Market Power and Macroeconomic Fluctuations *

Matthias Gnewuch †

University of Bonn

Job Market Paper

[Click Here for the Most Recent Version]

November 1, 2022

Abstract

Crises affect firms unequally. For example, natural disasters disrupt only those firms that are located in a specific region. The current paper studies the aggregate effects of shocks to a subset of firms in many industries – referred to as *asymmetric supply shocks*. Based on a model with oligopolistic competition and firm heterogeneity, the paper shows that an economy with a lower intensity of competition among firms is less resilient to asymmetric supply shocks. The reason is the behavior of unharmed firms which face a higher demand for their goods. With more market power, these firms find it optimal to respond by raising prices more and expanding production less. Therefore, the volatility of both output and markups is higher when the economy is less competitive. The main mechanism is supported by evidence from firm-level as well as time-series data: Higher markups are associated with a higher volatility.

Keywords: market power, oligopoly, firm heterogeneity, asymmetric supply shocks, aggregate fluctuations, competition policy

JEL Classification: E32, E61, D43, L13

*I would like to thank Christian Bayer, Keith Kuester and Donghai Zhang for their invaluable support and guidance. Moreover, I would like to thank Rüdiger Bachmann, Florin Bilbiie, Jeff Campbell, Dean Corbae, Alessandro Ferrari, Basile Grassi, Matthias Hoelzlein, Joachim Jungherr, Julian Schärer as well as seminar participants at the University of Notre Dame, the RTG-2281 Summer School 2022, the University of Bonn and the 1st PhD Workshop on Imperfect Competition in Macroeconomics for helpful comments and suggestions. I thank the DFG for supporting my research as part of the research training group RTG 2281 "The Macroeconomics of Inequality".

†University of Bonn, Institute for Macroeconomics and Econometrics, Kaiserplatz 7-9, 53113 Bonn, Germany. Email address: matthias.gnewuch@gmail.com.

1 Introduction

Crises affect firms unequally. Typically, their direct effects concentrate on some subset of firms, as the following examples illustrate. Natural disasters, such as floods or earthquakes, disrupt the production only of firms that are located in a specific region. Shortages of natural gas concern only firms that rely on this particular source of energy in their production process. Financial crises directly affect only firms that rely on external financing to fund their operations. All of these supply disruptions are neither aggregate nor industry-specific. Instead, they affect some firms more than others *within many industries*. I collectively refer to such shocks as *asymmetric supply shocks*. The current paper investigates the aggregate effects of asymmetric supply shocks. Most importantly, the aggregate consequences are shown to depend on the intensity of competition among firms. A less competitive economy is less resilient to asymmetric supply shocks.

The current paper builds a model with oligopolistic competition and firm heterogeneity in order to study the aggregate effects of asymmetric supply shocks. I show analytically that a low intensity of competition among firms makes an economy more vulnerable to asymmetric supply shocks. The mechanism relies on the profit-maximizing behavior of firms. When an adverse shock, such as a natural disaster, disrupts the production of a subset of firms, their unharmed competitors consequently face a higher demand for their goods. When these firms have high market power, they find it optimal to primarily raise prices instead of expanding production. In contrast, firms with low market power primarily raise production, not prices, and thereby help to stabilize aggregate output. I calibrate the model to the U.S. economy and find that the welfare losses from asymmetric supply shocks increase substantially when the intensity of competition falls. This finding is particularly concerning in view of the rise in market power documented by De Loecker et al. (2020). I also derive implications for competition policy. The mechanism implies that fostering competition among firms not only reduces markups, but also stabilizes the economy. Finally, I test the main mechanism in firm-level as well as time-series data. The evidence supports the main mechanism: I find that higher markups are associated with a higher volatility.

More in detail, to investigate the aggregate effects of asymmetric supply shocks, I study model environments with two essential features. First, firms have market power and compete strategically within narrow industries. Second, competing firms are heterogeneous and there are asymmetric supply shocks. These shocks affect one or more firms differently than one or more other firms within industries and thereby change the distribution of sales across firms. To introduce strategic competition among firms, I build on the oligopoly framework of Atkeson and Burstein (2008). There is a large number of industries and, in each of them, a small number of firms which produce differentiated goods. Due to the limited number of firms in each industry, firms have market power and compete strategically. Their degree of market power, and thus their profit-maximizing production decision, depends on the intensity of competition. For example, when there are few competitors in an industry, firms produce little and sell at a high markup because consumers have few alternatives to their product.

For the main analysis, I consider a very tractable form of firm heterogeneity. That is, there are only two types of firms, active and inactive ones. The total number of firms is fixed, but the share of active firms fluctuates over time. Exogenous changes in this share of active firms constitute asymmetric supply shocks. Since all active firms are identical, a decrease in the number of active firms constitutes a change in the distribution of sales because fewer firms produce larger amounts each. Intuitively, changes in the share of active firms can be interpreted as the result of regional shocks, such as natural disasters or strikes. Each region is home to a share of firms of each industry. Therefore, when some region is hit by a natural disaster, the fraction of active firms in each industry falls.

In this framework, I analytically derive the aggregate effects of asymmetric supply shocks, i.e. changes in the fraction of active firms, in partial equilibrium. I show that a given change in the fraction of active firms has larger effects on aggregate output and the aggregate markup when the intensity of competition among firms is low. The reason is the profit-maximizing behavior of the remaining active firms, which suddenly face a higher demand and have more market power. A reduction in the number of active firms from 4 to 3 (i.e. by 25%) gives the remaining firms substantially more market power. Therefore, they raise prices substantially and expand production relatively little. As a result, the aggregate markup rises and aggregate output falls substantially. In contrast, a reduction in the number of active firms from 40 to 30 (i.e. also by 25%), gives the remaining firms only a small increase in market power. Thus, they barely increase prices and primarily expand production. Thereby, they help to stabilize aggregate output. This result is reminiscent of Gabaix (2011), even though the mechanism is distinct. Gabaix (2011) shows that the aggregate effects of idiosyncratic shocks dissipate when the number of firms becomes very large. This is intuitive as a shock to 1 firm out of 4 firms can be expected to have a larger aggregate effect than a shock to 1 firm out of 40 firms. In contrast, I consider shocks that affect a fraction of firms (1 out of 4 or 10 out of 40) and therefore do not vanish by a law of large numbers when the number of firms becomes very large.

A corollary to the main result is that when the number of firms becomes very large (high intensity of competition), asymmetric supply shocks become irrelevant for aggregate outcomes. This finding connects to the irrelevance of firm heterogeneity in models without market power but decreasing returns to scale, as discussed in Khan and Thomas (2008), Winberry (2021) and Koby and Wolf (2020). In these frameworks, firm heterogeneity becomes irrelevant when the returns to scale – governed by an exogenous parameter – are close to constant. The important difference is that the number of firms, which determines the relevance of asymmetric supply shocks in the framework used in this paper, is not policy-invariant and can in principle be affected by competition policy.

Next, I estimate the welfare cost to a representative household of asymmetric supply shocks in general equilibrium. In line with the partial equilibrium results, I find that the welfare cost increases exponentially when the intensity of competition falls. Moreover, I decompose the total welfare cost into two components. First, asymmetric supply shocks cause fluctuations in consumption and labor and thereby reduce welfare for a risk-averse household. Second, asymmetric supply shocks further reduce welfare by bringing average consumption below steady-state consumption. This happens because output is a concave function of the number of active firms. As both cost components ultimately result from the market power of firms, both become larger when the intensity of competition decreases.

Thereafter, I investigate optimal competition policy in the face of asymmetric supply shocks. I assume that a government authority chooses the intensity of competition. While I have discussed the benefits of a higher intensity of competition extensively, so far the model did not include a cost to a higher number of firms. Therefore, I now assume that each firm – active or not – incurs a per-period operating cost, similar to Jaimovich and Floetotto (2008). It is straightforward to see that optimal competition policy depends on the volatility of asymmetric supply shocks. When competition policy takes their presence into account, it optimally prescribes a higher number of firms and thereby makes consumption both higher on average and more stable.

This tractable form of firm heterogeneity with active and inactive firms is useful to derive and illustrate the aggregate effects of asymmetric supply shocks. However, I emphasize that the main mechanism applies to a much broader class of models with some form of firm heterogeneity and supply disruptions that can be considered asymmetric supply shocks, because they change the distribution of sales across firms in an industry.[1] In particular, there are many firm heterogeneity frameworks which make some firms within an industry more exposed to certain fluctuations than other firms. First, when firms in an industry are located across several regions, region-specific shocks, such as natural disasters, strikes, or country-specific productivity shocks as in Atkeson and Burstein (2008), reallocate sales across firms.

[1]Motivated by the COVID-19 crisis, Guerrieri et al. (2022) investigate the macroeconomic implications of shocks to a subset of sectors of the economy. Despite also being labelled *asymmetric* shocks, these shocks are different from the asymmetric supply shocks considered in this paper. I consider shocks which affect firms asymmetrically *within* a given industry, but are not necessarily asymmetric *across* industries.

Second, when firms in an industry use different inputs or have different production functions, they are differently exposed to price changes or the availability of inputs. Shortages of natural gas immediately only affect firms that use natural gas, instead of oil, as a source of energy. Moreover, firms with a relatively labor-intensive production process are more exposed to wage changes than relatively capital-intensive competitors. Third, in models with financial frictions, such as Khan and Thomas (2013) or Ottonello and Winberry (2020), financially constrained firms are more exposed to aggregate shocks than unconstrained firms, because aggregate shocks affect the tightness of financial constraints. Therefore, a financial tightening reallocates sales from constrained to unconstrained firms. In addition, frameworks with endogenous entry and exit, such as Bilbiie et al. (2012), or with changes in the distribution of idiosyncratic shocks, e.g. due to volatility shocks as in Bloom (2009) or skewness shocks as in Salgado et al. (2019), give rise to changes in the distribution of sales and thus asymmetric supply shocks.

Finally, I provide empirical evidence which supports the main mechanism. The key insight from the model analysis is that when the intensity of competition among firms is low, markups are not only *high*, but also *volatile*. According to the model, this positive relationship between the level and the volatility of markups holds at the firm-level, at the industry-level, and at the aggregate level. I test this prediction in firm-level micro data from Compustat as well as in aggregate time-series data. I investigate the volatility of markups instead of the volatility of output for two reasons. First, a low intensity of competition unambiguously predicts a higher volatility of industry-level and aggregate output, but the relationship with the volatility of firm-level output is ambiguous. Second, output is affected by symmetric supply and demand shocks, while markups are not.

Building on the work of De Loecker and Warzynski (2012), De Loecker et al. (2020) and Baqaee and Farhi (2020), I estimate annual firm-level markups for firms included in Compustat. Using these estimates, I show that there is a positive correlation between the firm-level average markup and the firm-level markup volatility, as predicted by the model. However, according to the model, this relationship is not linear, but convex. To test this relationship by means of OLS, I derive a linear relationship. In particular, I show that the model predicts a *linear* relationship between the average level and the volatility of the *inverse* markup. Across a range of empirical specifications, I find support for this relationship in the data.

Moreover, I assess the relationship between the level of the markup and its volatility in aggregate time-series data. I employ the widely-used model of Smets and Wouters (2007) as a "measurement device" in order to obtain a quarterly time series of the aggregate (target) markup. This variable evolves over time due to exogenous price-markup shocks for which asymmetric supply shocks can provide a micro-foundation. These shocks explain around

7.4% of fluctuations in consumption and therefore provide a quantitatively relevant source of aggregate volatility. The estimated time series of the aggregate markup captures the recent increase in markups documented in firm-level data (De Loecker et al., 2020). However, the series shows that the markup has not only been high in recent years, but was also high during the 1970s. Both periods of a high aggregate markup coincide with a high volatility of the aggregate markup, supporting the main mechanism of the model.

Related Literature. This paper relates to several strands of the literature. First and foremost, it relates to the literature studying the macroeconomic implications of (rising) market power, markups, and industry concentration. De Loecker et al. (2020) and Covarrubias et al. (2020) document substantial increases in markups, industry concentration, and profit rates in the United States over the past decades and explore the extent to which these trends can explain a number of stylized facts, such as falling labor shares. While there exists plenty of evidence for the presence of these trends, their extent as well as the level of the aggregate markup remain subject to debate, as discussed in Basu (2019).

It is well-understood that – even absent aggregate fluctuations – rising market power and markups reduce consumer welfare because they impair the efficiency of the economy. Edmond et al. (2018) quantify the welfare costs of markups in a model with heterogeneous firms and endogenously variable markups. De Loecker et al. (2021) decompose the observed rise in markups into changes in technology and changes in market structure. The increase in markups is undesirable in either case, but changes in technology can also have positive effects when production is reallocated to more productive firms. Covarrubias et al. (2020) refer to this as "good concentration" whereas higher markups due to a change in market structure, i.e. a decline in competition, is considered "bad concentration".

More recently, attention has shifted to the implications of (rising) market power and markups for macroeconomic fluctuations. Burstein et al. (2020) extend the granular macroeconomic model of Gabaix (2011) to oligopolistic competition and show that variable markups dampen the aggregate effects of idiosyncratic shocks. Mongey (2021) and Wang and Werning (2022) study monetary shocks in dynamic oligopolies with price rigidities due to menu costs and Calvo stickiness, respectively. Mongey (2021) finds that the oligopoly economy amplifies monetary non-neutrality, compared to a monopolistic competition benchmark. Wang and Werning (2022) conclude that higher concentration leads to a higher degree of monetary non-neutrality. Colciago and Silvestrini (2022) find that monetary policy shocks have a larger effect on the number of firms, but a smaller effect on average productivity in sectors with low concentration, due to endogenous entry and exit in a model with constant markups.

I contribute to this literature by showing that a high intensity of competition makes an economy more resilient to all shocks that affect firms in an asymmetric manner, which I refer to as asymmetric supply shocks. Asymmetric supply shocks arise in a broad class of models with firm heterogeneity and economic fluctuations that reallocate market shares. Closely related are Jaimovich and Floetotto (2008) and Corhay et al. (2020), who consider the non-linear relationship between markups and the number of active firms in frameworks with homogeneous active firms within industries, endogenous entry and exit, and aggregate TFP shocks. In comparison, I show that this non-linearity matters for a much broader class of models with some form of firm heterogeneity and asymmetric supply shocks. Another closely related paper is Ferrari and Queirós (2022), who argue that more concentrated economies are more fragile in a framework with firm heterogeneity, endogenous firm entry and exit, and aggregate TFP shocks. Amplification of aggregate TFP shocks occurs via entry and exit of firms, which affects endogenous measured productivity. The strength of this mechanism depends on the dispersion of firm productivity, because this determines the number of firms close to the exit threshold. In comparison, I show that more concentrated economies are less resilient to all shocks that can be considered asymmetric supply shocks in a broad class of models with some form of firm heterogeneity – even in the absence of endogenous entry and exit and changes in measured productivity.

Any economic disturbance that changes the distribution of sales among firms can be considered an asymmetric supply shock. Therefore, this paper relates to several other strands of the literature. In particular, it relates to a large body of work investigating the transmission of macroeconomic shocks in models with firm-level financial frictions. For example, the presence of financial frictions in the models of Khan and Thomas (2013), Khan et al. (2016), and Ottonello and Winberry (2020) implies that any aggregate shock has an asymmetric component. Consequently, the main mechanism described in this paper becomes relevant as soon as the assumption of atomistic firms without market power – a common simplification in models with firm financial heterogeneity – is dropped.

Finally, this paper relates to the literature on competition policy and the optimal intensity of competition in macroeconomic models. Bilbiie et al. (2019) discuss the optimal number of varieties in a model with endogenous product creation and monopolistic competition, as well as how to implement the optimal allocation using taxes on consumption and dividends. Edmond et al. (2018) investigate the welfare consequences of a variety of subsidies in a model with firm heterogeneity and endogenous entry. Boar and Midrigan (2019) characterize optimal product market policy in an economy in which firms with market power are owned by a subset of heterogeneous households.

Organization. The remainder of this paper is organized as follows. Section 2 presents the main model with oligopolistic competition, firm heterogeneity, and asymmetric supply shocks. Section 3 demonstrates analytically and quantitatively that the aggregate effects of asymmetric supply shocks are larger when the intensity of competition among firms is low. Moreover, optimal competition policy is discussed. In Section 4, I provide empirical evidence in support of the main mechanism using firm-level data as well as time-series data. Section 5 concludes.

2 Model

In this section, I build a general equilibrium model with oligopolistic competition and firm heterogeneity. The purpose of the model is to study how the intensity of competition among firms matters for the aggregate effects of asymmetric supply shocks.

The core of the model is a supply side with two main features. First, firms have market power and compete strategically within narrow industries. Building on the framework of Atkeson and Burstein (2008), there is a large number of industries, but a small number of firms in each of them. Second, firms are heterogeneous and there are shocks which change the distribution of firms within industries. These shocks are referred to as asymmetric supply shocks, because they affect one or more firms differently than one or more other firms. Thereby, they change the distribution of sales across firms in an industry.

In Section 2.1, I describe a simple and tractable industry setup which introduces firm heterogeneity and asymmetric supply shocks in a parsimonious manner. Despite its simplicity, this setup suffices to illustrate the main results in Section 3. In Section 2.2, I explain the broader class of firm heterogeneity setups to which the main results apply. Thereafter, I integrate the simple industry setup into the supply side structure in Section 2.3 and explain how firms compete strategically. Finally, the representative household, which constitutes the intentionally simplistic demand side of the model, is presented in Section 2.4.

2.1 Simple Industry Setup

There exists a large number of industries j and within each industry, there are $\widetilde{N_j}$ firms, which are indexed by $i \in \{1,...,\widetilde{N_j}\}$. Each firm ij produces the intermediate good y_{ij} according to a constant-returns-to-scale production technology

$$y_{ijt} = z_{ijt} l_{ijt} \tag{1}$$

where z_{ijt} is a firm-specific component and l_{ijt} is the labor input. Firms are heterogeneous due to the firm-specific component, which is a binary variable, i.e. $z_{ijt} \in \{0,1\}$. Thus, there are only two types of firms. Firms with $z_{ijt} = 0$ have a labor productivity of 0 and therefore optimally shut down in period t. Hence, I will refer to firms as *active* ($z_{ijt} = 1$) and *inactive* ($z_{ijt} = 0$). The fraction of active firms is λ_t, such that the number of active firms in industry j in period t is

$$N_{jt} = \lambda_t \widetilde{N_j} \tag{2}$$

The fraction of active firms, λ_t, fluctuates over time. These fluctuations in λ_t constitute the *asymmetric supply shocks* in this simple setup. Regardless of the remaining features of the economy, the equilibrium distribution of sales across firms within the industry changes when λ_t changes. As an example, consider an industry with $\widetilde{N_j} = 4$ firms. When three firms are active, the equilibrium distribution of sales shares must be $\{⅓,⅓,⅓,0\}$, because all active firms are identical. With only two active firms, the equilibrium distribution of sales shares is $\{½,½,0,0\}$.

Many supply disruptions may force some firms in each industry to temporarily shut down and thus serve as a micro-foundation for λ_t. Regional shocks, such as natural disasters or strikes, provide an intuitive example. If the $\widetilde{N_j}$ firms are distributed equally across a number of regions and each region is hit by a regional shock from time to time, λ_t reflects the share of active regions. A low value of λ_t means that more regions than usual are inactive.

2.2 Overview of Asymmetric Supply Shocks

Asymmetric supply shocks are defined as shocks which – within an industry – affect one or more firms differently than one or more other firms and thereby change the distribution of sales across firms. In the previous subsection, I have presented regional shocks with active and inactive firms as a tractable example. However, there exists a fairly broad class of firm heterogeneity frameworks which give rise to supply disruptions that can be considered asymmetric supply shocks. In this subsection, I organize and discuss some of these frameworks. In contrast to the simple setup, many of these examples feature heterogeneity among active firms. In Appendix A.1, I therefore present a generalized industry setup which does not impose restrictions on z_{ijt} and thus allows for heterogeneity among active firms. The main results, derived using the simple setup in Section 3, are shown to hold in the generalized setup in Appendix A.2.

Heterogeneous Exposure. Many firm heterogeneity frameworks make some firms within an industry more exposed to certain disruptions than other firms. Three groups of examples appear particularly relevant. First, when firms in an industry are located across several regions, region-specific shocks, such as natural disasters, strikes or regional lockdowns, reallocate sales across firms. In a framework with multiple countries, such as Atkeson and Burstein (2008), country-specific productivity shocks also fall under this category. Second, when firms in an industry use different inputs or have different production functions, they are differently exposed to price changes or the availability of inputs. Shortages of natural gas immediately only affect firms which use natural gas instead of oil as a source of energy. Lockdowns in some part of the world, such as China, only affect firms getting inputs from this region. Moreover, firms with a relatively labor-intensive production process are more exposed to wage changes than relatively capital-intensive competitors. Third, firm-level frictions, in particular financial frictions as in Khan and Thomas (2013), Khan et al. (2016) and Ottonello and Winberry (2020), make firms differently exposed to aggregate shocks. Financial shocks to the tightness of borrowing constraints as in Khan and Thomas (2013) or Khan et al. (2016) immediately affect only firms which are "financially constrained", in contrast to other "financially unconstrained" firms. A financial tightening would thus reallocate sales from constrained to unconstrained firms. Other aggregate shocks, such as monetary policy shocks in Ottonello and Winberry (2020), endogenously change the tightness of borrowing constraints and therefore set in motion the same mechanism.

Idiosyncratic Shocks. Idiosyncratic shocks to productivity, demand, capital quality, or some other firm-level state variable can also be interpreted as asymmetric supply shocks. Except for some special cases, idiosyncratic shocks reallocate market shares between the firm facing the idiosyncratic shock and all other firms in the industry which are not directly affected. However, idiosyncratic shocks only matter for *aggregate* outcomes when firms are not atomistic, e.g. as in Burstein et al. (2020). In contrast, when there is a continuum of industries, as in Atkeson and Burstein (2008), idiosyncratic shocks "wash out" and do not have aggregate effects. Yet, shocks to the distribution of these idiosyncratic shocks still do have aggregate effects, because they change the distribution of sales in all industries. Examples of these asymmetric supply shocks include shocks to the dispersion (e.g. Bloom 2009, Bachmann and Bayer 2014, Ferrari and Queirós 2022) or skewness (e.g. Salgado et al. 2019) of idiosyncratic shocks.

Extensive Margin. Closely related to the simple setup is a class of models with homogeneous active firms and endogenous fluctuations in the number of active firms due to endogenous firm entry and exit (e.g. Bilbiie et al. 2012, Jaimovich and Floetotto 2008). In these frameworks, aggregate shocks, e.g. to aggregate productivity, affect firm entry and exit decisions and therefore the number of active firms. Thus, an otherwise perfectly symmetric aggregate shock becomes an asymmetric supply shock due to its propagation via endogenous entry and exit. A similar mechanism is at work in a class of models which features firms that endogenously choose the number of industries to enter or markets to serve (e.g. Sedláček and Sterk 2017). Symmetric aggregate shocks now affect how many markets any firm serves, and therefore the number of active firms in any market. Again, there is an asymmetric propagation of otherwise symmetric shock.

In sum, what all of these examples have in common is that there is some form of firm heterogeneity and a shock which changes the distribution of sales within industries. In Section 3, I investigate the aggregate effects of such asymmetric supply shocks. Beforehand, I integrate the simply industry setup into the larger supply side structure.

2.3 Supply Side

To study how the intensity of competition among firms matters for the aggregate effects of asymmetric supply shocks, I need a framework which not only features firm heterogeneity and asymmetric supply shocks, but also firms with market power. I therefore integrate the simple industry setup of Section 2.1 into the oligopolistic competition framework of Atkeson and Burstein (2008). The simple setup suffices to illustrate the main results regarding the effects of asymmetric supply shocks. An analysis of other asymmetric supply shocks is relegated to Appendix A.

The production side of the economy consists of three layers. There is a competitive final consumption good producer, a continuum of industries, and in each industry a small number of firms producing differentiated intermediate goods.

Consumption Good Production. A competitive final consumption good producer aggregates the industry goods Y_{jt} of a continuum of industries $j \in [0,1]$ according to

$$Y_t^C = \left[\int_0^1 Y_{jt}^{\frac{\eta-1}{\eta}} \, dj \right]^{\frac{\eta}{\eta-1}} \quad \text{with} \quad \eta > 1 \tag{3}$$

where Y_t^C is the quantity of the final consumption good.[2] The parameter η captures the elasticity of substitution *across* industries.

[2]The price index for the final consumption good is given by $P_t^C = \left[\int_0^1 P_{jt}^{1-\eta} dj \right]^{\frac{1}{1-\eta}}$ where P_{jt} is the price index for industry j.

Industry Good Production. The industry good Y_{jt} is an aggregate of the intermediate goods y_{ijt} produced by the N_{jt} active firms in industry j

$$Y_{jt} = N_{jt}^{\frac{1}{1-\rho}} \left[\sum_{i=1}^{N_{jt}} y_{ijt}^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}} \quad \text{with} \quad \rho > 1 \tag{4}$$

where the term $N_{jt}^{\frac{1}{1-\rho}}$ neutralizes love of variety effects[3], as in De Loecker et al. (2021).[4] The parameter ρ captures the elasticity of substitution *within* industries.

Intermediate Good Production. Industries are modeled as introduced in Section 2.1. That is, intermediate good firms operate a constant-returns-to-scale production technology $y_{ijt} = z_{ijt} l_{ijt}$, where the firm-specific component $z_{ijt} \in \{0,1\}$ creates active and inactive firms. The number of active firms, N_{jt}, is the product of the fraction of active firms, λ_t, and the number of firms, $\widetilde{N_j}$. Since active firms are homogeneous, *in equilibrium* they produce the same amount, i.e. $y_{ijt} = y_{jt} \; \forall \; i \in \{1,...,N_{jt}\}$, where y_{jt} is the output of any firm in industry j. Moreover, it follows from combining this insight with equation (4) that in equilibrium industry output is $Y_{jt} = N_{jt} y_{jt}$.[5]

Firm Optimization. The objective of intermediate good firms is to maximize profits, d_{ijt}, which are defined by

$$d_{ijt} = \left(\frac{p_{ijt}}{P_t} \right) y_{ijt} - w_t l_{ijt} \tag{5}$$

where p_{ijt} the price charged by firm i in industry j, P_t is the price index for the final consumption good, and w_t is the real wage. Firms compete by choosing quantities (Cournot competition[6]) and face the demand curve

$$\frac{p_{ijt}}{P_t} = \left(\frac{y_{ijt}}{Y_{jt}} \right)^{-1/\rho} \left(\frac{Y_{jt}}{Y_t^C} \right)^{-1/\eta} N_{jt}^{-1/\rho} \tag{6}$$

which results from optimizing behavior of industry and consumption good producers.

[3]Neutralizing love of variety effects is not necessary for Proposition 1, but simplifies the exposition. Without a love of variety, a change in the number of firms only affects market power and markups while leaving measured productivity unchanged. Changes in measured productivity are discussed in more detail in Appendix A.

[4]The price index for the industry good is given by $P_{jt} = N_{jt}^{\frac{1}{\rho-1}} \left[\sum_{i=1}^{N_j} p_{ijt}^{1-\rho} \right]^{\frac{1}{1-\rho}}$ where p_{ijt} is the price of the intermediate good produced by firm i in industry j.

[5]With homogeneous active firms, in equilibrium it must be the case that $p_{ijt} = p_{jt} \; \forall \; i \in \{1,...,N_{jt}\}$, where p_{jt} is the price of any firm in industry j. Combining this insight with the price index for the industry good yields $P_{jt} = p_{jt}$.

[6]The main result, Proposition 1, holds also under Bertrand competition. See Burstein et al. (2020) for a discussion and comparison of Cournot and Bertrand competition in a similar framework.

Under optimal behavior, firms set a (gross) markup over marginal costs, μ_{ijt}, which depends on the number of active firms in the industry

$$\frac{p_{ijt}}{P_t} = \mu_{ijt}(N_{jt})\,\frac{w_t}{z_{ijt}} \tag{7}$$

The optimal markup is a function of the number of active firms, because the demand elasticity faced by firm i in industry j, $\epsilon_{ijt}(N_{jt})$, is a function of the number of active firms

$$\mu_{ijt}(N_{jt}) = \frac{\epsilon_{ijt}(N_{jt})}{\epsilon_{ijt}(N_{jt}) - 1} \qquad \text{where} \qquad \epsilon_{ijt}(N_{jt}) = \left[\frac{1}{\eta}\frac{1}{N_{jt}} + \frac{1}{\rho}\left(1 - \frac{1}{N_{jt}}\right)\right]^{-1} \tag{8}$$

The demand elasticity is a weighted harmonic average of the elasticity of substitution across industries, η, and the elasticity of substitution within industries, ρ. This reflects that firms compete both *within* industries, where the relevant elasticity of substitution is ρ, and *across* industries, where the relevant elasticity of substitution is η. Firms internalize that their actions affect not only their own demand, but also demand for the industry good. The weight given to the elasticity of substitution across industries, η, is $\frac{1}{N_{jt}}$, which equals the market share of a single firm in the industry. This reflects that when there are fewer firms, any one firm becomes larger and has a larger influence on industry demand. Therefore, the demand elasticity depends on the number of active firms.

Combining the optimal markup (8) with the demand curve (6) yields equilibrium firm output

$$y_{ijt} = \mu_{ijt}(N_{jt})^{-\eta} w_t^{-\eta} \frac{Y_t^C}{N_{jt}} \tag{9}$$

Aggregation. In equilibrium, all active firms in an industry choose the same output quantity, y_{ijt}, and the same markup, μ_{ijt}. Combining the equation for industry output with the equation for optimal firm-level output (9) yields

$$Y_{jt} = N_{jt} y_{ijt} = N_{jt} \mu_{ijt}(N_{jt})^{-\eta} \left(\frac{w_t}{Z_t}\right)^{-\eta} \frac{Y_t^C}{N_{jt}} \tag{10}$$

Moreover, the industry markup, which is defined as the ratio of industry sales to labor payments[7], is equal to the markup of the any active firm

$$\mu_{jt}(N_{jt}) = \mu_{ijt}(N_{jt}) \tag{11}$$

[7]Formally, as shown in Burstein et al. (2020), the industry markup, defined as $\mu_{jt} = \frac{(P_{jt}/P_t)Y_{jt}}{w_t L_{jt}}$ can be rewritten as a sales-weighted harmonic average of firm markups.

To keep the model as parsimonious as possible, I assume that all industries $j \in [0,1]$ are homogeneous. That is, the number of firms in each industry j is $\widetilde{N}_j = \widetilde{N}$ and the number of active firm in each industry is $N_{jt} = N_t = \lambda_t \widetilde{N}$. Therefore, in equilibrium, industry output and the industry markup are identical for all j, i.e. $Y_{jt} = Y_t$ and $\mu_{jt} = \mu_t \; \forall \; j \in [0,1]$, where Y_t is the industry output and μ_t the industry markup of any industry.[8] Combining this insight with equation (3), it follows that

$$Y_t^C = Y_t \tag{12}$$

Moreover, the aggregate markup, defined as the ratio of aggregate sales and labor payments, is equal to the industry markup

$$\mu_t^C = \mu_t \tag{13}$$

Finally, it is important to point out that aggregate productivity, TFP, in this economy is constant and thus not affected by asymmetric supply shocks:

$$TFP = \frac{Y_t^C}{L_t} = 1 \tag{14}$$

where $L_t = N_t l_{ijt}$. As discussed in Appendix A, there are of course examples of asymmetric supply shocks that do affect productivity.

2.4 Household

There is a representative household which consumes the final consumption good, C_t, supplies labor, L_t, and owns all firms in the economy. The household has Epstein-Zin preferences and maximizes

$$W_t = u(C_t, L_t) + \beta \left(\mathbb{E}_t W_{t+1}^{1-\alpha} \right)^{1/(1-\alpha)} \tag{15}$$

where the risk aversion parameter α allows specifying a coefficient of relative risk aversion which differs from the intertemporal elasticity of substitution.[9] The period utility function is standard,

$$u(C_t, L_t) = \frac{C_t^{1-\sigma}}{1-\sigma} + \psi \frac{(1-L_t)^{1-\chi}}{1-\chi} \tag{16}$$

The household maximizes (15) subject to a sequence of budget constraints

$$C_t = w_t L_t + D_t \tag{17}$$

[8]With homogeneous industries, in equilibrium $P_{jt} = P_t \; \forall \; j \in [0,1]$, where P_t is the price index of any industry. Combining this insight with the price index for the consumption good yields $P_t^C = P_t$.

[9]When $\alpha = 0$, the coefficient of relative risk aversion coincides with the intertemporal elasticity of substitution and Epstein-Zin preferences coincide with standard expected utility preferences.

where D_t subsumes dividends of all firms. Optimization gives rise to a standard wage-Euler equation

$$C_t^{\sigma}\psi(1-L_t)^{-\chi}=w_t \tag{18}$$

2.5 Stochastic Process

The only source of aggregate uncertainty are changes in the fraction of active firms, λ_t. λ_t must remain on the interval $(0,1]$, such that the number of active firms, N_t, remains above 0 and below $\widetilde{N}$.[10] To implement this, λ_t is the logistic transformation of an otherwise standard AR(1) process:

$$\epsilon_t^{\lambda}=(1-\rho_\lambda)\overline{\lambda}+\rho_\lambda\epsilon_{t-1}^{\lambda}+\sigma_\lambda\nu_t \qquad \text{with } \nu_t\sim\mathcal{N}(0,1) \tag{19}$$

$$\lambda_t=\frac{1}{1+e^{-(\epsilon_t^{\lambda}-\overline{\lambda})}} \tag{20}$$

where $\overline{\lambda}$ is the steady-state value of λ_t, σ_λ determines the volatility of shocks to λ_t, and ρ_λ their persistence.

3 Model Analysis

I now study the aggregate implications of asymmetric supply shocks. Throughout, I pay particular attention to how the intensity of competition among firms matters for the aggregate effects of these shocks. First, I derive some analytical results in partial equilibrium. Thereafter, I calibrate the model in order to obtain quantitative results in general equilibrium. Finally, I investigate optimal competition policy in the face of asymmetric supply shocks.

3.1 Analytical Results

I begin by characterizing analytically the effects of an asymmetric supply shock, i.e. a change in the fraction of active firms, λ_t. I do so in partial equilibrium, meaning that the real wage, w_t, and demand for the final consumption good, $Y_t^{C,D}$, are held fixed.

First, it is important to notice that the partial equilibrium effects on aggregate output and the aggregate markup are inextricably linked. To see this, consider the following decompo-

[10] To avoid integer constraints, I assume that the number of active firms, N_t, is a continuous variable. Therefore, λ_t can take any value on the interval $(0,1]$ and not only values from the set $[1/\widetilde{N}, 2/\widetilde{N}, ..., 1]$. Integer constraints are an issue only in the simple setup. In the generalized setup, outlined in Appendix A.1, the *effective* number of firms is anyways a continuous variable, because active firms can be heterogeneous.

sition of the elasticity of aggregate output (suppy), $Y_t^{C,S}$, with respect to λ_t:

$$\begin{aligned}
\frac{\mathrm{dlog}(Y_t^{C,S})}{\mathrm{dlog}(\lambda_t)} &= \underbrace{\frac{\mathrm{dlog}(N_t)}{\mathrm{dlog}(\lambda_t)}}_{\text{Direct Effect}} + \left(\underbrace{\frac{\mathrm{dlog}(y_t)}{\mathrm{dlog}(\mu_t)}\frac{\mathrm{dlog}(\mu_t)}{\mathrm{dlog}(N_t)}\frac{\mathrm{dlog}(N_t)}{\mathrm{dlog}(\lambda_t)} - \frac{\mathrm{dlog}(N_t)}{\mathrm{dlog}(\lambda_t)}}_{\text{Spillover Effect}}\right) \\
&= \underbrace{\frac{\mathrm{dlog}(y_t)}{\mathrm{dlog}(\mu_t)}}_{-\eta}\frac{\mathrm{dlog}(\mu_t)}{\mathrm{dlog}(N_t)}\underbrace{\frac{\mathrm{dlog}(N_t)}{\mathrm{dlog}(\lambda_t)}}_{1} \\
&= -\eta\frac{\mathrm{dlog}(\mu_t)}{\mathrm{dlog}(N_t)} \qquad (21)
\end{aligned}$$

The (positive) direct effect reflects that aggregate output increases because more firms are active. The (negative) spillover effect reflects that all active firms produce less. The total effect boils down to the elasticity of the markup with respect to the number of active firms and a constant. The same elasticity governs the effect of the asymmetric supply shock on the aggregate markup:

$$\frac{\mathrm{dlog}(\mu_t^C)}{\mathrm{dlog}(\lambda_t)} = \frac{\mathrm{dlog}(\mu_t)}{\mathrm{dlog}(N_t)}\underbrace{\frac{\mathrm{dlog}(N_t)}{\mathrm{dlog}(\lambda_t)}}_{1} \qquad (22)$$

Thus, the effects on the aggregate markup and aggregate output are closely linked, which comes as no surprise, given that both effects are the result of the firm-level price-quantity trade-off. The central elasticity of the markup with respect to the number of active firms is

$$\frac{\mathrm{dlog}(\mu_t)}{\mathrm{dlog}(N_t)} = -\frac{\mu_t}{N_t}\left(\frac{1}{\eta} - \frac{1}{\rho}\right) \qquad (23)$$

Under the standard parameter restriction $\rho > \eta$, an increase in the number of active firms decreases the markup.[11] From this it follows that an increase in λ_t increases aggregate output and reduces the aggregate markup:

$$\frac{\mathrm{dlog}(Y_t)}{\mathrm{dlog}(\lambda_t)} > 0 \qquad (24)$$

$$\frac{\mathrm{dlog}(\mu_t^C)}{\mathrm{dlog}(\lambda_t)} < 0 \qquad (25)$$

Importantly, however, the central elasticity of the markup with respect to the number of active firms (Equation 23) depends on the time-invariant intensity of competition among firms, captured by $\widetilde{N}$. This observation leads to the main analytical result, summarized in

[11] This parameter restriction states that the elasticity of substitution is higher within industries than across industries. Intuitively, the consumer is more willing to substitute a Coke and a Pepsi, than a soda and a t-shirt.

Proposition 1.

Proposition 1. *In a more competitive economy (higher number of firms $\widetilde{N}$), an asymmetric supply shock has a smaller absolute effect on aggregate output and the aggregate markup:*

Proof.

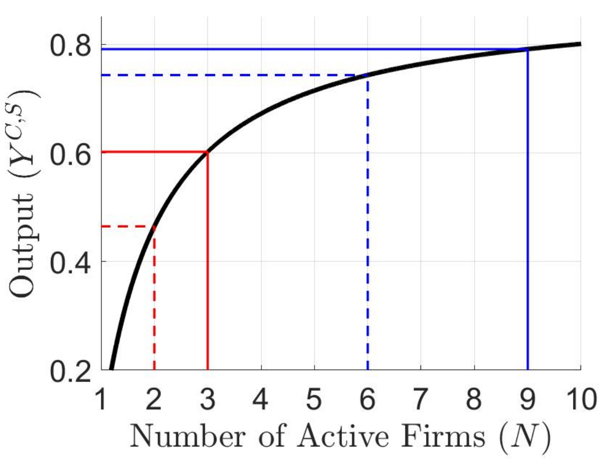
$$\frac{d\left(\frac{d\log(\mu_t)}{d\log(N_t)}\right)}{d\widetilde{N}} = \frac{\mu_t}{N_t}\left(\frac{1}{\eta}-\frac{1}{\rho}\right)\left[1+\frac{\mu_t}{N_t^2}\left(\frac{1}{\eta}-\frac{1}{\rho}\right)\right] > 0$$

From this, it follows that

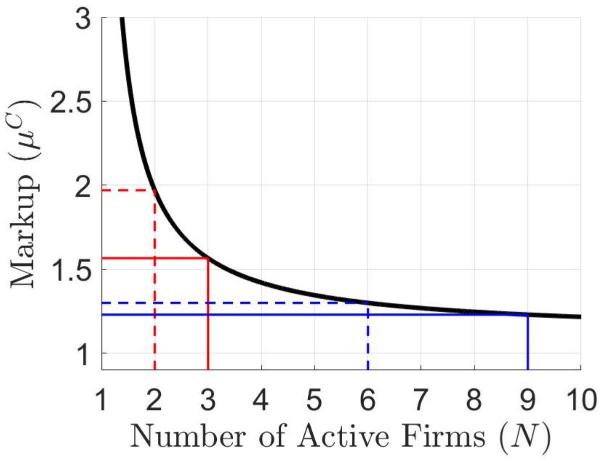
$$\frac{d\left(\frac{d\log(Y_t^{C,S})}{d\log(\lambda_t)}\right)}{d\widetilde{N}} = -\eta\frac{d\left(\frac{d\log(\mu_t)}{d\log(N_t)}\right)}{d\widetilde{N}} < 0$$
$$\frac{d\left(\frac{d\log(\mu_t^{C})}{d\log(\lambda_t)}\right)}{d\widetilde{N}} = \frac{d\left(\frac{d\log(\mu_t)}{d\log(N_t)}\right)}{d\widetilde{N}} > 0$$

□

Intuition. The intuition for Proposition 1 goes as follows. Suppose there is a decrease in the number of active firms by 50% (λ falls from 1 to 0.5). The direct effect is that industry output falls by 50%. However, the remaining 50% of firms now face more demand – both in absolute terms and relative to total industry demand. They respond to this by increasing output and by increasing prices (i.e. markups). The combination of the two depends on the increase in market power that the remaining firms experience, which again depends on the increase in their market share. In an economy with 100 firms per industry, the market share of remaining firms grows from 1% to 2%, which implies a fairly small increase on market power. In an economy with 2 firms per industry, the market share of the remaining firm grows from 50% to 100%, which constitutes a large increase in market power. Therefore, the remaining firm raises its markup by more and output by less than the remaining firms in the economy with 100 firms. In consequence, the total effect on aggregate output and the aggregate markup are larger in the economy with only 2 firms to start with. Figure 1 illustrates a similar example. Panel (a) plots aggregate output ($Y^{C,S}$) as a function of the number of active firms (N). The solid red line depicts an economy with a low intensity of competition, the number of firms being 3. The dashed red line illustrates the fall in output when the number of active firms falls by 33% (λ falls from 1 to ⅔). The solid blue line depicts an economy with a higher intensity of competition, the number of firms being 9. When the number of active firms falls by 33%, as illustrated with the blue dashed line, output falls, but much less than in the economy with a low intensity of competition. This illustrates the main result that a given shock (decrease in λ from 1 to ⅔) has smaller aggregate effects when the intensity of competition is high. Panel (b) plots the aggregate markup as a function of the number of active firms (N). Analogously to panel (a), the same shock has a smaller aggregate effect when the intensity of competition is high.



(a) Output



(b) Markup

Figure 1: Intuition for Proposition 1

Notes: This figure illustrates the effects of a drop in the number of active firms by 33%. The black lines depict aggregate output (left panel) and the aggregate markup (right panel) as a function of the number of active firms, using $\rho = 10$ and $\eta = 1.13$, which are the parameters used in the quantitative model below. The red lines refer to a low-competition economy with 3 firms, the blue lines refer to a high-competition economy with 9 firms. Solid lines depict the initial state ($\lambda = 1$), dashed lines depict the state after the drop in the number of active firms ($\lambda = 2/3$).

Irrelevance of Asymmetric Supply Shocks. A straightforward implication of Proposition 1 is that when the number of firms, $\widetilde{N}$, gets very large, asymmetric supply shocks become irrelevant for aggregate outcomes. This limit case is the familiar monopolistic competition setup in which firms charge a constant markup of $\mu_{ijt} = \frac{\rho}{\rho-1}$. This result, which is summarized in Corollary 1.1, connects to the literature on aggregation in heterogeneous-firm models, which has shown that when firms are atomistic – a common assumption – and profit functions become linear, firm-level frictions becomes irrelevant for aggregate outcomes. (Koby and Wolf, 2020; Winberry, 2021) Starting with Khan and Thomas (2008), the focus has been on firm-level capital adjustment costs generating "lumpy investment behavior", although Koby and Wolf (2020) have shown more recently that the irrelevance result also applies to firm-level financial frictions. In these frameworks, curvature in the profit function is governed by the exogenous parameter determining the degree of (decreasing) returns-to-scale. Thus, the curvature of profit functions is exogenous. Corollary 1.1 states that in the framework discussed in this paper, asymmetric supply shocks equally become irrelevant when profit functions become linear. However, as curvature of profit functions stems from market power, it is not exogenous, but endogenously depends on the intensity of competition among firms. This implies that the aggregate relevance of asymmetric supply shocks and firm heterogeneity more generally is not policy-invariant. In Section 3.4, I therefore discuss optimal competition policy in the face of asymmetric supply shocks.

Corollary 1.1. *When the number of firms becomes very large ($\widetilde{N} \to \infty$), asymmetric supply shocks become irrelevant for aggregate output and the aggregate markup.*

$$\lim_{\widetilde{N}\to\infty} \frac{dlog(Y_t^{C,S})}{dlog(\lambda_t)} = 0 \quad and \quad \frac{dlog(\mu_t^C)}{dlog(\lambda_t)} = 0 \tag{26}$$

Alternative Asymmetric Supply Shocks. As I show in Appendix A, it is straightforward to extend Proposition 1 to different firm heterogeneity frameworks and different asymmetric shocks. All that is necessary is to replace the number of active firms with the *effective number of firms*. Intuitively, the effective number of firms is the number of homogeneous firms which delivers the same industry concentration as a given distribution of heterogeneous firms.[12] Asymmetric supply shocks change the effective number of firms and Proposition 2 shows that a given change has larger aggregate effects when the intensity of competition is low.

3.2 Calibration

The model is calibrated as summarized in Table 1. The parameterization of the household follows Rudebusch and Swanson (2012). The discount factor is set to $\beta = 0.99$, which generates an annual real interest rate close to 4%. The labor disutility parameter, ψ, is chosen such that the household spends a third of its time endowment working. The curvature of the utility function with respect to consumption is set to $\sigma = 2$, which implies an intertemporal elasticity of substitution (IES) of 0.5. The curvature of the utility function with respect to labor is set to $\chi = 3$, which implies a Frisch elasticity of labor supply of $2/3$. The risk aversion parameter, is set to $\alpha = -148.3$. The resulting coefficient of relative risk aversion is 75 as in Rudebusch and Swanson (2012).

The steady-state fraction of active firms is normalized to $\overline{\lambda} = 0.5$. Then, the number of firms per industry is calibrated to $\widetilde{N} = 7.46$, such that the steady-state number of active firms in an industry matches the median effective number of firms in a market calculated in Mongey (2021).[13] The elasticity of substitution within industries is set to $\rho = 10$ as in Atkeson and Burstein (2008), Mongey (2021), and Wang and Werning (2022). The elasticity of substitution across industries is calibrated to $\eta = 1.13$ in order to generate a steady-state (sales-weighted) average markup of 1.45. This roughly corresponds to the average value

[12] Formally, the effective number of firms is the inverse of the Herfindahl–Hirschman concentration index.
[13] To calculate this number, a market is defined as an IRI product category within a state.

observed between 2000 and 2010 according to De Loecker et al. (2020). While being a fairly high markup value, it is still substantially below the latest value of 1.62 reported for the year 2016.[14] Finally, the volatility and persistence of asymmetric supply shocks are calibrated to match the observed fluctuations in the detrended (log) labor share. Note that in the model, the labor share is equal to the inverse of the gross markup. The persistence parameter is set to $\rho_\lambda = 0.95$, which generates an auto-correlation of 0.71. The volatility parameter is set to $\sigma_\lambda = 0.052$, which implies a standard deviation of 1.04%.[15]

Param.	Description	Value	Target / Source
Household			
β	Discount factor	0.99	$r^{ann} \approx 4\%$
ψ	Labor disutility	1.64	$L_{SS} = 1/3$
σ	Curvature of util. w.r.t. C	2	IES = 0.5
χ	Curvature of util. w.r.t. L	3	Frisch elasticity = 2/3
α	Risk aversion parameter	-148.3	CRRA = 75 (Rudebusch and Swanson, 2012)
Firms			
$\overline{\lambda}$	Share of active firms in SS	0.5	Normalization
$\widetilde{N}$	Number of firms per ind.	7.46	$N_{SS} = 3.73$ (Mongey, 2021)
ρ	Elast. of subst. within ind.	10	Atkeson and Burstein (2008)
η	Elast. of subst. across ind.	1.13	Avg. $\mu = 1.45$ (De Loecker et al., 2020)
ρ_λ	Persist. of fluct. in λ	0.95	$\rho(\log(\text{Labor Share}) = 0.71$ (detrended)
σ_λ	SD of innovations to λ	0.052	$\sigma(\log(\text{Labor Share})) = 1.04\%$ (detrended)

Table 1: Calibration

3.3 Welfare Cost of Asymmetric Supply Shocks

It is instructive to decompose the welfare cost of asymmetric supply shocks into two components. First, as shown previously in partial equilibrium, asymmetric supply shocks cause fluctuations in aggregate output and the aggregate markup. In general equilibrium, these fluctuations result in fluctuations in consumption and labor. Since the household is risk-averse, such fluctuations in consumption and labor reduce welfare. I refer to this component as the "volatility effect".

However, asymmetric supply shocks not only cause fluctuations in aggregate output and the aggregate markup, but also affect the *average* state of the economy. To see this, note that aggregate output is a concave function of the number of active firms, as shown in panel (a) of Figure 1. Due to this concavity, fluctuations in the number of active firms bring *average*

[14] There exists a wide range of estimates of the aggregate markup for the U.S. economy. This is due to difficulties in both, measuring and aggregating firm-level markups. See Edmond et al. (2018), Basu (2019), and De Ridder et al. (2021).

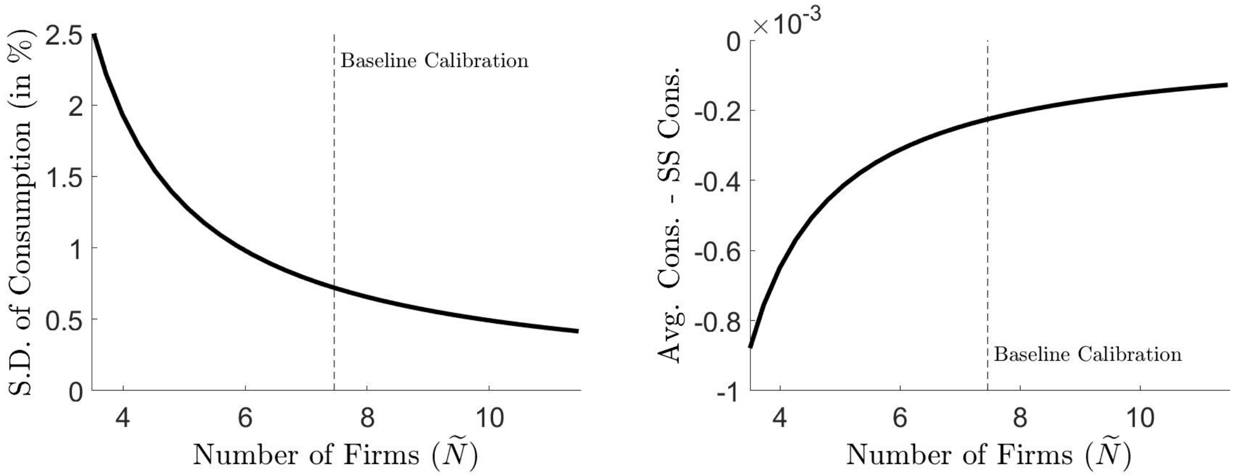
[15] To calculate these targets, I use the series "Nonfarm Business Sector: Labor Share for All Employed Persons" (PRS85006173) from FRED. Both the data and the model-generated data are detrended using an HP-filter with smoothing parameter $\lambda = 1600$.

output below *steady-state* output. By the same logic, the average markup exceeds the steady-state markup, because the markup is a convex function of the number of active firms. Since markups are distortionary and steady-state output and consumption are sub-optimally low, an even lower average consumption level reduces welfare. This component is referred to as the "mean effect".

The Role of Competition. The main point of this paper is that a higher intensity of competition among firms, which in this model is equivalent to a higher number of firms, is welfare-improving because it reduces the cost of asymmetric supply shocks. To support this argument, I first of all inspect both components of the welfare cost in isolation. The left panel of Figure 2 shows how the volatility of consumption, constituting the volatility effect, depends on the intensity of competition. Along the x-axis, the intensity of competition, i.e. the number of firms $\widetilde{N}$, changes while all other parameters are held fixed. Evidently, when there is less competition (low $\widetilde{N}$), the volatility of consumption increases in a convex manner. For example, when the number of firms is reduced by 50%, the volatility of consumption roughly triples. This shows that the main insight from Proposition 1 – competition makes the economy more resilient – holds in general equilibrium. The right panel of Figure 1 shows how the difference between average consumption and steady-state consumption, which constitutes the mean effect, depends on the intensity of competition. Very similar to the volatility effect, this effect increases in a convex manner when the intensity of competition falls. The similarity is not surprising in light of the fact that both effects are caused by the same non-linearity as shown in equation (21).

Finally, Figure 3 presents estimates of the welfare cost of asymmetric supply shocks for a range of competition intensities. At the baseline calibration, the household would be willing to give up around 0.035% of steady-state consumption to erase asymmetric supply shocks. Of this total cost, around 69% are due to the mean effect and 31% are due to the volatility effect. These modest numbers both for the total cost as well as for the volatility effect reflect the small cost of business cycles in models with a simple representative household, an observation dating back to Lucas (1987). Adding features such as countercyclical income risk as in Storesletten et al. (2001) or high and persistent individual consumption risk as in De Santis (2007) to the household would greatly amplify the cost of business cycles and therefore of asymmetric supply shocks. However, since these features would also complicate the analysis, I refrain from doing so and instead emphasize that the estimated welfare cost most likely presents a fairly low lower bound.

Either way, the main point of this paper is not about the level of the welfare cost of asymmetric supply shocks, but about how the cost varies with the intensity of competition



(a) Volatility Effect

(b) Mean Effect

Figure 2: Volatility & Mean Effects by Intensity of Competition

Notes: This figure illustrates how the cost of asymmetric supply shocks depends on the number of firms. The left panel plots the standard deviation of the log of aggregate consumption by the number of firms ("volatility effect"). The right panel plots the difference between average consumption and steady-state consumption by the number of firms ("mean effect"). The dashed vertical lines depict the baseline calibration ($\widetilde{N} = 7.46$).

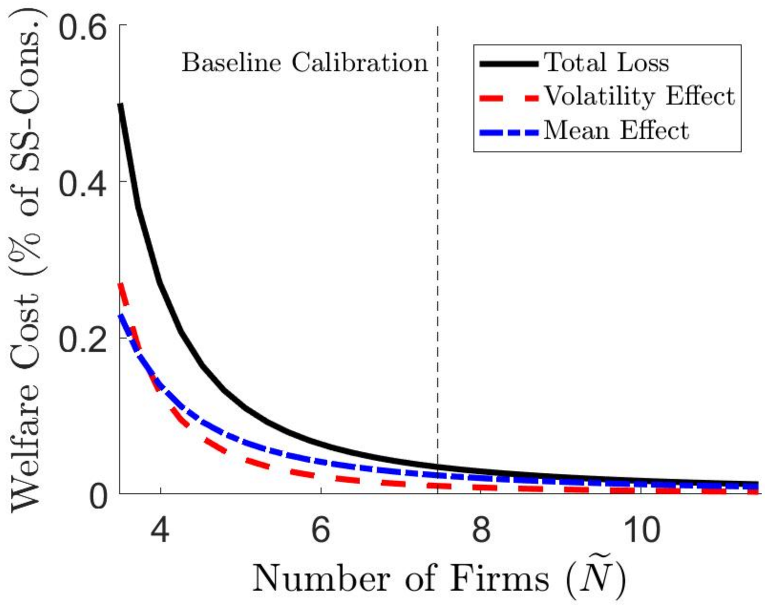


Figure 3: Welfare Cost of Asymmetric Supply Shocks by Intensity of Competition

Notes: This figure illustrates how the welfare cost of asymmetric supply shocks depends on the number of firms. The welfare cost is calculated as the share of steady-state consumption that the household would be willing to pay for the removal of asymmetric supply shocks. The black line depicts the total loss, the red line the volatility effect, and the blue line the mean effect. The dashed vertical line depicts the baseline calibration ($\widetilde{N} = 7.46$).

among firms. In line with Proposition 1 and the evidence in Figure 2, Figure 3 confirms that the welfare cost is monotonically decreasing in the number of firms. Note that this calculation does not include welfare gains from a higher/lower intensity of competition in steady state, but only from reducing the cost of asymmetric supply shocks. Two additional features of Figure 3 stand out. First, the welfare cost of asymmetric supply shocks is a convex function of the number of firms, reflecting the convexity of the markup itself, as evident from Figure 1. That is, increasing the number of firms by 1 reduces the welfare costs by 0.01 percentage points, whereas reducing the number of firms by 1 increases them by 0.017 percentage points. Second, the smaller the number of firms gets, the more important the volatility effect becomes. When the number of firms falls by 50% compared to the baseline calibration, the volatility effect even quantitatively dominates the mean effect.

3.4 Optimal Competition Policy

I now add to the model a government authority which decides on a competition policy. I assume that the government can choose a time-invariant number of firms in an industry, $\widetilde{N}$, and thereby the intensity of competition among firms. Thus far, there are benefits to competition, but no "cost" of having a high number of firms in the economy. Therefore, optimal competition policy would simply be $\widetilde{N} = \infty$. To make optimal competition policy less trivial, I henceforth assume that there is a cost to a having a high number of firms. Such a cost could arise because firms incur overhead operating costs as in Jaimovich and Floetotto (2008) or because there is firm churn and a sunk entry cost as in Bilbiie et al. (2012).

Costly Firms. I assume that each firm incurs a per-period operating cost δ, akin to the framework of Jaimovich and Floetotto (2008), but with two key differences. First, the cost is paid for and the number of firms is chosen in a welfare-maximizing way by the government. Thus, the number of firms is optimal, which is not necessarily the case if entry is a private decision. Second, the number of firms is time-invariant and therefore does not respond to shocks. The cost function takes the following functional form:

$$F_N = \delta \left(\widetilde{N} - \widetilde{N}_{oSS} \right) \tag{27}$$

The parameter determining the marginal cost of an additional firm is calibrated to $\delta = 0.0015$, such that the original steady state is socially optimal *in the absence* of asymmetric shocks. Moreover, the subtraction of the number of firms in the original steady state, $\widetilde{N}_{oSS}$, sets the total cost to zero, such that the steady state is exactly the same as without this cost. The government runs a balanced budget in each period and finances the operating cost with lump-sum taxes raised from the household

$$T_t = F_N \tag{28}$$

I assume that there are no other policy options available, such as a labor subsidy, which would solve the markup distortion entirely. (Bilbiie et al., 2019)

Quantitative Analysis. Absent asymmetric supply shocks, the government trades-off the cost of a high number of firms with the static benefit of a high number of firms. Absent shocks, a higher number of firms increases welfare, because it reduces markups and thereby the distortion in the household's consumption-labor decision. As explained in Bilbiie et al. (2019) among others, with positive markups, leisure is too cheap and therefore, labor supply and consumption are too low. The presence of asymmetric supply shocks adds two benefits of a high number of firms, as explained in the previous subsection, while leaving the cost unchanged. Therefore, adding shocks must lead the government to choose a higher intensity of competition. Quantitatively, I find that the government increases the number of firms in the economy by 1.1% when asymmetric supply shocks, calibrated as before, are introduced. Thereby, the planner reduces the standard deviation of (log) consumption by 1.45% (volatility effect) and decreases the gap between average and steady-state consumption by 1.53% (mean effect). In addition, steady-state consumption rises by 0.08% which reflects the static benefit of competition. Steady-state output rises by 0.11%. The gap between the changes in consumption and output is due to the higher total operating cost. As discussed above, these numbers should be interpreted as a lower bound given that the model features very low costs of business cycles.

4 Empirical Evidence

The main insight from the preceding analysis is that when the intensity of competition among firms is low, markups are not only *high*, but also *volatile*. According to the model, this positive relationship between the level and the volatility of markups holds at the firm-level, at the industry-level, and at the aggregate level. In this section, I test this prediction in firm-level micro data (Section 4.1) as well as aggregate time-series data (Section 4.2).

4.1 Evidence from Firm-Level Data

The fundamental source of the main result is the non-linear relationship at the firm-level between the markup and the market share.[16] Therefore, when facing the same shocks, firms with a higher markup due to a higher market share have a more volatile markup. I now turn to firm-level micro data from Compustat to investigate this relationship. Across a range of

[16] In the simple industry setup used in the model above, the market share only depends on the number of active firms ($s_{ijt} = \frac{1}{N_{jt}}$), which is why the markup was a function of the number of active firms only.

specifications, I indeed find that firms with higher average markups have more volatile markups, as predicted by the model.

Throughout, I focus on the volatility of firm-level markups instead of the volatility of firm-level output. While the model predicts a monotone relationship between the intensity of competition and the volatility of output at the industry-level and the aggregate level, this is not the case at the firm-level.[17] For the same reason, there is no monotone relationship between the intensity of competition and the volatility of firm-level sales.[18]

4.1.1 Data & Markup Estimation

I use annual firm-level data from Compustat North America. The data treatment is described in detail in Appendix B.1. Markups are estimated according to the production approach due to De Loecker and Warzynski (2012), which I summarize in Appendix B.2. Details of the implementation of this estimation are relegated to Appendix B.3. All steps broadly follow Baqaee and Farhi (2020) and De Loecker et al. (2020).

4.1.2 Basic Correlations

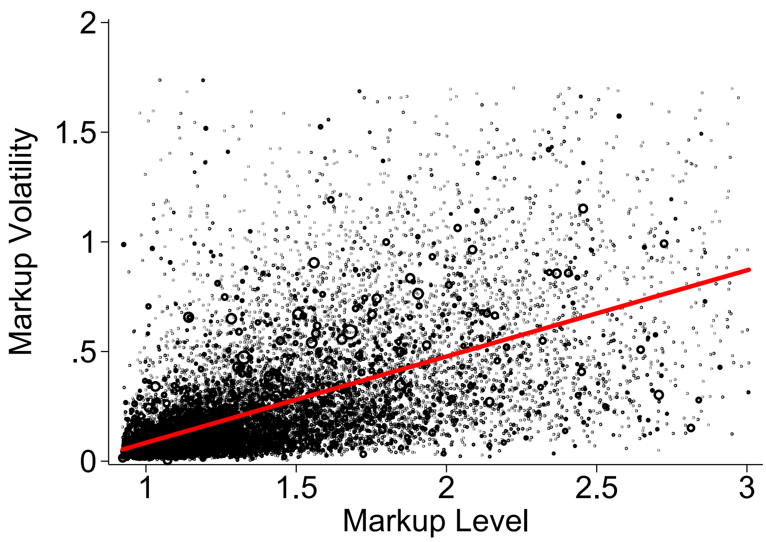


Figure 4: Firm-Level Markups & Markup Volatility

Notes: Each circle depicts one firm. The markup level is the median markup. The markup volatility is the interquartile range of the markup. Both variables are trimmed (1%). Firms are weighted by their average sales share. The red line shows the linear fit.

[17]To see this, note that the elasticity of firm-level output with respect to the number of active firms is $\frac{\text{dlog}(y_{ijt})}{\text{dlog}(N_{jt})} = -\eta \frac{\text{dlog}(\mu_{ijt})}{\text{dlog}(N_{jt})} - 1$, which can be positive or negative. Thus, a change in the intensity of competition, which changes $\frac{\text{dlog}(\mu_{ijt})}{\text{dlog}(N_{jt})}$, can increase or decrease the volatility of firm-level output.

[18]The elasticity of firm-level sales with respect to the number of active firms is $\frac{\text{dlog}(p_{ijt}y_{ijt})}{\text{dlog}(N_{jt})} = (1 - \eta)\frac{\text{dlog}(\mu_{ijt})}{\text{dlog}(N_{jt})} - 1$, which can be positive or negative.

First of all, I document basic correlations between the level and the volatility of firm-level markups. In the face of the caveats that come with the dataset and the estimation of markups, I primarily use measures that are not sensitive to outliers. As a measure of the level of the markup, I use the median markup of firm i over time. As a measure of the volatility of the markup, I use the interquartile range of markups of firm i.

Figure 4 displays the relationship between the level and the volatility of markups. There is a clear positive relationship, as predicted by the model. Table 2 confirms this positive relationship by regressing the interquartile range of markups on the median markup and a constant. There is an economically and statistically significant positive relationship, irrespective of whether firms are weighted by their average sales share (column 1) or not (column 2), or industry fixed effects are included (column 3). A potential concern might be that firm-specific trends induce a correlation between the level and the volatility of the markup. To address this, I compute the volatility after taking out a firm-specific linear trend (column 4) and from changes in markups (column 5). In addition, Figure A.2 and Table A.1 show that the results are robust to using log markups.

	(1) IQR (μ)	(2) IQR (μ)	(3) IQR (μ)	(4) IQR (μ)	(5) IQR ($d\mu$)
Median (μ)	0.393***	0.392***	0.375***	0.282***	0.173***
	(0.025)	(0.007)	(0.025)	(0.024)	(0.015)
Constant	-0.309***	-0.278***	-0.318***	-0.219***	-0.133***
	(0.030)	(0.009)	(0.058)	(0.029)	(0.018)
Observations	12282	12282	12282	12257	12259
R^2	0.386	0.307	0.456	0.344	0.357
Weights	Yes	No	Yes	Yes	Yes
Industry FE	No	No	Yes	No	No
Linear Trend	No	No	No	Yes	No

Table 2: Firm-Level Markups & Markup Volatility

Notes: Each column displays coefficients from a separate regression: $Volatility_i(\mu_{it}) = \beta_0 + \beta_1 * Level_i(\mu_{it}) + \epsilon_i$. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All variables are trimmed (1%).

4.1.3 Testing the Model

The model not only predicts that the volatility of the firm-level markup increases with the level of markup, but also that this relationship has a particular non-linear shape. To test whether there is evidence in the data for this particular shape, I derive a linear relationship, which can then be estimated by OLS, between a measure of the volatility and and a measure of the level of the markup from the model.

While the markup itself is a non-linear function of the market share (s_{ijt}), the inverse of the markup is a linear function thereof

$$\mu_{ijt}^{-1} = \frac{\rho - 1}{\rho} - \frac{\frac{\rho}{\eta} - 1}{\rho} s_{ijt} \tag{29}$$

This fundamental model equation cannot be estimated in Compustat data, because there is no information on the sales share in the relevant market.[19] Burstein et al. (2020) present evidence for this relationship in French administrative data, however.

The market share fluctuates around some long-run average ($\overline{s_{ij}}$)

$$s_{ijt} = \overline{s_{ij}} \epsilon_{ijt} \tag{30}$$

where ϵ_{ijt} follows some distribution.[20] It follows that the standard deviation of the inverse markup, $\sigma(\mu_{ijt}^{-1})$, can be written as

$$\sigma(\mu_{ijt}^{-1}) = \frac{\frac{\rho}{\eta} - 1}{\rho} \overline{s_{ij}} \sigma(\epsilon_{ijt}) \tag{31}$$

where $\sigma(\epsilon_{ijt})$ is the standard deviation of ϵ_{ijt}. Since I do not observe the market share, $\overline{s_{ij}}$, I use equation (29) to get

$$\sigma(\mu_{ijt}^{-1}) = \frac{\rho - 1}{\rho} \sigma(\epsilon_{ijt}) - \overline{\mu_{ij}^{-1}} \sigma(\epsilon_{ijt}) \tag{32}$$

Hence, the model predicts a negative linear relationship between the average level and the volatility of the *inverse* markup.

I estimate equation (32) to obtain an estimate of $\sigma(\epsilon_{ijt})$. To do so, I now use the standard deviation and mean as measures of the volatility and level of the markup. Table 3 displays the results. Across specifications, there is a significant negative relationship between the level and the volatility of inverse markups, as predicted by the model. The estimate for $\sigma(\epsilon_{ijt})$ ranges from 0.039 (column 4) to 0.054 (column 1). I include the same set of specifications as in Table 2, i.e. weighted by the average sales share (column 1), unweighted (column 2), with industry fixed effects (column 3), taking out a linear firm-level trend (column 4), and in changes (column 5).[21] Column 6 estimates the interquartile range of ϵ_{ijt} which should be

[19] The relevant market share would be the market share within a narrowly defined industry. Of course, it is possible to assign Compustat firms to industries, as done in Baqaee and Farhi (2020), but this is likely a far too coarse definition of a market.

[20] In the context of the simple model, the average market share would be $\overline{s_{ij}} = \frac{1}{N_j \overline{\lambda}}$ and the shock would be $\epsilon_{ijt} = \frac{\overline{\lambda}}{\lambda_t}$.

[21] Note that column 5 does not estimate $\sigma(\epsilon_{ijt})$, but $\sigma(d\epsilon_{ijt})$. These two statistics coincide only if there is no persistence in ϵ_{ijt}. The relatively smaller estimate for $\sigma(d\epsilon_{ijt})$ in column 5 suggests that there is persistence in ϵ_{ijt}, as assumed in the model outlined in Section 2.

	(1) SD (μ^{-1})	(2) SD (μ^{-1})	(3) SD (μ^{-1})	(4) SD (μ^{-1})	(5) SD ($d\mu^{-1}$)	(6) IQR (μ^{-1})
Mean (μ^{-1})	-0.054***	-0.054***	-0.053***	-0.039***	-0.021***	
	(0.008)	(0.003)	(0.008)	(0.007)	(0.006)	
Median (μ^{-1})						-0.099***
						(0.014)
Constant	0.117***	0.130***	0.110***	0.086***	0.065***	0.183***
	(0.007)	(0.002)	(0.010)	(0.006)	(0.005)	(0.012)
Observations	12250	12250	12250	12250	12249	12250
R^2	0.037	0.025	0.231	0.030	0.010	0.049
Weights	Yes	No	Yes	Yes	Yes	Yes
Industry FE	No	No	Yes	No	No	No
Linear Trend	No	No	No	Yes	No	No

Table 3: Firm-Level Inv. Markups & Inv. Markup Volatility

Notes: Each column displays coefficients from a separate regression: $Volatility_i(\mu_{it}^{-1}) = \beta_0 + \beta_1 * Level_i(\mu_{it}^{-1}) + \epsilon_i$. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All variables are trimmed (1%).

less sensitive to outliers.

In terms of magnitudes, the estimated $\sigma(\epsilon_{ijt})$ is somewhat smaller but certainly "in the same ballpark" as its model counterpart, $\sigma(\overline{\lambda}/\lambda_t) = 0.083$. The discrepancy might have several reasons. On the one hand, the model might overstate the volatility of asymmetric supply shocks by assigning *all* fluctuations in the labor share to them. On the other hand, the relationship might be understated in Compustat data, which is a very particular sample of firms.[22]

4.2 Evidence from Aggregate Data

According to the model, the positive relationship between the level and the volatility of markups holds not only at the firm-level, but also at the aggregate level. To test this relationship at the aggregate level, I now turn to time-series data.

However, measuring a time series of the aggregate markup for the U.S. economy is not straightforward. One approach is to aggregate the firm-level markups estimated for firms in Compustat, as done in De Loecker et al. (2020). There are, however, three issues with this method. First, an estimated firm-level markup is only available for a small and non-random

[22]To give an example why this relationship might be understated in Compustat, note that Compustat includes primarily large firms which most likely operate in more than one market. Ideally, we would then estimate the relationship of interest using market-specific markups instead of one firm-level markup. Of course, this is not possible with the data at hand. The firm-level markup, which is possible to estimate, will be an average of the market-specific markups and therefore have a lower volatility than the market-specific markups.

subsample of firms, i.e. firms in Compustat. Second, quarterly data is scarce in Compustat before the 1980s. Therefore, time series over long horizons can only be computed at an annual frequency. Third, there are different ways to aggregate firm-level markups that lead to very different aggregate patterns, as discussed in Edmond et al. (2018). Therefore, I choose a different, indirect approach to measuring the aggregate markup which avoids these issues. That is, I employ the widely-used medium-scale DSGE model of Smets and Wouters (2007) as a "measurement device" to obtain a quarterly time series for the aggregate markup from 1957 to 2019. An additional benefit of this approach is that it provides an estimate of the share of consumption fluctuations that can be attributed to asymmetric supply shocks. This is because asymmetric supply shocks can be interpreted as a micro-foundation for *price-markup shocks*, which are a common element of medium-scale DSGE models, as I explain next.

4.2.1 Asymmetric Supply Shocks and Price-Markup Shocks

The model presented in Section 2 does not include any price rigidities, which constitutes an important difference to New Keynesian models such as the one used in Smets and Wouters (2007). Therefore, firms always sell their products at the optimal (i.e., profit-maximizing) markup over the current marginal cost (see Equation 7), thus, at the optimal price. When some friction or cost to price adjustments is introduced, as in Smets and Wouters (2007), this is not the case anymore, and a gap between the actual markup, μ_{ijt}, and the optimal ("target") markup, μ^*_{ijt}, can arise. While the actual markup is affected by all sorts of shocks, the target markup fluctuates around its steady-state value only due to exogenous shocks, referred to as price-markup shocks.[23]

In the model of Section 2, the target markup, $\mu^*_{ijt} = \frac{\epsilon_{ijt}}{\epsilon_{ijt}-1}$ also fluctuates over time. However, its changes are not exogenous, but arise endogenously in frameworks with oligopolistic competition because asymmetric supply shocks affect the demand elasticity, ϵ_{ijt}.[24] Therefore, I henceforth interpret asymmetric supply shocks as a micro-foundation for price-markup shocks.[25] The similarity between endogenous markup fluctuations and "cost-push" shocks is also pointed out by Bilbiie et al. (2014).[26] I abstract from *interactions* between oligopolistic

[23]For example, in Smets and Wouters (2003), the target markup fluctuates around its steady-state value ($\mu^*_{SS} = 1 + \lambda_p$) according to: $\mu^*_t = 1 + \lambda_p + \nu^p_t$ where ν^p_t is i.i.d.-normal.

[24]In the simple model, the demand elasticity is a function of the number of active firms, N_{ijt}, see Equation (8). In the general model, it is a function of the market share, s_{ijt}, see Equation (34).

[25]Of course, ideally one would establish this link by calculating a correctly-weighted aggregate concentration measure and comparing its fluctuations with the estimated price-markup shocks. Gutiérrez et al. (2021) compute an aggregate concentration measure from Compustat firms for the period 1989 - 2015 (see their Figure A.5, Panel F), which aligns quite well with the time series of the price-markup shock estimated below (see Figure 5). However, to the best of my knowledge, there is no longer time series available or one which covers the entire firm sector.

[26]Bilbiie et al. (2014) make this point in a model in which the markup depends on the mass of firms due to

competition and price stickiness, as investigated in Mongey (2021) and Wang and Werning (2022).

4.2.2 Estimation

Assuming that price-markup shocks reflect asymmetric supply shocks, I take the estimation of a linearized medium-scale DSGE model in Smets and Wouters (2007) "off-the-shelf" in order to estimate a time series of the target markup and to get a sense of the quantitative relevance of price-markup shocks. As the linearization of the model eliminates any non-linearities, it is irrelevant that the model of Smets and Wouters (2007) does not feature the non-linearity highlighted in this paper, i.e. the aggregate markup becoming more volatile when its level is higher. In principle this non-linearity calls for a non-linear estimation, which however is outside the scope of this paper.

I estimate the linearized model of Smets and Wouters (2007) using an updated sample from 1957Q1 to 2019Q4 as in Bayer et al. (2020). The only adjustment of the model, also following Bayer et al. (2020), is the removal of the moving-average components of the wage- and price-markup shock processes. That is, I specify AR(1) processes for the wage- and price-markup shocks, instead of ARMA(1,1) processes.[27]

I find price-markup shocks (i.e., asymmetric supply shocks) to explain up to 7.4% of the fluctuations in consumption. Thus, these disturbances are a quantitatively relevant source of aggregate fluctuations. Due to the high persistence of the shocks, the share of fluctuations explained at short horizons is substantially lower than at longer horizons. These numbers are in line with the literature. For example, Smets and Wouters (2007) find price-markup shocks to explain up to 12% of fluctuations in output. In addition, Bayer et al. (2020) find that the importance of price-markup shocks increases once the representative household is replaced by heterogeneous households. Interestingly, they find price-markup shocks to also be a key driver of fluctuations in income and wealth inequality.

Figure 5 plots the smoothed time series for the target markup (black line) alongside the rolling-window standard deviation of innovations to the target markup (red line). Two observations stand out. First, the markup was high not only in recent years, but also during the 1970s. Second, both periods of a high markup coincide with a high standard deviation of innovations to the markup. The correlation of the two series is 0.49. This finding supports

a preference specification which features an "exponential love-of-variety".

[27]The price-markup shock process is estimated to be quite persistent, with a posterior mode of the autoregressive parameter of $\rho_\mu = 0.8665$, in line with Smets and Wouters (2007) and Bayer et al. (2020), who both find ρ_μ to be around 0.9. The posterior mode of the standard deviation of innovations to the price-markup shock is 0.0431, which is lower than the 0.14 estimated in Smets and Wouters (2007). Still, I estimate a higher volatility of the price-markup shock as the process does not include a moving-average term as in Smets and Wouters (2007).

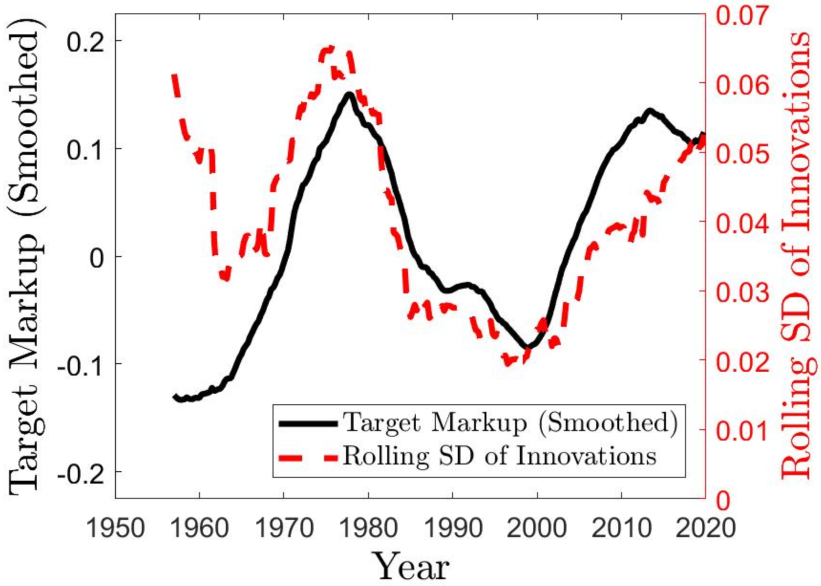


Figure 5: Markups & Markup Volatility in Estimated DSGE Model

Notes: The time series for the target markup results from the estimation of the model of Smets and Wouters (2007). Sample: 1957Q1 - 2019Q4. The smoothed target markup and the rolling-window standard deviation of innovations are both calculated over a symmetric 7-year window.

the prediction of the model that the volatility of the aggregate markup is higher when its level is high.

4.3 Discussion: Conditional Evidence

Both subsections, using firm-level data as well as time-series data, present *unconditional* correlations. Complementarily, Ferrari and Queirós (2022) present evidence *conditional* on a particular asymmetric supply shock. They show that after the financial crisis, which can be interpreted as an asymmetric supply shock (see Section 2.2), labor shares fell more strongly in industries which were more concentrated at the onset of the crisis. This suggests that markups increased more strongly in industries with a lower intensity of competition, which is in line with the main result of this paper.

5 Conclusion

In this paper, I have shown that a high intensity of competition among firms makes an economy more resilient to asymmetric supply shocks. The key mechanism relies on the profit-maximizing behavior of firms which compete strategically within narrow industries. In response to negative shocks to their competitors, firms with lots of market power find it optimal to not make up for the drop in total output, but to raise the prices of their goods instead. In contrast, firms with little market power find it optimal to primarily raise their production and thereby stabilize total output.

This mechanism provides an additional reason why the secular increases in market power, markups, and industry concentration, documented by De Loecker et al. (2020) and Covarrubias et al. (2020), are troubling. They not only reduce consumer welfare in static economies (Edmond et al., 2018), but also increase the extent of macroeconomic fluctuations, which further reduces welfare of risk-averse consumers. Therefore, competition policy must take into account that by leaning against those trends, it can not only reduce markups, but also provide macroeconomic stabilization.

I emphasize that the key mechanism is relevant for all supply disruptions that change the distribution of sales among firms within industries – and thereby change industry concentration and the "effective" number of firms. A broad class of models with firms heterogeneity gives rise to such disruptions, referred to as asymmetric supply shocks. Some well-known examples include models with firm financial heterogeneity (Khan and Thomas, 2013; Ottonello and Winberry, 2020), granular firm-level shocks (Burstein et al., 2020), or volatility shocks (Bloom et al., 2018).

In light of this wide range of supply disruptions that can cause fluctuations in market shares and industry concentration, two interesting questions for future research emerge. First, what are quantitatively the main drivers of changes in concentration at the industry-level and at the aggregate level? Second, do the main drivers of and the extent of fluctuations in concentration depend on the intensity of competition among firms, as has been observed for banks (Corbae and D'Erasmo, 2021)? Addressing both questions requires linking product market data (prices, quantities) with firm-level information (productivity, financial access) as done in Gilchrist et al. (2017) or Suveg (2021).

References

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

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031.

Bachmann, R. and C. Bayer (2014). Investment dispersion and the business cycle. *American Economic Review 104*(4), 1392–1416.

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics 135*(1), 105–163.

Basu, S. (2019). Are price-cost markups rising in the United States? A discussion of the evidence. *Journal of Economic Perspectives 33*(3), 3–22.

Bayer, C., B. Born, and R. Luetticke (2020). Shocks, frictions, and inequality in US business cycles. *CEPR Discussion Paper No. DP14364*.

Bilbiie, F. O., I. Fujiwara, and F. Ghironi (2014). Optimal monetary policy with endogenous entry and product variety. *Journal of Monetary Economics 64*, 1–20.

Bilbiie, F. O., F. Ghironi, and M. J. Melitz (2012). Endogenous entry, product variety, and business cycles. *Journal of Political Economy 120*(2), 304–345.

Bilbiie, F. O., F. Ghironi, and M. J. Melitz (2019). Monopoly power and endogenous product variety: Distortions and remedies. *American Economic Journal: Macroeconomics 11*(4), 140–74.

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

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018). Really uncertain business cycles. *Econometrica 86*, 1031–1065.

Boar, C. and V. Midrigan (2019). Markups and inequality. *NBER Working Paper No. 25952, National Bureau of Economic Research*.

Burstein, A., V. M. Carvalho, and B. Grassi (2020). Bottom-up markup fluctuations. *NBER Working Paper No. 27958, National Bureau of Economic Research*.

Colciago, A. and R. Silvestrini (2022). Monetary policy, productivity, and market concentration. *European Economic Review 142*.

Corbae, D. and P. D'Erasmo (2021). Capital buffers in a quantitative model of banking industry dynamics. *Econometrica 89*(6), 2975–3023.

Corhay, A., H. Kung, and L. Schmid (2020). Competition, markups, and predictable returns. *The Review of Financial Studies 33*(12), 5906–5939.

Covarrubias, M., G. Gutiérrez, and T. Philippon (2020). From Good to Bad Concentration? US Industries over the Past 30 Years. *NBER Macroeconomics Annual 34*, 1–46.

De Loecker, J., J. Eeckhout, and S. Mongey (2021). Quantifying market power and business dynamism in the macroeconomy. *NBER Working Paper No. 28761, National Bureau of Economic Research*.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics 135*(2), 561–644.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–71.

De Ridder, M., B. Grassi, and G. Morzenti (2021). The hitchhiker's guide to markup estimation. *Mimeo, London School of Economics*.

De Santis, M. (2007). Individual consumption risk and the welfare cost of business cycles. *American Economic Review 97*(4), 1488–1506.

Edmond, C., V. Midrigan, and D. Y. Xu (2018). How costly are markups? *NBER Working Paper No. 24800, National Bureau of Economic Research*.

Ferrando, A., P. McAdam, F. Petroulakis, and X. Vives (2021). Product market structure and monetary policy: evidence from the Euro Area. *ECB Working Paper No. 2632, European Central Bank*.

Ferrari, A. and F. Queirós (2022). Firm heterogeneity, market power and macroeconomic fragility. *Mimeo, University of Zurich*.

Furceri, D., R. Duval, M. M Tavares, and R. Lee (2021). Market power and monetary policy transmission. *IMF Working Papers 2021*(184).

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

Gilchrist, S., R. Schoenle, J. Sim, and E. Zakrajšek (2017). Inflation dynamics during the financial crisis. *American Economic Review 107*(3), 785–823.

Guerrieri, V., G. Lorenzoni, L. Straub, and I. Werning (2022). Macroeconomic implications of covid-19: Can negative supply shocks cause demand shortages? *American Economic Review 112*(5), 1437–74.

Gutiérrez, G., C. Jones, and T. Philippon (2021). Entry costs and aggregate dynamics. *Journal of Monetary Economics 124*, S77–S91.

Jaimovich, N. and M. Floetotto (2008). Firm dynamics, markup variations, and the business cycle. *Journal of Monetary Economics 55*(7), 1238–1252.

Khan, A., T. Senga, and J. K. Thomas (2016). Default risk and aggregate fluctuations in an economy with production heterogeneity. *Mimeo, The Ohio State University*.

Khan, A. and J. Thomas (2008). Idiosyncratic shocks and the role of nonconvexities. *Econometrica 76*, 395–436.

Khan, A. and J. K. Thomas (2013). Credit shocks and aggregate fluctuations in an economy with production heterogeneity. *Journal of Political Economy 121*, 1055–1107.

Koby, Y. and C. K. Wolf (2020). Aggregation in heterogeneous-firm models: A sufficient statistics approach. *Mimeo, Princeton University*.

Lucas, R. E. (1987). *Models of Business Cycles*. Blackwell.

Meier, M. and T. Reinelt (2022). Monetary Policy, Markup Dispersion, and Aggregate TFP. *Mimeo, University of Mannheim*.

Mongey, S. (2021). Market structure and monetary non-neutrality. *NBER Working Paper No. 29233, National Bureau of Economic Research*.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–97.

Ottonello, P. and T. Winberry (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica 88*, 2473–2502.

Rudebusch, G. and E. Swanson (2012). The bond premium in a DSGE model with long-run real and nominal risks. *American Economic Journal: Macroeconomics 4*(1), 105–43.

Salgado, S., F. Guvenen, and N. Bloom (2019). Skewed business cycles. *NBER Working Paper No. 26565, National Bureau of Economic Research*.

Sedláček, P. and V. Sterk (2017). The growth potential of startups over the business cycle. *American Economic Review 107*(10), 3182–3210.

Smets, F. and R. Wouters (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association 1*(5), 1123–1175.

Smets, F. and R. Wouters (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review 97*(3), 586–606.

Storesletten, K., C. I. Telmer, and A. Yaron (2001). The welfare cost of business cycles revisited: Finite lives and cyclical variation in idiosyncratic risk. *European Economic Review 45*(7), 1311–1339.

Suveg, M. (2021). Does firm exit raise prices? *IFN Working Paper No. 1414, Research Institute of Industrial Economics*.

Wang, O. and I. Werning (2022). Dynamic oligopoly and price stickiness. *American Economic Review 112*(8), 2815–49.

Winberry, T. (2021). Lumpy investment, business cycles, and stimulus policy. *American Economic Review 111*, 364–396.

Appendix

A Model Appendix

A.1 A Generalized Industry Setup

In this section, I describe a generalized industry setup that allows for various forms of firm heterogeneity and asymmetric supply shocks. The behavior of the industry good producers and the final consumption good producer remains unchanged. I continue to assume that within each industry j, there are $\widetilde{N_j}$ firms, which are indexed by $i \in \{1, ..., \widetilde{N_j}\}$. Each firm ij produces the intermediate good y_{ij} according to the constant-returns-to-scale production technology $y_{ijt} = z_{ijt} l_{ijt}$.

Firm heterogeneity still originates from the firm-specific component z_{ijt}. The crucial difference with respect to the simple industry setup in Section 2.1 is that no restrictions are imposed on z_{ijt}. In particular, z_{ijt} is not restricted to be a binary variable anymore. Many types of firm heterogeneity and asymmetric supply shocks map into this setup with appropriate choices for the firm-specific component. Some important examples, including financial shocks, are discussed below in Section A.3. Beforehand, I explain firm behavior and aggregate outcomes in this industry setup.

Firm Optimization. Firms continue to maximize profits (Equation 5) under Cournot competition subject to the demand curve (Equation 6). Under optimal behavior, firms still set a markup over marginal costs which depends on the intensity of competition in their industry (see Equation 7). However, with firm heterogeneity among active firms, the number of active firms, N_{jt}, is not sufficient to characterize the intensity of competition and market power anymore. Instead, a firm's market power and thus its optimal markup now depends on its market share defined by

$$s_{ijt} = \frac{p_{ijt} y_{ijt}}{P_{jt} Y_{jt}} \tag{33}$$

In particular, the optimal markup (see Equation 8) becomes

$$\mu_{ijt}(s_{ijt}) = \frac{\epsilon_{ijt}(s_{ijt})}{\epsilon_{ijt}(s_{ijt}) - 1} \qquad \text{where} \qquad \epsilon_{ijt}(s_{ijt}) = \left[\frac{1}{\eta}s_{ijt} + \frac{1}{\rho}(1 - s_{ijt})\right]^{-1} \tag{34}$$

As discussed in Section 3.1, in partial equilibrium, firms' price-setting and production decisions are inextricably linked. In general equilibrium, production is demand-determined and firms supply any quantity at the price p_{ijt}. Therefore, I focus on the markup (price) decision in this section. It is helpful to rewrite the optimal markup (Equation 34) as

$$\mu_{ijt} = \frac{\rho}{\rho - 1}\left[1 - \frac{\frac{\rho}{\eta} - 1}{\rho - 1}s_{ijt}\right]^{-1} \tag{35}$$

In addition, combining the definition of the market share (Equation 33) with the price equation (Equation 7) and the demand curve (Equation 6), yields the following expression for the market share

$$s_{ijt} = \frac{z_{ijt}^{\rho-1}\mu_{ijt}^{1-\rho}}{\sum_{k=1}^{\widetilde{N}_j} z_{kjt}^{\rho-1}\mu_{kjt}^{1-\rho}} \tag{36}$$

Given the firm-specific component z_{ijt} for all firms i, equations (35) and (36) can be used to solve for all firms' market shares, s_{ijt}, and markups, μ_{ijt}, in period t. Importantly, this is possible regardless of the features of the distribution of firm-specific components, z_{ijt}.

Industry Aggregates. The industry markup, defined by $\mu_{jt} = \frac{(P_{jt}/P_t)Y_{jt}}{w_t L_{jt}}$, can be rewritten, using Equation (35), as a function of the Herfindahl–Hirschman index (HHI), a measure of industry concentration

$$\mu_{jt} = \frac{\rho}{\rho - 1}\left[1 - \frac{\frac{\rho}{\eta} - 1}{\rho - 1}HHI_{jt}\right]^{-1} \tag{37}$$

where the HHI is calculated as the sum of squared market shares, $HHI_{jt} = \sum_{i=1}^{\widetilde{N}_{jt}} s_{ijt}^2$.

In the case of asymmetric supply shocks with heterogeneity among active firms, there is not only an effect on the industry markup μ_{jt}, but also on industry productivity Z_{jt}. Industry productivity is defined by

$$Z_{jt} = \frac{Y_{jt}}{L_{jt}} = \frac{N_{jt}^{\frac{1}{1-\rho}}\left[\sum_{i=1}^{\widetilde{N}_{jt}} \mu_{ijt}^{1-\rho}\right]^{\frac{\rho}{\rho-1}}}{\sum_{i=1}^{\widetilde{N}_{jt}} \mu_{ijt}^{-\rho}} \tag{38}$$

where again $N_{jt}^{\frac{1}{1-\rho}}$ is the term arising from the cancellation of love of variety effects.[28] It is easy to see that with symmetric firms, $Z_{jt} = 1$ and changes in N_{jt}, as in the main text, do not affect Z_{jt}. However, as soon as there are love-of-variety effects or active firms are heterogeneous, industry productivity is affected by asymmetric supply shocks. The analytical results below focus on the effect on markups, however.

Aggregation of Industries. In the following, I discuss the effect of asymmetric supply shocks on the industry markup, μ_{jt}. Under the assumption that all industries are identical, as made in the main text, the aggregate markup equals the industry markup: $\mu_t^C = \mu_{jt}$.[29] Thus, the effect of an asymmetric supply shock on the industry markup equals the effect on the aggregate markup. However, in principle the setup allows for industry heterogeneity and thus shocks which affect only a subset of industries.

A.2 Analytical Results

To see that the analytical results derived in the main paper generalize to this more general framework, it is helpful to define the effective number of firms as

$$N_{jt}^{eff} = HHI_{jt}^{-1} = \left(\sum_{i=1}^{\widetilde{N}_{jt}} s_{ijt}^2 \right)^{-1} \tag{39}$$

Intuitively, the effective number of firms is the number of homogeneous firms which results in the same intensity of competition (and thus the same industry concentration) as a given distribution of heterogeneous firms. Using Equation (37), the industry markup is

$$\mu_{jt} = \frac{\rho}{\rho - 1} \left[1 - \frac{\frac{\rho}{\eta} - 1}{\rho - 1} \left(N_{jt}^{eff} \right)^{-1} \right]^{-1} \tag{40}$$

whereas in the baseline model, the industry markup can be written as

$$\mu_{jt} = \frac{\rho}{\rho - 1} \left[1 - \frac{\frac{\rho}{\eta} - 1}{\rho - 1} N_{jt}^{-1} \right]^{-1} \tag{41}$$

Comparing these two equations, it is obvious that a change in the effective number of firms in the generalized setup has exactly the same effect on the industry markup as a change in

[28] As before, the number of active firms is defined as the number of firms with a positive market share. All firms with $z_{ijt} > 0$ have a positive market share.

[29] Moreover, assuming that all industries are identical, aggregate productivity equals industry productivity: $Z_t = Z_{jt}$. See Burstein et al. (2020) for a more detailed discussion of industry heterogeneity in a similar framework.

the number of active firms in the baseline model. Proposition 2 therefore extends Proposition 1 to the generalized setup.

Proposition 2. *In a more competitive industry (higher steady-state effective number of firms $\widetilde{N_j^{eff}}$), an asymmetric supply shock (log-change in effective number of firms) has a smaller absolute effect on the industry markup:*

$$\frac{d\left(\frac{dlog(\mu_{jt})}{dlog(N_{jt}^{eff})}\right)}{d\widetilde{N_j^{eff}}} > 0$$

Proof. This follows immediately from replacing N with N^{eff} in the first line of the proof of Proposition 1. □

Numerical Example. Intuitively, Proposition 2 shows that a given percent change in the effective number of firms (industry concentration) has a larger effect on the industry markup when the steady-state effective number of firms (industry concentration) is low to begin with. To provide a numerical example, consider an industry with two firms in which market shares are reallocated from $s = \{0.5, 0.5\}$ to $s = \{0.4, 0.6\}$. The HHI increases from 0.5 to 0.52, i.e. by 4%. In consequence, the industry markup increases by 3.2%, using the parameter values from Table 1. In contrast, if the industry was populated by four firms, i.e. splitting each firm in two, the same reallocation of market shares would be from $s = \{0.25, 0.25, 0.25, 0.25\}$ to $s = \{0.2, 0.2, 0.3, 0.3\}$. The HHI increases from 0.25 to 0.26, which again is an increase of 4%. However, the resulting change in the industry markup would only be 1.13%, so a lot less than in the industry with only two firms, in line with Proposition 2.

Discussion. Many economic disturbances cause changes in industry concentration as in the numerical example above. Some of these are discussed in more detail below in Section A.3. Depending on the example, the effect of an asymmetric supply shock on industry concentration may however also depend on the intensity of competition. To see this, consider the following decomposition of the effect of an arbitrary asymmetric supply shock ϵ_t^A on the industry markup:

$$\frac{\mathrm{dlog}(\mu_{jt})}{\mathrm{dlog}(\epsilon_{jt}^A)} = \underbrace{\frac{\mathrm{dlog}(\mu_{jt})}{\mathrm{dlog}(HHI_{jt})}}_{A} \underbrace{\frac{\mathrm{dlog}(HHI_{jt})}{\mathrm{dlog}(\epsilon_t^A)}}_{B} \tag{42}$$

Proposition 2 has established that the former elasticity ("A") is decreasing in the intensity of competition.[30] However, whether the latter elasticity ("B") also depends on the intensity of competition, depends on the precise example. In the baseline model, this was not the case, because $\frac{\text{dlog}(N_t)}{\text{dlog}(\lambda_t)} = 1$ (see Equation 22). In the financial-friction-example below, this is the case. However, in sum, the elasticity of the industry markup with respect to the asymmetric supply shock continues to decrease in the intensity of competition.

A.3 Examples of Asymmetric Supply Shocks

The previous subsection has shown that a given percent change in industry concentration (or, the effective number of firms) has larger aggregate effects when industry concentration is high (or, the effective number of firms is low) to begin with. Now, I show that many types of firm heterogeneity map into the generalized setup explained in Section A.1 and give rise to changes in industry concentration, such that they can be considered asymmetric supply shocks. I first discuss in detail a framework with firm heterogeneity due to financial frictions. Thereafter, I briefly describe additional examples.

Financial Frictions. In heterogeneous-firm models with financial frictions, such as Khan and Thomas (2013), Khan et al. (2016) or Ottonello and Winberry (2020), aggregate shocks affect concentration via their effect on the tightness of the financial constraint. To illustrate this, I consider a model in which firms produce, using capital and labor, according to the production function $y_{ijt} = k_{ijt}^{\theta} l_{ijt}$ with $\theta < 1$. Capital is purchased at the end of the previous period, such that at time t, only labor can be adjusted. Defining $z_{ijt} = k_{ijt}^{\theta}$, it is evident, that this example maps into the generalized industry setup outlined above. In addition, I assume that within each (homogeneous) industry there are two types of firms: financially constrained firms and financially unconstrained ones. Financially unconstrained firms ("u") choose the optimal level of capital k_{jt}^{*}. I normalize $z_{ujt} = k_{jt}^{*\,\theta} = 1$. Financially constrained firms ("c") can only purchase capital up to a (always binding) limit, γ_t, such that $k_{cjt} \leq \gamma_t \leq k_{jt}^{*}$. Hence, $z_{cjt} = \gamma_t^{\theta} \leq 1$.

In this setup, financial shocks, i.e. changes in the tightness of the financial constraint γ_t, directly affect financially constrained firms (z_{cjt}), but not financially unconstrained firms (z_{ujt}). To illustrate the effects, I provide a numerical example: All firms are initially of the same size ($z_{cjt-1} = 1$) and a negative financial shock hits, such that $z_{cjt} < 1$. I compare two economies, one with a high intensity of competition ($N = 10$) and one with a low intensity of competition ($N = 5$). In both economies, 20% of firms are financially constrained. Figure

[30] Note that $\text{dlog}(HHI_{jt}) = \text{dlog}(N_{jt}^{eff})$.

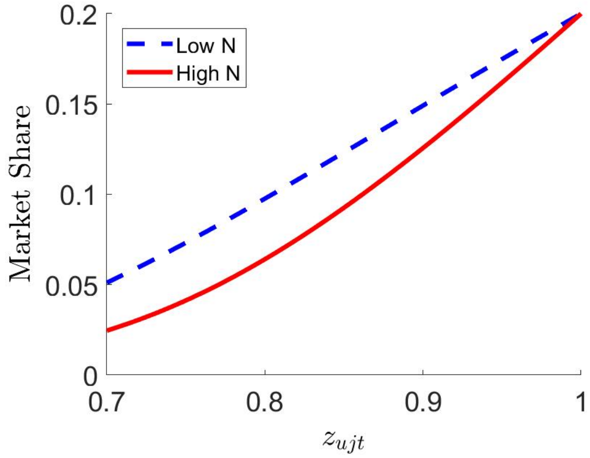
A.1 plots the effects for a range of values for z_{cjt}. Panel (a) shows that the financial tightening (decrease in γ_t, so decrease in z_{cjt}) leads to a fall in the market share of these 20% of constrained firms. Vice versa, the market share of the remaining 80% of unconstrained firms increases, as shown in panel (b). This leads to an increase in concentration, i.e. the HHI, as shown in panel (c). Notably, the increase in concentration is larger (in percent) when there are more firms (red line). This is because with more firms, it is easier to substitute away from the 20% of constrained and hence less productive firms.

In sum, it is therefore ambiguous whether a higher intensity of competition stabilizes the economy, or not. On the one hand, higher competition *increases* the effect of a financial tightening on industry concentration (panel (c)). On the other hand, a given percent change in concentration leads to a larger percent change in the aggregate markup according to Proposition 2. Panel (d) shows that in this example, the stabilizing effect of competition dominates and the financial tightening has a smaller effect on the aggregate markup when there are more firms. In addition, the decrease in TFP is smaller when competition is intense, as shown in panel (e). The reason is the same as before: the industry as a whole is better able to substitute away from the constrained firms when there are many firms. It follows that the effect of the financial tightening on aggregate consumption is smaller when there is a high intensity of competition, as shown in panel (f).

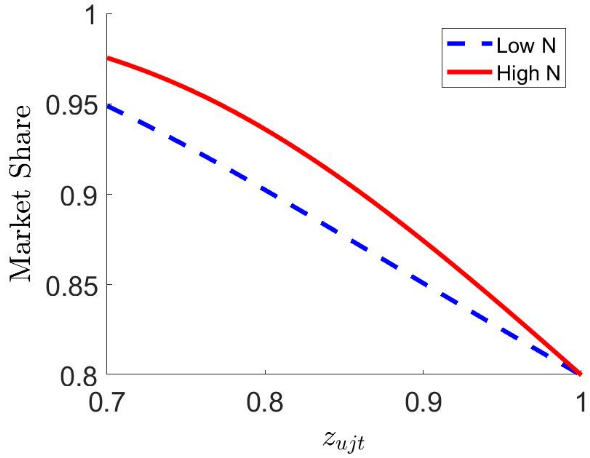
In models with endogenous financial constraints, e.g. Ottonello and Winberry (2020), all aggregate shocks affect concentration and thus the industry markup. For the particular case of monetary policy shocks, supporting empirical evidence can be found in Meier and Reinelt (2022). Ferrando et al. (2021) and Furceri et al. (2021) also empirically investigate the transmission of monetary policy shocks via firms considering both financial frictions and market power.

Idiosyncratic Shocks. Perhaps the most obvious example of asymmetric supply shocks are idiosyncratic shocks to productivity, demand, capital quality, or some other firm-level state variable. For the case of idiosyncratic productivity shocks, one would simply define z_{ijt} as firm productivity, which follows some exogenous process. For the case of capital quality shocks, one would define $z_{ijt} = \epsilon^{CQ}_{ijt} k_{ijt}$, where k_{ijt} is capital and ϵ^{CQ}_{ijt} is some a capital quality shock, drawn from some exogenous distribution. Either way, the idiosyncratic shock affects the firm-specific component of one firm, while not directly affecting all other firms in the industry. Therefore, market shares are reallocated and – expect for knife-edge cases – industry concentration changes and thus the industry markup.[31] However, idiosyncratic shocks only matter for *aggregate* outcomes, i.e. the aggregate markup, when firms are not atomistic,

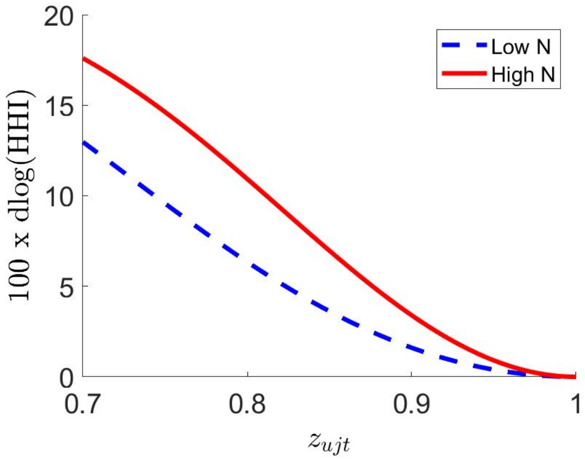
[31]See Burstein et al. (2020), Proposition 1 for a result regarding the *sign* of the change in the industry markup.



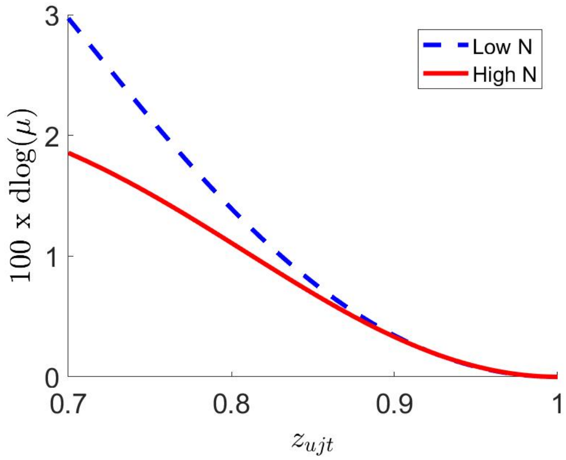
(a) Market Share: Constrained Firms



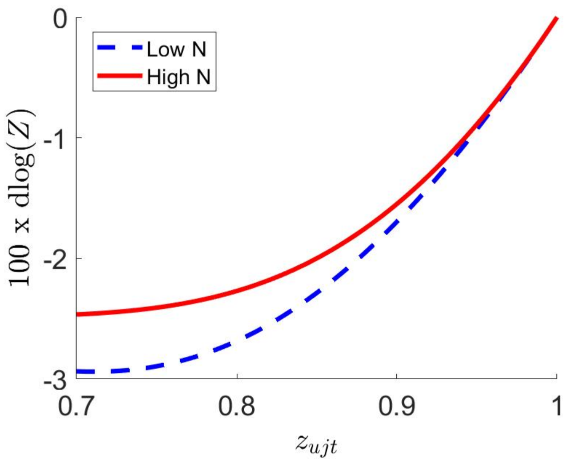
(b) Market Share: Unconstrained Firms



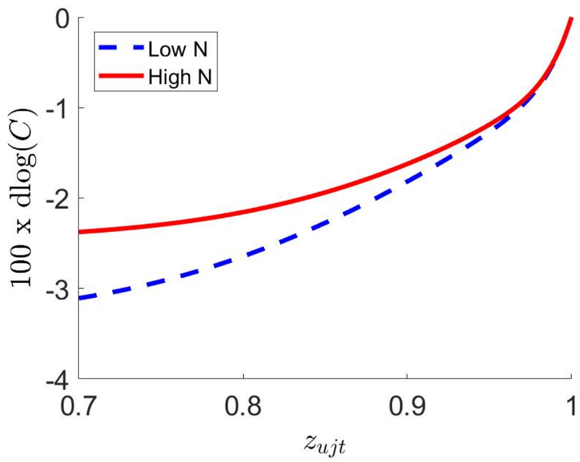
(c) Concentration



(d) Aggregate Markup



(e) TFP



(f) Consumption

Figure A.1: Financial Tightening as Asymmetric Supply Shock

e.g. as in Burstein et al. (2020) due to a large but finite number of industries. In contrast, when there is a continuum of industries, as in Atkeson and Burstein (2008), idiosyncratic shocks "wash out" and do not have aggregate effects. Yet, shocks to the *distribution* of these idiosyncratic shocks still do have aggregate effects, because they change the distribution of firm-specific components and thus reallocate sales shares in *all* industries. Examples of these asymmetric supply shocks include shocks to the dispersion (e.g. Bloom 2009, Bachmann and Bayer 2014, Ferrari and Queirós 2022) or skewness (e.g. Salgado et al. 2019) of idiosyncratic productivity (shocks).

Multi-Country Setups. In the two-country model of Atkeson and Burstein (2008), country-specific TFP shocks are asymmetric supply shocks, because they only affect firms from one country. This can be represented with the general setup by setting $z_{ijt} \in \{z_t^A, z_t^B\}$, where z_t^A and z_t^B are the country-specific productivity levels for countries A and B, respectively. A change in country-specific TFP reallocates market shares between the firms of the two countries and therefore changes concentration and the industry markup.

B Data Appendix

B.1 Data Treatment

For the empirical analysis in Section 4, I use annual firm-level data from Compustat North America. The data treatment described here broadly follows Baqaee and Farhi (2020) and De Loecker et al. (2020). From the beginning, I exclude

1. firms not incorporated in the United States (based on FIC)
2. missing industry (NAICS) or non-classifiable industry (NAICS = 99)
3. missing or non-positive sales (SALE), cost of goods sold (COGS), or total assets (AT)

Deflators. Sales and cost of goods sold are deflated using the price index for gross output from KLEMS. Capital expenditures and capital are deflated using the price index for gross fixed capital formation from KLEMS. These deflators are available from 1970 to 2015 which limits the analysis to these years.

B.2 Markup Estimation

The fundamental issue is that markups, defined as the price of a good divided by its marginal cost, are not observable. While the price of a good is typically observable, the marginal cost of producing it is not. Basu (2019) provides a summary and discussion of the various approaches developed in the literature to deal with this issue. I estimate firm-level markups using the popular "production function approach" due to De Loecker and Warzynski (2012). This approach is also used in De Loecker et al. (2020) and Baqaee and Farhi (2020).

I briefly describe the "production function approach" to estimating firm-level markups, largely following the exposition of De Loecker and Warzynski (2012). A firm i at time t produces output y_{it} using the production technology $y_{it} = F_{it}(z_{it}, k_{it}, v_{it})$, where z_{it} is productivity, k_{it} is capital and v_{it} is a variable input factor. The firm minimizes costs, subject to producing the quantity $\overline{y_{it}}$. The Lagrangian function is

$$L_{it} = r_{it}k_{it} + p_{it}^{v}v_{it} + \lambda_{it}\left(\overline{y_{it}} - y_{it}(\cdot)\right) \tag{43}$$

where the factor prices r_{it} and p_{it}^{v} are taken as given by the firm and λ_{it} is the Lagrange multiplier. Note that λ_{it} reflects the marginal cost of production, because $\frac{\partial L_{it}}{\partial Q_{it}} = \lambda_{it}$. The first-order condition for the variable input is

$$\frac{\partial L_{it}}{\partial v_{it}} = v_{it} - \lambda_{it}\frac{\partial y_{it}(\cdot)}{\partial v_{it}} = 0 \tag{44}$$

Rearranging and expanding this condition yields

$$\underbrace{\frac{p_{it}}{\lambda_{it}}}_{\mu_{it}} = \underbrace{\frac{\partial y_{it}(\cdot)}{\partial v_{it}}\frac{v_{it}}{y_{it}}}_{\theta_{it}^{v}} \underbrace{\frac{p_{it}y_{it}(\cdot)}{p_{it}^{v}v_{it}}}_{s_{it}^{v}{}^{-1}} \tag{45}$$

Thus, the markup, μ_{it} is simply the product of the output elasticity on the variable input, θ_{it}^{v}, and the inverse of the share of the variable input's expenditure in total sales, $s_{it}^{v}{}^{-1}$.

Estimating Output Elasticities. While the share of inputs in total sales can easily be calculated in most datasets, output elasticities need to be estimated. There exists an extensive literature on the estimation of production functions. I follow the implementation in Baqaee and Farhi (2020), who use the methodology of Olley and Pakes (1996) with the correction proposed by Ackerberg et al. (2015). I briefly outline the main idea, building on the aforementioned papers, which provide a much more complete description. Assuming a Cobb-Douglas functional form for $F_{it}(z_{it}, k_{it}, v_{it})$ and taking logs, the production function can be written as

$$\log(y_{it}) = \beta_0 + \theta^{k}\log(k_{it}) + \theta^{v}\log(v_{it}) + z_{it} + \epsilon_{it} \tag{46}$$

where z_{it} is a productivity shock, which is observed by the firm before choosing the variable input. ϵ_{it} is measurement error observed only after choosing inputs. The presence of the productivity shocks would bias an OLS estimate of β_l. The idea of Olley and Pakes (1996) is to control for z_{it} using investment i_{it} as a "proxy" variable, because it is observable and under mild assumptions monotonically increasing in z_{it}. This enables estimating the output elasticity on the variable input, θ^{v}, using GMM.

B.3 Implementation

I now describe how the markup estimation described in Appendix B.2 is applied to the Compustat data treated as described in Appendix B.1. As before, the steps broadly follow Baqaee and Farhi (2020) and De Loecker et al. (2020).

Output Elasticities. The output elasticity of a variable input is estimated using Olley and Pakes (1996) and the correction of Ackerberg et al. (2015). I use (log) sales as output variable, (log) cost of goods sold (COGS) as variable input, (log) capital (PPEGT) as state variable, and (log) investment (CAPX) as proxy variable. Additional controls are SIC-3-digit and SIC-4-digit sales shares to deal with the issue that sales do not measure quantities, but revenue.

To deal with outliers, the top and bottom 5% of the year-specific COGS-to-sales and XSGA-to-sales ratios are excluded for the estimation of the production function. I estimate time-varying elasticities by using 11-year rolling windows. Choosing relatively long windows ensures fairly stable parameter estimates. Moreover, I estimate industry-specific elasticities, grouping industries based on 2-digit NAICS codes.

Markups. Having estimated output elasticities of cost of goods sold, the markup is simply computed as the product of the elasticity with the inverse expenditure share on COGS, i.e. $\frac{SALE}{COGS}$.

Firm-Level Dataset. Inspecting the distribution of firm markups shows that the dataset includes many outliers. Therefore, I trim the top and bottom 7.5% of the year-specific markup distribution. Even after this, estimated markups range from 0.52 to 7.45, which are extreme values. In addition, I exclude firms for which less than 6 observations are available. This ensures that I can calculate measures of markup volatility at the firm-level, such as the interquartile range and the standard deviation.

B.4 Additional Results

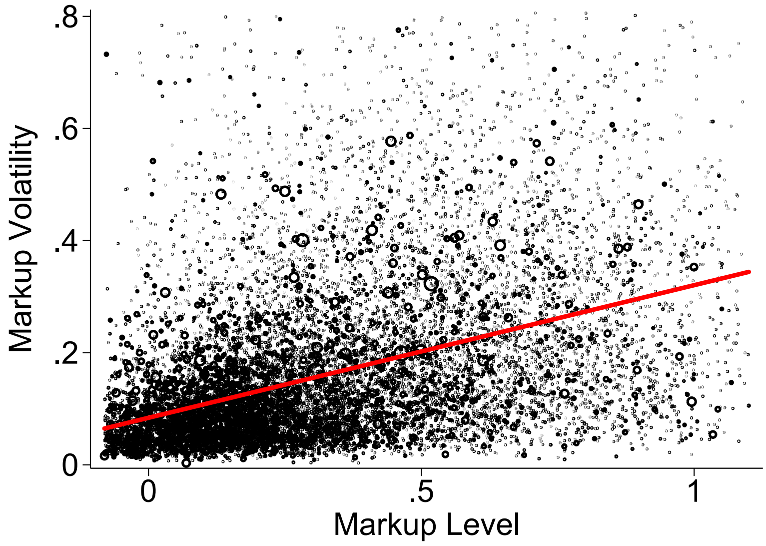


Figure A.2: Firm-Level (Log) Markups & (Log) Markup Volatility

Notes: Each circle depicts one firm. The markup level is the median log markup. The markup volatility is the interquartile range of the log markup. Both variables are trimmed (1%). Firms are weighted by their average sales share. The red line shows the linear fit.

	(1) IQR (log μ)	(2) IQR (log μ)	(3) IQR (log μ)	(4) IQR (log μ)	(5) IQR (dlog μ)
Median (log μ)	0.237***	0.218***	0.226***	0.164***	0.089***
	(0.017)	(0.005)	(0.017)	(0.016)	(0.008)
Constant	0.084***	0.107***	0.069*	0.062***	0.044***
	(0.005)	(0.002)	(0.039)	(0.004)	(0.002)
Observations	12257	12257	12257	12250	12251
R^2	0.211	0.146	0.316	0.188	0.161
Weights	Yes	No	Yes	Yes	Yes
Industry FE	No	No	Yes	No	No
Linear Trend	No	No	No	Yes	No

Table A.1: Firm-Level (Log) Markups & (Log) Markup Volatility

Notes: Each column displays coefficients from a separate regression: $Volatility_i(\mu_{it}) = \beta_0 + \beta_1 * Level_i(\mu_{it}) + \epsilon_i$. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All variables are trimmed (1%).