WORKING PAPER 2023:1

Dispersion over the business cycle: passthrough, productivity, and demand?

Mikael Carlsson
Alex Clymo
Knut-Eric Joslin

IFAU INSTITUTE FOR EVALUATION OF LABOUR MARKET AND EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

DISPERSION OVER THE BUSINESS CYCLE: PASSTHROUGH, PRODUCTIVITY, AND DEMAND

Mikael Carlsson,[a] Alex Clymo,[b] Knut-Eric Joslin[c]

January 2023

Abstract

We characterize the dispersion of firm-level productivity and demand shocks over the business cycle using Swedish microdata including prices and analyse the consequences for firms and the aggregate economy. Demand dispersion increases by more than productivity dispersion in recessions. Productivity shocks pass through incompletely to prices and have limited effect on sales dispersion. Demand shocks explain most of the variation in sales dispersion. In a heterogeneous-firm model matching the micro facts, demand dispersion has unambiguously negative effects on output via increased uncertainty and 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, variable markups, business cycles, dispersion, uncertainty, passthrough, adjustment costs.

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

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

[b]University of Essex. Email: a.clymo@essex.ac.uk.

[c]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, 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) and the Handelsbankens Forskningstiftelse (grant P18-0082). 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.

Contents

1 Introduction **1**

2 Data construction and summary **6**

3 Production and demand framework **9**
3.1 Productivity . . . 10
3.2 Demand . . . 12
3.3 Quality checks on estimated TFPQ and demand . . . 16
3.4 Comparability of TFPQ and demand shock variation . . . 17

4 The dispersion of productivity and demand shocks **17**
4.1 Measuring cyclical dispersion . . . 17
4.2 Average level of productivity and demand shock dispersion . . . 18
4.3 Cyclicality of productivity and demand shock dispersion . . . 19

5 The role of shocks for endogenous outcomes **21**
5.1 Passthrough specification . . . 21
5.2 Passthrough results . . . 23
5.3 Variance decomposition of sales and price growth . . . 25
5.3.1 Variance decomposition specification . . . 26
5.3.2 Variance decomposition results . . . 27
5.3.3 Robustness and extensions . . . 31
5.4 Summary of all empirical results . . . 32

6 Quantitative Model **33**
6.1 Environment . . . 33
6.2 Production and adjustment costs . . . 34
6.3 HJB and solution . . . 35
6.4 Steady state results: can the model generate sensible passthrough and dispersion? . 36
6.5 Aggregate results: response to increase in dispersion . . . 40

7 Conclusion **48**

A Data Construction Appendix **52**
A.1 Data sources and construction . . . 52
A.1.1 Sources . . . 52
A.1.2 Firm register data . . . 52
A.1.3 Other data sources . . . 58
A.1.4 Data cleaning and assembly . . . 58
A.1.5 Samples . . . 61
A.2 Construction of key variables . . . 63
A.3 Data description . . . 65
A.3.1 Output . . . 65
A.3.2 Firm variables . . . 66

B Productivity and Demand Estimation Appendix **73**
B.1 Productivity shock estimation . . . 73
B.1.1 Estimating the production function . . . 73
B.1.2 TFPQ Cyclicality . . . 76
B.1.3 Replication of Bloom et al. (2018): Distributions during the Great Recession . 79
B.1.4 TFPQ Autocorrelation . . . 79
B.2 Demand shock estimation . . . 80
B.2.1 Demand curve estimation . . . 80
B.2.2 Demand shock cyclicality . . . 84
B.2.3 Autocorrelation of demand . . . 86
B.3 Corroboration of TFPQ and demand shocks . . . 87
B.3.1 TFPQ corroboration . . . 87
B.3.2 Demand corroboration . . . 90

C Passthrough and Variance Decomposition Appendix **91**
C.1 Passthrough . . . 91
C.2 Variance Decomposition . . . 91
C.2.1 Specification . . . 91
C.2.2 Quantification . . . 94
C.2.3 Parameter robustness . . . 94
C.2.4 Sample . . . 97
C.2.5 Time-varying passthrough . . . 99
C.2.6 Sectoral variance decompositions . . . 99

D Quantitative Model Appendix **102**
D.1 Proof that x is sufficient state variable . . . 102
D.2 Investment first order condition . . . 104
D.3 Numerical solution details . . . 105
D.4 Model without adjustment costs . . . 106
D.5 Details of other exercises . . . 110
D.6 Further model plots and tables . . . 111

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.[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 build on the substantial progress made in the literature, and delve deeper into the nature of cyclical dispersion using rich Swedish register data. 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 exactly firms react to each type of shock, and hence how different forms of dispersion contribute 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 and demand. Our first contribution is to disentangle these distinct sources of cyclical dispersion. We estimate production functions and demand curves and directly measure productivity and demand shocks at the firm level. We find that the dispersion of both physical productivity 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 under-react to changes in their productivity ("incomplete passthrough") and raise their prices in response to positive demand shocks. Using variance decompositions, we show that the increased sales dispersion observed during recessions is associated almost entirely with increased dispersion of demand. Increases in productivity dispersion are instead absorbed in markup dispersion.

Finally, we build a dynamic heterogeneous-firm model with non-convex adjustment costs in order to understand the aggregate implications of cyclical dispersion. Building on Bloom et al. (2018), we incorporate idiosyncratic shocks to both productivity and demand dispersion. A key feature of the model is that firms face non-constant elasticities of demand (in the spirit of Kimball 1995) which we estimate on our data. The model replicates well our empirical findings on shock transmission and delivers novel insights about how the economy responds to an aggregate dispersion (a.k.a. uncertainty) shock. Given our demand estimates, firms care more about demand uncertainty than productivity uncertainty. Moreover, increased productivity dispersion now decreases (rather than increases) aggregate output because markup-induced misallocation overturns the "volatility overshoot" effect described by Bloom (2009).

1. When we measure TFP in the same fashion as Bloom et al. (2018), we reproduce their results. To illustrate this, we reproduce Figures 1 and 2 from their paper in Table 16 in the Appendix.

2. The rise in dispersion of shocks is systematically associated with higher uncertainty, a concept of renewed importance as the COVID crisis has raised many uncertainty indicators to their highest levels on record. Altig et al. (2020) show that all of their considered economic uncertainty indicators rise during the COVID pandemic in the US and UK, many of which to their highest recorded levels. Barrero et al. (2021) find a sharp rise in cross-firm equity return dispersion during the pandemic.

Empirical Contribution The empirical components of the project are based on register data from Swedish manufacturing firms for the years 1998-2013. Sweden experienced two recessions during this period. 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. Consistent with the existing literature, we find that both episodes were accompanied by a rise in the dispersion of revenue productivity growth. The recessions were also associated with increases in domestic measures of uncertainty, such as the EPU index (Armelius, Hull, and Köhler 2017). Going beyond the existing literature, however, we use our data to discriminate between firm-level developments in physical productivity and firm-level developments in demand. We are able to construct a convincing measure of physical productivity because we have data on prices and capacity utilization. In turn, our measure of physical productivity enables us to construct a convincing measure of demand shocks. Building on the approach from Foster, Haltiwanger, and Syverson (2008), we use our measure of physical productivity to estimate demand. Because changes in physical productivity affect marginal cost, productivity innovations can be used as an instrumental variable for price. We show that both physical productivity dispersion and demand dispersion increased during the Great Recession. The interquartile range of productivity growth was 35% higher in 2009 than in non-recession years, while demand growth was 56% higher. As part of our estimation exercises, we also document significant deviations from the benchmark log-linear constant elasticity of substitution (CES) demand specification. As we discuss below, a non-constant elasticity creates different roles for each type of shock by introducing a "real ridigity" in pricing decisions.

Can our productivity and demand measures be given structural interpretations? To validate our shocks, we provide additional corroboration using other data sources. With respect to productivity, we show that positive productivity growth is related to process innovations in manufacturing reported by firms in Eurostat's Community Innovation Survey. Our productivity measure is thus associated with the type of innovations that we expect to improve production efficiency (rather than, say, new product innovations). With respect to demand, we find that managers' reports of "insufficient demand" in the Business Cycle Statistics survey conducted by Statistics Sweden are systematically associated with reductions in our demand measure. Our demand measure thus seems to reflect the level of demand actually experienced by the firm.

We use our productivity and demand shocks to understand firm behavior. We conduct a series of empirical exercises. The first is a log-linear passthrough estimation. The purpose of the passthrough exercise is to describe how firms adjust their prices in response to TFPQ and demand shocks. The approach is akin to that used by De Loecker et al. (2016). With respect to productivity, we find only a moderate effect on prices. The coefficient on TFPQ is at most 0.3 regardless of how we estimate passthrough and typically smaller. This means that a 1% improvement in TFPQ lowers prices by less than 0.3%. With respect to demand, we find that firms raise their prices between 0.2% to 0.3% in response to a 1% increase in demand. Both of these findings are in contrast to the predictions of a benchmark model. In particular, a monopolistically competitive firm with constant returns to scale (CRS) production and (log-linear) CES demand would set a constant markup over marginal cost if prices and inputs can be freely adjusted. In contrast to what we find, this benchmark model implies complete passthrough of TFPQ to prices and no passthrough from demand to prices. An important motivation for our theoretical model is thus to account for the passthrough results. As we discuss below, our theoretical model is able to account for these findings in a parsimonious fashion.

Our second empirical exercise comprises a set of variance decompositions. We use these decompositions to investigate the relative importance of the two shocks for business cycle cyclicality. We combine our estimated passthrough equations with a log-linear approximation of the demand curve. This provides a semi-structural link between shocks and sales, and describes how the variance of sales will respond to changes in the variances of productivity and demand. The main result from this exercise is that demand shocks are the main driver of sales growth dispersion, both on average and over the business cycle, in a range of specifications. Demand shocks explain about half to two-thirds of sales growth dispersion on average, and increased dispersion of demand shocks during the Great Recession explains 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. Through the lens of a demand curve, the low sensitivity of prices to TFPQ shocks means that it is difficult for changes in TFPQ dispersion to contribute to sales dispersion. The attenuated effect on prices means that even large changes in TFPQ will have limited effect on demand. Instead, TFPQ dispersion appears to be absorbed in markup 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.

Theoretical Contribution The last component of the paper is a quantitative theoretical model. We use the model not just to study how firms respond to dispersion, but to understand how firms are affected by the uncertainty associated with dispersion. Like Bloom (2009), we include non-convex adjustment costs in the model. This creates "wait and see" behavior in response to increases in uncertainty. A novelty of model is that we also incorporate non-CES demand curves into this framework. These demand curves create separate roles for demand and productivity uncertainty. Specifically, we employ a demand framework that builds on Gopinath, Itskhoki, and Rigobon (2010). This model allows the elasticity to differ along the demand curve in the manner of a Kimball (1995) aggregator: Firms lose more customers when they raise their price than they gain when they lower their price, which attenuates the benefit of adjusting prices in response to changes in marginal costs. This unresponsiveness is a type of "real rigidity" (Ball and Romer 1990; Klenow and Willis 2016).

We first validate our model by showing that it is able to rationalize our main empirical findings, despite these being untargeted. The model predicts low passthrough from TFPQ to prices (21-33%) and and non-zero passthrough from demand shocks to prices (6-15%). The incomplete passthrough from TFPQ to prices follows mainly from our estimated demand curves, which deviate significantly enough from CES to explain why firms choose to adjust their prices so little in response to changes in productivity. The model generates passthrough from demand shocks to prices due to the presence of input adjustment costs. Intuitively, firms in their inaction regions do not change their production in response to a small demand shock, and instead must change their price in order to convince customers to continue purchasing their existing quantity. In both cases, the model predictions are close to what we observe in the data. Moreover, the model generates sensible dispersion in sales and prices relative to the dispersion of underlying shocks.

Since our model generates sensible predictions for firm-level behavior, we then use it to understand the aggregate implications of dispersion and uncertainty. We model aggregate uncertainty shocks as an increase in the dispersion of firm-level TFPQ and demand shocks consistent with increases seen in our data. Overall, we find that uncertainty shocks produce large declines in aggregate output, with demand driving more of the decline than TFPQ. Notably, however, firms care more about demand uncertainty than TFPQ uncertainty. In our set-up, demand uncertainty is a far more important driver of wait-and-see behavior than productivity uncertainty. When demand is sufficiently non-CES, firms do not adjust their production much in response to TFPQ shocks. Demand shocks directly shift the amount firms can sell and consequently their desired investment. Demand shocks are therefore powerful drivers of wait-and-see behavior. In contrast, TFPQ uncertainty does not worry firms much because firms do not anticipate large irreversible investments in response to the future realisations of these shocks. Instead firms absorb these shocks in markups.

Finally, we find that non-CES demand overturns the "volatility overshoot" effect described by Bloom (2009). This is the finding that in standard models, despite the fact that uncertainty (meaning the knowledge that shocks are becoming more dispersed) reduces aggregate output, realised dispersion itself later increases aggregate output. The volatility overshoot effect arises when firms' optimal sales are convex in their underlying productivity shocks. In contrast, we show that our estimated demand curve overturns this result. For our estimated parameters, the increase in *realised* TFPQ dispersion leads to a fall in output, due to an increase in markup dispersion and hence misallocation. Put differently, non-CES demand makes firms' decisions concave in TFPQ, and an increase in dispersion leads firms with negative TFPQ shocks to lose more sales than firms with positive TFPQ shocks gain. Our analysis thus nuances existing work on uncertainty shocks (Bloom 2009; Bachmann and Bayer 2013; Bloom et al. 2018) by revealing distinct roles for demand and TFPQ dispersion: Demand dispersion hurts output via uncertainty, while productivity dispersion hurts output via realised dispersion.[3]

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. With respect to shocks, existing research has demonstrated that the dispersion of TFPR is countercyclical (Bachmann and Bayer 2013; Kehrig 2015; Bloom et al. 2018). Our contribution is that 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. Countercyclicality also holds for most outcomes. Davis, Haltiwanger, and Schuh (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.[4] These patterns tend to hold in our data (we provide an illustration in the next section). In particular, we characterize the cyclicality of price and sales dispersion, and show how they are

3. In contrast, realised demand shock dispersion does not strongly affect aggregate output in our framework, since demand shocks linearly affect sales in the absence of adjustment costs. 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.

4. A related literature studies the effect of aggregate uncertainty (meaning uncertainty about aggregate, rather than idiosyncratic, shocks) on the economy. For recent contributions see, for example, Basu and Bundick (2017), Berger, Dew-Becker, and Giglio (2019), and Den Haan, Freund, and Rendahl (2021).

linked to the dispersion of underlying shocks.

Second, we contribute to a literature that estimates how firms respond to idiosyncratic shocks. As firm-level price data has become available to researchers, a number of recent papers have, like us, investigated the separate roles of demand and TFPQ shocks. Foster, Haltiwanger, and Syverson (2016) estimate firm-level demand and productivity and show that demand contributes more to firm growth over the lifecycle. Hottman, Redding, and Weinstein (2016) and Eslava and Haltiwanger (2020) estimate both shocks and additionally perform variance decompositions to investigate their role in explaining differences in sales growth across firms. Relative to these papers, we abstract from issues of multi-product firms, but additionally investigate non-CES demand and show that it can rationalise our estimates of incomplete passthrough. We apply our variance decompositions over the business cycle, while these papers investigate the firm cross section and lifecycle. De Loecker et al. (2016), Pozzi and Schivardi (2016), and Haltiwanger, Kulick, and Syverson (2018) estimate demand and TFPQ shocks, and also estimate passthrough from shocks to prices.

Related to our work, Kaas and Kimasa (2021) estimate productivity and demand shocks indirectly through the lens of a novel customer-capital model featuring both product and labor market frictions, and argue that both shocks are important for capturing the joint dynamics of prices, employment, and productivity across firms. They subject their model to an increase in demand and productivity uncertainty, and argue that increased demand uncertainty is a plausible feature of recessions, since it, unlike increases in productivity uncertainty, decreases aggregate output in their estimated model.

Carlsson, Messina, and Skans (2016) use a panel-VAR analysis to analyse how firms respond to permanent idiosyncratic demand and TFPQ shocks, focusing on the response of employment via hiring and firing flows. We find, as stressed by many of the above papers, that demand appears to play a larger role in driving firm behaviour than productivity. We show that this also applies to shock dispersion over the business cycle, highlighting the need to understand the underlying sources of firm-level demand shocks over the business cycle.[5]

Our main focus is on demand and productivity shocks. We find that dispersion in these objects accounts for the majority of the idiosyncratic variation as measured by our variance decomposition. Nevertheless, the residual "price wedge" in our variance decomposition plays an important role. This suggests that other factors are also important. One possibility is financial constraints. For example, Gilchrist et al. (2017) show in a dataset similar to ours that firms who were financially constrained during the financial crisis raised prices. They thus identify a shock that would be subsumed in the residual price wedge in our variance decomposition. Consistent with the literature, we also find evidence for time-varying passthrough responsiveness to shocks. Berger and Vavra (2019) argue that this can explain the correlation between price dispersion and exchange-rate passthrough.[6] We additionally present suggestive evidence of time-varying passthrough in our data and (endogenously) in our model. At longer horizons Decker et al. (2020) show that declining business dynamism in the US is driven by declining responsiveness (i.e. passthrough) to shocks, rather than declining volatility of idiosyncratic shocks.

5. There is a large literature estimating cost passthrough. For additional literature, see the references in Haltiwanger, Kulick, and Syverson (2018), footnote 2.

6. Additionally, Berger and Vavra (2019) build a menu-cost model which features non-CES demand curves. They calibrate the deviation from CES to match average passthrough in their data, while we estimate our demand curves directly and show they endogenously generate a sensible level of passthrough. We do not model price stickiness, but instead focus on adjustment costs in factor inputs and how they interact with non-CES demand.

Third, we relate to papers estimating or using non-CES demand curves, contributing with a novel application—how non-CES changes the aggregate effects of cyclical dispersion—and by directly estimating demand curves. Non-CES demand curves have found important applications in many fields of economics due to their ability to generate incomplete passthrough or heterogeneous markups.[7] 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.

Papers estimating non-CES demand curves directly using firm-level price and quantity data are rare. Closely related to our method is Haltiwanger, Kulick, and Syverson (2018), who estimate a Hyperbolic Absolute Risk Aversion demand curve using US firm-level data. They find significant deviations from CES, as we do. Our approaches differ in that we use fixed effects to hold the average elasticity of demand constant for all firms, while they allow for permanent differences in the elasticity across firms. In a trade context, Arkolakis et al. (2018) use aggregate trade data instrumented with tariffs to estimate demand curves at the 10-digit product level, and also find statistically significant deviations from CES demand.

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 heterogeneous-firm literature is vast (early work includes Hopenhayn 1992; Khan and Thomas 2008) and includes many of the papers referenced in the paragraphs above. Our contribution is to build a heterogeneous-firm model combining non-convex adjustment costs with an estimated non-CES demand curve and to study firm behavior and the role 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 set up our estimation and measurement framework. In Section 4 we estimate TFPQ and demand shocks and discuss their cyclicalities. In Section 5 we estimate passthrough and perform our variance decomposition, and finally in Section 6 build our structural dynamic model. 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. We construct our key variables using (1) bookkeeping data from financial statements, (2) 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. This survey is harmonized with the EU *Structural Business Statistics* and, in principle, covers the universe of Swedish industrial firms. Our sample is thus based on manufacturing firms engaged in the production of goods.[8] Our product-level price and quantity data are retrieved from Statistics Sweden's *Industrins Varuproduktion* (IVP) survey. The raw data from the IVP survey are reported at the 8-digit

7. Some recent examples include the following: Félix and Maggi (2021) show that non-CES demand is needed to explain the rise in employment among (especially the most productive) incumbent firms in response to a reform which lowered entry costs in Portugal. Berger and Vavra (2019) show that non-CES demand with a time-varying elasticity is needed to explain the positive correlation between price dispersion and passthrough in US data. In representative agent models, see Lindé and Trabandt (2018) for a recent contribution using non-CES demand to match the slope of the Phillips Curve. See the references in Arkolakis et al. (2018) for non-CES contributions in the trade literature.

8. These are firms classified by the European Union's NACE system as manufacturing (Section C), sectors 10-33 in the Swedish Industrial Classification (SNI).

product level according to the Combined Nomenclature (CN). Overall, our IVP data includes observations from about 10,000 unique firms. Our utilization data are taken from *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. Additional details about each survey are provided in Appendix A.1.

Our main sample is based on firm-year observations for which we have complete data for investment, prices, and capacity utilization. In total, this 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 quite skewed. While the median firm has 107 employees, the average firm has 278 employees. Descriptive statistics are presented in Table 8 in the Appendix.

We provide robustness results based on other samples including for a balanced panel for the period 1999-2010.

Our basic firm variables are sales s, a firm-level price P^f, number of employees l, intermediate goods m, degree of factor utilization u, and capital k. In brief, s is given by firm turnover deflated by a sectoral price index; firm-level price P^f is given by a price index computed based on product-level data; 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; and capital k is computed according to a perpetual inventory approach. We provide further discussion of our price and utilization data in conjunction with our discussion of TFPQ in the next section. Further discussion of the data, variable construction, and samples can be found in Appendices A.1, A.2, and A.3.

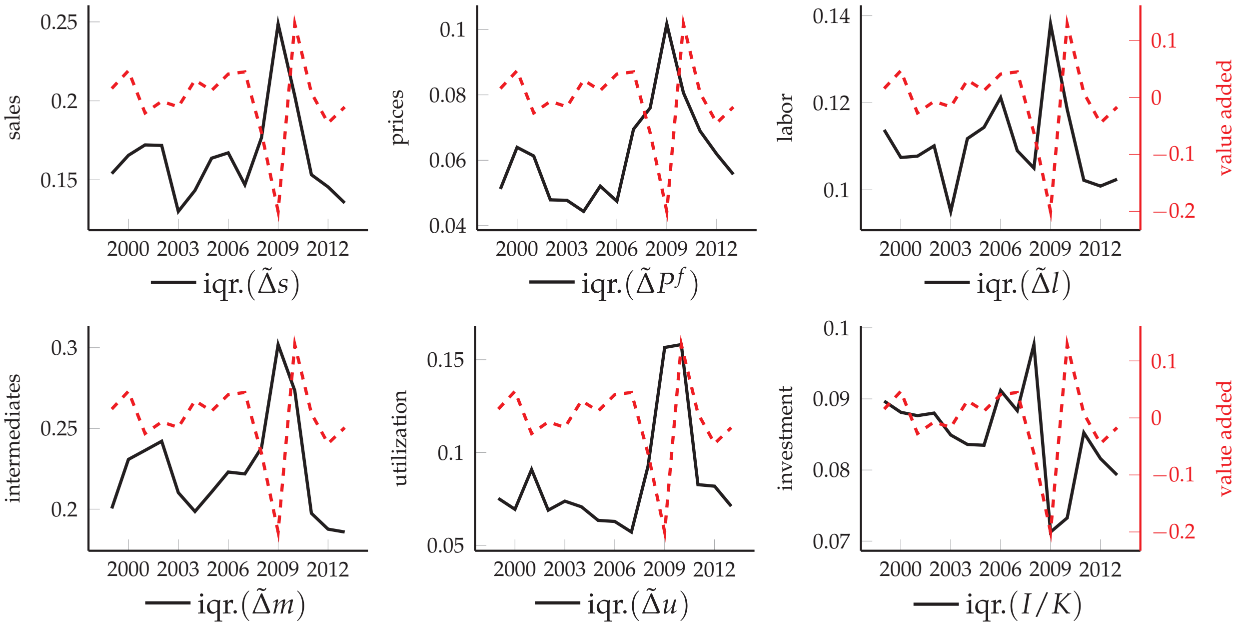
Cyclicality of firm variables Most firm choice variables exhibit countercyclical dispersion. This holds for sales, labor, and use of raw materials, as well as for prices and capacity utilization. There is a dramatic rise in dispersion during the Great Recession and a smaller, though still meaningful, rise in dispersion around 2001. The only variable that clearly deviates from the pattern of countercyclical dispersion is investment. In our data, investment is pro-cyclical.

The cyclicality of dispersion is illustrated for key variables in Figure 1. For each variable, we compute the within-sector-year dispersion of firm level growth as measured by the interquartile range in each year: We demean firm growth (measured as log-changes) by the average growth in the relevant sector during the relevant year. We denote this transformation by $\tilde{\Delta}$. We then compute our dispersion measure based on the transformed data.[9] For investment, we compute an analogous within-sector dispersion but based on the investment to capital ratio rather than based on a growth rate. For all variables, the black lines thus reflect firm-level cyclicality divorced from across sector effects. We maintain the focus on firm-level dispersion throughout the paper. All dispersion measures are demeaned at the sector-year level unless otherwise noted. To emphasize

9. Specifically, to calculate the within-sector moments we first demean each growth rate by subtracting the sector-year mean growth rate to define $\tilde{\Delta} x_{i,t} \equiv \Delta x_{i,t} - \bar{\Delta} x_{j,t}$, where $\bar{\Delta} x_{j,t}$ is the mean growth rate at time t across all firms in sector j. Taking the standard deviation or interquartile range across all firms in a given year gives us our yearly within-sector dispersion measures, $\text{std}_t(\tilde{\Delta} x_{i,t})$ and $\text{iqr}_t(\tilde{\Delta} x_{i,t})$. We can use identical procedures for any other chosen moments. For clarity, when we refer to across firm moments, we are referring to moments of their log changes, unless otherwise stated.

the cyclicality, we present each dispersion measures alongside 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).[10]

Figure 1: Dispersion of firm variables, 1999-2013



Each plot shows the interquartile range (IQR) across firms of log changes for key variables, calculated each year. The IQRs are computed within sector as $\text{iqr}_t(\tilde{\Delta} x_{i,t})$ for each variable $x \in \left\{s, P^f, l, m, u, k\right\}$. Sales s is given by firm turnover deflated by a sectoral producer price index. Price P^f is the 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 sectoral producer price index. Factor utilization u is based on managerial surveys. And capital k is computed according to a perpetual inventory approach. Complete descriptions of each variable are provided in the appendix. To indicate the Swedish business cycle, each plot also includes the average growth rate of value added v deflated by a sectoral producer price index.

The average level of dispersion is substantial. This highlights the importance of idiosyncratic firm shocks (rather than aggregate or sectoral shocks) in driving firm-level outcomes. The interquartile range of $\tilde{\Delta} s$ is 0.17 on average, and annual growth rates as high or low as $\pm 10\%$ are common. The dispersion of price changes $\tilde{\Delta} P^f$ is relatively smaller but still substantial, with an average IQR of yearly price growth of around 0.07.

Dispersion is, moreover, strongly counter-cyclical. To quantify this, we compare the level of dispersion in 2001 and 2009 to the average level of dispersion in all other years.[11] In 2009, the iqr of $\tilde{\Delta} s$ is 58% larger than the average, the iqr of $\tilde{\Delta} P^f$ is 79% larger than the average, the iqr of $\tilde{\Delta} l$ is 26% larger than the average, the iqr of $\tilde{\Delta} m$ is 39% larger than the average, and the iqr of $\tilde{\Delta} u$ is 99% larger than the average.[12] In 2001, the same comparison to the average shows increases of 9% for $\tilde{\Delta} P^f$, 8% for $\tilde{\Delta} s$, 8% for $\tilde{\Delta} m$, and 16% for $\tilde{\Delta} u$. The one exception (besides investment) is labor for which the iqr of growth is 2% smaller than the average in 2001.

10. Equivalent plots based on the standard deviation are presented in the Appendix in Figure 14. The patterns are the same.

11. These comparisons and others are reported in Table 11 in the Appendix.

12. The dispersion of the investment to capital ratio, in contrast, is 18% smaller than the average.

These findings support the general conclusion of increased dispersion in recessions. Given that this dispersion is within-sector, it raises the question of what drives the rise in dispersion: Is it a rise in dispersion of demand shocks, or TFP shocks? Or something else entirely, such as frictional "wedges" or changes in how firms respond to shocks? We address these questions in the remainder of the paper.

3 Production and demand framework

The key empirical challenge that we tackle is to distinguish changes in quantities that arise due to developments in real productivity from changes that arise due to fluctuations in demand. We are able to make progress on this challenge because of our access to high-quality firm register data. Using data on prices and capacity utilization, we construct a measure of firm-level real productivity. We then use our productivity measure to estimate demand, employing an approach similar to that used by Foster, Haltiwanger, and Syverson (2008).

Access to price data enables us to measure developments in physical productivity over time. We construct a firm-level price index and use it to deflate nominal production. This approach is analogous to that suggested by Smeets and Warzynski (2013). We compute real value added, $v_{i,t}$, by deflating our firm-level value-added measure $V_{i,t}$ by our firm-level price index $P_{i,t}$. This yields a measure of production quantity, which we then use to construct both productivity and demand shocks.

Our firm-level price is constructed as a chained Laspeyres index based on goods-level price and sales data.[13] The use of a price index is necessary because most firms produce multiple goods; the chained specification is necessary because firms adjust their product portfolio from year to year. One disadvantage of the chained approach is that comparisons at distant time horizons may degrade as the set of products changes. Another issue associated with chained price indices is the possibility of compounding errors that arise when mis-measurement in a given link propagates to subsequent observations. We view the first problem as limited because most firms are observed for a limited number of years and maintain a stable set of core products. To address the second issue, we perform targeted cleaning of the price data, and rebase firm price indices whenever we encounter an extreme value.

We are careful to never directly compare prices indices across firms. Such comparisons are problematic because of possible differences in product definition or product quality. Instead, we focus on within firm effects and rely on first differences, fixed effects, and normalisations to soak up permanent differences in prices across firms. Our results are thus based on relative price changes within the firm over time.[14] While this approach is straightforward in principle, the use of the firm identifiers provided in the data is occasionally problematic. The reason is that some firms undergo changes that categorically change the scale or nature of their productive activities. Some of these changes are observable in our data. For example, the firm may open or close production facilities, or be involved with mergers or acquisitions. We therefore define our "firm" panel identifiers in a conservative fashion. Specifically, we assign new firm identifiers if the number of plants within a firm changes, if there is an extreme change in the level of one or

13. Additional details and description of our price data are provided in Appendices A.1 and A.3.

14. A potential source of mismeasurement that we are unable to control for are changes in quality. Nevertheless, we posit that year to year changes in quality are sufficiently small as to not significantly bias our estimated year-to-year shocks.

more variables, or if there is a one or more year gap in the observation of the firm.

3.1 Productivity

Production function specification Our aim is to measure quantity productivity (TFPQ) at the firm level. Our 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:

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

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, $v_{i,t}$; the right-hand side is "true" usage of capital and labour ($u_{i,t}k_{i,t}$ and $u_{i,t}l_{i,t}$).

The key object of interest is $z^u_{i,t}$. $z^u_{i,t}$ is intended to measure physical productivity (TFPQ). For this interpretation to be legitimate, both inputs and outputs must be properly measured.[15] To measure output correctly requires a real measure of production. We therefore create our output variable $v_{i,t}$ using a firm specific price. To measure input usage appropriately requires a measure of factor usage, $u_{i,t}$. We therefore use a measure of capacity utilization reported by managers in the Business Cycle Statistics for Industry survey. In this survey, Managers are asked to assess their degree of capacity utilization relative to intended production intensity, expressed as a percentage (it was possible to report utilization in excess of "100%", see Appendix A.1 for details).

The utilization adjustment is critical in this study. If we do not correct for utilization, 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. In the absence of a utilization adjustment, this will look like a negative TFPQ shock. We substantiate this effect in our data. We perform a regression of firm-level utilization on an indicator for "insufficient demand" (also from the Business Cycle Statistics survey), an interaction term for the Great Recession, and sector-time and firm fixed effects. We find that firms reporting insufficient demand report 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 (see Table 5 in the Data Appendix for details of this regression).

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 $\gamma_{K,j} + \gamma_{L,j} = 1$.[16] CRS thus seems plausible and justifies the

15. The intended interpretation of $z^u_{i,t}$ also requires that the Cobb-Douglas specification provide a reasonable approximation of the "true" production function. Like most of the literature, we assume that this is the case, at least for small changes in labor and capital.

16. Other studies have also concluded that CRS is a reasonable assumption, in particular those that attempt to account for factor utilization (Basu 1996; Cette et al. 2015; Shapiro 1993).

estimation of $\gamma_{K,j}$ and $\gamma_{,j}$ based on each factor's share in total costs. Control function results are presented in Appendix B.1.

We measure real labour costs $c^l_{i,t}$ as total payments to labour C^L deflated by P^f. 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 and i_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 (our so-called "full" sample).

Our cost shares are consistent with the literature. We get an average cost share for labour of 0.735, with variation across industries.[17] The highest labour cost shares is 0.89 in sector 26 (electronics). The lowest labour cost share is 0.307 in sector 19 (petroleum). Cost shares for each of our 22 sectors are presented in Appendix B.1.

We use the sector-specific elasticities together with firm-level measures of value-added, labor, capital, and factor utilization to compute our productivity residuals. The labor variable is taken directly from our bookkeeping data which reports employment in terms of average full-time worker equivalents. We construct the capital variable using a perpetual inventory method. The perpetual inventory value is often 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. Our value added measure is based on the economic definition: Turnover plus the change in the stock of partially finished goods ($D_{i,t}$), minus the use of raw materials and consumables ($M_{i,t}$): $V_{i,t} = S_{i,t} + D_{i,t} - M_{i,t}$.[18] We then deflate by the firm price index to yield real value added. As discussed above, our firm-level utilization measure is based on a business cycle survey in which managers assess various aspects of the business environment and how it has affected production.[19]

Robustness and extensions Besides $z^u_{i,t}$, we also estimate two other TFPQ measures. The first is "raw" TFPQ, which we denote by $z_{i,t}$. $z_{i,t}$ is analogous to $z^u_{i,t}$ but is not adjusted for utilization. This TFPQ measure has been used in other studies that have access to price data. If production is constant returns to scale, then there is simple relationship between $z_{i,t}$ and $z^u_{i,t}$. $z^u_{i,t}$ is raw TFPQ scaled by the reported utilization rate: $z^u_{i,t} = z_{i,t} - \log u_{i,t}$. In other words, $z_{i,t}$ is equivalent to $z^u_{i,t}$ if there is full utilization ($u_{i,t} = 1$).

We also produce a measure of TFPQ using an alternative utilization adjustment. Although our basic utilization adjustment is conceptually simple and empirically transparent, other approaches to capacity utilization are possible. We therefore produce an alternative utilization adjusted TFPQ

17. Production function estimation based on control function approaches produces similar estimates when implemented on a balanced panel, and extremely similar estimates when we adjust for utilization.

18. We maintain the convention that nominal variables are expressed in uppercase and real in lowercase letters, where possible.

19. Variable definitions and variable construction are discussed in Appendices A.1 and A.2. TFP defintions and construction is discussed in Appendix B.1.

measure which allows a flexible relationship between utilization and production and also takes advantage of additional business cycle information. In this "projection approach" we regress raw TFPQ onto a fourth-order polynomial of the utilization and firm and year fixed effects. Details can be found in Appendix B.1. We denote the productivity measure based on this approach by $z^{u,p}_{i,t}$. In effect, this specification takes managers' utilization reports less literally and instead uses them as a potentially biased measure of true input utilization.[20] In general, results based on $z^{u,p}_{i,t}$ are close to those based on $z^{u}_{i,t}$.[21] For reasons of transparency and parsimony, we therefore focus on the simple linear utilization adjustment in our main exposition. Results based on the flexible utilization adjustment are instead presented in the appendix.

For comparison with the literature and as a baseline for evaluating our results, we also produce a measure of revenue productivity, so-called "TFPR." We denote this measure by $a_{i,t}$. We measure TFPR using the production function $V_{i,t} = e^{a_{i,t}} k_{i,t}^{\gamma_{K,j}} l_{i,t}^{\gamma_{L,j}}$ and continue to estimate $\gamma_{K,j}$ and $\gamma_{L,j}$ using the cost share approach. Many studies of dispersion have focused on this measure of shocks. TFPR and TFPQ are linked by the relationship $a_{i,t} = z_{i,t} + \log p_{i,t}$. Hence, TFPR can be thought of as a measure of underlying TFPQ confounded by the price at which a firm sells its goods.[22]

Overall, we work with three measures of TFPQ: Raw TFPQ $z_{i,t}$ and two utilization-adjusted measures $z^{u}_{i,t}$ and $z^{u,p}_{i,t}$. We also have a measure of TFPR, $a_{i,t}$ which can be utilization adjusted in a similar fashion as the TFPQ measures.

3.2 Demand

Demand function specification Our baseline model of demand is the constant elasticity specification (CES). 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. Specifically, we adapt the demand curve proposed by Gopinath, Itskhoki, and Rigobon (2010)—henceforth GIR—to our setting. We extend the GIR model to handle demand shocks and show how it can estimated.

The GIR demand curve 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.

Like GIR, we specify a model of demand that is parameterized by θ and η, and that (potentially) features a non-constant elasticity of demand:

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

20. For example, it could be that managers are more likely to notice or report larger changes in utilization relative to smaller changes.

21. Results based on $z^{u,p}_{i,t}$ tend to be intermediate between the results for $z_{i,t}$ and $z^{u}_{i,t}$. This holds for demand estimates and dispersion measures etc.

22. Measuring TFPR using cost shares and assuming constant returns to scale is the approach taken by Bloom et al. (2018). Another approach is to explicitly suppose a CES demand curve for a firm's goods, and derive a production function for revenue in which the elasticities are scaled by the elasticity of demand.

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 shock.

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.[23] $\hat{p}_{i,t}$ 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.[24]

The elasticity of demand in specification (2) is given by $\tilde{\theta}(\hat{p}) = \frac{\theta}{1-\eta\hat{p}}$. Similar to GIR, $\eta > 0$ 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 $\hat{\varepsilon}(\hat{p}) \equiv \partial \log \tilde{\theta}(\hat{p}) / \partial \log p = \frac{\eta}{1-\eta\hat{p}}$.[25] 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 in model (2). It forms the basis for our main analyses in conjunction with our productivity measures. $\epsilon_{i,t}$ describes the idiosyncratic level of demand for firm i at time t. In the context of our demand curve, $\epsilon_{i,t}$ captures changes in a firm's ability to sell holding their price constant. We thus identify a positive (negative) demand shock if a firm sells more (less) in a given year without a corresponding reduction (increase) in price. Although un-modelled, demand shocks reflect changes in the size of a firm's customer base or in customers' ability or willingness to pay.

In addition to model (2), we also work with a baseline CES model. The CES 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 version of the demand curve imposes that the elasticity of demand is constant at θ at all times.

Demand shifter estimation The estimation of demand curves is a classic econometric challenge. Because prices and quantities are jointly determined in market equilibrium, a regression of quantities on prices will seldom identify the true relationship. To overcome this challenge, we employ the approach of Foster, Haltiwanger, and Syverson (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 unre-

23. The baseline GIR demand curve has that low price firms will always face a lower demand elasticity than high price firms. While this is an important idea with much empirical support, we are instead interested in how the elasticity changes *within firm* when they change their price, and abstract from permanent differences in elasticities.

24. This demeaning has no effect on the estimates in the CES special 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 results are robust to simply demeaning the firms price using the firm's own average price.

25. For these calculations, recall that $\hat{p} = \log p - c$ for some firm- and time-specific constant c.

lated 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 TFPQ as an instrument for price. This is exactly the same procedure as used by Foster, Haltiwanger, and Syverson (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

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

The coefficients in the regression $\log q_{i,t} = b_1\hat{p}_{i,t} + b_2\hat{p}_{i,t}^2 + \hat{\alpha}_i + \hat{\mu}_{j,t} + \hat{\epsilon}_{i,t}^*$ 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 CES model which is (log) linear. We estimate (2) using demeaned log TFPQ and its square as instruments for the relative price and relative price squared.[26] Note that we denote shocks from the second-order approximation by ϵ^* to differentiate them those estimated from the CES-model.

We present results based on using raw TFPQ ($z_{i,t}$) as an instrument and based on using utilization-adjusted TFPQ ($z_{i,t}^u$) as an instrument. In theory, $z_{i,t}^u$ should produce better estimates because it is a superior measure of real productivity.

The specification presented in this section is pooled for the whole economy, yielding a single θ estimate which can be interpreted as the average demand elasticity across all sectors. In the appendix, we provide demand coefficients estimated sector by sector. In general, the sectoral estimates are close to the pooled estimate. We favor the pooled estimate as it eases the presentation of the variance decompositions in section 5.[27] Recall that each firm is still assigned its sector-specific reference price, and we use sector-year fixed effects to control for sector-specific changes in demand.

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 use raw TFPQ $z_{i,t}$ as an instrument, while columns 2 and 4 use utilization adjusted TFPQ ($z_{i,t}^u$) as an instrument. Comparison of of column 1 and column 2, and of column 3 and 4 thus shows the effect of our utilization adjustment on our demand estimates.

Results for the CES model yield a demand elasticity of about 4 when using raw TFPQ as an insturment (column 1) and about 3 when using utilization adjusted TFPQ as an instrument (column 2). These estimates are in line with existing estimates based on Swedish data: Carlsson, Messina, and Nordström Skans (2021) report elasticities around 3, in agreement with Heyman, Svaleryd, and Vlachos (2013).[28] Notably, using raw TFPQ as an instrument appears to bias the demand elasticities upwards. This makes sence. If TFPQ is calculated without correcting for utilization, a positive demand shock raises utilization and hence measured raw TFPQ. This breaks the independence assumption in the IV estimation, since the instrument (TFPQ) becomes correlated with the error term (the demand shock). The utilization adjustment thus plays a potentially important role.

26. Haltiwanger, Kulick, and Syverson (2018) estimate an approximation to their HARA demand curve using a similar IV approach.

27. We also estimate variance decompositions for our largest sectors and find similar results.

28. Our estimates also lie in the range of elasticities for the US estimated in Foster, Haltiwanger, and Syverson (2016), who report elasticities between 0.68 and 3.29.

Table 1: Demand estimation results

	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$
$\hat{p}_{i,t}$	-3.94***	-2.99***	-3.86***	-2.94***	-2.98***	-2.07***
	(0.24)	(0.20)	(0.22)	(0.20)	(0.25)	(0.25)
$\hat{p}^2_{i,t}$			-6.59***	-6.28***		
			(1.72)	(1.60)		
$\mathbf{1}(\hat{p}_{i,t} > 0)\hat{p}_{i,t}$					-1.84***	-1.79***
					(0.45)	(0.41)
Implied Structural Parameters						
θ	3.94***	2.99***	3.86***	2.94***		
	(0.24)	(0.20)	(0.22)	(0.20)		
η			3.42***	4.27***		
			(0.814)	(1.038)		
iv	$z_{i,t}$	$z^u_{i,t}$	$z_{i,t}$	$z^u_{i,t}$	$z_{i,t}$	$z^u_{i,t}$

Table 1 gives demand estimates based on 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. $\mathbf{1}(\hat{p}_{i,t} > 0)\hat{p}_{i,t}$ denotes the interaction between an indicator variable for an above average price and $\hat{p}_{i,t}$. 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 that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively. The first two columns give results from the basic CES demand curve estimation, model 3. Columns three and four present results for our non-linear approximation, model 4. The last two columns present results from a piece-wise linear specification. The difference between each pair of regressions is the choice of instrumental variable. Columns 1, 3, and 5 use the "raw" TFPQ measure $z_{i,t}$. Columns 2, 4, and 6 use instead the utilization adjusted TFPQ measure $z^u_{i,t}$. We indicate the choice of instrumental variable in the bottom row ("iv").

Results for our general demand specification are shown in columns three and four. The second order term is statistically significant at the 0.1% level in both cases. 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? For $\theta = 2.94$ and $\eta = 4.27$, 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.[29] If a firm's elasticity rises when it raises its price, it gains little revenue from raising its price because the firm loses many customers. Conversely, if a firm's elasticity falls when it lowers its price, the revenue gains from lowering prices are also small because the firm gains few customers. In such a world, firms find it optimal to change their prices little even when shocks change their marginal costs. As we show in our theoretical work, this is an economically meaningful departure from the CES assumption which can explain much of the incomplete passthrough which we will

29. 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).

later find in our data.

We thus estimate two measures of firm-level demand shocks: One from our general demand curve, denoted $\epsilon^*_{i,t}$, and one from the CES model, denoted $\epsilon_{i,t}$. At the aggregate level, the dispersion and cyclicality of the two series are similar, and many results are robust to the choice of specification. Although the non-linear model better describes the data, both series of shocks are important in the continuation. The CES model provides comparison to the literature and is the benchmark to which we compare the non-linear demand results. The CES model also provides the intuitive basis of the price passthrough exercises in Sections 5.1 and 5.2.

Robustness and extensions The finding that demand exhibits non-constant elasticity is an integral part of our analysis. One worry is that the results are driven by functional form. We therefore reproduce our results using an alternative specification. Rather than using a squared term to capture the non-linearity, we instead estimate a piecewise-linear model:

$$\log q_{i,t} = b_1 \hat{p}_{i,t} + b_2 \mathbf{1}(\hat{p}_{i,t} > 0)\hat{p}_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,t}. \tag{5}$$

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 demeaned at the firm level). In this specification, we instrument prices using demeaned TFPQ and its interaction with an indicator for being above average. The results from model (5) are presented in the final two columns of Table 1. We find that β_2 is significantly different from zero at the 0.1% level. This indicates that firms face different demand elasticities when their prices are below versus above average. Using z^u as an instrument, we find an elasticity of 2 when prices are below average, and 4 when above average.

Another concern is that the results are driven by the choice of sample. One may wonder if the results are driven by the Great Recession or are perhaps affected by entry and exit of firms. To investigate these possibilities we re-estimate our model on various subsamples. We find that results are quantitatively similar if we exclude the Great Recession (years 2008 and 2009) or if we instead use a balanced panel. These results are presented in Appendix B.2.

Finally, the results do not seem to be driven by outliers, are robust to further winsorizing the price data, and are similar if we restrict the analysis to only single plant firms.

3.3 Quality checks on estimated TFPQ and demand

Do the shocks that we measure match our intended interpretations? Do positive productivity shocks in fact reveal improved ability to produce? Do negative demand shocks genuinely reflect decreased ability to sell at a given price? A benefit of our register data is that we can "sense check" our estimated TFPQ and demand values using information from other firm surveys. Overall, we find corroborative evidence in support of structural interpretations of both of our shocks.

To corroborate our productivity shocks, we use microdata from the Swedish implementation of Eurostat's Community Innovation Survey (CIS). We find that positive TFPQ growth is associated with process innovations reported in the CIS. Firms are asked in the CIS survey whether they have introduced "new or significantly improved manufacturing methods" and if they have introduced "new or significantly improved supporting activities."[30] We use these variables as an indicator for process innovations in a regression of TFPQ growth. We find that firms that report

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

process innovations experience about an 8% increase in TFPQ relative to firms that do not report process innovations. Notably, we do not find a relationship between TFPQ growth and product innovations. This supports the interpretation of our TFPQ shocks as changes in manufacturing or production ability. The main drawback of this analysis is that results are based on a small sample. The CIS survey covers a limited number of firms, and the regressions are based on about 500 observations in total. This forces us to deviate from using sector-time fixed effects and instead use sector and year fixed effects. Complete results and additional discussion are presented in the Appendix B.3.

What about our demand shocks? To corroborate our demand shocks, we use an "insufficient demand" indicator available in the Business Cycle Statistics for Industry microdata. In this survey, firms that report less than 100% capacity utilization are asked to provide a reason (reasons for low capacity utilization include disruptions related to lack of workers, difficulty obtaining raw materials, and insufficient demand). Consistent with expectations, we show that our estimated demand shocks correlate strongly with a firm's self-reported "insufficient demand" indicator in a sector-time fixed effects regression. If a firm reports insufficient demand, this predicts a demand shock which is 8% lower than the firm's average. Hence, our demand shocks seem to capture changes in demand that are consistent with those perceived by managers. For additional details, see Appendix B.3.

3.4 Comparability of TFPQ and demand shock variation

Our demand measures are computed from a regression using firm and sector-time fixed effects. In contrast, our TFP measures are computed from the production function, and are not purged of fixed effects. In order to be comparable with our demand measures, we therefore remove the across-firm and sector-time variation from our TFP measures. This is either done explicitly within regressions, or implicitly by using our transformation $\tilde{\Delta}$ to take the first difference of TFP and then subtract the sector-time average. This approach consistent with our focus on idiosyncratic variation.

4 The dispersion of productivity and demand shocks

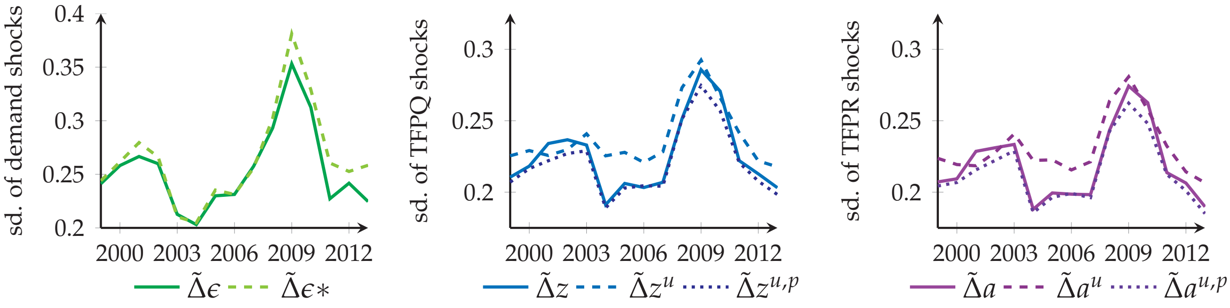
In this section, we present results related to the dispersion of firm-level shocks. Not only do TFPQ and demand dispersion both exhibit substantial volatility on average, but both series are also countercyclical. We elaborate on both findings below. Results from this section form the foundation for the remainder of the paper. We use the dispersion measures from this section in our variance decomposition exercises and to calibrate our dynamic model.

4.1 Measuring cyclical dispersion

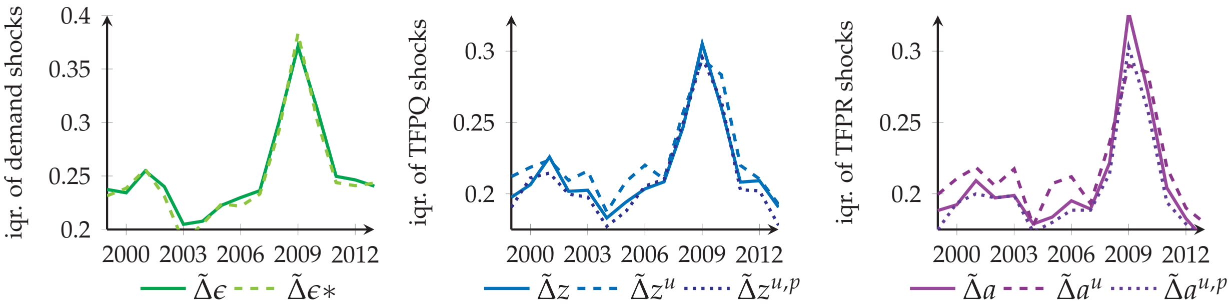
We measure shocks to TFPQ and demand as log changes relative to the previous year. To construct our dispersion measures, we demean by the relevant average sector-year growth and then compute the standard deviation and interquartile range. Dispersion in thus measured in terms of growth rates and reflects genuine heterogeneity across firms divorced from aggregate or sectoral volatility.

Our main results are summarized in Figure 2. This figure illustrates the within-sector interquartile range (top row) and standard deviation (bottom row) of each shock over time. The left panels show demand, the middle panels present TFPQ, and the right panels plot TFPR. To faciliate comparison with TFPQ, the three TFPR measures have been computed in an analogous fashion as $z^u_{i,t}$, $z_{i,t}$ and $z^{u,2}_{i,t}$, but based on sales rather than production quantities.

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



(a) Standard deviations



(b) Interquartile range

Each panel of this figure shows times series for the dispersion of demand, TFPQ, or TFPR. The figures in panel (a) plot the standard deviation across firms of log changes, calculated each year. The standard deviations are computed within sector as $\text{sd}_t(\tilde{\Delta}x_{i,t})$ for each variable x as explained in the text. Panel (b) is constructed in a similar fashion but using the interquartile range as the dispersion measure. The left panels plot demand shocks ($\tilde{\Delta}\epsilon$ and $\tilde{\Delta}\epsilon^*$), the center panels TFPQ shocks ($\tilde{\Delta}z$, $\tilde{\Delta}z^u$, and $\tilde{\Delta}z^{u,p}$), and the right panels TFPR shocks ($\tilde{\Delta}a$ and utilization adjusted measures $\tilde{\Delta}a^u$ and $\tilde{\Delta}a^{u,p}$ are constructed in a corresponding fashion as $\tilde{\Delta}z^u$ and $\tilde{\Delta}z^{u,p}$).

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 utilization adjusted 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 utilization adjusted TFPQ and 0.238 for demand.

With respect to our productivity measures, TFPQ and TFPR exhibit similar dispersion and are highly correlated. If we compare the middle and right panels in Figure 2, we see that the average level of dispersion is similar. The correlations between analogous series—such as $\tilde{\Delta} a_{i,t}$ and $\tilde{\Delta} z_{i,t}$—are about 0.9.[31] Perhaps surprisingly, the average level of dispersion is also robust to the measurement error associated with utilization. Although the dispersion of utilization adjusted productivity (z^u, solid line) tends to be greater than the dispersion of raw productivity (z, dashed line), these differences are small.

With respect to demand, ϵ and ϵ^* are highly correlated and exhibit similar average levels of dispersion. At the firm level, these measures have a correlation greater than 0.95. Importantly, demand volatility is at least as large as that seen for productivity—and often larger. This holds regardless of whether we measure dispersion based on the standard deviation or the interquartile range, and is not affected by the choice of demand measure. Since demand shocks are of a similar magnitude to the dispersion in TFPQ growth, this highlights the potential for demand to drive dispersion in outcomes across firms.

4.3 Cyclicality of productivity and demand shock dispersion

Productivity and demand are characterized by counter-cyclical volatility. Table 2 presents percentage changes in dispersion for the two recessions relative to the average computed across all other years. 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: Cyclicality of dispersion

	shocks								outcomes			
	$\tilde{\Delta} z$		$\tilde{\Delta} z^u$		$\tilde{\Delta} \epsilon$		$\tilde{\Delta} \epsilon^*$		$\tilde{\Delta} s$		$\tilde{\Delta} p$	
	sd	iqr	sd	iqr	sd	iqr	sd	iqr	sd	iqr	sd	iqr
2001	5.8	9.1	-3.9	3.0	8.0	7.2	9.5	9.8	17.0	9.2	-4.0	5.1
2009	29.2	47.5	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 1-8) shows the results for shocks (z and ε) while the right side of the table (columns 9-12) 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.

31. This result accords with the findings of Blackwood et al. (2021). They indirectly measure TFPQ from revenue data using a markup-based approach and also find high correlation between (cost share) TFPR and (revenue function) TFPR and similar variance. Since we directly measure TFPQ, we can decompose TFPR into TFPQ and price (using $a_{i,t} = z_{i,t} + \log p_{i,t}$) and directly confirm that the reason for the similarity between TFPR and TFPQ is that price changes are in fact relatively unresponsive to changes in TFPQ. This is something we explore further in the remainder of our paper in our work on passthrough.

A central contribution of this paper is the quantification of demand shock dispersion. Table 2 shows that demand dispersion is more cyclical than productivity dispersion, especially after adjusting productivity for utilization. For example, during the Great Recession, the IQR of $\tilde{\Delta}\epsilon$ was 56% above average, whereas the IQR of $\tilde{\Delta}z^u$ 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.[32]

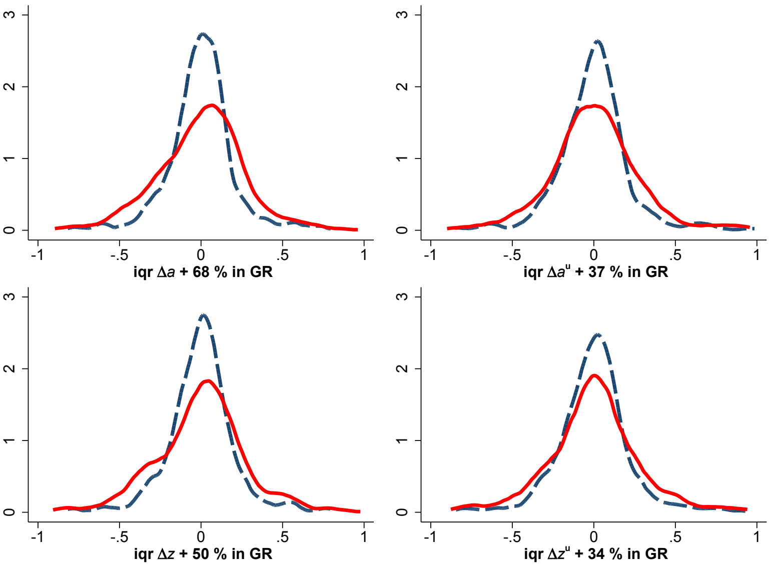
The role of utilization and prices for measurement cyclical productivity dispersion The dispersion in the various productivity measures is comparable on average. Nevertheless, we see in Table 2 that the utilization adjustment is quantitatively important for measuring the cyclicality of productivity dispersion. Inspecting the rise in the interquartile range of TFPQ in 2009 relative to non-recession years, we find that raw TFPQ dispersion rose by 47.5%, while the utilization adjusted measure rose by only 35.5%. Our main utilization adjustment thus reduces the rise in TFPQ dispersion by about a quarter.[33] This follows from the large rise in the dispersion of utilization across firms seen in Figure 1: Failing to account for the larger changes in utilization, which are correlated with a firm's change in sales, overstates the increase in the dispersion of productivity.

In most contexts, price and capacity utilization are not widely available. We therefore conclude this section by showing how price and utilization corrections affect measured dispersion. In Figure 3, we plot the journey from TFPR on the top left to utilization-adjusted TFPQ on the bottom right. In each panel the dashed blue line gives the distribution across firms in 2006, and the solid red line the distribution in 2009. The IQR of raw TFPR (top left panel) rises by 68% between these two years, which is much larger than the 34% rise in utilization-adjusted TFPQ (bottom right panel). What explains this difference? Is it that TFPQ adjusts TFPR for prices, or that we also adjusted for utilization? To evaluate the role of prices, we present raw TFPQ in the bottom left panel. Raw TFPQ adjusts TFPR for prices without adjusting for utilization. We see that prices reduce dispersion somewhat, from 68% to 50%. To evaluate the role of utilization, we present a measure of TFPR adjusted for utilization in the upper left. This has a dramatic effect of measured dispersion: When we correct for utilization, TFPR dispersion increases by only 37%. This is only slightly larger than the 34% dispersion increase for true utilization adjusted TFPQ. Together, these pictures suggest that correcting for utilization is more important than correcting for prices if one hopes to properly measure the dispersion of productivity shocks.

32. Despite rising by less, the fact that TFPQ dispersion rises in recessions when well-measured (accounting for both price and utilization changes) is, to the best of our knowledge, also novel. Of course, the remaining rise in TFPQ dispersion could be spurious due to measurement error or incorrectly specifying the production function. For example, Blackwood et al. (2021) investigate non-Cobb Douglas production functions as one possible source of misclassification.

33. The effect of the utilization adjustment is still present, but smaller, when looking instead at the standard deviations as the cyclical measure. In our empirical characterization of shocks, we focus on the IQR as our main measure as it is less likely to be affected by outliers.

Figure 3: From TFPR to utilization-adjusted TFPQ



This figure illustrates the role of prices and utilization adjustment for measuring productivity shocks. Each panel in the figure shows smoothed kernel densities for log changes for a firm-level productivity measure. The changes are computed within sector. The dashed black lines give the change measured in 2006 (i.e. between 2005 and 2006) and the red line the change measured in 2009. In the top left panel, we show the distribution of TFPR. In the bottom left panel, we show raw TFPQ. In the top right panel, we show utilization adjusted TFPR. And in the bottom right panel, we show utilization adjusted TFPQ. The bottom left and the top right panels thus show intermediate steps between TFPR and utilization-adjusted TFPQ.

5 The role of shocks for endogenous outcomes

In this section, we assess the role of shocks for driving sales and prices over the business cycle. We present two main exercises. In the first exercise, we estimate a log-linear passthrough equation. The goal of this exercise is to establish how shocks transmit to prices. The passthrough equation describes how firms adjust their prices in response to TFPQ and demand innovations. We find that the effects on prices are relatively moderate. With respect to productivity, we find that a 1% improvement in TFPQ lowers prices by less than 0.3%. With respect to demand, we find that a 1% increase in demand causes firms to increase their prices between 0.2% and 0.3%.

The second exercise is a "semi-structural" variance decomposition. The purpose of this second exercise is to describe how cyclical dispersion in sales and prices can be statistically attributed to dispersion in TFPQ and demand. This exercise relies on our demand and passthrough results, and can be given a structural interpretation. In terms of results, we find that demand plays a significant role in driving countercyclical volatility, while TFPQ is less important.

5.1 Passthrough specification

The passthrough equation that we estimate is similar to that in De Loecker et al. (2016):

$$\log p_{i,t} = \beta_z z^u_{i,t} + \beta_\epsilon \epsilon_{i,t} + \alpha_i + \mu_{j,t} + \tau_{i,t}. \tag{6}$$

Equation (6) 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.[34] 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, (6) can be estimated by OLS because there is no endogeneity problem.

The passthrough equation is consistent with a benchmark static pricing model in which prices are set as a constant markup over marginal cost. Consider a firm that faces constant demand elasticity θ and that produces using a constant returns to scale technology with TFPQ $z^u_{i,t}$. If the firm can adjust its price and use of inputs costlessly, and is a price taker in input markets, then the firm should set prices as a constant markup over marginal cost: $p_{i,t} = (\theta/(\theta-1))mc_{i,t}$, where $mc_{i,t} = c_{j,t}/z^u_{i,t}$ for some (possibly sector specific) weighted input price $c_{j,t}$. In logs this gives $\log p_{i,t} = -\log z^u_{i,t} + \log c_{j,t} + \log(\theta/(\theta-1))$. If this model is correct, then it can be estimated by (6). Notice in addition that the benchmark model implies $\beta_z = -1$ and $\beta_\epsilon = 0$, i.e. "complete" passthrough of TFPQ shocks and no passthrough of demand shocks. TFPQ shocks directly affect prices via the impact on marginal cost, while demand shocks only affect the quantity of sales, not the price. Since we can measure firms' prices directly, and consistent with this framework, in this paper we take a simple measure of markups using prices and TFPQ. Specifically, changes in $\log p_{i,t} + \log z^u_{i,t}$ give changes in markups over TFPQ, so that a firm with complete passthrough $\beta_z = -1$ would have a constant markup.[35]

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. If increased dispersion in $\tau_{i,t}$ drives most of the increased dispersion in endogenous outcomes, this provides evidence against simple models in which firm-level dispersion in sales and prices is directly driven by dispersion in shocks. It would instead suggest that cyclical dispersion in outcomes is generated by cyclical distortions to firm behavior.[36] The passthrough equation thus allows us to investigate both whether demand or TFPQ is a more important driver of cyclical dispersion, and whether shocks or distortions play a larger role.

34. Relative to De Loecker et al. (2016), we add the demand shock as a potential driver of price setting in our passthrough equation.

35. We assume constant returns to scale in production and define a simple measure of markups relative to TFPQ. An alternative approach additionally estimates marginal costs (accounting for returns to scale and fixed factors of production) and measures markups relative to estimated marginal cost. See for example De Loecker et al. (2016). In our approach, since firms' price indices may have different base years, $\log p_{i,t} + \log z^u_{i,t}$ does not capture the level of markups for a given firm, but still captures the within-sector change in markups over time. Marginal costs may change due to changes in TFPQ, $z^u_{i,t}$, or factor prices, $c_{j,t}$. If all firms in the same industry pay the same factor prices (or their factor prices are different but satisfy $c_{i,t} = \alpha_i c^s_{j,t}$ for some constant α_i and sector specific $c^s_{i,t}$) then within-sector changes in $\log p_{i,t} + \log z^u_{i,t}$ capture within-sector changes in the markup distribution, consistent with the first-differenced within-sector analysis in this paper.

36. 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.

5.2 Passthrough results

We present our passthrough results in Table 3. The first three columns of Table 3 give our baseline estimates of β_z and β_ϵ. The remaining columns present extensions and robustness exercises.[37] In the first column, we show estimates from an OLS fixed-effects estimation of (6). In the second column, we estimate the same fixed-effects model but include lagged values of TFPQ and demand as instruments. The inclusion of lagged values as instruments can help to correct for measurement error by relying on persistent changes. This is also the approach favored by De Loecker et al. (2016). In the third column, we estimate (6) in first differences (i.e. one year) rather than levels, again using OLS.

Across all three specifications, two main results emerge: Firms do not completely pass through TFPQ shocks to prices, and firms do pass through demand shocks to prices. The passthrough from TFPQ to prices is -0.124 in the OLS specification and only -0.097 in the first difference specification (columns 1 and 3). This implies that a firm only lowers its price by about 1% in response to a 10% reduction in costs. In contrast, we find passthrough of -0.24 when using the IV approach (column 2). This is substantially higher—albeit still very far from the benchmark of complete passthrough. Since the IV approach focuses on persistent changes in TFPQ, 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. The passthrough from demand to prices is more consistent across specifications, ranging between 0.209 and 0.235 (columns 1 to 3). This means that a demand shock of 10% leads to 2% higher prices. In other words, firms increase their prices in response to an increased ability to sell at a given price.

Our main results are inconsistent with the simple static pricing model. The finding of incomplete TFPQ passthrough ($\beta_z > -1$) suggests that firms allow their markup to rise rather than adjusting their price in proportion to the reduction in costs. It also appears that firms raise their price and markup when they receive positive demand shocks. These findings are important for understanding how demand and productivity shocks translate into firm behaviour. A meaningful contribution of the theoretical model that we present in the next section is that we are able to rationalize these findings.

Our passthrough 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, Kulick, and Syverson (2018). Our finding of positive passthrough from demand shocks to prices is also comparable with estimates in Pozzi and Schivardi (2016) and Haltiwanger, Kulick, and Syverson (2018)—though we find higher demand passthrough overall.

37. In Appendix C.1 we present the same table but estimated instead on a balanced panel. Results are comparable.

Table 3: Passthrough estimates

	$\ln p_{i,t}$	$\ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\ln p_{i,t}$	$\Delta \ln p_{i,t}$
$z^u_{i,t}$	-0.124***	-0.240***			-0.294***	
	(0.006)	(0.024)			(0.071)	
$\epsilon_{i,t}$	0.227***	0.235***			0.249***	
	(0.005)	(0.009)			(0.015)	
$\Delta z^u_{i,t}$			-0.0965***	-0.103***		-0.119***
			(0.004)	(0.005)		(0.004)
$\Delta \epsilon_{i,t}$			0.209***	0.225***		0.221***
			(0.005)	(0.005)		(0.003)
$\mathbf{1}(\Delta z^u_{i,t} < 5\%)$						0.0327***
						(0.005)
$\mathbf{1}(\Delta z^u_{i,t} > 95\%)$						0.0384***
						(0.005)
$\mathbf{1}(\Delta \epsilon^u_{i,t} < 5\%)$						-0.0326***
						(0.005)
$\mathbf{1}(\Delta \epsilon^u_{i,t} > 95\%)$						-0.00908
						(0.005)
N	15042	10132	11108	8873	7466	11108
iv	no	L.z L.ϵ	no	no	L2.z L2.ϵ	no
sample	all	all	all	$\lvert\Delta \ln P^f\rvert > 0.01$	all	all

The tables presents passthrough results estimated on our main sample. $p_{i,t}$ denotes firm i's relative price in year t, while $z^u_{i,t}$ and $\epsilon_{i,t}$ denote firms i's TFPQ and demand in year t. First differences are indicated by Δ. The terms of form $\mathbf{1}(\Delta x < 0.5\%)$ denote interactions between a shock x and an indicator for being in either in the lowest 5% og greatest 95% of the shock distribution. 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. Level of significance at the 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars. The first and second columns show results for the estimation of the passthrough equation 6 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 6 estimated in first differences. Column 4 is the same model as column 3, but excludes observations for which nominal price changes are less than 1% ($|\Delta \ln P^f| < 1\%$). Column 5 presents the same results as column 2, but instead using the two-year lag of the shocks as instruments. Column 8 repeats the first difference regression but allows for different coefficients for extreme large and small changes in TFPQ and demand. Below the estimation results, the bottom panel presents information on the number of observations (N), the use of instrumental variables (iv), and whether the sample excludes small price changes (sample).

Robustness and extensions Incomplete TFPQ passthrough can be explained by sticky prices but also by "real rigidities." In the case of sticky prices, firms would like to adjust their prices but have limited scope for doing so, as is the case in Calvo, Rotemberg, or menu cost models. In the case of real rigidities, firms are reluctant to adjust their prices but not because prices are sticky but rather because the environment limits the benefit of such adjustments, even in response to large shocks.

We choose to focus on real rigidities. Although there is evidence of price stickiness in our data—see for example, Carlsson and Skans (2012) and Carlsson (2017)—real rigidities are important even in the presence of price stickiness. Moreover, we perform two exercises that suggest that firm-level price stickiness does not provide a complete explanation for incomplete passthrough. To begin with, if incomplete passthrough was primarily driven by Calvo or menu-cost price rigidity, then the incomplete passthrough results would reflect a mixture of firms who adjust their prices and those who do not. To evaluate this possibility, we re-estimate our passthrough equations but drop observations with small price changes (less than 1% in absolute value). The results are presented in column 4 of Table 3. Notably, the measured passthrough is similar to the baseline results despite the fact that the estimates are based on only those firms who actively adjust their prices. Second, Rotemberg adjustment costs would imply that firms gradually adjust their prices in response to a shock, meaning that passthrough would be initially incomplete, but greater at longer horizons. While a deep investigation of the dynamics of passthrough would require a more fully fledged analysis, we provide suggestive evidence that passthrough is incomplete even at longer horizons. To do this, we re-run the IV specification using two-year lagged shocks as the instruments. This estimates the response of prices to shocks which are persistent enough to be predicted at a two year lag. As presented in column 5 of Table 3, we find higher passthrough, about -0.3, but the increase is moderate.[38]

Another question of interest is whether passthrough varies for small versus large shocks. In the final column of Table 3 we repeat our first difference specification, but allow the coefficients to vary for extreme shocks, defined as being below the 5th or above the 95th percentile in the shock distribution. We find that, 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.

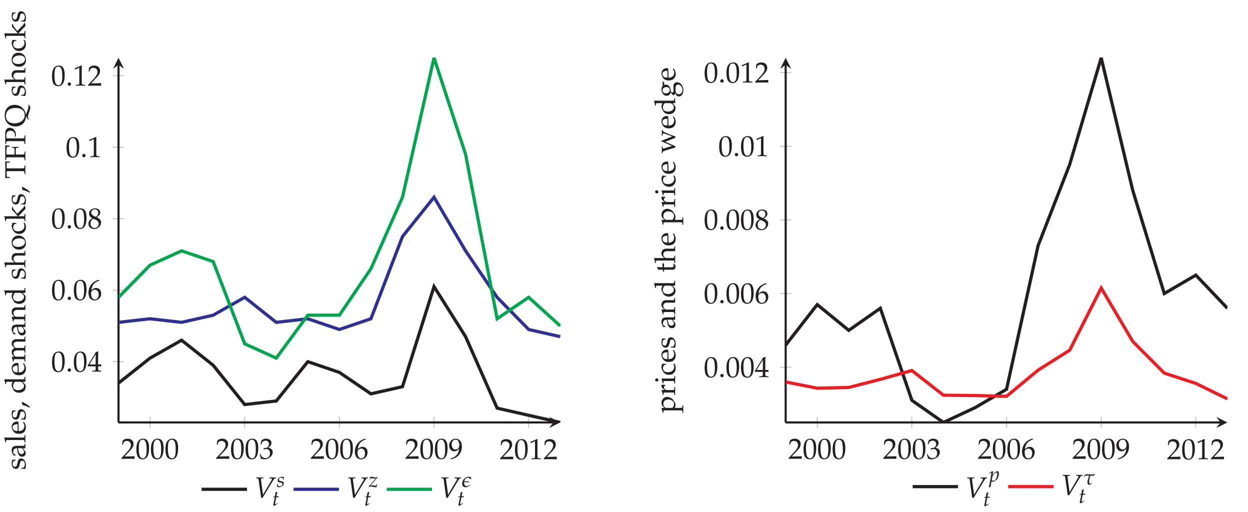
5.3 Variance decomposition of sales and price growth

To evaluate the economic importance of shocks for dispersion from an empirical perspective, we next perform 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. We refer to these exercises as semi-structural because they rely on our demand estimates and our behavioral passthrough equation. In other words, how the underlying distributions of shocks drives sales and price dispersion is determined by the parameters that we estimate. We explain the structural channels through which the shocks transmit to outcomes below.

Figure 4 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 any 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 dispersion 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

38. We have also investigated this by re-running the first difference specification with two and three year differences in prices and shocks. We found near identical parameters at these longer horizons to the one-year difference specification.

Figure 4: 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.

(in black and red, respectively). Here we see a sharp spike in the dispersion of the price wedge in 2009 and little cyclicality otherwise. Note that the variances are presented in different plots because of differences in scale.

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 (6), subtracting sector-year means from both sides, and taking the variance across all firms yields

$$V_t(\tilde{\Delta} p_{i,t}) = \beta_z^2 V_t(\tilde{\Delta} z_{i,t}) + \beta_\epsilon^2 V_t(\tilde{\Delta}\epsilon_{i,t}) + V_t^{p,resid}. \tag{7}$$

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: $V_t^{p,z} \equiv \beta_z^2 V_t^z$ measures the contribution of TFPQ dispersion, $V_t^{p,\epsilon} \equiv \beta_\epsilon^2 V_t^\epsilon$, measures the contribution of demand dispersion, and $V_t^{p,resid}$ measures the residual variance. $V_t^{p,resid}$ contains the covariances between shocks as well as the variance of the price wedge.[39] To ease the exposition going forward, we write this equation more compactly as

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

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

39. Specifically, $V_t^{p,resid} \equiv V_t(\tilde{\Delta}\tau_{i,t}) + \beta_z\beta_\epsilon \text{cov}_t(\tilde{\Delta} z_{i,t}, \tilde{\Delta}\epsilon_{i,t}) + \beta_z \text{cov}_t(\tilde{\Delta} z_{i,t}, \tilde{\Delta}\tau_{i,t}) + \beta_\epsilon \text{cov}_t(\tilde{\Delta}\epsilon_{i,t}, \tilde{\Delta}\tau_{i,t})$. See Appendix C.2.

of demand given by the CES model (3). 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 (6) to give

$$\log s_{i,t} = (1-\theta)\beta_z z^u_{i,t} + ((1-\theta)\beta_\epsilon + 1)\,\epsilon_{i,t} + (1-\theta)\tau_{i,t}. \tag{9}$$

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 the sales decomposition equation:

$$V_t(\tilde{\Delta} s_{i,t}) = (1-\theta)^2\beta_z^2 V_t(\tilde{\Delta} z_{i,t}) + ((1-\theta)\beta_\epsilon + 1)^2 V_t(\tilde{\Delta}\epsilon_{i,t}) + V_t^{s,resid}. \tag{10}$$

Again, $V_t^{s,resid}$ is the residual variance, containing the covariance terms and variance of the price wedge,[40] and we write the decomposition compactly as

$$V_t^s = V_t^{s,z} + V_t^{s,\epsilon} + V_t^{s,resid}. \tag{11}$$

This expression allows us to decompose the time-varying variance of sales growth using the time varying variances of shocks, passed through parameters of the passthrough equation and demand equations. There are three components of the decomposition: $V_t^{s,z} \equiv (1-\theta)^2\beta_z^2 V_t^z$ measures the contribution of TFPQ dispersion, $V_t^{s,\epsilon} = ((1-\theta)\beta_\epsilon + 1)^2 V_t^\epsilon$, measures the contribution of demand dispersion, and $V_t^{s,resid}$ measures the residual variance.

There are three parameters linking shocks to variances. β_z and β_ϵ measure how responsive firms' prices are to their shocks, and consequently how much these shocks can plausibly explain movements in prices. For sales, the demand elasticity θ measures how elastic sales is to changes in prices, and therefore to changes in the shocks which drive prices.[41]

5.3.2 Variance decomposition results

We present our variance decomposition results in Figure 5. Each plot in this figure shows a time series of the relevant variable—either sales or prices—along with the time series of each component of the decomposition—TFPQ, demand, and residual components. The residual components include one term that represents the contribution from the variance of the price wedge and three covariance terms between shocks and the price wedge. Decompositions of sales are shown on the left side of the figure and decompositions of prices are shown on the right of the figure. We illustrate our main results in the top row (panel a). For these results, we use the passthrough

40. Specifically, $V_t^{s,resid} \equiv (1-\theta)^2 V_t(\tilde{\Delta}\tau_{i,t}) + \beta_z(1-\theta)\,((1-\theta)\beta_\epsilon+1)\,\mathrm{cov}_t(\tilde{\Delta} z_{i,t}, \tilde{\Delta}\epsilon_{i,t}) + \beta_z(1-\theta)^2\mathrm{cov}_t(\tilde{\Delta} z_{i,t}, \tilde{\Delta}\tau_{i,t}) + (1-\theta)\,((1-\theta)\beta_\epsilon+1)\,\mathrm{cov}_t(\tilde{\Delta}\epsilon_{i,t}, \tilde{\Delta}\tau_{i,t})$. See Appendix C.2.

41. 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.

estimates from the first differences specification (taken from Table 3 column 3). We focus on the first difference estimates because the variance decomposition is conducted in first differences. The first-difference passthrough coefficients thus give the most relevant estimate of how firms change their prices (in first differences) in response to a change in the shocks. In panel (b), we show the same variance decomposition but using the IV estimates of passthrough (taken from Table 3 column 2). We present these results as a robustness exercise.

As is evident in the sales decomposition, demand shocks play a dominant role in driving sales growth dispersion. On average, demand accounts for 63% of sales dispersion, and is the most important of all the components. Demand also drives the cyclicality of sales dispersion. During the Great Recession, the variance of sales growth rose by 82% from 2008 to 2009, before gradually recovering through 2010 and 2011. The left panel of Figure 5(a) shows that this increase in sales dispersion, V_t^s, is mostly driven by the contribution of demand dispersion, $V_t^{s,\epsilon}$. In fact, comparing 2009 to non-recession years, we find increased demand dispersion explains 80% of the increase in sales dispersion.[42] 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. In contrast, TFPQ explains almost nothing of sales growth dispersion on average, and hence plays no role for cyclicality in either the Great Recession or during 2001. Demand dispersion thus appears to be much more important for explaining both the level and cyclicality of sales dispersion than TFPQ dispersion.

Given that demand dispersion explains most of the rise in sales dispersion and TFPQ explains little, what explains the remaining half? The remainder is driven by the residual terms. The variance of the price wedge, which measures movements in prices not correlated with changes in TFPQ or demand, accounts for around 25% of sales dispersion on average. This suggests that frictions or shocks other than demand and TFPQ may be quite important for explaining year-to-year changes in firms' sales. Alternatively, this could reflect misspecification of the econometric model used in the decomposition. During the Great Recession, the rise in the variance of the price wedge also contributes meaningfully to the rise in sales dispersion. The remaining rise in sales dispersion can be explained by a rise in the contribution of the correlation between demand shocks and the price wedge. We discuss this further below in conjunction with an analysis of time-varying passthrough.[43]

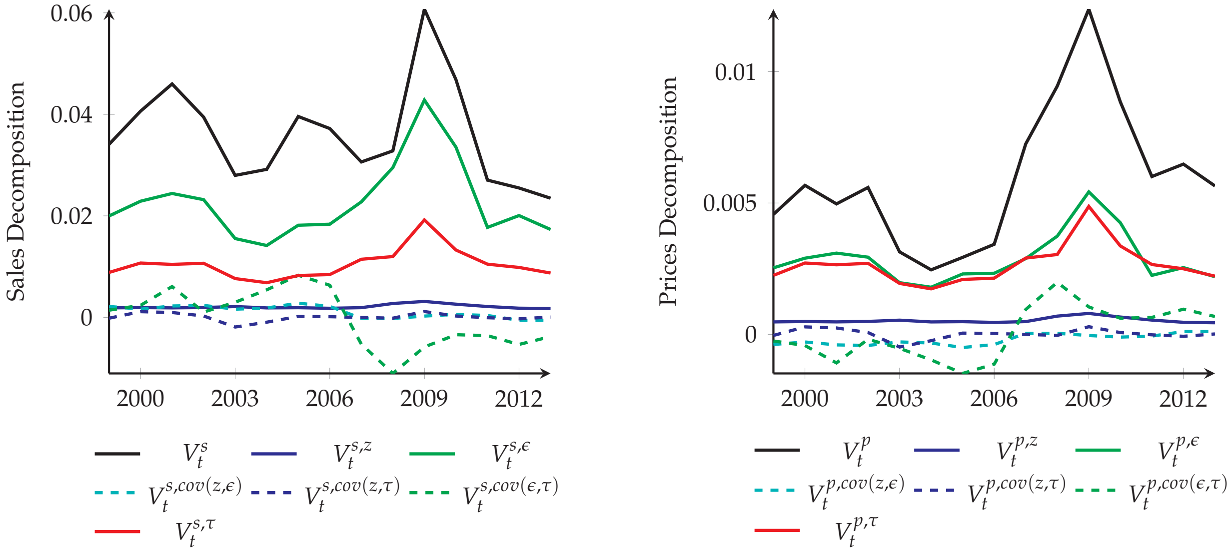
The picture is similar for price dispersion. As seen in the right panel of Figure 5(a), demand plays an important role while TFPQ does not. On average, demand accounts for about 50% of price dispersion, and 40% of the increase in price dispersion in 2009 relative to non-recession years. TFPQ, in contrast, explains only 10% of price dispersion on average and almost none of its movements over the cycle. For prices, the variance of the price wedge has significant explanatory power on average. The price wedge also spikes between 2008 and 2009. The correlation between demand shocks and the price wedge again plays an important cyclical role, but we defer discussion of this to the section on time-varying passthrough.

Overall, demand dispersion is more important than TFPQ dispersion in explaining the variance of firm level endogenous outcomes (sales and prices) both on average and over the cycle. In addition, there remains significant unexplained dispersion, which we attribute to the price wedge.

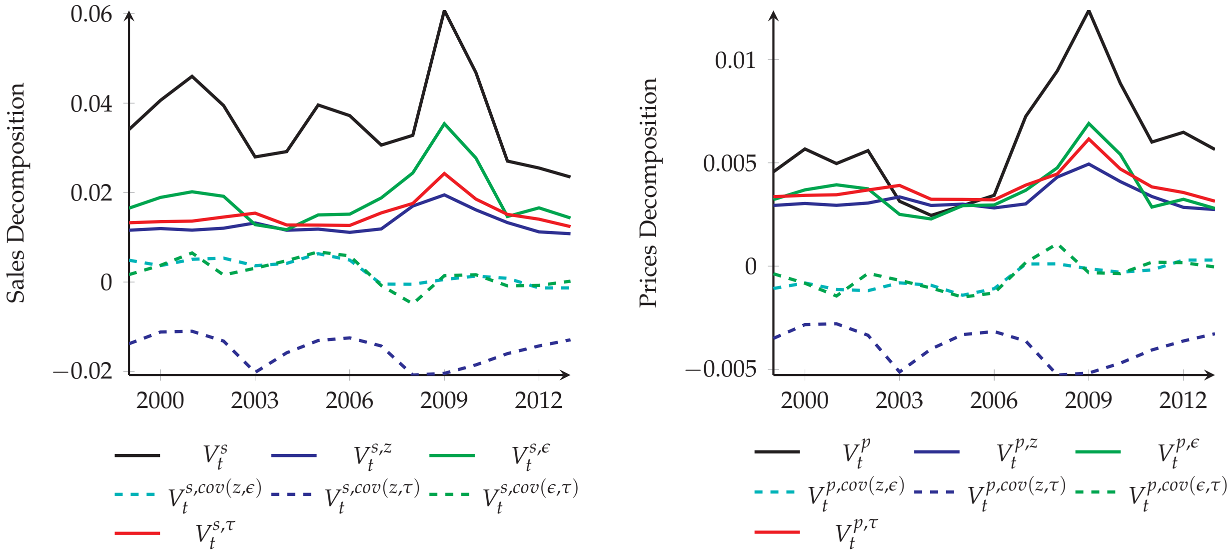
42. See Table 33 for these calculations. Appendix C.2 contains additional discussion and quantification.

43. Our estimation of (6) in first differences over the whole sample implies that $\tilde{\Delta}\tau_{i,t}$ is uncorrelated with $\tilde{\Delta}z_{i,t}$ and $\tilde{\Delta}\epsilon_{i,t}$ over the whole sample. This does not preclude the growth rates from being correlated within any given year. On average the correlations are small, but they do play a role in recession years.

Figure 5: Variance Decompositions of Sales and Prices



(a) Passthrough estimated in first differences



(b) Passthrough estimated using i.v.

Figure 5 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.

Although we focus on the roles of demand and productivity shocks in the remainder of the paper, the source of the price wedge is an important topic for future work.[44]

Why is demand dispersion more important than TFPQ dispersion? The role of passthrough What explains these results? Why is demand the dominant driver of sales dispersion over the cycle, while TFPQ is largely irrelevant? We are able to provide economic explanations for these findings because our variance decomposition is specified in terms of two structural equations.

The limited contribution of TFPQ to dispersion in endogenous variables follows directly from the low TFPQ passthrough that we estimate in the data. Recall that the contribution of TFPQ dispersion to sales dispersion is given by $V_t^{s,z} \equiv (1-\theta)^2\beta_z^2 V_t^z$, and that our passthrough and demand curve coefficients are $\beta_z = -0.0965$ and $\theta = 2.99$ respectively. This implies that $V_t^{s,z} \simeq 0.04 \times V_t^z$. Hence, the variance of TFPQ is shrunk by a factor of about forty. Thus, the substantial levels of TFPQ dispersion that we observe are simply not transmitted into sales dispersion due to the structure of the relationships between TFPQ, prices, and sales. The main cause is the low passthrough from TFPQ to prices: Since $\beta_z^2 \simeq 0.01$, this factor shrinks TFPQ dispersion (V_t^z) by about a factor of one hundred.

The economic intuition for this result is that TFPQ does not directly affect sales. TFPQ only indirectly affects sales via the price a firm sets. TFPQ shocks create price movements according to (6) 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 on average, and therefore also explains the inability of rising TFPQ dispersion in the Great Recession to explain the increase in sales dispersion.[45]

Demand dispersion is the most important driver of sales growth dispersion. Passthrough from demand shocks to prices is $\beta_\epsilon = 0.209$, and the contribution of a given level of demand dispersion to sales dispersion is thus $V_t^{s,\epsilon} = ((1-\theta)\beta_\epsilon + 1)^2 V_t^\epsilon \simeq 0.4 \times V_t^\epsilon$. 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 comparison $V_t^{s,z} = 0.04 V_t^z$). There are two reasons for the large multiplier. The first and most important reason is that 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. Secondly, 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. Combined with high demand dispersion on average, and a large increase during the Great Recession, this explains both the average and cyclical importance of demand shocks for sales dispersion.

Demand shocks are also more important than TFPQ shocks in the price decomposition for very similar reasons: The larger passthrough from demand to prices than TFPQ to prices, combined with the larger increase in demand dispersion during recessions.

44. 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%.

45. Note that the variance being a squared concept is particularly punishing to the role of TFPQ, because taking the variance squares the already small passthrough estimate. Measured in terms of standard deviation or IQR, the contribution of TFPQ is higher. For example, see our model-based interquartile range decomposition in Section 6.5.

5.3.3 Robustness and extensions

Functional form One possible concern is that the variance decomposition results are biased—or an artefact of—the assumption of (log-) linear demand. [46] This concern is perhaps especially salient because a non-constant elasticity demand model appears to better match the data. However, we do not find evidence to substantiate this concern. To begin with, the CES model appears to properly characterize the underlying distribution of demand shocks. We find a similar degree of demand dispersion regardless of whether we measure demand shocks based on a CES model or based on a more flexible demand specification.[47] What about the transmission of shocks to outcomes? 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 assumption on demand.

Passthrough coefficients Another possible concern is that the results are driven by the low degree of TFPQ-passthrough. Because greater TFPQ-passthrough boosts the role of TFPQ dispersion, the choice of TFPQ passthrough estimate will play a role for the variance decomposition results. To understand the sensitivity of the results to this choice, we therefore re-compute the variance decompositions using our largest estimated TFPQ-passthrough values. In the IV-estimation, we find a passthrough between 0.2 and 0.3. Consistent with intution, when we use our largest passthough estimate TFPQ dispersion plays a more meaningful role in the decompositions. This is especially the cases for prices. In periods in which TFPQ dispersion rises—such as the Great Recession—TFPQ plays a substantial role in driving volatility of price dispersion. Nevertheless, demand remains important for both prices and sales, and is the major driving force for sales dispersion.

One caveat of this exercise is that we use the same measured variance of TFPQ shocks in the exercise. Since firms clearly do not respond to all TFPQ shocks with this higher passthrough—as the IV passthrough estimate applies only to TFPQ shocks which turn out to be more persistent—it would perhaps be more appropriate to use a different measure of TFPQ shocks for a variance decomposition using the IV passthrough estimates. This measure of TFPQ dispersion would likely be smaller, so these results likely only give upper bounds on the contributions of TFPQ to sales and price dispersion.[48]

Time-varying passthrough Our variance decompositions rely on passthrough coefficients for demand and TFPQ that are constant over time. However, time-varying passthrough is a real and important possibility. For example, Berger and Vavra (2019) investigate whether increases in

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

47. This is likely due to the fact that the CES model provides a first order approximation of more complex demand functions.

48. This issue is evident when looking at the contribution of the covariance between TFPQ changes and price wedge changes in the decompositions. The contribution is very negative on average, and increasingly so in the Great Recession, in contrast to the first-difference approach where the term is always close to zero. This reflects the fact that the IV passthrough estimate differs from the true correlation between first-differenced TFPQ and τ.

shock dispersion or increases in responsiveness to shocks (i.e. passthrough) drives countercyclical dispersion in endogenous variables. In our case, time-varying passthrough is important because cyclical changes in passthrough will affect our variance decompositions.

To investigate the possibility of time-varying passthrough, we estimate passthrough on a year by year basis. These estimates are presented in Appendix C.2. 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.[49] Time-varying passthrough also explain why there is a large contribution from the correlation between demand shocks and the price wedge in 2009 in our variance decompositions (see Figure 5).

Taking the possibility of time-varying passthrough seriously, we re-compute our variance decompositions using using passthrough coefficients estimated period by period (see Figure 21 in the Appendix). We find that our main results are robust to this approach. In fact, the contribution of demand shocks to the increased sales dispersion in the Great Recession is now even larger because firms change their prices relatively less in response to demand shocks during 2009. It also means that there is no longer any contribution from the the correlation between the demand shocks and price wedge. This correlation is eliminated by construction when passthrough is re-estimated every year.

5.4 Summary of all empirical results

We present below a summary of our main empirical findings from Sections 2 to 5:

1. Dispersion rises in recessions. Measured within sector, the dispersion of sales, prices, employment, and intermediates, and utilization all rise during the Great Recession and (to a lesser extent) the 2000-2002 growth slowdown.

2. We find significant dispersion in TFPQ and demand shocks on average. This is true within sector, and even after correcting TFPQ for prices (by definition) and utilization, and demand for a non-CES specification.

3. We estimate statistically significant and economically meaningful departures from CES demand. In particular, the elasticity of demand rises as firms raise their price relative to their average price

4. Both TFPQ and demand dispersion are cyclical, rising in recessions. Demand shock dispersion rises more than TFPQ shock dispersion, and correcting for utilization reduces the cyclicality of TFPQ shock dispersion.

5. Passthrough from TFPQ shocks to prices is incomplete, as firms lower prices by between 10-30% of the rise in TFPQ. This is in contrast to the basic CES static optimal policy, which implies 100% passthrough.

6. Firms raise their prices by between 21-24% in response to a demand shock which would otherwise have raised their quantity sold by 100% had they kept their price fixed. This is

49. However, we also find systematically higher demand passthrough overall in the latter half of our sample as compared with the first half.

also in contrast to the basic CES model, where firms do not adjust prices in response to demand shocks.

7. Due to low estimated passthrough from TFPQ to prices, a semi-structural variance decomposition exercise finds that TFPQ shocks explain very little of the variance of price or sales changes across firms. The rise in sales dispersion in recessions is mainly caused by the large rise in demand shock dispersion.

In the remainder of the paper, we build a heterogeneous firm model which is able to replicate these facts, and use it study the aggregate implications of shocks to dispersion and uncertainty.

6 Quantitative Model

In most existing theoretical work related to dispersion, productivity and demand shocks have isomorphic effects. However, our main empirical results—including the importance of demand shocks for business cycle variability, the incomplete passthrough of TFPQ shocks to prices, and the rejection of the CES demand specification—indicate the need for a model in which demand and TFPQ shocks can play different roles. In this section, we therefore build a quantitative heterogeneous firm model. Our goals are twofold: The first is to investigate what features are needed to match our empirical conclusions regarding dispersion and passthrough. Our second purpose is to evaluate the economic importance of distinguishing between demand and productivity shocks. In particular, we treat dispersion as uncertainty and investigate "wait and see" behaviour related to both productivity and demand.

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 appear to have non-constant elasticities of demand. This invalidates the usual result, derived under the assumption of CES, that TFPQ and demand shocks are isomorphic and can be studied as a single shock to TFPR. This extension will also allow the model to generate passthrough from TFPQ shocks to prices in line with the data.

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 with multiple shocks. Moreover, it is justified given that general equilibrium effects are likely muted in the short-run, as discussed in Bloom et al. (2018). These simplifications are helpful, given that the firm's problem has an additional state variable coming from the distinction between TFPQ and demand shocks, and will have three exogenous, and one endogenous, states. Second, we simplify the adjustment cost structure to reduce the dimensionality of the firm's problem.

Time is continuous and indexed by t. The 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 i 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:

$$\log q = \frac{\theta}{\eta} \log(1 - \eta \log p) + \varepsilon. \tag{12}$$

The model is calibrated to have $\mathrm{E}[\log p] = 0$, so that θ denotes 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. The labour stock also depreciates at rate δ, which is assumed to be the same as capital depreciation for simplicity.[50] The stock of workers evolves according to $\dot{l} = h - \delta l$. Define the "overall scale" of a firm as $x \equiv k^{\alpha}l^{1-\alpha}$, which is the total amount of inputs weighed by their elasticities. Notice that output is therefore simply given by the linear function of x, $q = zx$. In the interests of simplicity, 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 the single state variable x, rather than both l and k separately.[51]

To do this, all non-convex adjustment costs are placed on the adjustment of the overall scale of the firm, x, rather than the individual factors. Specifically, capital can be bought and sold at price p_k. Labour can be hired at cost a (which is recouped when workers are fired). Capital and labour adjustment costs (such as resale loss from capital, firing costs, fixed costs) are instead placed on adjusting the overall size x. If $\dot{x}$ denotes the desired rate of change in x, we can define an investment rate for x as

$$\dot{x} = i_x - \delta x, \tag{13}$$

where i_x is investment in overall scale. In the appendix we derive how this is split into investment in each factor. The cost (on top of p_k and a) of investment in scale at rate i_x is assumed to be

$$c(i_x, x) = \begin{cases} \frac{\kappa}{2}\frac{(i_x - \delta x)^2}{x} & i_x > \delta x \\ 0 & \delta x \geq i_x \geq 0 \\ -\underline{\kappa} i_x + \frac{\kappa}{2}\frac{i_x^2}{x} & i_x < 0. \end{cases} \tag{14}$$

Here κ controls quadratic adjustment costs, which are paid for investment rates above the rate of depreciation, or for disinvestment.[52] $\underline{\kappa}$ is a partial irreversibility cost, meaning that the resale price of inputs is $\underline{\kappa}$ less than the purchase price. Total static cashflow is $cf = pq - wl - p_k i - ah - c(i_x, x)$.

50. In our baseline calibration this choice is without loss of generality since we will assume no hiring costs directly on the labour stock, making the choice of depreciation rate irrelevant.

51. Having adjustment costs on both capital and labor is important for generating large wait and see effects, as with adjustment costs only on one factor 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.

52. 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.

Both TFPQ, z, and the demand shock, ε, follow Markov processes with stochastic volatility. Starting with TFPQ, 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:

$$z' = (1-\rho_z)\mu_z + \rho_z z + \sigma_z(s)u_z, \qquad u_z \sim N(0,1), \tag{15}$$

where 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, for demand the firm draws a new level at rate λ^ε from an AR(1) process. 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

$$e^{\varepsilon'} = (1-\rho_\varepsilon)\mu_\varepsilon + \rho_\varepsilon e^\varepsilon + \sigma_\varepsilon(s)u_\varepsilon, \qquad u_\varepsilon \sim N(0,1), \tag{16}$$

where the parameters are defined symmetrically to the process for TFPQ above. This formulation is equivalent to redefining the demand curve to have the demand shifter enter as $\log q = \frac{\theta}{\eta}\log(1 - \eta \log p) + \log \hat{\varepsilon}$, with $\hat{\varepsilon}$ being normally distributed.

6.3 HJB and solution

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

Static cashflow can be written 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, which is just an average of the costs of investment in capital and labour.

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

$$\begin{aligned} rv(x,z,\varepsilon,s) = \max_{i_x} \pi(x,z,\varepsilon) - p_x i_x - c(i_x,x) + v_x(i_x - \delta x) + \lambda^z \left(\mathrm{E}_{z'}[v(x,z',\varepsilon,s)|z,s] - v(x,z,\varepsilon,s)\right) \\ + \lambda^\varepsilon \left(\mathrm{E}_{\varepsilon'}[v(x,z,\varepsilon',s)|\varepsilon,s] - v(x,z,\varepsilon,s)\right) + \lambda^s(s)\left(v(x,z,\varepsilon,s_{-1}) - v(x,z,\varepsilon,s)\right). \end{aligned} \tag{17}$$

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 proceeded 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. The solution to the investment problem is given in the appendix.

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 found to be constant over time, as it is in our exercises.

6.4 Steady state results: can the model generate sensible passthrough and dispersion?

As well as distinguishing between demand and supply shocks, our model's novelty lies in combining features of the price setting literature—namely a pricing decision with a non-CES demand curve—with features of the wait and see literature—namely adjustment costs in factor choices at the firm level. Thus we first investigate whether the combination of these features helps the model to replicate hard-to-match features of the data. For these exercises we first 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.[53]

Calibration One unit of time corresponds to one year. We choose a discount rate of $r = -\log(1 - 0.05)$, implying a 5% yearly discount rate. The capital depreciation rate is set to $\delta = -\log(1 - 0.1)$ to imply a 10% annual depreciation rate. We choose $\alpha = 0.255$ to match the capital share of costs in our dataset. We use the mean of demand μ_ϵ to normalise aggregate capital to $K = 1$ in the ergodic distribution. The real wage w is chosen to normalise aggregate labour to $L = 1$ in the ergodic distribution.[54] μ_z is chosen to shift mean TFPQ, and hence the mean price such that the log average (relative) price set by firms in the ergodic distribution is equal to zero.

Our demand curve parameters are taken from our estimates in Section 3.2, and represent the key departure of our model from a standard CES demand (or equivalent decreasing returns to scale) model. We choose $\theta = 3$ and $\eta = 4.3$ to allow for a non-constant elasticity of demand in line with our estimates. With $\eta > 0$, firms face an increasing demand elasticity when they raise their price.

In order to remain comparable with the existing literature, and focus on how our new demand specification changes the propagation of uncertainty shocks, we do not estimate adjustment costs using our dataset, and instead take the values used in Bloom et al. (2018). We normalise the purchase price of capital to $p_k = 1$. Recall that we place all non-convex adjustment costs on the adjustment of overall scale, x. Accordingly, we set the direct linear hiring cost to zero ($a = 0$) and represent all hiring costs using the costs on x. 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 Bloom et al.'s (2018) values 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 $\underline{\kappa} = 0.34 p_k (b^*)^{\alpha-1} + 2 \times 0.018 w (b^*)^\alpha = 0.3565$.[55] We choose not to use convex adjustment costs for calibration purposes, and set them close to zero.[56]

Finally, we calibrate our idiosyncratic shock processes to match moments of our observed

53. The results are almost identical if we focus on the ergodic distribution of a long simulation within the low uncertainty state of the full model with two uncertainty values, which we calibrate in the next section.

54. 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.

55. 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.

56. 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. Without it, firms outside of their inaction regions would immediately jump their capital to the new optimum, requiring more complicated numerical methods. We solve the model using methods based on Achdou et al. (2022). With our low value of κ, firms adjust their capital very quickly, and reach the new optimum within half a month.

yearly firm-level TFPQ and demand data. In comparing our model shocks (which evolve in continuous time) to our shocks in the data (which are measured yearly and might suffer from time-aggregation and measurement issues) we follow Bloom et al. (2018) and generate model-simulated yearly data constructed in the same way as our yearly data is. The model processes are chosen to match features of the data, and details are given in the appendix. 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, since firms draw new shocks on average once per year, implies a yearly autocorrelation of shocks of roughly 0.8 and 0.6 respectively, in line with what we estimate on our data.[57] 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 these dispersions we measure demand and TFPQ exactly as we would in the data, computing yearly measures which account for time aggregation and measurement error. See Appendix D.3 for more details and a table containing our calibrated parameters. We target an IQR of demand and TFPQ innovations both of 0.2, which corresponds roughly to the values in the years before the Great Recession (see Figure 2(b)).

Model validation – Dispersion We first validate our model's ability to generate sensible dispersion in endogenous variables in response to the dispersion in demand and TFPQ measured in the data. In the top row of Table 4(a) we give the IQRs of sales and prices, along with the shocks, in steady state. Despite being completely untargeted, the model generates dispersion of both very similar to the data. The IQR of sales growth is 0.195 in the model, close to the 0.17 in the 2005 data, and for prices the model and data are 0.064 and 0.056.

This success is not guaranteed, and arises endogenously from how strongly firms respond to shocks. To see this, in the second row of the table we provide the same dispersions calculated in a recalibrated model where the demand curve is assumed to be CES ($\eta = 0$). This model generates a much higher dispersion of price changes – an IQR of 0.107 – from the same dispersion in shocks. While the CES model does still generate a similar IQR of sales growth, it is only the full model with non-CES demand which can match both sales and price dispersion simultaneously. To understand why, we now turn to discussing passthrough.

Model validation – Passthrough A crucial finding of our empirical work was that passthrough from shocks to prices deviated from that implied by a simple static CES optimization model. 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.

Starting with passthrough from TFPQ, a striking feature of the model is how successfully it is able to generate low passthrough from TFPQ shocks to prices. 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%, and is much closer to the values of around 10% to 25% which we estimated on the data in Section 5.2.

57. We estimate the autocorrelation of the idiosyncratic demand shock to be 0.625 on our data using the Anderson-Hsaio method. For TFPQ, which is likely measured with more error, we find a range of autocorrelation estimates depending on the method used, and select 0.8 as a sensible value in the range of these estimates which is consistent with values used in the literature.

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
R^2:	78%	50%	59%
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.

The low passthrough in our model follows mainly from the estimated non-CES demand curve, and to see this clearly we can look at the optimal policies in a special case of our model with no adjustment costs ($\kappa = \underline{\kappa} = 0$), which reduces to 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:

$$\log p \simeq -\frac{\theta}{\theta + \eta} \log z. \tag{18}$$

Derivations are in Appendix D.4. This yields a passthrough equation directly comparable to our estimated equation, (6), allowing us to compare how the model's predictions for optimal price setting compare with the price setting behavior we observe in the data. When demand is non-CES passthrough is incomplete since, for any $\eta > 0$, the absolute size of the coefficient TFPQ must be less than one: $\frac{\theta}{\theta+\eta} > -1$. Intuitively, when $\eta > 0$ a firm's elasticity of demand rises as it increases its price. This captures the idea that it is hard for firms to easily gain new customers by lowering their price, and easy for them to lose existing customers by raising their price. This implies that firms find it less appealing to change their price in response to productivity changes, because lowering your price brings little extra revenue if quantity sold does not increase much, and raising your price brings little extra revenue of quantity sold decreases a lot. Hence firms adjust prices less than one-for-one to changes in productivity.

Following the data, our coefficients $\theta = 3$ and $\eta = 4.3$ give statically-optimal passthrough from TFPQ to prices of $\theta/(\theta + \eta) = 41\%$ according to this approximation. Thus, even abstracting from the full model, the departures from a CES demand curve that we estimated already rationalises a large part of the incomplete passthrough from TFPQ shocks to prices that we see in the data.

The full model implies passthroughs which are even lower, and hence closer to the data, as adjustment costs further inhibit passthrough. That factor adjustment costs affect passthrough, including generating passthrough from demand shocks to prices, is simple yet intuitive. The results can be seen with a CES demand curve, which we therefore use to demonstrate the result. Using the production function $q = zx$, inverting the CES limit of (12), taking $\alpha_{fe} = 0$ for clarity, allows us to express a firm's price as

$$\log p = -\frac{1}{\theta}(\log z + \log x) + \frac{1}{\theta}\epsilon. \tag{19}$$

This differs from (18) because the input x has not been optimized. Thus, in the absence of changes in the firm's input use, we necessarily see (incomplete) passthrough from both demand and TFPQ shocks to prices. Intuitively, if they do not change their input use, a firm's quantity sold is fixed at $q = zx$, and they must adjust their price in order to convince customers to purchase that quantity. If a firm chooses not to change their inputs due to adjustment costs, a TFPQ shock leads to passthrough of $1/\theta = 33\%$, in contrast to the static optimisation which leads to 100% passthrough with CES. This explains why adjustment costs in the full model further reduce TFPQ passthrough, on top of the effect of non-CES demand. Comparing passthrough estimates from the full model, our model delivers lower passthrough from TFPQ to prices than the CES model in all three regression specifications considered. For example, the CES model generates 82% passthrough in the IV specification, while the non-CES model gives passthrough of 33%.

Moving on to passthrough from demand shocks to prices, (18) reveals that in the absence of adjustment costs, the statically optimal price does not respond to demand shocks, giving passthrough of zero. This is true both for the CES ($\eta = 0$) and non-CES ($\eta > 0$) model. In this kind of framework the demand shock simply shifts the number of units that can be sold, but not their optimal price. Hence, the static model cannot explain why firms change their prices in response to demand shocks, as we saw in the data where passthrough from demand shocks was around 20%. However, we see in Table 4(b) that firms do adjust their prices in response to demand shocks in our full model including adjustment costs. The coefficients are smaller than the data, ranging from 5.5% to 14.8% depending on the specification, with the first difference specification giving the highest value which explains over half of the passthrough seen in the data. Adjustmen costs explain this: the simple example in (19) shows that if a firm does not adjust its inputs at all in response to a demand shock, it must instead move its price, with a passthrough of $1/\theta = 33\%$. The partial adjustment of inputs in response to shocks, which adjustment costs add to our model, can explain up to half 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 6 and its discussion in the next section.[58]

In summary, these exercises validate that our model is able to replicate well the novel features of the data that we documented. Firstly, it features both demand and TFPQ shocks at the firm level. Secondly, it generates passthrough from these shocks in line with the data, allowing it to match the dispersions in sales and price changes that we see in the data.

58. 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). Firstly, they show that decreasing returns to scale in production dampen TFPQ passthrough and increase demand passthrough. With CES demand and overall returns $\beta_K + \beta_L = \gamma$ optimal static passthrough becomes $\beta_z = -1/(\gamma + \theta(1-\gamma))$ and $\beta_\epsilon = (1-\gamma)/(\gamma + \theta(1-\gamma))$. For our estimated demand elasticity of $\theta = 3$, even quite strong decreasing returns to scale of $\gamma = 0.8$ would imply passthrough of $\beta_z = -0.71$ and $\beta_\epsilon = 0.14$, leaving most of incomplete TFPQ passthrough unexplained, but actually potentially explaining around 3/4 of demand passthrough. Secondly, since we are focusing on short run changes, this decreasing returns to scale is very interpretable as adjustment costs, as adjustment costs act as in increase in shadow marginal costs in the short run even if the unconstrained production function is CRS. Pozzi and Schivardi (2016) provide evidence that a firms ability to reorganize, a proxy for adjustment costs, correlates with passthrough as predicted by this framework.

6.5 Aggregate results: response to increase in dispersion

In this section we return to our full model with a high ($s = 2$) and low ($s = 1$) dispersion state. We follow Bloom (2009) and Bloom et al. (2018) and consider how the increase in dispersion represents a fundamental increase in *uncertainty* for the firm, and study its aggregate impacts.

Calibration of uncertainty process The calibration of the full model 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(b)). The peak increase in utilization-adjusted TFPQ is around 30%, and for demand it is around 60%, 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 distribution of the low uncertainty state.

The final feature of the process for uncertainty that needs to be calibrated is the persistence of the high and low uncertainty regimes. For our baseline calibration, we note that in our dataset, major recession events happened roughly eight years apart, and that dispersion is high 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 thus confirm that all of our results are robust to using the estimates for the US from Bloom et al.'s (2018), who estimate the persistence of the regimes using nearly 40 years of data.[59]

Policy function – (Dis)investment thresholds and inaction region The left and centre panels of Figure 6 give slices of a key firm policy function. Specifically, we define $\underline{x}(z, \epsilon, s)$ as the investment threshold, such that firms have positive investment ($i^x(x, z, \epsilon, s) > 0$) for current x below this value. $\bar{x}(z, \epsilon, s)$ 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, until it hits the investment threshold.

The central panel plots these functions across values of the demand shifter, with the productivity shock held at its central value. The solid lines plot the thresholds when uncertainty is in the low state, $s = 1$, with the investment threshold in blue and the disinvestment in red. The inaction region is wide, and simulations reveal a 25.4% yearly inaction rate on average (see Footnote 65). Consider a firm with $\epsilon = 0$, who has let their inputs depreciate down to the investment threshold (roughly $x = 1$). Following a move to $\epsilon = -0.2$, which means a 20% fall in demand, they choose not to actively disinvest, but let their inputs gradually depreciate. Given an annual depreciation rate of 10%, this implies a 10% reduction in output per year until the firm reaches the new opti-

59. 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 robustness section for details of the results with this calibration.

mum. Following the discussion in the last section this means the firm will have to initially lower its price in response to the demand shock, creating passthrough from demand to prices. The firm then gradually raises its price back towards the initial level as it gradually lowers its production. This effect is asymmetric, as the firm would instead choose to immediately invest in response to a positive demand shock, as the investment threshold is strongly upwards sloping in demand.[60]

The left panel plots the investment and disinvestment thresholds across productivity levels, with the demand shock held at its central value. We see the same inaction feature, but more importantly we see that the optimal input level is less responsive to productivity than demand, and even has a non-monotonic relationship. We discuss these features in more detail when we discuss our aggregate experiment. Finally, in both panels we also plot the thresholds in the high uncertainty state ($s = 2$), given as the dashed lines. We see that the inaction regions are wider in the high uncertainty state, meaning that when uncertainty is high firms more often choose to "wait and see" rather than undertaking costly and partially irreversible investment. The widening is driven mostly by a lowering of the investment threshold, meaning that when uncertainty rises firms allow their inputs to depreciate further, lowering production and hence aggregate output.

Model-based variance decomposition Before exploring the aggregate dynamics of an uncertainty shock, we return to our variance decomposition exercise, now through the lens of our model. 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. This was done through the lens of a log-linear CES demand curve and estimated passthrough equations, and in this section 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.[61] The first row shows that, 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 then 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 11). 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.[62]

In the final two rows of Table 4(a) we perform two counterfactual experiments. In the first, we recalibrate the model such that only TFPQ uncertainty is raised in the high uncertainty state. Specifically, we set $\sigma_\epsilon(2) = \sigma_\epsilon(1)$, and keep the estimated values of $\sigma_z(s)$.[63] In the second we do the opposite, and only raise demand uncertainty. Note that since we compute the IQRs of the shocks using simulated and time-aggregated data, raising each shock does lead to small increases in the measured IQR of the other shock, which demonstrates why such a procedure is important.

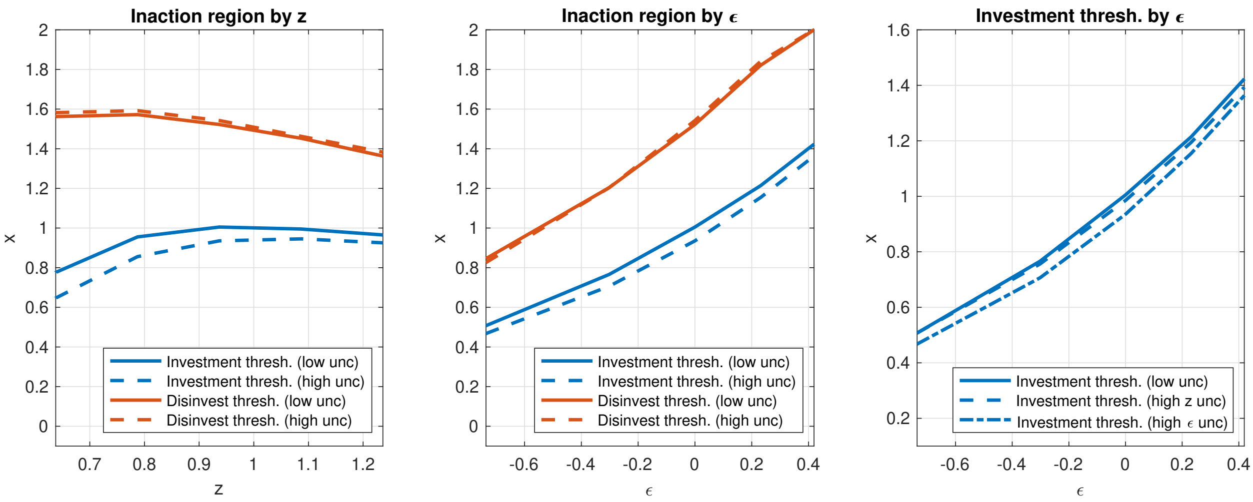
60. See Figure 27 in the appendix for plots of the impulse responses to positive and negative idiosyncratic demand and TFPQ shocks in the model.

61. 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 – see Figure 3 where the tails of the sales distribution are long even in the pre-recession period. For this reason we consider the results based on IQRs to be more robust to outliers, and focus on them in the text.

62. 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.

63. All other parameters are held constant, apart from μ_z which we adjust to ensure that $\mathrm{E} \log p = 0$ in the calibration at $s = 1$. Holding μ_z constant instead has no effect on the results.

Figure 6: Policy functions: Inaction regions by state and uncertainty level

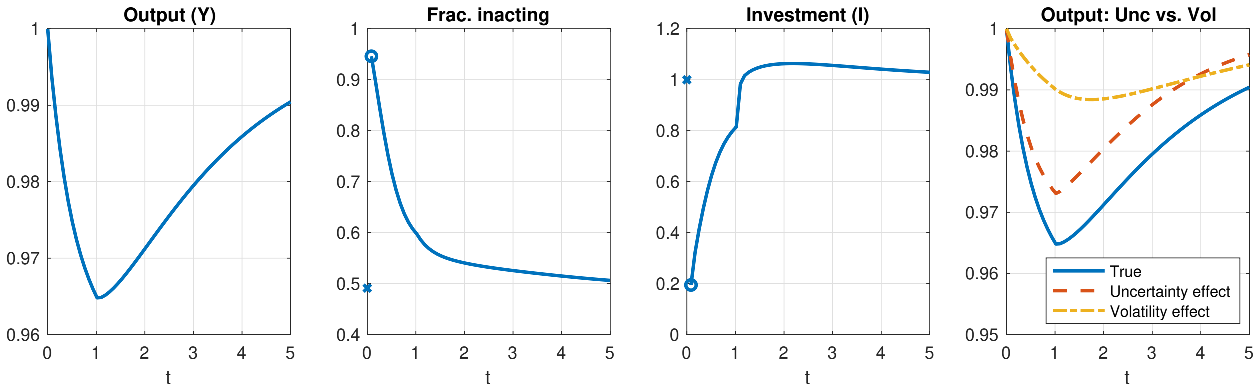


The left and centre panel give slices of the firm policy functions in the low uncertainty state (solid lines, $s = 1$) and high uncertainty state (dashed lines, $s = 2$). $\underline{x}(z, \epsilon, s)$ gives the investment threshold, such that firms have positive investment for current x below this 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.

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. 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. 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. Overall, this exercise serves to validate the model, as well as to emphasise the key result that increased demand shock dispersion is the main driver of increased sales dispersion during recessions.

Aggregate experiment To understand the aggregate implications of a rise in dispersion, we simulate a recession experiment in the model. Specifically, we consider the full distribution of firms, all of whom face the same level of uncertainty, which starts in the low state, $s = 1$. 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 our data sources which are yearly. From then on, the economy reverts back to the low uncertainty state at rate $\lambda^s(2)$.[64] Let $\mu_t(k, z, \varepsilon)$ denote the distribution of firms at time t within a given simulation. Aggregates are computed by integrating over this distribution, with GDP given by the integral over sales, and so on. All plots are averages across all possible realisations of the aggregate uncertainty process from time $t = 1$ onwards.

Figure 7: 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 top left panel gives the uncertainty state, top right gives aggregate output, bottom left the fraction of firms in their inaction regions at that instant, and bottom right the aggregate investment rate.

Aggregate response to increased dispersion The results of this exercise are given in Figure 7. The left panel shows the impact of the uncertainty shock on aggregate output, which falls by 3.5% in response to the rise in uncertainty. The remaining three panels explain the source of this fall. The centre-left 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.[65] As firms gradually adjust this number falls, and mostly recovers within one year. Consequently, aggregate investment (centre-right) falls, driving a fall in capital and labour which causes output to fall.

That rises in uncertainty can cause an aggregate fall in output is already well known and shown by, for example, Bloom (2009). However, our model differs significantly from previous work due to the presence of the non-CES demand curve, and this leads both to different aggregate effects, and to different transmission mechanisms. Firstly, in Bloom (2009), rising uncertainty leads to a short term fall in output but will eventually lead output to rise in an effect dubbed the "volatility overshoot". In our model, there is no such overshoot, and rising dispersion leads output to fall in both the short and medium term.[66]

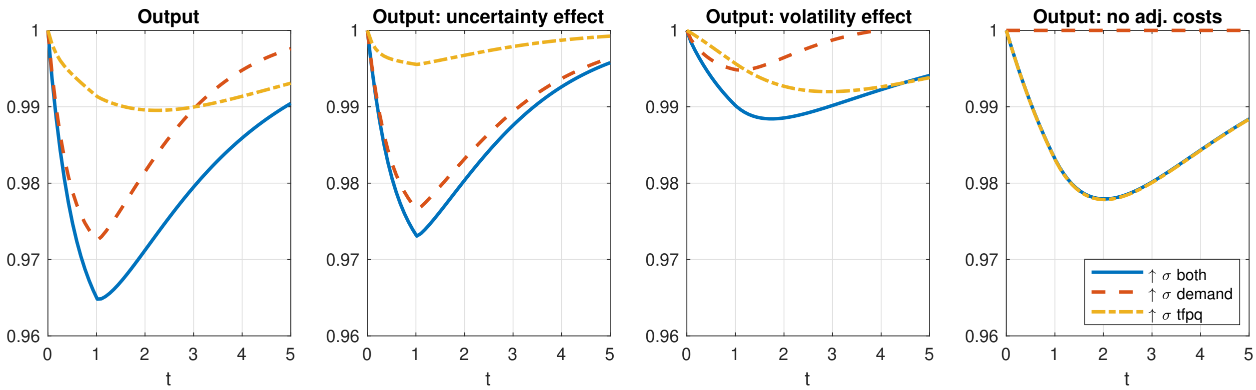
64. We have also simulated a permanent rise in uncertainty, and the results and intuitions are the same. Combined with the robustness using the Bloom et al. (2018) persistence values, the results appear robust to reasonable changes in the persistence of uncertainty shocks.

65. We plots the instantaneous inaction rate, which differs from the inaction rate as measured at, e.g., 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.

66. This is true even in response to a permanent rise in dispersion, and hence the result does not depend on the

Secondly, in the right panel of Figure 7 we plot the counterfactual paths for output from the "uncertainty" and "volatility" effects, as defined by Bloom (2009). The uncertainty effect simulates an economy where agents believe that uncertainty has increased and that they are in state $s = 2$, but where shocks are still drawn from the less uncertain distribution ($s = 1$) so that the realised volatility does not actually increase. Conversely, the volatility effect simulates an economy where firms still believe they are in the low uncertainty state ($s = 1$) but shocks are in fact drawn from the high uncertainty ($s = 2$) distribution. Intuitively, the uncertainty effect captures changes in firm behavior due to anticipation of higher uncertainty, in particular decreased investment due to wait and see behavior. The volatility effect instead holds firm behaviour constant and 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, meaning that both the fear of higher uncertainty *and* the effect of higher realised volatility contribute to lowering GDP in response to an uncertainty shock. This result is novel, as in Bloom (2009) the volatility effect is actually positive, and drives the medium term rise in overall output. Thus, our model amplifies the effect of uncertainty shocks by reversing the volatility effect, and we explain why in more detail below.

Figure 8: Model response to increase in demand or TFPQ uncertainty separately



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. 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.

To investigate the role of demand and TFPQ dispersion in driving the recession, we repeat our simulations in the counterfactual economies where only demand or TFPQ uncertainty are increased separately, The results are given in Figure 8. In each panel, the baseline calibration where uncertainty of both shocks increases is given in solid blue, and the alternative models in dashed red and dash-dotted yellow. The left panel plots output, and we see that the majority of the output decline is driven by the increased dispersion in demand. This explains almost all of the fall in output within the first year, and the majority of the fall for the first three years. Thus, our first result is that the negative first-moment effects of dispersion are mostly driven by demand dispersion, rather than supply dispersion.

persistence of shock.

Understanding wait and see behavior – The role of demand and productivity We next dig into the uncertainty effect, to see whether it is driven more by demand or TFPQ uncertainty. The second panel of Figure 8 decomposes the uncertainty effect into the role of each shock. Strikingly, increased demand uncertainty induces a large fall in output, of over 2%, while TFPQ uncertainty induces very little, at only 0.5%, implying that demand uncertainty is a much more important driver of wait and see effects. Partly this reflects that demand uncertainty rises more in the recession, but just as important is how firms react to each kind of uncertainty.

To understand why wait and see behavior responds less to TFPQ uncertainty, we return to the policy functions plotted in the left panel of Figure 6. A novel feature of our model is that optimal input size – measured by the position of the investment threshold – is very unresponsive to TFPQ shocks. This follows from our estimated non-CES demand curve: 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). This is the low passthrough we identified earlier, and since firms move their prices relatively little in response to TFPQ shocks, their quantity sold is also unresponsive, as is their input need. 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 to raise demand and sell more units. At some point, higher TFPQ causes optimal input scale to fall, as the firm is unable to increase the quantity of output sold as easily, and so increased productivity means the firm actually requires less units of inputs to produce the same amount of goods.[67]

Overall, optimal input use is very unresponsive to TFPQ in the non-CES model, with the investment threshold remaining near one across a wide range of values of productivity. Demand shocks, on the other hand, induce large changes in optimal scale, as seen in the centre panel of Figure 6. This is because demand shocks directly move how many units a firm can sell at a given price, and hence its target level of output, and the required capital and labour stocks to produce it. Since wait and see behaviour is driven by the fear of setting inputs at the wrong level, and consequently having to pay non-convex adjustment costs to correct your mistake, demand uncertainty drives more wait and see behavior. Intuitively, if prices are unresponsive to productivity, productivity uncertainty creates uncertainty about *markups*, while demand uncertainty creates uncertainty about *quantity sold*, and it the latter which drives wait and see behavior.

Understanding the volatility effect – the role of demand and productivity We now move on to investigating the volatility effect, which the third panel of Figure 8 decomposes into the separate effects of demand and TFPQ dispersion. In contrast to the uncertainty effect, it is TFPQ dispersion which is the main driver of the negative volatility effect, with a peak output fall of nearly 1%. While the effect from demand dispersion is slightly larger in the first year, it fades more quickly, leaving the more persistent impact of TFPQ dispersion to be the dominant force from year one onwards. 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 of these shocks.

To understand this difference, we first return to the simpler static-optimization problem of the model without adjustment costs. The final panel of Figure 8 plots the simulation of aggregate output in this model subject to the same shock processes as the full model. We see output declines

67. 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 23 in the appendix for details, and a comparison to the CES model.

by over 2% in response to the rise in dispersion. Considered independently, the rise in demand dispersion has exactly *zero* effect on output in the absence of adjustment costs, while the rise in TFPQ dispersion explains *all* of the decline. To understand why, a first order approximation to the solution of the static model yields the following expression for optimal sales:[68]

$$\log s \simeq (\theta - 1)\frac{\theta}{\theta + \eta}\log z + \epsilon \implies Y = \mathrm{E}\left[s\right] \simeq \mathrm{E}\left[z^{\frac{\theta(\theta-1)}{\theta+\eta}} e^{\epsilon}\right], \tag{20}$$

where each firm's optimal sales are $s \simeq z^{\frac{\theta(\theta-1)}{\theta+\eta}} e^{\epsilon}$, and averaging over firms yields aggregate output. First consider the model without demand uncertainty, and 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 $z^{\theta-1}$, raising 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 effect.[69] 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 $Y \simeq \mathrm{E}\left[z^{0.82}\right]$. Now a firm's optimal sales is actually *concave* in productivity. This means that lucky firms now expand by less than unlucky firms contract, and hence an increase in dispersion will output lower aggregate output. This result is, to the best of our knowledge, novel, and follows from the fact that incomplete passthrough introduces *markup dispersion* in response to increased productivity dispersion. In particular, following an increase in TFPQ dispersion, firms adjust their prices by less than their TFPQ adjusts. This leads firms to absorb the dispersion in increased markup dispersion, which introduces misallocation, which reduces aggregate output.[70] This moderates the Oi-Hartman-Abel effect and, for $\eta > \theta(\theta - 2)$, as we estimate, actually overturns it.[71]

Finally, we assumed that the demand shocks were drawn such that e^{ϵ} is normally distributed. This is because, 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

68. To do this, we combine our first order approximation of the firm's optimal price setting behavior, (18), 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.

69. Oi (1961), Hartman (1972), and Abel (1983).

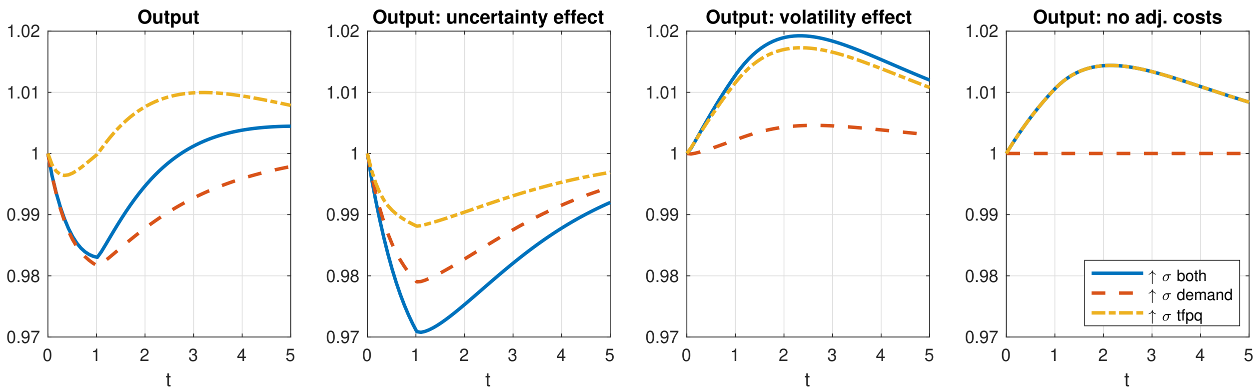
70. To see that it is markup dispersion that causes output to fall, rather than the non-linearity of the demand curve itself, recall that these first order approximations use the linear CES demand curve. For additional intuition, consider the efficient solution to the non-CES model without adjustment 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 $\eta > 0$.

71. This result is derived from a first order approximation, but we verify in Figure 23 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 8 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.

this distribution, an increase in the dispersion of ϵ has no effect on $\mathrm{E}\left[e^{\epsilon}\right]$, and hence on aggregate output. This assumption is natural, but also helpfully highlights that demand dispersion on its own does not necessarily lead to first order effects on output in the same way that TFPQ dispersion does. Overall, the key difference here is that TFPQ dispersion is passed through a demand curve, creating nonlinear effects, while demand dispersion is not. These nonlinear effects are positive for a CES demand curve, and negative for our estimated non-CES demand curve.[72]

Comparison to CES demand model To highlight how our model differs from a CES model, in Figure 9 we repeat our main plot but for the model solved under the assumption of CES demand ($\eta = 0$). Following the discussions above, in the third panel we see that the volatility effect is now positive, rather than negative, for TFPQ shocks. Moreover, since firms adjust their prices and output more to TFPQ shocks under CES, the uncertainty effect from TFPQ shocks is now over twice as large (second panel, dash-dotted yellow). The changes to how firms respond to demand shocks are smaller, with the large uncertainty effect almost unchanged. Overall, non-CES demand amplifies the total effect of uncertainty shocks on aggregate output by around 40% (peak fall of 2.5% versus 1.75%) mainly by overturning the offsetting Oi-Hartman-Abel effect.

Figure 9: Alternative CES model: response to increase in demand or TFPQ uncertainty



The plots give the aggregate response of the CES ($\eta = 0$) 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. 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.

Time-varying passthrough In Table 35 in the appendix, 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,

72. Is 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. For both demand and TFPQ dispersion, the volatility effect (third panel) differs from the model with no adjustment costs (fourth panel). The difference 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.

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. See the appendix for more details.

Robustness We perform several robustness exercises in the appendix. Firstly, we show the results are robust to using Bloom et al.'s (2018) higher persistence of uncertainty shocks. Secondly, we provide an alternative model where demand shocks affect the elasticity of demand, which we calibrate to fully match the passthrough from demand shocks to prices. This reduces the size of the wait and see effect from demand shocks, as firms use demand shocks to adjust prices more and hence quantity sold and input requirements less. Finally, 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 over the business cycle. In particular, we consider the distinct roles of demand and productivity and provide three novel contributions to our understanding of the cyclicality of dispersion and uncertainty.

First, we document that both demand dispersion and physical total-factor-productivity (TFPQ) dispersion are countercyclical. We are able to measure demand and TFPQ shocks separately since we observe prices and the degree of factor utilization at the firm level. This is a unique feature of our analysis which allows us to go behind revenue productivity (TFPR) measures and refine our understanding of cyclical dispersion. Importantly, we find that demand dispersion is more cyclical than TFPQ. In addition, accounting for utilization reduces the cyclicality of TFPQ dispersion. This suggests that demand is a prominent driver of dispersion over the business cycle, which we confirm using variance decomposition exercises.

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. This suggests that "real rigidities" can be economically important. In fact, deviations from CES help rationalize the finding of incomplete passthrough from TFPQ shocks to prices and provides a natural explanation for: i) the relative unimportance of TFPQ dispersion in driving the cyclical dispersion of endogenous variables in our variance decompositions, and ii) the aggregate first-order effects of uncertainty shocks in our model exercises.

Finally, we embed our estimated demand curve into a heterogeneous-firm model with non-convex input adjustment costs, following Bloom et al. (2018). We use this model to study the aggregate effects of idiosyncratic dispersion (and uncertainty) of both productivity and demand shocks. In the model, the non-constant elasticity of demand dramatically shapes the transmission of uncertainty shocks to aggregate first moments such as real GDP. 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 does not because firms allow their markups to fluctuate. Nevertheless, TFPQ dispersion is still harmful for aggregate output even though the uncertainty effect is limited. The reason is that it induces markup dispersion which leads to misallocation.

Our results highlight how measuring and modelling demand can help us understand firm behavior and the business cycle. Future work could investigate additional implications of these results, such as their implications for price rigidities, the cyclicality of markups, or firm-level factor utilization. As we provide a direct estimate of the degree of "real rigidities" coming from firms' demand curves, a natural next step is to combine our model with work on nominal rigidities. This would provide a data-disciplined investigation of whether real rigidities can generate meaningful monetary non-neutralities, extending the work of Ball and Romer (1990), Kimball (1995), and Klenow and Willis (2016).

References

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2022. "Income and wealth distribution in macroeconomics: A continuous-time approach." *The Review of Economic Studies* 89 (1): 45–86.

Altig, Dave, Scott Baker, Jose Maria Barrero, Nicholas Bloom, Philip Bunn, Scarlet Chen, Steven J Davis, Julia Leather, Brent Meyer, Emil Mihaylov, et al. 2020. "Economic uncertainty before and during the COVID-19 pandemic." *Journal of Public Economics* 191:104274.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2018. "The Elusive Pro-Competitive Effects of Trade." *The Review of Economic Studies* 86 (1): 46–80.

Armelius, Hanna, Isaiah Hull, and Hanna Stenbacka Köhler. 2017. "The timing of uncertainty shocks in a small open economy." *Economics Letters* 155:31–34.

Bachmann, Rüdiger, and Christian Bayer. 2013. "'Wait-and-See' business cycles?" *Journal of Monetary Economics* 60 (6): 704–719.

———. 2014. "Investment dispersion and the business cycle." *American Economic Review* 104 (4): 1392–1416.

Ball, Laurence, and David Romer. 1990. "Real Rigidities and the Non-Neutrality of Money." *The Review of Economic Studies* 57 (2): 183–203.

Barrero, Jose Maria, Nicholas Bloom, Steven J Davis, and Brent H Meyer. 2021. "COVID-19 is a persistent reallocation shock." In *AEA Papers and Proceedings,* 111:287–91.

Basu, Susanto. 1996. "Procyclical productivity: increasing returns or cyclical utilization?" *The Quarterly Journal of Economics* 111 (3): 719–751.

Basu, Susanto, and Brent Bundick. 2017. "Uncertainty Shocks in a Model of Effective Demand." *Econometrica* 85 (3): 937–958.

Berger, David, Ian Dew-Becker, and Stefano Giglio. 2019. "Uncertainty Shocks as Second-Moment News Shocks." *The Review of Economic Studies* 87 (1): 40–76.

Berger, David, and Joseph Vavra. 2019. "Shocks versus Responsiveness: What Drives Time-Varying Dispersion?" *Journal of Political Economy* 127 (5): 2104–2142.

Blackwood, G Jacob, Lucia S Foster, Cheryl A Grim, John Haltiwanger, and Zoltan Wolf. 2021. "Macro and Micro Dynamics of Productivity: From Devilish Details to Insights." *American Economic Journal: Macroeconomics* 13 (3): 142–72.

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

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

Carlsson, Mikael. 2017. "Microdata Evidence on the Empirical Importance of Selection Effects in Menu-Cost Models." *Journal of Money, Credit and Banking* 49 (8): 1803–1830.

Carlsson, Mikael, Julián Messina, and Oskar Nordström Skans. 2021. "Firm-level shocks and labour flows." *The Economic Journal* 131 (634): 598–623.

Carlsson, Mikael, Julián Messina, and Oskar Nordström Skans. 2016. "Wage adjustment and productivity shocks." *The Economic Journal* 126 (595): 1739–1773.

Carlsson, Mikael, and Oskar Nordström 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 (4): 1571–95.

Cette, Gilbert, Nicolas Dromel, Rémy Lecat, and Anne-Charlotte Paret. 2015. "Production factor returns: the role of factor utilization." *Review of Economics and Statistics* 97 (1): 134–143.

Davis, Steven J, John Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction.* 1st ed. Vol. 1. The MIT Press.

De Loecker, Jan, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik. 2016. "Prices, markups, and trade reform." *Econometrica* 84 (2): 445–510.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2020. "Changing Business Dynamism and Productivity: Shocks versus Responsiveness." *American Economic Review.*

Den Haan, Wouter J., Lukas B. Freund, and Pontus Rendahl. 2021. "Volatile hiring: Uncertainty in Search and Matching Models." *Journal of Monetary Economics,* ISSN: 0304-3932.

Eslava, Marcela, and John Haltiwanger. 2020. "The Life-cycle Growth of Plants: The Role of Productivity, Demand and Wedges." *NBER Working Paper,* no. 27184.

Félix, Sónia, and Chiara Maggi. 2021. "What is the Impact of Increased Business Competition." *Working paper.*

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?" *American Economic Review* 98 (1): 394–425.

———. 2016. "The Slow Growth of New Plants: Learning about Demand?" *Economica* 83 (329): 91–129.

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

Gopinath, Gita, Oleg Itskhoki, and Roberto Rigobon. 2010. "Currency choice and exchange rate pass-through." *American Economic Review* 100 (1): 304–36.

Haltiwanger, John, Robert Kulick, and Chad Syverson. 2018. "Misallocation Measures: The Distortion That Ate the Residual." *NBER Working Paper,* no. 24199.

Heyman, Fredrik, Helena Svaleryd, and Jonas Vlachos. 2013. "Competition, Takeovers, and Gender Discrimination." *ILR Review* 66 (2): 409–432.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and firm Dynamics in Long Run Equilibrium." *Econometrica* 60 (5): 1127–1150.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity." *The Quarterly Journal of Economics* 131 (3): 1291–1364.

Kaas, Leo, and Bihemo Francis Kimasa. 2021. "Firm dynamics with frictional product and labor markets." *International Economic Review* 62 (3): 1281–1317.

Kehrig, Matthias. 2015. "The Cyclicality of Productivity Dispersion." https://doi.org/http://dx.doi.org/10.2139/ssrn.1854401.

Khan, Aubhik, and Julia K. Thomas. 2008. "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics." *Econometrica* 76 (2): 395–436.

Kimball, Miles S. 1995. "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit, and Banking* 27 (4).

Klenow, Peter J., and Jonathan L. Willis. 2016. "Real Rigidities and Nominal Price Changes." *Economica* 83 (331): 443–472.

Lindé, Jesper, and Mathias Trabandt. 2018. "Should we use linearized models to calculate fiscal multipliers?" *Journal of Applied Econometrics* 33 (7): 937–965.

Melander, Ola. 2009. "The effect of Cash Flow on investment: an empirical Test of the balance Sheet Channel."

Mongey, Simon, and Jerome Williams. 2017. "Firm Dispersion and Business Cycles: Estimating Aggregate Shocks Using Panel Data." *Working paper,* http://www.simonmongey.com/uploads/6/5/6/6/65665741/mongey_williams_firm_dispersion_business_cycle.pdf.

Pozzi, Andrea, and Fabiano Schivardi. 2016. "Demand or Productivity: What Determines Firm Growth?" *RAND Journal of Economics,* 608–630.

Shapiro, Matthew D. 1993. "Cyclical productivity and the workweek of capital." *The American Economic Review* 83 (2): 229–233.

Smeets, Valerie, and Frederic Warzynski. 2013. "Estimating productivity with multi-product firms, pricing heterogeneity and the role of international trade." *Journal of International Economics* 90 (2): 237–244.

Vavra, Joseph. 2014. "Inflation Dynamics and Time-Varying Volatility: New Evidence and an Ss Interpretation." *The Quarterly Journal of Economics* 129 (1): 215–258.

APPENDICES

A Data Construction Appendix

A.1 Data sources and construction

In this section we describe our data sources and raw variables, data cleaning and assembly, and samples. Variable construction and data description are presented in their own sections.

A.1.1 Sources

To construct a firm-level panel dataset containing accounting, price, and utilization data, we combine microdata at the firm level, plant level, and product level. Our main firm-level data are from the *Företagens Ekonomi* (FEK) survey. This survey includes variables such as sales, number of employees, expenditure on labor and investments, capital stock, inventory, and raw materials and intermediates. Data on product-level prices and quantities are retrieved from the Statistic Sweden's *Industrins Varuproduktion* (IVP) survey. Product in this survey are specified at the 8-digit level according to the Combined Nomenclature (CN). Data related to utilization and the business environment are taken from the *Konjunkturstatistik för Industrin* (KFI) survey, which is quarterly survey of managers.

We provide background about each data source in the next subsections. Additional documentation is available via the webpage of Statistics Sweden. High level descriptions of the datasets and variables tend to be available in English. However, detailed documentation, including sampling methodology, is available only in Swedish.

A.1.2 Firm register data

Our firm register data come from Statistics Sweden (SCB) and the Swedish Tax Authority (Skatteverket). We are able to combine the various datasets because persistent firm and plant identifiers are used across surveys. For a significant number of firms, we have access to a rich set of variables. These variables range from basic bookkeeping information, to detailed information about individual products, about production and market conditions, and even about innovation activity.

For firm bookkeeping variables, we have almost universal coverage of the Swedish industrial firms. Variables such as turnover and number of employees are compiled from tax returns. For other surveys, only a sample of firms is available. These other surveys use various sampling schemes and typically aim for representativeness with respect to economic activity (i.e. sector) and firm size. The one exception is large firms. The largest firms are often deliberately included (i.e. not sampled) or over-sampled. This over-representation of large firms is true of all of our register datasets. As a consequence, large firms tend to be more persistent in our data than small firms.

We conduct our analysis at the firm level as defined by legal accounting entity. However, data from manufacturing firms are often reported at a subsidiary level, typically defined by mutually exclusive geographic location. We refer to these production units as plants. A single firm may be comprised of multiple plants and it is typically possible to produce 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) Many of our firm variables are from the *Företagens Ekonomi* (FEK) survey. The FEK survey is compiled on an calendar year basis and includes information on sales, number of employees, expenditures on labor, gross and net investment, capital stock for structures and equipment, use of raw materials, and inventories. The sample that we have access to covers the period 1996 to 2013. This entails about 14 million firm-year observations. Summary statistics for main bookkeeping variables on a per employee basis are show in subsection A.3.2.

The FEK are collected in order to describe the non-financial structure of the economy and to facilitate comparisons over time with respect to production, investments, and profitability. Among other uses, the FEK data are used to construct national accounts. The FEK conform to the European Union's *Structural Business Statistics* (SBS) framework. This ensures the consistency of specific target variables with respect to international standards. Documentation in English can be found here: Structural business statistics. The main variables in the FEK survey come from financial statements collected by the Swedish Tax Agency (*Skatteverket*). Basic firm information, balance sheet, and income statement data are delivered to the Swedish Tax Agency via a standardized statement of accounts (*standardiserade räkenskapsutdrag*, SRU) that can be automatically constructed from financial statements based on the BAS chart of accounts (*BAS-kontonplanen*) used by 95% of firms in Sweden. This ensures a high degree of standardization and comparability across firms. In addition, SCB directly surveys data from the largest firms as determined by criteria such as employment, sales and number of production facilities. For example, SCB directly surveyed 576 firms in 2009. Direct collection of data guarantees the quality of the data for the most important companies.

The FEK data includes basic bookkeeping variables for the universe of active firms. Active firms are identified by having made a tax payment in a given year, e.g. payroll tax or VAT payments. Although a small number may fail to get counted because they enter the target population very late in the calender year, these exceptions are not consequential for the representativeness of the sample (see the following SCB report: Addressing coverage and measurement errors using multiple administrative data sources).

Although basic information is available for nearly all firms, detailed information about revenue, investments and assets is only collected for a subsample of firms. These data are collected via auxiliary surveys which SCB then integrates into the FEK data. For example, the SpecRR survey—which includes data on turnover at the product level—collects data based on its own stratified sampling scheme based on sectors, which then get integrated into the main FEK data. In particular, we use information on assets and investments to construct our capital variable.

The main variables that we take from this survey are presented below. We include the notation we use throughout the paper (though we suppress firm and year subscripts), the English and Swedish names of the raw variable, and a short description.

- S, Net sales (*Nettoomsättning*): Net sales—also know as total net turnover—is given by gross sales minus allowances, discounts, and returns. This variable corresponds to the total value of market sales of goods and services. This variable excludes financial income and income classified as other operating income; it also excludes operating subsidies received from public authorities or the European Union.

- l, Number of employees (*Antal anställda*): Number of employees refers to the average number of employees converted to full-time equivalents.
- C^l (Total personnel costs, *Summa personal kostnader*): Personnel costs include salaries and other direct remuneration, as well as taxes and employees' social security contributions retained by the employer, and employer's compulsory and voluntary social contributions.
- C^M, Raw materials, consumables, and goods for resale (*Summa kostnader for ravaror och handelsvaror*): This is our cost of goods sold (COGS) variable. The item includes the acquisition value of raw materials and merchandise, as well as costs associated with subcontracting.
- I (Total gross investment, *Summa brutto investeringar*): Gross investment in tangible goods, mainly property, plant, and equipment. Included are new and existing tangible capital goods with a useful life of more than one year. This includes non-produced tangible goods such as land. This varaible excludes investments in intangible and financial assets.
- E (Plant, machinery, equipment and tools, *Maskiner och inventarier*): This item includes plant, machinery and other technical equipment and tools for production as well as equipment, tools, fixtures and fittings.
- I^E, gross investment in plant, machinery, equipment and tools (*Brutto investeringar maskiner och inventarier*): The year's purchases of tangible fixed assets related to machinery, equipment, or plant. This variable includes new purchases and expenditures that permanently raise the value of assets.
- B, Buildings, land improvements and land (*Byggnader, markanläggningar och mark*): This item includes buildings, land improvements and land.
- I^B, Gross investment in buildings and land (*Brutto investeringar byggnader och mark*): The year's purchases of buildings and land recorded under tangible assets. This variable includes new purchases and expenditures that permanently raise the value of the assets.
- D, Change in stocks of work in progress, finished goods and work on contract (*Förändring av lager av produkter i arbete, färdiga varor*): The item summarises changes in inventory and work in progress. This arises because of the lag between production and invoicing. Changes due to normal obsolescence are included (i.e. depreciation of an item due to it being damaged or obsolete or similar).
- T, Inventories, etc. (*Varulager m.m.*): This item includes all kinds of goods held in stock and goods and services produced or provided for on own account or on behalf of others. Examples can be raw materials, semi-finished goods, commodities or securities.

Goods price data: Production of Commodities and Industrial Services (IVP) A crucial part of our analysis involves the use of a firm-specific price index. Only a few other studies related to dispersion have access to such measures. Our price data come from the Production of Commodities and Industrial Services survey (*Industrins varuproduktion*, IVP). 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. Specifically, the IVP dataset provides product level revenue, quantity and price data, collected at the plant level. All firms with at least 20 employees are included, although smaller firms are included for certain sectors, in certain years, or if they have sufficiently high revenue. Product classification is based on the EU's Combined Nomenclature (CN) at the 8-digit level. For certain goods, Sweden also provides an additional alphabetic digit of differentiation, thus resulting in some nine-digit identifiers that consist of a CN code plus a letter. We thus have price data at a fine-grained level.

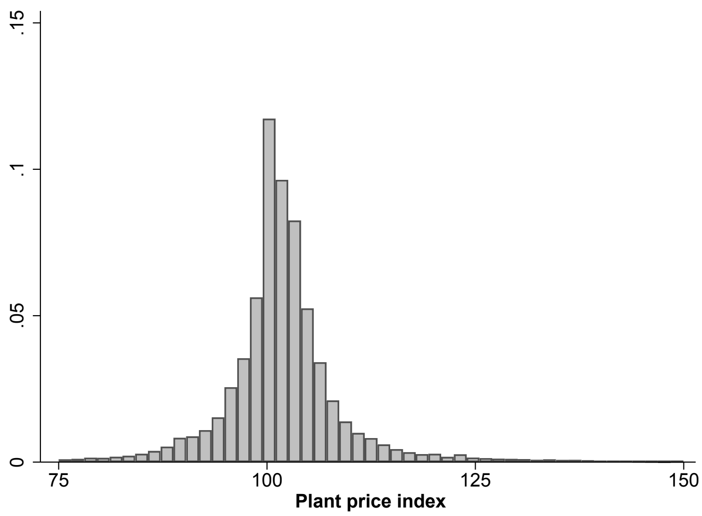
We use the IVP data to construct firm level price indices. Since most firms produce multiple goods, it is necessary to aggregate up from good-specific prices to a firm-specific price index. We do this in two different ways, though both approaches yield similar results. In our main approach, we aggregate plant level indices to the firm level using sales weights. For robustness, we also construct firm indices by aggregating directly at the product level across plants. In both cases, we use chained indices. This is necessary approach because most firms adjust their product portfolio over time. Because the 8-digit CN specification is quite fine-grained, incidental changes in classification are not infrequent. Using chained indices thus facilitate year to year comparisons.

In our main approach, we rely on plant level price indices (*ArbstIndex*) provided in the IVP data. These plant level indices are computed as chained Laspeyres indices based on goods level data. Specifically, the chained index $\tilde{P}_{j,t}$ for plant j in period t is computed:

$$\tilde{P}_{j,t} = \frac{\sum_k^K P_{k,j,t} Q_{k,j,t-1}}{\sum_k^K P_{k,j,t-1} Q_{k,j,t-1}}.$$

where $P_{k,j,t}$ and $Q_{k,j,t}$ denote respectively the price and quantity of product k at plant j in year t (we omit the firm index i). 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. The distribution is shown in Figure 10.

Figure 10: The 1% to 99% distribution of plant price indices



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:

$$\tilde{P}_{i,t} = \sum_j \frac{S_{j,t}}{\sum_j^J S_{j,t}} \tilde{P}_{j,t}.$$

where $S_{j,t}$ denote the sales of plant j in year t. Note that we re-base the plant price indices to 1 at the beginning of a series of consecutive series of firm observations, i.e. set the price associated with the first observation to $\tilde{P}_{j,t_0} = 1$, where t_0 indicates the year in which a firm "run" appears (as we explain below, we define a firm as a stable set of plants). The cumulative price change between the first period in which a firm is observed and period t is then given by

$$P^f_{i,t} = 1 \cdot \tilde{P}_{i,t_0+1} \cdot \ldots \cdot \tilde{P}_{i,t-1} \cdot \tilde{P}_{i,t}.$$

$P^f_{i,t}$ is our raw measure of firm-level price.

For robustness, we also construct firm-level indices by first aggregating revenue and quantities across products. This approach yields similar results overall, and identical results for single plant firms. However, we favor using SCB's *ArbstIndex* because the data coverage in our plant level sample is more comprehensive than data coverage at the product level.

It is important to note that we cannot control for differences across firms associated with differences in quality or exact product definition. For this reason, we are careful never to directly compare prices across firms in our empirical work. We use firm fixed effects, first differences, or normalisations in order to soak up permanent differences in prices across firms which could be due to differences in product definition or quality. Our estimates only rely on relative price changes within the firm over time. We posit that from one year to the next, changes in quality within a small are likely to be small enough so as not to significantly bias our estimated demand shocks.

In the IVP data, there are a small number of apparently redundant observations for which two distinct plant identifiers within the same firm are associated with identical production, sales, and price data in a specific year. We handle these redundant observations by retaining the plant-year data associated with the longest running plant (if both plants are present for an equal number of years, we keep the plant with the highest id number).

Utilization Data: Industrial Capacity Utilization survey We also use variables from SCB's Industrial capacity utilisation survey (Industrins kapacitetsutnyttjande) which is part of the Business Cycle Statistics for Industry survey (*Konjunkturstatistik för industrin*, KFI). In this survey, production facility managers evaluate various standardized measures of the business environment, including degree of capacity utilization. The Industrial Capacity Utilization survey is conducted quarterly and employs a stratified sampling scheme based on industry and firm size. In total, each survey covers 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. During the period 1998-2009, the target population included all firms with at least 10 employees. Unfortunately, from 2010 and onwards, utilization is imputed for firms with less than 50 employees. Additional information about this survey can be found on SCB's website: scb.se/en/data-collection/surveys/business-cycle-statistics-for-industry/. At the sector level, the degree of capacity utilization is available at the quarterly frequency back to at least 1990.

We rely on two variables from the business cycle survey in particular. The first is a measure of capacity utilization (*kapacitetsutnyttjande*). This variable 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.[73] Importantly, prevailing production

73. There is a small degree of variation in the precise wording across surveys. In some versions of the survey, staffing

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 argument is relevant for furloughs as compared to planned downsizing. Furloughs reflect changes in capacity utilization while downsizing reflects a change in the baseline level of full utilization. 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 second variable that we use from the business cycle survey is an indicator of whether low capacity utilization is primarily the result of "insufficient demand" (*otillräcklig efterfrågan*). Unfortunately, the meaning of "insufficient demand" is not made explicitly clear.

To aggregate the utilization and insufficient demand variables to the firm level, we compute the average across production facilities using revenue weights.[74] To convert the quarterly data to the annual frequency, we average the firm-level observations within the year. This yields two firm-level variables:

- u, average firm-level capacity utilization: The average level 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.
- $\mathcal{I}(\check{\epsilon})$, share of plants reporting insufficient demand: This variable ranges between 0 and 1. Overall, about 30% of firms report insufficient demand at all plants.

Table 5: Relationship between utilization and insufficient demand

	$\ln u$
$\mathcal{I}(\check{\epsilon})$	-0.154***
	(0.00501)
$\mathcal{I}(\check{\epsilon}) \times recession$	-0.109***
	(0.0142)

This table presents the relationship between utilization and insufficient demand estimated in our main sample (15,042). $\ln u$ denotes the logarithm of firm-level utilization, $\mathcal{I}(\check{\epsilon})$ denotes insufficient demand, and $\mathcal{I}(\check{\epsilon}) \times recession$ is an interaction between the insufficient demand variable and an indicator for the years 2001 and 2009. The regression includes firm and sector-year fixed effects. Standard errors are clustered at the firm level and given in parentheses. Both coefficients are significant at the 0.001 level as indicated three (***) stars.

There is a strong relationship between these two variables at the firm level. In Table 5, we show regressions of utilization on the insufficient demand variable and an interaction between

is not explicitly mentioned in the definition. However. that staffing should be taken into account seems evident given that managers are told to evaluate capacity utilization based on the working hours and shifts that can be considered normal

74. 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.

insufficient demand and a "recession indicator" for the years 2001 and 2009. In general, 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.

A.1.3 Other data sources

Besides the datasets described in the main text, we also use variables from a number of other sources. This includes prices indices from Statistics Sweden and depreciation rates taken from the literature:

- P_t^s, producer price index (PPI): For our sectoral price index, we use the producer price index available via Statistics Sweden's Statistical Database: Prisindex i producent- och importled (PPI). The PPIs are defined for activities and can be matched to sectors at the 2-, 3-, and 4-digit level. However, because of the small size of certain sectors, the 3-digit and 4-digit specification is not always available. As noted above, the PPIs are constructed from the IVP data.
- P_t^i, investment price index: We construct our investment price index based on price changes for gross fixed capital investment. These price changes are available from SCB at the two-digit sector level: Fasta bruttoinvesteringar.
- δ_t^s, depreciation rates: Depreciation rates 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.

We also use data from a few other sources—including the Community Innovation Survey—in conjunction with auxiliary exercises. We discuss these datasets together with those exercises.

A.1.4 Data cleaning and assembly

We aggregate our data to the firm-year level. Quarterly data, such as our utilization data, are averaged within calendar year. For multi-plant firms, we weight plant level data by the share of firm sales whenever we are interested in firm averages. An important example is our firm level price index which is constructed from a weighted average of plant level price indices.

Our main analyses rely on panel techniques and comparisons over time. This means that we use data from firms that are observed in two or more consecutive periods. 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 practice, we therefore define our own "firm" panel identifiers. We 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. This means that our firm identifiers 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. The idea is to remove plants that make only a limited contribution to overall firm activity. 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. This enables us to retain a number of large firms in our data. Note that removal of marginal plants affects primarily the price and utilization data which are based on the plant level averages (revenue and total employment data are measured directly at the firm level). Pruning will therefore only have an effect if prices growth or utilization is are systematically different among mariginal plants as compared to retained plants. Since this does not appear to be the case, we conclude that pruning improves the representativeness of our dataset because it facilitates the inclusion of more large firms.

We perform several rounds of data cleaning and data preparation. We drop a small number of observations that are nonsensical, implausible, or missing data on key variables. For example, we discard observations that report positive values when only negative values make sense in our analysis. We also drop observations for which key variables are zero. Specifically, we drop data with negative sales, negative tangible fixed assets, less than 1 employee, or zero in total personnel costs. In total, this entails dropping only a small fraction of the data.

The presence of extreme values is a more challenging issue. Because most of our analyses are at the firm-level, the natural way to identify and clean extreme values is based on changes at the firm (or plant) level. It is likely that some extreme increases or decreases in a given variable reflect either miscoding or a categorical change to the firm that is un-observable in the data.

- **Bookkeeeping data**: For our bookkeeping data, we handle extreme values based on per employee growth in the key variables. 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 $\chi_{i,t}$:

$$\chi_{i,t} = \sqrt{\sum_x dx_{i,t}^2}.$$

  χ is large when one or more of the variables exhibits a large change. Note that we construct χ based on variables that we expect to be fairly stable over time. For example, a ten-fold increase in the level of capital is difficult to reconcile with a model of the firm as a stable production unit. In contrast, we do not trim based on utilization data or inventories because it is plausible that certain durable goods manufacturers could experience an enormous drop in production during the Great Recession.

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.

- **Price data**: Some of the price data appear to be affected by mismeasurement. One source of mismeasurement relates to changes in units. For instance, products occasionally change their units from kilograms (kg) to thousands of kilograms (1000 kg). In the presence of such a change, we would expect about a thousand-fold change in the quantity as compared to earlier years. For some observations, however, the expected change in the order of magnitude only shows up after a couple years. The result is mismeasurement in the quantity data. To handle issues with the price data, we therefore trim observations associated with the 1.5% most extreme firm level price changes in every year.

Assembly At the overarching level, we construct our data in two stages. In the first stage, we combine plant-level datasets and aggregate to the firm level. This includes price and utilization data. Then, in the second stage, we combine these data with our firm-level bookkeeping data and perform sample selection. Overall, there are three main bottlenecks. 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 arises from the availability of price data. We use the price data to get TFPQ. This cuts the sample down to about 50,000 observations. Finally. the third bottleneck is our utilization data which we use to improve our TFPQ measure. This leaves us with about 15,000 observations in total. We describe the steps in our data construction below.

- Our price indices cover 101,951 plant-year observations. The price indices are defined as chained Laspeyres indices.
- Our utilization data are also given at the plant level. However, the utilization data are given at the quarterly frequency. To convert to the annual frequency, we use the simple average of quarterly observations. After doing so, the utilization data cover 60,442 plant observations. Because the sampling scheme varies between the price data and the utilization data, only 38,481 plant-year observations include both price and utilization data.
- To aggregate price and utilization data to the firm level, we weight plant observations by their share of firm sales. We construct these sales weights using product level production data from the IVP data. We also discard marginal firms. Marginal firms are firms that only account for a negligible fraction of firm sales. After discarding the marginal firms, we are left with 91,465 plant observations for which price data are available and 37,278 observations for which both price and utilization data are available.

- Once we collapse the plant level data to the firm level, we are left with 63,983 firm level observations with price data and 20,190 firm level observations with both price and utilization data.
- In the last step before we merge the plant level datasets with the firm data, we build firm price indices based runs of consecutive firm-year observations. At this juncture we clean extreme observations from the price data (we do not perform any cleaning of the plant level price data). We trim 1.5% of the most extreme values from each tail of the price change distribution. This leaves us with 62,401 usable firm observations of which 19,258 include utilization data.
- Our firm bookkeeping data cover 13,821,831 observations. We merge the bookkeeping dataset to the aggregated plant level data and then combine it with the various sectoral price indices.
- The firm bookkeeping data contain some nonsense observations. We drop observations for which turnover is negative, labor expenditures or number of employees is zero, or raw materials is missing or has the wrong sign. This results in a loss of 1,537 observations.
- Our analysis focuses on manufacturing firms and we drop observations associated with sectors unrelated to production. In total, we drop 4,080 observations in sectors related to either services, construction or mining. We also drop sectors 9, 12, and 19 because these sectors include only a limited numbers observations. The criteria that we use for sectors is that there be at least 4 firms in every year. Note that the median number of observations in a sector-year is 183.
- We construct our measure of capital stock based on the perpetual inventory method. Because investment data is available for only a subset of data, we are forced to drop another fraction of the data. When a continuous run of observations is interrupted due to missing capital data, we assign a new firm identifier and re-base the within firm capital and price series.
- As explained above, we handle extreme values from the bookkeeping data by constructing a multidimensional measure of firm growth based on growth in capital, value-added, raw materials and expenditure on labor. Whenever this measure increases or decreases sharply, we define a new firm run.
- We keep only firm runs of at least two consecutive years. This means dropping observations that are missing either price data, capital data, or that have consecutive extreme changes in the bookkeeping variables. At the conclusion of our data construction, we have 48,047 observations for which price data is available and 15,044 observations for which both price and utilization data is available.

A.1.5 Samples

We work with a number of different samples. The reasons are two-fold. The first reason is that utilization data is only available for a limited number of firms. The second reason is that there may be differences between unbalanced and balanced samples. In particular, results based on the balanced samples reflect the sample of firms that survive over time. Specifically there are four samples that we use at various junctures.

- *Full sample*: The "full" sample is based on observations for which both investment and price data are available—but not capacity utilization data. This sample thus requires only the FEK and IVP data. This sample is an unbalanced panel for the period 1996-2013. It includes about 48,000 observations. As the most representative sample, we use this data to characterize the overall development of the Swedish economy and to estimate cost shares (which we use to compute TFP measures).
- *Main sample*: Our main sample is subsample of the Full sample for which capacity utilization data is available. The utilization data is only available since 1998 and covers a limited number of firms. In addition, the utilization sample is biased toward large firms. Overall, this sample covers the period 1998-2013 and includes about 15,000 observations. The average firm size is also substantially larger than in the full sample.
- *Balanced sample*: Our third sample is the 12-year balanced subsample of firms based on the Full sample. This sample describes the set of firms that are continuously present throughout the period 1999-2010. We use this sample for robustness exercises.

Table 6 presents the number of observations per year for each sample, and the total number of observations N. In total we have 48,047 observations for which price data is available. When we limit the sample to firms that also have utilization data available, we have about 15,000 observations. When we limit the sample to the balanced sample, we have about 8,700 observations.

Table 6: Number of firms per year by sample.

	Full	Main	Balanced
1996	2717		
1997	2879		
1998	2717	885	
1999	2663	985	726
2000	2655	982	726
2001	2689	1103	726
2002	2785	1146	726
2003	2747	1077	726
2004	2718	1033	726
2005	2730	973	726
2006	2718	978	726
2007	2763	973	726
2008	2781	972	726
2009	2699	823	726
2010	2554	839	726
2011	2518	768	
2012	2507	808	
2013	2207	699	
N	48047	15044	8712

How many unique firms are present in the various samples? In our largest sample, we have close to 8,000 unique observations with an average panel length of about 6. When limiting the sample to include utilization data, this falls to 3,181 observations, with an average panel length of about 5. If we require a balanced panel, our analysis includes 726 observations per year overall, and 520 firms that have both price and utilization measures present.

A.2 Construction of key variables

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

Some variables can be used without adjustment (e.g. employees l) or after deflation by a price index. 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:

- Real cost of labor: $c^l = \frac{C^l}{P^s}$
- Real cost of raw materials: $m = \frac{C^M}{P^s}$

We also use the sectoral price index P_t^s to compute our measure of relative price:

- Relative price: $p = \frac{P^f}{P^s}$

We construct our output and capital variables based on their economic definitions. For gross output and value added we deflate the nominal value by the firm price index. This is similar to the approach proposed by Smeets and Warzynski (2013):

- Real output: Nominal gross output, Q, is total production that year, computed as the value of sales, S, plus the change in inventories D, giving $Q \equiv S + D$. To get units of output, we then deflate using the firm price index $P^f f$.

$$q = \frac{S + D}{P^f}$$

- Real value-added: Our value-added measure is the difference between the output measure and the cost of raw materials, consumables, and goods for resale (C^M). Real value added is gross output less the value of raw materials, $M_{i,t}$ deflated by the firm price.[75]

$$v = \frac{S + D - C^M}{P_f}$$

SCB also provides there own production and value-added measures. Whether we use our own measures or those provided by SCB has only a minor impact on our quantitative results. We favor our own measures because we can be certain of their definitions.

Our real capital stock measure, k, is based on combined capital stock for structures and equipment. We construct our capital series for structure and equipment using a perpetual inventory

75. This formulation implicitly deflates a firm's intermediate inputs using their own sales price, rather than the price index of the intermediates. However, the price of a firm's intermediates is rarely known, meaning that intermediates must either be deflated by a firm's own price, or some aggregate price. We deflate intermediates by the firm's own price when estimating value added production functions, and by the sectoral price level when estimating gross output production functions. This produces similar results.

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. Although we have the book value of capital $K_{i,t}$ in each year, the accounting value of capital tends to be a poor measure of the amount of capital used in production. In particular, firms have an incentive to depreciate the value of capital because this reduces taxable income.

The main question when applying this method is how to initialize the capital series. Many of our firms are only present in the data for a limited number of periods. Because of this, the initial level of capital has a persistent effect on the stock of capital. To address this issue, 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 as the sum of real investment plus the depreciated value of capital stock, unless the book value in a given period exceeds the value given by the PIM.

Sectors Sector classification is important as we often de-mean or include sector fixed effects in our analyses. Swedish firms are classified according to the Swedish Standard Industrial Classification (SNI). 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. The SNI system associates with each type of economic activity a numeric code. The first two digits specify activity at an aggregated level. Each additional digit provides more specificity. For example, the manufacture of motor vehicles, trailers and semi-trailers is grouped under code 29, while the manufacture of only motor vehicles is specified under code 29.10. The SNI classification also include a fifth digit that provide an additional level of articulation. For example, in the SNI system code the manufacture of passenger cars and other light motor vehicles (29.101) is distinguished from the manufacture of trucks and other heavy motor vehicles (29.102).

We focus on SNI sectors 10-33 which comprise industrial production (NACE group C "manufacturing"). The sectoral distribution of observations is shown in Table 7. 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.

When using the SNI codes, two main challenges arise in practice. The first relates to changes in the classification system. During the period of study, the SNI classification has been revised twice: There was a minor revision in 2002 and a more substantial revision in 2007. This means that the identifier of a given economic activity may have changed over time. The second challenge arises when firms are reclassified from one sector to another. This means that the same firm may be associated with two distinct SNI codes.

To handle revisions of the SNI codes, we harmonize all SNI codes to the 2007 standard using reconciliation tables provided by SCB (note that the SNI92 classification must be first updated to the SNI02 standard before updating to the 2007 standard). In many cases, this harmonization is straightforward. Occasionally, however, codes are either "merged" or "split." For example, a code in the SNI02 scheme may be associated with two or more distinct codes in the SNI07 classification (a splitting of a code). In such cases, a decision must be made about how the earlier codes will be updated. We handle this in two ways. If a firm is present across revisions, then we can exploit this to harmonize the SNI codes retroactively to correspond to the most recent classification. This

Table 7: Number of observations by two-digit SNI sectors.

10 Manufacture of food products	1406
11 Manufacture of beverages	86
13 Manufacture of textiles	272
14 Manufacture of wearing apparel	102
15 Manufacture of leather and related products	64
16 Manufacture of wood and products of wood and cork, except furniture;...	1352
17 Manufacture of paper and paper products	907
18 Printing and reproduction of recorded media	133
20 Manufacture of chemicals and chemical products	862
21 Manufacture of pharmaceutical products and pharmaceutical preparations	159
22 Manufacture of rubber and plastic products	830
23 Manufacture of other non-metallic mineral products	591
24 Manufacture of basic metals	721
25 Manufacture of fabricated metal products, except machinery and equipment	1342
26 Manufacture of computer, electronic and optical products	634
27 Manufacture of electrical equipment	646
28 Manufacture of machinery and equipment not elsewhere classified	1707
29 Manufacture of motor vehicles, trailers and semi-trailers	915
30 Manufacture of other transport equipment	316
31 Manufacture of furniture	753
32 Other manufacturing	471
33 Repair and installation of machinery and equipment	593

task is made easier by the presence of *both* SNI02 and SNI07 codes in the 2008 data. However, in the absence of such information, we reconcile the SNI code to the code most commonly observed among firms in the first year of the revision. This means that codes from SNI92 that are split in the 2002 revision are associated with the most commonly observed code in 2002, and similarly for SNI02 codes that are split in the 2007 revision.

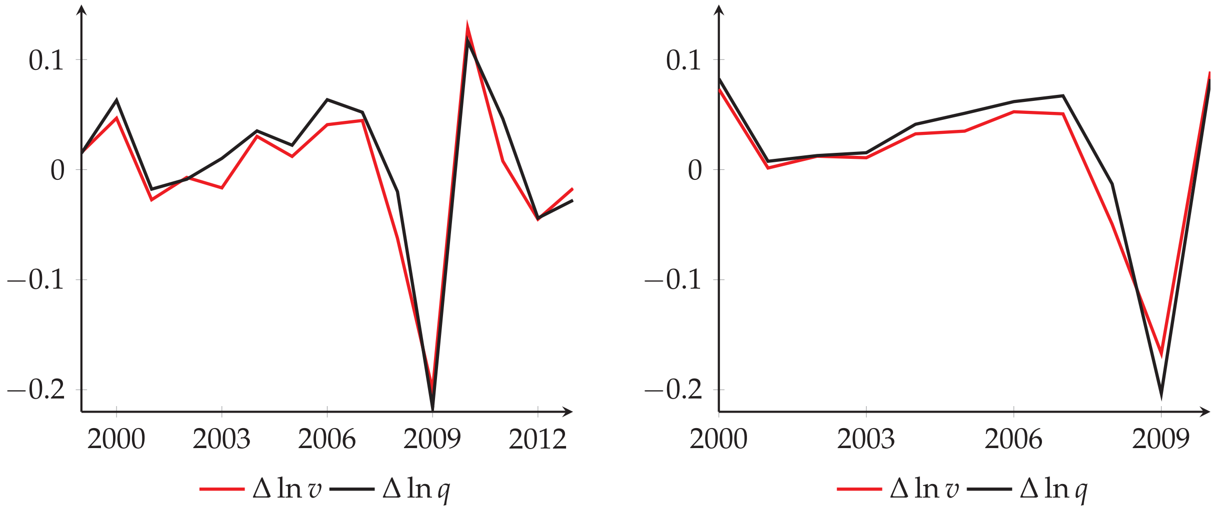
Although most firms are consistently identified with a single sector, in cases where the sector identification changes, we pick the most commonly observed sector with a firm run. We thus use firm identifiers to harmonize industry codes within consecutive series of observations, picking the most commonly observed sector affiliation.

A.3 Data description

A.3.1 Output

We measure firm output by real value-added $v_{i,t}$. We also have a measure of gross output $q_{i,t}$. We present value added growth $\Delta \ln v$ together with output growth $\Delta \ln q$ in Figure 11. Value added is show in red and output in black. The left panel shows the average growth in our main sample, while the right panel shows average growth in our balanced panel. Value added and output track each other closely. A comparison of the main sample, on the left, and the balanced panel, on the right, shows similar developments over time. In both samples, there is weak growth around 2001 and a deep contraction in 2009.

Figure 11: Aggregate output, main and balanced samples



(a) Main sample

(b) Balanced panel

The red line shows the average annual firm-level growth in value added $\Delta \ln v$. The black line shows the average annual firm-level output growth $\Delta \ln q$. The left panel is based on our main sample. The right panel is based on the balanced sample.

The dispersion of firm output growth as measured by the interquartile range is shown in Table 12 for our main sample. The IQR is computed after removing sector-year growth. There is a small increase in the dispersion of value added in 2001 and a sharp increase in dispersion in 2009. Notably, dispersion remains high, and actually reaches its highest level in 2010. The reason for the large increase in 2010 appears to be a rebound effect in which firms that experienced negative growth in 2009 experience positive growth in 2010.

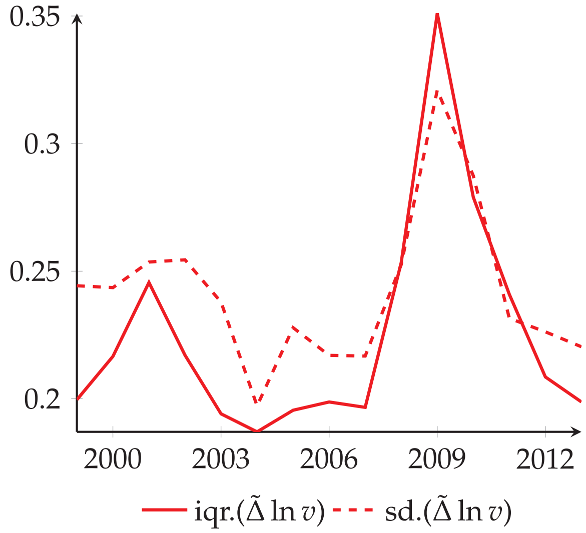
A.3.2 Firm variables

Table 8 shows descriptive statistics for basic firm variables: Employees (l), sales (S), real value added (v), capital (k), and intermediates (m). We present each variable relative to the number of firm employees (with the exception of number of employees). For examples, sales per worker is computed as S/N. This facilitates comparisons across firms and samples. All values are reported in units of 1000 SEK.

The top panel presents basic statistics for each sample. We include the mean of each variable, the value at the 25th-percentile in the distribution, and the value at the 75th-percentile in the distribution (denoted p25 and p75, respectively). Most of the variables are comparable across samples on a per employee basis. The most systematic difference between the samples relates to the presence of utilization data. The main sample—which includes utilization data while the full sample does not—includes larger firms on average. The other notable feature is that the mean values are often larger than the 75-th percentile. This is due to the fact that most variables have a substantial positive skew.

In the bottom panel, we present more detailed information about our main sample. We present the 1st-percentile, the 25th-percentile, the median, the 75th-percentile, the 99th-percentile, and

Figure 12: Dispersion of value added growth



The dispersion of firm-level value added growth $\Delta \ln v$ measured by the interquartile range is presented in solid red. The dispersion of firm-level value added growth $\Delta \ln v$ measured by the standard deviation is presented in dashed red. Firm-level growth is de-meaned by sector-year prior to computing the dispersion statistics.

skewness and kurtosis statistics. What is most clear, is the substantial heterogeneity across firms. For instance, the largest 1% of firms are about 20 times larger than the median firm. This long right tail of firms is also reflected in the skewness statistic.

Prices and utilization Table 9 presents descriptive statistics for firm-level prices and firm-level capacity utilization. Firm-level price P^f and firm-level relative price p are presented in terms of growth: $\Delta \ln P^f$ and $\Delta \ln p$. We present the data in terms of growth because price comparisons are only meaningful within firm.

The nominal prices P^f exhibit positive growth over time on average (mean growth of 1.68% and median growth of 0.95%), substantial kurtosis, and positive skewness. Large price increases are more common than large price decreases. Nevertheless, negative nominal price growth is often observed. At the 25th-percentile, firms are reducing prices by nearly 2%. Relative prices p, 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.

For utilization, we present both the degree of utilization u (in percent) and utilization growth $\Delta \ln u$. As indicated in the bottom row, it is fairly common to observe less than 100% utilization. The average utilization is about 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.

The distributions of relative price changes and firm utilization are show in Figure 13.

How common are price changes? How common is full utilization? In Table 10 we address a number of such questions. The percentage of observations for which we do not observe any change in the nominal firm price is 3.3%. If we consider the share of observations for which the

Table 8: 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		

	Main						
	p1	p25	p50	p75	p99	skewness	kurtosis
employees	18	55	107	246	2442	14	245
sales	600	1260	1819	2758	9047	7	115
value added	249	578	790	1124	3183	56	3186
capital	57	409	780	1515	12329	9	143
intermediates	117	607	1000	1742	7335	12	295

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.

nominal price change is less than $\pm$ 1%, this increases to 21.1%. In other words, nominal prices change by more than 1% for about 80% of firms. With respect to utilization, firms report no change in the degree of utilization about 18% of the time ($\Delta \ln u = 0$). About a quarter of observations are associated with 100% or more utilization ($u \geq 100$).

Cyclicality Most firm variables are characterized by countercyclical volatility. This pattern is clear during the Great Recession, when dispersion increases sharply for sales, prices, employment, intermediates, and utilization. The pattern holds during the 2001 recession, though the tendency is less strong overall and does not hold for prices. The main exception to countercyclical dispersion is investment. Investment exhibits a degree of procyclicality during the 2001 and 2009 recessions, but exhibits somewhat ambiguous cyclicality overall.

We illustrate the cyclicality of firm variables based on growth rates in Figure 14. Growth rates are computed as log changes for all variables apart 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. In the main text, we present the time series of the interquartile range. Figure 14, is the same figure as in the main text, but based on the standard deviation rather than the interquartile range. The results are similar.

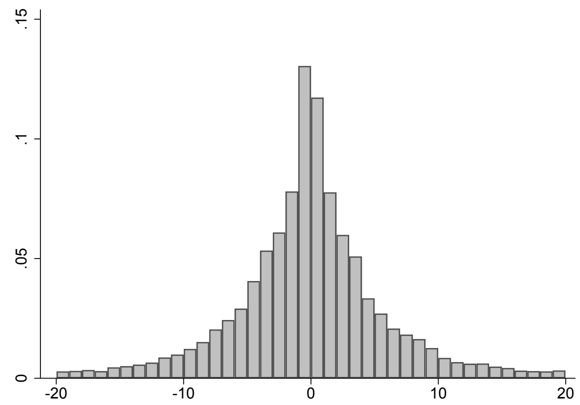
We provide quantification of the cyclicality in Table 11. For each variable, the table presents the change in the dispersion of the variable in 2001 and 2009 relative to the average across all

Table 9: Summary statistics for prices and utilization

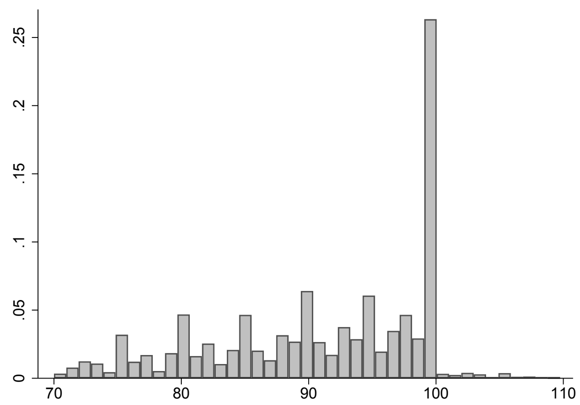
	mean	sd	p1	p25	p50	p75	p99	skewness	kurtosis
$\Delta \ln P^f$	1.68	8.42	-19.87	-1.80	0.95	4.38	31.14	1.21	10.13
$\Delta \ln p$	0.01	7.95	-23.77	-3.09	-0.04	2.81	25.96	0.32	9.04
Δu	-0.42	13.77	-43.29	-3.69	0.00	3.30	37.16	-0.59	45.36
u	87.93	13.31	43.50	80.50	91.00	99.05	105.00	-1.43	6.13

Summary statistics for prices and utilization are based on our main sample. $\Delta \ln P^f$ 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.

Figure 13: Distributions of relative price changes and capacity utilization based



(a) Firm-level relative price growth, $\Delta p_{i,t}$



(b) Firm-level utilization, u

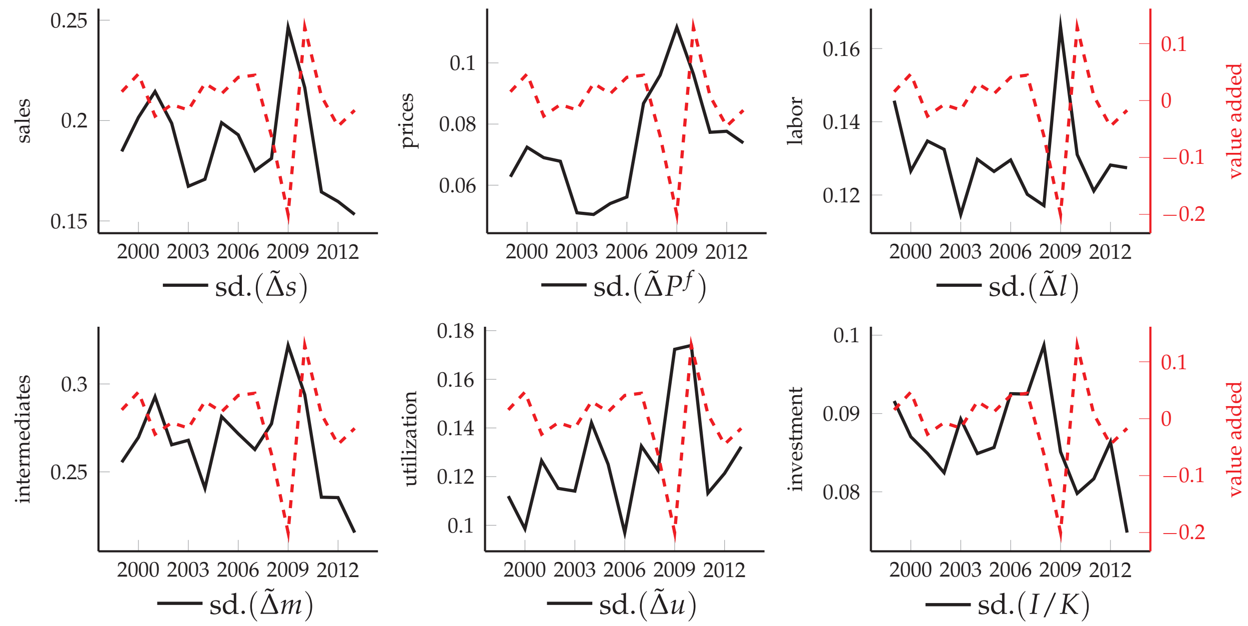
This figure presents the distributions of relative price changes and degree of capacity utilization based on our main sample. The left panel shows relative annual relative price changes expressed in percent. The right shows degree of utilization also expressed in percent.

other years when excluding those two years. In the Table 11(a), we present the change using the standard deviation. In Table 11(b), we present the change based on the interquartile range. For example, the first row on the left tells us that the standard deviation of sales growth increased 18% in 2001 and 33% in 2009 measured relative to the average in 1998-2013 when excluding 2001 and 2009.

There are only 5 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.

Skewness and kurtosis measures are presented in Table 12. These statistics are presented for three periods: The years 2001 and 2009, and the remaining "non-recession" periods. Note that these are not comparisons (as in Table 11), but the actual statistics. The reason is that skewness is occasionally close to zero in the non-recession period, which makes percentage changes difficult

Figure 14: Dispersion of firm variables, 1999-2013 (standard deviation)



This figure is the equivalent to Figure 1 in the main text, but measuring dispersion using the standard deviation rather than interquartile range. It shows the standard deviation (sd) across firms of log changes for key variables, calculated each year. The standard deviations are computed within sector as $\text{iqr}_t(\tilde{\Delta} x_{i,t})$ for each variable $x \in \left\{s, P^f, l, m, u, k\right\}$. Sales s is given by firm turnover deflated by a sectoral producer price index. Price P^f 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. And capital k is computed according to a perpetual inventory approach. Complete descriptions of each variable are provided in the appendix. To indicate the Swedish business cycle, each plot also includes the growth rate of aggregate value added v, defined as turnover plus changes in inventory and unfinished goods minus the value of intermediate goods.

Table 10: Price and utilization basic statistics

$\Delta \ln P^f = 0$	3.3 %
$\lvert\Delta \ln P^f\rvert < 1\%$	21.1 %
$\Delta \ln u = 0$	18.0 %
$u \geq 100$	24.1 %

$\Delta \ln P^f = 0$ denotes the share of observations for which nominal prices are the same as in the previous year. $\lvert\Delta \ln P^f\rvert \leq 1\%$ is the share of observations for which the nominal price change is less than $\pm 1\%$. $u \geq 100$ indicates the share of firms with at least full utilization. And $\Delta \ln u = 0$ is the percentage of observations for which utilization is the same as the previous year. All statistics are based on our main sample.

Table 11: Cyclicality of firm variables

	2001	2009
sd.($\tilde{\Delta} s$)	17	35
sd.($\tilde{\Delta} P^f$)	-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.($\tilde{\Delta} P^f$)	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 while the right table provides the same comparisons but based on the interquartile range. All measures have been de-meaned by sector-year.

to interpret.

In general, it is difficult to identify systematic developments with respect to skewness and kurtosis. There is perhaps some tendency that the skewness becomes more negative during the recessions—which is the case for sales, prices, and utilization in both 2001 and 2009—there are many exceptions. Likewise, there is some tendency toward lower kurotsis during recessions, but this only holds for some variables.

Table 12: Skewness and kurtosis of firm variables

		non-recession	2001	2009
ΔS	skewness	-0.01	-0.25	-0.32
	kurtosis	11.02	10.91	4.86
ΔP^f	skewness	0.50	-0.18	0.25
	kurtosis	8.50	6.79	5.92
Δl	skewness	-0.38	0.31	-0.86
	kurtosis	10.53	14.31	12.04
Δm	skewness	0.02	0.36	-0.49
	kurtosis	11.29	11.19	6.33
Δu	skewness	-0.54	-1.31	-1.15
	kurtosis	53.70	13.00	6.34
i/k	skewness	2.30	1.83	3.07
	kurtosis	11.36	7.04	18.15

This table presents skewness and kurtosis measures for main firm variables. Statistics are presented for 2001, for 2009, and based on all other years excluding 2001 and 2009 (“non-recession”).

B Productivity and Demand Estimation Appendix

B.1 Productivity shock estimation

B.1.1 Estimating the production function

The starting point for our productivity analysis is the quantity production function for firm i in sector j

$$\log v_{i,t} = z_{i,t} + \gamma_{K,j} \ln k_{i,t} + \gamma_{L,j} \ln l_{i,t}.$$

Here, $v_{i,t}$ is real value added computed using a firm specific price—i.e. $v_{i,t}$ a measure of production quantity—and $\beta_{k,j}$ and $\beta_{l,j}$ denote sector-specific output elasticities for capital and labor. For this specification, $z_{i,t}$ describes quantity productivity (TFPQ). We refer to $z_{i,t}$ as "raw" TFPQ because capital and labor aren't adjusted for the intensity of factor usage. To get an improved measure of TFPQ, we therefore include a utilization adjustment:

$$\ln v_{i,t} = z^u_{i,t} + \gamma_{K,t}(\ln u_{i,t} + \ln k_{i,t}) + \gamma_{L,t}(\ln u_{i,t} + \ln l_{i,t}).$$

$z^u_{i,t}$ is our favored measure of TFPQ. If production is approximately constant returns to scale ($\gamma_{K,j} + \gamma_{L,j} = 1$), then there is a simple relationship between raw TFPQ $z_{i,t}$ and $z^u_{i,t}$:

$$z^u_{i,t} = z_{i,t} - \ln u.$$

In other words, our favored TFPQ measure utilization is equivalent to a simple adjustment of raw TFPQ.

The productivity measure $z^u_{i,t}$ relies on a simple utilization adjustment. We also consider an alternative way to account for utilization that is both more flexible and incorporates additional business cycle information. We denote this alternative productivity measure by $z^{u,p}_{i,t}$. We compute this utilization adjustment by regressing "raw" TFPQ on a fourth order polynominal of $\ln u$, an insufficient demand variable, ($\mathcal{I}_{i,t}$), a recession indicator (r), the interaction $\mathcal{I}_{i,t} \times r$, and the pairwise interaction between the four utilization terms and the three other variables (18 terms in total).[76] In other words, we estimate $z^{u,p}_{i,t}$ based on the expression

$$z_{i,t} = z^{u,p}_{i,t} + \beta_u \ln u + \beta_{u^2}(\ln u)^2 + \beta_{u^3}(\ln u)^3 + \beta_{u^4}(\ln u)^4 + \beta_{\mathcal{I}}\mathcal{I}_{i,t} + \beta_r r + \beta_{r\times\mathcal{I}}(\mathcal{I}_{i,t} \times r) + \text{interactions}.$$

This utilization adjustment allows for a more flexible interaction between reported utilization than the simple utilization adjustment. It also takes advantage of other business cycle information. Both of these changes will in principle . Although we find some evidence that this measure more effectively accounts for utilization—and thus discriminates better between between productivity and demand—the conclusions are the similar for both utilization measures. Notably, the $z^{u,p}_{i,t}$ results tend to be intermediate between the results based on $z_{i,t}$ and the results based on $z^u_{i,t}$. For instance with respect to the cyclicality of the dispersion of TFP shocks (which we discuss below), $\tilde{\Delta} z_{i,t}$ exhibits the largest increase in dispersion during recessions, $\tilde{\Delta} z^u_{i,t}$ exhibits the smallest increase in dispersion during recessions, and $\tilde{\Delta} z^{u,p}_{i,t}$ exhibits an increase in dispersion that is between that observed for raw TFPQ z and utilization adjusted TFPQ z^u.

We also construct a set of revenue productivity measures ("TFPR") that are analogous to our TFPQ measures. The difference is that we use revenue rather than output when we estimate our

76. For brevity, we use the notation $\mathcal{I}$ for the insufficient demand variable rather than the $\mathcal{I}(\check{e})$ notation from above).

TFPR measures. For example, we estimate a TFPR measure $a_{i,t}$ based on the model

$$\ln s_{i,t} = a_{i,t} + \gamma_{K,j} \ln k_{i,t} + \gamma_{L,j} \ln l_{i,t}.$$

where $s_{i,t}$ is value added deflated by a sectoral price index. The TFPR measure $a_{i,t}$ is equivalent to our "raw" TFPQ measure $z_{i,t}$. In similar fashion, we compute utilization adjusted TFPR measures $a^{u}_{i,t}$ and $a^{u,p}_{i,t}$. These TFPR measures are interesting in comparison to our TFPQ and also in comparison to the literature.

Cost share approach We estimate the output elasticity of capital and labor based on the cost shares. This approach is widely employed in the literature. Our firm-level cost of labor is given by the real personnel costs, $c^{l}_{i,t}$. Our measure of 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. This rate is based on the cost of a typical basket of consumer goods and services in a given year.
- $\Delta_{Aaa,t}$: A 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.

To get the sector specific output elasticities, we then compute 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:

$$\gamma_{K,j} = \frac{\sum_t \sum_{J(i)} c^{k}_{i,t}}{\sum_t \sum_{j(i)} c^{k}_{i,t} + c^{l}_{i,t}}$$
$$\gamma_{L,j} = \frac{\sum_t \sum_{J(i)} c^{l}_{i,t}}{\sum_t \sum_{j(i)} c^{k}_{i,t} + c^{l}_{i,t}}.$$

Table 13 shows the sectoral cost shares that we use to compute our productivity measures in combination with $l_{i,t}$ and $k_{i,t}$. We estimate the cost shares on our Full sample rather than our main sample because the Full sample includes over three times more observations—including additional observations of firms in the main sample. For example, sector 10 (Manufacture of food products) has over 5000 observations in the full sample, but only about 1400 in our main sample. We list the number of observations in each sector alongside $\gamma_{K,j}$ and $\gamma_{L,j}$.

Control function estimation We also estimate an aggregate production function using a control function methodology. The control function results are not dissimilar from the cost share results. When using the cost share approach, we estimate an average capital elasticity of 0.265 and average labor elasticity of 0.735. When using the control function approach, we find capital elasticities that

Table 13: Cost shares by two-digit sector

	Obs	$\gamma_{K,j}$	$\gamma_{L,j}$
10	5089	0.674	0.326
11	206	0.582	0.418
13	781	0.721	0.279
14	330	0.856	0.144
15	137	0.858	0.142
16	4723	0.620	0.380
17	1876	0.449	0.551
18	450	0.711	0.289
20	1913	0.626	0.374
21	318	0.602	0.398
22	2911	0.725	0.275
23	2068	0.753	0.247
24	1466	0.697	0.303
25	6564	0.765	0.235
26	1908	0.882	0.118
27	2098	0.864	0.136
28	6377	0.815	0.185
29	2758	0.665	0.335
30	772	0.877	0.123
31	2395	0.787	0.213
32	1682	0.821	0.179
33	2964	0.818	0.182

This table presents cost shares for capital ($\gamma_{k,j}$) and labor ($\gamma_{l,j}$) computed for each of the 22 industrial sectors that are present in the main sample. The first column denotes the two-digit sector, the second column presents the number of observations in the sector, and the last two columns present the cost shares. The capital cost share is computed based on user cost of capital in constant prices. The labor cost share is based on total remuneration to labor in constant prices.

are slightly smaller—about 0.2 in our balanced panel—and labor elasticities that range from about 0.6 to about 0.9 depending on which sample and control function technique is chosen.

Concretely, when using the control function approach, we estimate a Cobb-Douglas production function using the Levinsohn–Petrin method including sector and year fixed effects. In all regressions, we use real value added $v_{i,t}$ as our measure of output, utilization capital as our "state" variable, (uk), utilization adjusted labor as our "free" variable, and the value of intermediates as our "proxy" variable. We perform this estimation in Stata using the `prodest` package. In Table 14, we presents results for three different samples with and without the Ackerberg–Caves–Frazer (acf) correction. "balanced u" indicates a 12-year sample in which utilization data is available for every firm in every year.

The results are affected by both the use of the acf correction and the choice of sample. The acf correction increases the labor elasticity and, via that channel, causes the estimates to be slightly increasing returns to scale rather than decreasing returns to scale (there is no effect on the estimated capital coefficient). This effect is present for all three samples. For example, comparing the first

Table 14: Control function estimation, inputs adjusted for utilization

	$\ln v_{i,t}$	$\ln v_{i,t}$	$\ln v_{i,t}$	$\ln v_{i,t}$	$\ln v_{i,t}$	$\ln v_{i,t}$
$\ln ul$	0.671***	0.925***	0.604***	0.855***	0.626***	0.692***
	(0.025)	(0.002)	(0.036)	(0.003)	(0.070)	(0.007)
$\ln uk$	0.158***	0.158***	0.191***	0.192***	0.262***	0.270***
	(0.001)	(0.001)	(0.017)	(0.004)	(0.049)	(0.006)
dataset	main	main	balanced	balanced	balanced (u)	balanced (u)
acf	no	yes	no	yes	no	yes

This table presents production function estimates using the Levinsohn-Petrin method. All regressions include sector and year fixed effects. $v_{i,t}$ denotes real value added, ul denotes utilization adjusted labor, and uk denotes utilization adjusted capital. Standard errors are clustered at the firm level and given 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 the next to last row, we indicate the choice of sample. In the bottom row, we indicate whether the Ackerberg–Caves–Frazer correction is used.

two columns, we see that the use of acf correction increases the labor estimate from 0.671 to 0.925. The choice of sample also plays a systematic role. When a balanced panel is used rather than an unbalanced panel, we estimate smaller labor elasticities and larger capital elasticities. For example, we estimate an elasticity of 0.925 in our main sample as compared to 0.855 for our balanced sample when using the acf-correction.

In Table 15 we present the same exercises as Table 14 but without adjusting the inputs for degree of utilization. We include these estimates for comparison to our main (utilization adjusted) results and to the wider literature for which utilization data has not typically be available.

Table 15: Control function estimation, inputs not adjusted for utilization

	$\ln v_{i,t}$	$\ln v_{i,t}$	$\ln v_{i,t}$	$\ln v_{i,t}$	$\ln v_{i,t}$	$\ln v_{i,t}$
l	0.754***	0.987***	0.683***	0.949***	0.628***	0.717***
	(0.027)	(0.000)	(0.025)	(0.000)	(0.080)	(0.005)
k	0.150***	0.139***	0.164***	0.164***	0.241***	0.292***
	(0.006)	(0.000)	(0.005)	(0.001)	(0.012)	(0.005)
dataset	main	main	balanced	balanced	balanced (u)	balanced (u)
acf	no	yes	no	yes	no	yes

This table presents production function estimates using the Levinsohn-Petrin method. All regressions include sector and year fixed effects. $v_{i,t}$ denotes real value added, l denotes units of labor, and k denotes units of capital. This specification does not include a utilization adjustment. Standard errors are clustered at the firm level and given 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 the next to last row, we indicate the choice of sample. In the bottom row, we indicate whether the Ackerberg–Caves–Frazer correction is used.

B.1.2 TFPQ Cyclicality

We quantify the cyclicality of TFPQ in Table 16. For each TFPQ measure, Table 16 presents the percentage change in dispersion in 2001 and 2009 relative to the average in all other years. For example, in 2009, the standard deviation of utilization adjusted TFPQ (z^u) is about 24.6% higher than the average. The dispersion measures are computed in the following fashion: For each TFPQ measure, we compute TFPQ shocks based on firm-level differences demeaned by the sector-year TFPQ growth. We then compute the standard deviation and interquartile range of TFPQ shocks in each year.

Table 16: Cyclicality of TFPQ dispersion

	2001	2009
sd.($\tilde{\Delta} z$)	5.8	29.2
sd.($\tilde{\Delta} z^u$)	-3.9	24.6
sd.($\tilde{\Delta} z^{u,p}$)	2.2	26.4

(a) Standard Deviation

	2001	2009
iqr.($\tilde{\Delta} z$)	9.1	47.5
iqr.($\tilde{\Delta} z^u$)	3.0	35.5
iqr.($\tilde{\Delta} z^{u,p}$)	6.7	46.9

(b) Interquartile Range

These tables presents percentage changes for TFPQ shock dispersion for the 2001 and 2009 recessions relative to the average over all other years. The left table shows the results for standard deviation while the right table show results based on the interquartile range.

The main take-away from Table 16 is that TFPQ dispersion increases during recessions. The only exception to this pattern is TFPQ (z^u) in 2001 when measured using the standard deviation. It is notable that the percentage change in dispersion is in all cases larger when measured using the interquartile range as compared to the standard deviation. Dispersion is not driven primarily by outliers. Consistent with intuition, we also find that controlling for utilization moderates measured changes in dispersion. Properly accounting for variation in the intensity of factor usage reduces mis-measurement of TFPQ in recession periods.

Table 17: Cyclicality of TFPR dispersion

	2001	2009
sd.($\tilde{\Delta} a$)	6.3	27.6
sd.($\tilde{\Delta} a^u$)	-4.6	22.7
sd.($\tilde{\Delta} a^{u,p}$)	2.3	24.4

(a) Standard Deviation

	2001	2009
iqr.($\tilde{\Delta} a$)	7.4	68.1
iqr.($\tilde{\Delta} a^u$)	4.8	39.0
iqr.($\tilde{\Delta} a^{u,p}$)	4.6	58.4

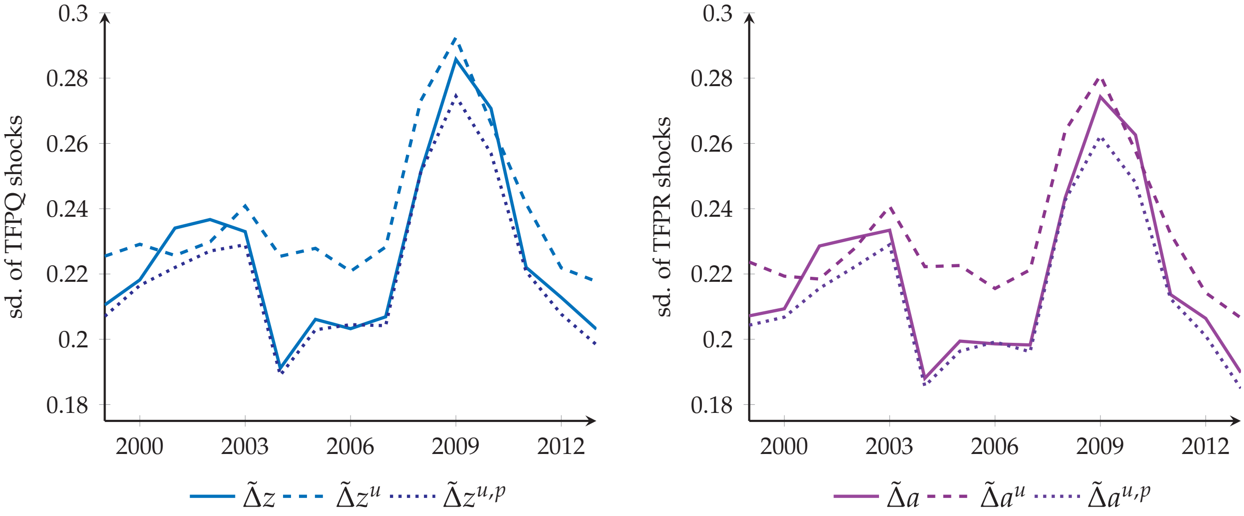
(b) Interquartile Range

These tables presents percentage changes for TFPR shock dispersion for the 2001 and 2009 recessions relative to the average over all other years. The left table shows the results for standard deviation while the right table show results based on the interquartile range.

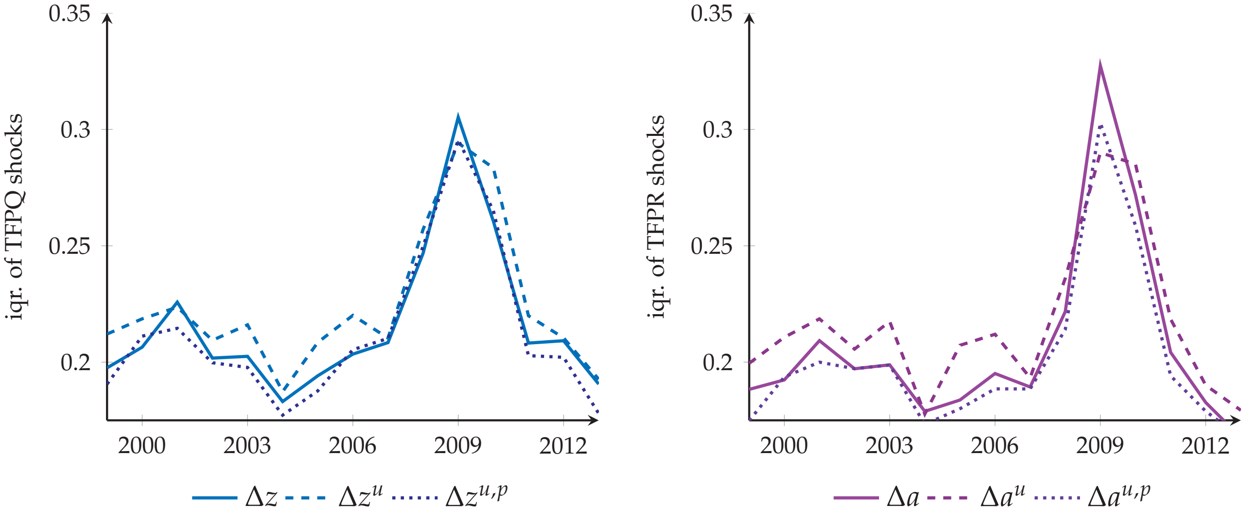
Table 17 presents measures of cyclicality for the TFPR shocks $\tilde{\Delta} a$, $\tilde{\Delta} a^u$, and $\tilde{\Delta} a^{u,p}$. Overall, results are comparable to those for TFPQ. As we found for TFPQ, we find that controlling for utilization plays an important role in reducing measured dispersion. In fact, controlling for utilization appears to play a more meaningful role than controlling for prices. In other words, dispersion of $\tilde{\Delta} a^u$ and $\tilde{\Delta} z^u$ are more similar than $\tilde{\Delta} z$ and $\tilde{\Delta} z^u$. This insight may be informative for other studies that hope to use TFPR as a proxy for TFPQ.

We illustrate how the dispersion of productivity shocks evolves over time in Figure 15. This figure presents the same dispersion measures as Table 16. The left side of the figure presents our

Figure 15: Volatility of TFP dispersion, 1999-2013



(a) Standard deviation



(b) Interquartile range

Each panel shows time series for dispersion of productivity shocks. The left side presents the TFPQ measures $\tilde{\Delta}z$, $\tilde{\Delta}z^u$ and $\tilde{\Delta}z^{u,p}$ while the right side presents the TFPR measures $\tilde{\Delta}a$, $\tilde{\Delta}a^u$ and $\tilde{\Delta}a^{u,p}$.

TFPQ measures while the right side presents the analogous TFPR measures. The top row of the figure presents dispersion measured by the standard deviation while the bottom row presents dispersion measured by the IQR.

Skewness The distribution of TFPQ shocks is close to symmetric in general. This can be seen in Table 18. This table presents skewness and kurtosis statistics. In the non-recession period and during 2009, skewness is always smaller in magnitude than 0.36. Although there is more skewness in 2001, it is still moderate. Hence, we find limited support for the conjecture that asymmetry in TFPQ shocks is driving asymmetry in outcomes.

Table 18: Skewness and kurtosis of TFPQ

		non-recession	2001	2009
$\tilde{\Delta} z$	skewness	-0.31	-0.93	0.08
	kurtosis	10.38	11.54	4.78
$\tilde{\Delta} z^u$	skewness	-0.20	-0.35	0.03
	kurtosis	12.28	8.40	5.03
$\tilde{\Delta} z^{u,p}$	skewness	-0.36	-0.69	0.08
	kurtosis	10.65	10.57	5.18

This table shows the skewness and kurtosis of TFPQ shocks during 2001, 2009, and the average for 1998-2013 when excluding the recession years ("non-recession").

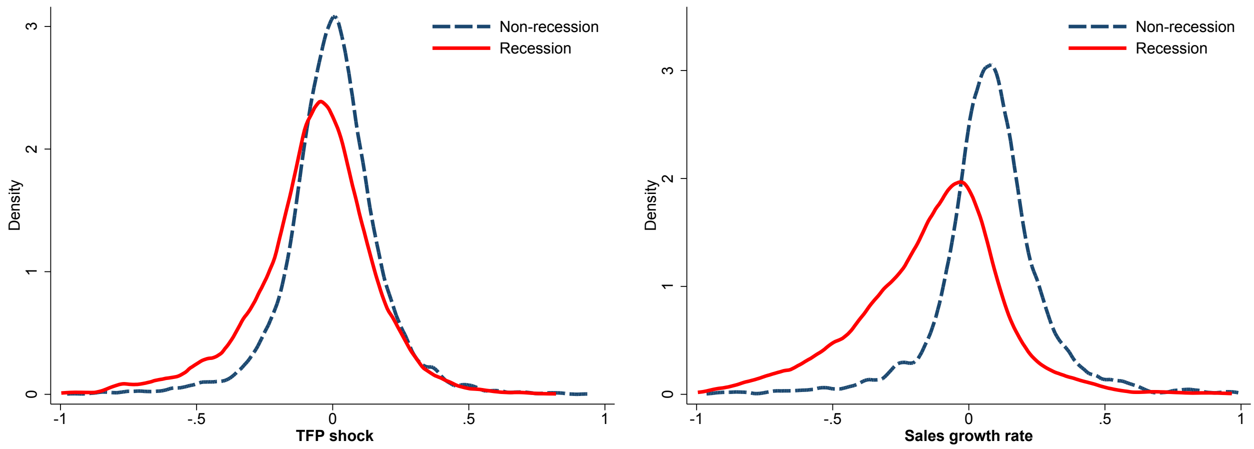
B.1.3 Replication of Bloom et al. (2018): Distributions during the Great Recession

For comparison with the literature, we reproduce Figure 1 and Figure 2 from Bloom et al. (2018). To do this, we estimate a log-AR(1) process for establishment-level revenue total-factor productivity (TFPR), computed using sectoral factor shares, and without utilization correction. We plot the distribution of the calculated innovations to the AR(1) process across firms pre-crisis (2005-6) in comparison with the crisis (2008-9) in the left panel of Figure 16. In the right panel we do the same for sales growth dispersion. We find similar results for the Swedish economy as Bloom finds for the US during the Great Recession. In the US, the variance of establishment-level TFPR shocks increases 76% in the recession. In Sweden, the the firm-level variance increases by 113%. For sales, the increase is 152% in the US, and 89% in Sweden.

B.1.4 TFPQ Autocorrelation

We present autocorrelations for our TFPQ measures in Table 19. All estimates are based on our main dataset. The first three columns present results for "raw" TFPQ while the last three columns present results for our utilization adjusted TFPQ measure. The difference across the columns is the choice of instrument. In columns 2, 3, 5, and 6, we use either the second lag or the third lag of the TFPQ measure as an instrument for the first lag. All regressions include sector-year fixed effects, use standard errors clustered at the firm level, and are estimated on the subsample of the main sample for which at least three lags are available.

Figure 16: Reproduction of Bloom et al. (2018) Figures 1 and 2



This figure reproduces Figures 1 and 2 from Bloom et al. (2018). The left hand side presents the distribution of TFP shocks while the right hand side presents the distribution of firm sales growth. TFP shocks are specified as the residuals from an AR(1) process, while TFP itself is measured using a cost share approach and a measure of output based on a sectoral price index (rather than a firm specific price). The red line shows the distribution in the recession (2008-2009) while the dashed blue line shows the distribution in a representative non-recession period (2005-2006). The distributions are estimated using a kernel density approach (command `kdensity` in Stata) using an Epanechnikov kernel function.

Table 19: TFPQ autocorrelation estimates

	$z_{i,t}$	$z_{i,t}$	$z_{i,t}$	$z^u_{i,t}$	$z^u_{i,t}$	$z^u_{i,t}$
$z_{i,t-1}$	0.909***	0.965***	0.985***			
	(0.011)	(0.008)	(0.008)			
$z^u_{i,t-1}$				0.908***	0.967***	0.978***
				(0.011)	(0.008)	(0.008)
iv	no	$z_{i,t-2}$	$z_{i,t-3}$	no	$z_{i,t-2}$	$z_{i,t-3}$

Because of possible firm heterogeneity, we also estimate the autocorrelation of TFPQ using a dynamic panel approach. As in the in Anderson-Hsiao approach, in Table 20 we regress the difference ($\Delta z^u_{i,t} = z^u_{i,t} - z^u_{i,t-1}$) on the lagged difference ($\Delta z^u_{i,t-1} = z^u_{i,t-1} - z^u_{i,t-2}$) and then use values from earlier periods as instruments. The choice of lags strongly affects the estimates. If we use $z^u_{i,t-2}$, $z^u_{i,t-3}$, or $z^u_{i,t-4}$, we get quite different values.

B.2 Demand shock estimation

B.2.1 Demand curve estimation

In our main results, we estimate demand using z^u as an instrument in the presence of firm and sector-year fixed effects. These choices are important. Below, we show how alternative choices with respect to instrumental variable and specification of fixed effects affect our demand results. In particular, the use of real productivity as an instrument and the inclusion of firm fixed effects are crucial for re-producing our estimates.

The choice of sample, in contrast, does not meaningfully affect our results. We find similar

Table 20: TFPQ autocorrelation estimates, Anderson-Hsiao method

	$\Delta z^u_{i,t}$	$\Delta z^u_{i,t}$	$\Delta z^u_{i,t}$
$\Delta z^u_{i,t-1}$	0.366***	0.731**	1.244**
	(0.0753)	(0.261)	(0.478)
iv	$z^u_{i,t-2}$	$z^u_{i,t-3}$	$z^u_{i,t-3}$

This table presents autocorrelation estimates for utilization adjusted TFPQ ($z^u_{i,t}$) based on an Anderson-Hsiao method. All regressions include sector-year fixed effects. Standard errors are clustered at the firm level and given in parentheses. Level of significance at that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively. The difference between the three columns in the choice of instrument. In the first column, the second lag is used as an instrument for $\Delta z^u_{i,t-1}$. In the second column, the third lag is used as an instrument for $\Delta z^u_{i,t-1}$. And in the third column, the fourth lag is is used as an instrument for $\Delta z^u_{i,t-1}$. The choice of instrument is specified in the bottom row.

estimates to those presented in the main text if we instead use a balanced sample. The same is true if we exclude the recessions. Conditional on the choice of specification, our results are stable. We also document this in the following subsection.

Choice of instrumental variable In the main text, we present demand estimates based on our main sample, using "raw" TFPQ and utilization adjusted TFPQ as instrumental variables (z and z^u). In Table 21, we compare the results for z and z^u to other choices with respect to instrumental variable. As in the main text, all specifications in this table include firm and sector-year fixed effects, use standard errors clustered at the firm level, and are based on our main sample. Choice of instrument (or lack thereof) is indicated in the bottom row.

Table 21: Demand estimation, the role of instrumental variables

	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$
$\hat{p}_{i,t}$	-0.713***	6.090***	-3.944***	-2.985***	-3.603***
	(0.0481)	(0.761)	(0.235)	(0.202)	(0.216)
iv	OLS	$a_{i,t}$	$z_{i,t}$	$z^u_{i,t}$	$z^{u,p}_{i,t}$

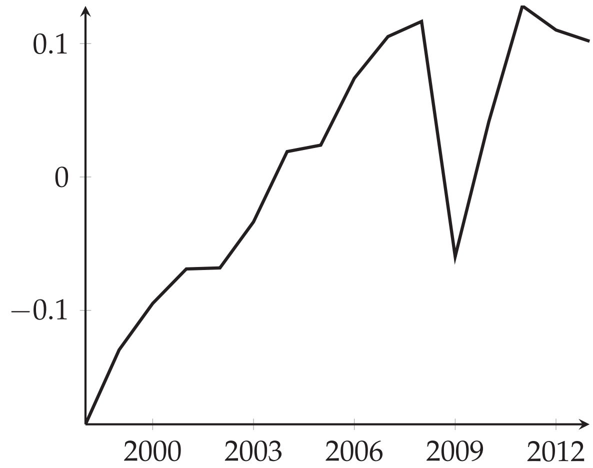
This table presents estimates of our baseline demand model based on different choices with respect to choice of instrumental variable. The choice of instrumental variable is indicated in the bottom row. All estimates are based on our main sample ($N = 15,042$). $q_{i,t}$ denotes firm i's real sales in 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 that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively.

In the first column, we show the OLS results, i.e. not using an instrumental variable. This yields a statistically significant but absurd result since it implies an inelastic demand elasticity. This finding is not surprising because there is no reason to expect demand to be identified in this specification. In the second column, we present results using TFPR ($a_{i,t}$) as an instrument. Again, we find a nonsense result: The estimated demand elasticity has the wrong sign. This finding arises because TFPR confounds productivity and demand effects, and it emphasizes the importance of using a measure of real productivity as an instrument.

In columns 3 and 4, we reproduce the estimates from the main text using z and z^u as instruments. Although both estimates are reasonable, using utilization adjusted TFPQ reduces the estimated coefficient meaningfully relative to raw TFPQ. In column 5, we present estimates using the TFPQ measure based on our alternative utilization adjustment, $z^{u,p}$. Notably, if we use this more flexible utilization adjustment, we get an intermediate result between the z result and the z^u result. This may indicate that the simple utilization adjustment is a bit too coarse.

Fixed effects In Figure 17, we show the average of sector-year fixed effects in each year. Our main sample includes observations in 22 sectors. What is clear, is that the sector-year fixed effects capture most of the aggregate development in the Swedish economy. This lends credibility to our firm-level results. We identify demand after removing this variation. This may be especially important with respect to how our dispersion measures are interpreted. Our dispersion measures reflect changes in the firm-level dispersion of shocks and outcomes.

Figure 17: Average sector-year fixed effect from demand estimation, 1998-2013



This figure plots the average sector-year fixed effect for each year in the period 1998-2013. We take the sector-year fixed effects from our favored demand specification which uses $z^u_{i,t}$ as an instrumental variable and includes firm and sector-year fixed effects. We compute the average sector-year effect as the simple average across sectors in each year, i.e. not weighted by sector size.

In the main text, we include firm and sector-year fixed effects in all our demand regressions. But how important is the choice of fixed effects? We show in Table 22 how different choices with respect to fixed effects affects our demand estimates.[77] The first column shows results without fixed effects, the second column shows results for year fixed effects only, and the third column shows results for sector fixed effects only. In all three cases, the results differ meaningfully from our favored specification. As shown in column 4, firm level fixed effects play a significant role in identifying the effect of prices on quantities. When we also include sector-time fixed effect, along with firm fixed effects—our favored specification, and the basis for our main results—the estimated coefficient changes slightly but remains fairly close to that when only using firm level fixed effects.

Sample Results and analyses in our main text are based on an unbalanced sample of firms. However, one may wonder whether entry or exit of firms affects our results. To evaluate this pos-

77. Note that we use $\ln p_{i,t}$ rather than the sector-year demeaned $\hat{p}$ as our price variable

Table 22: Demand estimation, the role of fixed effects

	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$
$\ln p_{i,t}$	0.535	0.681	-2.388***	-3.665***	-2.985***
	(0.493)	(0.512)	(0.289)	(0.284)	(0.202)
Fixed effects	none	t	s	i	i & s-t

This table presents estimates of our baseline demand model, but based on different choices with respect to fixed effects. The choice of fixed effects is indicated in the bottom row: t indicates year fixed effects, s indicates sector fixed effects, i indicates firm fixed effects, and s-t indicates sector-year fixed effects. All estimates are based on our main sample ($N = 15,042$) and use $z^u_{i,t}$ as an instrumental variables. $q_{i,t}$ denotes firm i's real sales in year t and $p_{i,t}$ denotes firm i's relative price in year t. Standard errors are clustered at the firm level and given in parentheses. Level of significance at that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively.

sibility, we therefore re-estimate demand based on a balanced sample of firms. These results are show in Table 23. We find similar results as in the main text, though the elasticities estimated on the balanced panel are systematically smaller in magnitude. For example, when estimated on the balanced panel and using z^u as an instrument, the estimated elasticity is reduced in magnitude from about -3 in the main text to -2.6. These differences may be attributed to differences in the balanced panel sample, which has a greater proportion of large firms and successful firms. Regardless, the estimates based on the balanced and unbalanced samples are sufficiently close that we favor using the unbalanced sample which is more representative of the overall distribution of firms in the Swedish economy than the balanced panel.

Table 23: Demand estimation, balanced sample

	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$
$\hat{p}_{i,t}$	-3.674***	-2.573***	-3.667***	-2.583***
	(0.245)	(0.314)	(0.244)	(0.327)
$\hat{p}^2_{i,t}$			-5.252***	-4.350*
			(1.530)	(2.123)
N	8666	3313	8666	3313
iv	$z_{i,t}$	$z^u_{i,t}$	$z_{i,t}$	$z^u_{i,t}$

This table presents demand estimates based on a balanced sample. $q_{i,t}$ 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 that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively. The number of observations (N) and the choice of instrumental variable (*iv*) is indicated in the bottom panel of the table.

Another worry is that non-systematic or transitory factors, in particular related to the Great Recession, distort our demand estimates. This may especially be a problem for the non-linear demand specification. For instance, if the Great Recessions creates peculiar and unsystematic relationships between prices and quantities, this limits the general applicability of our results. For robustness, we therefore re-estimate our main demand estimates but exclude 2001 and 2009 from the sample. These results are shown in Table 24. These results are close to those based on the full sample. In other words, it does not seem that our demand estimates are simply a by-product of randomness associated with recessions.

Table 24: Demand estimation, main sample excluding recessions

	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$	$\ln q_{i,t}$
$\hat{p}_{i,t}$	-3.869***	-2.944***	-3.789***	-2.906***	-2.945***	-1.923***
	(0.230)	(0.199)	(0.222)	(0.204)	(0.264)	(0.257)
$\hat{p}^2_{i,t}$			-6.418***	-7.207***		
			(1.893)	(1.924)		
$\mathbf{1}(\hat{p}_{i,t} > 0)\hat{p}_{i,t}$					-1.763***	-2.007***
					(0.500)	(0.474)
iv	$z_{i,t}$	$z^u_{i,t}$	$z_{i,t}$	$z^u_{i,t}$	$z_{i,t}$	$z^u_{i,t}$

This table presents the same demand estimates as the in the main text, but excluding the recession years 2001 and 2009 from our main sample ($N = 13,117$). $q_{i,t}$ 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. $\mathbf{1}(\hat{p}_{i,t} > 0)\hat{p}_{i,t}$ denotes the interaction between an indicator variable for an above average price and $\hat{p}_{i,t}$. 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 that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively. The first two columns give results from the basic CES demand curve estimation, model 3. Columns three and four present results for our non-linear approximation, model 4. The last two columns present results from a piece-wise linear specification. The difference between each pair of regressions is the choice of instrumental variable. Columns 1, 3, and 5 use the "raw" TFPQ measure $z_{i,t}$. Columns 2, 4, and 6 use instead the utilization adjusted TFPQ measure $z^u_{i,t}$. We indicate the choice of instrumental variable in the bottom row ("iv").

Sectoral estimates To what extent is there heterogeneity across sectors? In our main analyses, we use a demand elasticity estimated for the entire economy. This elasticity is estimated using firm and sector-year fixed effects. Do the results differ if we instead estimate demand on a sector by sector basis? In Table 25, we present summary statistics for parameters θ and η estimated 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.

We find average and median demand elasticities that are consistent with our pooled estimates. There is some heterogeneity across sectors, as can be seen from the 25th percentile to 75th percentile range. This heterogeneity is most prominent for the η estimate, whereas the interquartile range for θ is moderate. We also find that the demand coefficient θ in a typical sector tends to be smaller than the mean, i.e. positive skewness. In the baseline demand estimates, there are few extreme values. For the non-linear approximation, in contrast, we find some outliers with respect to η. It may be that it is difficult to estimate this parameter, especially if there aren't enough observations or variability in the data.

Although interesting in their own right, we also use the sectoral demand estimates as part of a robustness exercise in which we estimate variance decompositions on a sector by sector basis. This exercise is presented below.

B.2.2 Demand shock cyclicality

Demand dispersion is unambiguously countercyclical. In Table 26, we present the percentage increases in demand dispersion during the 2001 and 2009 recessions as compared with the rest of

Table 25: 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.

the sample. We include the demand shocks measured from the baseline model ($\tilde{\Delta}\epsilon$) and from the non-linear approximation ($\tilde{\Delta}\epsilon^*$). We also include $\tilde{\Delta}\epsilon^{raw}$ and $\tilde{\Delta}\epsilon^{raw,*}$, which are identical but estimated using the "raw" TFPQ ($z_{i,t}$) as an instrument rather than utilization adjusted TFPQ ($z^u_{i,t}$). For each of the demand measures, we compute the log differences and demean by average sector-year growth. We then compute our dispersion measure on the transformed data. We see substantial increases in dispersion whether measured via the standard deviation or the interquartile range. Moreover, results are similar whether we measure dispersion using our baseline demand model (ϵ) or from the non-linear model (ϵ^*).

Table 26: Cyclicality of demand

	2001	2009
sd.($\tilde{\Delta}\epsilon^{raw}$)	4.3	46.0
sd.($\tilde{\Delta}\epsilon^{raw,*}$)	6.9	52.1
sd.($\tilde{\Delta}\epsilon$)	8.0	43.1
sd.($\tilde{\Delta}\epsilon^*$)	9.5	49.3

(a) Standard Deviation

	2001	2009
iqr.($\tilde{\Delta}\epsilon^{raw}$)	10.5	67.3
iqr.($\tilde{\Delta}\epsilon^{raw,*}$)	9.3	69.3
iqr.($\tilde{\Delta}\epsilon$)	7.2	56.2
iqr.($\tilde{\Delta}\epsilon^*$)	9.8	64.1

(b) Interquartile Range

This table presents percentage increases in demand shock dispersion for the 2001 and 2009 recessions relative to the average over all other years. ϵ and ϵ^* denote demand estimated from the baseline model and the non-linear model, respectively. In the top two rows, the demand measures are estimated using the "raw" TFPQ measure, as indicated by the *raw* superscript. The bottom two rows present results using utilization adjusted TFPQ as an instrument. The left table shows the results for standard deviation while the right table show results based on the interquartile range. All dispersion statistics are computed after removing sector-year variation across sectors. All statistics are based on the main sample.

Perhaps the most interesting observation is the difference between the standard deviation and the interquartile range: In nearly all comparisons, the increase in dispersion is larger when measured based on the interquartile range. This illustrates that increased dispersion during recessions is not driven by outliers.

We present in Table 27 kurtosis and skewness statistics for demand shocks. Kurtosis exhibits a clear pattern. Relative to non-recession years, kurtosis decreases in recessions. Moreover, the smallest kurtosis estimates are seen in 2009. This is unsurprising and consistent with the notion that dispersion increases during recessions. In contrast, the skewness statistics tell a more complicated story. While the skewness of demand shocks is unambiguously more negative in 2001 as compared to the non-recession period, the results are mixed in 2009. In 2009, if we measure demand based on our baseline model, we find that the shocks become more negatively skewed. However, in 2009, if we instead use shocks from the non-linear demand model, we find the opposite: Skewness becomes more positive. How we measure demand shocks is thus important. Taking into account "real rigidities" changes our understanding of the Great Recession. Specifically, it suggests that most firms were negatively affected by demand shocks in this period, rather than just a long left tail of firms doing poorly.

Table 27: Skewness and kurtosis of demand

		non-recession	2001	2009
$\tilde{\Delta}\epsilon^{raw}$	skewness	0.164	-0.250	0.104
	kurtosis	7.570	5.573	4.324
$\tilde{\Delta}\epsilon^{raw,*}$	skewness	0.186	-0.526	0.647
	kurtosis	9.048	6.723	5.706
$\tilde{\Delta}\epsilon$	skewness	0.094	-0.296	-0.041
	kurtosis	7.753	6.775	3.968
$\tilde{\Delta}\epsilon^{*}$	skewness	0.119	-0.477	0.523
	kurtosis	9.034	7.177	5.543

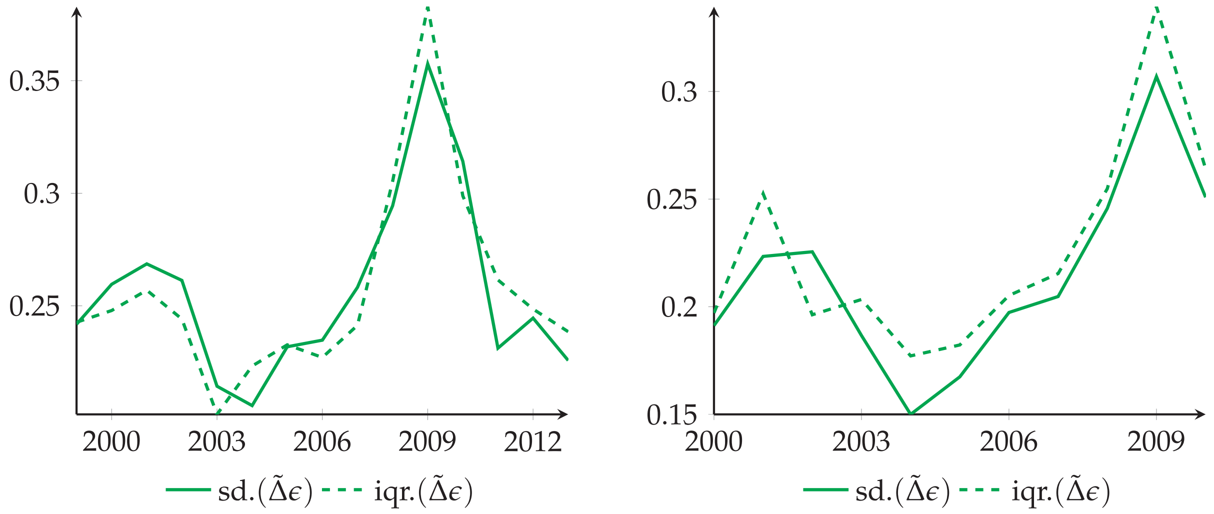
This table shows the skewness and kurtosis of demand shocks during 2001, 2009, and the average for 1998-2013 when excluding the recession years ("non-recession").

Time series for the standard deviation and the interquartile range of ϵ are presented in Figure 18. In the left panel we show the results from the main text, while in the right panel we show the results based on a balanced panel. With respect to the cyclicality of demand, the finding is perhaps even stronger in the case of a balanced panel.

B.2.3 Autocorrelation of demand

We estimate the autocorrelation of demand in our main dataset using an Anderson-Hsiao approach. Specifically, we regress the difference ($\Delta\epsilon_{i,t-1} = \epsilon_{i,t} - \epsilon_{i,t-1}$) on the lagged difference ($\Delta\epsilon_{i,t} = \epsilon_{i,t-1} - \epsilon_{i,t-2}$) and then use values from earlier periods as instruments. We consider various lags as instruments ($\epsilon_{i,t-2}$, $\epsilon_{i,t-3}$, and $\epsilon_{i,t-4}$), but this does not strongly affect our estimates. Overall, the autocorrelation of demand seems to be a bit larger than 0.6.

Figure 18: Volatility of demand



(a) Main sample (b) Balanced panel

This figures illustrates the volatility of demand. The right panel shows the dispersion of demand shocks estimated on our main sample, 1999-2013. The left panels show the dispersion of demand shocks estimated on the balanced sample, 2000-2010. The solid green line shows the standard deviation, sd.($\tilde{\Delta}\epsilon$)). The dashed green line shows the interquartile range, iqr.($\tilde{\Delta}\epsilon$).

B.3 Corroboration of TFPQ and demand shocks

Our TFPQ and demand measures are building blocks for our analyses. They are also interesting in their own right. In this section, we corroborate each measure. We show that each measure is associated with other variables that match the intended structural interpretations. Developments in our TFPQ measure are related to improved production processes, while developments in our demand measure are match managerial evaluations of the level of demand.

If TFPQ is properly measured, it should be associated with developments in the physical productivity at a firm. Consistent with this intuition, we show that process innovations reported in the Community Innovation Survey predict changes in our TFPQ measure. Firms that report improvements in manufacturing or supporting activities experience about 8% higher TFPQ growth on average. We do not find a comparable association with product innovations. Thus we find that the type of innovation activity that one would expect to increase production efficiency is exactly the kind that predicts our TFPQ measure.

If demand is properly measured, it should be associated with changes in the ability of the firm to sell its products. To check this hypothesis, we exploit a variable from SCB's Capacity Utilization survey in which managers evaluate the business environment. We find that negative demand shocks are associated with perceived low demand by managers. Specifically, demand growth falls by about 9% at firms where managers begin to report "insufficient demand".

B.3.1 TFPQ corroboration

Community Innovation Survey We use SCB microdata collected as part of the European Union's Community Innovation Survey (*Innovationsverksamhet i Sverige*) in our TFPQ corroboration exer-

Table 28: Autocorrelation of demand

	$\Delta\epsilon_{i,t}$	$\Delta\epsilon_{i,t}$	$\Delta\epsilon_{i,t}$
$\Delta\epsilon_{i,t-1}$	0.623***	0.624***	0.750***
	(0.0364)	(0.0505)	(0.0665)
iv	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-4}$	$\varepsilon_{i,t-4}$

This table presents autocorrelation estimates for demand ($\epsilon_{i,t}$) based on an Anderson-Hsiao method. All regressions include sector-year fixed effects. Standard errors are clustered at the firm level and given in parentheses. Level of significance at that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively. The difference between the three columns in the choice of instrument. In the first column, the second lag is used as an instrument for $\Delta\epsilon_{i,t-1}$. In the second column, the third lag is used as an instrument for $\Delta\epsilon_{i,t-1}$. And in the third column, the fourth lag is is used as an instrument for $\Delta\epsilon_{i,t-1}$. The choice of instrument is specified in the bottom row.

cises.[78] The CIS is a survey of innovation activity in firms. The CIS questionnaire includes sections related to firm organization, innovation activities, research and development, and activity in domestic and international markets. The CIS includes specific questions related to product and process innovations. We rely on four questions in particular, two related to product innovations and two related to process innovations:

- *Product innovations:* Did your firm introduce during the period t to $t+2$:
  - New or significantly improved goods, not including aesthetic changes or goods that are bought from other firms for resale
  - New or significantly improved services
- *Process innovations:* Did your firm introduce during the period t to $t+2$:
  - New or significantly improved methods of production or manufacturing
  - New or significantly improved supporting activities, e.g. maintenance systems, purchasing or accounting systems, or other computer technology

For each question, the possible responses are either yes or no. The questionnaire includes supplemental information providing definitions and examples. For instance, the questionnaire includes a list of industrial process innovations that includes examples such as new CAD systems or new blades for lumber mills. The questionnaire also specifies aspects of production that should not be considered product or process innovations. For examples, routine upgrades should not be considered product innovations, and expanded production capacity that resembles existing production methods should not be considered process innovation.

The CIS survey is conducted every other year and we have access to five surveys covering the period 2002-2012 (CIS4, CIS2006, CIS2008, CIS2010, and CIS2012). The sampling scheme is random within strata defined by sector and firm size. All firms with at least 250 employees are included, while firms with fewer than 10 employees are not part of the sample. Although CIS data is available for about 60% of firms in our sample, many firms are only present in a single CIS survey. This limits the number of observations that are available in practice.

78. More information is available via the Eurostat website: https://ec.europa.eu/eurostat/web/microdata/community-innovation-survey.

An idiosyncratic feature of the CIS survey is that managers are asked to evaluate the *three* year period preceding the survey. For example, the CIS4 survey covers the years 2002-2004 while the CIS2006 survey covers 2004-2006. This means that the surveys overlap and responses to subsequent surveys may reflect the same innovations. For example, an innovation reported in 2004 may be reported in both CIS4 and CIS2006. We address this in our empirical specification by defining changes as a two-period difference.

After merging the CIS data with our main dataset, we aggregate to match the frequency of the CIS sample. For TFPQ, we compute the average level of TFPQ during a survey period. Our dataset is thus a panel with firm-survey dimensions. As explained above, the CIS surveys overlap in time period. We therefore compute TFPQ growth as a forward-lag difference, i.e. a two period difference. This guarantees that growth in TFPQ is not based on the same years. However, it means that we are left with only about 500 usable observations.

Corroboration Results from the corroboration exercise are presented in Table 29. This tables shows estimates from a regression of TFPQ growth on two innovation indicators. The "process" indicator is based on whether the firm reported either new methods of production or new supporting activities in the CIS. The "product" indicator is based on whether the firm reported new goods or services in the CIS. We view these indicators as denoting changes in the "stock" of innovations, with product and process innovations accumulating differently and playing different roles. Importantly, we expect that the factor productivity improves as process innovations accumulate. To measure TFPQ growth we use a forward-backward difference (computed as explained above), which we then try to predict based on the indicators. Along with the main regressors, we also include time and sector fixed effects. We would have preferred to use sector-time fixed effects, but are constrained by the lack of observations available in this exercise. Standard errors are clustered at the firm level.

Table 29: The relationship between TFPQ shocks and product and process innovations

	$\Delta z_{i,t}$	$\Delta z^{u}_{i,t}$	$\Delta z^{u,p}_{i,t}$
process	0.083^{**}	0.072^{*}	0.082^{**}
	(0.031)	(0.032)	(0.030)
product	0.048	0.015	0.041
	(0.031)	(0.034)	(0.031)
N	492	492	489

Estimates are based on a firm-time panel in which time period is based on CIS survey. TFPQ growth, indicated by Δ, is computed as the one period forward-one period backward change. As a consequence, inclusion in the analyses requires that a firm is observed in at least three consecutive CIS surveys ($N = 492$). All specifications include sector and time fixed effects. Standard errors are clustered at the firm level and given in parentheses. Level of significance at that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively.

For each of our three TFPQ measures, we find significant evidence of an association between process innovations and TFPQ growth. In each case, a report of a process innovation predicts about 8% higher growth in TFPQ. Notably we do not find comparable associations with product innovations. Consistent with our intended interpretation of our TFPQ measure, it seem to reflect efficiency gains rather than the introduction of new products or improvements in product quality. Also notable is that a similar regression but using demand growth rather than TFPQ growth yields no significant results. For process innovations, this makes sense because process innovations should be related to supply-side factors.

B.3.2 Demand corroboration

As part of the capacity utilization survey conducted by SCB, managers evaluate various features of the business environment as experienced by their firm. Among these is an option related to "insufficient demand." Although a specific definition of "insufficient demand" is not provided, a plausible interpretation is that sales or orders are lower than expected given the price. This is consistent with situations in which utilization is low due to lack of demand.

To corroborate demand, we investigate whether our demand measure is related to reports of insufficient demand reported by managers. Specifically, we regress demand shocks on the change in the insufficient demand indicator, denoted $\Delta\mathcal{I}(\check{\epsilon})$. We use $\Delta\mathcal{I}(\check{\epsilon})$ rather than $\mathcal{I}(\check{\epsilon})$ itself because we only expect changes to be associated with developments in demand. For example, a firm that has persistent low demand may report insufficient demand in consecutive periods even though the firm has not experienced any demand shocks. We include sector-year fixed effects. Results are shown in Table 30. Consistent with the intended structural interpretation, a firm that begins to report insufficient demand experiences about 8% lower growth than what otherwise would be expected.

Table 30: The relationship between demand shocks and reported "insufficient demand"

	$\Delta\epsilon$
$\Delta\mathcal{I}(\check{\epsilon})$	-0.081***
	(0.01)

Estimates are based on our main sample ($N = 10,118$). $\Delta\epsilon$ denotes firm level demand shocks. Demand is estimated using utilization adjusted TFPQ, $z^u_{i,t}$. $\Delta\mathcal{I}(\check{\epsilon})$ denotes the change in the insufficient demand indicator. All specifications include sector-year fixed effects. Standard errors are clustered at the firm level and given in parentheses. Level of significance at that 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars respectively.

C Passthrough and Variance Decomposition Appendix

C.1 Passthrough

Sample Are our passthrough results affected by estimation in an unbalanced sample? In Table 31, we re-produce the passthrough estimates from the main text but based on a balanced panel. This includes re-estimating demand and computing demand shocks based on the balanced panel only (the demand estimate based on the balanced panel—presented above—is slightly smaller in magnitude than that estimated in our main sample). In general, the results are similar. For demand, the estimated results are close to identical. For TFPQ we see some small differences. Some of the β_z estimates increase while others decrease.

Time-varying passthrough Time-varying passthrough is an interesting and potentially important issue. In Table 32, we present passthrough estimated on a year by year basis for our main sample. In other words, we estimate $z_{i,t}$ and $\epsilon_{i,t}$ as normal, but run our passthrough regression for each year individually rather than for the while sample. In this regressions, we control for sector effects: $\log p_{i,t} = \beta_t^z z_{i,t}^u + \beta_t^\epsilon + \mu_j + \tau_{i,t}$. This yields passthrough coefficients for each year between 1999 and 2013. We find that TFPQ passthrough is countercyclical while demand passthrough is procyclical. We discuss these estimates further in conjunction with a variance decomposition exercise below. This exercise, we rely on the year by year passthrough estimates to compute the decompositions components for sales and prices (see Figure 21).

C.2 Variance Decomposition

C.2.1 Specification

For reference, we re-present our variance decomposition and the variance decomposition terms below.

Price decomposition The variance decomposition of prices is based on our passthrough equation.

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

Taking differences of this equation and removing sector-year growth from each variable yields:

$$\tilde{\Delta} \log p_{i,t} = \beta_z \tilde{\Delta} z_{i,t}^u + \beta_\epsilon \tilde{\Delta} \epsilon_{i,t} + \tilde{\Delta} \tau_{i,t}.$$

Taking the variance of both sides then yields

$$\text{Var}_t(\tilde{\Delta} p_{i,t}) = \beta_z^2 \text{Var}_t(\tilde{\Delta} z_{i,t}^u) + \beta_\epsilon^2 \text{Var}_t(\tilde{\Delta} \epsilon_{i,t}) + \text{Var}_t^{p,resid}.$$

where $\text{Var}_t(\tilde{\Delta} p_{i,t})$ is the dispersion of prices, $\text{Var}_t(\tilde{\Delta} z_{i,t})$ and $\text{Var}_t(\tilde{\Delta} \epsilon_{i,t})$ are the volatilities of the shocks, and $\text{Var}_t^{p,resid}$ is a residual term. The residual is composed of a set of covariances, as well as the volatility of the price wedge itself:

$$\text{Var}_t^{p,resid} \equiv \text{Var}_t(\tilde{\Delta} \tau_{i,t}) + \beta_z \beta_\epsilon \text{Cov}_t(\tilde{\Delta} z_{i,t}^u, \tilde{\Delta} \epsilon_{i,t}) + \beta_z \text{Cov}_t(\tilde{\Delta} z_{i,t}^u, \tilde{\Delta} \tau_{i,t}) + \beta_\epsilon \text{Cov}_t(\tilde{\Delta} \epsilon_{i,t}, \tilde{\Delta} \tau_{i,t}).$$

Table 31: Passthrough estimates, balanced panel

	$\ln p_{i,t}$	$\ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\ln p_{i,t}$	$\Delta \ln p_{i,t}$
$z^u_{i,t}$	-0.145***	-0.189***			-0.192**	
	(0.0160)	(0.0349)			(0.0656)	
$\epsilon_{i,t}$	0.226***	0.233***			0.246***	
	(0.0122)	(0.0203)			(0.0296)	
$\Delta z^u_{i,t}$			-0.108***	-0.114***		-0.107***
			(0.00951)	(0.0106)		(0.00896)
$\Delta \epsilon_{i,t}$			0.209***	0.224***		0.224***
			(0.0111)	(0.0112)		(0.00802)
$\mathbf{1}(\Delta z^u_{i,t} < 5\%)$						0.00207
						(0.0135)
$\mathbf{1}(\Delta z^u_{i,t} > 95\%)$						-0.00672
						(0.0137)
$\mathbf{1}(\Delta \epsilon_{i,t} < 5\%)$						-0.0287*
						(0.0129)
$\mathbf{1}(\Delta \epsilon_{i,t} > 95\%)$						-0.0202
						(0.0129)
N	3313	2553	2612	2131	2135	2612
iv	no	L.z L.ϵ	no	no	L2.z L2.ϵ	no
sample	all	all	all	$\lvert\Delta \ln P^f\rvert > 0.01$	all	all

This table presents the same passthrough regressions as the main text but based on a balanced panel. $p_{i,t}$ denotes firm i's relative price in year t, while $z^u_{i,t}$ and $\epsilon_{i,t}$ denote firms i's TFPQ and demand in year t. First differences are indicated by Δ. The terms of form $\mathbf{1}(\Delta x < 0.5\%)$ denote interactions between a shock x and an indicator for being in either in the lowest 5% og greatest 95% of the shock distribution. 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. Level of significance at the 0.05, 0.01, or 0.001 levels are indicated by one (*), two (**), or three (***) stars. The first and second columns show results for the estimation of the passthrough equation 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 the passthrough equation estimated in first differences. Column 4 is the same model as column 3, but excludes observations for which nominal price changes are less than 1% ($|\Delta \ln P^f| < 1\%$). Column 5 presents the same results as column 2, but instead using the two-year lag of the shocks as instruments. Column 8 repeats the first difference regression but allows for different coefficients for extreme large and small changes in TFPQ and demand. Below the estimation results, the bottom panel presents information on the number of observations (N), the use of instrumental variables (iv), and whether the sample excludes small price changes (sample).

Table 32: Passthrough estimates, year by year

	1999	2000	2001	2002	2003	2004	2005
	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$
$\Delta z^u_{i,t}$	-0.0900***	-0.123***	-0.114***	-0.102***	-0.0489***	-0.0616***	-0.0842***
	(0.00734)	(0.00832)	(0.00854)	(0.00775)	(0.00608)	(0.00623)	(0.00690)
$\Delta \epsilon_{i,t}$	0.197***	0.196***	0.174***	0.203***	0.173***	0.145***	0.138***
	(0.00686)	(0.00738)	(0.00722)	(0.00686)	(0.00688)	(0.00691)	(0.00683)
N	845	748	706	880	847	811	794

	2006	2007	2008	2009	2010	2011	2012	2013
	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$	$\Delta \ln p_{i,t}$
$\Delta z^u_{i,t}$	-0.0896***	-0.0955***	-0.0929***	-0.114***	-0.102***	-0.0957***	-0.0866***	-0.0964***
	(0.00745)	(0.00853)	(0.00686)	(0.00974)	(0.00844)	(0.00867)	(0.00872)	(0.00834)
$\Delta \epsilon_{i,t}$	0.156***	0.243***	0.264***	0.229***	0.224***	0.239***	0.248***	0.241***
	(0.00711)	(0.00757)	(0.00639)	(0.00806)	(0.00719)	(0.00922)	(0.00800)	(0.00808)
N	770	748	793	594	666	599	646	661

This table presents passthrough estimates estimated on a year by year basis using our main sample. Passthrough is estimated in first differences including sector 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.

This yields the terms of our price decomposition:

$$\begin{aligned}
V_t^p &= \text{Var}_t(\tilde{\Delta} p_{i,t}) \\
V_t^{p,z} &= \beta_z^2 \text{Var}_t(\tilde{\Delta} z^u_{i,t}) \\
V_t^{p,\epsilon} &= \beta_\epsilon^2 \text{Var}_t(\tilde{\Delta} \epsilon_{i,t}) \\
V_t^{p,\tau} &= \text{Var}_t(\tilde{\Delta} \tau_{i,t}) \\
V_t^{p,\text{Cov}(z,\epsilon)} &= \beta_z \beta_\epsilon \text{Cov}_t(\tilde{\Delta} z^u_{i,t}, \tilde{\Delta} \epsilon_{i,t}) \\
V_t^{p,\text{Cov}(z,\tau)} &= \beta_z \text{Cov}_t(\tilde{\Delta} z^u_{i,t}, \tilde{\Delta} \tau_{i,t}) \\
V_t^{p,\text{Cov}(\epsilon,\tau)} &= \beta_\epsilon \text{Cov}_t(\tilde{\Delta} \epsilon_{i,t}, \tilde{\Delta} \tau_{i,t}).
\end{aligned}$$

Sales decomposition The starting point for the sales decomposition is the observation that $\log s_{i,t} = \log p_{i,t} + \log q_{i,t}$. In other words, the expression for sales combines the price equation and the demand expression $\log q_{i,t} = -\theta \log p_{i,t} + \epsilon_{i,t}$. This yields

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

Taking differences and removing sector-year variation yields

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

Computing the variances then yields:

$$V_t^s = V_t^{s,z} + V_t^{s,\epsilon} + V_t^{s,\tau} + V_t^{s,\text{Cov}(z,\epsilon)} + V_t^{s,\text{Cov}(\epsilon,\tau)} + V_t^{s,\text{Cov}(z,\tau)},$$

where the terms are given by:

$$\begin{aligned}
V_t^s &= \text{Var}_t(\tilde{\Delta} s_{i,t}) \\
V_t^{s,z} &= (1-\theta)^2 \beta_z^2 \text{Var}_t(\tilde{\Delta} z_{i,t}) \\
V_t^{s,\epsilon} &= (1+(1-\theta)\beta_\epsilon)^2 \text{Var}_t(\tilde{\Delta}\epsilon_{i,t}) \\
V_t^{s,\tau} &= (1-\theta)^2 \text{Var}_t(\tilde{\Delta}\tau_{i,t}) \\
V_t^{s,\text{Cov}(z,\epsilon)} &= (1-\theta)\beta_z(1+(1-\theta)\beta_\epsilon)\text{Cov}(\tilde{\Delta} z_{i,t}), \tilde{\Delta}\epsilon_{i,t})) \\
V_t^{s,\text{Cov}(z,\tau)} &= (1-\theta)^2\beta_z \text{Cov}(\tilde{\Delta} z_{i,t}), \tilde{\Delta}\tau_{i,t})) \\
V_t^{s,\text{Cov}(\epsilon,\tau)} &= (1-\theta)(1+(1-\theta)\beta_\epsilon)\text{Cov}(\tilde{\Delta}\epsilon_{i,t}), \tilde{\Delta}\tau_{i,t})).
\end{aligned}$$

C.2.2 Quantification

In Table 33, 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.

The top panel quantifies a number of observations from our variance decompositions. To begin with, price and sales dispersion (V^p and V^s) tends to increase during recessions. The only exception is prices during 2001. In 2001, price dispersion actually fell by a small amount. This is because the increase in price dispersion actually preceded the 2001 recession slightly. We also see that the dispersion of shocks tends to increase. Here, the only exceptions are the components related to TFPQ in 2001. For neither prices nor sales, does the variance attributable to z^u increase during 2001.

In the bottom panel of Table 33, we quantify how the variance of the shocks changes relative to the variance of sales and prices during recessions. This provides a measure of what portion of the change in the variance of sales and prices that can be attributed to the shocks. The most clear effect is the strong cyclical effect of demand in driving cyclical dispersion. For example in the last column, we compute the change in $V^{s,\epsilon}$ relative to the change in V^s. In 2001, about 27% of the total increase in sales volatility is attributable to demand. In 2009 this is even larger. In the Great Recession, about 80% of the total increase in sales dispersion is attributable to demand. We also see in the second column, that demand plays a substantial role for prices during the Great Recession, explaining about 40% of the rise in price dispersion. With respect to TFPQ, we see that it plays a small role in the Great Recession, explaining about 4% of the increase for both price and sales variance.[79]

C.2.3 Parameter robustness

In the main body of the text, we present variance decomposition results that rely on parameters that we estimate in our data. We use a demand elasticity of about $\theta = 3$ in all our variance de-

79. The remaining entries in the table are more difficult to interpret. For instance, the variance of prices actually fell marginally during 2001. Comparisons of the components $V^{p,z}$ and $V^{p,\epsilon}$ with this small negative change are difficult to interpret as meaningful.

Table 33: 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^{u} and ϵ as defined by the ratio of changes in 2001 and 2009 presented in the top panel.

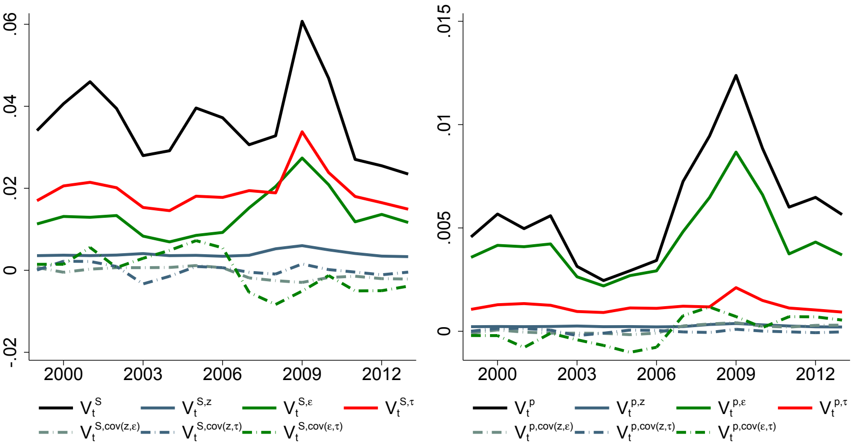
compositions. Our passthrough estimates for TFPQ (β_z) range between -0.1 and -0.3, depending on specification and sample, while our passthrough estimates for demand (β_ϵ) are all close to 0.225. We view these estimates as our most credible. Nevertheless, it is worth evaluating how our variance decomposition results change when different parameter values are used. In Figure 19, we present two robustness exercises in which we impose values for our structural parameters. Imposing a parameter affects the variance decomposition mechanically via its role in the computation of the variance decomposition components. Imposing a parameter also affects the variance decomposition via other channels. For example, when we impose a value for θ, this changes the demand shocks that we measure, with consequences for both the volitility of demand and for the passthrough estimation (β_z, β_ϵ, and τ).

In the top panel, we re-estimate the variance decomposition but impose a larger demand elasticity. This is an valuable exercise because a larger θ is not implausible. A larger elasticity would, for instance, imply lower static markups. Concretely, we impose a demand elasticity of $\theta = 5$, i.e. 40% higher than our actual estimate. This reduces the passthrough coefficients marginally ($\beta_z = -0.066$, $\beta_\epsilon = 0.173$), increases demand volatility, and reduces τ volatility.

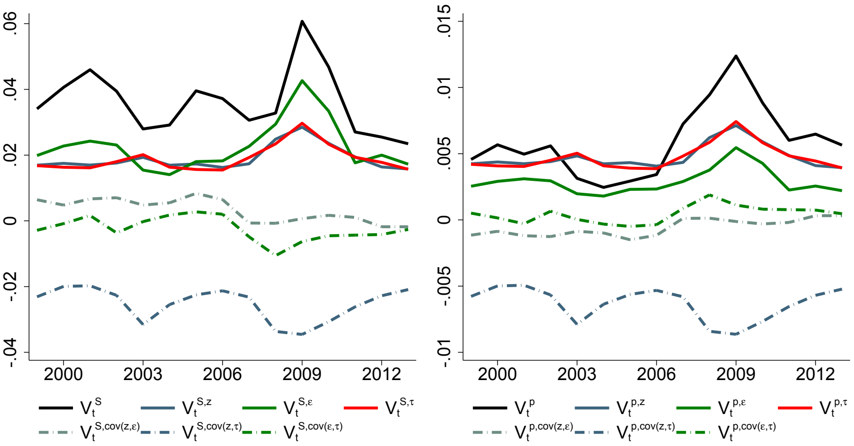
The variance decomposition results for $\theta = 5$ are in line with the results in the main text. Demand remains the most important driver of sales and price dispersion, though the relative importance for sales and prices changes somewhat. Demand is less important for sales—due to the reduction in β_ϵ—but more important for prices—due to the increase in $\text{Var}(\tilde{\Delta}\epsilon_{i,t})$. TFPQ remains inconsequential, in part because of lower passthrough β_z. Perhaps the most notable difference is the more prominent role of the price wedge in driving the cyclicality of sales. Relative to the main text, $V_t^{s,\tau}$ is larger overall (because of θ) and exhibits more cyclicality around 2001 and 2009 related to developments in $\text{Var}_t(\tilde{\Delta}\tau_{i,t})$.

The passthrough estimates we use in the main text are small. This has consequences for the

Figure 19: Variance decomposition, parameter robustness



(a) Variance decomposition, $\theta = 5$



(b) Variance decomposition, $\beta_z = -0.3$

The left panels present sales decompositions and the right panels price decompositions. All decompositions are based on passthrough estimated in first differences. The top row presents variance decompositions in which we impose a demand elasticity of $\theta = 5$. The bottom row presents variance decompositions in which we impose TFPQ passthrough of $\beta_z = -0.3$. This is about three times higher than what we estimate in our data. 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.

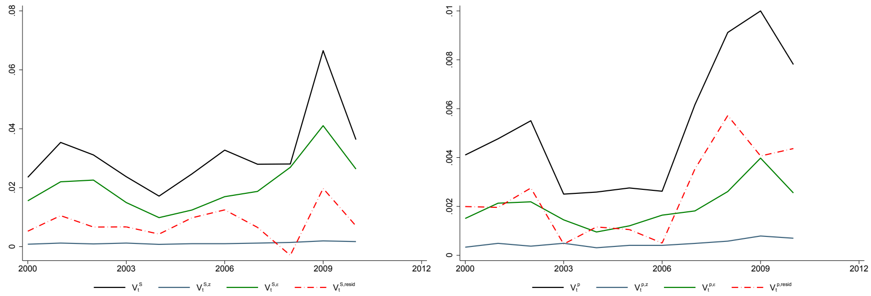
variance decomposition because low TFPQ passthrough implies that firms do not adjust their prices much in response to shocks—and prices are the only channel by which TFPQ shocks affect sales (notice the presence of β_z in all the variance decomposition components related to TFPQ). But what if true passthrough from TFPQ to prices is higher? For instance, perhaps measurement error makes it difficult to measure the actual degree of passthrough? We investigate this in the bottom panel of Figure 19. Here, we re-do the variance decomposition but impose TFPQ passthrough that is three times larger than the passthrough that we estimate in first differences. In other words, we use a TFPQ passthrough of -0.3 rather than -0.1. This directly increases the role of TFPQ for both sales and prices because β_z is larger. However, it also has a large impact on the covariance terms. In contrast to our favored specification in which this term is small, the covariance between z and τ now plays a substantial role. This is evidence of misspecification and suggests that $\beta_z = -0.3$ is too large. Too much TFPQ passthrough forces τ to be large and thus creates a meaningful degree of covariance between TFPQ and τ. Consistent with this conclusion, we see that $V^{s,\mathrm{Cov}(z,\tau)}$ becomes more negative exactly when the dispersion of TFPQ increases (in 2003 and 2008-2009). For this reason, we are sceptical that underestimated TFPQ passthrough is important for the results.

C.2.4 Sample

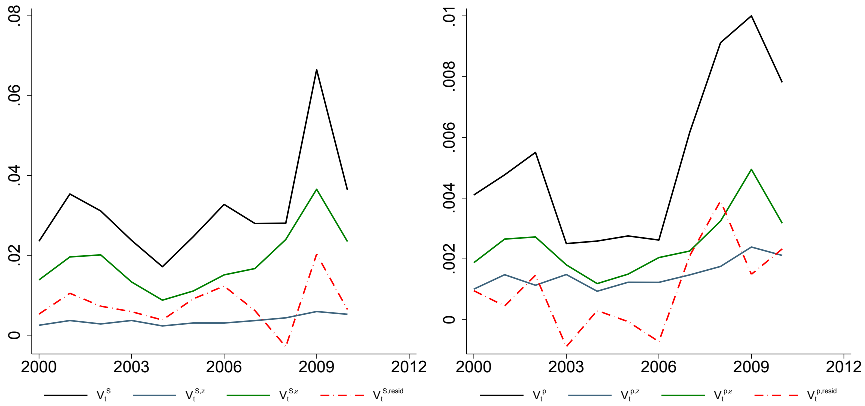
The variance decomposition results are robust to using a balanced sample. In Figure 20, we reproduce the variance decompositions from the main text but using a balanced sample for the period 2000-2010. In this exercise, we re-estimate demand and passthrough in the same balanced panel. As in the main text, the left side shows sales decompositions while the right side presents price decomposition. The top panel shows the variance decomposition using passthrough estimated based on first differences. while the bottom panel shows the variance decomposition using passthrough estimated using an instrumental variable. In contrast to the main text, however, we combine the τ and covariance terms into a single aggregate residual term. This residual is indicated by a dashed red line.

Results are comparable to those in the main text. Demand shocks can explain much of the overall volatility and cyclicality of sales and prices. As before, TFPQ volatility has limited explanatory power.

Figure 20: Variance decompositions, balanced sample



(a) Passthrough estimated in first differences



(b) Passthrough estimated using i.v.

In contrast to the variance decompositions presented in the main text, the variance decompositions presented in this figure are based on a balanced sample. 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.

C.2.5 Time-varying passthrough

We find some evidence of time-varying passthrough. In panel (a) of Figure 21 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).

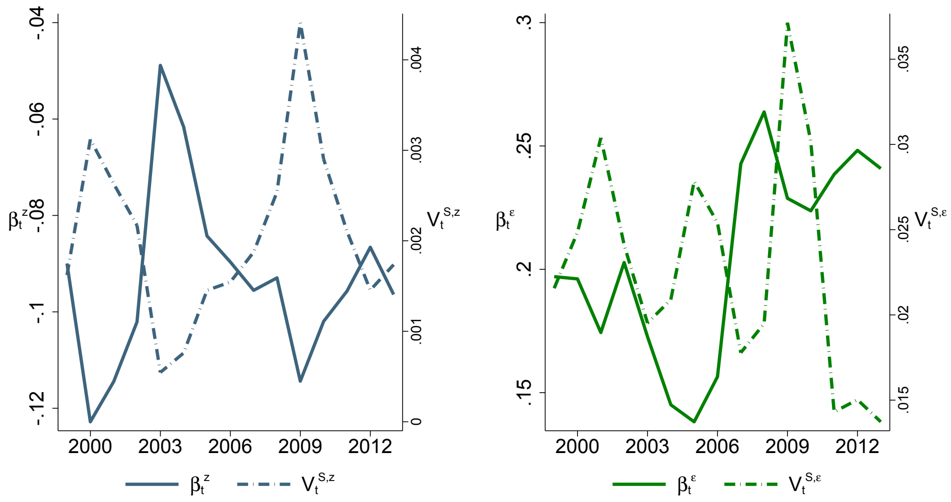
To what extent does time-varying passthrough change our variance decomposition? To begin with, it makes the estimation simpler because year-by-year estimation forces the covariance between the shocks and τ to be zero. There thus only five components of interest.

Results from a variance decomposition using passthrough estimated year by year is shown in panel (b) of Figure 21. 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.

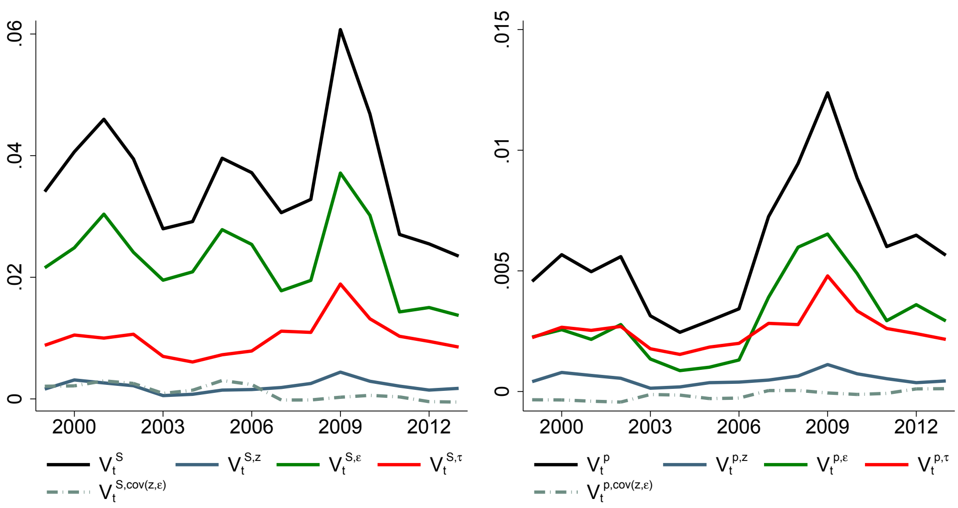
C.2.6 Sectoral variance decompositions

Our main results are pooled across manufacturing sectors. One may therefore wonder if our results disguise meaningful heterogeneity. Although we remove sector-year variation, there will still be differences across sectors with respect to dispersion. We there re-estimate demand, passthrough, and variance decompositions on a sector by sector basis. In Figure 22, we show variance decompositions of sales estimated for our four largest sectors, Food and Beverage on the top left, Wood Products on the top right, Metal Products on the bottom left, and Machinery and Equipment on the bottom right (sectors 10, 16, 25, and 28, respectively). In all four sectors, demand plays a major role during the Great Recession. In most cases, demand also appears to be important in 2001. In none of the sectoral decompositions does TFPQ play a significant role. Key features of the aggregated variance decomposition are thus reproduced on a sector by sector basis.

Figure 21: Variance decompositions, time varying passthrough



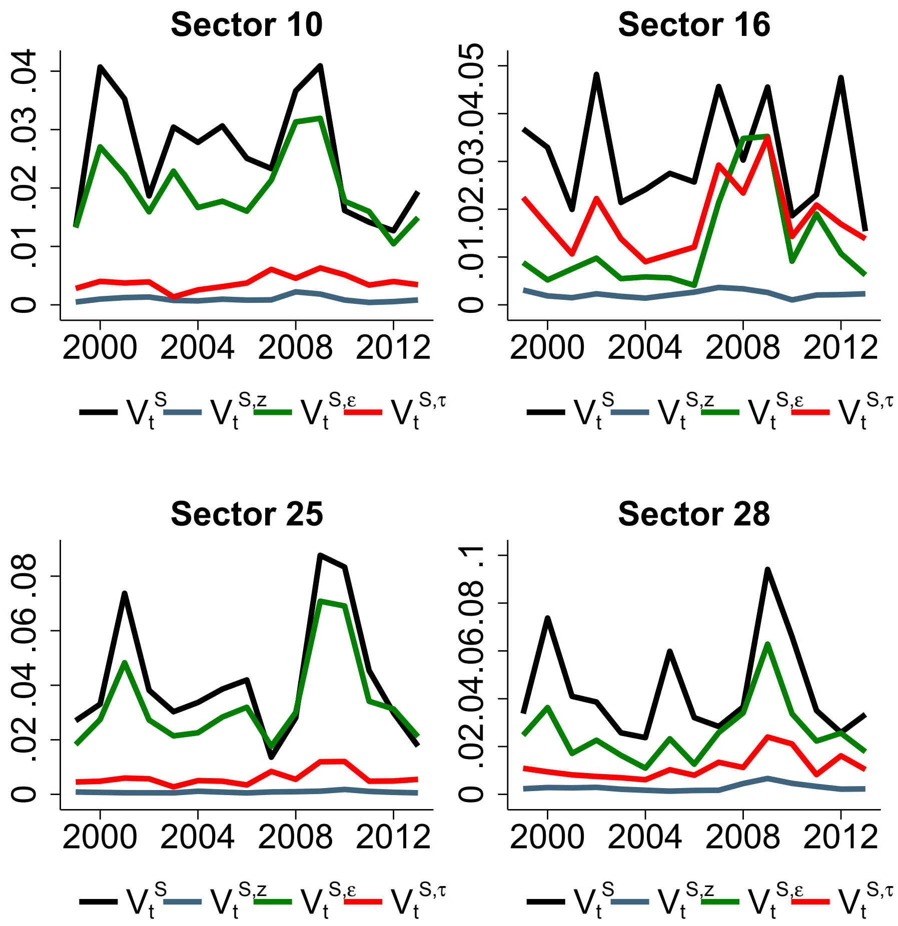
(a) Time-varying passthrough coefficients



(b) Variance decompositions: Time-varying passthrough

The figures in panel (a) plot 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, for comparison. The figures in panel (b) give 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.

Figure 22: Variance decompositions, by sector



Sales decompositions for the four largest sectors in our data: Food and beverage (sector 10), wood products (sector 16), metal products (sector 25), and machinery and equipment (sector 28).

D Quantitative Model Appendix

D.1 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 follows:

$$l = x(l/k)^\alpha = xb^\alpha, \qquad k = xb^{\alpha-1}. \tag{21}$$

Output is simply $q = zx$. The law of motion for x in terms of capital and labor evolution is:

$$\dot{x} = \dot{k}\alpha k^{\alpha-1}l^{1-\alpha} + \dot{l}(1-\alpha)k^\alpha l^{-\alpha} = x\left(\frac{\dot{k}}{k}\alpha + \frac{\dot{l}}{l}(1-\alpha)\right) = x\left(\frac{i}{k}\alpha + \frac{h}{l}(1-\alpha) - \delta\right). \tag{22}$$

For expositional clarity we can define an investment rate for x so that we can make a simple law of motion:

$$i_x \equiv \dot{x} + \delta x \implies \dot{x} = i_x - \delta x. \tag{23}$$

The law of motion for b is:

$$\dot{b} = \dot{k}\frac{-l}{k^2} + \dot{l}/k = b\left(\frac{\dot{l}}{l} - \frac{\dot{k}}{k}\right) = b\left(\frac{h}{l} - \frac{i}{k}\right). \tag{24}$$

Now we can combine these to solve for investment rates in terms of i_x and $\dot{b}$:

$$i = i_x b^{\alpha-1} - \dot{b}(1-\alpha)xb^{\alpha-2}, \tag{25}$$

$$h = i_x b^\alpha + \dot{b}\alpha x b^{\alpha-1}. \tag{26}$$

Recall that investment goods cost p_k and the linear hiring / firing cost is a. Then total hiring costs are

$$p_k i + ah = i_x\left(p_k b^{\alpha-1} + ab^\alpha\right) + \dot{b}x\left(a\alpha b^{\alpha-1} - p_k(1-\alpha)b^{\alpha-2}\right), \tag{27}$$

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

$$cf = pq - wl - p_k i - ah - c(i_x, x) = \pi(x,b) - i_x\left(p_k b^{\alpha-1} + ab^\alpha\right) - \dot{b}x\left(a\alpha b^{\alpha-1} - p_k(1-\alpha)b^{\alpha-2}\right) - c(i_x, x), \tag{28}$$

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:

$$rv(k,l) = \max_{i,h} \hat{\pi}(k,l) - ip_k + ah - c(i_x, k^\alpha l^{1-\alpha}) + v_k(i - \delta k) + v_l(h - \delta l), \tag{29}$$

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:

$$rv(x,b) = \max_{i_x, \dot{b}} \pi(x,b) - i_x\left(p_k b^{\alpha-1} + ab^\alpha\right) - \dot{b}x\left(a\alpha b^{\alpha-1} - p_k(1-\alpha)b^{\alpha-2}\right) - c(i_x, x) + v_x(i_x - \delta x) + v_b\dot{b}. \tag{30}$$

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:

$$\frac{\partial}{\partial \dot{b}} = -x\left(a\alpha b^{\alpha-1} - p_k(1-\alpha)b^{\alpha-2}\right) + v_b(x,b) = 0. \tag{31}$$

This defines an optimal $b(x)$, once the derivative $v_b(x,b)$ is known. The challenge of writing the value function with x as the only state is being able to solve for the optimal $b(x)$ function analytically. Fortunately, the solution is very simple. Guess that the solution will have b independent of x, and so the optimal value is just a constant b^*. Plug this value into the FOC

$$x\left(a\alpha(b^*)^{\alpha-1} - p_k(1-\alpha)(b^*)^{\alpha-2}\right) = v_b(x,b^*). \tag{32}$$

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 $f()$.

To establish that this is true, take the envelope condition of the HJB with respect to b:

$$rv_b(x,b) = R_b(x,b) - i_x\left(a\alpha b^{\alpha-1} - (1-\alpha)p_k b^{\alpha-2}\right) + v_{xb}(i_x - \delta x) + \dot{b}(...). \tag{33}$$

Note that at the optimal b^* we have $\dot{b} = 0$, allowing us to drop the final term:

$$rv_b(x,b^*) = R_b(x,b^*) - i_x\left(a\alpha(b^*)^{\alpha-1} - (1-\alpha)p_k(b^*)^{\alpha-2}\right) + v_{xb}(x,b^*)(i_x - \delta x). \tag{34}$$

Use the b FOC (32) to replace b^*:

$$rv_b(x,b^*) = R_b(x,b^*) - i_x\frac{v_b(x,b^*)}{x} + v_{xb}(x,b^*)(i_x - \delta x). \tag{35}$$

If it is the case that $v_b(x,b^*) = xf(b^*)$ then we have $v_b(x,b^*)/x = f(b^*)$ and $v_{bx}(x,b^*) = f(b^*)$. Plug these in, along with $R_b = -\alpha w x b^{\alpha-1}$, to give

$$f(b^*) = -\frac{\alpha w(b^*)^{\alpha-1}}{r+\delta}, \tag{36}$$

which solves for the guessed $f()$ function. Plug this back into (32) to get the solution for b^*:

$$b^* = \frac{1-\alpha}{\alpha}\frac{(r+\delta)p_k}{(r+\delta)a + w}. \tag{37}$$

Reassuringly, this is the standard formula for the optimal ratio between k and l in a model where both can be frictionlessly adjusted.

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 (30) to create a HJB for $v^*(x)$:

$$rv^*(x) = \max_{i_x} \pi(x,b^*) - i_x p_x - c(i_x,x) + v^*_x(i_x - \delta x). \tag{38}$$

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. This is the relevant value function which we use in the main text, which corresponds to the value function for any firm who starts with b set at the optimal level. That is, the value (17) in the text is an extension of (38) to include idiosyncratic and aggregate shocks.

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^*. We can calculate the full value function $v(x, b)$ for such a firm with any non-optimal value of b. Suppose a firm has current state (x, b). How does the firm jump to (x, b^*)? The firm currently has capital and labour:

$$l = xb^\alpha, \qquad k = xb^{\alpha-1}, \tag{39}$$

and must jump to

$$l^* = x(b^*)^\alpha, \qquad k^* = x(b^*)^{\alpha-1}. \tag{40}$$

The total cost of doing so is

$$p_k(k^* - k) + a(l^* - l) = p_k x\left((b^*)^{\alpha-1} - b^{\alpha-1}\right) + ax\left((b^*)^\alpha - b^\alpha\right). \tag{41}$$

This means the value of having this initial value of b is

$$v(x, b) = v^*(x) - p_k x\left((b^*)^{\alpha-1} - b^{\alpha-1}\right) - ax\left((b^*)^\alpha - b^\alpha\right), \tag{42}$$

which is just the cost of jumping the state from b to b^* plus the value of then continuing with those states, which is $v^*(x)$.

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 (in this case for all s) then the optimal capital-labor ratio b^* is a constant. This makes the model identical to a model with a Leontief production function $q = \min\{b^*k, l\}$, since with Leontief the capital-labor ratio is also optimally held constant at b^*. In our quantitative exercises, we consider a partial equilibrium with constant prices, so the results are equivalent to those with a Leontief production function. If the factor prices change across aggregate states, the ratio b^* would change across states according to the formulas above.

D.2 Investment first order condition

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

$$c(i_x, x) = \begin{cases} \frac{\kappa}{2}\frac{(i_x - \delta x)^2}{x} & i_x > \delta x \\ 0 & \delta x \geq i_x \geq 0 \\ -\underline{\kappa} i_x + \frac{\kappa}{2}\frac{i_x^2}{x} & i_x < 0, \end{cases} \tag{43}$$

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 is not differentiable at the boundaries:

$$c_{i_x}(i_x, x) = \begin{cases} \kappa\frac{i_x - \delta x}{x} & i_x > \delta x \\ 0 & \delta x > i_x > 0 \\ -\underline{\kappa} + \kappa\frac{i_x}{x} & i_x < 0. \end{cases} \tag{44}$$

Studying the HJB (17), the first order condition for investment within any region where a local optimum exists is

$$-p_x - c_{i_x}(i_x, x) + v_x = 0. \tag{45}$$

Firstly, suppose that an investment rate $i_x > \delta x$ was optimal. Then, the first order condition implies that this optimal value must be

$$-p_x - \kappa \frac{i_x - \delta x}{x} + v_x = 0 \implies \frac{i_x}{x} = \delta + \frac{v_x - p_x}{\kappa}. \tag{46}$$

This optimal value only satisfies $i_x > \delta x$ if $v_x > p_x$. Secondly, suppose that an investment rate $i_x < 0$ was optimal. Then, the first order condition implies that this optimal value must be

$$-p_x + \underline{\kappa} - \kappa \frac{i_x}{x} + v_x = 0 \implies \frac{i_x}{x} = \frac{v_x - (p_x - \underline{\kappa})}{\kappa}. \tag{47}$$

This optimal value only satisfies $i_x < 0$ if $v_x < p_x - \underline{\kappa}$. Notice that $\underline{\kappa}$ acts as a lower resale price for the input bundle: you can buy inputs for price p_x but must sell them for $p_x - \underline{\kappa} < p_x$. For $p_x - \underline{\kappa} < v_x < p_x$ neither investment nor disinvestment are optimal and the firm sets $i_x = 0$. For $v_x = p_x$, 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 $i_x = \delta x$. Stitching these functions together yields the final policy function:

$$\frac{i_x}{x} = \begin{cases} \delta + \frac{v_x - p_x}{\kappa} & v_x \geq p_x \\ 0 & p_x > v_x > p_x - \underline{\kappa} \\ \frac{v_x - (p_x - \underline{\kappa})}{\kappa} & p_x - \underline{\kappa} \geq v_x. \end{cases} \tag{48}$$

D.3 Numerical solution details

Numerical implementation We solve the model using continuous time numerical methods which draw heavily from Achdou et al. (2022). We use their finite difference methods, including upwinding for the solution of the investment policy function. 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. Ergodic distributions and the aggregate simulations are calculated using the grid based simulation procedure that forms part of the Achdou et al. (2022) method.

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 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. Firstly, we time aggregate the data for real sales, capital, and labor as mentioned above. We construct TFPQ as the Solow residual using our estimated capital share of 25.5% from the data. Since there is no notion of (and therefore simulated data on) variable factor 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 dataset, in Table 19 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. 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.

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 "full model" calibration allows aggregate uncertainty to fluctuate between $s = 1$ and $s = 2$ as in the data. Parameters for both calibrations are provided in Table 34, where it can be seen that most parameters are identical across the two calibrations. In the full model, 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.

D.4 Model without adjustment costs

Optimal static markup: In this section we derive the policy functions for the model with no adjustment costs. This yields a static maximization problem and policy functions where prices and sales depend only on the current TFPQ and demand shock. The derivations are done for an extended model where demand shocks can influence the elasticity of demand. To recover the model from the text set $\eta_\epsilon = 0$.

Consider a simple static model of price setting. Firms face (log) TFPQ and demand shocks z and ϵ as measured in the data. 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 in our regressions. Firms maximize static profit, $\Pi \equiv (p - mc)q$ subject to their demand

Table 34: 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. See text for further details.

curve. Define a firm's markup over marginal cost as $\mu \equiv p/mc$.

Consider the specification of demand

$$\log q = \frac{\theta}{\eta} \log \left(1 - \eta \log p + \eta_\epsilon \epsilon\right) + \epsilon. \tag{49}$$

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, $\mathrm{E}z$.

Static profits are defined as

$$\begin{aligned} \Pi &= (p - \frac{1}{z})q \\ &= (p - \frac{1}{z})e^{\epsilon} \left(1 - \eta \log p + \eta_\epsilon \epsilon\right)^{\frac{\theta}{\eta}}. \end{aligned}$$

From which we get the first-order condition

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

which can be rearranged to yield

$$\left(1 - \eta \log p + \eta_\epsilon \epsilon\right) = \theta \left(1 - \frac{1}{\mu}\right). \tag{50}$$

Recalling the definition of the markup $\mu = pz \Rightarrow p = \mu/z$:

$$1 - \eta \log \mu + \eta \log z + \eta_\epsilon \epsilon = \theta \left(1 - \frac{1}{\mu}\right). \tag{51}$$

(51) pins down the level of the optimal markup μ, which we denote $\mu^*(z, \epsilon)$.

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. This is a convenient feature of our demand specification.

In the main model we assume that $\eta_\epsilon = 0$. In this case, the markup first order condition simplifies to

$$1 - \eta \log \mu + \eta \log z = \theta \left(1 - \frac{1}{\mu}\right). \tag{52}$$

and the optimal markup is now a function only of the TFPQ shock and not the demand shock: $\mu^*(z)$.

Optimal static passthrough (linearized model): To investigate passthrough, we take a log linear approximation to the optimal markup equation, (51). Simply replace $\frac{1}{\mu}$ with the approximation $\frac{1}{\mu} \simeq 1 - \log \mu$ in (51), and subtract the equation evaluated in steady state to give:

$$\hat{\mu} = \frac{\eta}{\theta + \eta} \hat{z} + \frac{\eta_\epsilon}{\theta + \eta} \hat{\epsilon}, \tag{53}$$

where $\hat{x} \equiv \log x - \log x_{ss}$. Noting that $\hat{p} = \hat{\mu} - \hat{z}$ gives the final passthrough equation in terms of prices:

$$\hat{p} = \frac{-\theta}{\theta + \eta} \hat{z} + \frac{\eta_\epsilon}{\theta + \eta} \hat{\epsilon}. \tag{54}$$

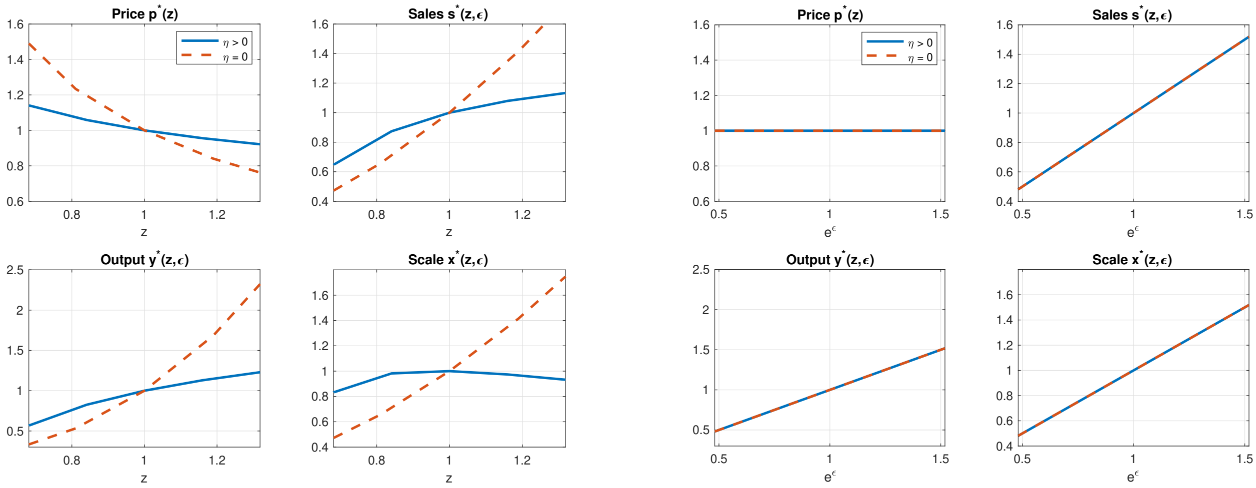
This equation is comparable to our estimated passthrough equation, as discussed in the main text. Equation (18) in the text corresponds to the special case of this equation with $\eta_\epsilon = 0$. For CES demand, the special case where $\eta = \eta_\varepsilon = 0$, this reduces to the standard result that the optimal markup is a constant and given by $\mu = \mu^{ces} = \theta/(\theta - 1)$.

Policy function plots and further discussion: In Figure 23 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. In particular, the demand curve has $\theta = 3$ and $\eta = 4.3$ in the non-CES model, and $\theta = 3$ and $\eta = 0$ in the CES model.

In both cases we are using our demand curve with $\eta_\epsilon = 0$, which means that the optimal markup is a function only of TFPQ: $\mu^*(z)$ is implicitly defined by equation (52). We plot the true non-linear policy functions, rather than linear approximations. We then back out the implied values of the other variables from the equations of the model. The price is $p^*(z) = \mu^*(z)/z$, quantity sold $q^*(z, \epsilon)$ comes from the demand curve, sales is $s^*(z, \epsilon) = p^*(z) q^*(z, \epsilon)$, and scale is $x^*(z, \epsilon) = q^*(z, \epsilon)/z$.

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. To understand why non-CES demand reduces the response of aggregate output to uncertainty about TFPQ, the bottom right panel plots the optimal scale policy function without adjustment costs.

Figure 23: Policy functions (without adjustment costs) for $\eta > 0$ vs. $\eta = 0$



(a) Policy functions across z for $\varepsilon = \varepsilon_{ss}$

(b) Policy functions across ε for $z = z_{ss}$

Figures plot 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 are plotted relative to their average value.

When demand is CES, the firm strongly increases its scale (capital and labor) stock when productivity rises. This is because high productivity allows the firm to lower its price, and increase its customer base by taking advantage of this lower price. With non-CES demand the optimal policy is very different. Recall that with non-CES demand the firm's elasticity of demand falls when it tries to lower its price. Thus, lowering the price brings in relatively fewer customers, representing the idea that it is hard to grow the customer base quickly.

Thus, when a firm's productivity rises there is little incentive to lower its price, which it does less than in the CES case (this is incomplete passthrough, shown in the smaller response of the price in the top left panel). Accordingly, the firm does not raise output as much when productivity rises (bottom left panel) and does not need to increase capital and labor as strongly. In fact, the effect here is so strong that optimal scale may actually fall when productivity rises, because the same amount of output can be produced with fewer inputs.[80] Overall, regardless of the sign of the response, scale is much less responsive to productivity when demand is non-CES rather than CES, as shown in the bottom right panel. 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, choosing instead to absorb the productivity change in markups.

To understand why our estimated non-CES demand reversed the Oi-Hartman-Abel effect, in the top right panel we see that the estimated non-CES demand system ($\eta > 0$) makes optimal sales concave in productivity. This is in contrast to the CES demand system, where sales are convex in productivity. This is what reverses the Oi-Hartman-Abel effect that increased productivity dispersion raises output, which derived from the convexity of sales in productivity. Non-CES demand weakens this result by making sales less convex in productivity, and may even entirely reverse it (as it does for our estimated parameter values) if sales become concave in productivity.

80. At the estimated parameter values, scale is non-monotonic in productivity, and increases in productivity when productivity is low, then decreases in productivity when productivity is high.

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, denoted z_{ss}. As is to be expected, this form of non-CES demand (with $\eta_\varepsilon = 0$) leads to no difference in how demand shocks affect firm behavior relative to the CES model (when there are no adjustment costs). This is because demand shocks are a pure shifter here, and do not affect the elasticity of demand. Without adjustment costs, the price is optimally set only as some markup over marginal cost in both the CES and non-CES case, which is constant for the fixed level of z. Hence the price is unaffected by the demand shock in both the CES and non-CES cases. Sales, output, and capital are all simply linear in the demand shock in both cases. This explains why switching to non-CES demand has relatively minor effects on the transmission of the increase in demand uncertainty, as demand shocks affect the both models similarly.

D.5 Details of other exercises

Time-varying passthrough: In Table 35 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. In times of high dispersion, firms tend to wait and see, meaning that firms who do not receive idiosyncratic shocks allow their inputs to depreciate further than usual. However, the firms that do receive demand shocks tend to receive larger shocks. This makes them more likely to need to adjust their inputs in response to the shock, allowing them to keep their price roughly constant and reducing the correlation between price changes and demand shocks in times of high uncertainty. We also confirm this mechanism by repeating our passthrough regression which allows for different values for extreme shocks (final column of Table 37) on the model data. We find that, as in the data, passthrough is lower for larger values of the shocks in our model. This is true for both TFPQ and demand shocks, although for negative demand shocks the effect kicks in for the bottom 1% of shocks in our model, as opposed to the bottom 5% in the data. Nonetheless, 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 as part of the estimation.

Robustness exercises: In Figure 24 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 25 we plot the results for a very different calibration approach. In particular, we modify our demand curve in order to be able to very closely replicate the passthrough estimates we found in the data. Specifically, one concern with our baseline model is that, despite it generating passthrough much more in line with the data than a CES model, it still does not perfectly replicate the passthrough data. To investigate how important this is, we twist the demand curve in the model to force the model to match passthrough more closely. Of course, this exercise is not obviously superior to the main exercise, since while it does match passthrough more closely, it does so for a demand curve which is not formally estimated on the data. For this exercise, we use the extended demand curve (49). This demand curve allows the demand shock, ϵ, to not only shift the level of demand, but also the *elasticity* of demand. As seen in the static analysis without adjustment costs, this causes firms to permanently change their price when they receive a demand shock: see the passthrough in (54). Intuitively, if a positive demand shock also lowers the elasticity of demand, this will allow the firm to raise its price in response to the demand shock. This then lowers the increase in quantity sold following the demand shock. For this exercise, we manually set $\eta = 10$ and $\eta_\epsilon = 4.3$. This leads to estimated passthrough in the model (using the IV specification) of 21.8% for TFPQ and 17.8% for demand shocks. This brings the passthrough closer to the data, which had 24.0% for TFPQ and 23.5% for demand. The results are given in Figure 25. We see that the size of the output fall in response to the uncertainty shock is dampened relative to the baseline model. However, the main results are unchanged: the uncertainty effect is still mostly driven by demand shock uncertainty, and TFPQ dispersion still drives a negative OHA volatility effect.

Finally, in Figure 26 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. We see that the size of the output fall in response to the uncertainty shock is slightly dampened relative to the baseline model. Most noticeably, the size of the uncertainty effect becomes much smaller, reflecting the results of Bachmann and Bayer (2013) and Mongey and Williams (2017). However, the overall output fall from the shock falls by much less, because the volatility effect becomes much stronger. This reflects the reversed OHA effect, which is independent of wait and see behaviour. And with lower adjustment costs firms respond more to the rise in TFPQ dispersion, leading to larger output falls from the OHA effect. Despite these changes, the main results are still unchanged: the uncertainty effect is still mostly driven by demand shock uncertainty, and TFPQ dispersion still drives a negative OHA volatility effect.

D.6 Further model plots and tables

Table 35: Time-varying passthrough in the baseline model ($\eta = 5$)

	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

The 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 within the low and high uncertainty states respectively.

Table 36: Passthrough in the CES model ($\eta = 0$)

	$\log p$	$\log p$	$\Delta \log p$
$\log z$	-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

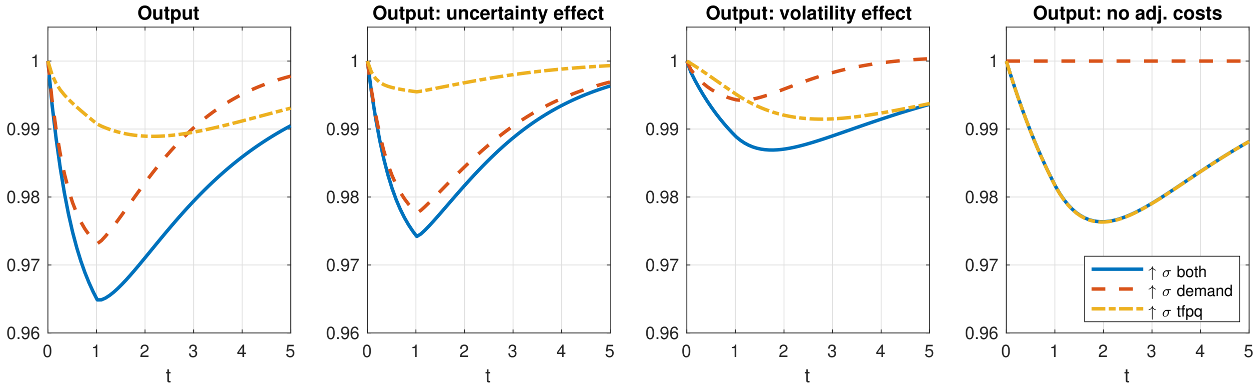
The table gives passthrough estimated on model simulated data for the CES version of the model. The data are time-aggregated to the yearly frequency. All coefficients are significant at at least the 0.1% level. The data are generated from long simulations of a single firm of 5,000 years in the steady state version of the model with constant uncertainty.

Table 37: Passthrough in the baseline model: extreme values

	$\Delta \log p$	$\Delta \log p$
$\Delta \log z$	-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%

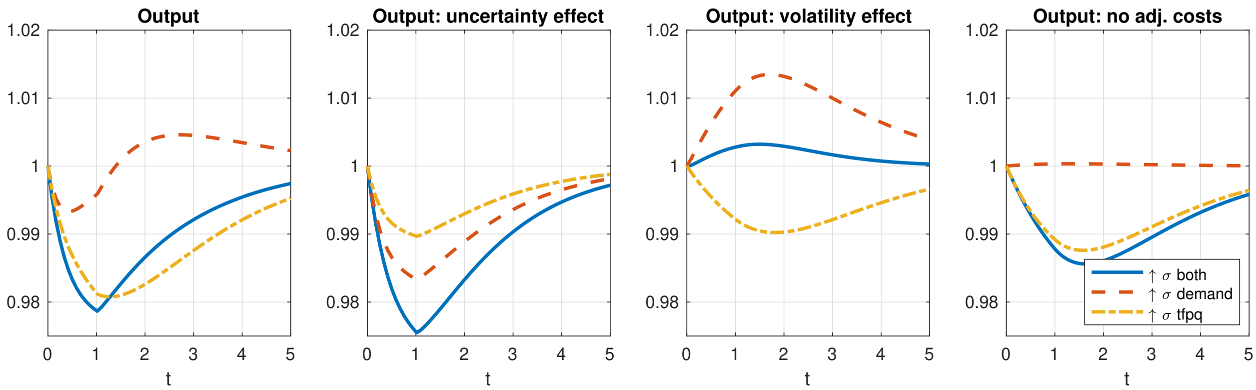
The table gives passthrough estimated on model simulated data, allowing for different coefficients for extreme values of the shocks. The data are time-aggregated to the yearly frequency. The first column defines extreme shocks as being in the bottom or top 5% of realised log changes, and the second column 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, which is insignificant at conventional levels. The data are generated from long simulations of a single firm of 5,000 years within the low uncertainty state.

Figure 24: Robustness: Bloom et al (2018) persistence of uncertainty regime



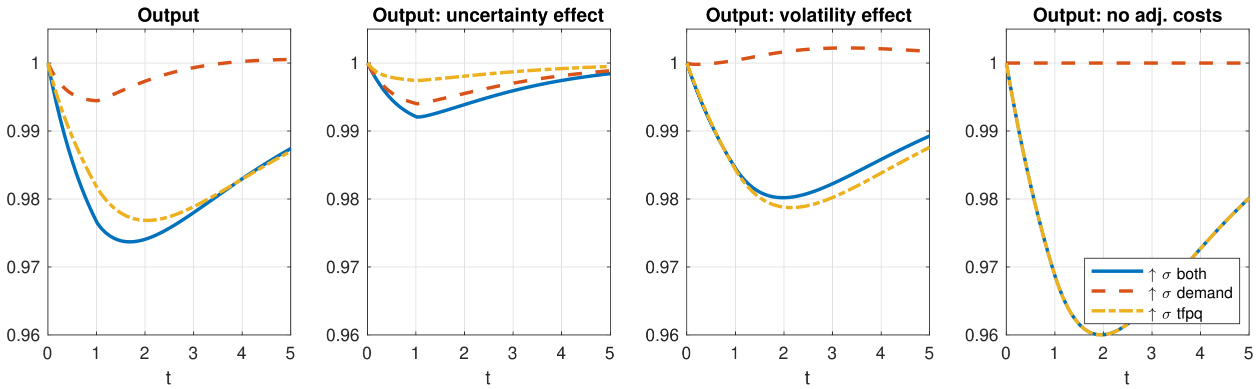
The plots give the aggregate response of the model to a switch to the high uncertainty state, $s = 2$, starting from the ergodic distribution when $s = 1$. Solid blue lines give the response to increased uncertainty in both shocks, dashed red is a version where only demand uncertainty rises in state 2, and dash-dotted yellow where only TFPQ uncertainty rises. 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.

Figure 25: Robustness: Model where demand shocks affect elasticity of demand



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. 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.

Figure 26: 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. 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.

Figure 27: Impulse response to idiosyncratic shocks: baseline model with $s = 1$

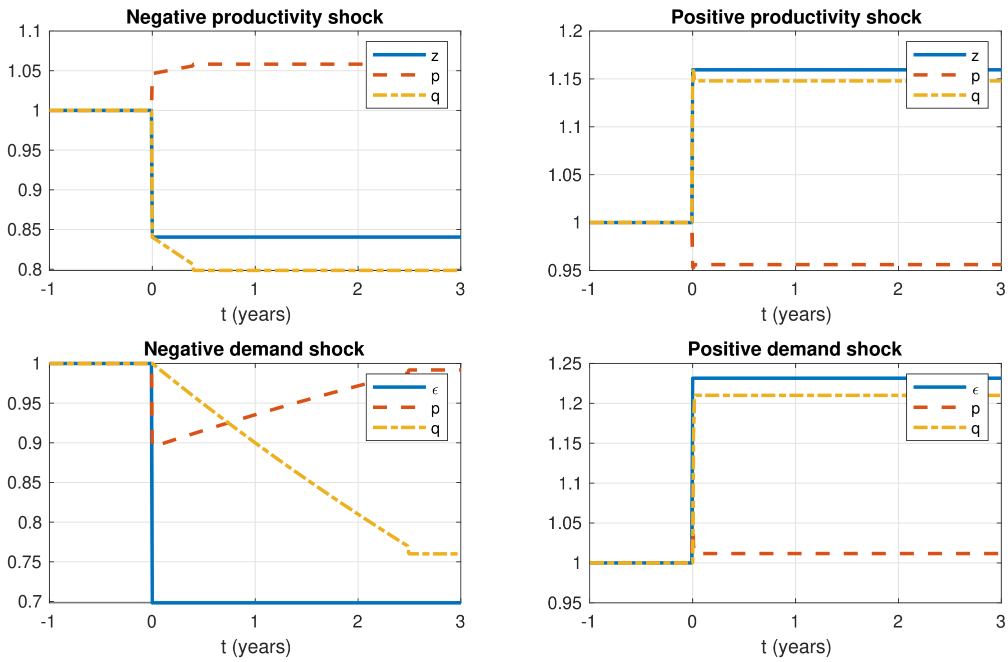
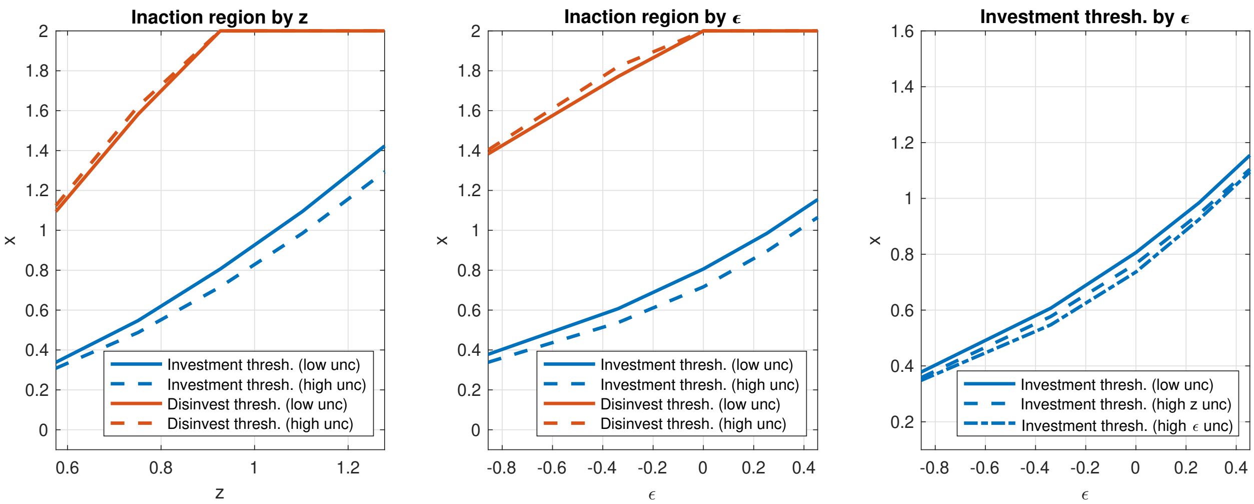


Figure plots the response of a firm to idiosyncratic demand and TFPQ shocks, represented as an impulse response. Top row gives the response to productivity shocks, and bottom to demand shocks. Left column gives response to negative shocks, and right to positive shocks. In all panels, the firm starts 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).

Figure 28: 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, $s = 2$). $\underline{k}(z, \epsilon, s)$ gives the investment threshold, such that firms have positive investment for current k below this value. $\bar{k}(z, \epsilon, s)$ gives the disinvestment threshold, such that firms disinvest for k above this value. For k 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. 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.