb-Douglas production function $q = zk^{\alpha}l^{1-\alpha}$, where l and k are labour and capital, and z is idiosyncratic physical total factor productivity (TFPQ). We suppress firm subscripts for readability. Let p denote the firm's price relative to the aggregate price level P. Firms face a common demand curve $q = d(p, \varepsilon)$, where ε is an idiosyncratic demand shifter. We assume we can invert the demand curve to get $p = p(q, \varepsilon)$. In our quantitative implementation, we use a demand curve consistent with our empirical work:

$$\log q = \frac{\theta}{\eta}\log(1 - \eta \log p) + \varepsilon. \tag{9}$$

The model is calibrated to have $\mathrm{E}[\log p] = 0$. θ is therefore the average demand elasticity. In the limit of CES demand ($\eta \to 0$), this reduces to $\log q = -\theta \log p + \varepsilon$.

Capital takes time to adjust. It depreciates at rate δ and is increased by investment i giving $\dot{k} = i - \delta k$. Labour is also potentially subject to hiring costs, and so we track the stock of labour at the firm. The hiring rate is denoted h, and the labour stock depreciates at rate δ, which is assumed to be the same as capital depreciation for simplicity.[21] The stock of workers thus evolves according to $\dot{l} = h - \delta l$. Capital can be bought and sold at price p_k while labour can be hired at cost a (which is recouped when workers are fired).

We seek a formulation of the firm's problem where we can represent non-convex adjustment costs to both capital and labour but only carry a single state variable x.[22] We therefore define the "overall scale" of a firm as $x \equiv k^{\alpha}l^{1-\alpha}$, i.e. total amount of inputs weighed by their elasticities. As a consequence, output can be described by the linear function $q = zx$. All non-convex adjustment costs (such as resale loss from capital, firing costs, fixed costs) are then placed on the adjustment of the overall scale of the firm, x, rather than the individual factors. We define an investment rate for x as

$$\dot{x} = i_x - \delta x, \tag{10}$$

where i_x is investment in overall scale and $\dot{x}$ denotes the desired rate of change in overall scale. In the appendix, we derive how this is split into investment in each factor. The cost (in

[21] This choice is without loss of generality in our baseline calibration since we assume no hiring costs directly on the labour stock, making the choice of depreciation rate irrelevant.

[22] Having adjustment costs on both capital and labor is important for generating large wait and see effects. Otherwise, the firm is able to accommodate shocks reasonably well by simply adjusting the other factor. Given the Swedish labor market structure, adjustment costs on both factors appear reasonable, but we also consider robustness to only placing adjustment costs on capital.

addition to p_k and a) of investment in scale at rate i_x is assumed to be

$$c(i_x, x) = \begin{cases} \frac{\kappa}{2}\frac{(i_x - \delta x)^2}{x} & i_x > \delta x \\ 0 & \delta x \geq i_x \geq 0 \\ -\underline{\kappa} i_x + \frac{\kappa}{2}\frac{i_x^2}{x} & i_x < 0. \end{cases} \tag{11}$$

Here κ controls quadratic adjustment costs, which are paid for investment rates above the rate of depreciation, or for disinvestment.[23] $\underline{\kappa}$ is a partial irreversibility cost, meaning that the resale price of inputs is $\underline{\kappa}$ less than the purchase price. Total static cashflow is $cf = pq - wl - p_k i - ah - c(i_x, x)$.

Both TFPQ and demand follow Markov processes with stochastic volatility. The firm draws a new level of TFPQ at rate λ^z. If a new value is drawn, it is drawn from an AR(1) process:

$$z' = (1 - \rho_z)\mu_z + \rho_z z + \sigma_z(s) u_z, \qquad u_z \sim N(0, 1), \tag{12}$$

ρ_z controls the autocorrelation of TFPQ and μ_z the mean. $\sigma_z(s)$ controls the standard deviation of innovations to TFPQ, which depends on the aggregate uncertainty state, s. Shocks are drawn from a normal distribution in order to avoid mechanical effects on mean productivity from changes in uncertainty. Similarly, the firm draws a new level of demand at rate λ^ε. The AR(1) process for demand is chosen so that mean output in the absence of adjustment costs is independent of the level of uncertainty, which, in turn, requires that e^ϵ is normal and an AR(1) process of the form

$$e^{\varepsilon'} = (1 - \rho_\varepsilon)\mu_\varepsilon + \rho_\varepsilon e^\varepsilon + \sigma_\varepsilon(s) u_\varepsilon, \qquad u_\varepsilon \sim N(0, 1), \tag{13}$$

where the parameters are defined in the same way as for the TFPQ process.

6.3 HJB and solution

Given our assumption on adjustment costs, the problem takes a simplified form with a single endogenous state variable, x. The firm will hold the capital-labour ratio constant at some optimal value b^*. Combining this with the definition of x, capital and labour are known linear functions of x: $l(x) = x(b^*)^\alpha$ and $k(x) = x(b^*)^{\alpha-1}$. The full statement of the problem, proofs, and investment policy function are relegated to Appendix D.

[23]Paying the quadratic cost only for investment rates above depreciation has no major effects on the results. This just ensures that the marginal quadratic cost in steady state (where $i = \delta k$) is exactly zero, simplifying some expressions for steady state calculations used as initial guesses in the numerical solution.

We write static cashflow as $cf = \pi(x,z,\epsilon) - i_x p_x - c(i_x, x)$, where $\pi(x,z,\epsilon) = p(zx,\epsilon)zx - wx(b^*)^\alpha$ is revenue less labour cost and $p_x \equiv p_k(b^*)^{\alpha-1} + a(b^*)^\alpha$ is the investment cost of x. p_x is just an average of the costs of investment in capital and labour.

The HJB equation describing firm value in terms of x can then be written as

$$rv(x,z,\varepsilon,s) = \max_{i_x} \pi(x,z,\varepsilon) - p_x i_x - c(i_x,x) + v_x(i_x - \delta x) + \lambda^z \left(\mathrm{E}_{z'}[v(x,z',\varepsilon,s)|z,s] - v(x,z,\varepsilon,s)\right)$$
$$+ \lambda^\varepsilon \left(\mathrm{E}_{\varepsilon'}[v(x,z,\varepsilon',s)|\varepsilon,s] - v(x,z,\varepsilon,s)\right) + \lambda^s(s)\left(v(x,z,\varepsilon,s_{-1}) - v(x,z,\varepsilon,s)\right). \quad (14)$$

Here, s is the current uncertainty state, and s_{-1} represents the other state, which is switched to at rate $\lambda^s(s)$. The terms preceded by λ^z and λ^ε denote the change in value following a jump in TFPQ and demand, respectively. The firm's only choice is the investment rate, i_x, with optimal values given by the policy function $i_x = i^x(x,z,\varepsilon,s)$. Given the non-convex adjustment costs, this will either take the value $i_x = 0$ if investment is not worthwhile, or a finite positive or negative value.[24]

6.4 Steady state results: Passthrough and dispersion

We investigate pricing with a non-CES demand curve in the presence of adjustments costs for factors at the firm level. Our model thus combines features of the price setting literature with features from the wait-and-see literature. In this section, we calibrate a steady state version of the model where uncertainty is constant ($s = 1$ and $\lambda^s(1) = 0$) and focus on cross-sectional moments. The goal is to show that our model matches well the empirical facts documented above.

Calibration We calibrate the model such that one unit of time corresponds to one year. We use a discount rate of $r = -\log(1-0.05)$ and a capital depreciation rate of $\delta = -\log(1-0.1)$. The capital elasticity α is set to 0.255 in order to match capital's share of costs in our data. We use the mean of demand μ_ϵ to normalise aggregate capital to $K = 1$ in the ergodic distribution, and we set the real wage w to normalise aggregate labour to $L = 1$ in the ergodic distribution.[25] μ_z is chosen to shift mean TFPQ such that the log average (relative) price set by firms in the ergodic distribution is equal to zero. The purchase price of capital p_k is normalized to 1.

[24] Alternatively, our model is equivalent to simply imposing a Leontief production function with a fixed ratio of capital and labor within each firm. This is true as long as the optimal ratio b^* is constant over time, as it is in our exercises.

[25] Since the model is solved in partial equilibrium, our calibration strategy targets only firm-level moments and steady-state aggregates, and not aggregate business cycle moments.

Our demand curve parameters are taken from our estimates in Section 3.2. This is the key departure of our model from a standard CES specification. Specifically, we use $\theta = 3$ and $\eta = 4.3$. Since $\eta > 0$, firms face an increasing demand elasticity when they raise their price.

We rely on non-convex adjustment cost values from Bloom et al. (2018). This ensures that our results are comparable with the existing literature, and it highlights how our demand specification changes the propagation of uncertainty shocks. Specifically, we set the resale loss of adjusting scale, $\underline{\kappa}$, to 0.3565. This combines values for the resale loss from capital (34%) and the spread between hiring and firing costs (1.8% of annual wages) taken from Bloom et al. (2018). Note that hiring costs are represented as costs on x because all non-convex adjustment costs are on overall firm scale (accordingly, the direct linear hiring cost a is set to zero). We choose not to use convex adjustment costs for calibration purposes, and set them close to zero.[26] Unlike Bloom et al. (2018), we do not include a fixed cost of investment.

We calibrate our idiosyncratic shock processes to match moments of our observed yearly firm-level TFPQ and demand data. As in Bloom et al. (2018), we generate model-simulated yearly data to correspond to our yearly data. We do this because our data is annual and may suffer from time-aggregation and measurement issues. We fix $\lambda^z = \lambda^\varepsilon = 1$ so that firms draw a new value of each shock on average once per year. We set the autocorrelation parameters for new shock draws to $\rho_z = 0.8$ and $\rho_\varepsilon = 0.6$ which implies a yearly autocorrelation of shocks in line with what we estimate on our data.[27] The standard deviations of the shocks are chosen to match the interquartile range of the log changes in demand and TFPQ in our data. Crucially, when calculating dispersion we measure demand and TFPQ exactly as we would in the data, computing yearly measures which account for time aggregation and measurement error. We target an IQR of demand and TFPQ innovations both of 0.2. This corresponds roughly to the values in the years before the Great Recession (see Figure 2). For more details about the calibration, including a table containing our calibrated parameters, see Appendix D.2.

Model validation We first validate our model's ability to generate sensible dispersion in endogenous variables in response to the TFPQ and demand dispersion measured in the

[26] Specifically, we set $\kappa = 0.0001$, and verify that further lowering the cost has no affect on the solution. Keeping a positive value of κ is helpful for the numerical solution of the model, as it implies that investment rates are finite allowing us to simply adapt the methods of Achdou et al. (2022).

[27] We estimate the autocorrelation of the idiosyncratic demand shock to be 0.625 on our data using the Anderson-Hsaio method. TFPQ is likely measured with significant error and we find a range of autocorrelation estimates depending on the method used. We select 0.8 as a sensible value based on our estimates. Moreover, this estimate is consistent with values used in the literature.

data. In the top row of Table 4(a) we give the IQRs of sales, prices, and shocks in steady state. Despite being untargeted, the model generates dispersion similar to that in the data: An IQR of sales growth of 0.195 and an IQR of price growth of about 0.064. Both of these are close to the average IQRs in non-recession years.

In the second row of Table 4(a), we provide the same dispersions calculated in a recalibrated model where the demand curve is assumed to be CES ($\eta = 0$). While the CES model generates a similar IQR for sales growth as in the non-CES model, the CES-model generates a much higher dispersion of price changes—an IQR of 0.107—from the same dispersion in shocks. Hence, only the model with non-CES demand can match both sales and price dispersion simultaneously. This finding can be explained by differences in passthrough relative to the CES model.

Table 4: Model performance: dispersion and passthrough

	$IQR(\Delta s)$	$IQR(\Delta p)$	$IQR(\Delta z)$	$IQR(\Delta \epsilon)$
Low σ state:				
Baseline ($\eta = 4.3$):	0.1950	0.0635	0.2000	0.2000
CES ($\eta = 0$):	0.2137	0.1072	0.2001	0.2002
Effect of $\uparrow \sigma$ in baseline model:				
$\uparrow \sigma_z, \uparrow \sigma_\epsilon$:	57%	25%	31%	61%
$\uparrow \sigma_z$:	15%	10%	20%	11%
$\uparrow \sigma_\epsilon$:	46%	15%	13%	55%

(a) Dispersion

	$\log p$	$\log p$	$\Delta \log p$
$\log z$	-0.3060	-0.3301	
$\log \varepsilon$	0.0901	0.0553	
$\Delta \log z$			-0.2081
$\Delta \log \varepsilon$			0.1482
Method:	OLS	IV	OLS

(b) Passthrough

These tables give moments from model simulated data. The left table gives interquartile ranges of firm-level log changes of yearly data, constructed as in our dataset. Baseline refers to our baseline model with non-CES demand, and CES to an alternative CES model. The right table gives passthrough estimated on model simulated data. The data are time-aggregated to the yearly frequency. All coefficients are significant at at least the 0.1% level. The data are generated from long simulations of a single firm of 5,000 years in the steady state version of the model with constant uncertainty.

A conclusion from our empirical work is that passthrough from shocks to prices deviates from that implied by a simple static CES optimization model. A success of our theoretical model is that it generates low passthrough from TFPQ shocks to prices in line with these findings. In Table 4(b), we calculate passthrough using the same methodology on model-generated data. The three columns gives passthrough coefficients estimated using OLS and IV in levels, and first differences. Measured in levels, the model generates passthrough of around 30%, and around 20% when measured in first differences. This contrasts starkly from the predictions of a frictionless model with CES demand, where passthrough should be 100%. The model results are instead much closer to the estimates in Section 5.2 which range between 10% and 25%.

The low passthrough in our model follows mainly from the estimated non-CES demand curve. To see why this is the case, consider the optimal policies in a special case of the model with no adjustment costs, $\kappa = \underline{\kappa} = 0$—in effect, a static profit maximization problem. Following GIR and Berger and Vavra (2019), a first-order approximation to the optimal markup first order condition yields the firm's optimal price as a log-linear function of their TFPQ only:

$$\log p \simeq -\frac{\theta}{\theta + \eta} \log z. \tag{15}$$

A derivation of this approximation can be found in Appendix D.3. This passthrough equation is directly comparable to our estimated equation (5). This allows us to evaluate how the model's predictions for optimal price setting compare with the price setting behavior we observe in the data. Whenever $\eta > 0$, passthrough is incomplete because $\frac{\theta}{\theta+\eta} > -1$. This result has a clear economic rationale. When $\eta > 0$, a firm's elasticity of demand rises as it increases its price. Firms thus find it less appealing to change their price in response to productivity changes. The benefits of lowering the price are limited because the quantity sold does not increase much, while the benefits of raising the price are limited because the quantity sold decreases considerably. Hence firms adjust prices less than one-for-one to changes in productivity.

Based on approximation (15), our empirical results imply a statically-optimal passthrough from TFPQ to prices of $\theta/(\theta + \eta) = 41\%$. The demand curve that we estimate thus rationalises quite low passthrough from TFPQ shocks to prices—nearly as low as what we estimate in our data—even when abstracting from adjustment costs.

Adjustment costs further inhibit passthrough. For a given level of input use, a firm's quantity sold is fixed at $q = zx$, and they adjust their price only to convince customers to purchase that exact quantity. If a firm chooses not to change their inputs due to adjustment costs, then the price passthrough of a TFPQ shock is further restricted.[28] This intuition is borne out in version of the model with adjustment costs. Adjustment costs make TFPQ passthrough even lower and hence closer to the data. However, the combination of both non-CES demand and adjustment costs are necessary to generate sufficiently low passthrough from TFPQ to prices to match our findings. For example, the CES model generates 82% passthrough in the IV specification (see Table A8(b)) while the non-CES model gives passthrough of 33%.

With respect to demand, we find passthrough of about 0.2 in our empirical work. Non-CES demand can not alone explain this result. As shown by approximation (15), the statically

[28] In the CES special case with our estimated $\theta = 3$ holding x fixed leads to passthrough of $1/\theta = 33\%$. In contrast, the static input optimisation implies 100% passthrough.

optimal price does not respond to demand shocks. This is true for both the CES ($\eta = 0$) and non-CES ($\eta > 0$) model. In this kind of framework, a demand shock changes the number of units that can be sold, but not the optimal price. Adjustment costs, on the other hand, can help to explain passthrough from demand to prices. If a firm does not adjust its inputs following a negative demand shock, then it must still sell the same quantity as before. But at the lower level of demand, the demand curve (9) implies that the firm must lower its price in order to convince customers to buy the same number of units. When we include adjustment costs in the model, we find demand passthrough that ranges between 5.5% and 14.8% depending on specification, see Table 4(b).

The partial adjustment of inputs in response to shocks can therefore explain a meaningful portion of the passthrough from demand shocks to prices seen in the data. For plots of the inaction regions which lead to this partial adjustment, see Figure 7 and its discussion in the next section.[29]

6.5 Aggregate response to increase in dispersion

In this section, we investigate a fully fledged version of the model in which there are two dispersion states, one low dispersion state ($s = 1$) and one high dispersion state ($s = 2$). As in Bloom (2009) and Bloom et al. (2018), increases in dispersion in the model are a source of fundamental uncertainty for the firm. We use this version of the model to study the aggregate impacts of dispersion.

Calibration of uncertainty process The calibration of the model with uncertainty shocks is identical to the calibration of the model in steady state, with the exception of the shock processes. We continue to use $\sigma_z(1)$ and $\sigma_\epsilon(1)$ to target IQRs of 0.2 for the log-changes in measured TFPQ and demand in the low uncertainty state. For the high uncertainty state, we base our calibration on the increase in dispersion seen in the Great Recession, which peaked in 2009 (see Figure 2). The peak increases in TFPQ and demand dispersion are around 30% and 60% respectively, and we use $\sigma_z(2) = 1.38\sigma_z(1)$ and $\sigma_\epsilon(2) = 1.90\sigma_\epsilon(1)$ to target these increases in time-aggregated shocks our model. Thus, in line with our findings in Section 4.3, demand dispersion increases by more the TFPQ uncertainty in times of high uncertainty. All other features of the model are calibrated as before, to match moments within the ergodic

[29]The idea that adjustment costs lead to reduced passthrough from TFPQ to prices and non-zero passthrough from demand to prices is related to ideas in Pozzi and Schivardi (2016). They show that decreasing returns to scale in production dampens TFPQ passthrough and increases demand passthrough.

distribution of the low uncertainty state.

The other feature of the uncertainty process that needs to be calibrated is the persistence of the high and low uncertainty regimes. In our data, the recessions happened roughly eight years apart, and the associated increase in dispersion lasted for one to two years. We thus choose $\lambda^s(1) = 1/8$ so that the high uncertainty state is entered on average every eight years, and $\lambda^s(2) = 1/1.5$ so that the high uncertainty state lasts one and a half years of average. Given our relatively short sample length, estimating the persistence of these regimes is challenging on our dataset. For robustness, we therefore confirm that our results hold if we instead use persistence estimates for the US taken from Bloom et al. (2018).[30]

Aggregate response to increased dispersion To understand the aggregate implications of a rise in dispersion, we simulate a recession experiment in the model. Specifically, we consider the distribution of firms, all of whom face the same level of uncertainty. We suppose that the economy has initially been in the low uncertainty state for a long time, so that we start from the ergodic distribution over firm states conditional on $s = 1$. At time $t = 0$, uncertainty switches to the high state $s = 2$. We suppose the economy remains in this state for a full year, in line with the frequency of our data. From then on, the economy reverts back to the low uncertainty state at rate $\lambda^s(2)$.[31] Let $\mu_t(k, z, \varepsilon)$ denote the distribution of firms at time t within a given simulation. Aggregates such as GDP are computed by integrating over this distribution. All plots are averages across all possible realisations of the aggregate uncertainty process from time $t = 1$ onward.

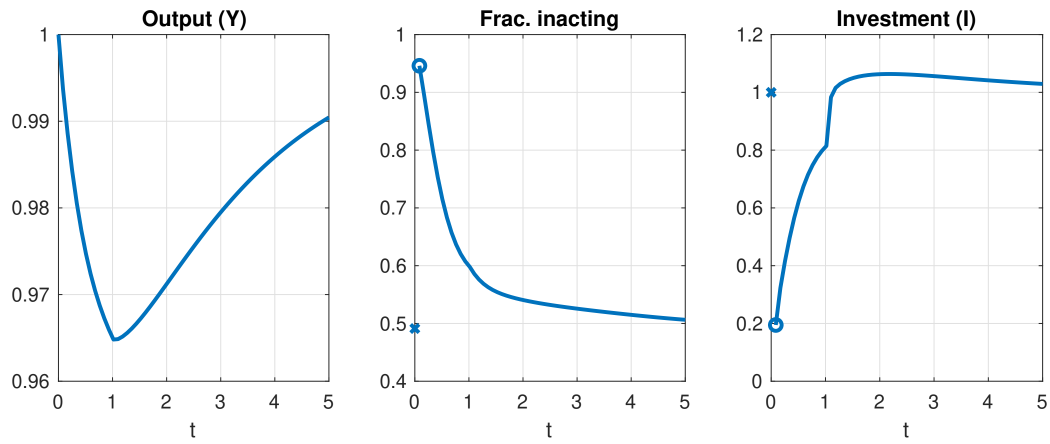
The results of this exercise are given in Figure 5. The left panel shows the impact of the uncertainty shock on aggregate output. We find that output falls by 3.5% in response to the rise in uncertainty. The remaining two panels explain the source of this fall. The centre shows the fraction of firms in their inaction regions, and hence not investing or hiring at that instant of time. Initially, around 50% of firms are inacting, but this jumps to 95% of firms following the rise in uncertainty.[32] As firms adjust this number falls gradually, and mostly recovers within one year. Concurrent with the rise of inactive firms, aggregate investment falls (right

[30] Bloom et al. (2018) estimate the persistence of the regimes using nearly 40 years of data. They report a 97.4% (94%) quarterly probability of remaining in the low (high) uncertainty state, which corresponds to $\lambda^s(1) = -4\log(0.974)$ and $\lambda^s(2) = -4\log(0.94)$ in our model. See the appendix for details.

[31] We have also simulated a permanent rise in uncertainty, and the results and intuitions are the same. The results are robust to reasonable changes in the persistence of uncertainty shocks.

[32] We plot the instantaneous inaction rate, which differs from the inaction rate as measured at a yearly frequency. Measured yearly inaction rates are lower, as firms move in and out of their inaction regions within a year. Defining inaction as having a yearly investment rate less than 1% in absolute value gives an inaction rate of 25.4% in the low uncertainty state in the model.

Figure 5: Model response to an increase in both demand and TFPQ uncertainty



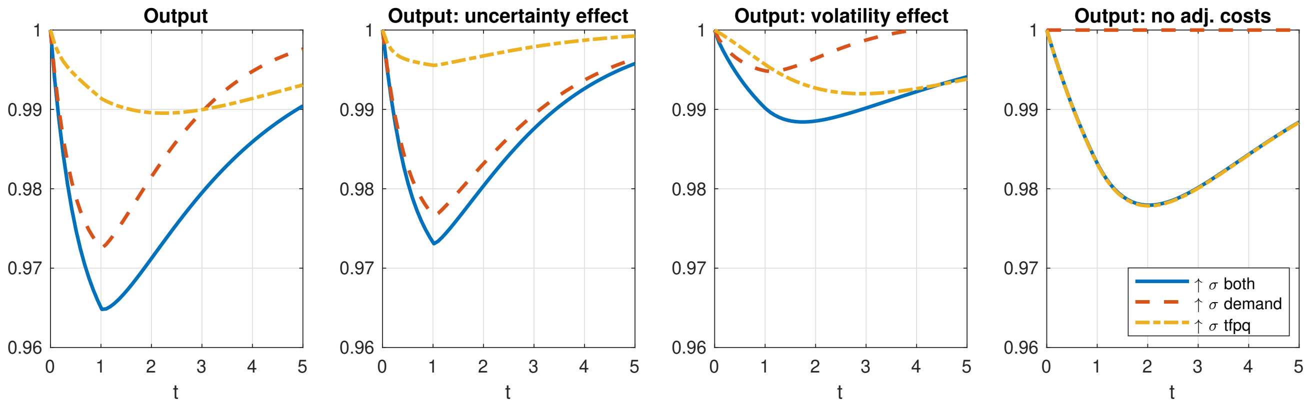
The plots give the aggregate response of the model to a switch to the high uncertainty state, $s = 2$, starting from the ergodic distribution when $s = 1$. For series which jump in response to the shock, crosses denote the pre-shock value and circles the values following the shock. The left panel gives aggregate output, middle the fraction of firms in their inaction regions at that instant, and right aggregate investment rate.

panel), capital and labour decline, and aggregate output is reduced.

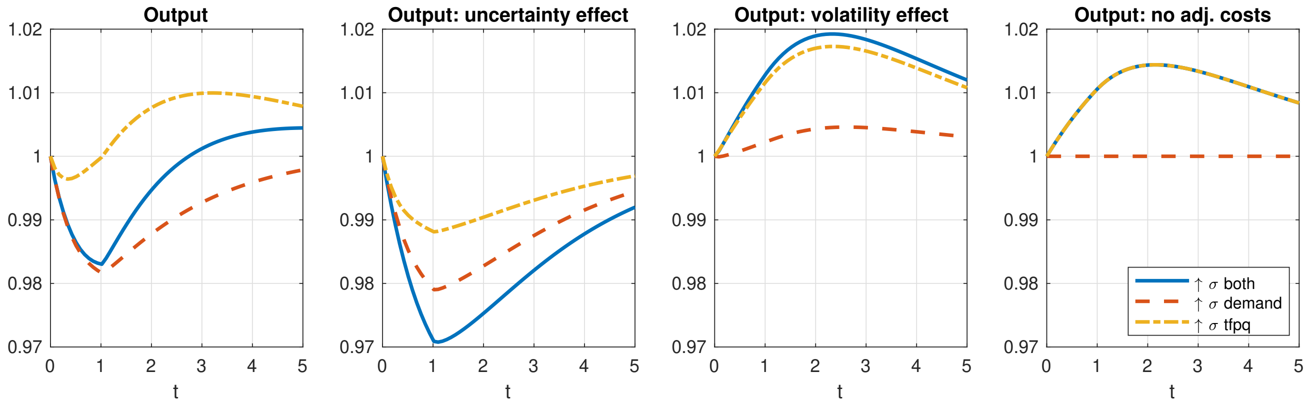
That rises in uncertainty can cause an aggregate fall in output is already well known. However, our model differs significantly from previous work due to the presence of the non-CES demand curve. This feature leads both to different aggregate effects, and to different transmission mechanisms. To see this, consider Figure 6(a). This plot shows the counterfactual paths for output from the "uncertainty" and "volatility" effects as defined by Bloom (2009). The uncertainty effect (centre-left panel) simulates an economy in which agents believe that uncertainty has increased ($s = 2$), but shocks are drawn from the less uncertain distribution ($s = 1$). The uncertainty effect captures changes in firms' precautionary behavior due to greater uncertainty about the value of future shocks. This is the aggregate effect of increased firm-level wait and see behavior, which we will discuss more below. Conversely, the volatility effect (centre-right panel) simulates an economy in which firms believe they are in the low uncertainty state ($s = 1$), but shocks are in fact drawn from the high uncertainty ($s = 2$) distribution. The volatility effect thus captures changes in aggregate outcomes due to firms actually drawing more extreme shocks when uncertainty is high.

We find that both effects are strongly negative in our model. Both the fear of higher uncertainty and the effect of realised volatility reduce GDP. This result is novel. In contrast, Bloom (2009) finds that the volatility effect is positive and that it drives a medium term rise in overall output. The different result in our model is a consequence of the non-CES demand

Figure 6: Model response to increase in demand or TFPQ uncertainty separately



(a) Estimated non-CES model: $\eta > 0$



(b) CES model: $\eta = 0$

The plots give the aggregate response of the model to a switch to the high uncertainty state, $s = 2$, starting from the ergodic distribution when $s = 1$. Panel (a) is our baseline estimated model with non-CES demand ($\eta > 0$) and panel (b) is a counterfactual model with CES demand ($\eta = 0$). Solid blue lines give the response to increased uncertainty in both shocks, dashed red is a version where only demand uncertainty rises in state 2, and dash-dotted yellow where only TFPQ uncertainty rises. The left panel gives output, the middle panels give the counterfactual output path from only the uncertainty and volatility effects respectively, and the right from a counterfactual model without adjustment costs.

curve. This can be seen in Figure 6(b), where we repeat the same exercise in the CES model ($\eta = 0$) and thereby reproduce the conclusion from Bloom (2009). Our first theoretical result is therefore that non-CES demand amplifies the total effect of uncertainty shocks on aggregate output by around 40% (peak fall of 2.5% versus 1.75%), mostly by reversing the positive volatility effect.

6.6 Understanding the mechanisms: inaction vs. markup dispersion

To investigate the mechanisms by which demand and TFPQ dispersion can drive recessions, we repeat our simulations in counterfactual economies in which only demand or only TFPQ uncertainty increases. The simulations are given in Figure 6(a) for the non-CES model and Figure 6(b) for the CES model. The dashed red and dash-dotted yellow lines give the effects of increasing only demand or only TFPQ dispersion, while the solid blue line shows results in which uncertainty of both shocks increases.

The majority of the output decline is driven by the increased dispersion in demand. This can be seen in the left panel of Figure 6(a). Increased demand dispersion explains almost all of the fall in output within the first year, and the majority of the fall for the first three years. Our second theoretical result is therefore that the negative first-moment effects of dispersion are mostly driven by demand dispersion, rather than supply dispersion.

A third theoretical result, which we discuss in detail below, is that demand dispersion affects aggregate output mostly through the uncertainty effect, while TFPQ dispersion does so mostly through the volatility effect.

Understanding the uncertainty effect The second panel of Figure 6(a) decomposes the uncertainty effect into the role of each shock. Increased demand uncertainty induces a large fall in output, while TFPQ uncertainty has a relatively minor effect. This is in part explained by the fact that the increase in demand dispersion is larger. However, the finding is also a consequence of how firms react to each kind of uncertainty.

The reason that TFPQ uncertainty is less important for wait and see behaviour than demand uncertainty can be understood from the firm-level policy functions, shown in slices in Figure 7. In this figure, $\underline{x}(z,\epsilon,s)$ denotes the investment threshold such that firms have positive investment ($i^x(x,z,\epsilon,s) > 0$) for current x below this value. Similarly, $\bar{x}(z,\epsilon,s)$ gives the disinvestment threshold, such that firms disinvest ($i^x(x,z,\epsilon,s) < 0$) for x above this value. For x between the two, the firm sets investment equal to zero. This is the inaction region where firms choose neither to invest nor disinvest due to the presence of non-convex adjustment costs. If a firm is inside its inaction region, its size will gradually decrease due to depreciation of its inputs. Once it hits the investment threshold, it will remain there until a shock hits. The inaction region is wide, and simulations reveal a 25.4% yearly inaction rate on average.

In the left and centre panels of Figure 7, we denote policy functions in the low and high uncertainty state with solid and dashed lines respectively. The shifting down of the investment threshold in the high uncertainty state is the driver of the uncertainty effect: The inaction region widens, and the extra firms which now fall into the inaction region let their inputs depreciate to a lower level before investing. Aggregate output declines because more firms choose to pause their investment while uncertainty is high. In the right panel, we focus on the investment threshold and its counterfactual values in response to an increase in demand or TFPQ uncertainty separately. We see that it shifts down further in response to increased demand uncertainty. This explains why increased demand dispersion drives a larger uncertainty effect than TFPQ.

The left panel of Figure 7 plots the investment and disinvestment thresholds across productivity levels, with the demand shock at its central value. A novel feature of our model is that optimal size is very unresponsive to TFPQ shocks, with the investment threshold remaining near one across a wide range of values of productivity. This follows from our estimated non-CES demand curve which makes low TFPQ passthrough optimal. Firms choose not to adjust their price or quantity sold much in response to idiosyncratic productivity shocks, because raising (lowering) their price leads their elasticity of demand to rise (fall). In fact, for our estimated demand curve, optimal capital is non-monotonic in productivity. For low TFPQ, raising TFPQ causes optimal capital to rise as the firm lowers its price in order to sell more units. At some point, however, higher TFPQ causes optimal input scale to fall as the firm is unable to easily increase the quantity of output sold. In this case, the firm actually requires less units of inputs to produce the same amount of goods when productivity increases.[33]

Demand shocks, on the other hand, induce large changes in optimal scale. This can be seen in the centre panel of Figure 7. This figure plots the policy functions across a range of demand shifter values, with the productivity shock at its central value. Demand shocks have a strong effect on scale because demand shocks directly affect how many units a firm can sell at a given price (i.e. shocks shift the demand curve rather than creating movement along the demand curve). This changes the target level of output, and the required capital and labour.

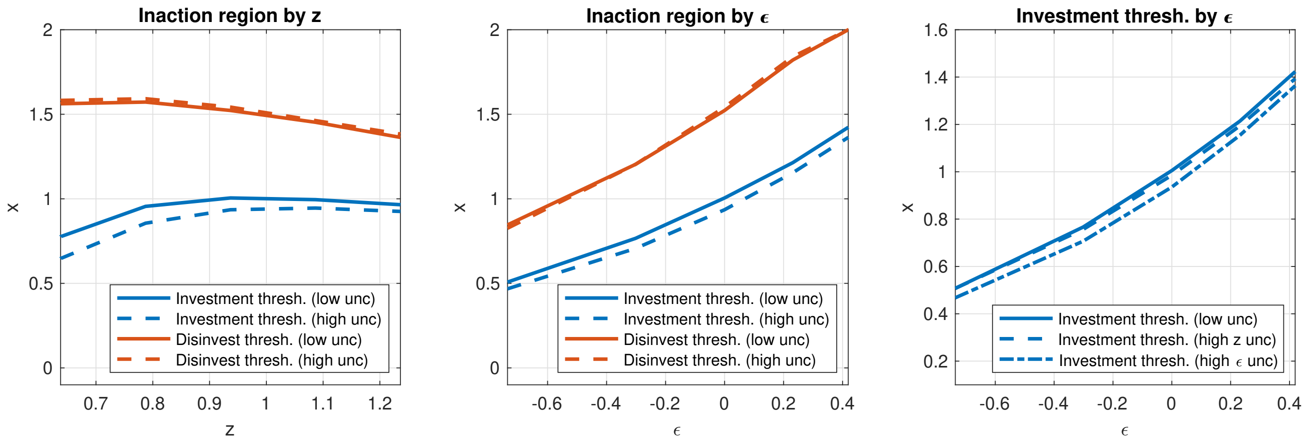
Wait and see behavior is driven by firms fearing having to pay the non-convex adjustment cost to adjust their input level in response to shocks. Because optimal prices are unresponsive to productivity, TFPQ uncertainty creates uncertainty regarding markups. While potentially unpleasant for firms, this does not induce major changes in investment through wait and see behavior since desired input use is relatively inelastic to TFPQ shocks.[34] On the other hand, demand uncertainty creates uncertainty about the quantity that can be sold and hence

[33]This can also be seen in the statically-optimal solution to the model without adjustment costs. As productivity rises, total quantity sold monotonically rises, while input use first rises and then falls. See Figure A3 in the appendix for details, and a comparison to the CES model.

[34]To see that it is non-CES demand which dampens the role of TFPQ uncertainty, note that the uncertainty effect from TFPQ dispersion is twice as large in the CES model than the non-CES model, as shown in Figure 6.

required input use. This causes strong wait and see effects and aggregate output declines from uncertainty.

Figure 7: Policy functions: Inaction regions by state and uncertainty level



The left and centre panel give slices of the firm policy functions in the low uncertainty state (solid lines, $s = 1$) and high uncertainty state (dashed lines, $s = 2$). $\underline{x}(z, \epsilon, s)$ gives the investment threshold, such that firms have positive investment for current x below this value. $\bar{x}(z, \epsilon, s)$ gives the disinvestment threshold, such that firms disinvest for x above this value. For x between the two, the firm sets investment equal to zero. The left panel plots these across z values for ϵ held at the central value, and vice versa for the central plot. These are plotted over ranges for z and ϵ covering at least 90% of their ergodic distributions in the low uncertainty state. The right panel repeats the investment threshold (plotted across ϵ values) for the counterfactual models where only z uncertainty (dashed line) or ϵ uncertainty (dash-dotted) rise in the high uncertainty state.

Understanding the volatility effect TFPQ dispersion is the main driver of the negative volatility effect. This is illustrated in the third panel of Figure 6(a), which decomposes the volatility effect into the distinct effects of demand and TFPQ dispersion. Demand plays a role during the first year, but the effect of demand fades after that. After the first year, the main driver of the volatility effect is TFPQ. Much of the peak output fall that occurs between year 1 and 2 is attributable to productivity. Given that the rise in TFPQ dispersion is smaller than the rise in demand dispersion, this represents a fundamental difference in the transmission of realised dispersion.

To understand this finding, consider the version of the model absent adjustment costs. The final panel of Figure 6(a) plots aggregate output in such a model but subject to the same shock processes as the fully fledged version. In this version of the model, output declines by over 2% in response to the rise in dispersion. This is entirely explained by the rise TFPQ dispersion, as demand dispersion has zero effect on output in the absence of adjustment costs. The following expression for optimal sales based on a first order approximation to the solution of the static model helps explain the result:[35]

$$\log s \simeq (\theta-1)\frac{\theta}{\theta+\eta}\log z+\epsilon \implies Y=\mathrm{E}\left[s\right]\simeq \mathrm{E}\left[z^{\frac{\theta(\theta-1)}{\theta+\eta}}e^{\epsilon}\right], \tag{16}$$

Each firm's optimal sales are $s \simeq z^{\frac{\theta(\theta-1)}{\theta+\eta}}e^{\epsilon}$, and averaging over firms yields aggregate output. This model provides an approximation to the path of the volatility effect in the model with adjustment costs.[36] Demand dispersion has no volatility effect because a mean preserving spread in demand shocks leads to more dispersion in sales across firms, but no change in the total amount sold.[37]

The reason why TFPQ dispersion leads to a negative volatility effect in our model, can be understood in a model without demand uncertainty. Set $\epsilon=0$ to yield $Y\simeq \mathrm{E}\left[z^{\frac{\theta(\theta-1)}{\theta+\eta}}\right]$. In the case of CES demand ($\eta=0$), this expression becomes exact and output is given by $Y=\mathrm{E}\left[z^{\theta-1}\right]$. Since we estimate $\theta>2$, this function is convex in TFPQ, meaning that an increase in the dispersion of z would raise the average value of $z^{\theta-1}$ and increase aggregate output. This is the well-known result that raising dispersion can actually raise aggregate output, which Bloom et al. (2018) refer to as the Oi-Hartman-Abel (OHA) effect (Oi, 1961; Hartman, 1972; Abel, 1983). This effect follows in our setting because optimal sales are convex in productivity when firms face either a downwards-sloping CES demand curve or, equivalently, decreasing returns to scale Cobb-Douglas production function. Intuitively, lucky firms expand by more than unlucky firms contract, so more dispersion raises aggregate output.

However, this result is entirely overturned for our estimated non-CES demand curve. To see this, for $\theta=3$ and $\eta=4.3$ we have $\frac{\theta(\theta-1)}{\theta+\eta}=0.82<1$ and hence $Y\simeq \mathrm{E}\left[z^{0.82}\right]$. A firm's optimal sales is now concave in productivity, meaning that lucky firms expand by less than unlucky firms contract. An increase in dispersion thus lowers aggregate output. To the best of our knowledge, this is a novel conclusion. Following an increase in TFPQ dispersion,

[35] To do this, we combine our first order approximation of the firm's optimal price setting behavior, (15), with a first order approximation of the non-CES demand curve itself. Taking a first order approximation of (2) around $\log p=0$ simply yields the CES demand curve, $\log q=-\theta\log p+\alpha_{fe}+\varepsilon$. Combining these two equations yields the result, where we set $\alpha_{fe}=0$ for expositional clarity.

[36] The difference between the two is driven by adjustment costs, which alter how firms respond to the shocks. For TFPQ shocks, adjustment costs reduce how much firms adjust their inputs, which shrinks the output fall relative to the no adjustment cost model. For demand shocks, the slight fall in output with adjustment costs is driven by larger changes in inputs for firms who actively adjust downwards in response to negative shocks than for those adjusting upwards for positive shocks.

[37] To make this point clear, we assume that demand shocks are drawn such that e^{ϵ} is normally distributed. As per the formula above, sales are proportional to e^{ϵ} rather than ϵ, and hence e^{ϵ} is the true measure of demand, in the sense of how many units a firm can sell at a given price. For this distribution, an increase in the dispersion of ϵ has no effect on $\mathrm{E}\left[e^{\epsilon}\right]$, and hence on aggregate output. If ϵ is normally distributed, a rise in demand dispersion causes output to rise, further increasing the difference between the volatility effect for TFPQ and demand shocks.

firms adjust their prices by less than their TFPQ adjusts due to incomplete passthrough. This increases markup dispersion and thereby exacerbates misallocation and reduces aggregate output.[38] This moderates the OHA effect and, for $\eta > \theta(\theta - 2)$, as we estimate, actually overturns it.[39] This explains why the volatility effect from TFPQ is negative in our model, while it is positive in the standard (i.e. CES) model.

Other exercises and robustness In Appendix D.4, we present results from other exercises. We first perform a model-based variance decomposition, which finds similar results to our empirical semi-structural variance decomposition. To investigate time-varying passthrough, we compare passthrough in the model in the low and high uncertainty state. We find that demand passthrough falls when uncertainty rises. This is in line with the evidence on time varying passthrough we discussed in Section 5, where we found suggestive evidence that demand passthrough falls in times of high dispersion.

Among other robustness results, we solve a version of the model where labor is not subject to adjustment costs, while capital is, in line with the model of Bachmann and Bayer (2013). In contrast to our baseline model, where adjustment costs are on both factors as in Bloom et al. (2018), the ability to costlessly adjust labor dampens the wait and see effect. Nonetheless, our main results that (1) demand shocks drive more wait and see behavior than TFPQ shocks, and (2) non-CES demand reverses the sign of the OHA effect, remain true.

7 Conclusion

In this paper, we use rich Swedish micro-data to investigate firm-level dispersion of demand and physical productivity (TFPQ) over the business cycle. First, we document that both demand dispersion and TFPQ dispersion are countercyclical. Importantly, we find that demand

[38]To see that it is markup dispersion that causes output to fall, rather than the non-linearity of the demand curve itself, recall that these first order approximations use the linear CES demand curve. For additional intuition, consider the efficient solution to the non-CES model without adjustment costs, which instead requires that price equals marginal cost, giving $p = c/z$ for some constant c. The same first order approximation then gives that efficient aggregate output is equal to $Y \simeq \mathrm{E}\left[z^{\theta-1}\right]$. Thus, efficient output always increases in response to an increase in TFPQ dispersion (since $\theta > 1$), while actual output decreases. The difference between the two models is that efficient output features no markups at any firms, while actual output features positive markups, which become more dispersed following the TFPQ dispersion increase as long as $\eta > 0$.

[39]This result is derived from a first order approximation, but we verify in Figure A3 that the true non-linear sales policy function is concave in productivity for our estimated values, and the decline in output reported in the final panel of Figure 6 comes from the true non-linear solution to the model. We additionally verify that the maximization problem is still well behaved, meaning that profit is concave in prices and there is an interior optimum price.

dispersion is more cyclical than TFPQ dispersion. Using variance decomposition exercises, we confirm that demand is a prominent driver of dispersion over the business cycle.

Second, we investigate how firm prices respond to productivity and demand shocks. We find significant and economically meaningful deviations from the benchmark (constant markup) pricing model. Firms respond little to productivity shocks but meaningfully to demand shocks. Motivated by this finding, we estimate a demand curve that allows for a non-constant elasticity of demand analogous to Kimball (1995). For the parameters that we estimate, firms lose more customers by raising their price than they gain by lowering their price, providing direct evidence for "real rigidities."

Finally, we embed our estimated demand curve into a heterogeneous-firm model with non-convex input adjustment costs, following Bloom et al. (2018), to study the aggregate effects of idiosyncratic dispersion and uncertainty. The non-constant elasticity of demand dramatically shapes the transmission of uncertainty shocks to aggregate first moments. In the model, demand uncertainty has a powerful effect on aggregate output, while TFPQ uncertainty is relatively inconsequential. The reason is that uncertainty related to demand leads to "wait and see" effects while TFPQ uncertainty is absorbed in markups. Nevertheless, TFPQ dispersion is still harmful for aggregate output because the induced markup dispersion leads to misallocation.

References

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

ACHDOU, Y., J. HAN, J.-M. LASRY, P.-L. LIONS, AND B. MOLL (2022): "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," *Review of Economic Studies*, 89, 45–86.

ARKOLAKIS, C., A. COSTINOT, D. DONALDSON, AND A. RODRÍGUEZ-CLARE (2018): "The Elusive Pro-Competitive Effects of Trade," *Review of Economic Studies*, 86, 46–80.

ARMELIUS, H., I. HULL, AND H. S. KÖHLER (2017): "The Timing of Uncertainty Shocks in a Small Open Economy," *Economics Letters*, 155, 31–34.

BACHMANN, R. AND C. BAYER (2013): "'Wait-and-See' Business Cycles?" *Journal of Monetary Economics*, 60, 704–719.

——— (2014): "Investment Dispersion and the Business Cycle," *American Economic Review*, 104, 1392–1416.

BACHMANN, R., K. CARSTENSEN, S. LAUTENBACHER, AND M. SCHNEIDER (2021): "Uncertainty and Change: Survey Evidence of Firms' Subjective Beliefs," NBER WP No. 29430.

BALL, L. AND D. ROMER (1990): "Real Rigidities and the Non-Neutrality of Money," *Review of Economic Studies*, 57, 183–203.

BASU, S. (1996): "Procyclical Productivity: Increasing Returns or Cyclical Utilization?" *Quarterly Journal of Economics*, 111, 719–751.

BERGER, D. AND J. VAVRA (2019): "Shocks versus Responsiveness: What Drives Time-Varying Dispersion?" *Journal of Political Economy*, 127, 2104–2142.

BLOOM, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685.

BLOOM, N., S. J. DAVIS, L. S. FOSTER, S. W. OHLMACHER, AND I. SAPORTA-EKSTEN (2022): "Investment and Subjective Uncertainty," NBER WP No. 30654.

BLOOM, N., M. FLOETOTTO, N. JAIMOVICH, I. SAPORTA-EKSTEN, AND S. J. TERRY (2018): "Really Uncertain Business Cycles," *Econometrica*, 86, 1031–1065.

CARLSSON, M. (2017): "Microdata Evidence on the Empirical Importance of Selection Effects in Menu-Cost Models," *Journal of Money, Credit and Banking*, 49, 1803–1830.

CARLSSON, M., A. CLYMO, AND K.-E. JOSLIN (2022): "Dispersion Over the Business Cycle: Passthrough, Productivity, and Demand," Sveriges Riksbank WP No. 414.

CARLSSON, M., J. MESSINA, AND O. NORDSTRÖM SKANS (2021): "Firm-level Shocks and Labour Flows," *Economic Journal*, 131, 598–623.

CARLSSON, M., J. MESSINA, AND O. N. SKANS (2016): "Wage Adjustment and Productivity Shocks," *Economic Journal*, 126, 1739–1773.

CARLSSON, M. AND O. N. SKANS (2012): "Evaluating Microfoundations for Aggregate Price Rigidities: Evidence from Matched Firm-Level Data on Product Prices and Unit Labor Cost," *American Economic Review*, 102, 1571–95.

CETTE, G., N. DROMEL, R. LECAT, AND A.-C. PARET (2015): "Production Factor Returns: The Role of Factor Utilization," *Review of Economics and Statistics*, 97, 134–143.

DAVIS, S. J., J. HALTIWANGER, AND S. SCHUH (1996): *Job Creation and Destruction*, The MIT Press.

DE LOECKER, J., P. K. GOLDBERG, A. K. KHANDELWAL, AND N. PAVCNIK (2016): "Prices, Markups, and Trade Reform," *Econometrica*, 84, 445–510.

DECKER, R. A., J. HALTIWANGER, R. S. JARMIN, AND J. MIRANDA (2020): “Changing Business Dynamism and Productivity: Shocks versus Responsiveness,” *American Economic Review*.

DHYNE, E., A. PETRIN, V. SMEETS, AND F. WARZYNSKI (2021): “Theory for Extending Single-Product Production Function Estimation to Multi-Product Settings,” mimeo Yale University.

ESLAVA, M. AND J. HALTIWANGER (2020): “The Size and Life-cycle Growth of Plants: The Role of Productivity, Demand and Wedges,” NBER WP No.27184.

FÉLIX, S. AND C. MAGGI (2019): “What is the Impact of Increased Business Competition,” Banco de Portugal WP No. 201904.

FOSTER, L., J. HALTIWANGER, AND C. SYVERSON (2008): “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?” *American Economic Review*, 98, 394–425.

——— (2016): “The Slow Growth of New Plants: Learning about Demand?” *Economica*, 83, 91–129.

GILCHRIST, S., R. SCHOENLE, J. SIM, AND E. ZAKRAJŠEK (2017): “Inflation Dynamics During the Financial Crisis,” *American Economic Review*, 107, 785–823.

GOPINATH, G., O. ITSKHOKI, AND R. RIGOBON (2010): “Currency Choice and Exchange Rate Pass-Through,” *American Economic Review*, 100, 304–36.

GUISO, L. AND G. PARIGI (1999): “Investment and Demand uncertainty,” *Quarterly Journal of Economics*, 114, 185–227.

HALTIWANGER, J., R. KULICK, AND C. SYVERSON (2018): “Misallocation Measures: The Distortion That Ate the Residual,” NBER WP No. 24199.

HARTMAN, R. (1972): “The Effects of Price and Cost Uncertainty on Investment,” *Journal of Economic Theory*, 5, 258–266.

HEYMAN, F., H. SVALERYD, AND J. VLACHOS (2013): “Competition, Takeovers, and Gender Discrimination,” *Industrial and Labor Relations Review*, 66, 409–432.

HOTTMAN, C. J., S. J. REDDING, AND D. E. WEINSTEIN (2016): “Quantifying the Sources of Firm Heterogeneity,” *Quarterly Journal of Economics*, 131, 1291–1364.

KAAS, L. AND B. F. KIMASA (2021): “Firm Dynamics with Frictional Product and Labor Markets,” *International Economic Review*, 62, 1281–1317.

KEHRIG, M. (2015): “The Cyclicality of Productivity Dispersion,” mimeo Duke University.

KIMBALL, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit, and Banking*, 27, 1241–1277.

KLENOW, P. J. AND J. L. WILLIS (2016): "Real Rigidities and Nominal Price Changes," *Economica*, 83, 443–472.

LINDÉ, J. AND M. TRABANDT (2018): "Should We Use Linearized Models to Calculate Fiscal Multipliers?" *Journal of Applied Econometrics*, 33, 937–965.

MELANDER, O. (2009): "The effect of Cash Flow on Investment: an Empirical Test of the Balance Sheet Channel," Sveriges Riksbank WP No. 228.

MONGEY, S. AND J. WILLIAMS (2017): "Firm Dispersion and Business Cycles: Estimating Aggregate Shocks Using Panel Data," mimeo University of Chicago.

OI, W. Y. (1961): "The Desirability of Price Instability under Perfect Competition," *Econometrica*, 29, 58–64.

POZZI, A. AND F. SCHIVARDI (2016): "Demand or Productivity: What Determines Firm Growth?" *RAND Journal of Economics*, 47, 608–630.

SHAPIRO, M. D. (1993): "Cyclical Productivity and the Workweek of Capital," *American Economic Review*, 83, 229–233.

SMEETS, V. AND F. WARZYNSKI (2013): "Estimating Productivity with Multi-Product Firms, Pricing Heterogeneity and the Role of International Trade," *Journal of International Economics*, 90, 237–244.

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

APPENDIX

Additional details may be found in the appendix to the earlier working paper Carlsson et al. (2022).

A Data and variables

A.1 Sources

Our firm register data come from Statistics Sweden (SCB) and the Swedish Tax Authority (Skatteverket). We exploit persistent firm and plant identifiers to combine data from multiple surveys. When necessary, we construct firm variables by aggregating across plants. For variables like revenue and employment this is straightforward as firm level variables can be (re-) produced by a simple summation across plants. In other cases, we produce a firm level aggregate by weighting the plant level observations by the plant's share of firm sales. This approach is relevant in the case of averages or indices. For example, we weight plant-level measures of utilization to get a firm-level average.

Structural Business Statistics (FEK) The *Företagens Ekonomi/Stuctural Business Statistics* survey includes basic bookkeeping variables for nearly all Swedish industrial firms. This survey also includes detailed information about investments and assets, but only for a subsample of firms. The sample from the FEK survey that we have access to covers the period 1996 to 2013 and includes about 14 million firm-year observations. The main variables that we take from this survey are net sales (S), number of employees (l), total personnel costs (C^L), raw materials (C^M), gross investment (I), investment in machinery and equipment (I^E), the value of machinery and equipment (E), investment in property and buildings (B), the value of property and buildings (B), change in stocks of work in progress (D), and inventories (T). These variables adhere to standard accounting conventions.[40]

Production of Commodities and Industrial Services (IVP) The *Industrins varuproduktion/Production of Commodities and Industrial Services* survey provides revenue, quantity, and price data for products defined at the 8-digit level according to the EU's combined nomenclature system.[41] Among other uses, the IVP data are used to construct the (domestic part) of the producer price index (PPI). The IVP data are available since 1996 and are based on a stratified sample with an overrepresentation of large firms.

[40] Additional information about this survey, including variable definitions can be found on the Statistics Sweden website: www.scb.se/.../foretagens-ekonomi/

[41] See www.scb.se/.../industrins-varuproduktion-ivp/ for more information and variable definitions.

We use plant level price indices (*ArbstIndex*) provided in the IVP data to build firm-level price indices.[42] These plant level indices are computed as chained Laspeyres indices based on goods level data. Before aggregating to the firm level, we remove the 0.5% most extreme plant price changes. After doing so, the 1% to 99% distribution ranges from 75 to 146, with a median of 101. To get a firm level price index $\tilde{P}_{i,t}$, we weight the price indices according to their plant's share of firm sales. Note that we re-base the plant price indices to 1 at the beginning of a series of consecutive series of firm observations. Our raw measure of firm-level price $P_{i,t}$ is thus the cumulative price change between the first period in which a firm is observed t_0 and period t: $P_{i,t} = \Pi_{j=t_0+1}^{t} \tilde{P}_{i,j}$.

Industrial capacity utilisation The *Industrins kapacitetsutnyttjande / Industrial capacity utilisation* survey is part of the Business cycle for industry (Konjunkturstatistik för Industrin, KFI) data collection framework. In this survey, managers report on their degree of capacity utilization and evaluate various standardized measures of the business environment. The survey is conducted at the quarterly frequency and is based on a stratified sample of about 2000 industrial firms. Firms with 200 employees or more are fully surveyed while smaller firms are randomly sampled from strata. Although the sampling unit is the firm, utilization data are reported at the level of production facilities.[43]

The measure of capacity utilization (*kapacitetsutnyttjande*) that we take from this survey is defined as the ratio of actual utilization to full utilization, expressed in percent. Full utilization means that machinery and staffing are fully employed under the prevailing production setup. Importantly, prevailing production setup is defined relative to the intended level of production. Consider a situation in which day shifts are normal. If the firm temporarily introduces night shifts, then utilization is above 100%. On the other hand, if a firm permanently adds night shifts, then this reflects a change in prevailing production setup. A similar logic is relevant for furloughs as compared to planned downsizing. Managers are in addition explicitly reminded (1) to disregard seasonal variations (e.g. summer vacations), (2) that capacity utilization can exceed 100%, (3) to evaluate capacity utilization based on the working hours and shifts that can be considered normal, and (4) if measures have been taken with the intention of changing production capacity, the new situation shall be considered normal.

The other variable we use from the business cycle survey is an indicator of whether low

[42] For robustness, we also construct firm-level price indices directly via aggregation from the product level. This approach yields nearly identical results and we therefore favor the IVP indices because of somewhat better data coverage.

[43] Additional information about this survey can be found on SCB's website: scb.se/en/data-collection/surveys/business-cycle-statistics-for-industry/.

capacity utilization is primarily the result of "insufficient demand" (*otillräcklig efterfrågan*).

To aggregate the utilization and insufficient demand variables to the firm level, we compute the average across production facilities using revenue weights.[44] To convert the quarterly data to the annual frequency, we average the firm-level observations within the year. This yields two firm-level variables. The first is the average level of capacity utilization u. The average degree of capacity utilization is about 88% and the median level of capacity utilization is about 91%. The standard deviation of firm level capacity utilization is 14.1%. The 1st-percentile is 40% utilization and the 99th-percentile is 105% utilization. The second variable is the share of plants reporting insufficient demand $\mathcal{I}(\check{e})$. This variable ranges between 0 and 1. Overall, about 30% of firms report insufficient demand at all plants.

There is a strong relationship between capacity utilization and the measure of insufficient demand. Based on a fixed effects regression with firm and sector-year fixed effects, we that a firm that reports insufficient demand exhibits 15% lower utilization in the same year. During the Great recession, this relationship is even stronger. For a firm that reports insufficient demand during the Great Recession, we expect 26% lower utilization. Both findings are significant at the 0.001 level.

Other data sources Besides the datasets described in the main text, we also use variables from a number of other sources. For a sectoral price index P_t^s, we use Statistics Swedens producer price.[45]. The PPIs can be matched to sectors at the 2-, 3-, and 4-digit level (though only the 2-digit level is available for some small sectors). As noted above, the PPIs are constructed from the IVP data. We construct an investment price index P_t^I based on price changes for gross fixed capital investment (*Fasta bruttoinvesteringar*). These price changes are available from SCB at the two-digit sector level: www.statistikdatabasen.scb.se. Depreciation rates δ_t^s for equipment and structures are based on Melander (2009). Melander uses depreciation rates from the Bureau of Labor Statistics (BLS) to construct depreciation rates for the Swedish Industrial Classification at the two-digit level. The depreciation rates for equipment vary by sector, while the depreciation rate for structures is constant across sectors.

[44]Typically, production facilities coincide with plants. However, plants are occasionally comprised on multiple production facilities. Because sales data is not available at the facility level, we use the average level of utilization across production facilities as the measure of plant utilization.

[45]This index is available via Statistics Sweden's Statistical Database: www.scb.se/.../prisindex-i-producent-och-importled-ppi/

A.2 Dataset definition and construction

Our analyses rely on panel techniques and comparisons over time. However, the firm identifiers in the data refer to a legal entity, even if that entity undergoes categorical changes. For instance, a firm may open or close production facilities and thereby fundamentally change the scale or nature of production. In our main analyses, we therefore assign a new identifier whenever there is reason to believe that there may have been a categorical change in the nature of the firm. Specifically, we give new firm identifiers whenever the set of plants within a firm changes, if there is an extreme change in the level of one or more variables (discussed further below), or if there is a one or more year gap in the observation of the firm. Firm identifiers therefore refer to a stable set of continuously operating entities. Defining the identifiers in a careful way is important in the case of multi-plant firms because we re-base the plant price indices to the same initial level when constructing the firm price index. We also use the identifiers to harmonize industry codes within consecutive series of observations, picking the most commonly observed sector affiliation. Defining firms as a stable set of plants ensures that comparisons over time make sense. However, the approach creates issues with respect to large firms. The reason is that large firms often open and close plants. If a new firm id is assigned every time a large firm changes its set of plants, new identifiers would be assigned in nearly every year. We therefore find it fruitful to prune "marginal" plants before aggregating plants to the firm level. Specifically, we exclude a plant if either (1) the plant accounts for less than 1% of sales, or (2) if the plant accounts for less than 5% of sales and is present in the data for only a single year.

To ensure the robustness of our approach to firm definition, we also produce results based on the firm identifiers provided in the data. This approach has the advantage that we retain longer consecutive series of observations, and we do not have to do any pruning of "marginal" plants. This sample ends up being about 6.5% larger. Results based on this sample are comparable to those based on our main sample. For example, demand estimates and cyclicality are similar (see Table A4). This sample is interesting as it allows us to investigate changing plant counts during our time series. In a simple descriptive analysis, we find that the average number of plants at a firm increased sharply in 2001. But we do not find evidence of cyclicality. Growth in the number of plants is close to zero in most other years. Perhaps surprisingly, we do not find much impact of the Great Recession on firm structure. This pattern is similar to what we find for number of products.

We perform several rounds of data cleaning and data preparation. We drop a small number of observations that are nonsensical, implausible, or missing data for key variables. To handle extreme values in our bookkeeping data, we trim based on a multidimensional measure based on growth in the key variables on a per employee basis. Let x denote the per employee value of the variables $\{V, K, E, C^M, C^L\}$ in constant prices. For each firm i we compute a measure of absolute change for each of these variable:

$$dx_{i,t} = \begin{cases} \frac{x_{i,t}}{x_{i,t-1}} & \text{if} \quad x_{i,t} > x_{i,t-1} \\ \frac{x_{i,t-1}}{x_{i,t}} & \text{if} \quad x_{i,t-1} > x_{i,t}. \end{cases}$$

dx_t is thus a measure that increases in the size of the relative change regardless of whether the change is an increase or decrease. Next, we construct the multidimensional measure $\chi_{i,t} = \sqrt{\sum_x dx_{i,t}^2}$. χ is large when one or more of the variables exhibits a large change. Based on χ, we then assign a new firm identifier for the 1% largest values of χ. This allows us to (potentially) handle systematic changes to the firm differently from measurement error (for example, incorrect units). For example, if we observe a single extreme change in a firm panel, then this potentially reflects a permanent change to the firm. It therefore makes sense to assign a new panel identifier. However, if we observe multiple extreme values for χ in a row, then it seems likely that there has been some measurement error. For example, if χ increases sharply in one year and decreases sharply in the next this seems to indicate a transitory disruption to firm data. Such data are discarded because we do not retain firms in our panel if they do not exist for multiple periods.

The other instance in which we perform trimming of extreme values is our price data, some of which appears to be affected by mismeasurement. In this dataset, we drop observations associated with the 1.5% most extreme annual price changes at the firm level.

Assembly and samples Our main dataset covers the period 1998-2013 and includes about 15,000 observations based on about 3000 unique firms. The assembly of this dataset is affected by three bottlenecks in particular. The first is the presence of investment data. We require investment data in order to construct our capital measure, but investment data is only available for a subsample of the FEK data. The second bottleneck is the availability of price data from the IVP, which we require to compute TFPQ. And the third bottleneck is the limited coverage of the Industrial Capacity Utilization survey which we require for the utilization adjustment. In addition to our main dataset, we conduct robustness exercises based on the sample of single product firms and a 12-year balanced panel.

A.3 Variables

We maintain the convention that nominal variables are expressed in uppercase and real in lowercase letters, where possible. We use i to index firms, t years, and s sectors.

For expenditure variables such as remuneration (C^l) and raw materials (C^M), we deflate nominal expenditure by a sectoral producer price index P^s defined at the four digit level when available and defined at the two-digit level if there are few observations. This yields the real cost of labor $c^l_{i,t} = \frac{C^l_{i,t}}{P^s_t}$ and the real cost of raw materials: $m_{i,t} = \frac{C^M_{i,t}}{P^s_t}$ We also use the sectoral price index P^s_t to compute firm relative price: $p_{i,t} = \frac{P_{i,t}}{P^s_t}$

We construct our output and capital variables based on their economic definitions. We compute real output as the value of sales, $S_{i,t}$, plus the change in inventories $D_{i,t}$, deflated by the firm price $P_{i,t}$. This is similar to the approach proposed by Smeets and Warzynski (2013). Our real value-added measure $v_{i,t}$ is then the difference between real output and the real cost of raw materials, consumables, and goods for resale (C^M): $v_{i,t} = \frac{S_{i,t}+D_{i,t}-C^M_{i,t}}{P_{i,t}}$.

Our real capital stock measure, k, is based on combined capital stock for structures and equipment. We construct our capital series for structures and equipment using a perpetual inventory approach (PIM) in which the capital stock is based on accumulated investment adjusted for depreciation, $k_t = (1-\delta)k_{t-1} + i^k$ where δ is a sector specific depreciation rate, k is real capital, and i^k is nominal investment deflated by the investment price index. We pick for the initial capital stock whichever of the "steady state" value or deflated current book value is larger, where we compute the steady state value as real investment in the initial period divided by the depreciation rate. The remaining periods are then computed according to the PIM, unless the book value is larger, in which case we use the book value.

We use the Swedish Standard Industrial Classification (SNI) to define our sector fixed effects. The SNI classification is the Swedish implementation of the *Statistical Classification of Economic Activities in the European Community* (NACE). The first four digits of the SNI codes are identical to NACE codes. We focus on SNI sectors 10-33 which comprise industrial production (NACE group C “manufacturing”). When using sector fixed effects, we include fixed effects at the two-digit sector level. When using sectoral price indices, we use 4-digit sector indices where available and otherwise 2-digit sector prices.

A.4 Data description

Table A1 shows descriptive statistics for basic firm variables relative to the number of firm employees (with the exception of number of employees itself). For example, sales per worker is computed as S/N. All values are reported in units of 1000 SEK. Perhaps the most notable feature of the data is the substantial positive skew, with mean values typically larger than the median.

Table A1: Descriptive statistics for firm variables

	Full			Main			Balanced		
	mean	p25	p75	mean	p25	p75	mean	p25	p75
employees	132	28	95	278	55	246	103	34	86
sales	1991	1073	2316	2301	1260	2758	1904	1098	2312
value-added	884	506	958	1073	578	1124	774	506	924
capital	1511	336	1398	1519	409	1515	1349	396	1365
intermediates	1254	508	1473	1434	607	1742	1170	521	1479
Observations	48047			15044			8712		

The top panel presents the mean and interquartile range for each sample. The bottom panel provides additional details about the distribution of variables in our main sample. With the exception of number of employees, all variables are measured in units of 1000 SEK per worker.

Table A2 presents descriptive statistics for prices and capacity utilization based on our main sample. Firm-level price $P_{i,t}$ and relative price $p_{i,t}$ are presented in terms of growth: $\Delta \ln P_{i,t}$ and $\Delta \ln p_{i,t}$. The nominal prices $P_{i,t}$ exhibit positive growth over time on average, substantial kurtosis, and positive skewness. Large price increases are more common than large price decreases. Relative prices $p_{i,t}$, in contrast, show little systematic growth on average and relative price growth is more symmetrically distributed than nominal price growth. As with nominal price growth, however, large relative price increases are more common than large price decreases. How common are price changes? For about 3.3% of observations, there is no change in the firm price from year to year. At the product level, however, sticky prices are common. About 50% of our observations are based on firms for which at least one product-level price is unchanged from the previous year.

For utilization, we present both the degree of utilization u (in percent) and utilization growth $\Delta \ln u$. Average utilization is 88% and the median utilization is 91%. The distribution of u is fairly skewed, reflecting a long negative tail of firms with low utilization. With respect to utilization changes Δu, however, skewness is less pronounced. Both increases and decreases in utilization are common. How common is full utilization? About a quarter of observations are associated with 100% or more utilization ($u \geq 100$).

Cyclicality We illustrate the cyclicality of growth of firm variables in Figure A1 (shown at the end of this section). Growth rates are computed as log changes for all variables apart

Table A2: Summary statistics for prices and utilization

	mean	sd	p1	p25	p50	p75	p99	skewness	kurtosis
$\Delta \ln P$	1.68	8.42	-19.87	-1.80	0.95	4.38	31.14	1.21	10.13
$\Delta \ln p$	0.01	7.95	-23.77	-3.09	-0.04	2.81	25.96	0.32	9.04
Δu	-0.42	13.77	-43.29	-3.69	0.00	3.30	37.16	-0.59	45.36
u	87.93	13.31	43.50	80.50	91.00	99.05	105.00	-1.43	6.13

Summary statistics for prices and utilization are based on our main sample. $\Delta \ln P$ denotes growth of firm-level nominal price and $\Delta \ln p$ denotes growth of firm-level relative price. u is the degree of capacity utilization and $\Delta \ln u$ denotes utilization growth. All growth rates are expressed in percent.

from investment. For investment, we present investment in period t relative to capital stock in period t (this ratio can be interpreted as the growth rate of capital). All growth rates are de-meaned by sector-year growth. The basic pattern is countercyclical dispersion. The only exceptions are prices and investment. Prices are countercyclical in 2009 but not clearly so in 2001. Investment is somewhat ambiguous overall, but procyclical in 2001 and 2009.

Table A3: Cyclicality of firm variables

	2001	2009
sd.($\tilde{\Delta} S$)	17	35
sd.($\tilde{\Delta} P$)	-4	56
sd.($\tilde{\Delta} l$)	6	30
sd.($\tilde{\Delta} m$)	12	23
sd.($\tilde{\Delta} u$)	2	39
sd.($I\tilde{/}K$)	-5	-4

(a) Standard Deviation

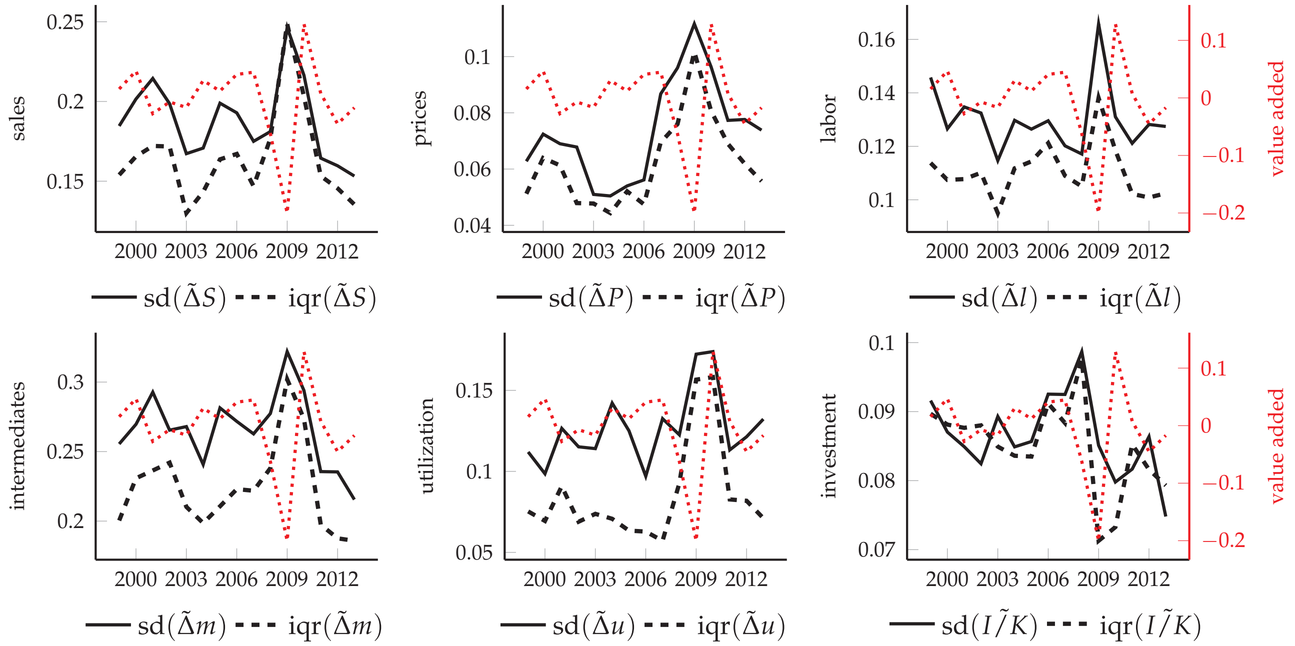
	2001	2009
iqr.($\tilde{\Delta} S$)	9	58
iqr.($\tilde{\Delta} P$)	8	79
iqr.($\tilde{\Delta} l$)	-2	26
iqr.($\tilde{\Delta} m$)	8	39
iqr.($\tilde{\Delta} u$)	16	99
iqr.($I\tilde{/}K$)	1	-18

(b) Interquartile Range

This table presents percentage changes in dispersion measures for the 2001 and 2009 recessions relative to the average over all other years. The left table shows comparisons based on the standard deviation and the right for the interquartile range. All measures have been de-meaned by sector-year.

We quantify the cyclicality in Table A3. For each variable, the table presents the change in the dispersion of the variable in 2001 and 2009 relative to the average across all other years when excluding those two years. In the Table A3(a), we present the change using the standard deviation. In Table A3(b), we present the change based on the interquartile range. There are only five exceptions to the finding of countercyclical volatility. As noted above, capital, if anything, displays procyclicality, becoming less dispersed during recessions. The other exceptions are price growth and employment growth in 2001. Although these variables are clearly countercyclical during the Great Recession, the findings are ambiguous for 2001. For both variables, the cyclicality is different when measured by the standard deviation as compared to the interquartile range.

Figure A1: Dispersion of firm variables, 1999-2013 (standard deviation)



This figure shows the standard deviation (sd) and interquartile range (iqr) across firms of log changes for key variables, calculated each year. Both dispersion measures are computed within sector based on $\tilde{\Delta}x_{i,t}$ for each variable $x \in \{S, P, l, m, u, I/K\}$. Sales S is given by firm turnover deflated by a sectoral producer price index. Price $P_{i,t}$ is given by a firm-level price index. Number of employees l is measured in full-time equivalents. Intermediate goods m is given by the value of the stock of raw materials and consumables deflated by a producer price index. Factor utilization u is based on managerial surveys. Investment is denoted I and capital K is computed according to a perpetual inventory approach. To indicate the Swedish business cycle, each plot also includes the growth rate of aggregate value added v.

B Productivity and Demand

B.1 TFPQ

We estimate TFPQ using a cost share approach. We compute sector specific output elasticities as the ratio of the factor cost to total cost, where costs are aggregated across firms within sector j across each year t in the sample: $\gamma_{K,j} = \frac{\sum_t \sum_{J(i)} c^k_{i,t}}{\sum_t \sum_{j(i)} c^k_{i,t} + c^l_{i,t}}$ and $\gamma_{L,j} = \frac{\sum_t \sum_{J(i)} c^l_{i,t}}{\sum_t \sum_{j(i)} c^k_{i,t} + c^l_{i,t}}$. The firm-level cost of labor is given by the real personnel costs, $c^l_{i,t}$, and the user cost of capital $c^k_{i,t}$ is given by $(r_t + \delta_j - \iota_{j,t} + \Delta_{Aaa,t}) \frac{K_{i,t}}{P^s}$ where r_t is the interest rate on a ten-year Swedish govenment bond provided by the Swedish central bank (*Sveriges Riksbank*) on their website; δ_j is a sector specific depreciation rate computed based on Melander (2009); $\iota_{j,t}$ is the aggregate Swedish inflation rate for average consumer prices taken from the IMFs World Economic

Outlook Database; and $\Delta_{Aaa,t}$ is the spread between 10-year treasury and Aaa bonds from the St. Louis Fred website: *Moody's Seasoned Aaa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity*; $\frac{K_{i,t}}{P^s}$ is the real capital in terms of sector output.

We get similar results if we instead use control function approaches to estimate factor elasticities. We get results that are close to constant returns to scale and capital elasticities that we estimate range between 0.16 and 0.27 when using the Ackerberg-Caves-Fraser correction. However, results can vary meaningfully depending on how exactly the control function technique is implemented.

B.2 Demand

The finding of a non-constant elasticity of demand is robust to both specification and sample.

Piece-wise linear specification We estimate a piecewise-linear model of the following form: $\log q_{i,t} = b_1 \hat{p}_{i,t} + b_2 \mathbf{1}(\hat{p}_{i,t} > 0)\hat{p}_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,t}$. This model allows for a different elasticity when the firm's price is above as compared to below average (recall that $\hat{p}_{i,t}$ is de-meaned at the firm level). In this specification, we instrument prices using demeaned TFPQ and its interaction with an indicator for being above average. We find that β_2 is significant at the 0.1% level and that firms face an elasticity of 2 when prices are below average, and 4 when above average.

Sample Table A4 presents demand estimates for a balanced panel of firms (columns 1 and two), when excluding the Great Recession from the main sample (columns 3 and 4), and if using the SCB firm identifiers provided in the data rather than our own (columns 5 and 6). Estimates are comparable to those in the main text.

Table A4: Demand estimation, balanced sample

	Balanced Panel		Excluding GR		Alternative firm defintion	
	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$
$\hat{p}_{i,t}$	-2.573***	-2.583***	-2.955***	-2.906***	-3.237***	-3.152***
	(0.314)	(0.327)	(0.199)	(0.204)	(0.215)	(0.219)
$\hat{p}^2_{i,t}$		-4.350*		-7.207***		-5.833***
		(2.123)		(1.924)		(1.952)

This table presents demand estimates for a balanced panel ($N = 3313$, columns 1 and 2) and when excluding the Great Recession ($N = 13,117$, columns 3 and 4). $q_{i,t}$ denotes firm i's real sales in year t and $\hat{p}_{i,t}$ denotes the log of firm i's relative price in year t de-meaned at the sector-year level. All specifications are estimated using $z_{i,t}$ as an instrument, include firm and sector-year fixed effects, and use standard errors clustered at the firm level. Standard errors are shown in parentheses. Level of significance at that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively.

In Table A5, we present summary statistics for parameters θ and η estimated separately for each of our 22 sectors. The top panel shows results for θ estimated for our baseline model of demand. The bottom panel shows results for θ and η based on our non-linear demand approximation. Average and median demand elasticities are consistent with our pooled estimates. There is some heterogeneity across sectors, especially with respect to η. Moreover, the demand coefficient θ in a typical sector tends to be smaller than the mean, i.e. positive skewness.

Table A5: Demand estimation by sector, summary statistics

	mean	p50	p25	p75	skewness
θ	3.26	2.77	2.26	4.11	0.74

(a) Baseline demand

	mean	p50	p25	p75	skewness
θ	3.89	2.73	2.29	4.38	3.29
η	7.42	5.37	2.85	12.81	-0.43

(b) Non-linear demand approximation

Summary statistics for sectoral demand elasticities estimated for each sector separately. Panel (a) present results for θ estimated for our baseline model of demand. Panel (b) presents estimates for θ and η for our non-linear demand approximation.

B.3 Quality checks on estimated TFPQ and demand

We find corroborative evidence in support of structural interpretations of both of our shocks.

To check our TFPQ measure, we exploit innovation measures taken from Eurostat's Community Innovation Survey (CIS). We find that positive TFPQ growth is associated with process innovations, defined as whether the firm has introduced "new or significantly improved manufacturing methods" or "new or significantly improved supporting activities."[46] Firms that report process innovations have about 7% greater TFPQ growth relative to firms that do not report process innovations.

To corroborate our demand measure, we use a measure of "insufficient demand" reported by managers in the Industrial Capacity Utilization Survey. In this survey, managers evaluate

[46] English versions of the CIS survey are available from the Eurostat website: eurostat/web/microdata/community-innovation-survey.

whether their firm has experienced "insufficient demand" during a given period. In a regression with sector-time fixed effects, we find that this indicator is associated with 8% lower demand growth than the firm average.

C Passthrough and Variance Decomposition

C.1 Passthrough

Differential transmission of extreme shocks We find some evidence that passthrough varies for small versus large shocks: For both demand and TFPQ, passthrough is smaller for extreme shocks. This means that firms adjust their prices proportionally less in response to large shocks, with the effect being similar for positive and negative shocks. If this is the case, then passthrough will be lower in times of high dispersion, since firms are receiving more extreme shocks. While not the main focus of the paper, we do present suggestive evidence that this is true in the data, and find that our model generates the same feature.

Robustness Our passthrough results are robust to various samples and approaches.

- We find similar estimates of β_z and β_ϵ if we reproduce the analysis (including the computation of shocks) on a balanced panel.
- In the sample of single product firms, we find passthrough for TFPQ of somewhat greater than -0.2 and passthrough of slightly less than 0.2 for demand.
- Passthrough is higher if we use the two-period lags of TFPQ and demand as instruments, but not by much: $\beta_z = -0.294$, $\beta_\epsilon = 0.249$.

Time-varying passthrough Time-varying passthrough is an interesting and potentially important issue, and we find some evidence of time-varying passthrough. In panel (a) of Figure A2 we plot passthrough coefficients estimated year by year alongside the relevant variances. On the left-hand side, we show β_z together with $V_t^{s,z}$. On the right-hand side, we show β_ϵ together with $V_t^{s,\epsilon}$. What this figure makes clear is that passthrough and volatility move in opposite directions, though with somewhat different interpretations for TFPQ and demand passthrough. On the left, we see that TFPQ passthrough increases during periods when volatility of TFPQ increases. From peak to trough, TFPQ passthrough more than doubles. This means that TFPQ passthrough is countercyclical. In contrast, demand passthrough seems to fall during periods of high volatility. This implies that demand passthrough is procyclical. With respect to demand, it is also interesting to observe that demand passthrough is systematically higher by about 25% in the second half of the period (roughly 2008 and after) as compared to the first half of the period (roughly 2007 and before).

C.2 Variance Decomposition

Quantification In Table A6, we quantify elements from our variance decomposition. In the top panel, we show the change in components of the price decomposition (V^p, $V^{p,z}$, and $V^{p,\epsilon}$) and the sales decomposition (V^s, $V^{s,z}$, and $V^{s,\epsilon}$) in 2001 and 2009 as compared to the non-recession average during all other periods, i.e. 1998-2013 excluding 2001 and 2009. We present the non-recession average in the bottom row and the changes relative to this baseline in the 2001 and 2009 rows. For instance, the first columns shows V^p. The average variance of V^p (excluding recession years) is 0.0055. During 2001, V^P fell by 0.0053 to 0.000497. During 2009, V^P increased by .00688 to 0.01238.

Table A6: Cyclicality of variance decomposition components

	ΔV^p	$\Delta V^{p,z}$	$\Delta V^{p,\epsilon}$	ΔV^s	$\Delta V^{s,z}$	$\Delta V^{s,\epsilon}$
2001	$-.00053$	$-.00004$	.00043	.01258	$-.00016$	.00338
2009	.00688	.00028	.00277	.02732	.00111	.02179
average	.00550	.00051	.00267	.03341	.00203	.02100

(a) Cyclicality of variance decomposition components

	$\frac{\Delta V^{p,z}}{\Delta V^p}$	$\frac{\Delta V^{p,\epsilon}}{\Delta V^p}$	$\frac{\Delta V^{s,z}}{\Delta V^s}$	$\frac{\Delta V^{s,\epsilon}}{\Delta V^s}$
2001	.075	-.810	-.013	.269
2009	.041	.402	.041	.800

(b) Explained variance shares

Panel (a) shows the absolute change in variance decomposition components during 2001 and 2009 as compared to the average over all other periods. The change in a given component x is denoted ΔV^x, where the top row denotes the change in 2001 and the bottom row denotes the change in 2009. For comparison, the average for all periodes 1998-2013 excluding 2001 and 2009 is included in the bottom row. Panel (b) presents the share of the changes in V^p and V^s that can be attributed to z and ϵ as defined by the ratio of changes in 2001 and 2009 presented in the top panel.

Robustness We perform a number of investigations to ensure the robustness of our variance decomposition results.

- **Parameters:** The variance decomposition is based on our estimates of θ and β_z. How sensitive are our results to these values? To investigate this question, we re-compute the variance decompositions but impose various (higher) values for these parameters.

These tests do not change the basic conclusions from the main analysis unless we choose radically different values. For both θ and β_z, increases affect the patterns in a reasonable fashion. Increases in θ dampen the role of TFPQ for sales and prices, increase the role of demand for prices, and reduce the role of demand for sales. Increases in β_z result in a greater role for TFPQ shocks and a relatively smaller role for demand, though at the cost of much larger contributions from the covariance terms.

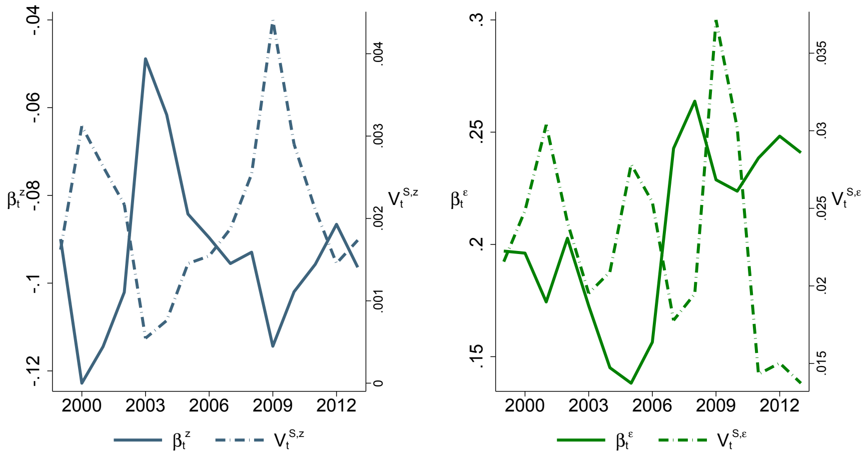
- **Sample:** The variance decomposition results based on a balanced sample are quite comparable those in the main text. Results also tend to be similar if we instead conduct the variance decomposition on a sector by sector basis. All of our largest sectors (food, wood products, metal products, and machinery and equipment), conform to the pattern of countercyclical dispersion mostly explained by demand.

- **Non-linear decomposition**: One possible concern is that the variance decomposition results are biased—or an artefact of—the assumption of (log-) linear demand. [47] Does the CES model mis-characterize how shocks transmit to prices? To investigate this possibility, we perform a non-linear variance decomposition based on our non-linear model of demand. This exercise reproduces the main patterns from our simple decompositions. Since neither the distribution of shocks nor the transmission of shocks seems sensitive to the CES assumption, we conclude that the variance decomposition exercise is robust to the linear demand assumption.

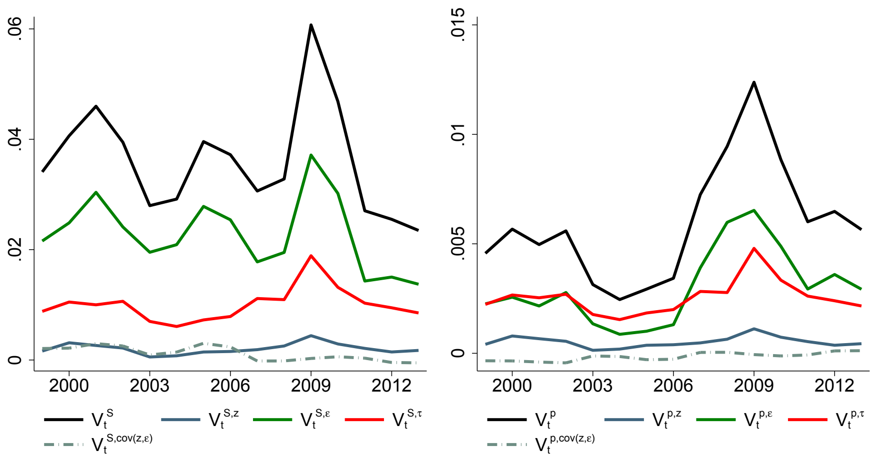
Time-varying passthrough To what extent does time-varying passthrough change our variance decomposition? To begin with, it makes the estimation simpler because year-by-year estimation forces the covariance between the shocks and τ to be zero. There thus only five components of interest. Results from a variance decomposition using passthrough estimated year by year is shown in panel (b) of Figure A2. For both sales (on the left) and prices (on the right), we find that the importance of demand is even greater than when imposing constant passthrough. The price wedge plays some role, though this is most pronounced in the great recession. TFPQ plays a minor role.

[47] Recall that we combine the passthrough equation with a linear demand curve to establish the functional relationship between sales and shocks.

Figure A2: Variance decompositions, time varying passthrough



(a) Time-varying passthrough coefficients



(b) Variance decompositions: Time-varying passthrough

Panel (a) plots the coefficients of the passthrough equation when estimated year-by-year in first differences, giving coefficients β_t^z and β_t^ϵ. Each plot also gives the variance of the shock itself. Panel (b) gives the results of the variance decomposition using these time-varying passthrough coefficients. The left panel is the sales decomposition and the right is the price decomposition, with definitions of the components given in the text.

D Quantitative Model Appendix

D.1 Further model details

Proof that x is sufficient state variable Our goal is to reduce the model to one that admits a single state variable, $x \equiv k^{\alpha} l^{1-\alpha}$. The other state will be the labour intensity of production $b \equiv l/k$, which we will prove is a redundant state variable. We can solve for the original state variables, k and l, given these two replacements, as $l = x(l/k)^{\alpha} = xb^{\alpha}$ and $k = xb^{\alpha-1}$. Output is simply $q = zx$. The law of motion for x in terms of capital and labor evolution is:

$$\dot{x} = \dot{k}\alpha k^{\alpha-1} l^{1-\alpha} + \dot{l}(1-\alpha)k^{\alpha} l^{-\alpha} = x\left(\frac{\dot{k}}{k}\alpha + \frac{\dot{l}}{l}(1-\alpha)\right) = x\left(\frac{i}{k}\alpha + \frac{h}{l}(1-\alpha) - \delta\right). \tag{17}$$

For expositional clarity we can define an investment rate for x so that we can make a simple law of motion: $i_x \equiv \dot{x} + \delta x \implies \dot{x} = i_x - \delta x$. The law of motion for b is:

$$\dot{b} = \dot{k}\frac{-l}{k^2} + \dot{l}/k = b\left(\frac{\dot{l}}{l} - \frac{\dot{k}}{k}\right) = b\left(\frac{h}{l} - \frac{i}{k}\right). \tag{18}$$

Now we can combine these to solve for investment rates in terms of i_x and $\dot{b}$: $i = i_x b^{\alpha-1} - \dot{b}(1-\alpha)xb^{\alpha-2}$ and $h = i_x b^{\alpha} + \dot{b}\alpha x b^{\alpha-1}$. Recall that investment goods cost p_k and the linear hiring / firing cost is a. Then total hiring costs are

$$p_k i + ah = i_x\left(p_k b^{\alpha-1} + ab^{\alpha}\right) + \dot{b}x\left(a\alpha b^{\alpha-1} - p_k(1-\alpha)b^{\alpha-2}\right), \tag{19}$$

where we later will add the exogenous costs of adjusting x, which we put in a function $c(i_x, x)$. For clarity, we prove the sufficiency of x for the special case without any idiosyncratic or aggregate shocks. The result extends naturally to the case with shocks. Current cashflow is $cf = pq - wl - p_k i - ah - c(i_x, x) = \pi(x, b) - i_x\left(p_k b^{\alpha-1} + ab^{\alpha}\right) - \dot{b}x\left(a\alpha b^{\alpha-1} - p_k(1-\alpha)b^{\alpha-2}\right) - c(i_x, x)$, where $\pi(x, b) = p(zx)zx - wxb^{\alpha}$ is revenue less labour cost. The original HJB with capital and labour as state variables is:

$$rv(k, l) = \max_{i,h} \hat{\pi}(k, l) - ip_k + ah - c(i_x, k^{\alpha} l^{1-\alpha}) + v_k(i - \delta k) + v_l(h - \delta l), \tag{20}$$

where $\hat{\pi}(k, l) = p(zk^{\alpha} l^{1-\alpha})zk^{\alpha} l^{1-\alpha} - wl$, and i_x is a known function of k, l, i, and h. We can equivalently define the HJB with x and b as state variables:

$$rv(x, b) = \max_{i_x, \dot{b}} \pi(x, b) - i_x\left(p_k b^{\alpha-1} + ab^{\alpha}\right) - \dot{b}x\left(a\alpha b^{\alpha-1} - p_k(1-\alpha)b^{\alpha-2}\right)$$
$$- c(i_x, x) + v_x(i_x - \delta x) + v_b \dot{b}. \tag{21}$$

Notice that value is linear in $\dot{b}$, so we know that b will jump to a given level for any x, since the optimal solution for $\dot{b}$ must be either zero (at the optimal b) or positive/negative infinity. Taking the first order condition with respect to $\dot{b}$ identifies the optimal level of b:

$$\frac{\partial}{\partial \dot{b}} = -x\left(a\alpha b^{\alpha-1} - p_k(1-\alpha)b^{\alpha-2}\right) + v_b(x,b) = 0. \tag{22}$$

This defines an optimal $b(x)$, which we can guess and verify is a constant b^*. Plug this into the FOC above to give:

$$x\left(a\alpha(b^*)^{\alpha-1} - p_k(1-\alpha)(b^*)^{\alpha-2}\right) = v_b(x,b^*). \tag{23}$$

We can see that for a constant b^* to satisfy the FOC, it must be that for any x, at b^* we can write the derivative of the value function as $v_b(x,b^*) = xf(b^*)$ for some function $f()$. Using the envelope condition for $v_b(x,b)$ and (23) gives $f(b) = -\frac{\alpha w(b)^{\alpha-1}}{r+\delta}$ and $b^* = \frac{1-\alpha}{\alpha}\frac{(r+\delta)p_k}{(r+\delta)a+w}$.

We can plug in this optimal value of b^* to restate the HJB with only one state variable. Let $v^*(x) = v(x,b^*)$. Then we can use the HJB (21) to create a HJB for $v^*(x)$:

$$rv^*(x) = \max_{i_x} \pi(x,b^*) - i_x p_x - c(i_x,x) + v^*_x(i_x - \delta x). \tag{24}$$

where $p_x \equiv p_k(b^*)^{\alpha-1} + a(b^*)^{\alpha}$ is the investment cost of x, which is just a weighted average of p_k and a. The optimal solution maximising over only x coincides with the solution of the full problem for any firm who starts with b set at the optimal level. That is, the value (14) in the text is an extension of (24) to include idiosyncratic and aggregate shocks.[48]

The same results go through with idiosyncratic shocks and demand shocks. If the factor prices p_k, r, and w are the same in all aggregate states (as is true in our results which are in partial equilibrium) then the optimal capital-labor ratio b^* is a constant. This makes the model identical to a model with a Leontief production function $q = \min\{b^*k, l\}$.

Optimal investment The solution for optimal investment is complicated by the non-convex adjustment costs. It is characterised by first order conditions within the investment and disinvestment regions, and thresholds determining when these regions are entered. The complication arises due to the adjustment cost function since $c_{i_x}(i_x,x)$ is needed for the first order conditions. This derivative takes different values in different regions, and the function

[48]For a firm who starts with $b \neq b^*$ their value is different but the policies are identical, apart from an initial instantaneous jump in b from its initial level to b^*.

is not differentiable at the boundaries:

$$c(i_x, x) = \begin{cases} \frac{\kappa}{2}\frac{(i_x-\delta x)^2}{x} & i_x > \delta x \\ 0 & \delta x \geq i_x \geq 0 \\ -\underline{\kappa} i_x + \frac{\kappa}{2}\frac{i_x^2}{x} & i_x < 0, \end{cases} \implies c_{i_x}(i_x, x) = \begin{cases} \kappa\frac{i_x-\delta x}{x} & i_x > \delta x \\ 0 & \delta x > i_x > 0 \\ -\underline{\kappa} + \kappa\frac{i_x}{x} & i_x < 0. \end{cases} \quad (25)$$

The maximization in the HJB (14) then gives the investment function as:[49]

$$\frac{i_x}{x} = \begin{cases} \delta + \frac{v_x - p_x}{\kappa} & v_x \geq p_x \\ 0 & p_x > v_x > p_x - \underline{\kappa} \\ \frac{v_x - (p_x - \underline{\kappa})}{\kappa} & p_x - \underline{\kappa} \geq v_x. \end{cases} \quad (26)$$

D.2 Numerical solution details

Numerical implementation We solve the model using continuous time numerical methods which draw heavily from Achdou et al. (2022). We discretise the state variable x with a grid of 201 nodes. We discretise the TFPQ and demand shock processes using a Rouwenhurst procedure with 7 nodes. The discretisation procedure follows Bloom et al. (2018) and picks the grid points in the low uncertainty state, and then in the high uncertainty state simply recalculates the new transition probabilities on the same grid.

Constructing real-world comparable data Certain calibration objects must be calculated on time-aggregated yearly data, constructed to be comparable to our Swedish data source. For this, we simulate a single firm for 5000 years in each uncertainty state. Since there is no permanent heterogeneity across firms in the model, this yields a distribution equivalent to simulating a large panel of firms. We construct yearly data following the data collection procedure of our datasets: capital is the capital stock at the end of the year. Labor, output, and sales are the total sum throughout the year. The yearly price is yearly sales divided by yearly quantity sold. To compute yearly demand shocks comparable to the data, we replicate a first order regression on the model. That is, for consistency with the numbers reported in the main text, we compute yearly demand shocks as if the demand curve was CES, despite the underlying demand curve in the baseline model being non-CES. This allows us to simply calibrate to the main IQRs reported in the text, and the data in the model and real word data

[49] For the knife-edge value $v_x = p_x$, the fact that firms pay no quadratic costs until $i_x = \delta x$ creates a slight complication. For this marginal value, the firm is indifferent about any investment rates between 0 and δx, and we assume as a tie breaking condition that the firm sets $i_x = \delta x$.

are treated identically. We impose a coefficient $\eta = 3$ on the data, regress to find the intercept, and then compute our demand shocks.

To compute yearly TFPQ shocks, we follow Bloom et al. (2018) and acknowledge that TFPQ is likely measured with error. We construct TFPQ as the Solow residual using our yearly data, using our estimated capital share of 25.5% from the data. Since there is no notion of utilization in our model, and we observe capital and labor directly, we do not utilization adjust the model-simulated data. Secondly, we add additional measurement error to our computed yearly TFPQ values. To calibrate the measurement error, we follow their procedure. Using our data, we estimate the autocorrelation of utilization adjusted TPFQ to be 0.908 in a standard AR(1) regression. Running the same regression instrumenting TFPQ with one year lagged TFPQ yields a coefficient of 0.967. Bloom et al. (2018) propose a measurement error model which implies that the relative standard deviation of an i.i.d. measurement error is given by $\sigma_{\text{measurement error}}/\sigma_{\log z} = (0.967/0.908 - 1)^{1/2} \simeq 30\%$. We thus add 30% measurement error to our TFPQ data (drawn from a normal distribution) before computing our IQRs and passthrough regressions. We do not add measurement error to our demand estimates since output and prices, the key input to measuring the value of the demand shock, are likely reported with much less error than underlying factor inputs or utilization.[50]

Calibration details We calibrate the model in two ways. Firstly, a "steady state" calibration holds uncertainty constant at $s = 1$ at all times. Secondly, the "uncertainty shock" calibration allows aggregate uncertainty to fluctuate between $s = 1$ and $s = 2$ as in the data. Parameters for both calibrations are provided in Table A7, where it can be seen that most parameters are identical across the two calibrations.[51]

Calibrating adjustment costs Our normalisations imply $b^* = 1$ and hence $p_x = p_k = 1$. Bloom et al. (2018) report using a resale loss of capital of 34%, a fixed cost of adjusting hours of 2.1% of annual sales, and hiring and firing costs of 1.8% of annual wages. The resale loss of adjusting x, $\underline{\kappa}$, is chosen to combine values from Bloom et al. (2018) for the resale loss from capital and the spread between hiring and firing costs. Since reducing x by one unit leads the firm to reduce k and l by $(b^*)^{\alpha-1}$ and $(b^*)^{\alpha}$ units respectively, we set

[50] Moreover, the persistence of the demand shock in an AR(1) regression does not increase when instrumenting with a lag, as it does for TFPQ, implying no measurement error using the Bloom et al. (2018) procedure.

[51] In the model with uncertainty shocks, the anticipation of the possibility of moving to the high uncertainty state affects behavior in the low uncertainty state, which leads to very small differences in the calibrated values of μ_z, μ_ϵ, $\sigma_z(1)$, and $\sigma_\epsilon(1)$ between the two calibrations. Forcing these parameters to be identical across the calibrations has no noticeable effect on the results.

$\underline{\kappa} = 0.34 p_k (b^*)^{\alpha-1} + 2 \times 0.018 w (b^*)^{\alpha} = 0.3565.$[52] We choose not to use convex adjustment costs for calibration purposes, and set them close to zero.[53]

D.3 Model without adjustment costs

Optimal static markup (no adjustment costs) Consider a simple static model of price setting. Firms face (log) TFPQ and demand shocks z and ϵ. We assume that a firm's real marginal cost, mc, is inversely proportional to its TFPQ: $\log mc = \log c - \log z$, where c is a sector-time specific constant across firms reflecting aggregate factor prices, as implied by cost minimization. We normalise $c = 1$ without loss of generality for these exercises, since it would be absorbed in a sector-time fixed effect. Firms maximize static profit, $\Pi \equiv (p - mc)q$ subject to their demand curve. Define a firm's markup over marginal cost as $\mu \equiv p/mc$. Consider the specification of demand $\log q = \frac{\theta}{\eta} \log(1 - \eta \log p) + \epsilon$. We consider a normalisation that the average log price is equal to zero, giving $\mathrm{E}[\log p] = 0$. The demand shock has zero mean: $\mathrm{E}\epsilon = 0$. The normalisation for prices is achieved by choosing an appropriate average level of TFPQ, Ez. Optimal price satisfies:

$$\frac{\partial \Pi}{\partial p} = (1 - \eta \log p)^{\frac{\theta}{\eta}} - \frac{\theta}{p} (1 - \eta \log p)^{\frac{\theta}{\eta} - 1} (p - \frac{1}{z}) = 0,$$

which can be rearranged to yield $(1 - \eta \log p) = \theta \left(1 - \frac{1}{\mu}\right)$. Recalling the definition of the markup, $\mu = pz \Rightarrow p = \mu/z$, gives

$$1 - \eta \log \mu + \eta \log z = \theta \left(1 - \frac{1}{\mu}\right). \tag{27}$$

(27) pins down the level of the optimal markup μ, which depends only on the TFPQ shock and we denote $\mu^*(z)$. Because $\epsilon = 0$ on average, and z is chosen such that $\log p = 0$ on average, we see that the "steady-state" markup is equal to the standard markup of $\mu = \frac{\theta}{\theta - 1}$, up to a Jensen's inequality correction.

Optimal static passthrough (linearized model) To investigate passthrough, we take a log linear approximation to the optimal markup equation, (27). Replace $\frac{1}{\mu}$ with the approximation $\frac{1}{\mu} \simeq 1 - \log \mu$ in (27) to give $\hat{\mu} = \frac{\eta}{\theta + \eta} \hat{z}$ where $\hat{x} \equiv \log x - \log x_{ss}$. Noting that $\hat{p} = \hat{\mu} - \hat{z}$

[52] We abstract from fixed costs of investment, which Bloom et al. (2018) additionally use, making our results relatively conservative as we exclude one form of non-convex adjustment cost.

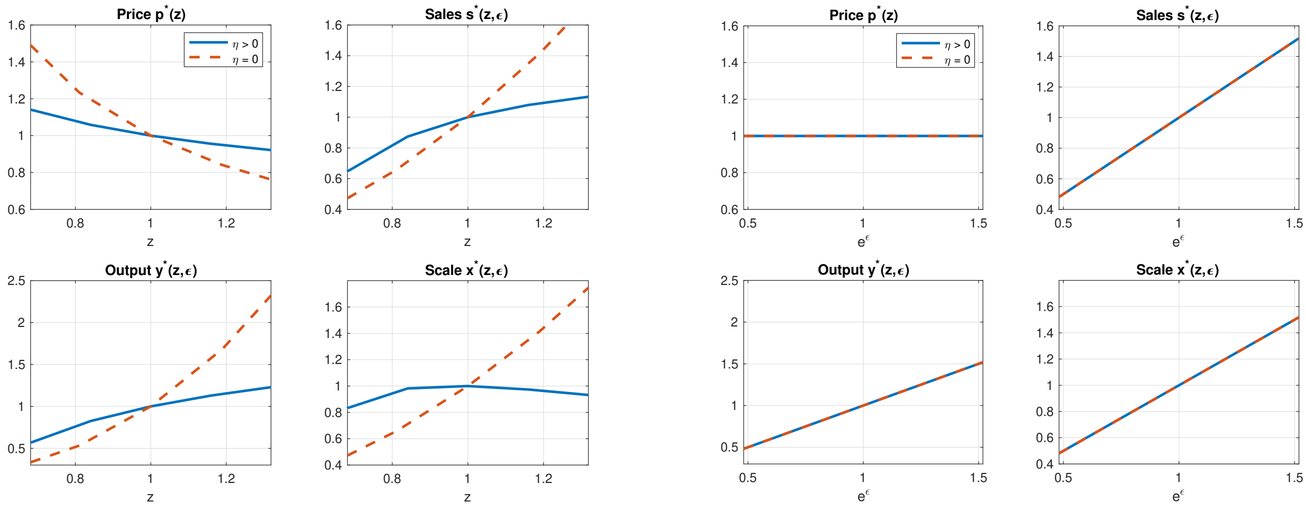
[53] Specifically, we set $\kappa = 0.0001$, and verify that further lowering the cost has no affect on the solution. Keeping a positive value of κ is helpful for the numerical solution of the model, as it implies that investment rates are finite allowing us to simply adapt the methods of Achdou et al. (2022).

gives the price passthrough equation:

$$\hat{p} = \frac{-\theta}{\theta + \eta} \hat{z} \tag{28}$$

Policy function plots and further discussion In Figure A3 we plot summaries of the policy functions in the CES and non-CES models without adjustment costs. Apart from setting $\underline{\kappa} = 0$, the calibration is exactly the calibration from the main text. $\mu^*(z)$ is implicitly defined by equation (27). We plot the true non-linear policy functions.

Figure A3: Policy functions (without adjustment costs) for $\eta = 4.3$ vs. $\eta = 0$



(a) Policy functions across z for $\epsilon = \epsilon_{ss}$ (b) Policy functions across ϵ for $z = z_{ss}$

Slices of the optimal policy functions in the model without adjustment costs. Panel (a) plots slices across TFPQ, z, holding demand, ϵ, at its average value (defined as the value in the central node of the discretized grid). Panel (b) plots slices across demand holding TFPQ at its average value. All variables normalized to one at mean.

Panel (a) plots a slice of the policy functions across levels of TFPQ, e.g. $s^*(z, \epsilon_{ss})$, for the average level of demand, denoted ϵ_{ss}. CES is given in dashed red, and non-CES in solid blue. Non-CES demand reduces the response of prices, sales, quantity, and scale to changes in TFPQ. Once adjustment costs are added, this is why firms care less about TFPQ uncertainty when demand is non-CES, since they do not plan to change capital or labor much in response to productivity anyway. In the top right panel we see that the estimated non-CES demand system makes optimal sales concave in productivity, in contrast to CES where sales are convex. This is what reverses the OHA effect, which derives from convexity. Panel (b) plots a slice of the policy functions across levels of demand, e.g. $s^*(z_{ss}, \epsilon)$, for the average level of TFPQ, z_{ss}. As expected, this form of non-CES demand leads to no difference in how demand shocks affect firm behavior relative to the CES model (when there are no adjustment costs) because demand shocks are a pure shifter and do not affect the optimal markup.

D.4 Details of other exercises

Model-based variance decomposition In our variance decompositions from Section 4 we investigated how changes in demand and TFPQ dispersion contributed to the changes in sales and price dispersion over the cycle. Here we perform a similar exercise through the lens of our non-linear model. Specifically in the bottom half of Table 4(a) we compare the IQRs of variables and shocks in the low and high uncertainty regimes.[54] As targeted in our calibration, the IQRs of TFPQ and demand changes are 31% and 61% higher respectively in the high uncertainty state. The model endogenously generates a 57% rise in the IQR of sales growth, which is comparable to the 58% rise seen in the data in the Great Recession (see Table A3). The model endogenously generates a 25% rise in the IQR of price growth, which is around 1/3 of the approximately 80% increase seen in the data.[55] In the final two rows of Table 4(a) we raise the dispersion of TFPQ and demand separately.[56] Through the lens of the model, increasing TFPQ and demand dispersion are approximately equally important for rising price dispersion, generating increases of 10% and 15% when moved independently.[57] Where the model does better is the increase in sales dispersion, and here the model attributes the bulk of the rise to demand dispersion: increasing demand dispersion alone generates a 46% rise in sales dispersion, while TFPQ dispersion alone only generates a 15% rise. This is in agreement with our semi-structural variance decomposition, which also finds demand to be the larger driver of the increase in sales dispersion.

Time-varying passthrough In Table A8(c) we compare passthrough in the model in the low and high uncertainty state. We find that demand passthrough falls when uncertainty rises, in line with the evidence on time varying passthrough we discussed in Section 5, where we found suggestive evidence that demand passthrough appears to fall in times of high dispersion. In the model, this channel operates through non-convex adjustment costs. When dispersion is high the firms that do receive demand shocks tend to receive larger shocks,

[54] We focus on IQRs, rather than variances, because the model is calibrated to match the rise in the IQR of shocks in recessions. The relative rise in variances are slightly different, most likely due to outliers in the sales distribution which raise the sales variance even in normal times. For this reason we consider the results based on IQRs to be more robust to outliers, and focus on them in the text.

[55] Part of the reason the model falls short at generating the full increase in price dispersion seen in the data is that part of the increase was driven by an increase in the dispersion of residual price changes uncorrelated with demand or TFPQ shocks (the price wedge, $\tau_{i,t}$) which is not a feature of our model.

[56] Note that since we compute the IQRs of the shocks using simulated time-aggregated data, raising each true shock dispersion does lead to small increases in the measured IQR of the other.

[57] This is because passthrough from TFPQ to prices in the model is higher than for demand, while the increase in demand dispersion is larger than the increase in TFPQ dispersion. In the data the passthrough from demand shocks to prices is higher, which explains why the model does not find demand to be more important, as in our semi-structural variance decomposition.

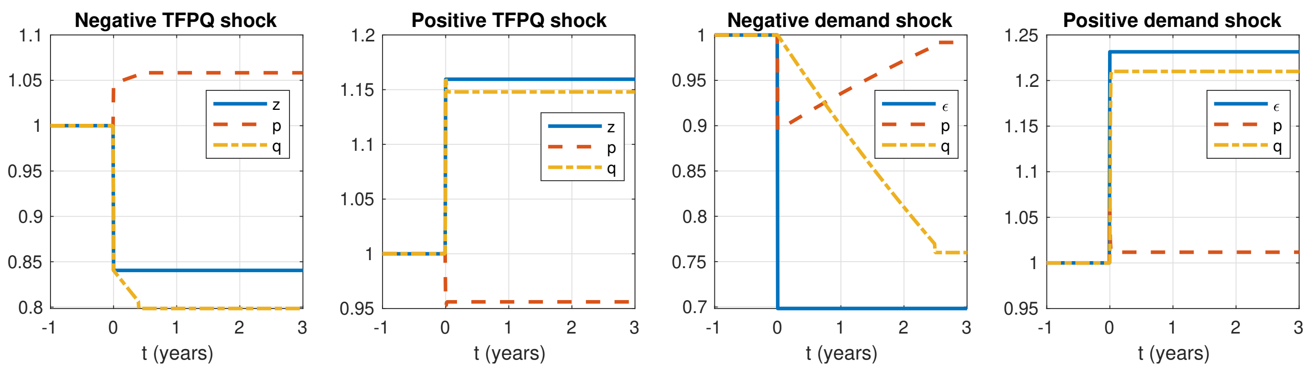
making them more likely to adjust their inputs and need to adjust their prices less.[58]

Robustness exercises In Figure A6(a) we repeat our main results for a version of the model using the estimates of Bloom et al. (2018) from the US for the persistence of the aggregate uncertainty state. The main results are quantitatively and qualitatively unchanged.

In Figure A6(b) we build a version of the model closer to Bachmann and Bayer (2013) and Mongey and Williams (2017). In this version, we suppose that adjustment costs are only paid for capital, and that labor can be freely adjusted each period. We keep the same value of adjustment costs as in the main text, but now load them all on capital adjustment. The size of the output fall in response to the uncertainty shock is slightly dampened relative to the baseline model. The size of the uncertainty effect becomes much smaller, reflecting their results. However, the volatility effect becomes much stronger, because with lower adjustment costs firms respond more to the rise in TFPQ dispersion, leading to larger output falls from the OHA effect. Our main results are thus unchanged: the uncertainty effect is still mostly driven by demand uncertainty, and TFPQ dispersion still drives a negative volatility effect.

D.5 Further model plots and tables

Figure A4: Impulse response to idiosyncratic shocks: baseline model with $s = 1$



In all panels, the firm starts having had both demand and TFPQ at their average values (central nodes) for a long time. At time 0 either productivity or demand jumps to a new value, and we plot the responses of the shock (z or ε) as well as the firm's price (p) and quantity sold (q).

[58]We also confirm this mechanism by repeating our passthrough regression which allows for different values for extreme shocks (final column of Table A8(a)) on the model data. We find that, as in the data, passthrough is lower for larger values of the shocks in our model. These two exercises suggest that the model is able to replicate the basic features of time-varying passthrough we saw in the data, despite this not being targeted.

Table A7: Parameter values and target moments

	Interpretation:	Model 1:	Model 2:	Source:
	Basic parameters:			
r	Discount rate	0.0513	0.0513	5% annual interest rate
δ	Depreciation rate	0.1054	0.1054	10% annual depreciation rate
α	Production function elasticity	0.255	0.255	Capital share of costs = 25.5%
θ	Demand elasticity	3	3	Estimated
η	Demand super-elasticity	4.3	4.3	Estimated
$\underline{\kappa}$	Scale downsizing cost	0.3565	0.3565	Bloom et al. (2018) (see text)
w	Real wage	0.4577	0.4577	Normalize aggregate $L = 1$ in low unc. state ($s = 1$)
a	Hiring costs	0	0	All adjustment costs placed on x
p_k	Capital price	1	1	Normalization
	Aggregate uncertainty process:			
$\lambda^s(1)$	Rate leave low unc. state	0	1/8	Model 1: Permanent. Model 2: Low state lasts 8 years on average
$\lambda^s(2)$	Rate leave high unc. state	—	1/1.5	Model 2: High unc. state lasts 1.5 years on average
	Idiosyncratic demand and TFPQ processes:			
μ_z	Average productivity	0.9345	0.9362	Normalize average $p = 1$ in low unc. state ($s = 1$)
μ_ϵ	Average demand	1.0057	1.0157	Normalize aggregate $K = 1$ in low unc. state ($s = 1$)
ρ_z	z autocorrelation	0.8	0.8	z yearly autocorrelation $\simeq 0.8$
ρ_ϵ	ϵ autocorrelation	0.6	0.6	ϵ yearly autocorrelation $\simeq 0.6$
λ_z	Rate new z drawn	1	1	New z drawn once per year on average
λ_ϵ	Rate new ϵ drawn	1	1	New ϵ drawn once per year on average
$\sigma_z(1)$	z std. in low unc. state	0.1264	0.1268	IQR measured yearly TFPQ growth 0.2 when $s = 1$
$\sigma_\epsilon(1)$	ϵ std. in low unc. state	0.2431	0.2465	IQR measured yearly ϵ growth 0.2 when $s = 1$
$\sigma_z(2)$	z std. in high unc. state	—	0.1751	IQR measured yearly TFPQ growth increases 30% when $s = 2$
$\sigma_\epsilon(2)$	ϵ std. in high unc. state	—	0.4689	IQR measured yearly ϵ growth increases 60% when $s = 2$

Calibrated parameter values and source moments. Model 1 refers to the model with no aggregate uncertainty shocks, where $s = 1$ at all times. Model 2 refers to the model with aggregate uncertainty shocks.

Table A8: Model performance: dispersion and passthrough

	$\Delta \log p$	$\Delta \log p$
$\Delta \log z$	-0.1792	-0.1992
$\Delta \log \varepsilon$	0.1525	0.1559
$\Delta \log z \times \mathbf{1}(\Delta \log z < x\%\text{ile})$	-0.0581	-0.0607
$\Delta \log z \times \mathbf{1}(\Delta \log z > (100 - x)\%\text{ile})$	-0.0577	-0.0444
$\Delta \log \varepsilon \times \mathbf{1}(\Delta \log \varepsilon < x\%\text{ile})$	0.0256	-0.0070
$\Delta \log \varepsilon \times \mathbf{1}(\Delta \log \varepsilon > (100 - x)\%\text{ile})$	-0.0381	-0.0433
Method:	OLS	OLS
Thresh:	5%/95%	1%/99%

(a) Passthrough in baseline: extreme values

	$\log p$	$\log p$	$\Delta \log p$
$\log z$	-0.6840	-0.8180	
$\log \varepsilon$	0.1913	0.1502	
$\Delta \log z$			-0.3307
$\Delta \log \varepsilon$			0.2635
R^2:	84%	65%	71%
Method:	OLS	IV	OLS

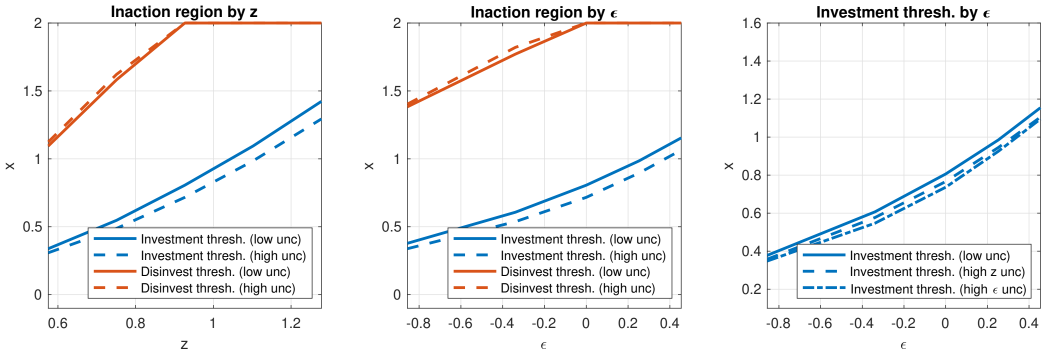
(b) Passthrough in the CES model ($\eta = 0$)

	Low σ:			High σ:		
	$\log p$	$\log p$	$\Delta \log p$	$\log p$	$\log p$	$\Delta \log p$
$\log z$	-0.3070	-0.3316		-0.2790	-0.3094	
$\log \varepsilon$	0.0899	0.0560		0.0626	0.0413	
$\Delta \log z$			-0.2083			-0.1934
$\Delta \log \varepsilon$			0.1471			0.1030
	OLS	IV	FD	OLS	IV	FD

(c) Time-varying passthrough in baseline ($\eta = 5$)

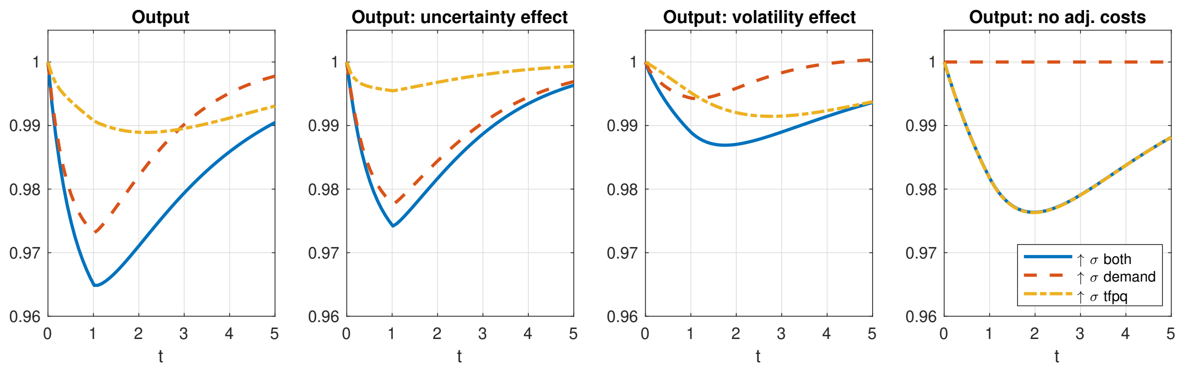
The table gives passthrough estimated on model simulated data. The data are time-aggregated to the yearly frequency. In panel (c), the first column defines extreme shocks as being in the bottom or top 5% of realised log changes, and the second column the bottom or top 1%. All coefficients are significant at at least the 0.1% level, except for the coefficient on $\Delta \log \varepsilon \times \mathbf{1}(\Delta \log \varepsilon < x\%\text{ile})$ in the right column of panel (a). The data are generated from long simulations of a single firm of 5,000 years within the low and high uncertainty states respectively.

Figure A5: CES model: Inaction regions by state and uncertainty level

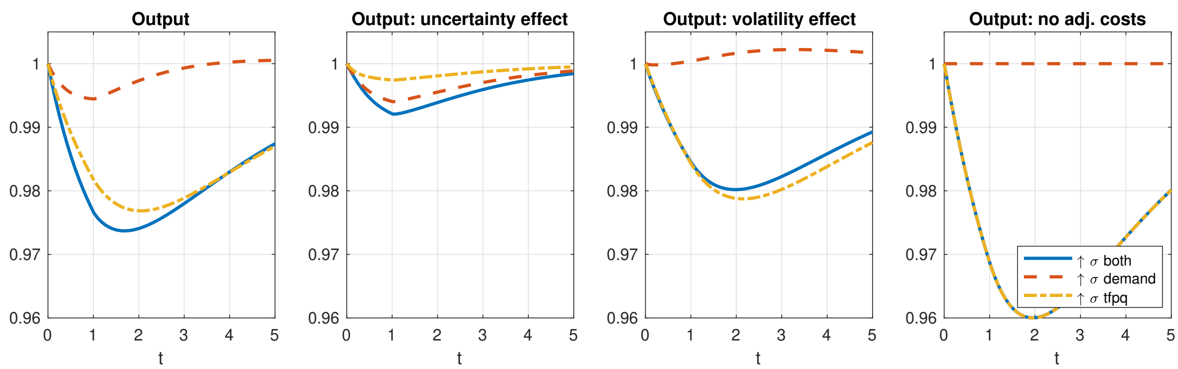


The left and centre panel give slices of the firm policy functions in the low uncertainty state (solid lines, $s = 1$) and high uncertainty state (dashed lines, $s = 2$). The right panel repeats the investment threshold (plotted across ϵ values) for the counterfactual models where only z uncertainty (dashed line) or ϵ uncertainty (dash-dotted) rise in the high uncertainty state. See Figure 7 for further details.

Figure A6: Robustness: IRFs for different model variants



(a) Robustness: Bloom et al (2018) persistence of uncertainty regime



(b) Robustness: Model with no adjustment costs on labor

The plots give the aggregate response of the model to a switch to the high uncertainty state, $s = 2$, starting from the ergodic distribution when $s = 1$. Solid blue lines give the response to increased uncertainty in both shocks, dashed red is a version where only demand uncertainty rises in state 2, and dash-dotted yellow where only TFPQ uncertainty rises. Subplots (a) and (b) give different model variants. Within each, the left panel gives output, the middle panels give the counterfactual output path from only the uncertainty and volatility effects respectively, and the right from a counterfactual model without adjustment costs.