Firm Heterogeneity, Market Power and Macroeconomic Fragility*

Alessandro Ferrari[†] and Francisco Queirós[‡]

[†]University of Zurich

[‡]University of Naples Federico II and CSEF

First Version: March 2019

This Version: January 2023

Latest Version Here

Abstract

We study how firm heterogeneity and market power affect macroeconomic fragility, defined as the probability of long slumps. We propose a theory in which the positive interaction between firm entry, competition and factor supply can give rise to multiple steady-states. We show that when firms are highly heterogeneous in terms of productivities, even small temporary shocks can trigger firm exit and make the economy spiral in a competition-driven poverty trap. We calibrate our model to incorporate the well-documented trends on rising firm heterogeneity in the US economy, and show that they significantly increase the likelihood and length of slow recoveries. We use our framework to study the 2008–09 recession and show that the model can rationalize the persistent deviation of output and most macroeconomic aggregates from trend, including the behavior of net entry, markups and the labor share. Post-crisis cross-industry data corroborates our proposed mechanism. We conclude by showing that firm subsidies can be powerful in preventing long slumps and can lead to up to a 21% increase in welfare.

JEL Classification: E22, E24, E25, E32, L16

Key words: firm heterogeneity, competition, market power, poverty traps, great recession

*This is a substantially revised version of a previously circulated paper under the title "Low Competition Traps". We thank Árpád Ábrahám, Pol Antràs, Fernando Broner, Jesús Bueren, Giacomo Calzolari, Vasco Carvalho, Andrea Colciago, Russell Cooper, David Dorn, Jan Eeckhout, Matteo Escudé, Luca Fornaro, Manuel García-Santana, Matteo Gatti, Basile Grassi, David Hemous, Giammario Impullitti, Nir Jaimovich, Tullio Jappelli, Chad Jones, Philipp Kircher, Omar Licandro, Ramon Marimon, Isabelle Mejean, Morten Olsen, Marco Pagano, Giacomo Ponzetto, Edouard Schaal, Florian Scheuer, Armin Schmutzler, Jaume Ventura, Nic Vincent, Xavier Vives and seminar participants at the EUI, UPF, CSEF, the University of Konstanz, EIEF, University of Nottingham, University of Bonn, Bank of Portugal, Bank of England, the EEA, SMYE, T2M, Swiss Macro Workshop and the Barcelona Summer Forum. All errors are our own.

1 Introduction

The US economy appears to be recovering more slowly from its recessions. The left panel of Figure 1 shows, for different recessions, the change in detrended output in the first two years after the trough. Over the postwar period the pace of recoveries has significantly slowed down. This has been especially clear for the last three recessions (Galí et al., 2012). As shown in the right panel of Figure 1, four years after the beginning of each recession, detrended output was still below its pre-crisis value. For the 2008-2009 crisis, the gap is substantial (about 10 log points below trend) and has been referred to as the *long slump* or *great deviation* (Hall, 2011). During the same period, the US experienced a significant increase in firm heterogeneity along several dimensions, such as productivity, size and markups. This is illustrated in Figure 2, which shows the evolution of the standard deviation of sales for US public firms.[1] In this paper we argue that the slowdown in the pace of US recoveries can be partially explained by the rise of firm heterogeneity.

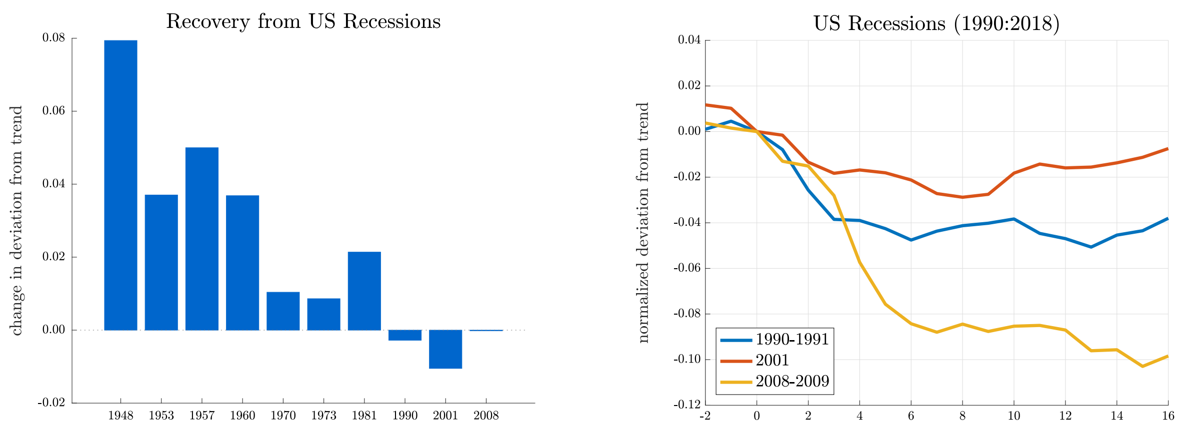


Figure 1: Recovery from US Recessions

Note: The left panel shows, for each US postwar recession, the change in detrended output in the first two years after the trough. We use quarterly real GDP per capita (in logs, from BEA) and compute a linear trend for the period 1947-2019. We consider all recessions that lasted longer than 6 months according to the NBER. The right panel shows the evolution of detrended output for the 1990-1991, 2001 and 2008-2009 recessions.

The main contribution of this paper is to show that rising firm heterogeneity can have a significant impact on business cycle fluctuations, and result in a higher probability of long slumps. We refer to such a probability as *macroeconomic fragility*. We show this in the context of an RBC model with oligopolistic competition, endogenous firm entry and elastic capital and labor supply.

[1]The figure focuses on listed firms, but increasing firm-level dispersion has been documented using census data, and focusing on narrowly defined industries. Several studies have documented rising firm differences in terms of i) revenue TFP (Andrews et al., 2015; Kehrig, 2015; Decker et al., 2018), ii) size (Bonfiglioli et al., 2018; Autor et al., 2020), and iii) markups (De Loecker et al., 2020; Calligaris et al., 2018; Díez et al., 2018). See Van Reenen (2018) for a summary of the recent findings.

We use a quantitative version of our framework to study the effects of the trends in rising firm heterogeneity. We find that these forces are quantitatively important, they can rationalize episodes such as the 2008-2009 recession and its aftermath and are consistent with cross-industry empirical evidence from the *great deviation*.

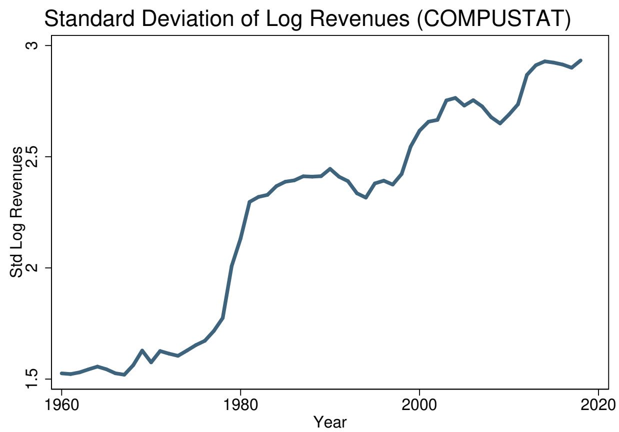


Figure 2: Dispersion in Sales for US Public Firms

Note: The figure shows the standard deviation of log sales for US public firms. Data is from COMPUSTAT.

At the heart of our model is a complementarity between competition and factor supply. First, economies with intense competition in product markets feature low profit shares and high factor shares and factor prices, which induce high factor supply. Second, high factor supply allows more firms to enter the market, which results in greater competition. This complementarity can give rise to multiple competition regimes or (stochastic) steady-states. A key contribution of our theory is to show that rising heterogeneity in idiosyncratic TFP can make transitions from high to low steady-states more likely to occur. When firm heterogeneity increases, large, productive firms expand, while small, unproductive firms contract. As a consequence, smaller negative shocks can be enough to trigger firm exit and a transition to a steady-state featuring lower competition, capital and output. An identical result is obtained when fixed costs of production increase. We characterize these results formally by showing that the minimum size (negative) shock required to trigger a transition from a high to a low steady-state decreases when firm TFP heterogeneity rises or when fixed costs increase.

To quantify these economic forces, we provide three calibrations of our model, where we target US firm-level moments in 1975, 1990 and 2007. These different calibrations differ in the level of fixed costs and in the degree of TFP dispersion (which both increase over time). We first find that, while the 1975 economy is characterized by a unimodal ergodic distribution of output, the 1990 and 2007 feature bimodal distributions (an indication of two stochastic steady-states). Second, when we subject the three economies to the same shocks, the 2007 economy exhibits significantly greater amplification and propagation. For example, we find that the 2007 economy experiences a recession greater than 10% every 75 years, while the 1990 and 1975 economies experiences one every 120 and 735 years respectively. This suggests that rising firm heterogeneity and fixed costs made the US economy significantly more fragile and more prone to long-lasting slumps.

We also use our model economy as a laboratory to study the Great Recession and its aftermath. The 2008 crisis was marked by a large and persistent deviation of output and other aggregates from trend, something unusual in the entire postwar period. For example, in 2019, output per capita was 14% below its pre–2007 trend, a deviation far larger and persistent than in previous recessions (Figure A.1). We ask whether our model can replicate such a quasi-permanent drop in aggregate output and other aggregate variables. To this end, we feed our 2007 economy a sequence of shocks calibrated to match the behavior of aggregate TFP in 2008–9 and study the economy's response. The model generates the observed persistent deviation from trend of GDP as well as of investment, hours and aggregate TFP. It also explains the sharp and persistent drop of the labor share after 2008. Importantly, when we subject the 1990 and 1975 economies to the same shocks, the model does not predict such a persistent deviation from trend.

We also provide empirical evidence in favor of our proposed mechanism. Our theory provides predictions in terms of cross-industry responses to the business cycle. In particular, for any two industries with the same number of firms, the one with a higher level of heterogeneity reacts more to a negative shock. We test this prediction in the data using concentration as a proxy for heterogeneity in US 6-digit industries. Consistent with our model predictions, we show that industries that were more concentrated in 2007 experienced greater cumulative declines in the number of firms, the labor share and economic activity over the 2008-2016 period.

In terms of policy lessons from our theory, we show that firm subsidies can be effective in preventing deep recessions, leading to a welfare gain of up to 21% in consumption-equivalent terms. The fundamental intuition behind this result is that in our economy firms' entry and exit decisions have externalities as they change market power. A planner may then find it beneficial to trade off the efficiency loss attached to less productive firms staying in the market with the reduction in the rents of more productive firms. This policy lesson is specific to our model and to the pivotal role of the extensive margin in shaping the degree of aggregate market power. Its logic, however, is more general. Cyclical changes in market power can amplify aggregate fluctuations. Policies that aim to keep the degree of product market competition high can prevent the economy from entering into low output regimes and thereby improve welfare.

Related Literature Our paper speaks to three different strands of the literature. First, it is related to the macroeconomic literature studying models of coordination failures (Cooper and John, 1988; Matsuyama, 1991; Benhabib and Farmer, 1994; Farmer and Guo, 1994; Herrendorf et al., 2000; Buera et al., 2021). While we are not the first to show how multiple equilibria and/or steady-states can arise in a context of imperfect competition and variable markups (Pagano, 1990; Chatterjee et al., 1993; Rotemberg and Woodford, 1995; Galí and Zilibotti, 1995; Jaimovich, 2007),[2] we contribute to this literature by studying the role of firm-level heterogeneity in shaping macroeconomic *fragility*. As we discuss, this concept is distinct from the *existence* of steady-state multiplicity. We also provide a quantification of our mechanism and link it to the 2008 crisis.[3]

Second, this paper relates to a large and growing literature documenting long-term trends in firm heterogeneity and market power. There are several signs that indicate rising market power in the US and other advanced economies. For example, Autor et al. (2020) use data from the US census to document rising sales and employment concentration, while Akcigit and Ates (2019) document a rise in patenting concentration. Other studies have documented a secular rise in price-cost markups. Using data from national accounts, Hall (2018) finds that the average sectoral markup increased from 1.12 in 1988 to 1.38 in 2015. De Loecker et al. (2020) document a steady increase in sales-weighted average markups for US public firms between 1980 and 2016.[4] This was driven by both an increasing share of large firms and by rising dispersion in the markup distribution. Close to our approach, De Loecker et al. (2021) argue that declining dynamism and rising market power can be explained, among other channels, by increasing productivity dispersion and fixed costs. Edmond et al. (2021) estimate that the welfare cost of markups can be large and represent a loss of up to 25% in consumption-equivalent terms.We contribute to this literature by investigating the business cycle implications of these trends, and in particular their impact on the 2008 crisis and the subsequent *great deviation*. We emphasize that increased firm heterogeneity and market power may have negative welfare consequences through an increase in macroeconomic fragility.

Lastly, this paper relates to the literature studying slow recoveries and the persistent impact of the 2008 crisis. A large part of this literature has focused on the secular decline in interest rates and/or the possibility of liquidity traps, which constrains monetary policy (Galí et al., 2012; Benigno and Fornaro, 2017; Guerrieri and Lorenzoni, 2017; Eggertsson et al., 2019; Mian et al., 2021), or on the long-run consequences of R&D decisions (Queralto, 2020; Benedetti-Fasil et al., 2021). Close to this paper, Schaal and Taschereau-Dumouchel (2018) build an RBC model with complementarities in capacity utilization choices among monopolistic firms to generate multiple steady-states. We too use coordination to obtain multiplicity and interpret the post–2008 deviation as a transition to a low steady-state. Differently from their work, we study the role of heterogeneity

[2]Without relying on multiple equilibria or steady-states, Cooper and John (2000), Etro and Colciago (2010), Bilbiie et al. (2012) and Gamber (2021) show that a combination of imperfect competition with endogenous entry amplifies aggregate fluctuations.

[3]Our paper speaks to the literature on the cyclicality of markups, which includes Rotemberg and Saloner (1986), Rotemberg and Woodford (1991), Jaimovich and Floetotto (2008), Bils et al. (2018), Nekarda and Ramey (2020) and Burstein et al. (2020).

[4]Edmond et al. (2021) show that a cost-weighted average markup displays a less pronounced trend. See also Traina (2018), Karabarbounis and Neiman (2019) and Bond et al. (2021) on trends in markups.

in productivities in oligopolistic markets with variables markups. We complement their analysis, and the aforementioned literature, by arguing that rising firm-level heterogeneity has increased the likelihood of slumps. Our theory can also account for number of trends observed after 2008, such as the acceleration in the labor share decline, the acceleration of markup growth, and the decline in the number of firms.

The rest of the paper is organized as follows. Section 2 sets the general framework and provides the first result relating to technology and fragility. Section 3 presents a growth model with oligopolistic competition and derives the main theoretical results. Section 4 discusses the calibration and presents the quantitative results. Section 5 provides and extended application to the US great recession and its aftermath, and presents the cross-industry empirical evidence. In Section 6, we study the welfare effects of fiscal policy in our model. Finally, section 7 concludes.

2 General Framework

In this section we introduce the concept of macroeconomic fragility. We do so in the context of a general economy, featuring a representative household, endogenous factor supply, and a large number of product markets (characterized by an endogenous number of firms, variable markups and love-for-variety). We proceed by first describing the economy. We then define macroeconomic fragility and study how it changes with firm-level heterogeneity. In Section 3 we impose additional structure to this general framework, by developing an RBC model with oligopolistic competition. This allows us to derive sharper predictions and analytical results.

2.1 Primitives

Demographics The economy contains an infinitely-lived representative household, who has time separable utility $U = \sum_{t=0}^{\infty} \beta^t u\left(C_t, L_t\right)$. The household can save by investing in capital, which depreciates at rate δ. The household earns rental and wage rates R_t and w_t, determined in competitive factor markets.

Technology There is a final good (used for consumption and investment), which is given by a homothetic aggregator of different products or varieties. The economy contains $I \in \mathbb{N}^+$ different categories of products. Within each category, there is a measure one of markets. For example, one possible product category is restaurants, within which there are many different geographic locations in which restaurants compete. We think of goods sold in different markets as different products which enter the final good aggregator.[5] All product markets are characterized by the same maximum number of players $M \in \mathbb{N}^+$, who produce differentiated varieties. Firms produce

[5]We adopt this formulation only for tractability reasons, to make the aggregate number of firms a continuously differentiable variable, in spite of the existence of a finite number of different types of markets I.

according to $\tilde{F}(\gamma, L, K) = \gamma \cdot F(L, K)$ where γ is idiosyncratic TFP and $F(L, K)$ is a constant returns to scale production function satisfying the Inada conditions. Operating in a product market of type i requires a per period fixed cost c_i (in units of the final good). Firms purchase factors of production in competitive factor markets, but imperfectly compete in product markets. Different product categories can be subject to different market structures (e.g. Cournot or Bertrand competition). The technology can be summarized by a matrix of Hicks-neutral idiosyncratic productivities $\Gamma := [\gamma_{ij}]_{(I \times M)}$ and a vector of fixed costs $\mathcal{C} := [c_i]_{(I \times 1)}$. Given these definitions, the technology set is summarized by $\Lambda := [\,\Gamma\,,\,\mathcal{C}\,]$. The aggregate mass of firms can be written as $n = \sum_{i=1}^{I} \sum_{m=1}^{M} \eta_{im}\, m$, where $\eta_{im} \in [0, 1]$ is the fraction of product markets i with $m \leq M$ firms. We assume that firms enter sequentially and that firms making higher profits enter first. Given these assumptions, the aggregate mass of firms n and the technology set Λ fully characterize the set of active firms. For exposition purposes, we work with a deterministic setting subject to one time unexpected shocks. In particular, we think of the possibility of shocks that destroy a fraction $1 - \chi \in [0, 1]$ of the capital stock. Aggregate uncertainty, by means of stochastic TFP, is introduced in the model we study in Section 3.

Definition 1 (Equilibrium)**.** *An equilibrium in this economy is a set of policies such that i) all agents optimize; ii) all active firms make no loss; iii) inactive firms would make a loss upon entry and iv) all markets clear.*

This economy admits an aggregate production function, which we characterize in Lemma 1.

Lemma 1 (Aggregate Production Function)**.** *The aggregate production function is given by*

$$F(\Lambda, K, n(\Lambda, K)) = \Phi(\Lambda, n(\Lambda, K))\, F(K, L),$$

where Λ is the technology set (as defined above), n is the mass of active firms, and $\Phi(\cdot)$ is aggregate TFP. Finally $F(K, L)$ is the CRS production function introduced above.[6]

Proof. See Appendix A.2. ∎

Aggregate TFP $\Phi(\Lambda, n(\Lambda, K))$ reflects two terms: a weighted average of firm level productivity γ_{ij} and love for variety. From the aggregate production function it is possible to characterize the economy's inverse demand for capital. Let Ω denote the aggregate factor share (i.e. the ratio of total labor and capital payments over gross output Y). As shown in Appendix A.2, the equilibrium rental rate can be written as

$$R(\Lambda, K) \;=\; \Omega(\Lambda, n(\Lambda, K))\, \Phi(\Lambda, n(\Lambda, K))\, F_K\big(K, L\big). \tag{1}$$

[6]Note that L is a function of the set of state variables (Λ, K). For ease of notation we suppress these.

It is convenient to define $\Theta(\Lambda, n) \coloneqq \Omega(\Lambda, n)\,\Phi(\Lambda, n)$, which represents a measure of aggregate factor prices (Appendix A.2). [7] Given the demographic structure of this economy, the long-run supply of capital is infinitely elastic and given by[8]

$$R^* = \beta^{-1} - (1-\delta)\,. \tag{2}$$

Therefore, the economy features multiple steady-states if the following equation admits more than one solution

$$R(\Lambda, K) = R^*. \tag{3}$$

The red curve in Figure 3 represents the capital demand schedule in eq. (1) for a particular economy. Assuming that the Inada conditions are satisfied, we have $R(\Lambda, 0) = \infty$ and $R(\Lambda, \infty) = 0$. Therefore, multiple steady-states occur only if (1) features at least one increasing part. As explained in Section B.1 of the Supplementary Material, there are different mechanisms that can make $R(\cdot)$ locally increasing in K. For example, suppose that firm entry increases in capital, so that competition and the aggregate factor share $\Omega(\Lambda, n)$ are also increasing in K. Then, if $\Omega(\Lambda, n)$ increases sufficiently fast in K in some region, it can counteract decreasing returns $F_K(\cdot)$ and make $R(\cdot)$ locally increasing. In Figure 3, the economy in red contains two stable steady-states, K_1^S and K_2^S, and an unstable one, K_1^U. The initial capital stock determines the unique path that the economy follows and the steady-state to which it converges.[9] We are particularly interested in the possibility of downward transitions across steady-states. Note that, if the economy starts close to the high steady-state K_2^S, a transition to K_1^S occurs if the economy is hit by a shock that destroys a fraction $1 - K_1^U/K_2^S$ of its capital stock. Thus, the likelihood of a downward transition depends on the distance between K_1^U and K_2^S. For example, under the alternative economy represented in blue, a smaller fraction of the capital stock needs to be destroyed for a downward transition to take place, $1 - \tilde{K}_1^U/\tilde{K}_2^S$. This second economy is characterized by a lower capital demand schedule for sufficiently high values of K. This suggests that a reduction in the aggregate demand for capital can be associated with a higher likelihood of downward transitions across steady-states. We next formalize the concept of *fragility* and explore different mechanisms that, by decreasing the aggregate demand for capital, can result in more likely downward transitions.

---

[7] This is equal to the Lagrange multiplier of the cost minimization problem for a firm with unit productivity.

[8] This can be obtained by evaluating the stationary Euler equation of the representative household.

[9] When the economy starts on the left of K_1^U it reaches K_1^S, otherwise it achieves K_2^S. Multiple paths would be possible if the curve in Figure 3 was a correspondence, and not a function. In this section, we exclude multiple paths by assumption.

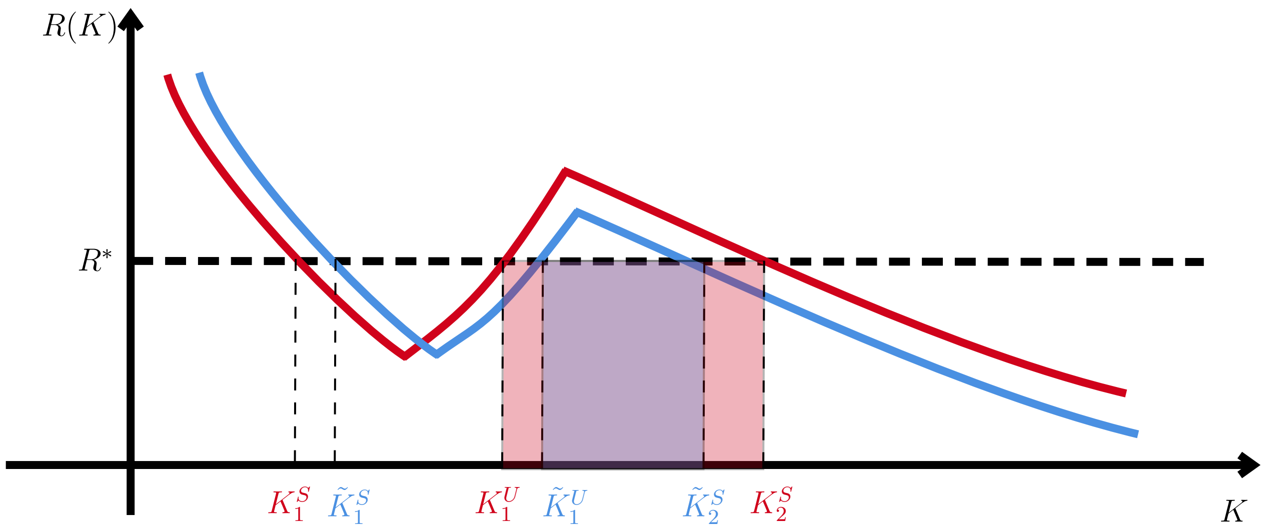


Figure 3: Steady-state multiplicity

2.2 Technological Change and Fragility

We start by defining the concept of fragility, which is central to our analysis. Let $\mathcal{K}$ denote the ordered set of all steady-states capital levels. Let $\mathcal{K}^{\mathcal{S}}$ be the collection of stable steady-states and $\mathcal{K}^{\mathcal{U}}$ be the collection of unstable steady-states in $\mathcal{K}$. Recall that, in our setting, the economy can be hit by a one time unexpected shock that destroys a fraction $1-\chi$ of the capital stock.[10] Our measure of fragility captures the proximity of a *stable* steady-state $\mathcal{K}_n^S$ to the preceding *unstable* steady-state $\mathcal{K}_{n-1}^U$. The closer these two steady-states are to each other, the lower is the minimum shock needed to trigger a downward transition.

Definition 2 (Fragility)**.** *Let $\chi_n \in [0,1]$ be defined as $\chi_n := \mathcal{K}_{n-1}^U/\mathcal{K}_n^S$. It follows that $1-\chi_n$ is the minimum size shock needed to make the economy move from $\mathcal{K}_n^S$ to $\mathcal{K}_{n-1}^S$. We call an increase in χ_n an increase in fragility.*

Two further observations should be made. First, our main focus is on *fragility*, which is different from the *existence* of multiplicity. Although the first requires the second, these are different concepts. Second, the notion of *fragility* is related to, but distinct from, the idea of the *stability* of a steady-state. We think of *fragility* as the possibility of downward transitions only. This is the size of the left partition of the basin of attraction of the steady-state, which is given by $\mathcal{K}_n^S - \mathcal{K}_{n-1}^U$. The notion of stability, instead, relates to the size of the whole basin of attraction and also accounts for the possibility of upward transitions. In Proposition 1 we state a sufficient condition for a technology shift to increase fragility, i.e. $\partial\chi_n/\partial\lambda > 0$. This condition says that fragility increases when the rental rate decreases for intermediate values of K (i.e. in the region between steady-states).

[10]The intuition behind all our results carries through under more general shock structures, with some adjustments. In general the capital level in a steady state depends on the shock realization. This would significantly increase the difficulty in defining a norm between $\mathcal{K}_n$ and $\mathcal{K}_{n-1}$, since their distance would become a random variable itself.

Proposition 1 (Comparative Statics)**.** *Let $\mathcal{K}_n^S$ be a stable steady-state and consider a technological shift $\partial\lambda$. Then $\partial\chi_n/\partial\lambda > 0$, i.e. the stable steady-state $\mathcal{K}_n^S$ becomes more fragile, if for $K = \{\mathcal{K}_{n-1}^U, \mathcal{K}_n^S\}$*

$$\frac{\partial}{\partial\lambda}\Theta\left(\Lambda, n\left(\Lambda, K\right)\right) = \Theta_\lambda\left(\Lambda, n\left(\Lambda, K\right)\right) + \Theta_n\left(\Lambda, n\left(\Lambda, K\right)\right) n_\lambda\left(\Lambda, K\right) < 0.$$

Using the definition of $\Theta(\cdot)$ we note that this is equivalent to

$$\underbrace{\Omega(\Lambda, n)\Phi_\lambda(\Lambda, n)}_{\Delta\ TFP} + \underbrace{\Omega_\lambda(\Lambda, n)\Phi(\Lambda, n)}_{\Delta\ factor\ share} + \underbrace{\left[\Omega_n(\Lambda, n)\Phi(\Lambda, n) + \Omega(\Lambda, n)\Phi_n(\Lambda, n)\right] n_\lambda\left(\Lambda, K\right)}_{\Delta\ number\ of\ firms} < 0.$$

The first two terms characterize the direct effect of $\partial\lambda$ on aggregate TFP and the factor share. The last term describes the effect of changes in the number of active firms.

Proof. See Appendix A.2 ■

Proposition 1 provides a sufficient condition for a (technology-driven) increase in fragility. There are three main channels highlighted in this result: i) changes in aggregate TFP (Φ), ii) changes in the factor share (Ω) and iii) changes in the number of firms (n).[11] To connect this result to the empirical patterns discussed in the introduction, we can think of $\partial\lambda$ as an increase in firm level heterogeneity. Through the lens of Proposition 1 such an increase can trigger higher fragility by reducing the factor share (due to greater market power), via the reallocation of activity towards less productive firms or by reducing the number of active firms. If firms' market shares are increasing in their productivity, as heterogeneity increases three things happen: *i)* a mechanical efficiency gain due to large firms becoming more productive and through positive reallocation; *ii)* a compression of factor shares as large firms can exert more market power; *iii)* entry is harder for firms outside the market, which can generate a loss of varieties and higher market power. Intuitively, if the anti-competitive forces *ii)* and *iii)* dominate the efficiency gains from higher heterogeneity *i)*, the return to capital decreases and it becomes harder to sustain a high steady-state.

The intuition behind the result described in Proposition 1 is similar in spirit to the findings of Baqaee and Farhi (2020) and Edmond et al. (2021) relating dispersion and aggregate TFP. As will become clearer in the model presented in Section 3, we put forth the additional effect on the stability of the economy with respect to aggregate fluctuations. Our economy features an endogenous amplification mechanism that can trigger non-linearities in the response to shocks. As firms become more heterogeneous these non-linearities become more salient and the economy can be more likely to experience quasi-permanent slumps. To dig deeper into the forces underlying the effect of heterogeneity on fragility we specify a RBC model in which firms compete oligopolistically. This additional structure allows us to provide sharper theoretical and quantitative results.

[11] We study these channels in detail in the Supplementary Materials through Remarks B.1, B.2 and B.3.

3 A Model with Firm Heterogeneity and Variable Markups

The model presented in this section builds upon the neoclassical growth model, with a representative household that supplies labor and capital. The technology side is comprised of a large number of product markets, where firms compete oligopolistically as in Atkeson and Burstein (2008). We start by describing the demand side and the technology structure. Then we analyze the equilibrium of a particular product market (taking aggregate variables as given). Finally, we characterize the general equilibrium.

3.1 Preferences

Time is discrete and indexed by $t = 0, 1, 2, \ldots$. There is a representative, infinitely-lived household with lifetime utility

$$U_t = \mathbb{E} \sum_{t=0}^{\infty} \beta^t \, U\left(C_t, L_t\right), \tag{4}$$

where $0 < \beta < 1$ is the discount factor, $C_t \geq 0$ is consumption of the final good and $L_t \geq 0$ is labor. We adopt the period utility function as in Greenwood et al. (1988)

$$U\left(C_t, L_t\right) = \frac{1}{1-\psi} \left(C_t - \frac{L_t^{1+\nu}}{1+\nu}\right)^{1-\psi}, \tag{5}$$

where $0 \leq \psi \leq 1$ is the inverse of the intertemporal elasticity of substitution and $\nu > 0$ is the inverse of the Frisch elasticity of labor supply.

The representative household contains many individual members, which are denoted by j. Each individual member can run a firm in the corporate sector. We assume that if two or more individuals run a firm in the same product market, they behave in a non-cooperative way – i.e. they compete against each other and do not collude. Nevertheless, all individuals pool together the profits they make. Hence there is a single dynamic budget constraint[12]

$$K_{t+1} = \left[R_t + (1-\delta)\right] K_t + W_t L_t + \Pi_t^N - C_t, \tag{6}$$

where K_t is capital, R_t is the rental rate, W_t is the wage rate and $\Pi_t^N = \sum_j \Pi_{jt}^N$ are the profits from all firms j net of fixed costs. Capital depreciates at rate $0 \leq \delta \leq 1$ and factor prices R_t and W_t are taken as given. The representative household maximizes (5) subject to (6). Our choice of GHH preferences implies that the aggregate labor supply is given by $L_t^S = W_t^{1/\nu}$.

[12] We assume that the economy features perfect financial markets. There is a stock market where individuals can trade firms (whose price equals the NPV of profits). Since stocks and capital must offer the same expected return $E_t\{R_{t+1}\}$, firm transactions among individuals do not affect the aggregate budget constraint.

3.2 Technology

There is a final good (the *numeraire*), which is a CES aggregate of I different product markets $Y_t = \left(\sum_{i=1}^{I} y_{it}^{\rho}\right)^{1/\rho}$, where y_{it} is the quantity of product market i, and $\sigma_I = 1/(1-\rho) > 1$ is the elasticity of substitution across product markets. I is assumed to be large, so that each individual product market has a negligible size in the economy. The output of each product market $i \in \{1, \dots, I\}$ is itself a CES composite of differentiated goods or varieties $y_{it} = \left(\sum_{j=1}^{n_{it}} y_{jit}^{\eta}\right)^{1/\eta}$, where n_{it} is the number of active firms in product market i at time t (to be determined endogenously) and $\sigma_G = 1/(1-\eta) > 1$ is the within product market elasticity of substitution. Following Atkeson and Burstein (2008), we assume that goods are more easily substitutable within product market than across product markets: $0 < \rho < \eta \leq 1$. Given these assumptions, the inverse demand for each variety j in product market i is given by

$$p_{ijt} = \left(\frac{Y_t}{y_{it}}\right)^{1-\rho} \left(\frac{y_{it}}{y_{ijt}}\right)^{1-\eta}. \tag{7}$$

We assume that in every product market i there is a maximum number of entrepreneurs $M \in \mathbb{N}$, so that $n_{it} \leq M$. Entrepreneur j can produce her variety by combining capital k_{ijt} and labor l_{ijt} through a CRS technology

$$y_{it} = \underbrace{A_t \, \gamma_{ij}}_{\tau_{ijt}} \left(k_{ijt}\right)^{\alpha} \left(l_{ijt}\right)^{1-\alpha}. \tag{8}$$

Note that the productivity of each entrepreneur τ_{ijt} is the product of two terms (i) a time-varying aggregate component A_t (common to all product markets and types) and (ii) a time-invariant idiosyncratic term γ_{ij}. We refer to A_t as aggregate productivity and to γ_{ij} as j's idiosyncratic productivity. Aggregate productivity follows an auto-regressive process $\log A_t = \phi_A \log A_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim N\left(0, \sigma_{\varepsilon}^2\right)$. Without loss of generality, we order idiosyncratic productivities according to $\gamma_{i1} \geq \gamma_{i2} \geq \cdots$. Labor is hired at the competitive wage W_t and capital at the rental rate R_t. Entrepreneur j can thus produce her variety at a marginal cost Θ_t/τ_{ijt}, where

$$\Theta_t := \left(\frac{R_t}{\alpha}\right)^{\alpha} \left(\frac{W_t}{1-\alpha}\right)^{1-\alpha} \tag{9}$$

is the marginal cost function for a Cobb-Douglas technology with unit productivity. We refer to Θ_t as the *factor price index*. In addition to all variable costs, the production of each variety entails a fixed production cost $c_i \geq 0$ per period (which may be different across product markets). Such a cost is in units of the *numeraire*.[13]

[13] This implies that fixed costs do not change with factor prices. We modify this assumption in Section 5.2.

3.3 Market Structure

We assume that all firms that enter (and thus incur the fixed cost c_i) play a static Cournot game: they simultaneously announce quantities, taking the output of the other competitors as given.[14] Therefore, each entrepreneur j solves

$$\max_{y_{ijt}} \quad \left(p_{ijt} - \frac{\Theta_t}{\tau_{ijt}}\right) y_{ijt} \qquad \text{s.t. } p_{ijt} = \left(\frac{Y_t}{y_{it}}\right)^{1-\rho} \left(\frac{y_{it}}{y_{ijt}}\right)^{1-\eta} \text{ and } \quad y_{it} = \left(\sum_{k=1}^{n_{it}} y_{kit}^{\eta}\right)^{\frac{1}{\eta}}. \tag{10}$$

The solution to (10) yields a system of n_{it} non-linear equations in $\{p_{ijt}\}_{j=1}^{n_{it}}$ (one for each firm)[15]

$$p_{ijt} = \underbrace{\frac{1}{\eta - (\eta - \rho)\, s_{ijt}}}_{\mu_{ijt}} \frac{\Theta_t}{\tau_{ijt}}, \tag{11}$$

where s_{ijt} is the market share of firm $j = 1, \ldots, n_{it}$ and μ_{ijt} is the markup. Eq. (11) establishes a positive relationship between market shares and markups. This follows from firms internalizing the impact of their size on the price they charge (p_{ijt}); large firms end up restricting output disproportionately more (relative to productivity), thereby charging a high markup. Rearranging (11), one can also see that market shares are a positive function of revenue TFP ($p_{ijt}\, \tau_{ijt}$). Our model thus features a positive association between revenue productivity, size and markups. A shock that increases dispersion in revenue TFP is also associated with greater dispersion in market shares and markups.

To conclude the description of the product market equilibrium, we need to determine the number of active firms n_{it}. To this end, let $\Pi\,(j, n_{it}, \Gamma_{it}, X_t) \coloneqq (p_{ijt} - \Theta_t/\tau_{ijt})\, y_{ijt}$ denote the gross profits of firm $j \leq n_{it}$ in product market i, when there are n_{it} active firms, given a productivity distribution $\Gamma_i \coloneqq \{\gamma_{i1}\,, \gamma_{i2}\,, \ldots\}$ and a vector of aggregate variables $X_t \coloneqq [A_t\,, Y_t\,, \Theta_t]$. The equilibrium number of firms must be such that (i) the profits of each active firm are not lower than the fixed cost c_i and (ii) if an additional firm were to enter, its profits would be lower than the fixed cost. Formally, an interior solution $n_{it}^* < M$ to the equilibrium number of firms must satisfy

$$[\Pi\,(n_{it}^*, n_{it}^*, \Gamma_i, X_t) - c_i]\; [\Pi\,(n_{it}^* + 1, n_{it}^* + 1, \Gamma_i, X_t) - c_i] \;\leq\; 0, \quad \forall i = 1, \ldots, I. \tag{12}$$

Lemma B.1 in Appendix B.2 provides an analytical characterization of the profit function under the special case of $\eta = 1$. We show that the profits of any firm j i) increase in its own idiosyncratic productivity γ_{ij} and ii) decrease in the idiosyncratic productivity of all the other firms γ_{ik}. This

[14]We follow Atkeson and Burstein (2008) and assume that firms enter sequentially in decreasing order of TFP.

[15]This system of first order conditions admits a close-form solution only in the limit case in which there is no differentiation within an product market ($\eta = 1$), as shown in the Section B.2.

means that, as top firms become more productive, small firms make lower profits and become closer to their exist threshold (*ceteris paribus*). This is key to understanding some of the aggregate results that we describe next on the behavior of the economy under higher firm heterogeneity.

3.4 General Equilibrium

Equilibrium Definition We start by defining an equilibrium for this economy. Denoting the history of aggregate productivity shocks by $A^t = \{A_t, A_{t-1}, ...\}$ we have the following definition.

Definition 3 (Equilibrium)**.** *An equilibrium is a sequence of policies* $\{C_t(A^t), K_{t+1}(A^t), L_t(A^t)\}_{t=0}^{\infty}$ *for the household, firm policies* $\{y_{ijt}(A^t), k_{ijt}(A^t), l_{ijt}(A^t)\}_{t=0}^{\infty}$*, and a set of active firms* $\{n_{it}(A^t)\}_{t=0}^{\infty}$ *with* $\forall i \in \{1, \ldots, I\}$ *such that i) households optimize; ii) all active firms optimize; iii) the slackness free entry condition in eq. (12) holds; iv) capital and labor markets clear.*

3.4.1 Static Equilibrium

We now describe the general equilibrium of this economy. We start by focusing on a static equilibrium, in which production and labor supply decisions are described, taking the aggregate level of capital K_t as given. Later on, we describe the equilibrium dynamics.

Aggregate Production Function Given a $(I \times M)$ matrix of idiosyncratic productivity draws $\mathbf{\Gamma}$ and a vector of active firms $\mathbf{N}_t \coloneqq \{n_{it}\}_{i=1}^{I}$, aggregate output can be written as

$$Y_t = A_t \, \Phi\left(\mathbf{\Gamma}, \mathbf{N}_t\right) \; L_t^{1-\alpha} K_t^{\alpha}. \tag{13}$$

The term $\Phi(\cdot)$ represents the endogenous component of aggregate TFP and is a function of the number of active firms, individual productivities and market shares. An analytic expression for $\Phi(\mathbf{\Gamma}, \mathbf{N}_t)$ is provided in Section B.3.

Aggregate Factor Share Let $\mathbb{C}_t \coloneqq W_t \, L_t + R_t \, K_t$ represent aggregate variable costs. We can write the aggregate factor share $\Omega(\cdot) \coloneqq \mathbb{C}_t / Y_t$ as a function of individual markups and market shares[16]

$$\Omega\left(\mathbf{\Gamma}, \mathbf{N}_t\right) = \sum_{i=1}^{I} \sum_{j=1}^{n_{it}} s_{it} \; s_{ijt} \; \mu_{ijt}^{-1}. \tag{14}$$

[16] The aggregate factor share is equal to the inverse of the aggregate markup $\mu(\cdot) \coloneqq Y_t / \mathbb{C}_t$.

Combining (14) and (11) we can write the aggregate factor share as a decreasing function of all product market-level Herfindahl-Hirschman index (HHI) of concentration

$$\Omega\left(\mathbf{\Gamma},\mathbf{N}_t\right)=\sum_{i=1}^{I}\ s_{it}\ \left[\eta-\left(\eta-\rho\right)\ HHI_{it}\right]. \tag{15}$$

where $HHI_{it} := \sum_{j=1}^{n_{it}}\ s_{ijt}^2$. This result, which is identical to the findings of Grassi (2017) and Burstein et al. (2020), highlights two important relationships. First, product markets with higher concentration have larger markups. Second, when highly concentrated product markets have large shares in the economy (large s_{it}) the economy's average markup is also high. In the special case of $\eta = 1$ and symmetric product markets we can characterize two important results about the aggregate factor share, which we formalize in the next lemma.

Lemma 2 (Aggregate Factor Share)**.** *Suppose that $\eta = 1$ and all product markets are identical and have n firms. Let $\Gamma_n = \{\gamma_1, \ldots, \gamma_n\}$ be the productivity vector of n the active firms. Then, the following holds:*

a) $\Omega\left(\Gamma_n, n+1\right) > \Omega\left(\Gamma_n, n\right)$,

b) $\Omega\left(\tilde{\Gamma}_n, n\right) < \Omega\left(\Gamma_n, n\right)$ *if $\tilde{\Gamma}_n$ is a mean-preserving spread of Γ_n .*

Proof. See Appendix A.3. ■

Part *a*) of the Lemma states that increasing the number of firms in the economy increases the aggregate factor share. This intuitive result goes through the competitive effect of a larger number of firms, which compresses markups and the profit share and therefore increases the factor share. Part *b*) of the lemma states that economies characterized by higher firm level heterogeneity feature lower factor shares. Suppose we increase firm heterogeneity by increasing the productivity of the best firm in a product market. This implies a reallocation of market shares from low to high productivity firms. In our model this also implies a change in markups. As a consequence of the reallocation and markup response, the average markup increases and the factor share decreases.

Factor Prices and Factor Markets We can write the aggregate demand schedules for labor L_t and capital K_t, the analogs of eq. (1), as

$$W_t = \left(1-\alpha\right)\ \Theta\left(\mathbf{\Gamma},\mathbf{N}_t\right)\ L_t^{-\alpha}\, K_t^{\alpha}, \tag{16}$$

$$R_t = \alpha\ \Theta\left(\mathbf{\Gamma},\mathbf{N}_t\right)\ L_t^{1-\alpha}\, K_t^{\alpha-1}. \tag{17}$$

It is convenient to write the factor price index as the product between the aggregate factor share and aggregate TFP. Combining eqs. (9), (13) and (17), we have

$$\Theta\left(\mathbf{\Gamma},\mathbf{N}_t\right) \;=\; \underbrace{\Omega\left(\mathbf{\Gamma},\mathbf{N}_t\right)}_{\substack{\text{aggregate}\\ \text{factor share}}}\; \underbrace{A_t\,\Phi\left(\mathbf{\Gamma},\mathbf{N}_t\right)}_{\text{aggregate TFP}}. \tag{18}$$

Aggregating over all firms' first order condition (11), we can express $\Theta\left(\mathbf{\Gamma},\mathbf{N}_t\right)$ as a function of the number of active firms (n_{it}), markups (μ_{ijt}) and individual TFP (γ_{ij})

$$\Theta\left(\mathbf{\Gamma},\mathbf{N}_t\right) = A_t \left\{\sum_{i=1}^{I}\left[\sum_{j=1}^{n_{it}}\left(\frac{\gamma_{ijt}}{\mu_{ijt}}\right)^{\frac{\eta}{1-\eta}}\right]^{\frac{1-\eta}{\eta}\frac{\rho}{1-\rho}}\right\}^{\frac{1-\rho}{\rho}}. \tag{19}$$

There are two aspects of eq. (19) that are worth highlighting. First, factor prices are decreasing in markups (holding the number of firms constant): higher markups drive a reallocation of income from factors of production to rents, resulting in lower $\Theta\left(\cdot\right)$. Second, factor prices are increasing in the number of active firms, which happens through two channels: *directly* through the increase in n_{it} (i.e. holding markups fixed) and *indirectly* through the decrease in markups.

The factor demand schedules in (17) can be combined with the labor and capital supplies

$$L_t^S = W_t^{1/\nu} \quad \text{and} \quad K_t^S = K_t \tag{20}$$

to determine the factor market equilibrium. Combining eqs. (17) and (20) with (13), we can write aggregate labor and output as a function of the aggregate capital stock K_t, the productivity distribution $\mathbf{\Gamma}$ and the set of active firms $\mathbf{N}_t$

$$L_t \;=\; \left[\left(1-\alpha\right)\Theta\left(\mathbf{\Gamma},\mathbf{N}_t\right)\right]^{\frac{1}{\nu+\alpha}}\, K_t^{\frac{\alpha}{\nu+\alpha}}, \tag{21}$$

$$Y_t \;=\; A_t\,\Phi\left(\mathbf{\Gamma},\mathbf{N}_t\right)\,\left[\left(1-\alpha\right)\Theta\left(\mathbf{\Gamma},\mathbf{N}_t\right)\right]^{\frac{1-\alpha}{\nu+\alpha}}\, K_t^{\alpha\frac{1+\nu}{\nu+\alpha}}. \tag{22}$$

Both aggregate labor L_t and output Y_t are increasing in the factor price index. Higher factor prices result in higher wages (through (17)) and hence a larger labor supply (through 20). We conclude the characterization of the static equilibrium by determining the set of active firms $\mathbf{N}_t$.

Equilibrium Set of Firms The number of active firms in each product market i is jointly determined by eqs. (19), (22) and the set of inequalities defined in (12). Such a joint system does not admit a general analytical characterization. Nonetheless, we can characterize the particular case in which $\eta = 1$ and all product markets are identical.[17] Proposition 2 states the conditions for

[17] The special case with $\eta = 1$ implies that goods are perfect substitutes within a product market.

a symmetric equilibrium across product markets.

Proposition 2 (Existence and Uniqueness of a Symmetric Equilibrium)**.** *Suppose that* $\eta = 1$ *and that all product markets have the same distribution of idiosyncratic productivities* $\Gamma_i = \Gamma$*. Then there exist two positive values* $\underline{K}(\Gamma, n)$ *and* $\overline{K}(\Gamma, n)$*, with* $\underline{K}(\Gamma, n) < \overline{K}(\Gamma, n)$ *such that when* $K_t \in [\underline{K}(\Gamma, n), \overline{K}(\Gamma, n)]$ *the economy can sustain a symmetric equilibrium with* n *firms in every product market. Furthermore, if the following holds*

$$\frac{\Phi(\Gamma, n)}{\Phi(\Gamma, n+1)} > \left[\frac{\Theta(\Gamma, n)}{\Theta(\Gamma, n+1)}\right]^{\frac{\rho}{1-\rho} - \frac{1-\alpha}{\nu+\alpha}} \qquad \forall n,$$

then there is a unique symmetric equilibrium in $[\underline{K}(\Gamma, n), \overline{K}(\Gamma, n)]$, $\forall n$*. When there are no productivity differences across firms, this condition is equivalent to* $\frac{\rho}{1-\rho} > \frac{1-\alpha}{\nu+\alpha}$.

Proof. See Appendix A.3. ■

Intuitively, K_t must be sufficiently large so that all existing n firms can break even, but cannot be too high, for otherwise an additional firm could profitably enter in at least one product market. Figure 4 illustrates the static equilibrium when all product markets are (ex-ante) identical. When capital is within the bounds $\left[\underline{K}(1), \overline{K}(1)\right]$, the economy is characterized by a monopoly in every product market ($n = 1$); both labor and capital increase in the capital stock, but in a concave fashion (because of decreasing returns). When capital is above $\overline{K}(1)$, at least one product market can sustain a duopoly ($n = 2$). The increase in competition translates into a higher factor price index and a higher labor supply. For this reason, output is locally convex on capital when $K_t \in \left[\overline{K}(1), \underline{K}(2)\right]$.[18]

The last part of the proposition provides a condition for uniqueness of symmetric equilibria, in the special case in which firms are identical. If the across product market elasticity of substitution $\sigma_I = 1/(1-\rho)$ is high, if the capital elasticity α is large or if the inverse Frisch elasticity ν is large, it is easier for the condition for uniqueness to be satisfied. We return to the discussion on uniqueness in the calibrated version of our model.

3.4.2 Equilibrium Dynamics

We next explore the dynamic properties of our economy. Denoting by s_t is the aggregate savings rate (from the household maximization problem), we can write the capital law of motion as

$$K_{t+1} = (1-\delta) K_t + s_t \cdot Y_t. \tag{23}$$

[18] In the regions $\left[\overline{K}(n), \underline{K}(n+1)\right]$, the economy is characterized by an asymmetric equilibrium in which some product markets have n firms, and some others have $n+1$ firms. The number of product markets with $n+1$ firms is such that the last firm exactly breaks even. See Section B.3 for details.

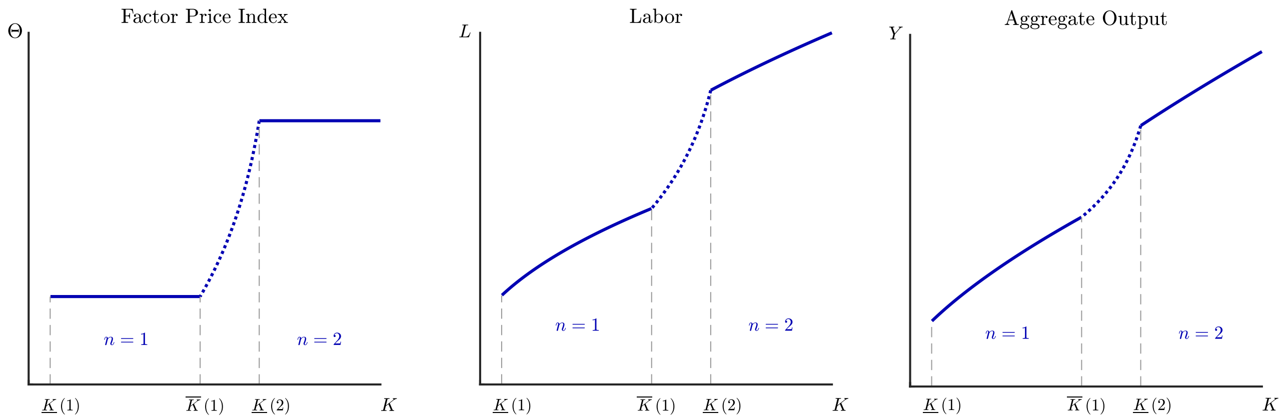


Figure 4: Static equilibrium

Note: the figure shows how the factor price index, aggregate labor and output move with capital. Solid segments represent a regions with a symmetric equilibrium across product markets, while dotted segments represent non-symmetric regions. We use $\alpha = 1/3$, $\rho = 3/4$, $\eta = 1$, $\nu = 2/5$, $\gamma_{ij} = 1$ and $c_i = 0.015$.

Even though we cannot provide an analytical characterization of s_t, we can establish that, in a steady-state, it is proportional to the aggregate factor share $\Omega(\mathbf{\Gamma}, \mathbf{N})$. Lemma 3 establishes that a more competitive market structure, by resulting in a larger factor share $\Omega(\mathbf{\Gamma}, \mathbf{N})$, leads to a higher steady-state savings rate (and hence capital supply).

Lemma 3 (Steady-State Savings Rate)**.** *In a steady-state, the following holds:*

$$s^* = \frac{\beta\,\delta}{1-(1-\delta)\,\beta}\;\underbrace{\alpha\,\Omega(\mathbf{\Gamma}, \mathbf{N})}_{\textit{capital share}}. \tag{24}$$

Furthermore, as a consequence of Lemma 2, s^* is increasing in the number of firms.

Proof. See Appendix A.3 ∎

The left panel of Figure 5 represents the law of motion (23) for a particular parameter combination. For the sake of exposition, we assume again that all product markets are (ex-ante) identical and that the economy is subject only to one time unexpected shocks (we relax these assumptions in the quantitative model). First, note that K_{t+1} is not globally concave in K_t. Second, the economy features multiple steady-states: there are two stable steady-states, K_1^* and K_2^*, and an unstable one, K_U.[19] The shape of the law of motion (and the existence of multiple steady-states) can be explained by the interaction between competition, factor prices and factor supply. Within the colored regions, the economy is characterized by the same market structure (same number of firms n_{t+1}). This

[19]It is important to note that this map is a function, rather than a correspondence, which would signal the possibility of dynamic multiplicity. In this setting the initial conditions fully determine the steady-state the economy will converge to.

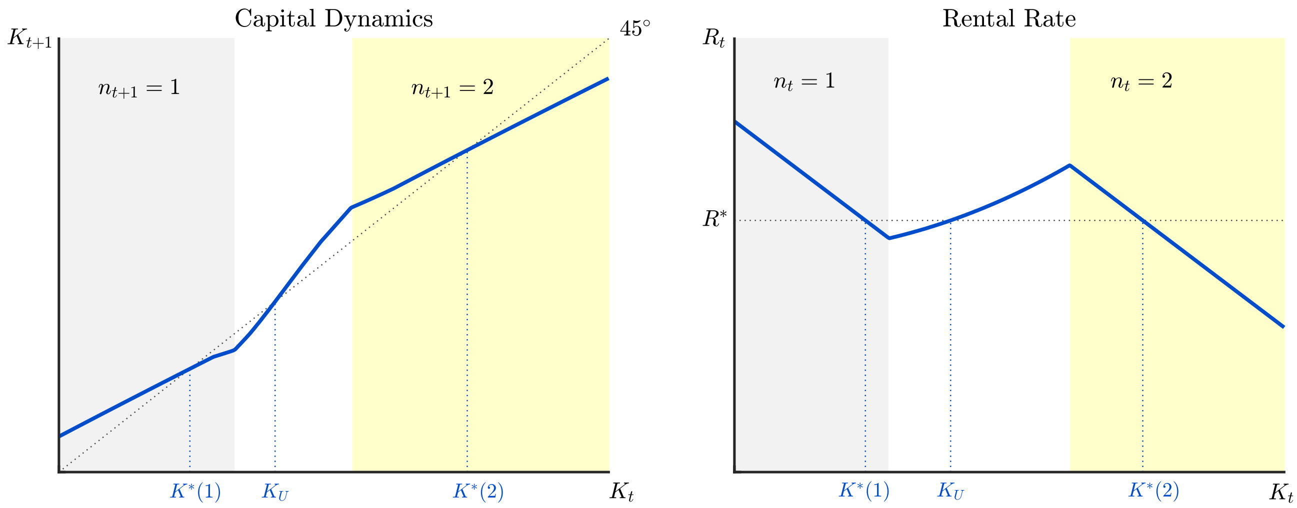


Figure 5: Law of Motion and Rental Rate Map

Note: This example features two stable steady states and an unstable one. We use $\psi = 1$, $\rho = 3/4$, $\eta = 1$, $\alpha = 1/3$, $\delta = 1$, $\nu = 2/5$ and $c_i = 0.015$.

ensures that the law of motion is concave within these regions, as in a standard neoclassical growth model. However, in the regions coinciding with changes in the market structure, the law of motion is convex. To understand this, consider again eq. (23). The law of motion can be convex in capital if at least one of two conditions holds: i) Y_t is convex in K_t or ii) s_t increases sufficiently fast in K_t. As already highlighted in Figure 4, Y_t can be convex in K_t because of the positive impact of competition on labor supply. On the other hand, more intense competition can also result in a larger savings rate s_t and hence a larger supply of capital. To sum up, relative to K_1^*, steady-state K_2^* is characterized by a more competitive market structure, and hence a larger supply of labor and capital by the representative household.

The right panel of Figure 5 represents the rental rate map of this economy. As discussed in Section 2, multiple steady-states occur whenever this map crosses the steady-state rental rate multiple times. A steady-state is characterized by a constant rental rate equal to $R^* = \beta^{-1} - (1-\delta)$. Proposition 3 characterises the conditions for the existence of multiple steady-states.

Proposition 3 (Existence of Multiple Steady States)**.** *Suppose that all product markets have the same distribution of idiosyncratic productivities $\Gamma_i = \Gamma$. The economy features multiple symmetric steady states if and only if there exists an $n \in \mathbb{N}$ such that*

$$\Theta\left(\Gamma, n\right)^{\frac{1+\nu}{\nu+\alpha}} \; \overline{K}\left(\Gamma, n\right)^{-\nu\frac{1-\alpha}{\nu+\alpha}} < \frac{\beta^{-1} - (1-\delta)}{\alpha \; (1-\alpha)^{\frac{1-\alpha}{\nu+\alpha}}} < \; \Theta\left(\Gamma, n+1\right)^{\frac{1+\nu}{\nu+\alpha}} \; \underline{K}\left(\Gamma, n+1\right)^{-\nu\frac{1-\alpha}{\nu+\alpha}}. \tag{25}$$

where $\underline{K}\left(\Gamma, n\right)$ and $\overline{K}\left(\Gamma, n\right)$ are defined in Appendix A.3.

Proof. See Appendix A.3. ∎

Proposition 3 formalizes the idea that multiplicity obtains if there exists an increasing segment of the rental rate map and this segment crosses the steady-state interest rate $R^* = \beta^{-1} - (1 - \delta)$. This condition depends on fundamental parameters such as the productivity distribution or the fixed production cost. For example, both $\underline{K}(\Gamma, n + 1)$ and $\overline{K}(\Gamma, n)$ are strictly increasing in the fixed cost c. Therefore, as fixed costs change, the economy may enter/exit a region of steady-state multiplicity. Throughout this section, we assume that product markets are ex-ante symmetric and that the condition of Proposition 3 is satisfied, so that multiple steady-states exist. In Section 4, we relax the assumption of product market symmetry and assess whether multiple steady-states arise or not under different calibrations of our model.

Finally, applying the definition of *fragility* from Section 2 to this particular economy, we have $\chi_2 = K_U / K^* (2)$. We next ask how changes in the technology set of the economy can affect *fragility*. We consider two comparative statics exercises: a mean-preserving spread (MPS) to the productivity distribution of active firms and an increase in fixed costs. These technological shifts can explain some of the micro trends that have taken place since the 1980s (e.g. rising concentration or markups). Next we study the aggregate consequences of rising heterogeneity and fixed costs, with a particular focus on the likelihood of persistent transitions.

3.5 Comparative Statics

Firm Heterogeneity We start by studying the impact of an MPS in the distribution of productivities on *fragility*. The next results obtain under two assumptions, which are necessary to have an analytical characterization of the model's steady-state(s) i) all product markets are ex-ante identical and ii) $\eta = 1$ (i.e. no within product market differentiation). We dispense with both assumptions in our quantitative application. We start by characterizing the response of the factor price index $\Theta(\cdot)$ to a productivity spread.

Lemma 4 (Mean Preserving Spread and Factor Prices)**.** *Let $\eta = 1$ and suppose that all product markets are identical and have n firms. Let $\Gamma_n = \{\gamma_1, \ldots, \gamma_n\}$ be the productivity vector of the n active firms and $\tilde{\Gamma}_n$ be a mean-preserving spread on Γ_n. Then $\Theta(\tilde{\Gamma}_n, n) < \Theta(\Gamma_n, n)$.*

Proof. See Appendix A.3. ■

This result states that, in a symmetric steady state with n firms, a spread of the productivity distribution decreases the aggregate factor price index. As formalized by the first part of Proposition 4, this is a sufficient condition for the stable steady-state level of capital to decrease.

Proposition 4 (Firm Heterogeneity and Fragility)**.** *Let $\eta = 1$ and suppose that all product markets are identical. Let $K^*(n)$ be a stable steady-state with n firms in every product market. Let $\Gamma_n = \{\gamma_1, \ldots, \gamma_n\}$ be the productivity vector of the n active firms and $\tilde{\Gamma}_n$ be a mean-preserving spread on Γ_n. The following holds* *a)* $K^*(\tilde{\Gamma}_n, n) < K^*(\Gamma_n, n)$.

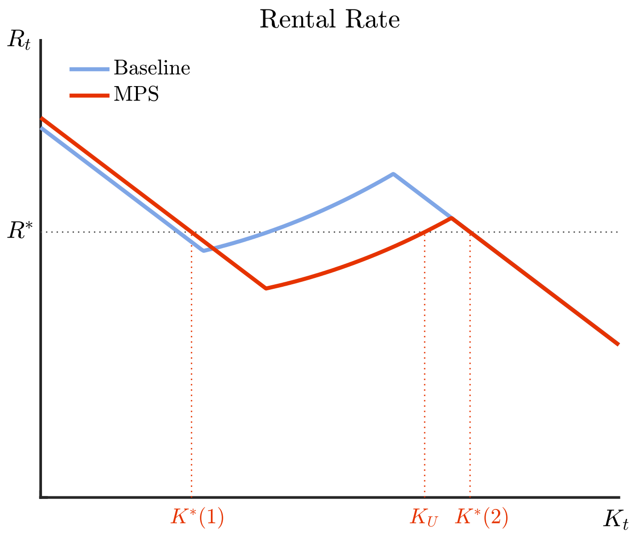
b) *If* $n = 2$ *and* $\rho > 1 - \dfrac{\nu(1-\alpha)}{1+\nu\alpha}$, *then* $K_U(\tilde{\Gamma}_2) > K_U(\Gamma_2)$.

Proof. See Appendix A.3. ■

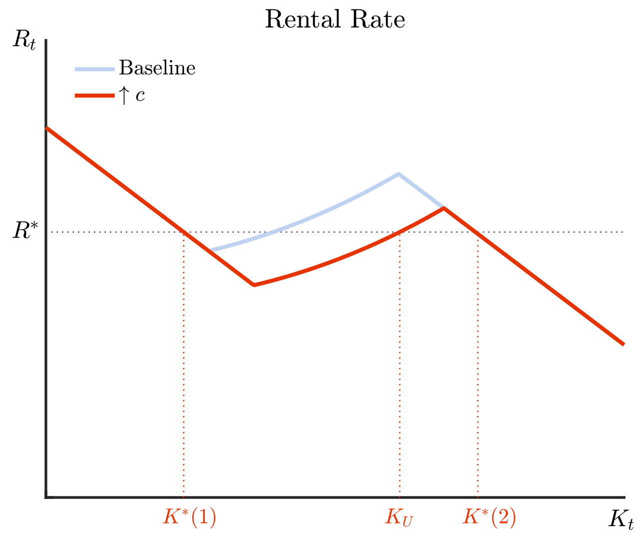
Corollary 1. *If the condition in Proposition 4b) is satisfied, the highest steady state* $K^*(\tilde{\Gamma}_2, 2)$ *becomes more fragile after a mean-preserving spread. Formally,* $\chi_2(\tilde{\Gamma}_2) > \chi_2(\Gamma_2)$.

Proposition 4 is the analog of Proposition 1 in Section 2. Before explaining its results, let us start by discussing the distributional consequences of an MPS. Fixing the number of firms, an MPS of idiosyncratic productivities, by increasing product market concentration, results in a lower factor share (eq. 15). This pushes the factor price index down, for a given level of aggregate TFP, as eq. (18) highlights. This is a *market power effect* associated with higher firm heterogeneity, which results in lower factor prices. However, aggregate TFP is likely to increase after an MPS: large, high productivity firms become even more productive and increase their market shares. This pushes the factor price index up (eq. 18). This is an *allocative efficiency effect*, which results in higher factor prices. The impact of an MPS on factor prices depends on the relative strength of these two forces. These effects are the more precise characterizations of the elements highlighted in Section 2. Proposition 4 specifies the result from Proposition 1 under an MPS of productivities.

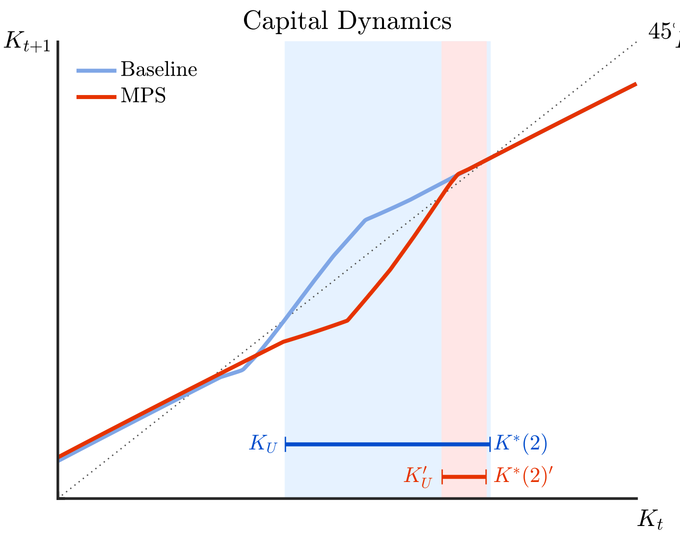
Proposition 4 formalizes two results. The first is that, if there exists a steady-state where all product markets are identical and have n firms, the steady-state level of capital necessarily shrinks after an MPS of the productivities of these n firms. This happens because the *market power effect* always dominates, so that $\Theta(\cdot)$ declines. The steady-state level must shrink, so that eq. (17) holds at the same R^*. The second part of Proposition 4 characterizes the behavior of the unstable steady-state K_U. It provides a sufficient condition under which the unstable steady-state moves rightward after an MPS. As the rental rate map is upward slopping at K_U, this steady-state moves rightward whenever the rental rate falls. This is illustrated in the left panels of Figure 6. Panel (a) shows that capital demand decreases for sufficiently high values of K_t. This translates into a higher proximity between the unstable steady-state K_U and the highest stable steady-state $K^*(2)$, as illustrated in panel (c). Proposition 4b) provides a sufficient condition under which the rental rate falls in the region in which it is increasing. The difficulty in establishing a result for this region is that, even when all product markets are ex-ante identical, there are ex-post differences: some have n firms, while others feature $n-1$ firms. For tractability reasons, Proposition 4b) focuses on the case of $n = 2$. Proposition A.1 in Appendix A.3 provides a result for general n. Proposition 4b) states that K_U increases when ρ is sufficiently high. A high value of ρ means that the degree of cross product market differentiation and of average markups are relatively low. Small productivity differences are magnified and entry becomes more difficult for a second player (which brings down aggregate factor shares and factor prices).



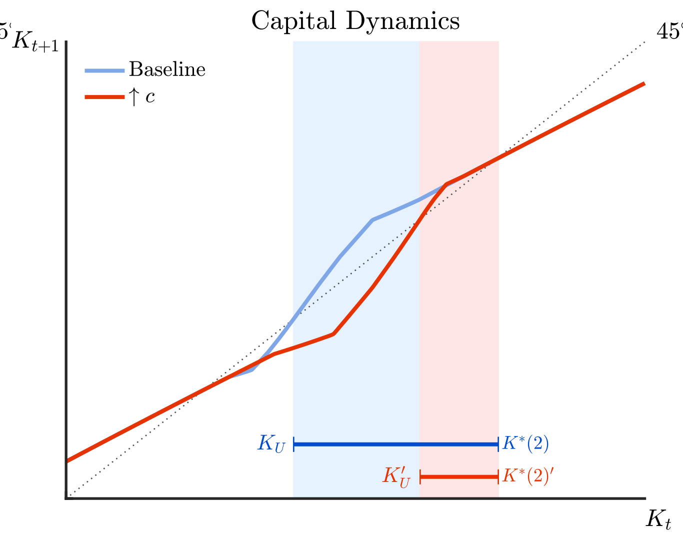
(a) Effect of an MPS on capital demand



(b) Effect of $\uparrow c$ on capital demand



(c) Effect of an MPS on the law of motion



(d) Effect of $\uparrow c$ on the law of motion

Figure 6: Comparative Statics

Panels (a) and (c) represent the effect of an MPS on capital demand and the law of motion of capital (in an economy with two steady-states). Panels (b) and (d) represent the effect of an increase in fixed costs. We use $\rho = 3/4$, $\eta = 1$, $\alpha = 1/3$ and $\nu = 2/5$.

Taken together, as highlighted in Corollary 1, the two parts of Proposition 4 provide a sufficient condition under which fragility increases after an MPS. Finally, Proposition 4 focuses on the consequences of an MPS with $n = 2$. By construction, we increase the productivity of the top firm and reduce the productivity of the second one, keeping their average unchanged. However, fragility can also increase if we increase the productivity of the top firm, leaving the productivity of the second firm unchanged.[20]

Rising Fixed Costs The decline in product market competition since the 1980s may also be explained, through the lens of our model, by rising fixed costs.[21] As fixed costs rise, they drive out of the market those firms that exactly break even. This results in higher markups and concentration. Their effect on the capital demand schedule can be understood in two steps. First, in a steady-state with n firms where all firms make strictly positive profits, the capital level of such a steady-state $\left(K_n^{SS}\right)$ is unaffected by a marginal increase in c_f. On the other hand, the larger fixed costs result in a rightward shift of the unstable steady-state. Recall that K_U belongs to a region where some firms are breaking even. These firms can break even with a higher fixed cost only if the capital stock is higher. Proposition 5 and Corollary 2 formalize these results. Differently from Proposition 4 we make no assumption of symmetry across product markets or perfect substitution.

Proposition 5 (Fixed Costs and Fragility)**.** *Let $K^*(\mathbf{\Gamma}, \mathbf{N})$ be a stable steady-state with a set of active firms $\mathbf{N}$. Then the following holds*

a) $\dfrac{\partial K^*(\mathbf{\Gamma}, \mathbf{N})}{\partial c} \leq 0$.

b) $\dfrac{\partial K_U(\mathbf{\Gamma}, \mathbf{N})}{\partial c} > 0.$

Proof. See Appendix A.3. ■

Corollary 2. *When fixed costs increase, any stable steady-state becomes more fragile:* $\dfrac{\partial \chi(\mathbf{\Gamma}, \mathbf{N})}{\partial c} > 0.$

Panels (b) and (d) of Figure 6 illustrate the effect of an increase in fixed costs. We represent again an economy where all product markets are ex-ante identical (same distribution of productivities and identical fixed costs). The two declining segments of the rental rate map represent a situation of full monopoly and full duopoly; since all firms make strictly positive profits in these regions, the equilibrium is unchanged as c increases marginally. In the increasing segment, where some firms are exactly breaking even, some are driven out of the market. The rental rate decreases and the unstable steady-state increases. As a result, the economy becomes more fragile (Corollary 2).

These results also shed light on the conditions of Proposition 3.A change in fixed costs can affect the existence of multiple steady-states. This is made clear in Figure 6: a sufficiently large increase in fixed costs can shift the rental rate map downward such that it crosses R^* only once. In such a case, only one steady-state exists, where all product markets are monopolies ($K^*(1)$). The level of capital $K^*(2)$ is no longer a steady-state. The opposite would happen if fixed costs were to experience a sufficiently large decrease. In such a case, only $K^*(2)$ would be a steady-state.[22]

[20]Figure B.1 in Supplementary Material Section B.4 shows one such example.

[21]This channel is also studied in De Loecker et al. (2021).

[22]Figure B.2 in Supplementary Material Section B.4 shows one such example.

Discussion We conclude by summarizing two keys insights of our theory, which are relevant to understanding the US growth experience after 2008. The first is that a complementarity between competition and factor supply can generate multiple competition regimes or steady-states. A transition from a high competition to a low competition regime can in many aspects describe the 2008 recession and the subsequent great deviation. The second insight is that changes in technology that result in larger market power (e.g. larger productivity differences across firms or larger fixed costs) make high competition regimes more difficult to sustain, and transitions to low competition traps more likely to occur. Our model therefore suggests that the US economy, experiencing a long-run increase in markups and concentration since the 1980s, became increasingly vulnerable to transitions like the one observed after 2008.

In the next section, we use a calibrated version of our model and study its quantitative predictions. We then ask whether it can replicate the behavior of the US economy in the aftermath of the 2008 crisis and finally study the welfare gain of policy interventions.

4 Quantitative Results

The goal of this section is to develop a quantitative version of the model built in Section 3. We use it to provide a quantification of the forces described earlier, and to evaluate policy counterfactuals.

There are two objects that we need to parametrize: the distributions of idiosyncratic productivities γ_{ij} and of fixed costs c_i. We assume that firms draw their idiosyncratic productivities from a log normal distribution with standard deviation λ, i.e. $\log \gamma_{ij} \sim \mathrm{N}(0, \lambda)$. Each product market i is characterized by M such draws. Since M is a finite number, product markets have different ex-post distributions of idiosyncratic productivities $\{\gamma_{ij}\}_{j=1}^{M}$.

Furthermore, we assume that there are two types of product markets: a fraction $f \in (0, 1)$ of all product markets face a fixed cost $c_i = c > 0$, whereas the remaining fraction $1 - f$ faces a zero fixed cost $c_i = 0$. We refer to the first as *concentrated* and the second as *unconcentrated* product markets. In product markets with a zero fixed cost, the extensive margin is muted as all potential M entrants are always active. However, these product markets do not necessarily operate close to perfect competition, as there can be large productivity differences across firms, resulting in high concentration and large markups for productive firms. Importantly, even if idiosyncratic productivities are drawn from the same distribution, there will be ex-post differences in productivity draws. This heterogeneity acts against the existence of steady-state multiplicity.[23]

[23] We assume that there is a common fixed cost c among all *concentrated* product markets. These product markets differ in their distributions of idiosyncratic TFP draws $\{\gamma_{ij}\}_{j=1}^{M}$ and may display a different number of active firms.

4.1 Calibration

The model is calibrated at a quarterly frequency. Our calibration strategy relies on the interpretation that, if two or more steady-states exist, the economy starts in the highest one. Some parameters are standard and taken from the literature. For the preference parameters, we set $\beta = 0.99$ and $\psi = 1$ to have log utility. We set $\nu = 0.352$, which implies a Frisch elasticity of 2.84; this corresponds to the average macro elasticity of hours reported by Chetty et al. (2011). We set the number of product markets at $I = 10,000$ and the maximum number of firms per product market at $M = 20$.[24] For technology parameters, we set the capital elasticity to $\alpha = 0.3$ and the depreciation rate to $\delta = 0.025$. There are 5 parameters that we need to calibrate internally: the within and across product market elasticities of substitution (σ_G and σ_I), the standard deviation of the log-normal productivity distribution for the pool of potential entrants (λ), the fixed cost for the *concentrated* sector (c), and the fraction of *concentrated* product markets (f).

Before describing in detail our calibration strategy we briefly discuss our algorithm, which is similar to De Loecker et al. (2021). We begin by drawing productivity values for all $20 \times 10,000$ potential firms in the economy. For the product markets without entry costs, these also represent the number of active firms, while for the product markets with $c_i > 0$ the free entry condition might not hold. To check that the slackness free entry condition holds at each pair of state variables (K_t, A_t), we start with a full matrix of firms. If any firm makes negative profits, we eliminate the firm with the largest negative value and recompute the equilibrium. We do so until all firms make non-negative profits. This procedure is more computationally demanding than the alternative one in which we start with an empty matrix and fill it with firms as long as they make positive profits. The reason behind our choice is that iterative deletion implies, in the case of multiple equilibria, we consistently select the one with the highest number of active firms. We view our approach as being conservative since the mechanism we propose has the largest potential effect when few firms are active in the market. Therefore our equilibrium selection algorithm provides us with a lower bound on the strength of our mechanism in case there are multiple static equilibria. The condition for static uniqueness presented in Proposition 2 was derived in the special case of $\eta = 1$ and product market and firm symmetry – which are not satisfied in this calibrated version of the model. The introduction of product market and firm heterogeneity makes uniqueness more likely to obtain.

We provide three different calibrations of our economy, where we target the same moments in different points in time: 1975, 1990 and 2007. Our calibration relies on the assumption that the elasticities σ_G and σ_I are time-invariant, while the other 3 parameters are allowed to vary over time. This allows us to compare features of the model in economies with different levels of firm heterogeneity and fixed costs. We target 4 moments in each of the three calibration years. We thus have 12 moments for 11 parameters (2 elasticities and 3x3 time-varying parameters). This gives us

[24] Contrarily to De Loecker et al. (2021) we fix this parameter across the different calibrations. We have also considered $M = 50$ and $M = 100$ and the results were identical.

an over-identified system, which we calibrate jointly.

Importantly, we design our calibration so that it does not depend on assumptions about a particular level of aggregation, i.e. how to map a market i in our model to a market i in the data. While for some tradeable-good firms the relevant competitive market might be a global 10-digit industry, for local service providers it might be as narrow as a neighborhood. Our calibration choices imply that we need not take a stance on this mapping and that we can think of some product markets in our model as a combination of *product market*$\times$*location* in the data (see Eeckhout, 2021; Benkard et al., 2021, for a more complete discussion).[25] For the two elasticities of substitution, we assume that they are time-invariant and calibrate them to match the sales-weighted average markup of public firms (as reported by De Loecker et al., 2020). Conditional on the other time-varying parameters, the elasticities are informative about the level of markups charged by firms in the economy. We calibrate the standard deviation of the exogenous productivity distribution (λ) to match dispersion in firm size. Specifically, we compute the standard deviation of log revenues for all firms in our economy and use as a target the corresponding moment in COMPUSTAT. Even though COMPUSTAT is not a representative sample of firms for the whole economy, we choose this dataset for two reasons. First, it covers all sectors in the economy (and is not restricted to specific sectors, such as manufacturing). Second, being a firm-level dataset, it allows us to obtain a measure of firm-level dispersion that does not depend on a particular level of aggregation (e.g. 4-digit NAICS).[26] Figure 2 shows the evolution of our dispersion measure in COMPUSTAT. We calibrate the fixed cost parameter to match the average ratio of fixed to total costs in COMPUSTAT. Following Gorodnichenko and Weber (2016), we define fixed costs as sum of 'Selling, General and Administrative Expenses', 'R&D Expenditures' and 'Advertisement Expenses'. We obtain a target level of 24.4% for 1975, 35.5% for 1990 and 41.4% for 2007 (see Figure B.8). To calibrate f, we target the fraction of aggregate employment in highly concentrated industries. In the calibrated model, *concentrated* product markets have fewer than 4 firms. In an ideal setting, we would have direct information on the aggregate employment share of markets with 4 or fewer firms. As discussed above, we think of an product market in our model as a market at the highest level of disaggregation (e.g. 10-digit NAICS or, in some case, as a product market-location pair). However, we only have information of concentration metrics at the 6-digit NAICS level. Since this will understate concentration levels in finer markets, we scale the market share of the 4 largest firms in 6-digit NAICS, by the share of the largest 8 firms. Then, we define an product market as *concentrated* if

[25]As Benkard et al. (2021) report, some industries can be too broad for some products (e.g. NAICS 325620 contains products such as after-shave, mouthwash or sunscreen). In other cases, highly substitutable products belong to different industries (e.g. metal cans or glass bottles).

[26]Some aggregate datasets, such as the BLS Multifactor Productivity Database, provide statistics on firm-level differences. However, these differences are computed within a 4-digit industry and data is restricted to the set of manufacturing industries.

the share of the top 4 to top 8 is above 90%.[27] Using data from the US Census, we find that 6.33% of aggregate employment is allocated to such 6-digit industries in 2007.

We also need to calibrate the two parameters governing the dynamics of aggregate productivity: the autocorrelation parameter ϕ_z and the standard deviation of the innovations σ_ε. We do so by targeting the first order autocorrelation and the standard deviation of output.[28] We calibrate these parameters using our 2007 model, and keep them unchanged in the 1975 and 1990 models.

Description	Parameter	Value			Source/Target
[A] External Parameters					
Capital elasticity	α		0.3		Standard value
Depreciation rate	δ		0.025		Standard value
Discount factor	β		0.99		Standard value
Inverse of Frisch elasticity	ν		0.352		Chetty et al. (2011)
Coefficient of risk aversion	ψ		1		log utility
Max number of firms per product market	M		20		
Number of product markets	I		10,000		
[B.1] Calibrated Parameters: Fixed					
Between product markets ES	σ_I		1.38		Sales-weighted average markup
Within product market ES	σ_G		11.13		Sales-weighted average markup
Persistence of z_t	ρ_z		0.950		Autocorrelation of log Y_t
Standard deviation of ε_t	σ_ε		0.003		Standard deviation of log Y_t
[B.2] Calibrated Parameters: Variable		1975	1990	2007	
Fraction of product markets with $c_i > 0$	f	0.110	0.135	0.140	Emp share concentrated industries
Standard deviation of γ_{ij}	λ	0.190	0.283	0.328	Std log revenues
Fixed cost ($\times 10^{-3}$)	c	0.47	0.96	1.34	Average ratio fixed/total costs

Table 1: Parameter Values

Table 1 reports our parameter values, while Table 2 reports our targeted moments, with their model counterparts. We obtain a value of 1.38 for the cross product market elasticity of substitution, and a value of 11.25 for the within product market elasticity. These are in line with the estimates from similar studies using US data, such as Edmond et al. (2021). The model is successful at matching the sales-weighted markups and dispersion in revenues in all three years. We slightly over-estimate the fixed cost ratio in 1975 and 1990 while we underestimate it in 2007.

[27] Alternative thresholds for the ratio of the top 4 to the top 8 (80% and 95%) yield identical results.

[28] We compute these moments for the entire postwar period 1947–2019. Consistent with our interpretation that the US economy moved to a different regime after 2008, we remove a linear trend computed for the period 1947–2007.

	1975		1990		2007	
	Data	Model	Data	Model	Data	Model
Sales-weighted average markup	1.28	1.26	1.37	1.38	1.46	1.44
Std log revenues	1.67	1.65	2.47	2.42	2.79	2.83
Average fixed to total cost ratio	0.244	0.271	0.355	0.376	0.414	0.397
Emp share *concentrated* product markets	-	0.066	-	0.071	0.063	0.068
Autocorrelation log GDP					0.978*	0.975
Standard deviation log GDP					0.061*	0.062

*computed over 1947:Q1-2019:Q4

Table 2: Targeted moments and model counterparts

Part [A] of Table 3 reports the evolution of some non-targeted moments. When computing aggregate markups with cost-weights, we find an average of 1.24 in the 1975 economy, and an average of 1.34 in 2007. While these values are slightly above the recent estimates by Edmond et al. (2021), the time variation is similar in the model and in the data (10pp increase in the model, against 9pp increase in the data). To characterize some of the aggregate consequences of the change in the three structural parameters (λ, c, f), we also report the evolution of aggregate TFP. Our model predicts an 16.7% increase in aggregate TFP, which represents roughly half of the 29.6% increase observed in the data.

Product markets facing positive fixed costs play an important role in our mechanism. Part [B] of Table 3 provides a characterization of these product markets in the three calibrated economies. Product markets with positive fixed costs consist mostly of monopolies and duopolies — the average number of firms is 1.97 in the 1975 economy, 1.60 in 1990 and 1.47 in 2007. This implies an average markup of 1.75 in 1975, 2.32 in 1990 and of 2.59 in 2007 in these product markets. These values are within the bounds of estimates for the markup distribution of US firms. For example, De Loecker et al. (2020) report that the 90th percentile of the (sales-weighted) markup distribution increased from 1.57 in 1975 to 2.25 in 2007. Concentrated product markets represent 6.8% of aggregate employment in the 2007 model, so approximately $1/15$ of aggregate employment is concentrated in monopolist or duopolist firms. We regard these numbers are reasonable for two reasons. First, in our model a market can be as disaggregated as narrow product market-location pair. Second, the right tail of the empirical markup distribution displays levels consistent with monopolies and duopolies, given our calibrated elasticities, which are similar to the values found in other studies.

Table 4 shows the distribution of HHIs obtained under the 2007 calibration. We also show the same moments reported by Benkard et al. (2021), who estimate concentration metrics for narrowly defined consumption-based product markets. The model closely matches both the 50th and the 90th percentiles of the empirical distribution of HHIs. An HHI of 0.5 is the one that would be obtained

	1975		1990		2007	
	Data	Model	Data	Model	Data	Model
[A] Total Economy						
Cost-weighted average markup	1.16	1.24	1.20	1.30	1.25	1.34
Aggregate TFP (log)	0.00	0.00	0.091	0.098	0.296	0.167
[B] Concentrated product markets						
Number of firms per product market		1.97		1.60		1.47
Sales-weighted average markup		1.74		2.32		2.59

Table 3: Non-targeted moments

Note: The cost-weighted average markup is from Edmond et al. (2021). Aggregate TFP is from Fernald (2012). The value of log TFP in 1975 is normalized to zero.

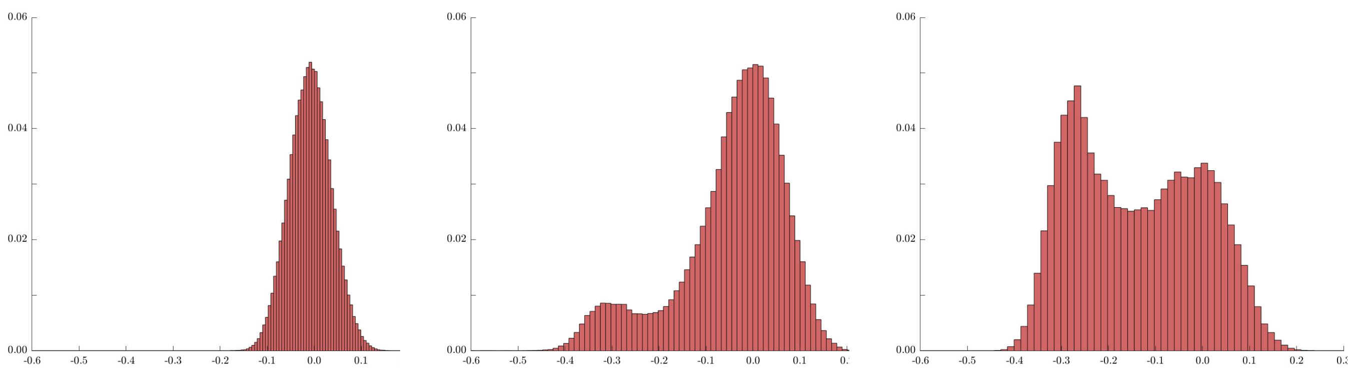
under a symmetric duopoly. This implies that both in our model and in the data 10% of the markets have concentration metrics equal or higher than that of a symmetric duopoly. Therefore, even if we do not impose a particular level of aggregation in our calibration, the model can replicate the distribution of concentration for consumption-based product markets.

	$p\,(10)$	$p\,(25)$	$p\,(50)$	$p\,(75)$	$p\,(90)$
model	0.135	0.160	0.200	0.271	0.501
Data: local market (Benkard et al., 2021)	0.0852	0.140	0.220	0.335	0.484
Data: national market (Benkard et al., 2021)	0.0863	0.132	0.214	0.319	0.456

Table 4: Distribution of HHI within product markets: model *versus* data

4.2 Quantitative Results

Having calibrated the model, we solve the full dynamic problem, approximating the policy function of the household, by iterating on the Euler equation. We describe the algorithm in Section B.5. We start by comparing the dynamic properties of the 1975, 1990 and 2007 economies. Figure 7 shows the ergodic distribution of log output; the distributions are centered around the highest mode, so that the horizontal axis represents output in percentage deviation from the highest steady-state. We highlight three important observations. First, while the ergodic distribution of the 1975 economy is unimodal, the other two economies feature bimodal distributions, implying that these two economies are characterized by multiple (stochastic) steady-states. Through the lens of our model, multiple competition regimes are possible in economies characterized by levels of markups and fixed costs observed in 1990 and 2007, but not in 1975, when markups and fixed costs were lower.[29] Second, relative to the 1990 economy, the 2007 model features a larger probability mass on the left, suggesting that the economy on average spends more time on the lowest regime, characterized by lower competition and output. Third, in the 2007 distribution, the two steady-states are also closer to each other — a transition from the high to the low regime implies a 19% reduction in output in 2007, as opposed to approximately 30% in 1990. While this means that transitions are less pronounced in 2007, it also implies that they are substantially more likely in 2007 than in 1990, as we discuss below. The 1975 economy, instead, behaves similarly to a standard RBC model. This economy can suffer temporary recessions, but these will not have long-lasting consequences. Table B.1 provides business cycle moments for the three economies.



(a) 1975 (b) 1990 (c) 2007

Figure 7: Ergodic distribution of output

Note: This figure shows the distribution of log output for the 1975, 1990 and the 2007 economies. We simulate each economy for 10,000,000 periods and plot output in deviation from the high steady state.

We next study the probabilities of deep recessions in the three economies. We simulate each economy 100,000 times for 40 and 100 quarters. We compute the fraction of simulations in which output experiences 10%, 15% or 20% drop relative to the high steady state level (for at least 4 consecutive quarters). The results are shown in Table 5. When running the 2007 economy for 40 quarters, output drops by at least 10% in 19.3% of the simulations, whereas the same figure for the 1990 economy is 9.6%. In the 1975 economy this type of deep recession is extremely rare as it only occurs in 0.7% of our simulations. Relative to 1990, the 2007 economy is approximately twice as likely to experience a 10% fall in output over a 10-year period. Over 100 quarters, the

[29]As highlighted in discussion of Proposition 3 and later of Figure 6, changes in fixed costs can affect the condition for the existence of multiple steady-states.

2007 economy appears about 1.67 times more likely to experience a 10% recession (35% probability in 2007, against 21.5% probability in 1990). The same probability for the 1975 economy is 3.4%. This implies that, in expectation, the 2007 economy experiences a recession larger than 10% every 75 years, the 1990 economy does so every 120 years and the 1975 economy every 735 years. When focusing on larger recessions, the difference is even starker: the 1975 economy never has output drops larger than 15%, while these events remain somewhat likely in the 2007 and 1990 economies. As a final remark on the ergodic properties of these economies, recall that the exogenous aggregate shocks process in unchanged across the three calibrations. The large observed differences in the ergodic behavior is fully driven by the endogenous response of the economies. The three calibrations only differ in parameters related to the firm heterogeneity and market structure. Therefore we attribute the observed differences in the business cycle properties of the three economies to the heterogeneity in competitive structure.

	1975 Model		1990 Model		2007 Model	
	T = 40	T = 100	T = 40	T = 100	T = 40	T = 100
Pr[10% recession]	0.007	0.034	0.096	0.215	0.193	0.350
Pr[15% recession]	0.000	0.000	0.018	0.082	0.068	0.206
Pr[20% recession]	0.000	0.000	0.002	0.035	0.011	0.104

Table 5: Probability of deep recessions

This table shows the probabilities of deep recessions in the 1975, 1990 and 2007 economies. Each economy starts in the highest steady-state and is simulated for $T = 40$ and $T = 100$ quarters. Each simulation is repeated 100,000 times. The table reports the fraction of simulations in which output is κ% below the initial value for at least 4 consecutive quarters.

The likelihood of a transition is related to the distance between the steady states of the economy. It is possible that, while slumps are more frequent in the 2007 economy, they are associted with lower drops in output, relative to the 1990 economy. It is not ex-ante clear which effect dominates and therefore whether the 2007 economy should be more or less volatile than the 1990 one. To investigate this, we report additional moments of our simulated economies in Table 6. The first row reports the average deviation of log output from the highest mode of the ergodic distribution. Since the 1975 distribution is symmetric around its unique mode, this deviation is zero. The 1990 and 2007 economies, on the other hand, feature sizeable deviations as they have multiple steady staes. In these economies, output is on average 5% and 12% below their respective high steady state. The second row of Table 6 shows that higher firm heterogeneity and fixed costs result in higher volatility. The higher likelihood of transitions dominates and relative 1975, the 1990 and 2007 economies are 2.5 and 3 times more volatile.

	1975 Model	1990 Model	2007 Model
$\overline{\log Y} - \log Y^H$	0.00	-0.05	-0.12
$\sigma_{\log Y}$	0.045	0.116	0.136

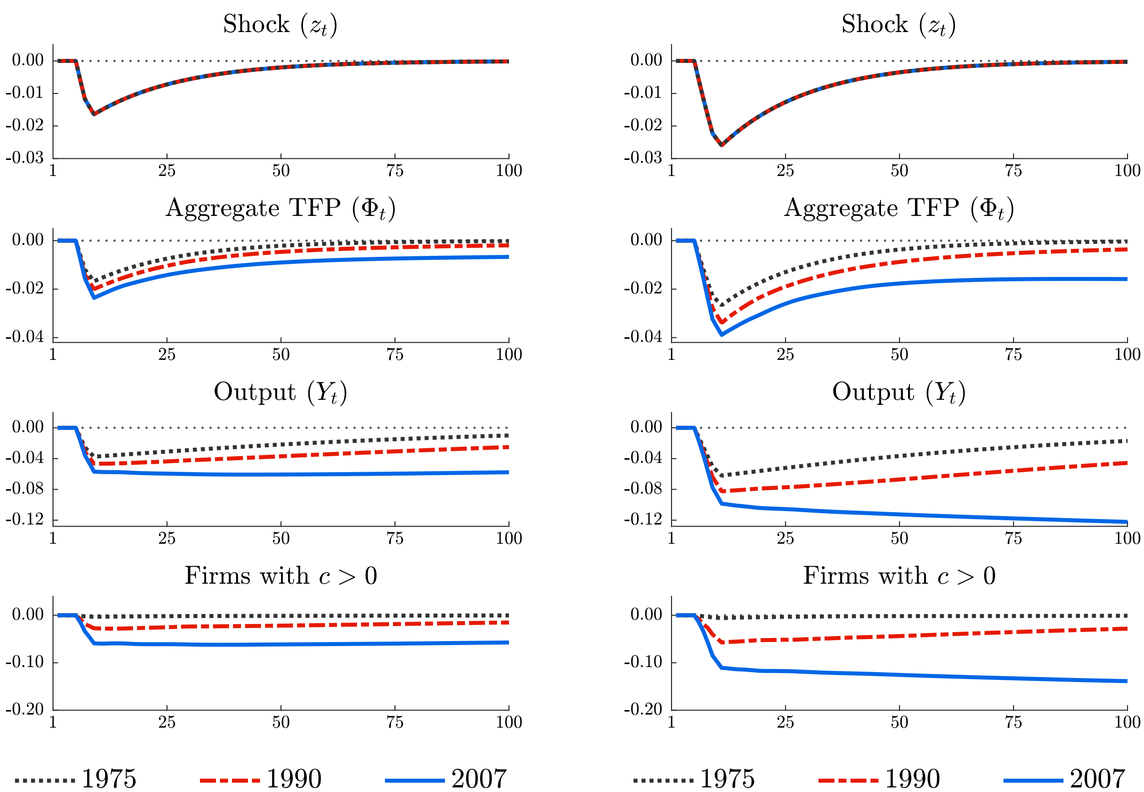
Table 6: Simulated distribution of $\log(Y_t)$

This table shows moments of the simulated distribution of $\log(Y_t)$. Each economy has been simulated for 10,000,000 periods. $\overline{\log Y} - \log Y^H$ is the average log deviation of output from the high steady state output. Lastly, $\sigma_{\log Y}$ is the standard deviation of log output.

Impulse Response Functions: Small Negative Shock We characterize the reaction of the economy to a small negative shock. We consider a shock to the innovation of the exogenous TFP process that is equal to $\varepsilon_t = -\sigma_\varepsilon$ and lasts for four quarters. Figure 8(a) shows the impulse responses for 1975, 1990 and the 2007 economy. The simulation of the transition dynamics covers 100 quarters. This shock generates different responses for the three economies. The 2007 economy exhibits both greater amplification and persistence. First, the 1975 economy experiences a 3.8% reduction in aggregate output after 5 quarters, against a 4.7% in 1990 and a 5.7% reduction in the 2007 economy. Second, after 100 quarters, the 1975 economy is 1.0% below the steady-state, while the 1990 economy is 2.5% below trend and the 2007 economy has a much more prolonged downturn, being still 5.6% below pre-crisis output. Eventually they converge back to the initial level.

The mechanism underlying such increased amplification and persistence can be understood by looking at the bottom panel, which plots the transition dynamics of the number of firms in *concentrated* product markets. In 2007, there is a much more significant reduction in the number of firms, due to the mechanisms outlined above: increased productivity dispersion and larger fixed costs make small, unproductive firms more sensitive to aggregate shocks. This additional action in the extensive margins generates both additional amplification and persistence. Quantitatively we have that on impact, in 1975, about 0.35% of firms in concentrated product markets exit, while this number is 5.9% for 2007. The slow net entry as the economy goes back to the original steady state drives the large persistence of the contraction.

Impulse Response Functions: Large Negative Shock The shock introduced above was small enough to make all three economies converge back to their initial steady-states, albeit in very different time horizons. We now study the effect of a larger shock. We repeat the same exercise for the three economies, but now introduce a negative shock $\varepsilon_t = -2\sigma_\varepsilon$, which lasts for six quarters. The dynamics are shown in Figure 8(b). As before, there is greater amplification and persistence in the 2007 economy. However, this economy now experiences a permanent drop in aggregate output, i.e. it transitions to a lower steady-state. In the example we consider, after 100 quarters, output is 12.2% below its initial value. Note, however, that the gap is still widening at the end of the sample. This is due to a permanent loss of 13.9% of active firms in concentrated product markets. The 1975 and 1990 economies, on the other hand, converge back to the pre-crisis levels of both firms and output.



(a) Small Negative Shock

(b) Large Negative Shock

Figure 8: Impulse Response Functions

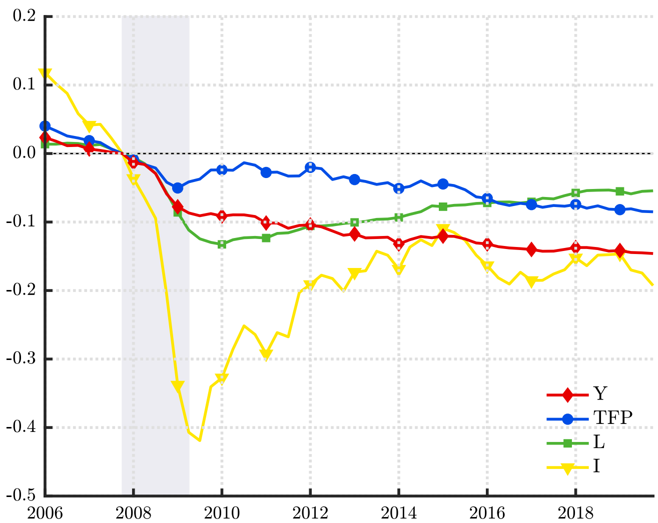
Note: The graphs show the IRFs to an exogenous TFP shock. In panel (a), we feed a shock $\varepsilon_t = -\sigma_\varepsilon$ that lasts for four quarters. In panel (b), we feed $\varepsilon_t = -2\,\sigma_\varepsilon$ during six quarters.

These results suggest that rising firm differences and fixed costs are a source non-linearity in the economy's response to aggregate shocks. This may appear to be inconsistent with the idea of a *great moderation* — namely, the fact that the volatility of aggregate output declined between 1980 and 2007. Note, however, that aggregate volatility in our economy is the product of two forces — exogenous volatility (TFP shocks) and endogenous amplification and persistence. If exogenous volatility declined over time, it is possible that aggregate volatility also declined in spite of larger amplification. There are reasons to think that exogenous aggregate volatility may have decreased over time – for example, because of demographic shifts (Jaimovich and Siu, 2009) or a rising share of low-volatility industries (Carvalho and Gabaix, 2013).

We next use our model as a laboratory to study the 2008 recession.

5 The 2008 Recession and Its Aftermath

In this section, we take a closer look at the 2008 recession and its aftermath. The left panel of Figure 9 shows the behavior of four aggregate variables from 2006 to 2019 — real GDP, real gross private investment and total hours (all in per capita terms), as well as aggregate TFP. All variables are in logs, detrended (with a linear trend computed over 1985–2007) and centered on 2007Q4. The four variables decline on impact and do not rebound to their pre-recession trends. For example, in the first quarter of 2019, real GDP per capita is 14.2% below trend (Table 7). Aggregate TFP experienced a 8.2% negative deviation from trend. Investment declines by more than 40% on impact, and then stabilizes at approximately 15-20% below the pre-crisis trend.



(a) Data (b) 2007 Model

Figure 9: The great recession and its aftermath

Note: The figure shows the evolution of key macro aggregates in the aftermath of the 2008 recession in the data (Panel a) and the model (Panel b). The model is subjected to a sequence of six quarter shocks $\{\varepsilon_t\}$ to match the dynamics of aggregate TFP in the data between 2008Q1:2009Q2. See Appendix A.1 for data definitions.

We then ask whether our model can replicate the behavior of these four variables. To this end, we feed the model a sequence of shocks ε_t (i.e. the innovation of A_t) that lasts for six quarters (2008Q1:2009Q2); these shocks are calibrated so that aggregate measured TFP in our model (i.e. $A_t\,\Phi_t$) matches the dynamics of its data counterpart over the same period. The economy starts at the high steady-state (with $z_t = 0$). We set the innovations to productivity to zero after 2009Q2 and let the economy recover. The right panel of Figure 9 shows the implied responses of output, aggregate TFP, employment and investment, generated by our model. As the figure shows, this series of shocks triggers a transition to the low steady-state. Our model provides a very good description of the evolution of the four variables. Output experiences a decline of 12.3% after 10

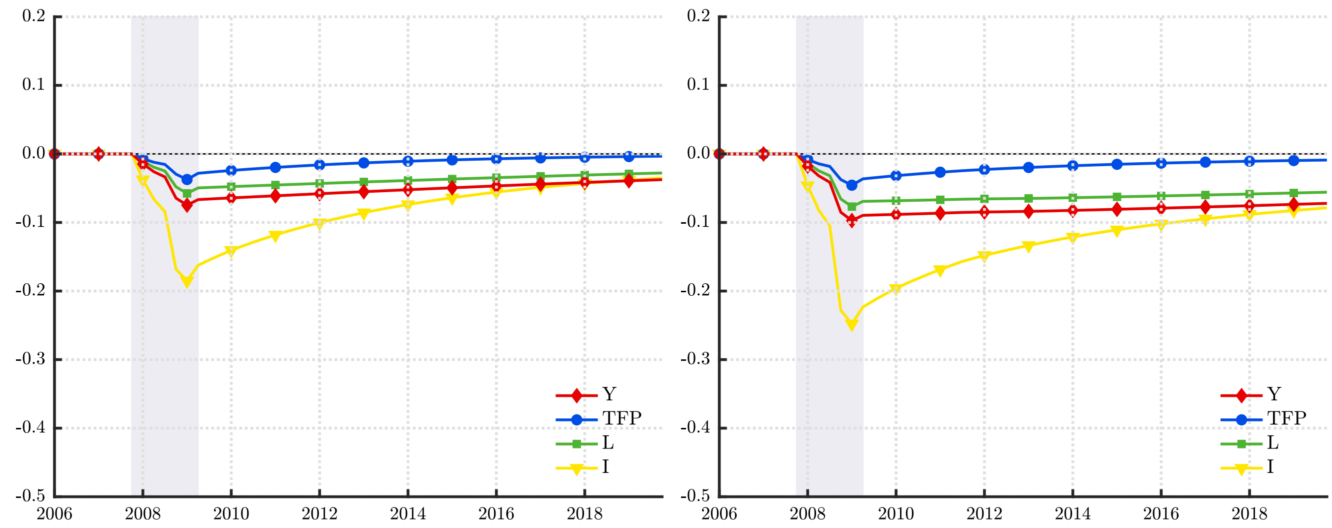
	Data			Model			
	2009Q4	2015Q1	2019Q1	2009Q4	2015Q1	2019Q1	2040Q1
Output	-0.088	-0.126	-0.142	-0.107	-0.118	-0.123	-0.142
TFP	-0.045	-0.026	-0.082	-0.039	-0.023	-0.019	-0.018
Hours	-0.130	-0.078	-0.055	-0.086	-0.094	-0.098	-0.113
Investment	-0.340	-0.109	-0.147	-0.247	-0.169	-0.151	-0.157

Table 7: The great recession and its aftermath

years, whereas hours drop by 9.8% (Table 7). Both reactions are of the same order of magnitude as observed in the data.[30] Additionally, our model predicts a 24.7% decline on impact for investment, and a 15.1% drop by 2019 (14.7% in the data). Finally, we observe a long-run decline of aggregate TFP, representing approximately 1/5 of the drop in the data (1.6% in the model, 8.2% in the data). We discuss the mechanisms underlying this result in the next section. The crisis experiment in our model can also replicate the still widening gap between output and its trend throughout our sample. Importantly, the model economy after 30 years shows a permanent deviation from the pre-crisis trend (last column of Table 7). Output is 14.2% lower while hours and investment are 11.3 and 15.7% below trend. The economy therefore features a change in steady-state to one with lower output and competition.

Next, as our first counterfactual, we ask whether the same sequence of aggregate TFP shocks used in the 2007 economy can also trigger a transition to the low steady-state in the 1975 and 1990 economies. We study this experiment to ask whether the deviation our model predicts for the 2007 economy is driven by an unusually large shock or by inherent fragility of the economy itself. Figure 10 and Table 8 show the transitional dynamics. Both economies exhibit substantially less amplification, and, importantly, they also revert to their pre-crisis steady-states. In economies with the 1975 and 1990 features, a negative aggregate shock of the magnitude required in our model to generate the 2008-2009 recession would not be large enough to induce a persistent deviation from trend. These economies would have experienced a faster reversal to trend. For example, as of 2019, output would be 3.8% and 7.2% below trend (in the 1975 and 1990 economies, respectively), against 12.3% of the 2007 model and 14.2% as observed in the data. When we evaluate the longer run behavior in 2040, we find that the 1975 economy would experience a 1% drop, while for the 1990 economy this figure is 3.7%. We conclude that the structural differences between the 1975,

[30] In the data, hours worked seem to recover faster than output, which seems inconsistent with a jobless recovery. Note, however, that i) we are showing variables in deviation from trend and that ii) hours worked were characterized by slower growth (and hence a flatter trend) before the crisis. For example, total hours worked were stagnant between 2000-2007.



(a) 1975 Model (b) 1990 Model

Figure 10: The *great recession* in the 1975 and 1990 models

Note: This figure shows the response of the 1975 and 1990 economies to the sequence of shocks used in Figure 9(b)

	1975 Model			1990 Model		
	2009Q4	2019Q1	2040Q1	2009Q4	2019Q1	2040Q1
Output	-0.065	-0.038	-0.010	-0.089	-0.072	-0.037
TFP	-0.025	-0.003	-0.000	-0.033	-0.009	-0.003
Hours	-0.048	-0.028	-0.007	-0.069	-0.056	-0.029
Investment	-0.148	-0.035	-0.006	-0.205	-0.079	-0.032

Table 8: The great recession and its aftermath

1990 and the 2007 economies (namely larger productivity differences and higher fixed costs) are key to understanding the 2008 crisis and the subsequent *great deviation*. Next we provide empirical evidence consistent with our proposed mechanism both at the aggregate and industry level.

5.1 Empirical Evidence

According to our model, a transition between steady-states is driven by a change in the competitive regime of the economy. We now provide evidence consistent with our model's mechanism. In particular, we show that after 2008 the US economy experienced i) a persistent decline in the number of active firms, ii) a persistent decline in the aggregate labor share and iii) acceleration in the aggregate profit share and markup. Furthermore, the model has stark predictions about how product markets starting with different levels of concentration should feature different responses to the negative shock. The extensive margin is more elastic in product markets with higher concentration to begin with; as a consequence, we should observe that these product markets experience a larger drop in output. We provide cross-industry evidence corroborating this prediction.

5.1.1 Aggregate Level Evidence

We begin by reviewing the evidence at the aggregate level on the number of firms in the economy, the behavior of markups, labor and profit shares and the dynamics of aggregate TFP.

Number of Firms Firms in our model are single-product and operate in only one market. This contrasts with the definition of firms in the data, which are often multi-product and operate in several markets (e.g. geographic segmentation). As a consequence, studying the dynamics of the number of firms probably provides a lower bound on the effects through the extensive margin as some firms might remain in market but remove products or close establishments. With this caveat in mind, we ask whether the evolution of the number of firms in the data qualitatively matches the prediction of our model. Figure 11 shows the evolution of the number of active firms (with at least one employee). As the figure shows, the number of active firms experiences a persistent deviation from trend after 2007. As of 2016, the number of active firms was 0.151 log points below trend. Such a persistent decline can also be observed within most sectors of activity (Section B.7). With the caveat that a firm in our model do not necessarily represents a firm in the data, we report the evolution of the number of firms in the *concentrated* sector. As shown in Table 9, this sector experiences a persistent decline in the number of firms of 0.134 log points.

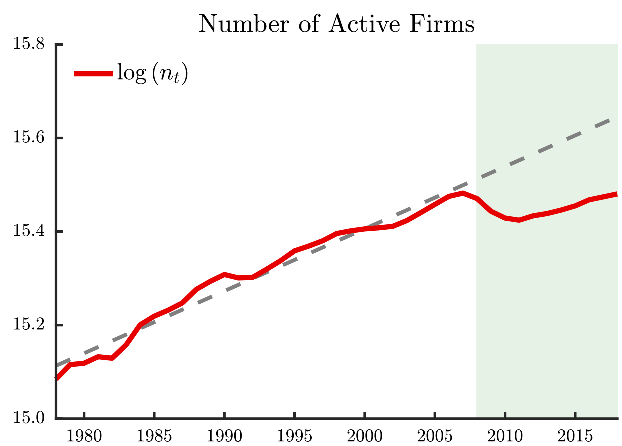


Figure 11: **Number of Firms: 1978-2018**

Note: The red line shows the number of firms with at least one employee (in logs). The dashed grey line shows a linear trend computed for the period 1978-2007. Data is from the US Business Dynamics Statistics.

We think that alternative models could replicate this figure only if they featured an active extensive margin, multiple steady states and predicted a transition during the 2007-08 recession. We argue that these features and the changes in the market structure, which we discuss next, corroborate our proposed mechanism.

Aggregate Markups, Labor and Profit Shares We now ask how our model compares to the data regarding the evolution of labor and profit shares after 2008. Figure 12 shows the evolution of the labor share, the profit share (both computed for the US business sector) and the aggregate markup series for publicly listed firms from De Loecker et al. (2020). The grey dashed lines represent linear trends computed for the period 1975–2007.

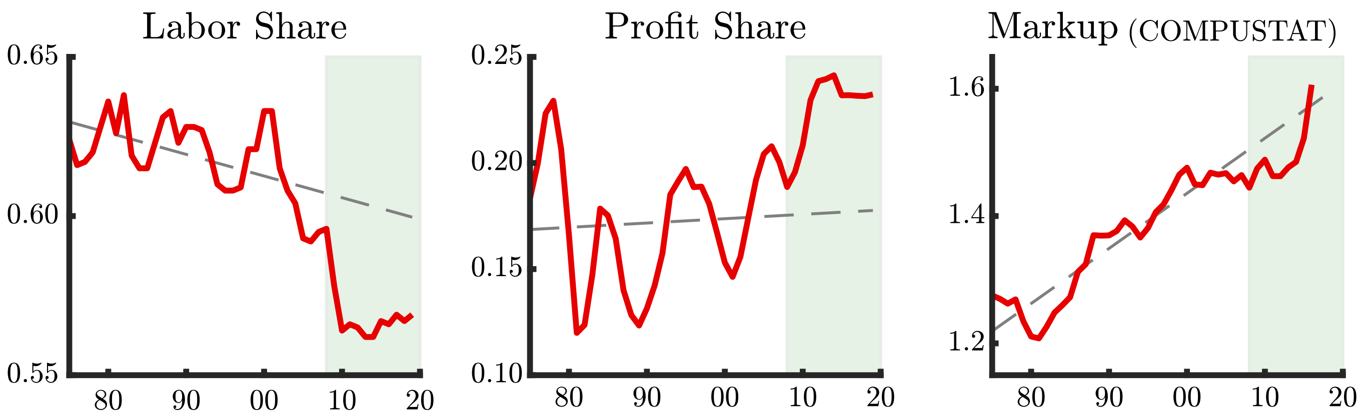


Figure 12: **Aggregate Markup and Labor and Profit Share: 1975-2019**

Note: This figure shows (i) labor share of the US business sector (from the BLS), (ii) the profit share of the US business sector (defined in Appendix A.1) and (iii) the aggregate markup series for COMPUSTAT firms from De Loecker et al. (2020). The dashed grey lines represent linear trends computed for the period 1975–2007.

Table 9 compares the evolution between 2007 and 2016 observed in the data and obtained in our model.[31] Our model predicts a 0.6pp decline in the aggregate labor share, approximately 17% of the observed decline between 2007 and 2016. If we account for a pre-crisis trend, we explain 22% of the deviation in 2016. For the profit share, we can explain 30% of its 3pp increase from trend. Markups increase by 4.1 points in our model, which represents 29% of the observed increase (14.2 points) and 64% of the deviation from the pre-crisis trend (6.4 points).[32]

Aggregate Productivity As shown in Figure 9, the post-2008 growth experience has also been characterized by a persistent decline of aggregate TFP from trend. Our model predicts a persistent

[31]The aggregate profit share in our model is net of fixed production costs.

[32]Other studies have also documented a sharp rise in markups in the post-crisis years. For example, Dopper et al. (2022) use US product-level data and document a 25% increase in markups between 2006 and 2019.

	Model	Data	
	$\Delta_{2007-2016}$	$\Delta_{2007-2016}$	$\Delta_{2007-2016} - \Delta_{\text{trend}}$
(log) Number of Firms (*concentrated* product markets)	-0.134	-0.014	-0.151
Labor Share	-0.005	-0.029	-0.023
Profit Share	0.009	0.032	0.030
Aggregate Markup	4.1	14.2	6.4

Table 9: Change in the number of firms and in income shares: model *versus* data

fall in aggregate TFP (though of smaller magnitude than the one observed in the data). In our model, this happens in spite of the exit of unproductive firms, which results in higher average firm-level TFP, the "cleansing effect of recessions" (see Figure B.3). There are two reasons explaining the decline in aggregate TFP. First, there is a *love-for-variety* effect, associated with the reduction in the number of firms. This can best be seen in the limit case in which all product markets have n firms with identical productivity τ. In such a case, aggregate TFP is equal to $\Phi = I^{\frac{1-\rho}{\rho}} n^{\frac{1-\eta}{\eta}} \tau$. Second, in a low competition trap, there is higher cross product market misallocation. This happens because markets with a larger contraction are the ones with positive fixed costs $c > 0$, i.e. whose output is already low.[33]

The macro trends discussed above suggest that, consistent with our model, market power accelerated after 2008. Alternative models would be able to generate these empirical pattern if they featured cyclical changes in the level of competition, so that, over the business cycle profit and factor shares would move. We next review the cross-sectional implications of our model and test them in the data.

5.1.2 Industry Level Evidence

According to our model, product markets featuring a larger concentration in 2007 should have experienced a larger contraction in 2008. This prediction follows from equation (11), which establishes a positive link between productivity, market shares and markups (for a given number of active firms). Therefore, if we take two markets with the same number of firms, the one featuring a more uneven distribution of productivities will have a larger dispersion in market shares and hence a larger concentration. In these markets, firms at the bottom of the distribution will be smaller and charge lower markups, and will hence be more likely to exit upon a negative shock. This prediction holds for a given number of firms n_{it}, so that, when measuring the correlation between concentration

[33] Figure B.4 (Section B.5) shows that the standard deviation of outputs, $\text{std}_i(\log y_{it})$, increases.

in 2007 and the size of the contraction in 2008, we must control for the number of firms in the industry.

We build a dataset combining the 2002 and 2007 US Census data on industry concentration to the Statistics of US Businesses (SUSB) and the Bureau of Labor Statistics (BLS) to obtain outcomes as employment, total wage bill and the number of firms at the industry level (6-digits NAICS). The final dataset includes 791 6-digit industries. In 2016, the median industry had 1,316 firms, 36,910 workers and a total payroll of $1,880 million.

To assess whether industries with a larger concentration before the crisis experienced a larger post-crisis decline, we estimate the following regression

$$\frac{\Delta y_{i,07-16}}{y_{i,07}} = \beta_0 + \beta_1 \text{ concent}_{i,07} + \beta_2 \text{ log firms}_{i,07} + \beta_3 \frac{\Delta y_{i,03-07}}{y_{i,03}} + a_s \mathbb{1}\{i \in s\} + u_i.$$

y_i is an outcome for industry i (for example total employment, total wage bill or total number of firms) and $concent_i$ is the share of the 4 largest firms (scaled by the share of the largest 50); we also control for the number of firms before the crisis ($\text{firms}_{i,07}$). Importantly, to control for possible differences in the pre-crisis dynamics, e.g. different growth opportunities in different industries, we include pre-trends in the regression. The outcomes always take the form of the annualized growth rate between 2007 and 2016 in a specific industry. In all regressions, we will also include sector fixed effects (a_s). The unit of observation is a 6-digit industry.

We start by studying the correlation between the change in employment between 2008 and 2016 and concentration in 2007. The results are compactly presented in Table 10 and robustness are shown in Appendix A.4. We find that more concentrated industries experienced lower employment growth in the aftermath of the great recession. Quantitatively, a 1pp higher pre-crisis concentration is associated with a 2pp lower employment growth rate between 2007 and 2016. This pattern holds irrespective of the inclusion of the number of firms in 2007. To address the concern that industries with larger concentration in 2007 could have already exhibited lower growth before the crisis, we include cumulative employment growth between 2003 and 2007 as a control; the results do not change. Finally, the results are also robust to the inclusion of sector fixed effects. While these results concern the evolution of employment growth, a similar pattern is found if we use the total wage bill instead . We also study the correlation between concentration and net entry after the crisis . Our findings suggest that a 1pp increase in the concentration measure is associated with a 2-3pp decrease in the post-crisis net entry. These results suggest that industries with larger concentration in 2007 experienced a larger contraction in activity after the crisis.

We conclude this section by providing evidence on the evolution of the labor share across industries. While the US census of firms provides data on total employment and the total number of firms for all 6-digit industries, it does not contain data on the labor share. We rely on data from the BLS 'Labor Productivity and Cost' programme (see Appendix A.1 for details). This database, however, only provides data on the labor share for a restricted group of industries. We

VARIABLES	(1) $\Delta \log \text{emp}_{07-16}$	(2) $\Delta \log \text{payroll}_{07-16}$	(3) $\Delta \log \text{firms}_{07-16}$	(4) $\Delta\text{labor share}_{07-16}$
concent_{07}	-0.0177***	-0.0189***	-0.0406***	-0.0314*
	(0.00682)	(0.00697)	(0.00635)	(0.0167)
$\log \text{firms}_{07}$	0.00193***	0.00164**	0.00119*	-0.00120
	(0.000706)	(0.000725)	(0.000661)	(0.00240)
$\Delta \log \text{y}_{03-07}$	0.0984***	0.0823***	0.0881***	0.169*
	(0.0241)	(0.0219)	(0.0270)	(0.0867)
Observations	769	773	791	98
R-squared	0.050	0.043	0.078	0.075

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Cross-industry regressions

Note: the table shows the results of regressing the growth rate of sectoral employment, payroll, number of firms and labor share between 2007 and 2016 on the measure of concentration in 2007 and controls. Robustness on these regressions are displayed in Appendix A.4.

find a negative relationship between the post-crisis change in the labor share and the pre-crisis level of concentration. Industries with larger concentration in 2007 experienced a larger drop in labor share between 2008 and 2016.

In summary, these results suggest that the structure of US product markets in 2007 is important to understand the consequences of the 2008 crisis. The results presented are purely cross-sectional — industries with a larger concentration in 2007, displayed a larger post-crisis contraction. We think, however, that they support one of the main insights of the model — rising concentration can have made the US economy more vulnerable to aggregate shocks.

Taken together we view the empirical evidence reported here as corroborating our proposed mechanism. The aggregate economy shows a long-lasting deviation and a widening gap from the pre-crisis trend and a change in the distribution of incomes from factor suppliers to firms, suggestive of an increase in market power. At the industry level we observe that industries respond differentially depending on the pre-existing market structure. Our model shows that the latter observation and its implications for the distribution of rents in the economy can explain the quasi-permanent deviation at the macro level.

5.2 Robustness Checks

In the previous section we described how a calibrated version of the model admits multiple stochastic steady states and performs well when tasked with replicating the behavior of the US economy in the aftermath of the 2008 recession. In this section we study two robustness checks on our model. First we subject the same economies to the aggregate shocks needed to match the 1990 recession. Next we check whether our modelling choices on the structure of the fixed cost, which is expressed in terms of the *numeraire* output good, matter for our results.

The 1990 Recession Through the lens of our model, the 2008 crisis made the US economy transition to a new steady-state. This fact has not been observed after any other postwar recession. This raises a natural question: what was special about the 2008 crisis? Was the shock hitting the economy in 2008 larger than in previous recessions? Or was the economy more fragile in 2008 and therefore more prone to experience a transition even for moderate shocks? Earlier in this section we showed that the same shocks underlying the 2008 recession in our model do not trigger transitions in the 1975 and 1990 economies. We show that this holds for other recessions. We repeat the experiment of Section 5 using the 1990 crisis. We feed the 1990 economy a sequence of shocks that replicates the dynamics of aggregate TFP during the 1990–1991 recession (1990Q3:1991Q1). We then take this same sequence of exogenous shocks and apply them to the 2007 economy. The results of this experiment are shown in the Supplementary Material Section B.6. When looking at the response of the 1990 economy, we observe a temporary decline in all variables, but followed by a gradual recovery to the previous steady-state. This contrasts with the response of the 2007 economy, which eventually experiences a transition to the lower steady-state. These results suggest that, rather than the consequence of an unusually large shock, the post-2008 deviation can be linked to an underlying market structure that made the US economy more fragile to negative shocks.

Variable fixed costs In the model presented so far firms pay a fixed cost $c > 0$ in units of the final good. This assumption implies that the cost of entry is independent of the state of the economy and, hence, of its competition regime and of factor prices. If fixed costs were to change with factor prices, entry could be cheaper (more expensive) in a low (high) competition regime, which could in principle eliminate steady-state multiplicity. To address this concern, we let firms hire labor and capital to pay the fixed cost. In particular we assume production entails a fixed cost $c = k_c^{\alpha} \, l_c^{1-\alpha}$. Under this assumption, firms pay an effective fixed cost $\Theta_t \cdot c$. Since the factor cost index Θ_t is increasing in the number of active firms, entry becomes more expensive as firms enter. Appendix B.9 shows the results for this version of the model. The existence of two competition regimes is preserved under this alternative assumption (for the 1990 and 2007 calibrations). We also obtain responses for the crisis experiment performed above and the results are qualitatively and quantitatively similar to the benchmark model.

6 Policy Experiment

We conclude by studying the role of policy in our model. In this economy, in addition to the standard (static) inefficiencies associated with markups and markup dispersion, market power carries additional negative consequences as it can trap the economy in a low competition regime.

We consider a government which grants an entry subsidy equal to a fraction $\tau_f \in [0,1]$ of the fixed cost, while levying a tax $\tau_\pi \in [0,1]$ on net profits $\tilde{\Pi}_{ijt} := \Pi_{ijt} - (1-\tau_f)\, c_i$ to balance the budget. First, note that by design the entry subsidy affects only *concentrated* product markets. Secondly, since the government is taxing only the amount of profits in excess of fixed costs, the tax $\tau_\pi > 0$ does not create a disincentive to entry. It also does not distort firm size conditional on the market structure, since firms' optimal size is given by the zero marginal profit condition $\pi' = 0 = (1-\tau_\pi)\pi'$. Finally, a positive entry subsidy will reduce the entry productivity thresholds. The planner faces, however, one key trade-off when subsidizing entry: while having more firms in the economy reduces markups and increases available varieties, it also implies that more resources are spent in fixed costs and that less productive firms enter the market.

We report welfare calculations for different levels of the entry subsidy in Figure 13 for our three economies. The analysis suggests that, in the 2007 economy, the government would find a subsidy of around 70% of the fixed cost optimal, implying a welfare gain of approximately 21% (in consumption equivalent terms). This value is within the range of estimates reported by the literature for the cost of markups. For example, Bilbiie et al. (2019) find a cost as high as 25%, while Edmond et al. (2021) find a welfare loss of 23.6% for an average cost-weighted markup of 1.25. For the 1990 economy the welfare effect of the policy takes a similar shape but the impact peaks at around 12%, while for 1975 the maximum welfare improvement is about 2% CEV.

It should be noted that the policy experiment we consider does not implement the first best allocation. We evaluate a simple firm subsidy and do not consider size-dependent taxes/subsidies that might be necessary to eliminate markup distortions. Furthermore, as mentioned earlier, we highlight another cost of market power: the fact that it can generate quasi-permanent recessions. Interestingly, Figure 13 suggests that welfare is very steep around $\tau_f = 0$, implying that relatively small entry subsidies can have a sizeable impact on welfare. For example, a 10% entry subsidy would be enough to generate a 8% gain in consumption equivalent terms. The reason is that even a relatively small subsidy can significantly shift the probability mass from the low to the high steady state. This intuition explains why the effect is partially and fully muted in the 1990 and 1975 economies. As the likelihood of moving to a low competition regime reduces, the welfare gain from the policy declines. In the 1975 economy, where the probability of quasi-permanent recessions is zero to begin with, the welfare effects are solely driven by trading off lower markups and more varieties with the more resources absorbed by fixed costs. Interestingly, the three economies also differ in terms of the welfare costs associated to an entry tax. In particular, such tax can have two complementary effects: i) it can force the economy in the low competition regime and ii) it can

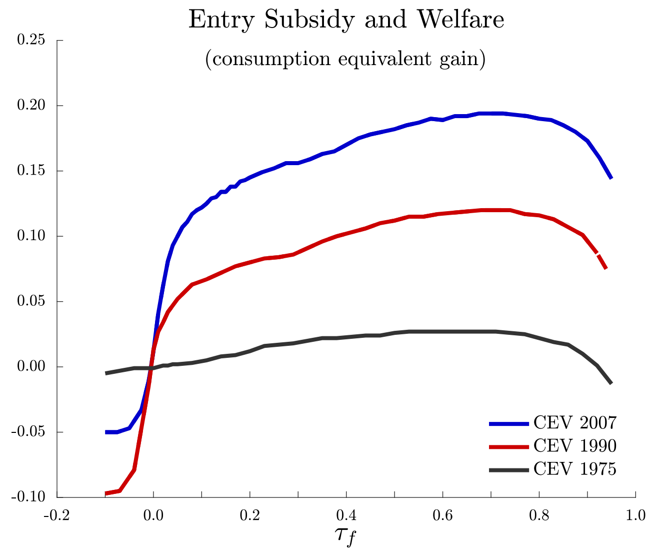


Figure 13: Welfare: consumption equivalent gain

Note: the figure shows the welfare impact (in consumption equivalent gains) of an entry subsidy equal to a fraction τ_f of fixed costs. For each level of τ_f, we simulate the economy 100,000 times and calculate average welfare.

worsen the welfare in the low steady state as fewer firms are able to survive. This rationale explains the large costs for the 1990 economy. Recall that in this calibration quasi-permanent recessions are unlikely but feature large output losses relative to the high steady state. These welfare costs are smaller for the 2007 economy as downward transitions entail smaller output losses and are more likely even before the tax.

Lastly, one can think about the optimal state-dependent subsidy. If the economy is hit by a large negative shock that triggers a steady state transition, the welfare benefit of such a subsidy can be very large. On the other hand, during a recession, profits are reduced, thereby making the budget constraint tighter. If the government could borrow intertemporally it would have large incentives to do so and to finance entry during downturns and pay back debt during booms. This suggests that, through the lens of our model, countercyclical firm subsidies can alleviate downturns by preventing the economy from falling into quasi-permanent recessions.

7 Conclusion

The US economy appears to have experienced a fundamental change over the past decades, with several studies and data sources indicating a reallocation of activity towards large, high markup firms. This observation has raised concerns in academic and policy circles about increasing market power, and it has been proposed as an explanation for recent macroeconomic *puzzles*, such as low aggregate investment, low wage growth or declining labor shares. Our model suggests that, besides their impact on factor shares and factor prices, that rising firm differences and greater market power can also have an impact on business cycles and provide an amplification and persistence mechanism to aggregate fluctuations. In particular, larger firm heterogeneity may have rendered the US economy more vulnerable to aggregate shocks and more likely to experience quasi-permanent slumps. Through the lens of our theory, such increased fragility may have been difficult to identify, as it manifests itself only in reaction to large shocks.

Our work also provides further motive for policies that curb market power. As we have shown, the endogenous response of the market structure to aggregate shocks act as an accelerant. On top of the standard static effects, any policy that reduces market power can have dynamic benefits in terms of the persistence and amplitude of aggregate fluctuations. These effects are particularly large if the economy is at risk of quasi-permanent slumps.

To keep the analysis simple, we have abstracted from a number of important features. One such example is that, in our model, firms solve a static problem and their productivity is permanent. The interaction between endogenous growth through innovation, market power and fragility is a natural next step for this area of research. Further, our model features one-sided market power. Recent models of oligopoly (see Azar and Vives, 2021) lend themselves to the study of the interaction between two-sided market power and the likelihood of quasi-permanent slumps. Lastly, recent work studies the interplay between competition and monetary policy (see Mongey, 2019; Wang and Werning, 2020; Fabiani et al., 2021). The question of how monetary policy, by changing the market structure, shapes the dynamic properties of an economy is an avenue for future research.

References

Akcigit, Ufuk and Sina T. Ates, "What Happened to U.S. Business Dynamism?," NBER Working Paper 25756 2019.

Andrews, Dan, Chiara Criscuolo, and Peter N. Gal, "Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries," OECD Productivity Working Papers 2015.

Atkeson, Andrew and Ariel Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, December 2008, *98* (5), 1998–2031.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *The Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

Azar, José and Xavier Vives, "General Equilibrium Oligopoly and Ownership Structure," *Econometrica*, 2021, *89* (3), 999–1048.

Baqaee, David Rezza and Emmanuel Farhi, "Productivity and misallocation in general equilibrium," *The Quarterly Journal of Economics*, 2020, *135* (1), 105–163.

Benedetti-Fasil, Cristiana, G Impullitti, O Licandro, and P Sedláček, "Heterogeneous firms, R&D policies and the long shadow of business cycles," 2021.

Benhabib, Jess and Roger E. A. Farmer, "Indeterminacy and Increasing Returns," *Journal of Economic Theory*, June 1994, *63* (1), 19–41.

Benigno, Gianluca and Luca Fornaro, "Stagnation traps," *The Review of Economic Studies*, 2017, *85* (3), 1425–1470.

Benkard, C. Lanier, Ali Yurukoglu, and Anthony Lee Zhang, "Concentration in Product Markets," *Working Paper*, 2021.

Bilbiie, Florin O., Fabio Ghironi, and Marc J. Melitz, "Endogenous Entry, Product Variety, and Business Cycles," *Journal of Political Economy*, 2012, *120* (2), 304–345.

_ , _ , and _ , "Monopoly Power and Endogenous Product Variety: Distortions and Remedies," *American Economic Journal: Macroeconomics*, October 2019, *11* (4), 140–174.

Bils, Mark, Peter J. Klenow, and Benjamin A. Malin, "Resurrecting the Role of the Product Market Wedge in Recessions," *American Economic Review*, April 2018, *108* (4-5), 1118–1146.

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

Bonfiglioli, Alessandra, Rosario Crinò, and Gino Gancia, "Betting on Exports: Trade and Endogenous Heterogeneity," *Economic Journal*, March 2018, *128* (609), 612–651.

Buera, Francisco J., Hugo Hopenhayn, Yongseok Shin, and Nicholas Trachter, "Big Push in Distorted Economies," NBER Working Paper 28561 2021.

Burstein, Ariel, Basile Grassi, and Vasco Carvalho, "Bottom-Up Markup Fluctuations," Working Paper 2020.

Calligaris, Sara, Chiara Criscuolo, and Luca Marcolin, "Mark-ups in the digital era," OECD Science, Technology and Industry Working Papers 2018.

Carvalho, Vasco and Xavier Gabaix, "The Great Diversification and Its Undoing," *American Economic Review*, August 2013, *103* (5), 1697–1727.

Chatterjee, Satyajit, Russell Cooper, and B Ravikumar, "Strategic complementarity in business formation: Aggregate fluctuations and sunspot equilibria," *The Review of Economic Studies*, 1993, *60* (4), 795–811.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber, "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins," *American Economic Review*, 2011, *101* (3), 471–475.

Cooper, Russell and Andrew John, "Coordinating Coordination Failures in Keynesian Models," *The Quarterly Journal of Economics*, 1988, *103* (3), 441–463.

_ **and** _ , "Imperfect competition and macroeconomics : Theory and quantitative implications," *Cahiers d'Économie Politique*, 2000, *37* (1), 289–328.

Decker, Ryan A, John C Haltiwanger, Ron S Jarmin, and Javier Miranda, "Changing business dynamism and productivity: Shocks vs. responsiveness," Technical Report, National Bureau of Economic Research 2018.

Díez, Federico J, Daniel Leigh, and Suchanan Tambunlertchai, "Global Market Power and its Macroeconomic Implications," IMF Working Papers, International Monetary Fund June 2018.

Dopper, Hendrik, Alexander MacKay, Nathan Miller, and Joel Stiebale, "Rising Markups and the Role of Consumer Preferences," *Working Paper*, 2022.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "How costly are markups?," Technical Report 2021.

Eeckhout, Jan, "Comment," *NBER Macroeconomics Annual*, 2021, *35* (1), 151–166.

Eggertsson, Gauti B, Jacob A Robbins, and Ella Getz Wold, "Kaldor and Piketty's facts: The rise of monopoly power in the united states," NBER Working Paper 24287 2018.

Eggertsson, Gauti B., Neil R. Mehrotra, and Jacob A. Robbins, "A Model of Secular Stagnation: Theory and Quantitative Evaluation," *American Economic Journal: Macroeconomics*, January 2019, *11* (1), 1–48.

Etro, Federico and Andrea Colciago, "Endogenous Market Structures and the Business Cycle," *Economic Journal*, 2010, *120* (549), 1201–1233.

Fabiani, Andrea, Luigi Falasconi, and Janko Heineken, "Monetary Policy and Corporate Debt Maturity," Technical Report 2021.

Farmer, Roger and Jang-Ting Guo, "Real Business Cycles and the Animal Spirits Hypothesis," *Journal of Economic Theory*, June 1994, *63* (1), 42–72.

Fernald, John G., “A quarterly, utilization-adjusted series on total factor productivity,” Working Paper Series 2012-19, Federal Reserve Bank of San Francisco 2012.

Galí, Jordi and Fabrizio Zilibotti, “Endogenous Growth and Poverty Traps in a Cournotian Model,” *Annals of Economics and Statistics*, 1995, *37-38*, 197–213.

Galí, Jordi, Frank Smets, and Rafael Wouters, “Slow recoveries: A structural interpretation,” *Journal of Money, Credit and Banking*, 2012, *44*, 9–30.

Gamber, William, “Entry and Employment Dynamics in the Presence of Market Power,” *Finance and Economics Discussion Series 2021-077, Board of Governors of the Federal Reserve System*, 2021.

Gorodnichenko, Yuriy and Michael Weber, “Are Sticky Prices Costly? Evidence from the Stock Market,” *American Economic Review*, January 2016, *106* (1), 165–199.

Grassi, Basile, “IO in I-O: Size, industrial organization, and the input-output network make a firm structurally important,” *Work. Pap., Bocconi Univ., Milan, Italy*, 2017.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W Huffman, “Investment, Capacity Utilization, and the Real Business Cycle,” *American Economic Review*, June 1988, *78* (3), 402–417.

Guerrieri, Veronica and Guido Lorenzoni, “Credit Crises, Precautionary Savings, and the Liquidity Trap,” *The Quarterly Journal of Economics*, 2017, *132* (3), 1427–1467.

Hall, Robert, “The Long Slump,” *American Economic Review*, 2011, *101* (2), 431–469.

_ , “Using empirical marginal cost to measure market power in the us economy,” Technical Report, National Bureau of Economic Research 2018.

Herrendorf, Berthold, Akos Valentinyi, and Robert Waldmann, “Ruling out multiplicity and indeterminacy: the role of heterogeneity,” *The Review of Economic Studies*, 2000, *67* (2), 295–307.

Jaimovich, Nir, “Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations,” *Journal of Economic Theory*, November 2007, *137* (1), 300–325.

_ **and Henry E. Siu**, “The Young, the Old, and the Restless: Demographics and Business Cycle Volatility,” *American Economic Review*, June 2009, *99* (3), 804–826.

_ **and Max Floetotto**, “Firm dynamics, markup variations, and the business cycle,” *Journal of Monetary Economics*, 2008, *55* (7), 1238–1252.

Karabarbounis, Loukas and Brent Neiman, "Accounting for factorless income," *NBER Macroeconomics Annual*, 2019, *33* (1), 167–228.

Kehrig, Matthias, "The cyclical nature of the productivity distribution," *Earlier version: US Census Bureau Center for Economic Studies Paper No. CES-WP-11-15*, 2015.

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

_ , _ , and Simon Mongey, "Quantifying market power and business dynamism," Technical Report 2021-54 2021.

Matsuyama, Kiminori, "Increasing Returns, Industrialization, and Indeterminacy of Equilibrium," *The Quarterly Journal of Economics*, 1991, *106* (2), 617–650.

Mian, Atif, Ludwig Straub, and Amir Sufi, "Indebted Demand," *The Quarterly Journal of Economics*, 2021, *136* (4), 2243–2307.

Mongey, Simon, "Market structure and monetary non-neutrality," *mimeo*, 2019.

Nekarda, Christopher J and Valerie A Ramey, "The Cyclical Behavior of the Price-Cost Markup," *Journal of Money, Credit and Banking*, 2020, *52* (S2), 319–353.

Pagano, Marco, "Imperfect Competition, Underemployment Equilibria and Fiscal Policy," *Economic Journal*, June 1990, *100* (401), 440–463.

Queralto, Albert, "A model of slow recoveries from financial crises," *Journal of Monetary Economics*, 2020, *114*, 1–25.

Rotemberg, Julio and Garth Saloner, "A supergame-theoretic model of price wars during booms," *New Keynesian Economics*, 1986, *2*, 387–415.

Rotemberg, Julio J and Michael Woodford, "Markups and the business cycle," *NBER macroeconomics annual*, 1991, *6*, 63–129.

_ and _ , "Dynamic general equilibrium models with imperfectly competitive product markets," *Frontiers of business cycle research*, 1995, *243.*

Schaal, Edouard and Mathieu Taschereau-Dumouchel, "Coordinating business cycles," *Working Paper*, 2018.

_ and _ , "Aggregate Demand and the Dynamics of Unemployment," *Working Paper*, 2019.

Traina, James, "Is aggregate market power increasing? production trends using financial statements," *Production Trends Using Financial Statements (February 8, 2018)*, 2018.

Van Reenen, John, "Increasing Differences Between Firms: Market Power and the Macro-Economy," *Working Paper*, September 2018.

Wang, Olivier and Iván Werning, "Dynamic oligopoly and price stickiness," Technical Report, National Bureau of Economic Research 2020.

Appendix A

A.1 Data Appendix

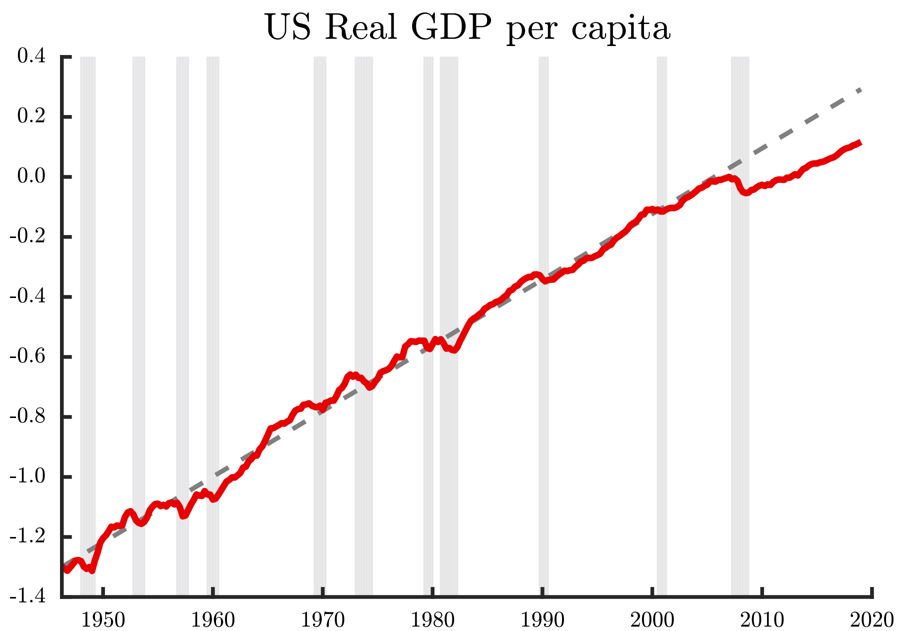


Figure A.1: The *Great Deviation*

Note: This figure shows real GDP per capita (from BEA). The series is in logs, undetrended and centered around 2007. The linear trend is computed for the 1947-2007 period.

Data Definition Table A.1 provides information on all the data sources used in Section 5.

Variable	Source
Real GDP	BEA – NIPA Table 1.1.3 (line 1)
Real Personal Consumption Expenditures	BEA – NIPA Table 1.1.3 (line 2)
Real Gross Private Domestic Investment	BEA – NIPA Table 1.1.3 (line 7)
Total Hours	BLS – Nonfarm Business sector: Hours of all persons
Aggregate TFP	Fernald (2012): Raw Business Sector TFP
Population	BEA – NIPA Table 2.1 (line 40)

Table A.1: Data sources

Aggregate Profit Share The aggregate profit share is computed as

$$\text{profit share}_t \;=\; 1 \;-\; \text{labor share}_t \;-\; \underbrace{\frac{R_t \cdot \mathrm{K}_t \;-\; \mathrm{DEP}_t}{\mathrm{VA}_t}}_{\text{capital share}}$$

where labor share_t is the labor share of the US business sector (from BLS), VA_t is the total value added of the US business sector (NIPA Table 1.3.5, line 2). K_t is the value of private fixed assets (including intangibles) of the US business sector (NIPA Table 6.1, line 1 - line 9 - line 10) and DEP_t is depreciation (NIPA Table 6.4, line 1 - line 9 - line 10). Finally, R_t is the required rate of return. We follow Eggertsson et al. (2018) and compute it as the difference between Moody's Seasoned BAA Corporate Bond Yield and a 5-year moving average of past CPI inflation (from BLS, used as a proxy for expected inflation).

Industry-level Labor Share We obtain data on the labor share at the industry level from the BLS 'Labor Productivity and Costs' (LPC) database. We calculate the labor share as the ratio of 'Labor compensation' to 'Value of Production'. Note that this ratio gives the share of labor compensation in total revenues, and not in value added.[34]

A.2 Proofs and Derivations of Section 2

Proof of Lemma 1

Proof. Let η_{ij} denote the input share of each firm, namely the fraction of a given input used by firm j in industry i out of the total amount of that input in the economy. Formally,

$$\eta_{ij} = \frac{k_{ij}}{K} = \frac{l_{ij}}{L}$$

Because of CRS, we have that

$$F\left(k_{ij}, l_{ij}\right) = \eta_{ij} F\left(K, L\right)$$

Using the fact that aggregate output must be equal to total firms' revenues

$$\begin{aligned} Y = &\; \textstyle\int_{ij} p_{ij}\, y_{ij} \\ = &\; \textstyle\int_{ij} p_{ij}\, \gamma_{ij}\, F\left(k_{ij}, l_{ij}\right) \\ = &\; \underbrace{\left(\int_{ij} p_{ij}\, \gamma_{ij}\, \eta_{ij}\right)}_{\Phi} F\left(K, L\right) \end{aligned}$$

[34]This ratio coincides with the 'Labor cost share' provided by the BLS. This variable is, however, available just for a restricted number of industries.

This establishes that an aggregate production function exists. The term Φ represents aggregate TFP, and is a function of individual firms' prices (p_{ij}), idiosyncratic productivities (γ_{ij}) and input shares (η_{ij}). It therefore depends on technology and on the set of active firms. ■

Derivation of Equation 1

Firms with productivity γ_{ij} solve

$$\max \quad p_{ij}\,\gamma_{ij}F\left(k_{ij},l_{ij}\right)-R\,k_{ij}-W\,l_{ij}.$$

The first order condition with respect to capital is

$$\begin{aligned}
&\frac{dp_{ij}}{dy_{ij}}F_k\left(k_{ij},l_{ij}\right)\gamma_{ij}F\left(k_{ij},l_{ij}\right)+p_{ij}\gamma_{ij}F_k\left(k_{ij},l_{ij}\right)=R\\
\Leftrightarrow\quad & p_{ij}F_k\left(k_{ij},l_{ij}\right)=\underbrace{\left[\frac{dp_{ij}}{dy_{ij}}\frac{y_{ij}}{p_{ij}}+1\right]^{-1}}_{\mu_{ij}}\frac{R}{\gamma_{ij}},
\end{aligned}$$

where

$$\mu_{ij}=\left[\frac{dp_{ij}}{dy_{ij}}\frac{y_{ij}}{p_{ij}}+1\right]^{-1}$$

is the markup, which is equal to the inverse of the firm's factor share

$$\omega_{ij}=\frac{1}{\mu_{ij}}=\left[\frac{dp_{ij}}{dy_{ij}}\frac{y_{ij}}{p_{ij}}+1\right].$$

Let Θ be the unit variable cost for a firm with unit productivity, or, equivalently, $\lambda_{ij}\gamma_{ij}=\Theta$, where λ_{ij} is the Lagrange multiplier of the cost minimization problem for a firm with productivity γ_{ij}. Then we have

$$\begin{aligned}
&p_{ij}F_k\left(k_{ij},l_{ij}\right)=\mu_{ij}\frac{R}{\gamma_{ij}}\\
\Leftrightarrow\quad &\underbrace{\mu_{ij}\frac{\Theta}{\gamma_{ij}}}_{p_{ij}}F_k\left(k_{ij},l_{ij}\right)=\mu_{ij}\frac{R}{\gamma_{ij}}\\
\Leftrightarrow\quad &\Theta F_k\left(k_{ij},l_{ij}\right)=R.
\end{aligned}$$

Because of CRS, $F_k\left(k_{ij},l_{ij}\right)$ only depends on the capital-labor ratio. Since all firms face the same factor prices we have

$$\frac{k_{ij}}{l_{ij}}=\frac{K}{L}\quad\Rightarrow\quad F_k\left(k_{ij},l_{ij}\right)=F_k\left(K,L\right),$$

which implies

$$\Theta F_k\left(K,L\right)=R. \tag{26}$$

This result states that the rental rate of the economy can be written as the unit variable cost index Θ times the marginal product of capital.

Aggregation

Now, we will show that $\Theta = \Omega\Phi$. Note that the following identity should hold in equilibrium

$$\int_{ij} \omega_{ij} p_{ij} \gamma_{ij} F\left(k_{ij}, l_{ij}\right) = \Omega\, \Phi F\left(K, L\right).$$

In words, the sum of factor payments from all firms, should be equal to aggregate factor payments in the economy. Note that

$$\omega_{ij} p_{ij} \gamma_{ij} = \underbrace{\frac{1}{\mu_{ij}}}_{\omega_{ij}} \underbrace{\mu_{ij} \frac{\Theta}{\gamma_{ij}}}_{p_{ij}} \gamma_{ij} = \Theta.$$

Substituting we obtain

$$\begin{aligned} & \textstyle\int_{ij} \Theta\, \eta_{ij} F\left(K, L\right) = \Omega\, \Phi F\left(K, L\right) \\ \Leftrightarrow \quad & \Theta\, F\left(K, L\right) \underbrace{\int_{ij} \eta_{ij}}_{=1} = \Omega\, \Phi F\left(K, L\right) \\ \Leftrightarrow \quad & \Theta = \Omega\, \Phi \end{aligned}$$

where η_{ij} is the input share of each firm (defined above in A.2). Combining the last equation with equation (26) we obtain

$$\Omega\, \Phi F_k\left(K, L\right) = R,$$

as stated in equation 1.

Proof of Proposition 1

Proof. First, note that, by Definition 2, $\frac{\partial \chi_n}{\partial \lambda} = -\frac{\partial \mathcal{K}_{n-1}^U / \mathcal{K}_n^S}{\partial \lambda}$. Next, note that a sufficient condition for the ratio $\mathcal{K}_{n-1}^U / \mathcal{K}_n^S$ to decrease in λ is that $R_\lambda < 0$ at $\mathcal{K}_{n-1}^U$ and $\mathcal{K}_n^S$. It therefore suffices to characterize a sufficient condition for the $R(\cdot)$ map to be locally decreasing in λ. To this end, we begin by using the analog of equation 3 for the wage: $W = \Theta(\Lambda, n) F_L$. By substituting in the labor supply L^S and inverting we have $L^S F_L^{-1} = \Theta$. As both functions on the LHS are non-decreasing in L it follows that $\text{sgn}\{L_\lambda\} = \text{sgn}\{\Theta_\lambda\}$. We are interested in $\partial R / \partial \lambda$ which is given by

$$R_\lambda = \Theta_\lambda F_K + \Theta F_{K\lambda}.$$

As $\text{sgn}\{F_{K\lambda}\} = \text{sgn}\{L_\lambda\} = \text{sgn}\{\Theta_\lambda\}$, $\Theta > 0$, $\text{sgn}\{R_\lambda\} = \text{sgn}\{\Theta_\lambda\}$, the statement follows. ■

A.3 Proofs and Additional Results for Section 3

Proof of Lemma 2

Proof. Using equation (15) and imposing $\eta = 1$ and symmetry we obtain $\Omega(\Gamma_n, n) = 1-(1-\rho)HHI_t$ where HHI_t is the Herfindhal-Hirschman Index of the economy. As $\rho < 1$ the aggregate factor share decreases in concentration, as measured by HHI. The latter is trivially decreasing in n, which proves the first part of the lemma. Furthermore, a mean-preserving spread of productivity, holding n fixed, implies a spread of the market shares distribution. Since HHI is a convex function of market shares, it necessarily increases after a spread of the productivity distribution. The statement follows. ■

Proof of Proposition 2

Proof. Suppose $\eta = 1$.When there are n active firms in a given industry, the profits of a firm with productivity γ_j are equal to

$$\Pi\left(\gamma_j, n, \Gamma, \Theta, Y\right) = \mathit{\Lambda}\left(\gamma_j, n, \Gamma\right)\Theta^{-\frac{\rho}{1-\rho}}Y \tag{27}$$

where $\mathit{\Lambda}\left(\gamma_j, n, \Gamma\right)$ has been defined in Appendix B.2. A symmetric equilibrium with n firms per industry is possible provided that

$$\begin{aligned} \mathit{\Lambda}\left(\gamma_n, n, \Gamma\right)\Theta^{-\frac{\rho}{1-\rho}}Y &\geq c \\ \mathit{\Lambda}\left(\gamma_{n+1}, n+1, \Gamma\right)\Theta^{-\frac{\rho}{1-\rho}}Y &\leq c \end{aligned}$$

Using equation (22), we can write the above inequalities as

$$\underline{K}\left(\Gamma, n\right) \leq K_t \leq \overline{K}\left(\Gamma, n\right), \tag{28}$$

where

$$\underline{K}\left(\Gamma, n\right) = \left\{\frac{c}{\mathit{\Lambda}\left(\gamma_n, n, \Gamma\right)}\left(1-\alpha\right)^{-\frac{1-\alpha}{\nu+\alpha}}\left[\Phi\left(\Gamma, n\right)\right]^{-1}\left[\Theta\left(\Gamma, n\right)\right]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}}\right\}^{\frac{\nu+\alpha}{\alpha(1+\nu)}} \tag{29}$$

$$\overline{K}\left(\Gamma, n\right) = \left\{\frac{c}{\mathit{\Lambda}\left(\gamma_{n+1}, n+1, \Gamma\right)}\left(1-\alpha\right)^{-\frac{1-\alpha}{\nu+\alpha}}\left[\Phi\left(\Gamma, n\right)\right]^{-1}\left[\Theta\left(\Gamma, n\right)\right]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}}\right\}^{\frac{\nu+\alpha}{\alpha(1+\nu)}}. \tag{30}$$

The condition on uniqueness can be derived by noting that it arises if and only if the following holds $\forall n$

$$
\begin{aligned}
& \underline{K}\left(\Gamma, n+1\right) > \overline{K}\left(\Gamma, n\right) \\
\Leftrightarrow \quad & \left[\Phi\left(\Gamma, n+1\right)\right]^{-1}\left[\Theta\left(\Gamma, n+1\right)\right]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}} > \left[\Phi\left(\Gamma, n\right)\right]^{-1}\left[\Theta\left(\Gamma, n\right)\right]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}} \\
\Leftrightarrow \quad & \frac{\Phi\left(\Gamma, n\right)}{\Phi\left(\Gamma, n+1\right)} > \left[\frac{\Theta\left(\Gamma, n\right)}{\Theta\left(\Gamma, n+1\right)}\right]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}}
\end{aligned}
$$

Therefore, when there are no productivity differences across firms, the condition becomes

$$
\begin{aligned}
& \left[\frac{\Theta\left(\Gamma, n\right)}{\Theta\left(\Gamma, n+1\right)}\right]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}} < 1 \\
\Leftrightarrow \quad & \frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha} > 0 \\
\Leftrightarrow \quad & \frac{\rho}{1-\rho} > \frac{1-\alpha}{\nu+\alpha}
\end{aligned}
$$

since $\Theta\left(\Gamma, n+1\right) > \Theta\left(\Gamma, n\right)$. ■

Proof of Lemma 3

Proof. In a steady-state we have a constant rental rate

$$R^{*} = \beta^{-1} - (1-\delta) \tag{31}$$

and

$$\delta K = sY \tag{32}$$

Combining these two equations with equation (17) we obtain

$$
\begin{aligned}
& \beta^{-1} - (1-\delta) = \alpha\, \Omega\left(\mathbf{\Gamma}, \mathbf{N}\right)\ \frac{Y^{*}}{K^{*}} \\
\Leftrightarrow & \beta^{-1} - (1-\delta) = \alpha\, \Omega\left(\mathbf{\Gamma}, \mathbf{N}\right)\ \frac{\delta}{s^{*}} \\
\Leftrightarrow & s^{*} = \frac{\delta\ \alpha}{\beta^{-1} - (1-\delta)} \Omega\left(\mathbf{\Gamma}, \mathbf{N}\right)
\end{aligned} \tag{33}
$$

■

Proof of Proposition 3

Proof. We have

$$R_t \;=\; \alpha\ (1-\alpha)^{\frac{1-\alpha}{\nu+\alpha}}\ \Theta\left(\mathbf{\Gamma}, \mathbf{N}_t\right)^{\frac{1+\nu}{\nu+\alpha}}\ K_t^{-\nu\frac{1-\alpha}{\nu+\alpha}}.$$

Let $\underline{R}(\Gamma, n)$ and $\overline{R}(\Gamma, n)$ be the rental rates at $\underline{K}(\Gamma, n)$ and $\overline{K}(\Gamma, n)$ respectively. Then, multiplicity obtains if there exists an $n \in \mathbb{N}$ such that

$$\overline{R}(\Gamma, n) < \underbrace{\beta^{-1} - (1-\delta)}_{R^*} < \underline{R}(\Gamma, n+1)$$

$$\Leftrightarrow \quad \Theta(\Gamma, n)^{\frac{1+\nu}{\nu+\alpha}} \; \overline{K}(\Gamma, n)^{-\nu\frac{1-\alpha}{\nu+\alpha}} < \frac{\beta^{-1} - (1-\delta)}{\alpha \; (1-\alpha)^{\frac{1-\alpha}{\nu+\alpha}}} < \; \Theta(\Gamma, n+1)^{\frac{1+\nu}{\nu+\alpha}} \; \underline{K}(\Gamma, n+1)^{-\nu\frac{1-\alpha}{\nu+\alpha}}.$$

■

Proof of Lemma 4

Proof. When all industries are identical and have n firms, the factor price index is equal to

$$\Theta(\Gamma_n, n) = \frac{n - (1-\rho)}{\sum_{k=1}^{n} \frac{1}{\gamma_k}} \tag{34}$$

As Θ is a concave function of γ_k, by the definition of MPS, we have that if $\tilde{\Gamma}$ is a MPS of Γ, then $\Theta(\tilde{\Gamma}, n) < \Theta(\Gamma, n)$. ■

Proof of Proposition 4

Proof. a) Let K_n^* be a steady-state where all industries are have n firms and common productivity distribution Γ_n. Using equation (17), we can define K_n^* as

$$R^* \; = \; \alpha \; (1-\alpha)^{(1-\alpha)/(\nu+\alpha)} \; \Theta(\Gamma_n, n)^{(\nu+1)/(\nu+\alpha)} \; (K_n^*)^{-\nu(1-\alpha)/(\nu+\alpha)} \tag{35}$$

Recall from Lemma 4 that $\Theta(\Gamma_n, n)$ declines after a MPS on Γ_n. Then K_n^* must also decline.

b) We provide a sufficient condition under which the unstable steady-state increases after an MPS. Note that the unstable steady-state increases whenever the increasing segment of the rental rate map lies strictly underneath the original one.

We know from the proof of Proposition 4.a that the new rental rate at $\underline{K}(2)$ is strictly lower than before. The proof involves two steps

[A] We derive a sufficient condition under which the new rental rate at $\overline{K}(1)$ is lower than before

[B] We show that the increasing segment of the rental rate map after an MPS cannot cross the previous one more than once. Thus, if the new segment starts and ends below the previous one, it can never go above it.

Proof of Part A

The free entry condition is

$$\Lambda(n)\,\Theta(n-1)^{-\frac{\rho}{1-\rho}}\;\Phi(n-1)\,K^{\alpha}\,L^{1-\alpha} \;=\; c_f \tag{36}$$

Using

$$L=[(1-\alpha)\,\Theta(n-1)]^{\frac{1}{\nu+\alpha}}\,K^{\frac{\alpha}{\nu+\alpha}} \tag{37}$$

we can rewrite the free-entry condition as

$$\begin{aligned}
&\Lambda(n)\,\Theta(n-1)^{-\frac{\rho}{1-\rho}}\;\Phi(n-1)\;K^{\alpha}\;[(1-\alpha)\,\Theta(n-1)]^{\frac{1-\alpha}{\nu+\alpha}}\,K^{\alpha\frac{1-\alpha}{\nu+\alpha}} \;=\; c_f\\
\Leftrightarrow&\Lambda(n)\,\Theta(n-1)^{-\frac{\rho}{1-\rho}}\;\Phi(n-1)\,[(1-\alpha)\,\Theta(n-1)]^{\frac{1-\alpha}{\nu+\alpha}}\,K^{\alpha\frac{\nu+1}{\nu+\alpha}} \;=\; c_f\\
\Leftrightarrow&K \;=\; \left[\frac{c_f\,(1-\alpha)^{-\frac{1-\alpha}{\nu+\alpha}}}{\Lambda(n)\,\Theta(n-1)^{\frac{1-\alpha}{\nu+\alpha}-\frac{\rho}{1-\rho}}\;\Phi(n-1)}\right]^{\frac{1}{\alpha}\frac{\nu+\alpha}{\nu+1}}
\end{aligned} \tag{38}$$

The interest rate is

$$\begin{aligned}
R=&\alpha\,\Theta(n-1)\;K^{\alpha-1}\,L^{1-\alpha}\\
=&\alpha\,\Theta(n-1)\;K^{\alpha-1}\;[(1-\alpha)\,\Theta(n-1)]^{\frac{1-\alpha}{\nu+\alpha}}\,K^{\alpha\frac{1-\alpha}{\nu+\alpha}}\\
=&\alpha\;(1-\alpha)^{\frac{1-\alpha}{\nu+\alpha}}\;\Theta(\mathbf{\Gamma},\mathbf{N}_t)^{\frac{\nu+1}{\nu+\alpha}}\;K^{\nu\frac{\alpha-1}{\nu+\alpha}}
\end{aligned} \tag{39}$$

Putting the two together

$$R=\alpha\;(1-\alpha)^{\frac{1-\alpha}{\nu+\alpha}}\;\Theta(n-1)^{\frac{\nu+1}{\nu+\alpha}}\;\left(\frac{\Lambda(n)\,\Theta(n-1)^{\frac{1-\alpha}{\nu+\alpha}-\frac{\rho}{1-\rho}}\;\Phi(n-1)}{c_f\,(1-\alpha)^{-\frac{1-\alpha}{\nu+\alpha}}}\right)^{\frac{\nu}{\nu+1}\frac{1-\alpha}{\alpha}} \tag{40}$$

implying

$$R^{\frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}}=\propto\;\Theta(n-1)^{\frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}\frac{\nu+1}{\nu+\alpha}}\;\Lambda(n)\,\Theta(n-1)^{\frac{1-\alpha}{\nu+\alpha}-\frac{\rho}{1-\rho}}\;\Phi(n-1) \tag{41}$$

where

$$\Theta(n)=g(n) \tag{42}$$

$$\Phi(n)=\frac{1}{\sum_{j=1}^{n}\frac{s_j}{\pi_j}} \tag{43}$$

$$\Lambda(n)=s_n^2\,[g(n)]^{\frac{\rho}{1-\rho}} \tag{44}$$

We can thus rewrite (40) as

$$R^{\frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}} = g\left(n-1\right)^{\frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}\frac{\nu+1}{\nu+\alpha}+\frac{1-\alpha}{\nu+\alpha}} \frac{s_n^2\left[\dfrac{g\left(n\right)}{g\left(n-1\right)}\right]^{\frac{\rho}{1-\rho}}}{\sum\limits_{j=1}^{n-1}\dfrac{\hat{s}_j}{\pi_j}} \tag{45}$$

where $\hat{s}_j$ is the market share of firm j in an industry with $n-1$ firms and s_j is the market share of that firm when there are n player in the industry. We want to show that the expression in (45) goes down when we do an MPS on n firms. The challenge is in the fact that the expression involves terms that refer to $n-1$ industries.

Under $n=2$, we have

$$g\left(1\right) = \rho\,\pi_1 \tag{46}$$

$$g\left(2\right) = \frac{1+\rho}{\dfrac{1}{\pi_1}+\dfrac{1}{\pi_2}} \tag{47}$$

so that

$$\begin{aligned} R^{\frac{\alpha}{1-\alpha}} &= \left(\rho\,\pi_1\right)^{\frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}\frac{\nu+1}{\nu+\alpha}+\frac{1-\alpha}{\nu+\alpha}-\frac{\rho}{1-\rho}} \left(\frac{1+\rho}{\dfrac{1}{\pi_1}+\dfrac{1}{\pi_2}}\right)^{\frac{\rho}{1-\rho}} \frac{\left[1-\dfrac{1+\rho}{\dfrac{1}{\pi_1}+\dfrac{1}{\pi_2}}\dfrac{1}{\pi_2}\right]^2}{\dfrac{1}{\pi_1}} \\ &= \propto \pi_1^{\frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}\frac{\nu+1}{\nu+\alpha}+\frac{1-\alpha}{\nu+\alpha}-\frac{\rho}{1-\rho}-1} \left[\pi_1-\pi_1\frac{1+\rho}{\dfrac{2x}{\pi_1}}\right]^2 \left[\frac{1+\rho}{\dfrac{1}{\pi_1}+\dfrac{1}{2x-\pi_1}}\right]^{\frac{\rho}{1-\rho}} \end{aligned} \tag{48}$$

The last term is decreasing on an MPS, since it is simply $g\left(2\right)$. The first term is decreasing on an MPS provided that

$$1+\frac{\rho}{1-\rho} > \frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}\frac{\nu+1}{\nu+\alpha}+\frac{1-\alpha}{\nu+\alpha} \Leftrightarrow \alpha < \frac{\nu+\rho-1}{\nu\left(2-\rho\right)} \tag{49}$$

since this MPS must result in higher π_1. We just need to evaluate the term in the middle. Note that we can rewrite it as

$$\pi_1 - \pi_1^2\frac{1+\rho}{2x} \tag{50}$$

where $2x \equiv \pi_1 + \pi_2$ is fixed by construction. The derivative of the expression above is

$$\frac{\partial}{\partial \pi_1} = 1 - 2\pi_1 \frac{1+\rho}{2x} \tag{51}$$

$$= 1 - \underbrace{\frac{\pi_1}{x}}_{>1} (1+\rho) < 0 \tag{52}$$

Therefore, for $n = 2$, the interest rate is always declining on an MPS provided that

$$\rho > 1 - \nu \frac{1-\alpha}{1+\nu\alpha} \tag{53}$$

Proof of Part B

Recall that the free entry condition is

$$\Lambda_j \Theta^{-\frac{\rho}{1-\rho}} \, \Phi \, K^{\alpha} \, L^{1-\alpha} = c_f \tag{54}$$

Aggregate TFP can be written as

$$\Phi = \frac{\left\{(1-m)\left[g(n-1)\right]^{\frac{\rho}{1-\rho}} + m\left[g(n)\right]^{\frac{\rho}{1-\rho}}\right\}^{\frac{1}{\rho}}}{(1-m)\left[g(n-1)\right]^{\frac{1}{1-\rho}} h(n-1) + m\left[g(n)\right]^{\frac{1}{1-\rho}} h(n)}$$

$$= \frac{\Theta^{\frac{1}{1-\rho}}}{(1-m)\left[g(n-1)\right]^{\frac{1}{1-\rho}} h(n-1) + m\left[g(n)\right]^{\frac{1}{1-\rho}} h(n)}$$

where

$$g(n) = \frac{n-(1-\rho)}{\sum_{j=1}^{n} \frac{1}{\pi_j}} \tag{55}$$

$$h(n) = \sum_{j=1}^{n} \frac{s_j}{\pi_j} \tag{56}$$

Now suppose that we do the MPS and have $\tilde{\Theta} = \Theta$ at the same K.[35] From the free entry condition

[35]We also have $\tilde{L} = L$, since L is a function of Θ and K.

and the expression for Φ, this is possible if

$$\frac{\Lambda_j}{\tilde{\Lambda}_j} = \frac{(1-m)\left[g(n-1)\right]^{\frac{1}{1-\rho}} h(n-1) + m\left[g(n)\right]^{\frac{1}{1-\rho}} h(n)}{(1-\tilde{m})\left[\tilde{g}(n-1)\right]^{\frac{1}{1-\rho}} \tilde{h}(n-1) + \tilde{m}\left[\tilde{g}(n)\right]^{\frac{1}{1-\rho}} \tilde{h}(n)}$$

$$\Leftrightarrow \frac{\Lambda_j}{\tilde{\Lambda}_j} = \frac{\left[g(n-1)\right]^{\frac{1}{1-\rho}} h(n-1) + m\left\{\left[g(n)\right]^{\frac{1}{1-\rho}} h(n) - \left[g(n-1)\right]^{\frac{1}{1-\rho}} h(n-1)\right\}}{\left[\tilde{g}(n-1)\right]^{\frac{1}{1-\rho}} \tilde{h}(n-1) + \tilde{m}\left\{\left[\tilde{g}(n)\right]^{\frac{1}{1-\rho}} \tilde{h}(n) - \left[\tilde{g}(n-1)\right]^{\frac{1}{1-\rho}} \tilde{h}(n-1)\right\}} \tag{57}$$

Rearranging this equation, we can write

$$\tilde{m} = a_1 + b_1 \cdot m \tag{58}$$

where a_1 and b_1 are some numbers (independent of K).

Furthermore, from $\tilde{\Theta} = \Theta$ we have

$$(1-m)\left[g(n-1)\right]^{\frac{\rho}{1-\rho}} + m\left[g(n)\right]^{\frac{\rho}{1-\rho}} = (1-\tilde{m})\left[\tilde{g}(n-1)\right]^{\frac{\rho}{1-\rho}} + \tilde{m}\left[\tilde{g}(n)\right]^{\frac{\rho}{1-\rho}}$$

$$\Leftrightarrow \left[g(n-1)\right]^{\frac{\rho}{1-\rho}} + m\left\{\left[g(n)\right]^{\frac{\rho}{1-\rho}} - \left[g(n-1)\right]^{\frac{\rho}{1-\rho}}\right\} = \left[\tilde{g}(n-1)\right]^{\frac{\rho}{1-\rho}} + \tilde{m}\left\{\left[\tilde{g}(n)\right]^{\frac{\rho}{1-\rho}} - \left[\tilde{g}(n-1)\right]^{\frac{\rho}{1-\rho}}\right\} \tag{59}$$

Rearranging this equation, we can write

$$\tilde{m} = a_2 + b_2 \cdot m \tag{60}$$

Combining (58) and (60), there is at most one pair $(m, \tilde{m})$ such that $\tilde{\Theta} = \Theta$. This establishes that $\tilde{\Theta}$ cannot cross Θ twice. ■

Proposition A.1 (Mean Preserving Spread and Fragility, general n)**.** *Let $\eta = 1$ and suppose that all industries are identical to start ($\gamma_{ij} = \gamma_j \; \forall i$). Let $K^*(n)$ be a steady-state with n firms. Let λ be a mean-preserving spread on the distribution $\{\gamma_1, \ldots, \gamma_n\}$ of active firms, such that for any $j = 1, \ldots, n-1$, γ_1/γ_j is unchanged. Then, if*

$$\rho > \frac{1+\nu\alpha}{1+\nu} \tag{61}$$

we have that

$$\frac{\partial \underline{B}(n)}{\partial \lambda} > 0. \tag{62}$$

Proof of Proposition A.1. Let us now provide a sufficient condition for general n. First note that we can write (45) as

$$R^{\frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}} = \underbrace{g\left(n-1\right)^{\frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}\frac{\nu+1}{\nu+\alpha}+\frac{1-\alpha}{\nu+\alpha}-\frac{\rho}{1-\rho}} g\left(n\right)^{\frac{\rho}{1-\rho}}}_{v} \underbrace{\frac{1}{\sum_{j=1}^{n-1}\frac{\hat{s}_j}{\pi_j}} s_n^2}_{z} \tag{63}$$

$$R^{\frac{\alpha}{1-\alpha}} = \underbrace{g\left(n-1\right)^{\frac{\alpha}{1-\alpha}-\frac{\rho}{1-\rho}} g\left(n\right)^{\frac{\rho}{1-\rho}}}_{v} \underbrace{\frac{1}{\sum_{j=1}^{n-1}\frac{\hat{s}_j}{\pi_j}} s_n^2}_{z} \tag{64}$$

The first term v is always decreasing on an MPS provided that

$$\frac{\rho}{1-\rho} > \frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}\frac{\nu+1}{\nu+\alpha} + \frac{1-\alpha}{\nu+\alpha} \Leftrightarrow \alpha < \frac{\rho\left(1+\nu\right)-1}{\nu} \tag{65}$$

To see it note that $g\left(n\right)$ is decreasing on an MPS. If $g\left(n-1\right)$ is increasing on an MPS, it immediately follows that v decreases on an MPS when the above condition is satisfied. If $g\left(n-1\right)$ is instead decreasing on an MPS, just rewrite v as

$$v = g\left(n-1\right)^{\frac{\nu+1}{\nu}\frac{\alpha}{1-\alpha}\frac{\nu+1}{\nu+\alpha}+\frac{1-\alpha}{\nu+\alpha}} \left[\frac{g\left(n\right)}{g\left(n-1\right)}\right]^{\frac{\rho}{1-\rho}} \tag{66}$$

and note that $\frac{g\left(n\right)}{g\left(n-1\right)}$ decreases on an MPS.

Thus, all we need to show is that z is also decreasing on an MPS. Note that we can write z as

$$z = \frac{s_n^2}{\sum_{j=1}^{n-1}\frac{\hat{s}_j}{\pi_j}} = \frac{1}{\sum_{j=1}^{n-1}\frac{\hat{s}_j}{\pi_j}\frac{1}{s_n^2}} \tag{67}$$

We know that s_n is decreasing on an MPS. A sufficient condition for z to be decreasing on an MPS is that

$$\frac{\hat{s}_j}{\pi_j}\frac{1}{s_n^2} \tag{68}$$

is increasing on an MPS for every $j = 1, 2, \ldots, n-1$.

Consider an MPS such that

$$\widetilde{\pi}_j = \gamma\,\pi_j \quad \forall j = 1, 2, \ldots, n-1 \tag{69}$$

and

$$\gamma \sum_{j=1}^{n-1} \pi_j + \widetilde{\pi}_n = \sum_{j=1}^{n-1} \pi_j + \pi_n \tag{70}$$

$$\Leftrightarrow \widetilde{\pi}_n = \pi_n - (\gamma - 1) \sum_{j=1}^{n-1} \pi_j \tag{71}$$

In this case we have

$$\frac{1}{\gamma} \underbrace{\frac{\hat{s}_j}{\pi_j}}_{\text{const}} \frac{1}{\widetilde{s}_n^2} \tag{72}$$

We need to show that

$$\gamma \underbrace{\left[1 - \frac{n-(1-\rho)}{\sum_{j=1}^{n-1} \frac{1}{\gamma \pi_j} + 1} \frac{1}{\pi_n - (\gamma-1) \sum_{j=1}^{n-1} \pi_j} \right]^2}_{\widetilde{s}_n} \tag{73}$$

decreases in γ. Note that we can rewrite it as

$$\gamma \underbrace{\left[1 - \frac{n-(1-\rho)}{\sum_{j=1}^{n-1} \frac{\pi_n - (\gamma-1)\sum_{j=1}^{n-1} \pi_j}{\gamma \, \pi_j} + 1} \right]^2}_{\tilde{s}_n} = \gamma \underbrace{\left[1 - \frac{n-(1-\rho)}{\left[\frac{1}{\gamma} \left(\pi_n + \sum_{j=1}^{n-1} \pi_j \right) - \sum_{j=1}^{n-1} \pi_j \right] \sum_{j=1}^{n-1} \frac{1}{\pi_j} + 1} \right]^2}_{\tilde{s}_n} \tag{74}$$

The derivative with respect to γ is

$$\tilde{s}_n^2 + \gamma 2 \tilde{s}_n \left\{ -(-1)\left[n-(1-\rho)\right] \frac{\left(\pi_n + \sum_{j=1}^{n-1} \pi_j \right) \sum_{j=1}^{n-1} \frac{1}{\pi_j} \left(-\frac{1}{\gamma^2} \right)}{\left(\left[\frac{1}{\gamma} \left(\pi_n + \sum_{j=1}^{n-1} \pi_j \right) - \sum_{j=1}^{n-1} \pi_j \right] \sum_{j=1}^{n-1} \frac{1}{\pi_j} + 1 \right)^2} \right\} \tag{75}$$

which is lower than zero if

$$\tilde{s}_n - \frac{2}{\gamma}\left[\frac{n-(1-\rho)}{\sum\limits_{j=1}^{n}\frac{\tilde{\pi}_n}{\gamma\pi_j}}\frac{1}{\sum\limits_{j=1}^{n}\frac{\tilde{\pi}_n}{\gamma\pi_j}}\left(\pi_n+\sum_{j=1}^{n-1}\pi_j\right)\sum_{j=1}^{n-1}\frac{1}{\pi_j}\right] < 0 \Leftrightarrow 1-\underbrace{\frac{n-(1-\rho)}{\sum\limits_{j=1}^{n}\frac{\tilde{\pi}_n}{\pi_j}}}_{(a)}\left[1+2\gamma\underbrace{\frac{\sum\limits_{j=1}^{n}\pi_j}{\tilde{\pi}_n}}_{>n}\underbrace{\frac{\sum\limits_{j=1}^{n-1}\frac{1}{\pi_j}}{\sum\limits_{j=1}^{n}\frac{1}{\pi_j}}}_{(b)}\right] < 0 \tag{76}$$

If suffices to prove that $(a) > 1/3$ and that $(b) > 1/n$.

To prove the first, note that

$$\tilde{s}_n = \frac{1}{1-\rho}\left[1-\frac{n-(1-\rho)}{\sum\limits_{j=1}^{n}\frac{\tilde{\pi}_n}{\pi_j}}\right] < \frac{1}{3} \Leftrightarrow \underbrace{\frac{n-(1-\rho)}{\sum\limits_{j=1}^{n}\frac{\tilde{\pi}_n}{\pi_j}}}_{(a)} > \frac{2+\rho}{3} > \frac{1}{3} \tag{77}$$

To prove the second, note that

$$\frac{\sum\limits_{j=1}^{n-1}\frac{1}{\pi_j}}{\sum\limits_{j=1}^{n}\frac{1}{\pi_j}} > \frac{1}{n} \Leftrightarrow \sum_{j=1}^{n}\frac{\pi_n}{\pi_j} > \frac{n}{n-1} \tag{78}$$

The last equation is implied by the fact that

$$1-\frac{n-(1-\rho)}{\sum\limits_{j=1}^{n}\frac{\pi_n}{\pi_j}} > 0 \Leftrightarrow \sum_{j=1}^{n}\frac{\pi_n}{\pi_j} > \underbrace{n-(1-\rho)}_{>2 \text{ under } n\geq 3} \tag{79}$$

which is needed for $s_n > 0$. This completes the proof. ■

Proof of Proposition 5

Proof. From equation (17) we can write

$$R_t \;=\; \alpha\;(1-\alpha)^{(1-\alpha)/(\nu+\alpha)}\;\Theta\left(\mathbf{\Gamma},\mathbf{N}_t\right)^{(\nu+1)/(\nu+\alpha)}\;K_t^{-\nu(1-\alpha)/(\nu+\alpha)} \tag{80}$$

where $\Theta\left(\mathbf{\Gamma},\mathbf{N}_t\right)$ is increasing in the number of active firms (as explained above). For a given steady-state K^*, the slackness free entry condition may or may not hold exactly. If it does hold exactly then, in response to a marginal increase in c, the number of firms will necessarily decrease and so will the level of capital at the steady-state. If it does not hold exactly then the level of capital will be unchanged as no firm will leave the market. The statement of part a) follows.

Second, at an unstable steady-state K_U, the rental rate is increasing in the capital stock. For this to happen, $\Theta(\mathbf{\Gamma}, \mathbf{N}_t)$ must be increasing in K at that point. Assuming I large, this only happens if some firm is exactly breaking even. Therefore, the rental rate at an unstable steady-state K_U necessarily declines after an increase in c, as stated in part b).

■

A.4 Regression Tables

	(1)	(2)	(3)	(4)
VARIABLES	$\Delta \log \text{emp}_{07-16}$	$\Delta \log \text{emp}_{07-16}$	$\Delta \log \text{emp}_{07-16}$	$\Delta \log \text{emp}_{07-16}$
$\text{concentration}_{07}$	-0.0223***	-0.0160**	-0.0177***	-0.0178**
	(0.00667)	(0.00688)	(0.00682)	(0.00732)
$\log \text{firms}_{07}$		0.00239***	0.00193***	0.00151
		(0.000705)	(0.000706)	(0.000983)
$\Delta \log \text{emp}_{03-07}$			0.0984***	0.0901***
			(0.0241)	(0.0247)
Observations	770	770	769	761
R-squared	0.014	0.029	0.050	0.064
Sector FE	NO	NO	NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.2: Change in Employment: 2007-2016

Note: the table shows the results of regressing the growth rate of sectoral employment between 2007 and 2016 on the measure of concentration in 2007. The table presents the results of progressively adding controls and, in the last column, sector fixed effects.

VARIABLES	(1) Δ log payroll$_{07-16}$	(2) Δ log payroll$_{07-16}$	(3) Δ log payroll$_{07-16}$	(4) Δ log payroll$_{07-16}$
concentration$_{07}$	-0.0231***	-0.0177**	-0.0189***	-0.0194***
	(0.00679)	(0.00702)	(0.00697)	(0.00749)
log firms$_{07}$		0.00203***	0.00164**	0.000991
		(0.000724)	(0.000725)	(0.00101)
Δ log payroll$_{03-07}$			0.0823***	0.0697***
			(0.0219)	(0.0225)
Observations	774	774	773	765
R-squared	0.015	0.025	0.043	0.054
Sector FE	NO	NO	NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.3: Change in Total Payroll: 2007-2016

Note: the table shows the results of regressing the growth rate of sectoral total payroll between 2007 and 2016 on the measure of concentration in 2007. The table presents the results of progressively adding controls and, in the last column, sector fixed effects.

VARIABLES	(1) Δ log firms$_{07-16}$	(2) Δ log firms$_{07-16}$	(3) Δ log firms$_{07-16}$	(4) Δ log firms$_{07-16}$
concentration$_{07}$	-0.0432***	-0.0391***	-0.0406***	-0.0231***
	(0.00608)	(0.00637)	(0.00635)	(0.00666)
log firms$_{07}$		0.00137**	0.00119*	0.00449***
		(0.000663)	(0.000661)	(0.000897)
Δ log firms$_{03-07}$			0.0881***	0.0808***
			(0.0270)	(0.0273)
Observations	791	791	791	782
R-squared	0.060	0.065	0.078	0.151
Sector FE	NO	NO	NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.4: Change in Number of Firms: 2007-2016

Note: the table shows the results of regressing the growth rate of the industry number of firms between 2007 and 2016 on the measure of concentration in 2007. The table presents the results of progressively adding controls and, in the last column, sector fixed effects.

VARIABLES	(1) Δlabor share$_{07-16}$	(2) Δlabor share$_{07-16}$	(3) Δlabor share$_{07-16}$	(4) Δlabor share$_{07-16}$
concentration$_{07}$	-0.0314*	-0.0319*	-0.0314*	-0.0301
	(0.0167)	(0.0168)	(0.0167)	(0.0196)
log firms$_{07}$		-0.00111	-0.00120	-0.00255
		(0.00240)	(0.00240)	(0.00335)
Δlabor share$_{03-07}$			0.169*	0.146*
			(0.0867)	(0.0871)
Observations	99	99	98	97
R-squared	0.035	0.037	0.075	0.111
Sector FE	NO	NO	NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.5: Change in Labor Share: 2007-2016

Note: the table shows the results of regressing the growth rate of sectoral labor share between 2007 and 2016 on the measure of concentration in 2007. The table presents the results of progressively adding controls and, in the last column, sector fixed effects.

Appendix B

Supplementary Material for

Firm Heterogeneity, Market Power and Macroeconomic Fragility

Not for Publication

B.1 General Framework

In this section we provide a taxonomy of the possible mechanisms behind multiplicity in our economy. We then provide to prove an intermediate lemma used in Proposition 1 and conclude by specifying three Remarks to exemplify the effect of heterogeneity on the fragility of the economy.

We start by noting that a necessary condition for multiplicity of steady states is that $\exists K^* : R_K(K^*) > 0$. It follows that the necessary condition can be rewritten as $\exists K^* : (\partial/\partial K)\, \Omega(\Lambda, n(\Lambda, K^*))\Phi(\Lambda, n(\Lambda, K^*))F_K(K^*, L(\Lambda, n, K^*)) > 0$. There are three main mechanisms underlying the possible locally increasing returns to capital.

Average Firm TFP In an economy with heterogeneous technologies and no love for variety, aggregate TFP can be written as a weighted average of firm-level productivities. The weights will depend on market shares. Average firm-level TFP can be increasing in K if a larger capital favors a reallocation towards more productive types.

Love for Variety In models with product differentiation, aggregate TFP typically increases in the number of available varieties. This reflects the fact that utility/welfare are themselves increasing in the number of available goods (i.e. there is *love for variety*). Take for simplicity an economy where firms operate with the same level of productivity $\gamma_{ij} = \gamma$, but with possibly different fixed costs c_{ij}. Each firm produces a differentiated good. A larger capital stock K can increase the incentives for the entry of new firms/goods, thereby making Φ (weakly) increasing in K. Examples of papers highlighting this channel as a source of multiple equilibria/steady-states include Schaal and Taschereau-Dumouchel (2019).[36]

Market Power In models featuring imperfect competition and variable markups, changes in the number of active players can have an impact on the distribution of income across factors of production and oligopoly rents. Take for example an economy where firms have identical fixed costs $c_{ij} = c$ but possibly different productivities π_{ij}. Assume further that firms enter sequentially in reverse order of productivity. If profit levels are increasing in the aggregate capital stock k (for

[36] Without relying on multiple equilibria or multiple steady-states, Cooper and John (2000) and Bilbiie et al. (2012) show that a combination of imperfect competition with endogenous entry can generate endogenous amplification and persistence of aggregate fluctuations.

a given set of players), a larger capital stock will result in a larger number of firms and lower markups. Lower markups in turn translate in a higher factor share Ω. This can establish a positive relationship between Ω and k. Examples of papers highlighting this channel as a source of multiple equilibria/steady-states include Pagano (1990), Chatterjee et al. (1993), Galí and Zilibotti (1995) and Jaimovich (2007).

Let us consider some special cases. The first remark discusses a mean-preserving spread to the distribution of idiosyncratic productivities in an economy with a fixed set of producers (i.e. no adjustment along the extensive margin). More precisely, we consider special case of mean-preserving spread in the transformation is monotonically increasing away from the median in the set of active producers.[37] We then consider a marginal change in γ.

Remark B.1 (Allocative efficiency). *In an economy with a fixed set of active producers, a technological shift $d\lambda$ increases fragility if*

$$\Omega_\lambda(\Lambda, n)\Phi(\Lambda, n) + \Omega(\Lambda, n)\Phi_\lambda(\Lambda, n) < 0 \quad \textit{for } K = \{\mathcal{K}_n, \mathcal{K}_{n+1}\}, n \textit{ even.} \tag{81}$$

Suppose for example that we consider a mean-preserving spread to the distribution of idiosyncratic productivities. Assume further that markups and market shares are both positive functions of productivities. If market shares are an non-decreasing function of productivities, then the second term is non-negative. To see this note that a mean-preserving spread, fixing the market share distribution, implies that the average productivity increases. Additionally the positive relationship between productivities and market share implies a reallocation from low to high productivity firms, reinforcing the increase in aggregate productivity. Secondly, if markups are themselves non-decreasing in market shares, then the first term is always non-positive. This result comes from the reallocation effect. As large firms become larger, they compress output to extract higher rents. In doing so they compress the factor share. Therefore fragility increases if the anti-competitive effect (first term) dominates the efficiency gains (second term) from increasing the dispersion of firm-level productivity. Through the lens of the taxonomy the first term in Remark B.1 represents the market power channel, while the second term is the average firm TFP channel.

We are also interested in exploring the consequences of an increase in fixed costs. To simplify the exposition, suppose that the economy only contains one industry type ($I = 1$), that all producers have identical productivity and there is no love for variety. In that case, aggregate productivity Φ is fixed and changes in the equilibrium rental rate will only happen through the aggregate factor share Ω.

[37] For example, we consider going from Π to $\tilde{\Pi}_{\gamma,S_\Pi} = (1 + \gamma \cdot S_\Pi) \circ \Pi$, with $\gamma > 0$, S_Π monotonically decreasing within a row and row-wise zero-sum, and $\circ$ denoting the Hadamard product. For a similar approach see Herrendorf et al. (2000).

Remark B.2 (**Market power**). *Consider an economy with one industry type ($I = 1$), identical producers and no love for variety. This economy will feature constant aggregate TFP, which we normalize to $\Phi(\Lambda, n) = \Phi$. Furthermore, because firms are identical, larger fixed costs only affect the aggregate factor share through changes in the mass of active firms, i.e. $\Omega_\lambda(\Lambda, n) = 0$. Therefore, in this economy a larger fixed cost generates greater fragility if*

$$\Omega_n(\Lambda, n) n_\lambda(\Lambda, k) < 0 \quad \textit{for } k = \{\mathcal{K}_n, \mathcal{K}_{n+1}\}, n \textit{ even.} \tag{82}$$

First, note that the aggregate factor share should be non-decreasing in the aggregate mass of firms, i.e. $\Omega_n(\Lambda, n) \geq 0$. As firms exit due to the higher fixed cost the surviving firms increase their market shares. In doing so they are able to increase their markups and compress the factor share. Second, the aggregate mass of firms must be non-increasing in fixed costs $n_\lambda(\Lambda, k) \leq 0$. This effect comes from firms being unable to cover the increased fixed costs and exiting. Therefore, larger fixed costs should generate increase fragility.

We also consider an economy with a constant markups and factor shares, to highlight how changes in the mass of firms can affect aggregate TFP.

Remark B.3 (**Love for variety**). *Consider an economy with constant markups and aggregate factor share $\Omega(\Lambda, n) = \Omega$. A technological shift $d\lambda$ increases fragility if*

$$\Phi_\lambda(\Lambda, n) + \Phi_n(\Lambda, n) n_\lambda(\Lambda, k) < 0 \quad \textit{for } k = \{\mathcal{K}_n, \mathcal{K}_{n+1}\}, n \textit{ even.} \tag{83}$$

Suppose for example that we consider a mean-preserving spread to idiosyncratic productivities. In that case, if there is a reallocation towards more productive firms, we have $\Phi_\lambda(\Lambda, n) \geq 0$, as well as $\Phi_n(\Lambda, n) \geq 0$ (love for variety) and $n_\lambda(\Lambda, n) \leq 0$ (if less productive firms are driven out of the market). If the second effect dominates (i.e. loss in number of varieties is stronger than the increase in average technical efficiency), fragility increases.

B.2 Industry Equilibrium with $\eta = 1$

Equilibrium Price and Output Suppose that $\eta = 1$. When n firms produce, we have a system of n first order conditions

$$p\left[1 - (1 - \rho) s_j\right] = \frac{\Theta}{\gamma_j} \tag{84}$$

Dividing the first order condition of firm j by that of firm 1 we obtain

$$s_j = \frac{1}{(1 - \rho)} \left\{1 - \frac{\gamma_1}{\gamma_j} \left[1 - (1 - \rho) s_1\right]\right\} \tag{85}$$

Note that

$$\sum_{k=1}^{n} s_k = 1 \Rightarrow \frac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}} = \gamma_1\left[1-(1-\rho)\,s_1\right] \tag{86}$$

Plugging the last equation into the first order condition of firm 1 we obtain

$$p = \frac{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}{n-(1-\rho)}\Theta \tag{87}$$

Total output is hence equal to

$$y = p^{-\frac{1}{1-\rho}}Y = \left[\frac{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}{n-(1-\rho)}\Theta\right]^{-\frac{1}{1-\rho}} Y \tag{88}$$

Market Shares Plugging the previous equation into the first order condition of firm j we have

$$s_j = \frac{1}{1-\rho}\left[1-\frac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}\frac{1}{\gamma_j}\right] \tag{89}$$

It is easy to verify that each firm's market share decreases in the total number of active firms. To see this, suppose that the number of firms increases from n to $n+1$. The new entrant will have a market share

$$s_{n+1} = \frac{1}{1-\rho}\left[1-\frac{n+1-(1-\rho)}{\sum\limits_{k=1}^{n+1}\frac{1}{\gamma_k}}\frac{1}{\gamma_{n+1}}\right] \tag{90}$$

which is non-negative provided that

$$\gamma_{n+1}\sum_{k=1}^{n+1}\frac{1}{\gamma_k} > n+1-(1-\rho) \tag{91}$$

and below one given that

$$\gamma_{n+1}\sum_{k=1}^{n+1}\frac{1}{\gamma_k} < \frac{1}{\rho}\left[n+1-(1-\rho)\right] \tag{92}$$

If we compare the market share of firm j when there n and $n+1$ firms in the market, we have

$$s_j\,|_{n+1} < s_j\,|_n \Leftrightarrow \gamma_{n+1}\sum_{k=1}^{n+1}\frac{1}{\gamma_k} > n-(1-\rho) \tag{93}$$

Note that the last condition is implied by (91).

Profits When there are n active firms, type γ_j makes production profits

$$\Pi\left(\gamma_j, n, \Gamma, \Theta, Y\right) = \underbrace{\frac{1}{1-\rho}\left[1-\frac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}\frac{1}{\gamma_j}\right]^2\left[\frac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}\right]^{\frac{\rho}{1-\rho}}}_{\equiv\Lambda(\gamma_j,n,\Gamma)}\Theta^{-\frac{\rho}{1-\rho}}Y \tag{94}$$

Lemma B.1. *When $\eta=1$, the profit function $\Pi\left(j, n_{it}, \Gamma_i, X_t\right)$ satisfies*

$$1)\frac{\partial\Pi\left(j,n_{it},\Gamma_i,X_t\right)}{\partial Y_t} > 0 \qquad 2)\qquad \frac{\partial\Pi\left(j,n_{it},\Gamma_i,X_t\right)}{\partial n_{it}} < 0 \quad,\ n_{it} > j \tag{95}$$

$$3)\frac{\partial\Pi\left(j,n_{it},\Gamma_i,X_t\right)}{\partial \gamma_{ij}} > 0 \qquad 4)\qquad \frac{\partial\Pi\left(j,n_{it},\Gamma_i,X_t\right)}{\partial \gamma_{ik}} < 0 \quad,\ \forall k\neq j. \tag{96}$$

Proof of Lemma B.1. We start by showing that $\Pi\left(\cdot\right)$ increases in γ_j

$$2\left[1-\frac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}\frac{1}{\gamma_j}\right]^{-1}\left\{-\frac{-\left[n-(1-\rho)\right]\left[-\left(\frac{1}{\gamma_j}\right)^2\right]}{\left(\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}\right)^2}\frac{1}{\gamma_j}+\frac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}\left(\frac{1}{\gamma_j}\right)^2\right\}+ \tag{97}$$

$$\frac{\rho}{1-\rho}\left[\frac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}\right]^{-1}\frac{-\left[n-(1-\rho)\right]\left[-\left(\frac{1}{\gamma_j}\right)^2\right]}{\left(\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}\right)^2} > 0 \tag{98}$$

$$\Leftrightarrow 2\left[1-\frac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}\frac{1}{\gamma_j}\right]^{-1}\left(\sum_{k\neq j}^{n}\frac{1}{\gamma_k}\right)+\frac{\rho}{1-\rho}\left[\frac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\frac{1}{\gamma_k}}\right]^{-1} > 0 \tag{99}$$

To prove points (ii) and (iii) it suffices to show that $\Lambda(\cdot)$ is decreasing in $[n-(1-\rho)]/\left[\sum_{k=1}^{n}\frac{1}{\gamma_k}\right]$

$$2\left[1-\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\gamma_k}}\frac{1}{\gamma_j}\right]^{-1}\left(-\frac{1}{\gamma_j}\right)+\frac{\rho}{1-\rho}\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\gamma_k}}\right]^{-1}<0 \tag{100}$$

$$\tag{101}$$

$$\Leftrightarrow \gamma_j\sum_{k=1}^{n}\frac{1}{\gamma_k}<\frac{2-\rho}{\rho}\left[n-(1-\rho)\right] \tag{102}$$

The last condition is implied by (92). ■

B.3 Derivations: General Equilibrium

B.3.1 Aggregate TFP

Aggregate TFP is given by

$$\Phi(\mathbf{\Gamma},\mathbf{N}_t)=\left[\sum_{i=1}^{I}\left(\sum_{j=1}^{n_{it}}\omega_{ijt}^{\eta}\right)^{\frac{\rho}{\eta}}\right]^{\frac{1}{\rho}}\left(\sum_{i=1}^{I}\sum_{j=1}^{n_{it}}\frac{\omega_{ijt}}{\tau_{ijt}}\right)^{-1}, \tag{103}$$

where

$$\omega_{ijt}\coloneqq\left[\sum_{k=1}^{n_{it}}\left(\frac{\mu_{ikt}}{\tau_{ikt}}\right)^{\frac{\eta}{1-\eta}}\right]^{\frac{\eta-\rho}{\eta}\frac{1}{1-\rho}}\left(\frac{\tau_{ijt}}{\mu_{ijt}}\right)^{\frac{1}{1-\eta}}. \tag{104}$$

B.3.2 Factor Prices and Factor Shares

We can aggregate firms' best responses, given by equation (11), to find an expression for the aggregate factor cost index. Given a $(I\times M)$ matrix of productivity draws $\mathbf{A}_t$ and a vector of active firms $\mathbf{N}_t\equiv\{n_{it}\}_{i=1}^{I}$, the equilibrium factor cost index is equal to

$$\Theta(\mathbf{A}_t,\mathbf{N}_t)=\left\{\sum_{i=1}^{I}\left[\sum_{j=1}^{n_{it}}\left(\frac{\tau_{ijt}}{\mu_{ijt}}\right)^{\frac{\eta}{1-\eta}}\right]^{\frac{1-\eta}{\eta}\frac{\rho}{1-\rho}}\right\}^{\frac{1-\rho}{\rho}}. \tag{105}$$

The aggregate factor share $\Omega(\cdot)=(W_t\,L_t+R_t\,K_t)/Y_t$ is equal to

$$\Omega(\mathbf{A}_t,\mathbf{N}_t)=\frac{\Theta(\mathbf{A}_t,\mathbf{N}_t)}{\Phi(\mathbf{A}_t,\mathbf{N}_t)}. \tag{106}$$

B.3.3 Asymmetric Equilibrium

When

$$\overline{K}\left(\Gamma,n\right) < K < \underline{K}\left(\Gamma,n+1\right) \tag{107}$$

there will be an asymmetric equilibrium at time $t+1$: some industries will contain n firms, whereas some industries will contain $n+1$ firms. The fraction of industries with $n+1$ will be pinned down by a zero profit condition for the marginal entrant in an industry with $n+1$ firms

$$\Lambda\left(\Gamma,\gamma_{n+1},n+1\right)\Theta^{-\frac{\rho}{1-\rho}}Y = c_i \tag{108}$$

The equilibrium is characterized by 4 variables: the fraction of the industries with $n+1$ firms (η), aggregate output (Y), aggregate productivity (Φ) and the aggregate cost index (Θ). These 4 variables are pinned down by the following 4 equations

$$Y = \Phi\left[\left(1-\alpha\right)\Theta\right]^{\frac{1-\alpha}{\nu+\alpha}} K^{\alpha\frac{1+\nu}{\nu+\alpha}} \tag{109}$$

$$\Phi = \frac{\left\{\left(1-\eta\right)\left[\frac{n-\left(1-\rho\right)}{\sum_{k=1}^{n}\frac{1}{\gamma_{1k}}}\right]^{\frac{\rho}{1-\rho}} + \eta\left[\frac{n+1-\left(1-\rho\right)}{\sum_{k=1}^{n+1}\frac{1}{\gamma_{2k}}}\right]^{\frac{\rho}{1-\rho}}\right\}^{\frac{1}{\rho}}}{\left(1-\eta\right)\left[\frac{n-\left(1-\rho\right)}{\sum_{k=1}^{n}\frac{1}{\gamma_{1k}}}\right]^{\frac{1}{1-\rho}}\left(\sum_{k=1}^{n}\frac{s_{1k}}{\gamma_{1k}}\right) + \eta\left[\frac{n+1-\left(1-\rho\right)}{\sum_{k=1}^{n+1}\frac{1}{\gamma_{2k}}}\right]^{\frac{1}{1-\rho}}\left(\sum_{k=1}^{n+1}\frac{s_{2k}}{\gamma_{2k}}\right)} \tag{110}$$

$$\Theta = \left\{\left(1-\eta\right)\left[\frac{n-\left(1-\rho\right)}{\sum_{k=1}^{n}\frac{1}{\gamma_{k}}}\right]^{\frac{\rho}{1-\rho}} + \eta\left[\frac{n+1-\left(1-\rho\right)}{\sum_{k=1}^{n+1}\frac{1}{\gamma_{k}}}\right]^{\frac{\rho}{1-\rho}}\right\}^{\frac{1-\rho}{\rho}} \tag{111}$$

$$\Lambda\left(\Gamma,\gamma_{n+1},n+1\right)\Theta^{-\frac{\rho}{1-\rho}}Y = c_i \tag{112}$$

s_{1k} is the market share of firm k in an industry with n firms, whereas s_{2k} is the market share of firm k in an industry with $n+1$ firms. They are defined in Appendix B.2.

B.4 The Baseline Model

Comparative Statics with $\uparrow \gamma_1$ and $\leftrightarrow \gamma_2$ (mean-increasing spread)

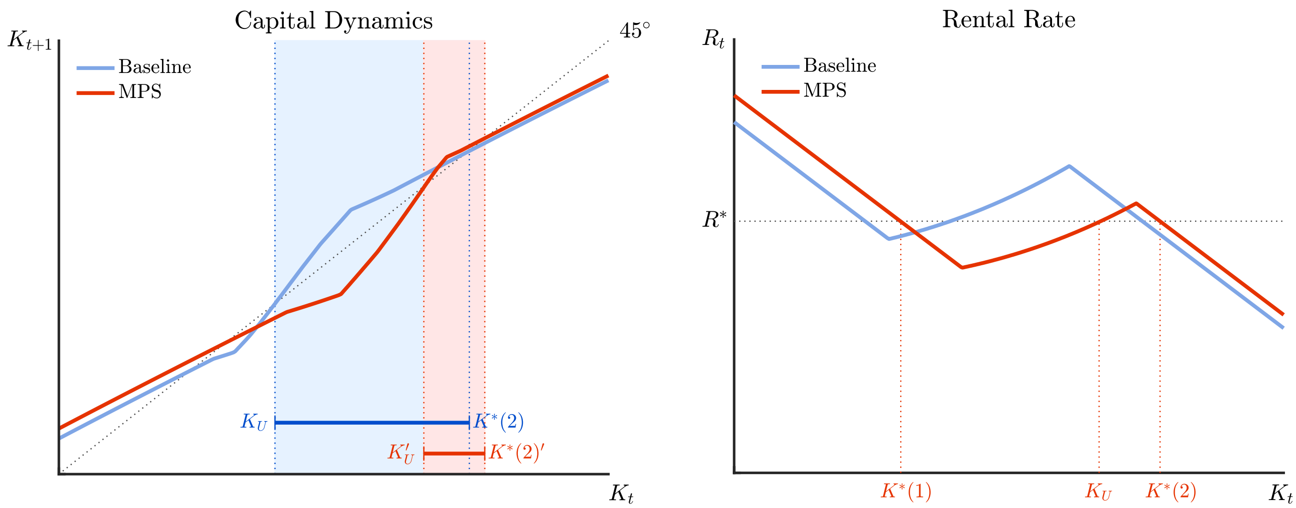


Figure B.1: Law of Motion and Rental Rate Map

This example features two stable steady states and an unstable one. We use $\psi = 1$, $\rho = 3/4$, $\eta = 1$, $\alpha = 1/3$, $\delta = 1$, $\nu = 2/5$ and $c_i = 0.015$.

Steady-State Multiplicity

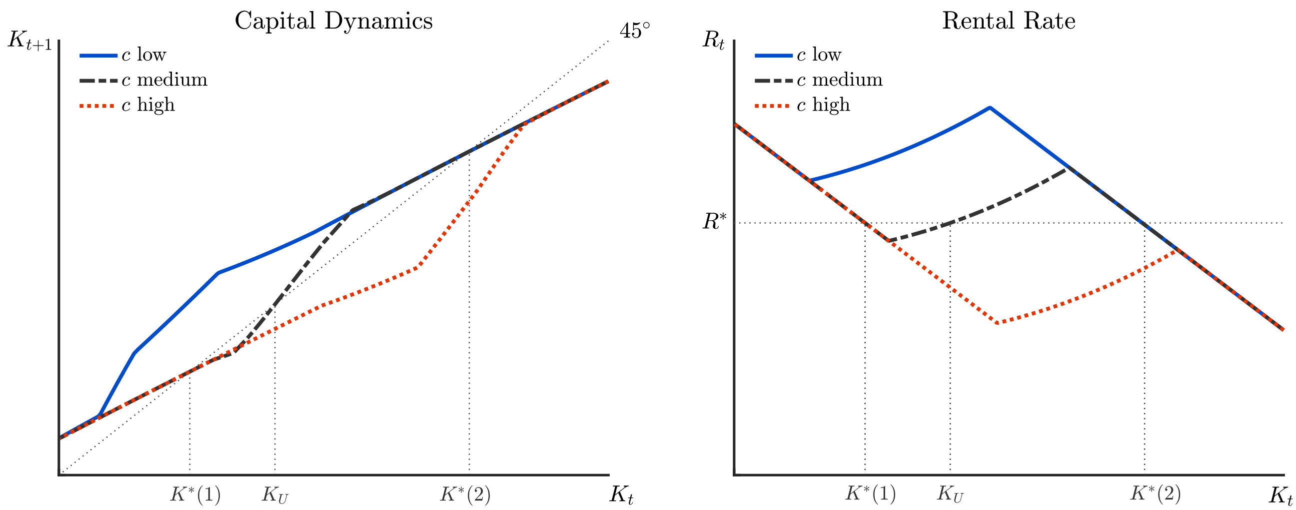


Figure B.2: Law of Motion and Rental Rate Map

This example features two stable steady states and an unstable one. We use $\psi = 1$, $\rho = 3/4$, $\eta = 1$, $\alpha = 1/3$, $\delta = 1$ and $\nu = 2/5$.

B.5 The Quantitative Model

B.5.1 Calibration

Steady-State We perform three different calibrations of our model – to match the average level of markups and its dispersion in 1975, 1990 and in 2007. We need to calibrate five technology parameters: the elasticity of substitution σ_I and σ_G (which are time-invariant), the log-normal standard deviation λ, the fixed production cost c and the fraction of industries with zero fixed cost f_{comp} (which are allowed to vary over time).

We start by specifying a grid with possible values for σ_I and σ_G. Then, for each pair (σ_I, σ_G), we specify a triplet (λ, c, f_{comp}), as well as a grid with values for the aggregate capital stock K. We then compute the aggregate equilibrium for each parameter combination $(\sigma_I, \sigma_G, \lambda, c, f_{comp})$ and for each value K.[38] We start by assuming that all firms are active, so that there are N firms in each of the I industries. We compute the aggregate equilibrium using equations (103) and (19). We then compute the profits net of the fixed cost that each firm makes

$$\left(p_{ijt} - \frac{\Theta_t}{\tau_{ijt}}\right) y_{ijt} - c_i$$

and identify the firm with the largest negative value. We exclude this firm and recompute the aggregate equilibrium. We repeat this iterative procedure until all firms have non-negative profits (net of the fixed production cost). If equilibrium multiplicity arises, this algorithm allows us to consistently select the equilibrium that features the largest number of firms.

For each triplet (λ, c, f_{comp}), we then have the general equilibrium computed for all possible capital values. The steady-state(s) of our economy correspond to the value(s) of K for with the rental rate R_t is equal to $\frac{1}{\beta} - (1 - \delta)$.

When multiple steady-states arise (as in the 1990 and 2007 economies), we compute model moments in the highest steady-state.

Data Definitions For the sales weighted-average markup, we use the series computed by De Loecker et al. (2020). The authors calculate price-cost markups for the universe of public firms, using data from COMPUSTAT. The markup of a firm j in a 2-digit NAICS sector s at time t is calculated as

$$\mu_{sjt} = \xi_{st} \cdot \frac{\text{sale}_{sjt}}{\text{cogs}_{sjt}}$$

where ξ_{st} is the elasticity of sales to the total variable input bundle, sale_{sjt} is sales and cogs_{sjt} is the cost of the goods sold, which measures total variable costs.

[38] Aggregate TFP e^{z_t} is assumed to be constant and equal to one.

B.5.2 Solution Algorithm for the Dynamic Optimization Problem

We now explain the algorithm we use for the dynamic optimization problem of the representative household. We take the calibrated parameters (λ, c) and form a grid for aggregate capital with $n_K = 70$ points. This grid is centered around the highest steady-state K_H^{ss}, with a lower-bound $0.5 \times K_H^{\text{ss}}$ and upper bound $1.5 \times K_H^{\text{ss}}$. We also form a grid for aggregate TFP, A. We use Tauchen's algorithm with $n_A = 11$ points, autocorrelation parameter ϕ_A and standard deviation for the innovations σ_ε (the last two parameters are calibrated, as explained in the main text). We compute the aggregate equilibrium for each value of K and A.

We next compute a numerical approximation for the household policy function, by iterating on the Euler equation. We start by making a guess for the savings rate

$$s\left(X_t\right) \coloneqq \frac{C\left(X_t\right)}{Y\left(X_t\right)}$$

for every combination of the vector of state-variables $X_t \coloneqq (K_t, A_t)$. Given a guess $s^{(n)}\left(X_t\right)\ \forall X_t$ for the savings rate, we use the Euler equation to obtain a new guess $s^{(n+1)}\left(X_t\right)$ as follows

$$\frac{1}{\left(1 - s^{(n+1)}\left(X_t\right)\right) Y\left(X_t\right) - \dfrac{W\left(X_t\right)^{(1+\nu)/\nu}}{1+\nu}} = \mathbb{E}_t \left\{ \frac{\beta\left[R\left(X_{t+1}\right) + (1-\delta)\right]}{\left(1 - s^{(n)}\left(X_{t+1}\right)\right) Y\left(X_{t+1}\right) - \dfrac{W\left(X_{t+1}\right)^{(1+\nu)/\nu}}{1+\nu}} \right\}$$

$$\Leftrightarrow \quad s^{(n+1)}\left(X_t\right) = 1 - \frac{1}{Y\left(X_t\right)} \left\{ \frac{W\left(X_t\right)^{(1+\nu)/\nu}}{1+\nu} + \left[\mathbb{E}_t \left\{ \frac{\beta\left[R\left(X_{t+1}\right) + (1-\delta)\right]}{\left(1 - s^{(n)}\left(X_{t+1}\right)\right) Y\left(X_{t+1}\right) - \dfrac{W\left(X_{t+1}\right)^{(1+\nu)/\nu}}{1+\nu}} \right\} \right]^{-1} \right\}.$$

We iterate on this procedure until

$$\left| s^{(n+1)}\left(X_t\right) - s^{(n)}\left(X_t\right) \right| \ < \ \epsilon \quad \forall X_t.$$

B.5.3 Business Cycle Moments

	Output	Consumption	Investment	Hours	TFP
		Correlation with Output			
Data: 1947-2019	1.00	0.95	0.76	0.67	0.71
Model: 1975 calibration	1.00	0.99	0.92	1.00	0.83
Model: 1990 calibration	1.00	1.00	0.96	1.00	0.88
Model: 2007 calibration	1.00	1.00	0.96	1.00	0.89
		Standard Deviation Relative to Output			
Data: 1947-2019	1.00	0.90	2.04	0.98	0.95
Model: 1975 calibration	1.00	0.95	1.54	0.74	0.25
Model: 1990 calibration	1.00	0.97	1.23	0.78	0.17
Model: 2007 calibration	1.00	0.98	1.20	0.79	0.15

Table B.1: Business Cycle Moments. All variables are in logs. Data variables are in per capita terms (except TFP) and in deviation from a linear trend computed over 1947-2007.

Table B.1 shows some business cycle moments for our two calibrated economies, as well as their data counterparts. To be consistent with our interpretation that the US economy transitioned to a lower steady-state after 2008, all data variables are in deviation from a linear trend computed over 1947-2007. This fact explains the large empirical correlation between consumption and output. Comparing our two calibrated economies, we see that both economies display the same correlations of consumption and hours with output. The 2007 economy displays, however, a significantly lower correlation of investment with output.

B.5.4 Aggregate Productivity

Average Firm Level TFP

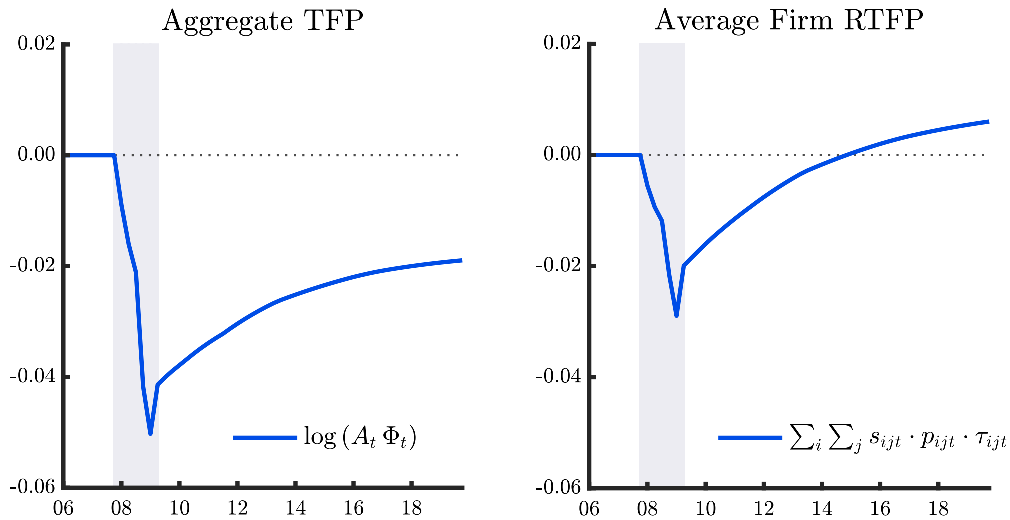


Figure B.3: Aggregate TFP versus Average Firm Level TFP

Note: The left panel shows aggregate TFP. The right panel shows a sales-weighted average of firm level revenue TFP $p_{ijt} \cdot \tau_{ijt}$.

Figure B.3 reports a sales-weighted average of firm level revenue TFP. A similar pattern emerges if one uses physical TFP instead.

Dispersion in Industry Output

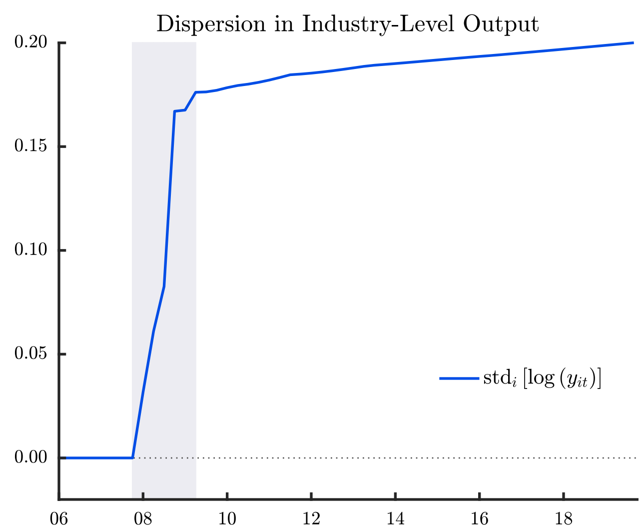
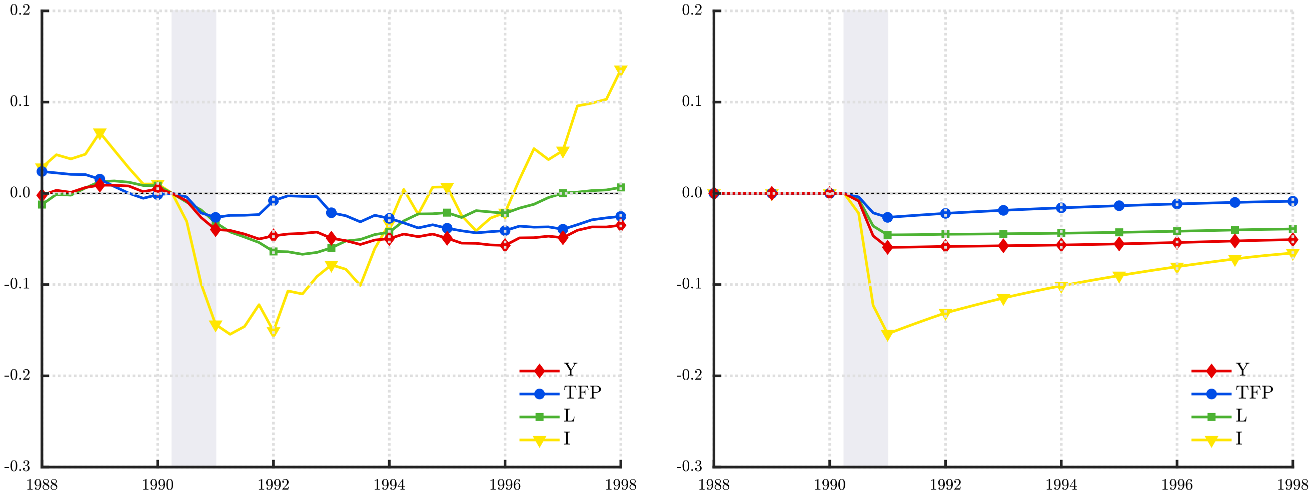


Figure B.4: Dispersion in $\log(y_{it})$

B.6 The 1990 Recession

The response in the 1990 economy



(a) 1990-1991 recession (data)

(b) The 1990-1991 shock in the 1990 model

Figure B.5: The 1990-1991 recession

The response in the 2007 economy

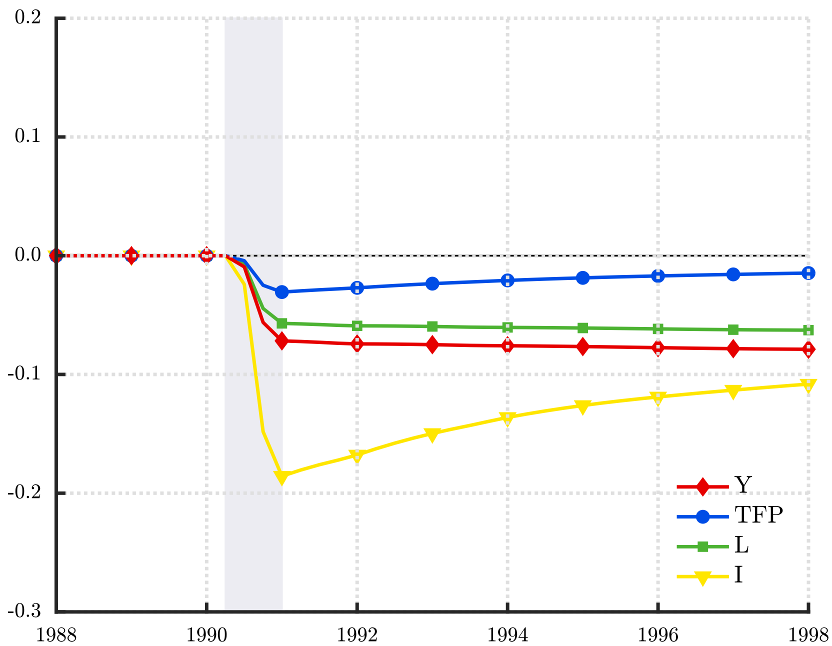


Figure B.6: The 1990-1991 shock in the 2007 model

B.7 Number of Firms per Sector

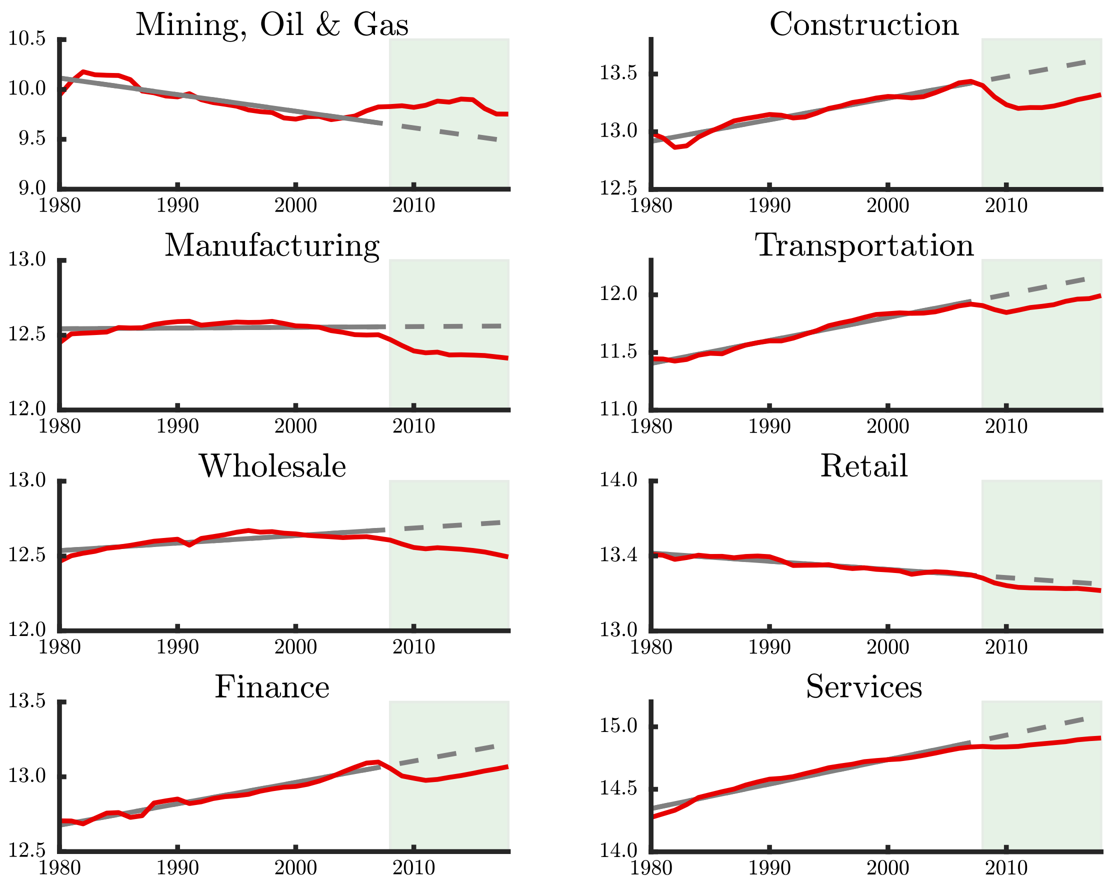


Figure B.7: **Number of Firms per Sector: 1980-2018**

Each panel shows the number of firms with at least one employee in each sector (in logs). For each series, the dashed grey line shows a linear trend computed over the 1980-2007 period. Data is from the US Business Dynamics Statistics

B.8 Fixed Costs

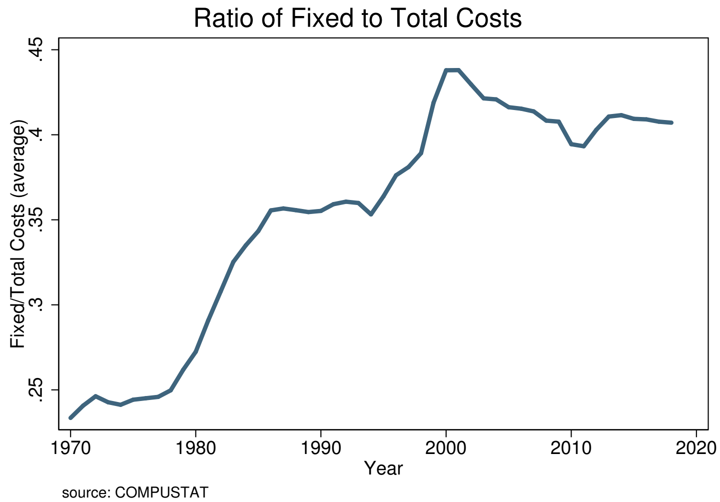


Figure B.8: Ratio of fixed to total costs
This figure shows the average ratio of fixed to total costs for COMPUSTAT firms. Following Gorodnichenko and Weber (2016), we define fixed costs as the sum of 'Selling, General and Administrative Expenses' (COMPUSTAT item XSGA), 'Advertising Expenses' (Compustat item XAD) and 'R&D Expenditures' (Compustat item XRD). Total costs are the sum of fixed costs and variable costs, where the latter correspond to the 'Cost of Goods Sold' (Compustat item COGS).

B.9 Robustness: Variable Fixed Costs

We assume that, each period, a fixed amount c_f of firms' output is lost

$$c_f = k_c^{\alpha} l_c^{1-\alpha}$$

Given these assumptions, firms need to pay a per per period fixed cost

$$\Theta_t \cdot c_f$$

where Θ_t is the factor price index.

Denoting by L_{yt} and K_{yt} the aggregate stocks of labor and capital used in the production, we have the following market clearing conditions for labor and capital

$$L_t = L_{yt} + N_t^c \cdot l_c$$
$$K_t = K_{yt} + N_t^c \cdot k_c$$

where N^c denotes the number of firms incurring c_f. Note that the optimal mix of l_c and k_c chosen by each individual firm satisfies

$$\frac{k_c}{l_c} = \frac{K_{yt}}{L_{yt}}$$

Calibration

Parameters not reported are as in the baseline calibration (Table 1).

Description	Parameter	Value			Source/Target
[A.1] Calibrated Parameters: Fixed					
Between-industry ES	σ_I		1.33		Sales-weighted average markup
Within-industry ES	σ_G		12.5		Sales-weighted average markup
Persistence of z_t	ρ_z		0.950		Autocorrelation of log Y_t
Standard deviation of ε_t	σ_ε		0.003		Standard deviation of log Y_t
[A.2] Calibrated Parameters: Variable		1975	1990	2007	
Fraction of industries with $c_i > 0$	f	0.105	0.140	0.130	Emp share concentrated industries
Standard deviation of γ_{ij}	λ	0.193	0.252	0.304	Std log market share
Fixed cost ($\times 10^{-4}$)	c	3.03	5.66	7.82	Average ratio fixed/total costs

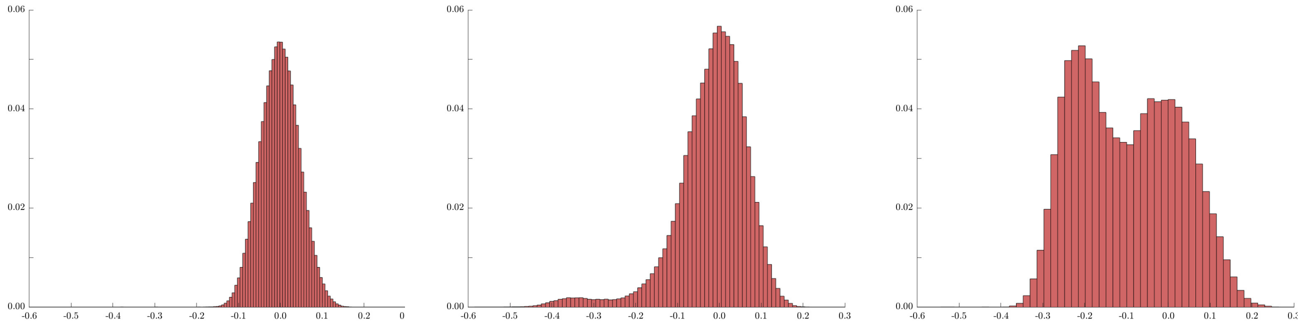
Table B.2: Parameter Values

	1975		1990		2007	
	Data	Model	Data	Model	Data	Model
Sales-weighted average markup	1.28	1.26	1.37	1.38	1.46	1.44
Std log revenues	1.67	1.74	2.47	2.36	2.79	2.92
Average fixed to total cost ratio	0.244	0.259	0.355	0.390	0.414	0.420
Emp share *concentrated* industries	-	0.067	-	0.072	0.063	0.060
Autocorrelation log GDP					0.978*	0.976
Standard deviation log GDP					0.061*	0.056

*computed over 1947:Q1-2019:Q4

Table B.3: Targeted moments and model counterparts

Ergodic Distributions

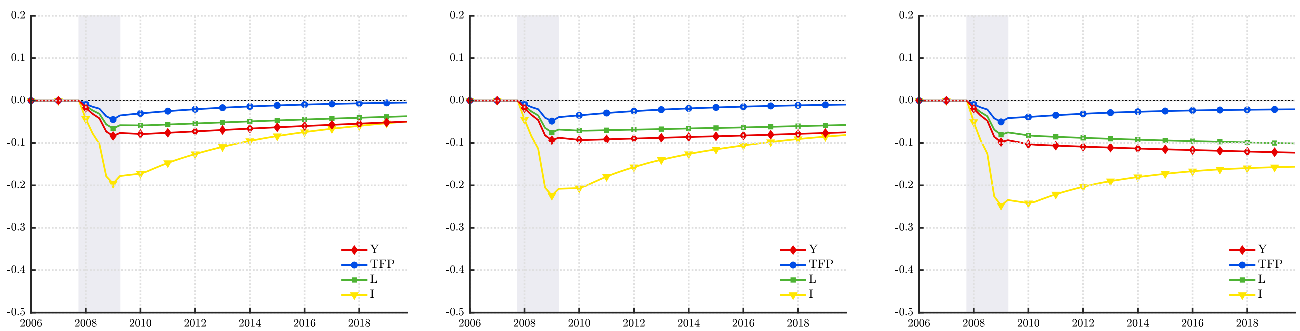


(a) 1975 (b) 1990 (c) 2007

Figure B.9: Ergodic distribution of output

Note: This figure shows the distribution of log output for the 1975, 1990 and the 2007 economies. We simulate each economy for 10,000,000 periods and plot output in deviation from the high steady state.

The 2008 Crisis



(a) 1975 Model (b) 1990 Model (c) 2007 Model

Figure B.10: The *great recession* in the 1975, 1990 and 2007 models

Note: The 2007 model is subjected to a sequence of six quarter shocks $\{\varepsilon_t\}$ to match the dynamics of aggregate TFP in the data between 2008Q1:2009Q2. This sequence of shocks is then fed in the 1975 and 1990 economies.