Endogenous Growth and Optimal Market Power [*]

Maria Voronina[†]

This version: November 11, 2021
Click here for the latest version

Abstract

We analyze the welfare effects of producer market power in frameworks with free entry and endogenous technical change. We show that the social planner cannot move the economy to the social optimum by simply eliminating firm market power in such settings. Thus, we suggest a second-best equilibrium concept that separates the welfare effects of markups from the impact of externalities generated by firms' investment and the inefficiencies associated with the decentralized entry. In addition, we decompose the distance to the first-best allocation into terms that separately measure the costs of sub-optimal markup distribution and sub-optimal investment policies. Our estimates indicate that the welfare losses due to market power are significant: the social planner can increase welfare by 20% by resetting markups to their socially optimal values. Sub-optimal markup distribution also accounts for 61% of the distance to the first best. We also analyze the evolution of misallocation in the US economy over the last four decades. We show that welfare costs of market power did not change significantly from 1980 to 2017. In order to calibrate our model, we re-estimate markups using a methodology that delivers consistent estimates under endogenous product prices and technical change. We find that the standard methodology underestimates the upward trend in markups by 5-10%. Our results suggest that the average cost-weighted markup in the US economy has increased by 19-24% over the last three decades.

---

[*] I am incredibly grateful to my advisors Pol Antras, Elhanan Helpman, and Marc Melitz, for their guidance and support. I am indebted to Emmanuel Farhi, who passed away in July 2020, for his help and insightful comments on the topic of this project. I also thank Martin Aragoneses, Gabriel Chodorow-Reich, Xavier Gabaix, Kenneth Rogoff, Ludwig Straub, Gabriel Unger, and all participants of the Harvard International Lunch and Harvard Macro Lunch for the helpful discussion.

[†] Email: mvoronina@g.harvard.edu. Harvard University, Department of Economics.

1. Introduction

> *“We do not live on the Pareto frontier, and we are not going to do so in the future. Yet policy decisions are constantly being made which can move us either toward or away from that frontier.”*
>
> Harberger (1964), *The Measurement of Waste*

Recent studies have documented several secular trends that are consistent with an increase in the market power of large US companies. Among these trends are the rise in concentration, the fall of the labor share, the decline in business dynamism, and the increase in markups. These findings raise concerns about the increasing influence of superstar firms in the US economy, and assessing the effects of the rise of large firms is a non-trivial task. To evaluate the welfare costs of market power, one needs to account for a variety of channels through which markups affect social welfare. Since Smith [1776], we know that the profit-maximizing behavior of large companies is rarely beneficial for the society in partial equilibrium models. However, the studies that analyze the economies with multiple sources of inefficiencies – e.g., the settings with free entry, economies of scale, or endogenous technological progress – offer a more nuanced perspective. A *certain* degree of market power might be beneficial for consumers because markups can counteract the effects of other frictions. For example, [Aghion et al., 2005] provide evidence in favor of an inverted-U relationship between competition and productivity growth. Although faster productivity growth does not necessarily generate an increase in consumer utility, we would expect Aghion et al. [2005] result to hold for welfare, as long the one-period output does not increase too steeply in the degree of competition. Relatedly, once we leave aside partial equilibrium settings, the impact of markups on welfare cannot be measured by computing the distance to the Pareto-efficiency frontier. In order to accurately evaluate the social cost of market power, we need to separate it from the effects of other inefficiencies.

This study’s primary purpose is to analyze the welfare effects of producer market power in the settings that feature endogenous technological progress and business dynamics. We start our analysis by building a general framework that flexibly models market power, innovation, and entry. We consider a discrete-time economy with a continuum of oligopolistic sectors. In this setting, product lines differ in relative productivity and relative stocks of fixed assets. Firms are heterogeneous in terms of the number of products they own and their goods’ types. Firm types follow a general Markov process, and their inter-temporal dynamics depend on the firm-level investment. The production process in our model involves both static inputs, represented by production labor and dynamic inputs. Overall, the setup of our model is quite general, and it encompasses many existing growth frameworks, including Romer [1990], Grossman and Helpman [1991b], and Klette and Kortum [2004].

In this project, we analyze the behavior of economies that move along balanced growth paths. On the decentralized balanced growth path, firms determine the investment and production employment levels, and consumers set the demand for final goods. In the first-best equilibrium, the social planner allocates dynamic and variable inputs to maximize social welfare, subject to technology constraints. To separate the impact of markups from the effects of other frictions, we also construct a second-best equilibrium that preserves the structure of the decentralized balanced growth path and, at the same time, features a ”socially-optimal” assignment of market power. The optimal distribution of market power is a distribution of markups that maximizes social welfare conditional on the laissez-faire industrial policy, without investment or entry subsidies. We then evaluate the welfare losses due to ”sub-optimal” market power by computing the distance between the decentralized allocation and the allocation achieved under the second-best optimum. This theoretical exercise is similar to the analysis performed in Dixit and Stiglitz [1977] as well as Baqaee and Farhi [2020b] and Baqaee and Farhi [2020a], who evaluate welfare losses due to misallocation in static settings. Importantly, the second-best optimum construct allows us to separate the welfare effects of market power from the effects of other frictions that include the entry wedge and various pecuniary and non-pecuniary externalities. In addition, we also consider a balanced growth path equilibrium in which the social planner can only alter investment values, but not the distribution of markups across final goods. The distance between this alternative second-best allocation and the first best is another measure of the importance of markups.

We show that in our framework, all balanced growth path allocations can be characterized concisely in terms of the equilibrium Markov chain transition kernel, the entrant type distribution, and variable surplus function that summarizes the results of short-run firm optimization. These objects comprise the set of sufficient statistics that define the reaction of social welfare, output, and productivity growth to friction and technology shocks in the models with endogenous productivity and input dynamics. We then describe the decentralized equilibrium, the first-best and second-best allocations in detail. The social planner can implement the first best only if they can alter both the values of markups and investment allocation. In the first-best equilibrium, markups are constant across producers, and their value depends only on the elasticity of the aggregate welfare with respect to producer mass and the aggregate profit rate. The social return of firms’ investment depends primarily on the elasticity of the equilibrium firm-type density with respect to R&D, capital expenditures, and expenditures on intangibles. Importantly, our analysis demonstrates that the social return to firms’ investment is generically not equal to the private return, even in the absence of knowledge spillovers and other non-pecuniary externalities. On the second-best balanced growth path, the social planner uses markups to reduce investment and variable inputs misallocation. The resulting markup levels differ across the producers. The deviation of the second-best markups from their average level depends on social and private returns on the producer’s investment.

Apart from computing the distances between the first-best, second-best and decentralized equilibria, we also endeavor to find out *why* market power causes misallocation. In other words, we analyze the channels through which markups affect welfare. First, we decompose the welfare losses due to market power into two welfare differentials associated with the misallocation of static and dynamic inputs, i.e., production labor and investment. This exercise allows us to directly compare the predictions of our model to the results of studies that feature static settings or dynamic settings without endogenous productivity dynamics or dynamic inputs. Furthermore, we show that investment misallocation affects welfare via four different mechanisms. First, lower investment levels have a direct effect on aggregate productivity growth and capital stock. Second, in the settings with product innovation, under-investment decreases the number of varieties available to consumers. The distribution of firm-level investment also determined the distribution of firm and product types. Thus, changes in firms’ R&D intensity or capital expenditure lead to the reallocation of sales and output across firm types. Similarly, changes in the type distributions also affect the allocation of variable inputs. We examine the importance of all these channels for the aggregate misallocation and the social costs of market power. We also provide a first-order policy-based decomposition for the distance to the first-best allocation: we separately evaluate the welfare effects of the ”sub-optimal” markup distribution and the sub-optimal investment policies.

To quantify our theoretical results, we calibrate our model using firm-level Compustat data, BDS data on entry and exit of firms, and aggregate data on output and population growth in the US. We want to analyze the evolution of misallocation in the US economy, and thus we consider two data samples that correspond to the early (1982-1997) and late (2002-2017) periods. The results of our counterfactual exercises are as follows. First and foremost, our analysis indicates that the costs of producer markups are high, regardless of the time period or the method used to measure the social cost or market power. For the late data sample, the social planner can increase social welfare by 20% by resetting markups to their second-best levels. The second-best equilibrium in which the social planner can alter investment also generates a 20% increase in social welfare. Moreover, misallocation due to market power accounts for 61% of the distance to the Pareto-efficiency frontier in the late sub-period. This means that conditional on the socially-optimal investment allocation, the welfare cost of sub-optimal markup distribution is equal to 22%. The second-best results for the early period are similar. The distance between the benchmark second best and the decentralized allocation is equal to 21%. If the social planner sets investment subsidies instead of markups, they can improve their objective by 18%. In contrast, the first-order decomposition of the distance to first best indicates that, in the early sub-sample, market power accounts only for 17% of total misallocation. Equivalently, the social cost of markups is equal to 6%.

Notably, in both sub-periods, the average values of the second-best markups are higher than in the decentralized equilibrium, and so is the markup variance. A closer look at the second-best markup values reveals that the mechanisms that determine the socially-optimal market power assignment differ across time periods despite similar magnitudes of welfare differentials. In the 1982-1997 sub-sample, the social planner uses markups primarily to ensure that entry and the allocation of variable factors of production are efficient. The resulting markup values are negatively correlated with the product TFP and capital stock and positively correlated with the entry rates into respective product types. Such a distribution of market power encourages entry and increases large companies’ cost shares and relative output. In contrast, in the late sub-sample, the social planner sets markups to induce firms to invest more in productivity and intangibles. In this case, the second-best markups increase with firm size and product state variables (TFP and capital). Accordingly, the primary source of second-best welfare gains in the late sub-sample is increased productivity growth, not the reallocation between production and entry. We hypothesize that the origin of these differences across periods is the increase in the importance of intangible assets in the determination of firm productivity and the subsequent rise in TFP dispersion. In our calibrations, the degree of under-investment rises significantly between the early and late sub-samples. Thus, the motivation of the social planner changes when they decide on the second-best markup values. It is likely that these differences in the production and innovation processes also determined the differences in the results of the first-best decomposition. The higher markups we observe in the late sub-sample are indeed more costly for society if the social planner can separately reset the investment rates to their optimal values.

A secondary contribution of this project consists of re-estimating the marginal[1] markup levels for the US economy using a methodology that generates consistent estimates of output elasticities in the presence of market power and endogenous TFP dynamics. Following Griliches and Regev [1995], Bond et al. [2021] show that conventional production function estimation methods deliver biased estimates whenever firms' sales are used as a measure of physical output and whenever firms set prices strategically. Moreover, while the standard proxy function approach can resolve the issue of the output price bias, it is challenging to construct a perfect proxy for output prices even in the simplest settings with Cobb-Douglas production and CES demand. In particular, we show that the proxy function strategy employed by De Loecker et al. [2020] is not valid in such settings. In order to derive consistent estimates of marginal markups, we employ the methodology that builds on the insights of Griliches and Regev [1995], De Loecker et al. [2016] and Gandhi et al. [2020]. Specifically, we use intertemporal and cross-sectional variation in the sectoral output levels at a 5-digit industry level to pin down the elasticity of substitution between products. In addition, we allow for the endogenous evolution of firm productivity in our production function estimation routine. Our results indicate that over the period from 1980 to 2017, the average cost-weighted markup increased by about 19-24% percentage points, as compared with a 14% increase suggested by the standard De Loecker et al. [2020] methodology. The output price bias has a modest effect on the markup estimates and their dynamics. In contrast, incorporating the endogenous determinants of firm productivity amplifies the markup trend by 2-6% depending on the specification.

1.1. Literature Review

This paper is connected to several strands of macro and IO research that focus on the phenomenon of market power, its origins and consequences.

[Welfare Effects of Market Power: Settings w/o Growth] The economic literature that relates producer market power to welfare outcomes has an impressively long history. The notion that the monopolist's behavior is harmful to consumers goes back at least to Smith [1776], and since then numerous studies, including Lerner [1934], Harberger [1954], and Dixit and Stiglitz [1977], have contributed to the field. In the model with CES demand and free entry, Dixit and Stiglitz [1977] find that the monopolistically competitive equilibrium generates an optimal allocation if the social planner cannot use lump-sum subsidies. Among recent studies, Restuccia and Rogerson [2008] analyze the effects of distortionary policies on welfare in a dynamic setting with aggregate capital accumulation. In their quantitative exercises, Restuccia and Rogerson [2008] show that the introduction of taxes and/or subsidies that increase price dispersion across producers can lead to up to 50% loss in aggregate output and TFP. Hsieh and Klenow [2009] provide a measure of allocative efficiency for a one-sector economy with the Cobb-Douglas production. In their applications, the authors show that distorted allocation of resources in China and India lead to the significant reduction in aggregate TFP levels. Baqaee and Farhi [2020b] evaluate the losses from misallocation in a general equilibrium multi-sector setting with production networks. Baqaee and Farhi [2020a] consider a more general static setting with multiple entry types and show that the values of markups that implement

---

[1] A marginal markup is equal to the ratio of the product price to its marginal cost. An average markup is equal to the ratio of price and the product's average cost.

the socially efficient allocation depend on the intensity of economies of scale. Similar to their earlier study, the authors find that the losses due to the misallocation are large. They also suggest that the presence of entry amplifies the misallocation effects. In contrast to Baqaee and Farhi [2020b] and Baqaee and Farhi [2020a], Edmond et al. [2021] setting features a dynamic economy with entry and endogenous capital accumulation. Despite the endogenous dynamics of capital stock and producer masses, the effects of market power in Edmond et al. [2021] are rather similar to the static settings with entry since there is no firm-level investment in either fixed assets or productivity. Capital is owned and accumulated by consumers and rented out by firms in each period. Thus, in Edmond et al. [2021] setting, both capital and labor act as variable inputs in the production of final goods, and market power leads to both under-employment of workers and under-usage of capital goods.

In terms of the welfare analysis, all the studies discussed above compare a decentralized equilibrium with a specific set of frictions and markups to a first-best optimum, implemented with a set of subsidies and taxes. Notably, both Baqaee and Farhi [2020a] and Edmond et al. [2021] rely on either lump-sum subsidies or non-linear subsidies that act as lump sum to implement the socially-efficient allocations. Thus, while the exercises implemented in Baqaee and Farhi [2020a] and Edmond et al. [2021] are certainly insightful, they might not provide us with an accurate estimate of the costs of markups. Market power is not the only friction in these settings: in the absence of strategic pricing behavior by firms, the first-best allocation is not feasible. In other words, if the social planner was only able to assign prices to firms' products, and if the other tools, such as the lump-sum transfers, were not available, the social planner would not be able to implement the social optimum. Thus, we suggest that it is inaccurate to label the difference in welfare levels between the social optimum and the decentralized allocation as the welfare cost of market power. Accordingly, in our analysis, we use a second-best optimum to pin down the welfare losses due to market power.

[Welfare and Market Power under Endogenous Growth] The literature on welfare and market power in endogenous growth settings is much younger than the studies that consider static settings. Grossman and Helpman [1991a] show that whenever firms can invest in the quality of their goods, the level of markups in the decentralized equilibrium could be either above or below the social optimum. At the same time, the authors also find that the CES markups generate socially-optimal allocation if the aggregate output growth is generated only by the expansion in the number of final goods, as in Judd [1985]. Aghion et al. [2005] provide evidence in favor of an inverted-U relationship between competition and productivity growth. Although faster productivity growth does not necessarily generate an increase in consumer utility, we would expect Aghion et al. [2005] result to hold for welfare, as long the one-period output does not increase too steeply in the degree of competition. Since then, the economic studies in this field have focused on evaluating the effects of various shocks on macro-outcomes, including the entry rate, productivity growth, and aggregate output. This branch of literature includes Aghion et al. [2019], Akcigit and Ates [2019], and Cavenaile et al. [2020]. While our analysis generates predictions about the origins of the rise of market power and the decline in business dynamism by construction, the primary focus of this study is on the welfare analysis of observed changes in markup levels and firm dynamics. To the extent of our knowledge, Cavenaile et al. [2020] is the only paper that attempts to evaluate the impact of the rise of market power on the aggregate misallocation. The authors find that the social welfare at the Pareto frontier is more than 120% higher than in the corresponding decentralized equilibrium. In contrast to Cavenaile et al. [2020], we aim to evaluate misallocation due to market power separately from the effects of other distortions.

[Origins of the Rise of Market Power and the Decline in US Business Dynamism] Davis et al. [2012] were the first to document a decline in "the pace of worker flows" in the US, and later on, the discussion shifted towards age and growth dynamics on the firm side. From the moment when economists discovered a decline in US "business dynamism," the academic economists linked it to the rise in market concentration, as in Hathaway and Litan [2014][2], and later on – to the rise in producer market power. A variety of explanations for these phenomena have been proposed in the literature. One of the prevalent hypotheses, developed in Bessen [2017], Crouzet and Eberly [2019] and de Ridder et al. [2019] focuses on changes in firms' production structure, and specifically, on the rise in the importance of intangible assets for the production decisions of US firms. Pugsley and Sahin [2019] suggest that the fall in the entry rates is a primary reason for a decline in the worker flows. A range of studies, including Karahan et al. [2019], Hopenhayn et al. [2018], Engbom et al., Peters and Walsh [2019], suggest that the decline in the US population growth rate has led to firm aging and the rise of industry concentration in US. Most of the trends mentioned above are also analyzed in Akcigit and Ates [2019] who propose that the secular trends are due to the decline in knowledge externalities between market leaders and laggards.

[Markups and Production Function Estimation] Finally, here we also need to mention the literature on the markup measurement and production function estimation. The increase in markups has been documented by several studies, including De Loecker et al. [2020] and Hall [2018], among others. Multiple other studies have confirmed the qualitative findings of De Loecker et al. [2020] and Hall [2018] under different assumptions on the production structure and the nature of inputs. However, the magnitudes of the aggregate rise in markups vary significantly across the studies. Recently, Bond et al. [2021] emphasize the importance of the output price bias in the production function and markup estimation. In our markup estimation methodology, we rely on the approach of Klette and Griliches [1996] to address this issue. Instead of using firm sales shares as proxies for the price variation, as suggested by De Loecker et al. [2020], we always specify a functional form for the demand system, and we estimate the parameters of consumer preferences jointly with production functions. We also augment the standard production function estimation routine to allow for the endogenous dynamics of firm productivity, in line with [Buettner, 2004] and Doraszelski and Jaumandreu [2013].

1.2. Roadmap

The rest of this paper proceeds as follows. Section 2 presents a toy model that illustrates the challenges we face in measuring the welfare effects of markups and our approach to addressing them. Section 3 lays out the setting of the general model. Section 4 describes equilibrium allocations for the decentralized equilibrium and socially-optimal balanced growth paths. Section 5 contains our theoretical results on welfare decompositions and comparative statics. Sections 6 and 7 describe the data we use in our structural

---

[2] On the rise in concentration, see Gutiérrez and Philippon [2016], Gutiérrez and Philippon [2017] and Grullon et al. [2019].

estimation exercises and the corresponding estimation strategy. In Section 8, we present the structural estimation results. Section 9 concludes.

2. Toy Model

This section presents a toy model that demonstrates our approach to evaluating the welfare effects of market power in frameworks with free entry and endogenous evolution of firm productivity. We argue that in such settings, the distance to the first-best allocation is not an accurate measure of welfare losses due to markups. We then suggest a second-best equilibrium concept that allows us to isolate the impact of market power on social welfare. We argue that the distance to the second-best allocation is a more appropriate measure of the costs associated with producer market power.

2.1. Toy Model: Setting

We consider a discrete time economy with one production sector. Final goods are produced by a mass $\mathcal{M}$ of monopolistically competitive firms, and the consumer's preferences across final good varieties are given by a CES(σ) aggregator:

$$Y_t = \left(\int_{\mathcal{M}} \left(y_{\theta t}\right)^{1-\frac{1}{\sigma}} \, \mathrm{d}\theta \right)^{\frac{\sigma}{\sigma-1}} . \tag{1}$$

The representative consumer also has CRRA inter-temporal preferences with an inverse elasticity of substitution ϑ:

$$\mathcal{W}_0 = \sum_{t=0}^{\infty} e^{(-\rho t)} \frac{\left(Y_t\right)^{1-\vartheta}}{1-\vartheta}. \tag{2}$$

In each period, the consumer supplies one unit of labor to producers.

Firm production function is given by $y_{\theta t} = a_{\theta t} l_{\theta t}$ for firm θ at time t. $a_{\theta t}$ denotes firm's productivity, and $l_{\theta t}$ – the amount of employed labor. We assume that initially all firms have the same level of TFP. Incumbent firms exit the production sector at an exogenous rate δ.

There is an unlimited mass of potential entrants. Entry entails a cost $\mathcal{L}_E$ that is measured in terms of labor units. Upon entry, firms are assigned a productivity level equal to the average TFP in the industry. In conjunction with our assumption on the initial TFP distribution, this implies that producers in this toy model setting are homogeneous: in each period, they have the same productivity level, employ the same number of workers in production, and invest at the same rate.

The evolution of firm productivity ($a_{\theta t}$) is deterministic:

$$a_{\theta(t+1)} = (1 + z_{\theta t})^{\omega} a_{\theta t}, \tag{3}$$

where ω is a fixed constant, and $z_{\theta t}$ represents investment of firm θ in period t, in terms of labor units.

2.2. Toy Model: Balanced Growth Path Analysis

In this section and the rest of the paper, we analyze the behavior of economies that move along balanced growth paths so that all the cross-sectional distributions are time-invariant, and all the average or aggregate variables grow at constant rates. In particular, the shares of labor allocated to production (Λ^Y), investment (Λ^Z), and entry (Λ^E) are constant. The balanced growth paths that we consider include the decentralized equilibrium, the first best, and the second-best. In this section, we omit time subscripts on all endogenous variables because we only consider balanced growth paths. Similarly, firm-type subscripts are omitted because producers are homogeneous.

[First Best] The first-best is defined as the balanced growth path that generates the highest possible welfare for consumers, subject to technology restrictions and labor market clearing. The first-best allocation can be described as follows

Proposition 2.1. *Suppose ω is small enough. The first-best balanced growth path exists, and it is unique. On the socially optimal balanced growth path, the labor shares solve*

$$\begin{aligned} \Lambda^{Y,FB} &= 1 - \frac{1}{\sigma}, \\ \Lambda^{Z,FB} &= \omega \frac{z}{1+z}\left(1 - \frac{1}{\sigma}\right)\Lambda^F, \\ \Lambda^{E,FB} &= 1 - \left(1 - \frac{1}{\sigma}\right)\left(1 + \omega \Lambda^F \frac{z}{1+z}\right). \end{aligned} \tag{4}$$

Here $\Lambda^F = \frac{\exp^{-\rho+(1-\vartheta)(1+z)^{\omega}}}{1-\exp^{-\rho+(1-\vartheta)(1+z)^{\omega}}}$ is the contribution of future periods' consumption to social welfare.

The intuition behind the first-best allocation is relatively straightforward. The share of production labor is pinned down by the tradeoff between the mass of producers $\mathcal{M}$ and per-firm output proportional to l. If each firm in the economy employs more workers, there should be fewer firms overall. This is essentially the same principle that pins down the optimal resource allocation in Dixit and Stiglitz [1977]. The share of investment labor is proportional to the elasticity of future productivity with respect to investment $\omega\frac{z}{1+z}$, and the weight of future periods' consumption in welfare Λ^F. The entry share is a residual.

[Implementation of First Best] Several sets of policies can implement the first-best allocation. In all the cases, to reach the social optimum, the social planner would need to alter producers' incentives in both short-term and dynamic optimization problems. They could do so by using a combination of an output tax and an investment subsidy. Alternatively, a combination of output tax and an entry subsidy/tax would also work. To note, if we allow the social planner to control markups instead of imposing output taxes, the value of the optimal markup would not be equal to either the standard CES markup $\frac{\sigma}{\sigma-1}$, or the value that implements first-best allocation.

[Decentralized Allocation with Fixed Markups] Now let us characterize the decentralized allocation that is generated under a fixed producer markup $\mu \geq 1$. In this case, it is convenient to fix the markup value because the allocation that we derive below describes both the *actual* decentralized equilibrium, in which firms set prices, and the second-best optimum, in which the social planner sets the markup. We discuss the second-best equilibrium concept in more detail in the next paragraph.

Solving for the decentralized labor allocation, we get:

Proposition 2.2. *Suppose the markup is is equal to μ. Then, the decentralized allocation solves*

$$\begin{aligned}\Lambda^{E,\mu} &= \Lambda^{Z,\mu}\delta\left(1+\frac{1}{\Lambda^{F\pi}}\right)\left(\frac{1}{\omega\left(\sigma-1\right)}\left(1+\frac{1}{z}\right)-\Lambda^{F\pi}\right),\\ \Lambda^{Z,\mu} &= \Lambda^{Y,\mu}\omega\left(\sigma-1\right)\left(\mu-1\right)\frac{z\Lambda^{F\pi}}{1+z}.\end{aligned} \tag{5}$$

$\Lambda^{F\pi}=\frac{(1-\delta)\exp^{-r}}{1-(1-\delta)\exp^{-r}}$ *is the contribution of future profits in the producer objective. The decentralized balanced growth path always exists and is unique.*

It is important to highlight the source of inefficiencies that generate the difference between the first-best and decentralized allocations in this setting. To ensure that the firms are homogeneous, we have assumed that the entrants have a productivity level equal to the average TFP in an industry. In the presence of such knowledge spillovers, the incumbents' investment choices are generically sub-optimal since the social returns on investment differ from the private returns. This discussion is similar to the welfare analysis presented in Grossman and Helpman [1991a] that suggests that the Dixit and Stiglitz [1977] result on the optimality of CES markups does not extend easily to the settings with endogenous productivity growth. Similar to canonical growth frameworks, our toy model also features business stealing and consumer surplus externalities. By now, multiple studies have shown that these externalities offset each other in the settings with CES demand and free entry[3]. Once we add firm-level innovation and dynamic inputs to the setting, the balance between business stealing and consumer surplus externalities is no longer optimal[4].

Importantly, Propositions 2.1 and 2.2 lead to a following useful observation

Corollary 1. *The ratios $\Lambda^{Z,FB}/\Lambda^{E,FB}$ and $\Lambda^{Z,\mu}/\Lambda^{E,\mu}$*

(i) do not depend on the markup value μ;

(ii) are generically not equal to each other.

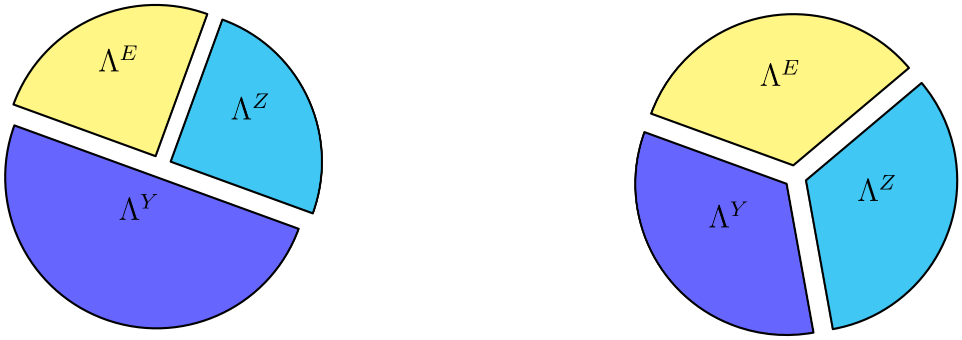
This corollary suggests that we can never reach the first-best allocation if we can only alter the markups on final goods. The investment in this setting is fixed, conditional on the entry cost, consumer preference parameters, TFP elasticity ω, and exit rate δ. Homogeneity of producers is crucial for this result. In the equilibrium, firm investment is proportional to the producer value function that is constant across firms. Furthermore, the free entry condition implies that the value of producer profits is proportional to the entry cost. This means that only the values $\mathcal{M}$ and l react to the markup shocks, the mass of producers, and per-firm employment readjust to keep the ratio $\Lambda^{Z,\mu}/\Lambda^{E,\mu}$ constant. Figure 1 depicts the labor allocation for two different levels of markups.

---

[3] Such results are presented, e.g., in Dixit and Stiglitz [1977], Grossman and Helpman [1991a], and Bilbiie et al. [2019].

[4] This is evident from the fact that in the settings similar to Atkeson and Burstein [2010] the first-best allocation is different from the decentralized equilibrium. Proof is provided upon request.

Figure 1: Toy Model: Labor Allocation Under Different Markup Values



(a) Low Markup

(b) High Markup

[Second Best] We define the second-best allocation as the solution to the following optimization problem:

$$\begin{aligned} \max_{\mu} \quad & \mathcal{W} = \frac{(l)^{1-\vartheta} (\mathcal{M})^{(1-\vartheta)\frac{\sigma}{\sigma-1}}}{1-\exp^{-\rho+(1-\vartheta)(1+z)^{\omega}}}, \\ \text{s.t.} \quad & (\mathcal{M})^{-1} = l + z + \delta\mathcal{L}_E, \\ & \delta\mathcal{L}_E = z\delta\left(1+\frac{1}{\Lambda^{F\pi}}\right)\left(\frac{1}{\omega(\sigma-1)}\left(1+\frac{1}{z}\right)-\Lambda^{F\pi}\right), \\ & z = l\omega(\sigma-1)(\mu-1)\frac{z\Lambda^{F\pi}}{1+z}. \end{aligned} \tag{6}$$

In the equation above, the first line represents the objective of the social planner, the second line contains the labor market clearing constraint, and the last two lines replicate the decentralized allocation conditions from Proposition 2.2. Informally, on the second-best balanced growth path, the social planner sets markups subject to the BGP feasibility constraints and the constraints imposed by the profit-maximizing behavior of firms. We hope that the motivation behind this second-best concept is relatively straightforward. Market power is defined as a producer's ability to control the prices of their goods. In the second-best, we take away this "power" from firms and reallocate it to the social planner, who then sets markups or prices at their discretion. This notion of second-best is quite similar to the one suggested by Dixit and Stiglitz [1977] in their analysis. We also view it as a natural extension of the welfare analysis performed by Baqaee and Farhi [2020b] and Baqaee and Farhi [2020a] in frameworks with exogenously fixed markups.

The second-best allocation can be characterized as follows:

Proposition 2.3. *The second best markup level solves*

$$\begin{aligned} \Lambda^{Y,\mu} &= 1-\frac{1}{\sigma}, \\ \mu &= 1+\frac{1}{\sigma-1}\left((\sigma-1)\omega\frac{z\Lambda^{F\pi}}{1+z}\left(1+\frac{\Lambda^{E,\mu}}{\Lambda^{Z,\mu}}\right)\right)^{-1}. \end{aligned} \tag{7}$$

As the proposition above suggests, at the second-best allocation, the share of production labor is the same as at first best: the social planner cannot readjust the distribution of workers between entry and investment, but they can still reset Λ^{Y} to its socially optimal value. Notably, the markup level that solves the equations in Proposition 2.3 is not equal to the optimal Dixit and Stiglitz [1977] markup value, i.e., it is generically not identical to the CES markup that firms would decide to impose in the "true" decentralized equilibrium. The deviation from the CES markup value depends on the ratio of shares $\Lambda^{E,\mu}/\Lambda^{Z,\mu}$ that determines the extent of misallocation at the second-best allocation and on the elasticity of firm profits with respect to investment z.

2.3. Toy Model: Takeaways

In the toy model that we have outlined above, the allocation of labor between investment and entry is independent of the markup value. Moreover, the distribution of resources in this economy is generically sub-optimal. Thus, we have suggested a way to isolate the welfare loss due to a sub-optimal level of markups from the misallocation generated by the discrepancy between social and private returns on investment. Specifically, we propose using the distance to the second-best allocation as a primary measure of the welfare costs of market power.

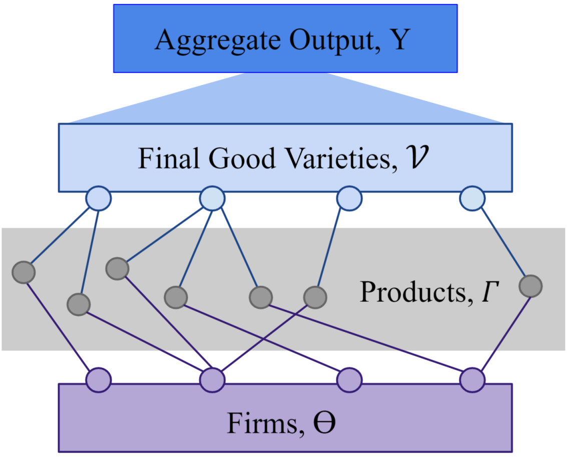
In a more general setting with heterogeneous firms, investment and entry cost shares depend on the value of markups imposed by different producer types. Still, the general logic of our toy model example does hold up. In a generic setting with free entry and endogenous productivity or capital dynamics, we cannot move the economy to the socially optimal balanced growth path by simply eliminating producer market power. The additional policy instruments are necessary to achieve the first best. Thus, we argue that the second-best comparisons are more appropriate if we want to analyze the welfare effects of markups – or any other inefficiencies.

3. General Model: Setting

In this section, we present the setting of the general theoretical model. Our framework is built in discrete time. The structure of the economy is similar to the models featured in Atkeson and Burstein [2008], Liu et al. [2019], Weiss [2019] and Cavenaile et al. [2020]. We assume that at each moment in time, the production sector contains a continuum of good varieties $\mathcal{V}_t$ at time t. Each variety contains a discrete number of product lines, and each good producer can own multiple products within different sectors. Similar to many growth frameworks based on Grossman and Helpman [1991b], this setup allows us to flexibly model market power at a firm level and preserves deterministic dynamics for all aggregate variables. Figure 2 illustrates this setup. We describe the definitions of product, sector, and firm types in the next subsection.

Our framework features both static and dynamic factors of production. We say that a production factor is static if firms can alter the amount of this factor instantaneously. Quantities of dynamic inputs available to producers in period t can only be affected by their investment in period $t-1$. In addition, we classify all inputs that accumulate over time as dynamic. This definition implies that dynamic inputs act as state variables in the short-run producer optimization, similarly to TFP. We casually refer to the static factor

Figure 2: Economy Structure for the General Setting



of production as "labor" and the dynamic factor – as "capital," although our setup admits a much broader interpretation. We introduce dynamic inputs in our model primarily to make our theoretical setting more consistent with the structural estimation framework and to simplify the interpretation of our counterfactual results. The role of fixed assets in production is further discussed in Section 3.3.

Population is homogeneous, and we typically assume that the size of the labor force is equal to one in the initial time period $t = 0$. Individuals supply labor to producers in a perfectly competitive labor market. In the benchmark, we assume that the labor supply is inelastic so that each worker generates one unit of labor input in each period. Population grows over time with a constant rate g_L.

Our setting also features free entry. Similar to the toy model, entry is a dynamic decision. Each potential entrant has an option to pay a fixed entry cost $\mathcal{L}_E$, enter the production sector in the next period, and draw a firm type from a fixed distribution. The entry cost is measured in terms of labor units, and it grows over time with a fixed rate g_E. The total number of entrants is determined endogenously. Product and firm exit in our setting are non-strategic so that there is no endogenous selection based on productivity and/or fixed assets.

3.1. Type Spaces

To keep notation concise, we index products, varieties, and firms by their types. A *"type"* is a summary of all relevant characteristics of a firm, product, or sector that (i) cannot be altered in the current period and (ii) acts as sufficient statistics for producers' decisions at a firm or product level. For example, under the definition of product types stated in this paragraph, all goods that belong to the same type have the same equilibrium markups and the same equilibrium employment levels. Similarly, firms that belong to the same type spend the same amount of funds on investment.

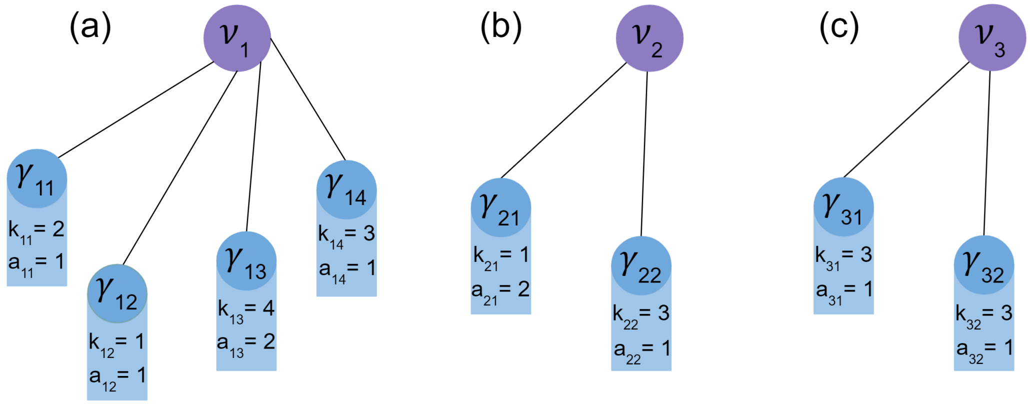
Our setting features homothetic consumer preferences and homothetic production. Thus, it is convenient to define product and producer types in terms of relative productivity and relative capital, denoted by

$a_{\gamma t}$ and $k_{\gamma t}$ for product γ in period t. $a_{\gamma t}$ and $k_{\gamma t}$ are defined in terms of absolute levels of TFP $\tilde{a}_{\gamma t}$ and capital stock $\tilde{k}_{\gamma t}$, and the corresponding un-weighted economy-wide averages $\mathcal{A}_t$ and $\mathcal{K}_t$[5]:

$$\begin{gathered} a_{\gamma t} = \frac{\tilde{a}_{\gamma t}}{\mathcal{A}_t}, \quad k_{\gamma t} = \frac{\tilde{k}_{\gamma t}}{\mathcal{K}_t}, \\ \text{where } \mathcal{A}_t = \mathbb{E}_{\gamma \in \Gamma_t}\left[\tilde{a}_{\gamma t}\right], \quad \mathcal{K}_t = \mathbb{E}_{\gamma \in \Gamma_t}\left[\tilde{k}_{\gamma t}\right]. \end{gathered} \tag{8}$$

Throughout the rest of this paper, we casually refer to the relative capital stock of a product as ”capital stock” and relative productivity of a good – as ”productivity” or ”TFP.”

Figure 3: Variety and Product Types



A *sector type* ν includes the number of product lines that belong to the sector, and a joint distribution of product-specific capital stocks and TFP. Figure 3 depicts three distinct variety types ν_1, ν_2 and ν_3. Variety ν_2 is different from variety ν_3 because of the differences in productivity and capital of goods γ_{21} and γ_{31}. A *product type* γ consists of a product-specific stock of capital, product-specific TFP and a variety type index ν. This definition implies that product types are variety-type specific: in Figure 3 γ_{14}, γ_{22} and γ_{32} belong to different product types despite the fact that all these products have similar productivities and capital stock levels.

Γ_t is the set of all products present in the economy in period t, and Γ_ν is the set of products within a sector for variety ν. Γ_ν is not indexed by the time period since the definition of variety types includes the distribution of capital and TFP across products. The same applies to the relative product characteristics a_γ and k_γ, and the product type densities within sectors $f_{\nu\gamma}$.

The definition of a firm type depends on properties of the investment process and assumptions on the dynamics of TFP and capital[6]. In our theoretical analysis, and thus, all our results will be formulated for a

---

[5] There are many equivalent ways to define relative producer types: for any constant-returns-to-scale function $h\left((s_{\gamma' t})_{\gamma' \in \Gamma_t}\right)$ and for any product characteristic s, the function $s_{\gamma t}/h\left(\{s_{\gamma' t}\}_{\gamma' \in \Gamma_t}\right)$ would evolve similarly to $s_{\gamma t}/\mathbb{E}_t\left[s_{\gamma' t}\right]$ along any balanced growth path of our setting. We use the unweighted expectations of product characteristics for the sake of simplicity only. All the results that we present in the subsequent sections are invariant to the relative type specifications.

[6] In a range of settings, firm types can be represented as multi-sets of product types owned by each firm. For example, this is the case when all the variety markets feature either monopoly or Bertrand competition with perfectly substitutable products, as in the canon models of Aghion and Howitt [1992], Grossman and Helpman [1991b] or Klette and Kortum [2004]. In other settings, firm types have to include information on the firms' positions in the economy's production network.

general Banach firm type space. We will use $\theta \in \Theta_t$ to index the producer types, where Θ_t denotes the set of all producer types present in the economy at time t.

3.2. Consumers

[Consumer Preferences] Preferences of consumers across final good varieties are given by a CES aggregator with elasticity of substitution $\sigma_{\mathcal{V}}$:

$$(Y_t)^{1-\frac{1}{\sigma_{\mathcal{V}}}} = |\mathcal{V}_t| \int_{\nu \in \mathcal{V}_t} (Y_{\nu t})^{1-\frac{1}{\sigma_{\mathcal{V}}}} f_{\nu t} \mathrm{d}\nu. \tag{9}$$

Here $Y_{\nu t}$ is the consumption index for goods that belong to variety ν, and $|\mathcal{V}_t|$ is the total mass of varieties available to consumers at time t; $f_{\nu t}$ denotes the density of variety type ν among all sectors at time t.

Similarly, the sector-level consumption is a CES (σ_Γ) aggregate of outputs produced by individual firms:

$$(Y_{\nu t})^{1-\frac{1}{\sigma_\Gamma}} = |\Gamma_\nu| \sum_{\gamma \in \Gamma_\nu} (y_{\gamma t})^{1-\frac{1}{\sigma_\Gamma}} f_{\gamma\nu}. \tag{10}$$

$f_{\gamma\nu}$ denotes the time-invariant density of product type γ among all products in sector of type ν. The density $f_{\gamma\nu}$ can be expressed as a ratio of product and sector type densities: $f_{\gamma\nu} = f_{\gamma t}/f_{\nu t}$.

[Consumer Optimization Problem(s)] In the short run, the representative consumer sets the demand for each final good by solving a standard output maximization problem, subject to their budget constraint. The inter-temporal consumer optimization is also standard:

$$\begin{aligned} \max_{\{Y_t\}_{t=0}^{\infty}, \{\alpha_t\}_{t=0}^{\infty}} \mathcal{W}_0 &= \sum_{t=0}^{\infty} \mathrm{e}^{-\rho t} \frac{(Y_t/L_t)^{1-\vartheta}}{1-\vartheta} \mathrm{d}t, \\ \text{s.t.} \quad \frac{Y_t}{L_t} P_t &= w_t + \alpha_t - \exp(-r_t)\alpha_{t+1}, \forall t. \end{aligned} \tag{11}$$

r_t is the log value of the interest rate, and α_t is the consumer asset holdings at time t. $\vartheta \geq 1$ is the inverse of the intertemporal substitution elasticity. In addition to the budget constraint, the consumer's problem is also subject to a standard no-Ponzi condition.

Importantly, unlike Edmond et al. [2021] or Weiss [2019], we assume that consumers do not own capital directly. Instead, the evolution of fixed asset stocks is determined by the dynamic producer optimization. A key consequence of this assumption is that market power has different and possibly opposing effects on production labor and capital in our setting. Whenever markups increase, the number of workers employed in production should fall, as higher prices reduce the consumer demand. At the same time, higher markups levels are likely to incentivize innovation and entry because firms are willing to invest more if they can earn higher surpluses.

3.3. Firms: Short Term

[Production] For each product $\gamma \in \Gamma_t$, the production function maps productivity, labor and capital stock to physical output:

$$y_{\gamma t} = a_\gamma \mathcal{A}_t \left(k_\gamma \mathcal{K}_t\right)^{\omega_K} \left(l_{\gamma t}\right)^{\omega_L}, \tag{12}$$

where a_γ is the relative good-specific productivity, and $k_\gamma \mathcal{K}_t$ and $l_{\gamma t}$ represent the amounts of capital and labor used in production at time t. ω_K and ω_L denote the output elasticities with respect to production factors. Also, we use $\xi = \omega_K + \omega_L$ to denote the returns to scale parameter of the production function.

There are several reasons why we explicitly include capital in our model. First, a few studies, including Bessen [2017], Crouzet and Eberly [2019], de Ridder et al. [2019] and Weiss [2019] have suggested that the rise of market power in the US can be explained by an increase in the importance of intangible capital. We want accurately evaluate the effects of the rise of intangibles on welfare and first and second-best allocations. To do so, we need to separate R&D and SGA expenditures from the costs associated with physical capital maintenance and accumulation. Second, by explicitly incorporating capital in our setting, we can make our theoretical model more similar to the structural estimation framework used to evaluate firm production functions and demand structure. Most of the existing production function estimation methods assume that physical capital and TFP are dynamic. The estimates obtained using these methods are inconsistent if capital is variable in the short run or functions as a part of composite TFP. Thus, this assumption simplifies the interpretation of our counterfactual exercises. Finally, if we separate capital and TFP at the firm level, we can adopt different assumptions for the dynamics of the aggregate stock of tangible assets and productivity. We can allow aggregate capital to accumulate linearly, as in most macro settings. Productivity can evolve in a log-linear fashion, as in the benchmark growth models.

[Market Power and Industry Structure] Our welfare analysis exercises focus on the first and second-best comparisons. In both first and second best, the social planner sets the values of markups, either directly or indirectly. This implies that the welfare implications of markups only depend on the markup values and not on their origins. Thus, as long as we focus on the welfare analysis only, we can w.l.o.g. consider a setting with fixed markups and marginal cost pricing, as in Baqaee and Farhi [2020b] and Baqaee and Farhi [2020a]:

Assumption 1. ***[Marginal Cost Pricing]*** *Producers take markups as given and choose employment and output levels that satisfy the following equation:*

$$p_{\gamma t} y_{\gamma t} = \frac{\mu_\gamma}{\omega^L} l_{\gamma t} w_t. \tag{13}$$

ω^L is the elasticity of output with respect to variable inputs (labor), and $p_{\gamma t}$ is the price that corresponds to the level of output $y_{\gamma t}$ whenever consumer optimization conditions hold.

Here it is important to note that, while the welfare comparisons that we consider are invariant to the nature of markups, results of other counterfactual exercises do depend on the reaction of markups to production or/and demand shocks. For example, suppose we want to analyze the welfare effects of increased capital intensity or a decline in demand elasticities. Such shocks inevitably change the magnitude of markups. The resulting welfare elasticity will vary depending on how much the markups change, i.e., the pass-through rates of shocks into markups. On the other hand, it is also true that pass-through rates do not affect the economy's reaction to all shocks that only affect firms' dynamic optimization. Such shocks include changes in innovation technology, knowledge spillovers, and exogenous growth rates of population and entry costs.

3.4. Firms: Dynamics

[Innovation and Investment] Incumbent firms control the dynamics of their state variables, including product-specific capital, product-specific TFP, and the number of active product lines by investing in either capital expenditures or various research projects that generate product and process innovations. Formally, we assume that there are several investment options available to firms. The set of these projects $\mathcal{I}$ is finite and time-invariant. Individual investment options are indexed by $\iota \in \mathcal{I}$. Firm-level investment in terms of labor units is represented by a vector $z_{\theta t} \in \mathbb{R}^{|\mathcal{I}|}$. Throughout our analysis, we assume that capital evolves independently from TFP and the number of active products. There exists a project $\iota_K \in \mathcal{I}$ such that investments in ι_K can only lead to changes in capital accumulation. Other projects can only influence TFP and the number of goods owned by the firm.

We assume that firm types follow a first-order Markov process. The transition probabilities between active firm types at time t are summarized by a kernel $\mathcal{P}_t : \Theta \times \eth \to [0, 1]$, where $\eth$ is a complete sigma-algebra on the firm type space. The elements of the map $\mathcal{P}_t$ are not fixed. Rather, the type transition probabilities depend on the state of the economy, the firm's own investment, and the distribution of investment across all producer types. Formally, we require that the following assumption holds:

Assumption 2. ***[Transition Probabilities]*** *The transition kernel $\mathcal{P}_t$ of the Markov process that describes dynamics of firm types can be specified in the following way:*

$$\mathcal{P}_t = P\left(z_{\Theta t} \exp^{-g_E t}, f_{\Theta t}, \mathcal{E}_t, \mathcal{K}_t \exp^{-g_E t}\right). \tag{14}$$

$z_{\Theta t} : \Theta \to \mathbb{R}^{|\mathcal{I}|}$ is the operator that maps firm types to investment levels. $f_{\Theta t}$ is the cross-sectional density of firm types at time t, $\mathcal{E}_t$ is the entry rate, and $\mathcal{K}_t \exp^{-g_E t}$ is the detrended aggregate capital stock.

We assume that the operator P is smooth in all the arguments. For simplicity, we also assume away corner solutions in firms' dynamic optimization, i.e., we consider the settings in which higher own investment in productivity increases the probability of product and process innovations and decreases the likelihood of exit. Similarly, own investment in the capital stock should increase the likelihood of obtaining larger capital stock in future periods. These conditions can be formalized in terms of the first-order stochastic dominance of the marginal distributions of product capital stock and productivity. Finally, to ensure the existence of balanced growth paths, we assume that the probability of transition to the ”exit” state is non-zero for all firm types. Later on, we will also impose additional ”concavity” restrictions on P to ensure the existence of a well-behaved equilibrium.

To clarify the meaning of Equation 14, let us note that for *any* setting with Markov dynamics transition probabilities between firm states can be expressed as functions of the control variables $z_{\Theta t}$, the state of the system, summarized by a set of time-dependent objects $\{f_{\Theta t}, \mathcal{A}_t, \mathcal{K}_t, \mathcal{M}_t, L_t\}$, and all model parameters. Thus, the specification above restricts the set of arguments of the probability transition kernel quite significantly. In this section, we provide a brief explanation for why each of the objects listed on the right-hand side of Equation 14 influences the evolution of producer state variables. We also argue that this specification is consistent with the assumptions imposed on productivity dynamics in many existing economic growth frameworks.

Equation 14 suggests that four factors influence firm type transitions. The first argument, $z_{\Theta t}$, captures three channels: first, own firm investment has a direct positive effect on either firm's productivity, capital stock, or/and a set of products lines available to the firm; second, investment done by other producers might affect firm's state due to either business-stealing or non-pecuniary externalities, depending on the model; finally, $z_{\Theta t}$ also determines the amplitude of pecuniary externalities generated by the investment. Similarly, the remaining arguments of the function P affect the transition probabilities via pecuniary and non-pecuniary externalities. The impact of pecuniary externalities in our setting is proportional to changes in the endogenous growth rates g_{At} and g_{Kt}. Thus, the magnitude of the pecuniary externality effects depends on all of the arguments of the operator P. The same goes for the non-pecuniary externalities. E.g., frequencies of firm types are likely to affect the probability of ”business stealing” by incumbent firms and expansions of incumbent firms into new markets. The entry rate $\mathcal{E}_{t+1}$ induces non-pecuniary externality effects on incumbents' states whenever the new firms are allowed to displace the incumbents upon entry or whenever the entrants gain access to the incumbent's sectors. The aggregate capital stock is included in the set of arguments of P because we want to allow for the additive capital accumulation process, as in most macro settings.

The specification of the innovation process stated above is rather abstract, so it is important to clarify which frameworks are covered by it. In classical settings of Aghion and Howitt [1992], and Grossman and Helpman [1991b], the frequency of product innovations depends only on firms' investment, and the probability of exit depends on the distribution of investment intensities across all firms in the industry, as well as the entry rates. The same holds for the Klette and Kortum [2004] model that allows firms to own multiple goods and thus features a multidimensional firm type space: in Klette and Kortum [2004], transitions between firm types are determined by the entry rate and distributions of the investment. The models with externalities, e.g., Akcigit et al. [2021] or Akcigit and Ates [2019], also are particular cases of our setting. The non-pecuniary externalities alter the shape of the $\mathcal{P}$ function, but not the set of its arguments.

[Entry] Firms that enter at time t have to hire $\mathcal{L}_{Et} = \mathcal{L}_E \exp^{g_E t}$ workers to set up the production process. This cost covers expenditures on the initial stock of fixed assets, as well as the productivity draw. $\mathcal{P}_{Et} : \eth \to [0, 1]$ denotes the distribution of firm types for entrants. $\mathcal{P}_{Et}$ is specified as follows:

Assumption 3. ***[Entrant Type Distribution]*** *The distribution of entrant types is given by a value of a function P_E:*

$$\mathcal{P}_{Et} = P_E\left(z_{\Theta t}\exp^{-g_E t}, f_{\Theta t}, \mathcal{E}_t, \mathcal{K}_t \exp^{-g_E t}\right). \tag{15}$$

P_E is smooth in all its arguments, and it also satisfies the following identities, for all possible values of its arguments:

$$\mathbb{E}_{\mathcal{P}_{Et}}[\tilde{a}_{\gamma t}] = \bar{A}_E \mathcal{A}_t, \quad \mathbb{E}_{\mathcal{P}_{Et}}[\tilde{k}_{\gamma t}] = \bar{K}_E \exp^{g_E t}. \tag{16}$$

Function P_E has the same set of arguments as P, and the motivation behind it is similar. In addition, we assume that the absolute levels of entrants' productivities are proportional to the aggregate productivity at the time of entry, $\mathcal{A}_t$, and that the detrended stocks of fixed assets for new firms are drawn from a fixed distribution. Although it might seem non-standard, this assumption is necessary for the existence of a balanced growth path with non-zero productivity growth. In different shapes or forms, it is present in many frameworks with endogenous growth. E.g., Sampson [2016] explicitly assumes that the entrants draw productivities from the existing TFP distribution. In the frameworks based on Aghion and Howitt [1992] creative destruction model, the entrants have to develop a good variety that supersedes the product of their closest competitor. This means that similarly to Sampson [2016], entrants' TFP levels grow proportionately to the aggregate TFP. Note also that under the assumptions imposed in this section there are no multiple equilibria in the entry stage.

[Producers' Dynamic Optimization] Potential entrants have to decide whether to abstain or to pay the cost $\mathcal{L}_E \exp^{g_E t}$ and get the type draw from distribution $\mathcal{P}_{Et}$. The free entry condition thus implies:

$$w_t \mathcal{L}_E \exp^{g_E t} = \mathcal{P}_{Et} V_{\Theta t}. \tag{17}$$

The incumbents' dynamic optimization problem is as follows

$$\begin{aligned} V_{\theta t} &= \max_{z_{\theta t}} \left\{ S_{\theta t} - z_{\theta t} w + \exp^{-r_t} \mathcal{P}_t V_{\Theta t+1} \right\}, \\ \text{s.t.} \quad S_{\theta t} &= \sum_{\gamma \in \Gamma_\theta} S_{\gamma t} = \sum_{\gamma \in \Gamma_\theta} p_{\gamma t} y_{\gamma t} - w_t l_{\gamma t} \end{aligned} \tag{18}$$

Here $S_{\gamma t}$ denotes the *variable surplus* generated by the sales of product γ, and $S_{\theta t}$ stands for the total variable surplus earned by firm θ. Variable surplus is a short-run objective, and the maximization of surpluses determines product-level employment, output, and markups. Profits, defined as the difference between variable surplus and investment costs $\pi_{\theta t} = S_{\theta t} - z_{\theta t} w$, act as a long-run objective. Thus profit levels determine investment levels and entry decisions. We assume that the liquidation value is equal to zero.

4. General Model: Balanced Growth Paths

This section describes the balanced growth path conditions for the decentralized equilibrium and the social optimum. The definition of BGP equilibria is as follows:

Definition 1. [Balanced Growth Paths] *A balanced growth path is formed by pair of positive allocation functions $l_\Gamma : \Gamma \to \mathbb{R}_+$ and $z_\Theta : \Theta \to \mathbb{R}_+^{|\mathcal{I}|}$. Function l_Γ specifies de-trended production employment for all* *the products in the economy, and z_Θ describes de-trended investment levels for all firm types and all investment types.*

The definition above implies that in period t, production employment is equal to $l_\Gamma \exp^{g_E t}$, and the investment levels are equal to $z_\Theta \exp^{g_E t}$. The cross-sectional distributions of sales shares, cost shares, markups, relative productivities, and relative capital stocks are time-invariant along the BGP, and all aggregate variables grow at constant rates. We normalize average variety-level sales in the economy to 1 in every period.

[Notation] In all subsequent sections, objects with subscripts Θ, $\mathcal{V}$ and Γ denote functions that map a respective type space to real numbers. As an example, $l_\Gamma : \Gamma \to \mathbb{R}_+$ assigns employment levels to all product types. Bold symbols with subscripts Θ, $\mathcal{V}$ and Γ denote linear diagonal operators that map a respective type space to itself. Elements of such diagonal operators are equal to the elements of the corresponding function, e.g., the elements of $\boldsymbol{l}_\Gamma$ are equal to the values of l_Γ. $\mathbb{E}_W\left[X\right]$ denotes a W-weighted expectation of X, so that $\mathbb{E}_W\left[X\right] = W'X/\left(W'\mathbb{1}\right)$.

4.1. Growth Rate Identities

In this paragraph, we derive growth rates for all aggregate and firm-level variables —- the expressions for the growth rates listed below hold for the decentralized equilibrium, first and second best. Formally,

Proposition 4.1. *[Aggregate Growth Rates] On any balanced growth path, the growth rates of the aggregate, firm- and product-level variables are equal to the values stated in Table 1.*

We start by characterizing the growth rates of real variables. First, since we want the shares of production, investment, and entry employment to remain constant, the population growth rate should be equal to the sum of growth rates of the mass of producers $\mathcal{M}$ and the average per-firm employment in each of the categories. Thus, we infer that the growth rate of the mass of firms should be equal to the difference $g_L - g_E$. This also implies that firm-level investment and production labor grow at the same rate as the entry costs. A setting with $g_E = g_L$ generates a constant mass of firms and a constant mass of products, as in the creative destruction models, e.g., Klette and Kortum [2004]. Under $g_E = 0$, per-firm employment and investment is constant, as in Judd [1985] or Romer [1990] models. The masses of varieties and products grow at the same rate as the mass of producers. The same applies to the mass of entering firms.

The growth rate of average product-level capital stock is determined the law of motion for capital. Whenever the expected value of fixed assets is linear in past real investment, the growth rate of product-level capital is equal to g_E[7]. Employment at a firm level also grows at a rate g_E.

In our setting, the aggregate output growth originates from two sources: expansion of the set of varieties produced in the economy and evolution of the output levels within products. The growth rate of product-level output is equal to $g_A + \xi g_E$, where g_A is the aggregate productivity growth. Thus, we have

---

[7] Here, we are implicitly assuming that productivity in the capital goods sector does not grow.

Table 1: Equilibrium Growth Rates

Variable	Growth Rate Value
1. Real Variables	
Mass of Producers, $\mathcal{M}_t$	$g_L - g_E$
Mass of Final Good Varieties, $\mathcal{V}_t$	$g_L - g_E$
Mass of Products, Γ_t	$g_L - g_E$
Product-Level Inputs, $l_{\gamma t}$ and $k_{\gamma t}\mathcal{K}_t$	g_E
Product-Level Output, $y_{\gamma t}$	$g_A + \xi g_E$
Aggregate Output, Y_t	$g_A + \xi g_E + \frac{\sigma_{\mathcal{V}}}{\sigma_{\mathcal{V}}-1}\left(g_L - g_E\right)$
2. Prices	
Wages, w_t	$-g_E$
Firm-Level Prices, p_t	$-g_A - \xi g^E$
CPI, P_t	$-g_A - \xi g_E - \frac{1}{\sigma_{\mathcal{V}}-1}\left(g^L - g_E\right)$

Notes: In the expressions listed in this table, ξ is the returns-to-scale parameter of firms' production function.

the following identity for the dynamics of the aggregate output:

$$g_Y = \underbrace{g_A + \xi g_E}_{\text{Within-Product Output Growth}} + \underbrace{\frac{\sigma_{\mathcal{V}}}{\sigma_{\mathcal{V}} - 1}\left(g_L - g_E\right)}_{\text{Love-of-Variety Effects}}. \tag{19}$$

The amplitude of the love-of-variety effects is determined by the inter-sectoral substitution elasticity. The scale of within-firm output growth depends on the returns-to-scale parameter ξ, and the endogenous growth rate of productivity g_A. Parameters $\sigma_{\mathcal{V}}$ and ξ thus measure direct effects of changes in the exogenous growth rates g_L and g_E on the output growth, *conditional* on the value of g_A. Whenever the population rate declines by $\%1$, the contribution of love-of-variety effects to the aggregate growth declines by $\frac{\sigma_{\mathcal{V}}}{\sigma_{\mathcal{V}}-1}$ %. The effect of changes in the entry cost growth depends on the magnitude of the economies to scale relative to the love-of-variety effects. Whenever returns to scale are large enough, increases in the growth rate of entry costs positively affect output growth, conditional on the growth rate of productivity. If g_E increases, labor is reallocated from the entrants towards incumbents, and this allows the economy to grow faster.

Now let us describe the dynamics of prices along the balanced growth path. Given the price normalization that we use, the aggregate nominal output is always equal to the mass of varieties $\mathcal{V}_t$. The price normalization also allows us to pin down the growth rate of wages and product-level prices. Since firm-level sales and markups are constant, the growth of product-level prices is equal to the negative product-level output growth rate. Similarly, the growth rate of wages is equal to the negative growth rate of per-product employment. The CPI growth rate is equal to the difference between the growth rate of product mass growth rate and the growth rate of aggregate output.

Using the values of aggregate growth rates, we can also derive the expression for the equilibrium interest rate:

$$\begin{aligned} r &= \rho + (\vartheta - 1)\left(g_Y - g_L\right) - g_E, \\ r &= \rho + (\vartheta - 1)\left(g_A + \xi g_E - g_L + \frac{\sigma_{\mathcal{V}}}{\sigma_{\mathcal{V}} - 1}\left(g_L - g_E\right)\right) - g_E. \end{aligned} \tag{20}$$

If consumer preferences exhibit the "love-of-variety" property, the interest rate is also increasing in the population growth rate, conditional on g_A. The sign of the partial derivative of r with respect to g_E depends on the relative magnitudes of parameters ξ and $\sigma_{\mathcal{V}}$, consistently with the discussion above.

4.2. Type Distributions along the BGP

In this section, we describe the cross-sectional type distributions that are generated by the balanced growth path allocations. To characterize the type densities, we first define the Neumann series operator for firm masses: let $\mathcal{P}$ denote the probability transition operator along some BGP,

$$\begin{aligned} \Psi^{\Theta} &= \sum_{n=0}^{\infty} \exp^{-g_M n} \mathcal{P}_E' \mathcal{P}^n = \mathcal{P}_E' \left(I - \exp^{-g_M} \mathcal{P}\right)^{-1}, \\ \text{where } \mathcal{P}^n\left(\theta, \theta'\right) &= \int \int ... \int \mathcal{P}\left(\theta, \hat{\theta}_1\right) \mathcal{P}\left(\hat{\theta}_1, \hat{\theta}_2\right) ... \mathcal{P}\left(\hat{\theta}_{n-1}, \theta'\right) \mathrm{d}\hat{\theta}_1 \hat{\theta}_2 ... \mathrm{d}\hat{\theta}_{n-1}. \end{aligned} \tag{21}$$

Operator Ψ^{Θ} tracks the *masses* of firm types that are present in the economy that follows a balanced growth path. Suppose $\mathbb{1}_{\mathcal{S}} : \mathcal{S} \to \mathbb{R}$ denotes the function that assigns a unit value to all elements of some space $\mathcal{S}$. Then, the number of producers that enter the economy in period $t - n$ and survive until t is proportional to $\exp^{-n g_M} \mathcal{P}_E' \mathcal{P}^n \mathbb{1}_{\Theta}$: the term $\exp^{-n g_M}$ represents the difference in the number of new firms between periods t and $t - n$, and the term $\mathcal{P}^n \mathbb{1}_{\Theta}$ captures type-dependent survival probabilities between t and $t - n$. The operator Ψ^{Θ} is a sum of terms $\exp^{-n g_M} \mathcal{P}_E' \mathcal{P}^n \mathbb{1}_{\Theta}$ for all past periods, and thus it contains information about all firms that survive until period 0.

We can then characterize the BGP entry rate $\mathcal{E}$, and distributions of producer and product types as follows:

$$\begin{aligned} \mathcal{E} &= \left(\Psi^{\Theta} \mathbb{1}_{\Theta}\right)^{-1}, \\ f_{\Theta} &= \mathcal{E} \Psi^{\Theta}, \\ \frac{\mathcal{V}}{\mathcal{M}} &= \mathcal{E}\left(\Psi^{\Theta} \Xi_{\Theta,\mathcal{V}} \mathbb{1}_{\mathcal{V}}\right) = \mathcal{E}\left(\Psi^{\mathcal{V}} \mathbb{1}_{\mathcal{V}}\right), \\ \frac{\Gamma}{\mathcal{M}} &= \mathcal{E}\left(\Psi^{\Theta} \Xi_{\Theta,\Gamma} \mathbb{1}_{\Gamma}\right) = \mathcal{E}\left(\Psi^{\Gamma} \mathbb{1}_{\Gamma}\right). \end{aligned} \tag{22}$$

In these identities, $\Xi_{\Theta,\mathcal{V}}$ and $\Xi_{\Theta,\Gamma}$ are time-invariant linear operators that map firm types to sector types and product types, respectively. The elements of operator $\Xi_{\Theta,\mathcal{V}}$ are equal to numbers of sectors of type ν in which firm θ is present, divided by the total number of products in the corresponding variety market. Similarly, the elements of $\Xi_{\Theta,\mathcal{V}\Gamma}$ are equal to the numbers of products of type γ owned by firm θ.

We also can derive the BGP expressions for the product and variety type densities, denoted by f_{Γ} and

$f_{\mathcal{V}}$ respectively, and the product entry rate $\mathcal{E}_\Gamma$. The entry rate at the product level is equal to the ratio of products owned by the entrants to the total product mass:

$$\begin{gathered}\mathcal{E}_\Gamma = \left(\Psi^\Gamma \mathbb{1}_\Gamma\right)^{-1}, \\ f_{\mathcal{V}} = \left(\Psi^{\mathcal{V}} \mathbb{1}_{\mathcal{V}}\right)^{-1} \Psi^{\mathcal{V}}, \;\; f_\Gamma = \mathcal{E}_\Gamma \Psi^\Gamma.\end{gathered} \tag{23}$$

Finally, $\mathcal{P}_E^\Gamma = \mathcal{P}_E \Xi_{\Theta,\Gamma}$ denotes the product distribution for entering firms.

[Type Transitions and Entry along the decentralized BGP] Recall that in Section 3.4 we have assumed that firm type transitions and the entry distribution depend on four factors: the relative type distribution, entry rate, de-trended investment, and de-trended aggregate capital stock. We can reduce the set of arguments of $\mathcal{P}$ and $\mathcal{P}_E$ even further for the economy that moves along a balanced growth path. In particular, note that the firm type transitions and the entry rates have to satisfy equation 22. In the next paragraph, we also show that the economy's capital intensity on the BGP is only a function of type densities, entry rate, and investment. Thus, it is w.l.o.g. to assume that the BGP transition kernel and the BGP entry type distribution only depend on the investment levels and the exogenously fixed growth rates g_L and g_E:

$$\begin{aligned}\mathcal{P} &= P^{\text{BGP}}\left(z_\Theta, g_E, g_L\right), \\ \mathcal{P}_{Et} &= P_E^{\text{BGP}}\left(z_\Theta, g_E, g_L\right).\end{aligned} \tag{24}$$

4.3. Capital Intensity and TFP Growth on the BGP

Since on the balanced growth path capital has to grow at a rate g_E, we can characterize the BGP value of $\mathcal{K}$ as follows:

$$\exp^{g_E} = \mathcal{E}_\Gamma \bar{K}_E / \mathcal{K} + \exp^{g_E}\left(1 - \mathcal{E}_\Gamma\right) \mathbb{E}_{\Psi^\Gamma - \mathcal{P}_E^\Gamma}\left[k_\gamma\right]. \tag{25}$$

The term $\mathcal{E}_\Gamma \bar{K}_E / \mathcal{K}$ captures the contribution of entrants to the aggregate capital stock. The second term on the right-hand side of Equation 25 is the expression for the average capital stock of surviving incumbents, times the incumbent share in the total product mass.

Operators $\mathcal{P}$ and $\mathcal{P}_E$ describe the relative firm type transitions and assignments. These functions don't contain any information about the magnitude of changes in the absolute productivity levels across firms. We need to introduce an additional operator that tracks expected product-level absolute TFP improvements over time[8]. Let this operator be denoted by $\mathcal{P}^{\Delta\mathcal{A}} : \Gamma \to \mathbb{R}$. The productivity growth g_A solves the following equation:

$$1 = \mathcal{E}_\Gamma \bar{A}_E + \exp^{-g_A}\left(1 - \mathcal{E}_\Gamma\right) \mathbb{E}_{\Psi^\Gamma}\left[a_\Gamma \boldsymbol{\mathcal{P}}^{\Delta\mathcal{A}}\right]. \tag{26}$$

This equation is in many aspects similar to the statistical decompositions of the productivity growth that have been developed in the literature[9], except for the fact that here we prefer to use population frequencies as weights instead of sales or cost shares. The first term on the right-hand side is entrants' contribution to aggregate productivity, and the second term is incumbents' contribution. Since we assume that the

---

[8] Here we assume that the productivity of products that exit the market is normalized to 0.

[9] E.g., see Baily et al. [1992], Griliches and Regev [1995], Olley and Pakes [1996], Foster et al. [2008], or Melitz and Polanec [2015].

economy moves along the BGP, the terms associated with changes in the sales or, more generally, weight distribution over time are equal to zero: only within-product productivity improvements contribute to aggregate TFP growth.

4.4. Decentralized Equilibrium

[Firm-level Optimization: Short Run] Consistently with Assumption 1, markups and product level employment are determined by the following condition:

$$p_\gamma y_\gamma \frac{\omega_L}{\mu_\gamma} = w l_\gamma. \tag{27}$$

ω_L is the elasticity of output with respect to labor l_γ, and μ_γ is the marginal markup of product γ. The variable surplus earned by product γ is equal to

$$S_\gamma = \left(\frac{\mu_\gamma}{\omega_L} - 1\right) w l_\gamma \tag{28}$$

In the expression above, the ratio μ_γ / ω_L is equal to the average markup of good γ. As it turns out, average markups act as an essential statistic in the description of the second-best allocation and most of our comparative statics results. Thus, we introduce an additional bit of notation: in the rest of the paper, the average markups are denoted by ζ_γ.

[Firm-level Optimization: Long Term] The equilibrium investment levels satisfy the following condition:

$$\forall \iota \in \mathcal{I}: \qquad w z_\Theta(\iota) = \exp^{-r} \Omega_{\mathcal{P}}^{Z\mathrm{Own}}(\iota)\, V_\Theta, \tag{29}$$

where $\Omega_{\mathcal{P}}^{\mathrm{Own}}(\iota)$ is the operator that contains the derivatives of transition probabilities with respect to the log of firms' *own investment* of type ι. $\Omega_{\mathcal{P}}^{\mathrm{Own}}$ denotes the sum of operators $\Omega_{\mathcal{P}}^{\mathrm{Own}}(\iota)$ for all investment types $\iota \in \mathcal{I}$. To ensure the (local) uniqueness of decentralized equilibrium, we need to impose an additional assumption on the transition probability matrix:

Assumption 4. ***["Concavity" of Transition Probabilities]*** *For all feasible investment levels and entry rates, the elements of the operator $\mathcal{P} - \Omega_{\mathcal{P}}^{\mathrm{Own}}$ are positive, and the norm of the operator $\mathcal{P} - \Omega_{\mathcal{P}}^{\mathrm{Own}}$ is below one.*

This restriction is analogous to the assumption on the convexity of investment costs typically imposed in other settings since, informally, it implies that the derivatives of transition probabilities are small enough in absolute value. Note that here by "own" investment, we mean the investment of a specific firm instead of the average investment within a corresponding firm type.

If the concavity restriction holds, we can express producer values and investment on the decentralized BGP in the following way:

$$\begin{gathered} V_\Theta = \left(\mathrm{Id} - \exp^{-r}\left(\mathcal{P}_\Theta - \Omega_{\mathcal{P}}^{Z\mathrm{Own}}\right)\right)^{-1} S_\Theta, \\ w z_\Theta(\iota) = \Omega_{\mathcal{P}}^{Z\mathrm{Own}}(\iota) \left(\exp^{r} \mathrm{Id} - \mathcal{P} + \Omega_{\mathcal{P}}^{Z\mathrm{Own}}\right)^{-1} S_\Theta. \end{gathered} \tag{30}$$

To keep the notation concise, we abbreviate the expression for investment as

$$wz_{\Theta}\left(\iota\right)=\Psi^{Z\mathrm{Own}}\left(\iota\right)S_{\Theta}. \tag{31}$$

The equations above relate dynamic firm choices to their short-term surplus levels S_{Θ}. The elasticities of firm's investment with respect to producers' short term surpluses are summarized by the operator $\Psi^{Z\mathrm{Own}}\left(\iota\right)=\Omega_{\mathcal{P}}^{Z\mathrm{Own}}\left(\iota\right)\left(\exp^{r}\mathrm{Id}-\mathcal{P}+\Omega_{\mathcal{P}}^{Z\mathrm{Own}}\right)^{-1}$.

We can also show that the producer value functions and investment levels satisfy several intuitive properties using these equations. E.g., conditional on the values of transition probabilities and surpluses, both value function and investment levels decline in the value of interest rates. Similarly, investment and values also increase whenever the exit probabilities decrease, i.e., when the elements of $\mathcal{P}$ increase uniformly. Finally, an increase in the short-term surpluses leads to an upward shift in the firm value function, and as a result, investment levels should increase.

We can also rewrite the free entry condition in terms of the short-term surplus levels:

$$\begin{aligned}w\mathcal{L}_{E}&=\exp^{r}\mathcal{P}_{E}\left(\exp^{r}\mathrm{Id}-\mathcal{P}+\Omega_{\mathcal{P}}^{Z\mathrm{Own}}\right)^{-1}S_{\Theta},\\ w\mathcal{L}_{E}&=\Psi^{E}S_{\Theta}.\end{aligned} \tag{32}$$

Thus, the value of entry has the same properties as the value function for incumbents: higher short-term surpluses, lower exit probabilities, and lower interest rates generate more entry, ceteris paribus.

The equations stated in this section lead to the following characterization of the decentralized equilibrium:

Proposition 4.2. ***[Decentralized Allocation]*** *The labor allocation functions that generate the decentralized balanced growth path solve the following system of equations:*

$$\begin{aligned}\zeta_{\gamma}l_{\gamma}w_{\gamma}&=p_{\gamma}y_{\gamma},\\ z_{\Theta}\left(\iota\right)&=\Psi^{Z\mathrm{Own}}\left(\iota\right)\frac{S_{\Theta}}{w},\\ \mathcal{L}_{E}&=\Psi^{E}\frac{S_{\Theta}}{w}.\end{aligned} \tag{33}$$

The operators Ψ^{E} and $\Psi^{Z\mathrm{Own}}\left(\iota\right)$ only depend on the investment functions and the exogenous growth rates g_E and g_L. At least one solution to this system of equations always exists. In addition, when the law of motion for capital stocks has a standard linear additive form, the existing balanced growth path equilibria will always feature strictly positive and finite aggregate capital stock value $\mathcal{K}_t$.

The existence of the BGP(s) follows from Schauder fixed point theorem since the set of possible allocation functions is compact, convex, and non-empty, and since we have assumed continuity for all the operators that are used to construct the equations in Proposition 4.2.

A formal definition for the ”linear additive” law of motion for capital stocks is as follows. We assume that the production technology for capital goods is linear in labor so that the amount of capital goods produced with $z^K_{\theta t}$ units of labor is equal to $A^K_t z^K_{\theta t}$, where A^K_t is some constant. At the product level, the law of motion for the capital state is as follows. For product γ of incumbent firm θ, conditional on the survival of the product, we have

$$\tilde{k}_{\gamma(t+1)} = \tilde{k}_{\gamma t}\left(1 - d_{\gamma t}\right) + z^K_{\theta\gamma t}, \tag{34}$$

where $d_{\gamma t}$ is a realization of random depreciation rate d^K with support $[0, 1)$. We introduce stochastic depreciation rates primarily to ensure that the transition probabilities are continuous in future types. For the entrants, the relative capital stock is simply a draw from a fixed distribution that is given by a marginal of $\mathcal{P}_E$. Note that the equation above explicitly states that capital investment is separate across products – this assumption is still in line with our setup.

4.5. Socially Optimal Balanced Growth Path

4.5.1. First Best

In the first best equilibrium, the social planner chooses labor allocation functions by maximizing the balanced growth path welfare:

$$\begin{aligned} \max_{\{l_\gamma\}_{\gamma\in\Gamma}, \{z_\theta\}_{\theta\in\Theta}} \quad & \left(1 - \exp^{-\delta + (1-\vartheta)(g_Y - g_L)}\right)^{-1} \frac{(Y/L)^{1-\vartheta}}{1-\vartheta}, \\ \text{s.t.} \quad & \mathcal{M}_E = \left(\mathcal{L}_E + \Psi^\Gamma l_\Gamma + \sum_{\iota\in\mathcal{I}} \Psi^\Theta z_\Theta(\iota)\right)^{-1}. \end{aligned} \tag{35}$$

To keep the definition of the first best concise, we have omitted all the technology constraints and the constraints that ensure that the economy is on the balanced growth path from the equation above. Note that here we have renormalized the aggregate productivity and the size of the labor force in the initial period to 1.

The first-best allocation is characterized by the following proposition:

Proposition 4.3. [First Best] *The allocations of production labor and investment that implement first best solve the following system of equations:*

$$\begin{aligned} FOC_{l_\gamma}: \quad & \lambda_\gamma \omega_L = \Lambda^Y \lambda^l_\gamma \frac{\sigma_\mathcal{V}}{\sigma_\mathcal{V} - 1}, \\ FOC_{z_\theta}: \quad & \Lambda^Z \lambda^Z_\theta d\log z_\theta + \Lambda^Z \mathbb{E}_{\lambda^Z_\Theta}\left[d\log\Psi^\Theta\right] + \Lambda^Y \mathbb{E}_{\lambda^l_\Gamma}\left[d\log\Psi^\Gamma\right] = \\ & \left(1 - \frac{1}{\sigma_\mathcal{V}}\right)\left(\Lambda^F dg_A + \omega_K d\log\mathcal{K}\right) + \mathbb{E}_{\lambda_\mathcal{V}}\left[d\log\Psi^\mathcal{V}\right], \end{aligned} \tag{36}$$

where λ^Z_θ is the share of investment type z_θ in all investment, and, similarly, λ^l_γ is the share of labor employed in production of good γ in all production labor. λ_ν and λ_γ are the sales shares of sector ν and product γ

in aggregate output, respectively. Λ^Z *and* Λ^Y *denote the aggregate shares of investment and production employment.* Λ^F *is the weight on future periods in social planner's objective, as in the toy model:*

$$\Lambda^F = \frac{\exp^{-\delta+(1-\vartheta)(g_Y - g_L)}}{1 - \exp^{-\delta+(1-\vartheta)(g_Y - g_L)}}. \tag{37}$$

For simplicity, in all subsequent derivations, we will assume that the social planners' problem features an interior solution, so that on the first-best balanced growth path, the function $z_\Theta : \Theta^{\mathcal{I}} \to \mathbb{R}$ is strictly positive. This assumption holds for our model calibration. Here we also abuse the notation a little and let z_θ denote the investment of firm θ in some project, i.e., z_θ is some element of the allocation function z_Θ that is defined on $\Theta^{\mathcal{I}}$. $\lambda^Z_\Theta : \Theta^{\mathcal{I}} \to \mathbb{R}^+$ denotes the function that contains the cost shares for all investment types and all firm types.

The first-order conditions of the social planner's optimization reveal several essential features of the first-best allocation. First, along the first-best BGP, the distribution of production labor is determined by the following equation:

$$\mathrm{FOC}_{l_\gamma}: \quad \lambda_\gamma \omega_L = \Lambda^Y \lambda^l_\gamma \frac{\sigma_{\mathcal{V}}}{\sigma_{\mathcal{V}} - 1}. \tag{38}$$

The condition above suggests that, along the decentralized balanced growth path, markups could lead to misallocation of production labor due to two reasons, as in Edmond et al. [2021]. First, markup heterogeneity across producers and goods generates discrepancies between relative employment levels at first-best allocation and decentralized equilibrium. Indeed, the markups that implement first-best equilibrium are always homogeneous across products. Second, unlike the static models without entry, the allocative efficiency of the economy that we consider depends on the level of markups. Even if markup dispersion is zero, the employment levels do not match with the social optimum unless markups satisfy the following optimality condition

$$\mu_\gamma (1 - \pi) = \frac{\sigma_{\mathcal{V}}}{\sigma_{\mathcal{V}} - 1}, \tag{39}$$

where μ_γ is the homogeneous producer markup, and π is the aggregate profit rate, defined as $\pi = 1 - Lw/PY$. Thus, the FOC for production labor implies that the markup that implements the first-best allocation is equal to the love-of-variety elasticity $\frac{\sigma_{\mathcal{V}}}{\sigma_{\mathcal{V}}-1}$ multiplied by the factor $1/(1-\pi)$. This condition is noticeably similar to the Dixit and Stiglitz [1977] result and other CES-based optimality conditions that have been derived in the literature. Just like in Dixit and Stiglitz [1977], the social planner in our setting faces a tradeoff between encouraging entry and increasing firm-level output, and the amplitude of love-of-variety effects regulates this tradeoff. The profit rate multiplier $1/(1-\pi)$ readjusts the Dixit and Stiglitz [1977] result in the presence of investment and firm type dynamics in our framework. In the endogenous growth settings with stochastic TFP dynamics, the profit rate is proportional to firms' net producer surplus. The profit rate is weakly positive, and it is strictly above zero if the type distribution among incumbents does not coincide with the entrant type distribution. Thus, in a sense, the aggregate profit rate is zero only if incumbents' investment is beneficial for society, i.e., if it leads to "better" firm type distribution. The equation above suggests that the first-best markups should be higher whenever the profit rate is high. Unless the profit rate in the economy is equal to zero, the value of first-best markup is higher than in the static Dixit and Stiglitz [1977] benchmark.

Reiterating Proposition 4.3, investment levels at the social optimum are determined by the following first-order condition:

$$
\begin{aligned}
\mathrm{FOC}_{z_\theta}: \quad \Lambda^Z \lambda_\theta^Z \mathrm{d}\log z_\theta = {} & \underbrace{\left(1-\frac{1}{\sigma_{\mathcal{V}}}\right)\Lambda^F \mathrm{d}g_A}_{\text{Direct Growth Rate Effect}} + \underbrace{\left(1-\frac{1}{\sigma_{\mathcal{V}}}\right)\omega_K \mathrm{d}\log\mathcal{K}}_{\text{Capital Accumulation Effect}} \\
& + \underbrace{\mathbb{E}_{\Psi^{\mathcal{V}}}\left[\mathrm{d}\log\Psi^{\mathcal{V}}\right]}_{\text{Love-of-Variety Effect}} + \underbrace{\mathbb{E}_{\lambda_\nu}\left[\mathrm{d}\log\Psi^{\mathcal{V}}\right] - \mathbb{E}_{\Psi^{\mathcal{V}}}\left[\mathrm{d}\log\Psi^{\mathcal{V}}\right]}_{\text{Sectoral Reallocation Effect}} + \\
& - \underbrace{\left(\Lambda^Z \mathbb{E}_{\lambda_\Theta^Z}\left[\mathrm{d}\log\Psi^\Theta\right] + \Lambda^Y \mathbb{E}_{\lambda_\Gamma^l}\left[\mathrm{d}\log\Psi^\Gamma\right]\right)}_{\text{Labor Reallocation Effects}}.
\end{aligned}
\tag{40}
$$

The social value of the investment, summarized by the right-hand side of Equation 40, will be important in deriving the first and second-order welfare decompositions. We use $\Psi_\Theta^Z(z_\Theta, l_\Gamma)$ to denote the function that evaluates the elasticity of welfare with respect to investment, given the allocation of production and investment labor:

$$
\mathrm{FOC}_{z_\theta}: \quad \lambda_\theta^Z = \Psi^Z\left(z_\Theta^{FB}, l_\Gamma^{FB}\right). \tag{41}
$$

Equation 40 demonstrates that investment affects the allocation via four key channels. First, higher investment values lead to faster productivity growth, and, in case of capital investment, higher aggregate capital stock. This direct effect of investment on consumer welfare is summarized by the first two terms in Equation 40. Second, investment affects entry and product innovation, and thus it influences the number of final good varieties available to consumers at each moment in time. These love-of-variety effects are summarized by the 3rd term in Equation 40. Finally, steady state investment also affects the densities of variety, product and firm types, and thus, changes in investment lead to reallocation of production across sectors, and reallocation of labor – across firm and product types. These reallocation effects are summarized by the last two terms in the equation above. So, the sectoral reallocation effect suggests that investment is beneficial for society if it leads to an increase in the number of high-output sectors relatively to the number of low-output sectors. At the same time, investment is costly for society if it increases the density of product types that hire a lot of workers either in production or in investment, because this means that the social planner would not be able to allocate as much labor to entry sector. This channel is summarized by the labor reallocation effects.

It is important to note that, due to the nature of the social planner's problem, Equation 40 ultimately describes the tradeoff between investment and entry that is in many ways similar to the tradeoff between production employment and entry illustrated by Equation 39. So, in the limit case when $\sigma_{\mathcal{V}} \to 1$, the terms that measure the productivity growth and capital accumulation effects are close to zero, as well as the sectoral reallocation effect. Investment is only beneficial if it generates new sectors, and the costs of investment are proportional to the labor reallocation effects – and the direct cost of investment.

[First-Best: Implementation] To introduce the second-best equilibria, it is useful to reformulate the first best planner's problem in the following way. Suppose that, instead of directly selecting the production and investment labor distributions, the social planner chooses the values of *policy variables* that affect $\{l_\gamma\}_{\gamma\in\Gamma}$ and $\{z_\theta\}_{\theta\in\Theta}$. Assuming that the policies available to the social planner are sufficiently effective, they can implement the first-best allocation by resetting the policy variables to their "optimal" values. To make this discussion more formal, let us consider an economy with marginal cost pricing. Also, suppose the social planner can influence the allocation of production labor by altering the fixed product markups. In addition, they can also use investment subsidies (denoted by β) to change the allocation of investment labor. Then, the firm optimization condition satisfies

$$\beta_\Theta \lambda_\theta^Z = \Psi^{Z\text{Own}} \frac{S_\Theta}{w}. \tag{42}$$

We can then show that the first-best allocation can be implemented by optimal markup levels and optimal investment subsidies. The values of the optimal policy variables are pinned down by the first-order conditions of both the decentralized equilibrium and the social planner's optimization. The discussion above shows that the optimal markup values are homogeneous across producers, and the first-best optimal markup level satisfies:

$$\mu_\Gamma^{FB} = \frac{\sigma_\mathcal{V}}{\sigma_\mathcal{V} - 1} \frac{1}{1-\pi^{FB}}, \tag{43}$$

where π^{FB} is the aggregate profit rate evaluated at the first-best allocation.

Similarly, the expression for the optimal investment subsidies is pinned down by the FOCs of the firm and social planner's optimization problems:

$$\beta_\Theta^{FB} = \Psi^{Z\text{Own},FB} \frac{S_\Theta^{FB}}{\lambda_\theta^{Z,FB} w^{FB}}. \tag{44}$$

Here the superscripts FB again indicate that the respective variables are evaluated at the first-best allocation.

There exist multiple combinations of policies that can implement the first best. As an alternative to directly setting markups, we could allow the social planner to impose output or sales taxes on final goods. Similarly, in the setting with the generalized oligopolistic competition, the social planner might be able to influence the industry structure by changing the anti-trust regulations in the production sector. In turn, we can substitute investment subsidies with lump-sum transfers at the firm type level or profit taxes. In general, to reach the first-best allocation, the social planner needs to freely readjust the incentives of producers to hire the production workers in the short run and their motivation to invest – in the long run. This means that the social planner requires two policy tools that can alter (i) the marginal revenues or marginal costs at the product level and (i) the marginal returns to investment. Here it is helpful to compare the analysis and discussion above to the Edmond et al. [2021] result that shows that the first-best allocation in their setting can be implemented with a single output subsidy. The reason why in this case, the social planner can achieve socially-optimal allocation by using a single policy is that Edmond et al. [2021] allow for arbitrary non-linear subsidy schedules. This means that the policymaker can simultaneously change the marginal and total revenues of each product. Moreover, the changes in producers' marginal incentives can be independent of changes in their surplus levels. In other words, the non-linear subsidy functions like a combination of a linear output tax/subsidy and a lump sum transfer.

4.5.2. Second Best

As suggested in the previous section, the economy that we consider features two types of frictions that affect production and investment labor allocation. Policy variables μ and β that regulate markup levels and investment subsidies, respectively, can influence the magnitude of these frictions. Under the first best solution concept, both policy variables are set at the optimal levels, and thus, social welfare reaches its global maximum. Here we define the *second best* balanced growth path as an equilibrium in which the social planner sets only one of the policies (markups). In contrast, the other wedge (investment subsidies) remains at the level prescribed by the decentralized equilibrium. It is important to note that the second-best allocation is to some extent invariant to the choice policy tools that the social planner uses. As long as the values of markups match the second best, the resulting allocation does not change. The same applies to the average markups, provided that producer marginal costs always increase in the level of output. Thus, we can w.l.o.g. focus on a more manageable problem in which the economy operates under marginal cost pricing, and the social planner sets the values of average markups.

The social planner's second-best optimization problem as follows:

$$\begin{aligned}
\max_{\zeta_\Gamma} \quad & \left(1-\exp^{-\rho+(1-\vartheta)(g_Y-g_L)}\right)^{-1}\frac{(Y/L)^{1-\vartheta}}{1-\vartheta}, \\
\text{s.t.} \quad & z_\Theta = \Psi^{Z\text{Own}}\frac{S_\Theta}{w}, \\
& y_\gamma p_\gamma = \zeta_\gamma l_\gamma w, \\
& \mathcal{L}_E = \Psi^E \frac{S_\Theta}{w}, \\
& (\mathcal{M}_E)^{-1} = \mathcal{L}_E + \Psi^\Gamma l_\Gamma + \sum_{\iota\in\mathcal{I}} \Psi^\Theta z_\Theta(\iota) .
\end{aligned} \tag{45}$$

The first-order conditions derived from this problem are in many ways similar to the comparative statics with respect to markup levels around the decentralized equilibrium:

Proposition 4.4. [Second Best] *The second-best markup levels solve the following system of equations:*

$$\begin{aligned}
\text{FOC}_{\mu_\Gamma}: \quad & \Lambda^Y \mathbb{E}_{\lambda^l_\Gamma}\left[\left(\frac{\lambda_\gamma}{\lambda^l_\gamma}\omega_L - \Lambda^Y\frac{\sigma_\mathcal{V}}{\sigma_\mathcal{V}-1}\right)\mathrm{d}\log l_\gamma\right] + \\
& + \Lambda^Z \mathbb{E}_{\lambda^Z_\Theta}\left[\left(1-\left(\lambda^Z_\Theta\right)^{-1}\Psi^Z_\theta\right)\mathrm{d}\log z_\theta\right] = 0.
\end{aligned} \tag{46}$$

The equation above describes the optimality condition in terms of the elasticities of labor and investment allocations. The first term is equal to the cost-share weighted average of deviations from the optimal production labor allocation. Because of the nature of the optimality condition for production labor, we can interpret this term as a measure of the deviation of markups from their first-best level. Similarly, the second term measures the deviation of investment levels from their socially optimal values. In the equation above, the operator Ψ_{θ}^{Z} should be evaluated at the second-best values of l_{Γ} and z_{Θ}.

Equation 46 conveys several important messages about the second-best allocation. First, unlike the first-best level of markups, the second-best markups are heterogeneous across products and firms. The differences between the first-best markup levels and their second-best values depend on the (i) deviation of investment from the social optimum, (ii) elasticity of investment and production labor allocation with respect to markups. Intuitively, in the second-best, the social planner is forced to use a limited set of policies to compensate for the misallocation of both production labor and investment. Also, since we would expect $\sigma_{\mathcal{V}}$ to be above one, and since the profit rate is always above zero for the endogenous growth frameworks, we would expect the second-best markup values to be above one.

We can characterize the elasticities $\mathrm{d}\log l_{\gamma}$ and $\mathrm{d}\log z_{\theta}$ as follows:

Proposition 4.5. ***[Propagation Equations: Second Best]*** *At the second best optimum, the allocation function differentials $d\log l_{\gamma}$ and $d\log z_{\theta}$ solve the system of differential equations described below, for all possible values of $d\log\zeta_{\theta}$. Let $\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}$ denote the operator that computes the deviation of the function value from its $\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}$-weighted average, then*

$$\begin{aligned}\left(Id_{\Gamma}-\omega_{L}\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\Phi^{Y}\right)d\log l_{\gamma} &= -\mathbb{1}_{\Gamma}\mathbb{E}_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Theta}}\left[\frac{d\log\Psi^{E}}{d\log z_{\Theta}}\right]d\log z_{\Theta}+ \\ &\quad -\left(\boldsymbol{\zeta}_{\gamma}-\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\right)\left(\boldsymbol{\zeta}_{\gamma}-Id\right)^{-1}d\log\zeta_{\gamma}, \\ \left(Id_{\Theta^{\mathcal{I}}}-\Phi_{ZZ}\right)d\log z_{\Theta} &=\mathbb{E}_{\Psi_{\Gamma}^{ZOwn}\boldsymbol{S}_{\Gamma}}\left[\frac{\zeta_{\gamma}}{\zeta_{\gamma}-1}d\log\zeta_{\gamma}\right]+\mathbb{E}_{\Psi_{\Gamma}^{ZOwn}\boldsymbol{S}_{\Gamma}}\left[d\log l_{\gamma}\right].\end{aligned} \tag{47}$$

In these equations, Φ^{Y} is the Hessian operator for the aggregate output with respect to the product outputs y_{Γ}, Ψ_{Γ}^{E} and Ψ_{Γ}^{ZOwn} are the maps of operators Ψ^{E} and Ψ^{ZOwn} on the product space, and the operator Φ_{ZZ} is defined as follows:

$$\Phi_{ZZ}=\mathbb{E}_{\Psi^{\mathrm{ZOwn}}\boldsymbol{S}_{\Theta}}\left[\frac{\mathrm{d}\log\Psi^{\mathrm{ZOwn}}}{\mathrm{d}\log z_{\Theta}}\right]. \tag{48}$$

Here it is useful to clarify the definition of the operator $\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}$. Formally, we have:

$$\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}=\mathrm{Id}_{\Gamma}-\left(\Psi_{\Gamma}^{E}S_{\Gamma}\right)^{-1}\mathbb{1}_{\Gamma}\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}. \tag{49}$$

The system of equations above has a well-defined solution if the norms of the operators $\omega_{L}\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\Phi^{Y}$ and Φ_{ZZ} are contraction mappings, or in other words, if the norms of these operators are below one. The norm of $\omega_{L}\Phi^{Y}$ is below one, since the elements of this operator are proportional to the elements of the aggregate output Hessian:

$$\omega_{L}\left[\Phi^{Y}\right]_{\gamma\gamma'}=\omega_{L}\frac{\partial\log Y}{\partial\log y_{\gamma}\partial\log y_{\gamma'}}. \tag{50}$$

For the norm of Φ_{ZZ} to be below one, it has to be the case that firm investments are, in a sense, sufficiently substitutable, i.e., for each $\theta\in\Theta$ and each $\iota\in\mathcal{I}$ $z_{\theta}\left(\iota\right)$ does not respond strongly to changes in other investment rates.

The equations above demonstrate how markup shocks alter the labor allocation. We start our discussion with an interpretation of the production labor equation. First, Equation 47 suggests that the direct effect of an increase in average markups is negative: an increase in the price-cost margins drives the employment down. In addition, an increase in markups triggers an increase in surplus levels, leading to an upward shift in wages. The increase in labor costs further reduces product-level employment. In the equations above, the effects of changes in production labor allocation on wages are described by the expectation term $\omega_L \mathbb{E}_{\Psi^E_\Gamma \boldsymbol{S}_\Gamma}\left[\Phi^Y \mathrm{d}\log l_\gamma\right]$. Notably, in settings without entry, the wage movements tend to offset the direct effects of markup shocks, not amplify them. Substitution effects between final goods, summarized by the operator $\omega_L \Phi^Y$ also lead to an amplification of markup shocks because the decline in the sales of product γ leads to an increase in other goods' outputs. Finally, the operator $\mathbb{1}_\Gamma \mathbb{E}_{\Psi^E_\Gamma \boldsymbol{S}_\Theta}\left[\frac{\mathrm{d}\log \Psi^E}{\mathrm{d}\log z_\Theta}\right]$ in Equation 47 describes the feedback loop between changes in investment and changes in production employment under the markup shocks. This term represents the effect of changes in investment $\mathrm{d}\log z_\Theta$ on wages. The sign of this term is ambiguous since higher levels of incumbents' investment could both increase and decrease the expected value of a firm for entrants, conditional on surplus levels. The direction of the feedback effect between $\mathrm{d}\log z_\Theta$ and $\mathrm{d}\log l_\Gamma$ is ambiguous since the impact of investment wages is essentially a ”composition” effect.

In contrast to production employment, the direct effect of markup shocks on investment is positive since an increase in surpluses induces producers to spend more to avoid the exit shocks and retain their position in the product markets. In addition to the direct effect of markup changes, the investment values also depend on the reaction of production labor to the shock: whenever production employment increases, producer surpluses increase, and so do the investment rates. The differentials $\mathrm{d}\log l_\gamma$ likely offset the direct effect of markup changes in the investment equation because the effect of markup shocks on labor is negative, and the feedback between investment and labor (in the labor equation) is not likely to be large in magnitude.

[Fundamentals for Comparative Statics] Equation 47 allows us to understand what fundamentals determine the second-best level of markups, average or marginal. The list of the primitives that define the ”optimal” markup distribution includes

(i) the statistics that typically affect shock propagation in the static settings, including the Hessian of the aggregate output and production function elasticity with respect to variable inputs; the elements of the aggregate output Hessian determine the quasi-demand elasticity at a product-type level, and the substitution elasticities, which are used as sufficient statistics in Baqaee and Farhi [2020b] and Baqaee and Farhi [2020a].

(ii) a range of statistics that are specific to the endogenous growth settings, including the type density operators Ψ^Θ and $\Psi^\mathcal{V}$, operators Ψ^E and $\Psi^{Z\mathrm{Own}}$ that determine the elasticity of investment and wages with respect to short term surplus rates, and the operator Φ_{ZZ} that depends on the substitutability between investment types.

In general terms, Proposition 4.5 suggests that the endogenous growth settings feature a distinct set of primitives that determine the reaction of the economy to technology shocks and/or changes in allocation frictions. These primitives include the transition kernel $\mathcal{P}$ and the entry type distribution $\mathcal{P}_E$.

5. Welfare Decompositions

This section presents several first-order welfare decompositions that allow us to analyze the channels through which producer market power affects social welfare.

5.1. Second Best

We use the distance to the second-best allocation as our primary measure of welfare costs of market power. In addition to computing this difference in a full non-linear model, we also conduct several exercises to understand why sub-optimal market power is harmful to consumers. Specifically, we compute the first-order approximation for the distance to the second best. Then we decompose the first-order welfare differential into the terms that correspond to different types of misallocation.

The total elasticity of social welfare with respect to markups can be derived from the characterization of the second-best equilibrium. First, note that, regardless of the assumptions on production and consumption structure, the first-order approximation of the difference between decentralized and second-best welfare levels satisfies the following equation:

$$\begin{aligned}\frac{\mathrm{d}\log|\mathcal{W}^{DE}| - \mathrm{d}\log|\mathcal{W}^{SB}|}{\vartheta-1} \approx{}& \Lambda^Y \mathbb{E}_{\lambda^l_\Gamma}\left[\left(\frac{\lambda_\gamma}{\lambda^l_\gamma}\omega_L - \Lambda^Y\frac{\sigma_\mathcal{V}}{\sigma_\mathcal{V}-1}\right)\frac{\mathrm{d}\log l_\gamma}{\mathrm{d}\log\mu_\Gamma}\log\frac{\mu^{SB}_\Gamma}{\mu_\Gamma}\right] + \\ &+ \Lambda^Z\mathbb{E}_{\lambda^Z_\Theta}\left[\left(1-\left(\lambda^Z_\theta\right)^{-1}\Psi^Z_\theta\right)\frac{\mathrm{d}\log z_\theta}{\mathrm{d}\log\mu_\Gamma}\log\frac{\mu^{SB}_\Gamma}{\mu_\Gamma}\right].\end{aligned} \tag{51}$$

This is a direct consequence of Proposition 4.4, since the second-best is defined by setting the elasticity of social welfare with respect to markups to zero. Moreover, whenever the production function has a Cobb-Douglas form, the conditions that determine the elasticities of allocation with respect to markups $\frac{\mathrm{d}\log l_\gamma}{\mathrm{d}\log\mu_\Gamma}$ and $\frac{\mathrm{d}\log z_\theta}{\mathrm{d}\log\mu_\Gamma}$ also can be evaluated based on our analysis in section 4.5.2:

Corollary 2. ***[Distance to 2nd Best]*** *The elasticities $\frac{d\log l_\gamma}{d\log\mu_\Gamma}$ and $\frac{d\log z_\theta}{d\log\mu_\Gamma}$ solve the system of propagation equations in Proposition 4.5 with $d\log\zeta_\gamma = d\log\mu_\gamma$.*

This result holds because, under the CD production, the elasticity of product-level output with respect to variable labor input is constant. There are no differences in capital intensity across firms. Thus, average markups are proportional to marginal markups. In contrast, under the CES production structure, the values of average markups depend on the distribution of relative stocks of fixed assets and the aggregate capital stock value. The interpretation of this Corollary replicates the discussion at the end of section 4.5.2. The direct effects of markup shocks are positive for investment and negative – for production employment. The sign and magnitude of the effects of changes in investment on production labor allocation are ambiguous.

The feedback between production employment and investment tends to mitigate the markup shocks.

Given the propagation equations for the allocation functions $\mathrm{d}\log l_\Gamma$ and z_Θ, Equation 51 decomposes the difference in welfare levels $\mathrm{d}\log|\mathcal{W}^{DE}| - \mathrm{d}\log|\mathcal{W}^{SB}|$ into the terms that measure misallocation of production labor relative to the second best, and the corresponding misallocation of investment. Furthermore, the component that evaluates the effect of investment misallocation can be further decomposed into the effects of markup changes on the productivity growth rates, aggregate capital stock, sector-type density and producer mass, as in Equation 40.

5.2. First Best

Edmond et al. [2021], and Cavenaile et al. [2020] evaluate welfare losses due to market power by computing the distance between the decentralized allocation and the Pareto efficiency frontier. Our analysis suggests that such comparisons do not provide an accurate estimate of the welfare costs of producer market power because the social planner cannot implement the first-best allocation by readjusting the markup values. Still, whenever the decentralized allocation is sufficiently close to the social optimum, the differences between the decentralized allocation and the first-best are informative about misallocation generated by markups. For example, suppose that, following our discussion in section 4, the social planner implements the first-best allocation by resetting product markups and by introducing an investment subsidy. Then, the distance to the first-best can be decomposed into the terms that represent the misallocation due to (i) sub-optimal distribution of markups, and (ii) sub-optimal investment policies: suppose β_Θ denotes the investment subsidy, then

$$\mathcal{W}^{FB} - \mathcal{W}^{DE} \approx -\frac{\mathrm{d}\log|\mathcal{W}^{DE}|}{\mathrm{d}\log\mu_\Gamma}\left(\mu_\Gamma^{FB} - \mu_\Gamma\right) - \frac{\mathrm{d}\log|\mathcal{W}^{DE}|}{\mathrm{d}\log\beta_\Theta}\left(\beta_\Theta^{FB} - \beta_\Theta\right). \tag{52}$$

The first term in this decomposition can be computed similarly to the second-best approximation considered above – the only difference is that in this case, the markup shocks use the first-best markup values. Then, the following proposition allows us to evaluate the second term that measures the effects of investment subsidies (or taxes):

Proposition 5.1. ***[Investment Subsidies]*** *The elasticity of welfare with respect to investment tax or subsidy β can be evaluated as follows:*

$$\begin{aligned}\frac{1}{\vartheta-1}d\log|\mathcal{W}| = \Lambda^Y\mathbb{E}_{\lambda_\Gamma^l}\left[\left(\frac{\lambda_\gamma}{\lambda_\gamma^l}\omega_L - \Lambda^Y\frac{\sigma_\mathcal{V}}{\sigma_\mathcal{V}-1}\right)d\log l_\gamma\right] + \\ + \Lambda^Z\mathbb{E}_{\lambda_\Theta^Z}\left[\left(1-\left(\lambda_\theta^Z\right)^{-1}\Psi_\theta^Z\right)d\log z_\theta\right].\end{aligned} \tag{53}$$

The propagation equations for allocation elasticities $\frac{d\log l_\Gamma}{d\log\beta_\Theta}$ and $\frac{d\log z_\Theta}{d\log\beta_\Theta}$:

$$\begin{aligned}\left(Id_\Gamma - \omega_L\Sigma_{\Psi_\Gamma^E\boldsymbol{S}_\Gamma}\Phi^Y\right)d\log l_\gamma &= -\mathbb{1}_\Gamma\mathbb{E}_{\Psi^E\boldsymbol{S}_\Theta}\left[\frac{d\log\Psi^E}{d\log z_\Theta}\right]d\log z_\Theta \\ \left(Id_\Theta - \Phi_{ZZ}\right)d\log z_\Theta &= -d\log\beta_\Theta + \mathbb{E}_{\Psi_\Gamma^{Z\,Own}\boldsymbol{S}_\Gamma}\left[d\log l_\gamma\right].\end{aligned} \tag{54}$$

The propagation equations listed above are similar to Propositions 4.5. The only difference is that in this scenario, we consider a change in investment subsidies $\mathrm{d}\log\beta_\Theta$. This shock has a direct negative effect on investment – since the increase in β_Θ is effectively a reduction in subsidies or an increase in taxes on investment. Investment policies do not affect labor allocation directly, but the distribution of production employment is still altered due to the changes in labor wage.

5.3. Policy-Invariant Decompositions

Welfare decompositions that we have presented so far imply that the social planner can alter the product markups. In the case of the first-best allocation, they can also set subsidies on investment. Still, as was mentioned in Section 4, first and second-best allocations can be achieved by altering different sets of policies. Thus it might be useful to evaluate the importance of labor and investment misallocation without relying on the specific assumptions about the policy tools available to the social planner. To that end, we can compute the first-order approximations of welfare differences based on differences in the allocation functions instead of policies. So, we compute the first and second-best allocation functions $(l_\Gamma^{FB}, z_\Theta^{FB})$ and $(l_\Gamma^{SB}, z_\Theta^{SB})$, and the corresponding Taylor expansion for the welfare: e.g., for the second-best allocation we have,

$$\begin{aligned}\frac{\mathrm{d}\log|\mathcal{W}^{DE}| - \mathrm{d}\log|\mathcal{W}^{SB}|}{\vartheta - 1} &\approx \Lambda^Y \mathbb{E}_{\lambda_\Gamma^l}\left[\left(\frac{\lambda_\gamma}{\lambda_\gamma^l}\omega_L - \Lambda^Y \frac{\sigma_\mathcal{V}}{\sigma_\mathcal{V} - 1}\right)\log\frac{l_\gamma^{SB}}{l_\gamma}\right] + \\ &+ \Lambda^Z \mathbb{E}_{\lambda_\Theta^Z}\left[\left(1 - \left(\lambda_\theta^Z\right)^{-1}\Psi_\theta^Z\right)\log\frac{z_\theta^{SB}}{z_\theta}\right].\end{aligned} \tag{55}$$

As before, the first term corresponds to the misallocation of production labor, and the second one – to misallocation of investment. We can expand the decomposition as in Equation 40, if necessary.

6. Data, Estimation and Calibration

To map our theoretical setting to the data, we need to choose values of the following parameters:

- Consumer preference parameters: $\sigma_\mathcal{V}$, σ_Γ, ρ and ϑ.
- Production function parameters ω_l, and ξ.
- Markup values (μ_γ) at the product level.
- Parameters that determine the dynamics of firm types, product-level TFP, and fixed assets.
- Aggregate growth rates of population, entry costs, and productivity.

In order to calibrate values of labor, capital and demand elasticities, and the distributions of markups and TFP in the economy, we estimate firm-level production functions and sectoral demand using the Compustat Annual Fundamentals data on firms' balance sheets, BEA KLEMS data on quantity and price indices, and BLS producer price index data. In our estimation, we allow for the endogenous dynamics of firm productivity and knowledge spillovers across firms. We use USPTO data on granted patents and patent citations to compute sector and firm-level ”knowledge-transfer” weights, defined similarly to Bloom et al. [2013]. To ensure that our results are robust to sectoral differences in markups, demand, and production processes, we implement our production function estimation routine separately for 15 upper-tier industries associated with 2-digit NAICS sectors. Table 16 Appendix H.3 shows the mapping between 15 upper-tier sectors, and BEA KLEMS classification. The 5-digit NAICS industries represent the varieties of final goods.

To calibrate the parameters that govern the innovation process, we need to define the product and firm types in the data. We assume that investments in capital, R&D, and intangible inputs are product-specific in the benchmark calibration. We also assume away scope economies within firms. Thus, firms’ investment choices are independent of the number of product lines owned by firms, and firm types are isomorphic to product types. In turn, product types are defined based on the capital stock of the product, its productivity, and the characteristics of the associated variety market. Section 6.3 describes the calibration of firm types, product types, and the Markov process operators in detail.

This project examines the dynamics of misallocation losses in the US economy over the last three decades. Thus, we calibrate our model using the data from two sub-samples. The ”early” sub-sample covers the period from 1982 to 1997, and the ”late” sub-sample covers 2002 to 2017.

6.1. Firm-Level Data

[Compustat Firm-level Data] Compustat is used primarily as a source of data on the usage of production inputs by US firms. We assume that, apart from unobserved TFP, physical output depends on three inputs, including capital and the input bundles represented by the Compustat data items ”COGS” and ”XSGA.” In general terms, the ”costs of goods sold” item includes the costs of materials, compensation for production workers, and all other expenditures directly related to the production of goods. ”XSGA” item consists of the accounting costs, advertising, delivery, and distribution expenses, as well as all other variable costs that scale up with physical output but are not directly related to production. Consistently with these definitions, and consistently with the literature[10], we treat the expenditures listed in the ”COGS” category as a measure of variable costs of production. We also assume that data item XSGA represents expenditures on dynamic inputs set either at the beginning of the current period t or at the end of the period $t-1$. A measure of capital stock is constructed via a perpetual inventory method using gross and net values of capital stock from Compustat (data items ”PPEGT” and ”PPENT”).

We also use Compustat data items as proxies for firm TFP. These items include reported capital expenditures, SGA expenditures, R&D expenditures, average R&D expenditures across all firms within the same KLEMS sector, average R&D expenditures across all other sectors, weighted with sector-level citation

---

[10] Traina [2018] assumes that XSGA and COGS represent expenditures on the same bundle of variable inputs and uses the sum XSGA+COGS as a measure of variable input usage. In their benchmark, De Loecker et al. [2020] assume that XSGA item represents fixed costs of production that don’t affect the value of firm output in the short run. In their robustness checks De Loecker et al. [2020] also show that the specifications that treat both XSGA and COGS as measures of variable inputs produce results that contradict the first-order conditions derived from firm optimization problems. We adhere to the compromise approach and treat XSGA as a dynamic input given this diversity of methods.

rates, as well as lagged values of these variables. To control for selection into Compustat, we reweigh firm-level observations when we compute the average R&D expenditures. The weights that we use are based on employment values reported in Compustat and the BDS data on the distribution of US firms across employment classes[11].

In our production function estimation routine, we assume that each Compustat observation corresponds to one product line. The structure of the data is naturally at odds with this assumption since Compustat data are collected at a firm level, and, as was pointed out by Bernard et al. [2010] and Ding [2020], multi-product and multi-industry firms comprise a significant share of the total US firm population. They produce more than half of the manufacturing output. Still, we attempt to control firms' product scope by excluding the firms that report owning business segments with different 3-digit NAICS codes. Our estimation framework also does not allow for demand heterogeneity, and thus, we filter out firms that report selling more than 25% of their output to foreign entities.

[UPSTO Data]In our theoretical framework and estimation, we allow for the presence of non-pecuniary externalities in TFP dynamics. We assume that the magnitude of external effects generated by firms θ''s research and θ is proportional to the R&D expenditures of firm θ' times the knowledge transfer rate between the firms θ and θ'. The UPSTO data on patents of US firms is used to evaluate the knowledge transfer rates. In our benchmark specification, we adhere to the methodology of Jaffe [1986], and Bloom et al. [2013] and compute the knowledge spillover weights based on the firm patent counts in different technology classes. In an alternative specification, the spillover weights are computed using the citation counts of patents.

6.2. Markup, Production Function and Demand Estimation

Our estimation procedure is based on the ACF-corrected production function estimation algorithm developed in Ackerberg et al. [2015], and R&D-augmented production function estimation suggested by Doraszelski and Jaumandreu [2013] and Buettner [2004]. We further augment the estimation procedure in order to address the critiques of Bond et al. [2021] and De Ridder et al. [2021] that concern on the potential effect of the output price bias on markup estimates. The estimation algorithm described below is implemented separately for 15 sector groups[12].

In the benchmark specification, all firm inputs, including COGS, capital, and XSGA, are aggregated via Cobb-Douglas, so that w.l.o.g. SGA's contribution to output could be thought of as a part of TFP. The consumer preferences are of the CES form. We describe the production function estimation procedure under non-linear production and demand specifications in Appendix Section H. The sales shares at the

---

[11] Firm-level employment weights are constructed as follows. First, given the Compustat data on employment, we compute the shares of Compustat firms within each BDS employment class, e.g., firms with 20 to 99 employees, firms with 100 to 499 employees, and so on. To avoid the small sample bias, we joined all the BEA employment classes with less than 20 employees into one class. Given the employment class shares in Compustat (denoted by $f_i^{L,\text{Compustat}}$) and in BDS (denoted by $f_i^{L,\text{BDS}}$), the computed weights are proportional to the ratios $f_i^{L,\text{BDS}}/f_i^{L,\text{Compustat}}$.

[12] The definitions of sector groups are listed in Table 16 in the Appendix: the sector groups are aligned with the 2-digit NAICS sectors whenever Compustat sample allows it.

product level are determined as follows:

$$\lambda_{\nu\gamma t} = \left(\frac{y_{\gamma t}}{Y_{\nu t}}\right)^{1-\frac{1}{\sigma_{\Gamma}}}, \tag{56}$$

Thus, the product sales shares primarily depend on the physical output produced by firms ($y_{\gamma t}$) and on the sectoral output index $Y_{\nu t}$. Note that this implies that firm-level and sectoral output can serve as precise controls for the output prices. To fit this equation to the data and pin down the demand parameters, we have to determine the production function specification and the generating process for TFP. Since firm productivity is unobserved, we also need to construct a proxy function for firm TFP.

[TFP] We assume that the data generating process for productivity is consistent with the assumptions stated in Section 3.4. Producer's state variables include productivity and capital stock and the "variety type" of the corresponding sub-sector. Firms' investment then is determined by the firm's state and firms' expectations over their competitors' investments, aggregate investment, and aggregate productivity growth. Formally, the best response investment functions satisfy

$$\begin{aligned} z^{\text{SGA}}_{\gamma t} &= \bar{Z}^{\text{SGA}}\left(a_{\gamma t}, k_{\gamma t}, \nu, Z\right), \\ z^{K}_{\gamma t} &= \bar{Z}^{K}\left(a_{\gamma t}, k_{\gamma t}, \nu, Z\right), \\ z^{A}_{\gamma t} &= \bar{Z}^{A}\left(a_{\gamma t}, k_{\gamma t}, \nu, Z\right). \end{aligned} \tag{57}$$

In the data, firms' R&D investment and SGA expenditure vary conditional on capital investment, and capital investment has a non-zero variance conditional on R&D and SGA expenditures. We assume that the unobserved differences in producer types generate this variation. E.g., it could be the case that some firms own technologies that function better with a larger stock of capital or with a larger stock of intangible assets; alternatively, such conditional variation in investment could also originate from the differences in costs of loanable funds. Thus, we will include SGA, capital and R&D investment in the set of TFP proxies. We assume that either conditional on $z^{K}_{\gamma t}$ and $z^{\text{SGA}}_{\gamma t}$, R&D expenditures are strictly monotone in product TFP $a_{\gamma t}$, or similarly, conditional on the value of R&D and SGA, the capital expenditures are a strictly monotone function of product TFP. Inverting either of the equations above conditional on the rest of the arguments of $\bar{Z}^{\text{SGA}}$, $\bar{Z}^{K}$ or $\bar{Z}^{A}$, and taking logs, we obtain the expression for the R&D-augmented TFP proxy function:

$$\log a_{\gamma t} = \mathfrak{A}\left(z^{\text{SGA}}_{\gamma t}, z^{K}_{\gamma t}, z^{A}_{\gamma t}, k_{\gamma t}, \nu, Z\right). \tag{58}$$

To note, due to the multi-dimensionality of the productivity and capital stock distributions within sectors, it is not possible to fully control for the differences in variety types. Instead, we will use the sub-sector-level moments of capital stock, capital expenditures and R&D to proxy for sector type differences.

[Output] Physical output $y_{\gamma t}$ is determined as follows:

$$\log y_{\gamma t} = \omega^{L}_{t} \log l_{\gamma t} + \omega^{K}_{t} \log k_{\gamma t} + \mathfrak{A}\left(z^{\text{SGA}}_{\gamma t}, z^{K}_{\gamma t}, z^{A}_{\gamma t}, k_{\gamma t}, \nu, Z\right). \tag{59}$$

Then, Equation 56 is mapped to the data in the following fashion:

$$\begin{aligned}\log \lambda_{\nu\gamma t} = \left(1-\frac{1}{\sigma_t}\right)\big(\omega^L_t \log l_{\nu\gamma t} + \omega^K_{\nu t}\log k_{\gamma t} + \\ + \mathfrak{A}\left(z^{\text{SGA}}_{\gamma t}, z^K_{\gamma t}, z^A_{\gamma t}, k_{\gamma t}, \nu, Z\right) - \log Y_{\nu t}\big) + e_{\gamma t}.\end{aligned} \tag{60}$$

$e_{\gamma t}$ is the measurement error term.

The estimation equation above relies on both firm-level data on inputs and investment, and the sectoral quantity indices. The coefficient on $\log Y_{\nu t}$ allows us to derive an estimate of product-level substitution elasticity σ_t, as was originally suggested by Klette and Griliches [1996] – later on, similar methodology was also adapted by De Loecker [2011] and Gandhi et al. [2020]. In turn, the coefficients on production inputs identify elasticities ω^L_t, ω^{SGA}_t and ω^K_t. In practice, we construct a proxy for $\log Y_{\nu t}$ using the BLS PPI data on 5-digit industry level and Census SUSB data on the sales shares of 5-digit industries in 2002, 2007, 2012 and 2017.

Consistently with the Ackerberg et al. [2015] methodology, equation 60 is estimated in two steps. In the first step, we identify and filter out measurement error term $e_{\gamma t}$ by non-parametrically regressing the producer sales shares on production inputs, sectoral quantity indices, and all the arguments that enter the TFP proxy function. In the second step, we exploit our assumptions on the dynamics of TFP to identify demand and production parameters and the parameters that determine the dynamics of productivity. The estimation equation has the following form: let $\log \hat{\lambda}_{t\nu\gamma} = \log \lambda_{t\nu\gamma} - e_{\gamma t}$ denote the fitted values of firm sales shares from the first step of the estimation, then

$$\begin{aligned}\psi_{\gamma t} &= \log \hat{\lambda}_{t\nu\gamma} - \left(1-\frac{1}{\sigma_t}\right)\left(\omega^L_{\nu t}\log l_{\gamma t} + \omega^K_{\nu t}\log k_{\gamma t} - \log Y_{\nu t}\right), \\ \psi_{\gamma t} &= \left(1-\frac{1}{\sigma_t}\right)\Bigg(\phi^{\text{Own}}_{\nu t}\log z^A_{\gamma(t-1)} + \phi^{\text{Own, 0}}_{\nu(t-1)}\mathbb{1}_{z^A_{\gamma(t-1)}=0} + \sum_{\tau=t-5}^{t-1}\phi^{\text{Ext},\tau}_{\nu t}\bar{z}^A_{\nu\tau} + \\ &\qquad + \omega^{\text{SGA}}_{\nu t} z^{\text{SGA}}_{\gamma t} + \sum_{\tau=t-5}^{t-1}\phi^{\text{Ext},\tau}_{\nu t}\bar{z}^A_t + F\left(\psi_{\gamma(t-1)}\right) + e^A_{\gamma t}\Bigg).\end{aligned} \tag{61}$$

Here $e^A_{\gamma t}$ is the idiosyncratic component of TFP that producers do not observe until period t, and $F(\cdot)$ is typically approximated by a 3-rd degree polynomial. $\bar{z}^A_{\nu\tau}$ denotes the average R&D expenditure within sector ν in period τ, and $\bar{z}^A_t$ – average expenditure across all sectors in the economy (weighted with patent citation rates). The term $\phi^{\text{Own}}_{\nu t}\log z^A_{\gamma t}$ summarizes the effect of producer's own R&D expenditures on their TFP, and the term $\phi^{\text{Own, 0}}_{\nu(t-1)}\mathbb{1}_{z^A_{\gamma(t-1)}=0}$ – the effect of reporting 0 R&D expenditures for the Compustat firms[13]. In turn, the sums $\sum_{\tau=t-5}^{t-1}\phi^{\text{Ext},\tau}_{\nu t}\bar{z}^A_{\nu\tau}$ and $\sum_{\tau=t-5}^{t-1}\phi^{\text{Ext},\tau}_{t}\bar{z}^A_{t}$, represent the effects of non-pecuniary externalities on the dynamics of firm TFP.

We estimate Equation 60 with a 7-year rolling window and within each upper-tier sector.

---

[13] We impute \$1000 R&D expense for firms that report zero R&D, so the coefficient $\phi^{\text{Own, 0}}_{\nu(t-1)}$ identifies the effect of reporting zero R&D relative to reporting \$1000 R&D expense. \$1000 is a minimum value of R&D investment reported in Compustat across all sectors and years.

[Markups, Surplus rates and Profits] We obtain the marginal markup estimates using De Loecker et al. [2020] formula:

$$\tilde{\mu}_{\gamma t} = \hat{\omega}_t^L \frac{p_{\gamma t} y_{\gamma t}}{\text{COGS}_{\gamma t}}. \tag{62}$$

$\hat{\omega}_{\nu\gamma t}^L$ is the estimate of the elasticity of firms' output with respect to variable inputs. $\hat{\omega}_t^L$ is constant under Cobb-Douglas production function specification.

Apart from the marginal markups, we also estimate the distributions of average markups and profit rates for Compustat firms. Average markups and profits are informative about evolution of producer market power in the US economy, and producer's incentives to innovate and accumulate capital. The ratio of sales to costs of goods sold is an appropriate measure of average markups in our environment. Profit rates are computed as a difference between establishments' volume of business and their total expenditure, normalized by sales.

6.3. Calibration: Firm Types, Innovation and Miscellaneous Parameters

[Firm Types and Sectoral Structure] We assume that firm TFP and capital stock can take five different values in each upper-tier sector. Also, each variety sector in our economy contains up to 3 product lines. Together these assumptions imply that there are around 3000 firm (and product) types in each upper-tier sector. Values of capital and TFP correspond to the quantiles of capital stock reported in the Compustat data and estimated productivity, respectively.

[TFP and Capital Dynamics, Entry and Exit] We assume that the transition probabilities between firm types are of logit form. So, the future value of capital stock for a firm with $k_{\gamma t}$ units of capital in period t is drawn from the following distribution: suppose $\tilde{k}$ and $\tilde{k} + \Delta > \tilde{k}$ represent two subsequent capital types, then

$$\mathbb{P}\left[k_{\gamma(t+1)} = \tilde{k} | k_{\gamma t}\right] = \frac{1}{1 + \exp^{\varsigma_{\tilde{k}} - \varsigma^Z \log Z_{\gamma t}^K - \varsigma^K k_{\gamma t}}} - \frac{1}{1 + \exp^{\varsigma_{\tilde{k}+\Delta} - \varsigma^Z \log Z_{\gamma t}^K - \varsigma^K k_{\gamma t}}}. \tag{63}$$

In the equation above, thresholds $\varsigma_{\tilde{k}}$ and $\varsigma_{\tilde{k}+\Delta}$ vary across capital types $\tilde{k}$, and satisfy $\varsigma_{\tilde{k}+\Delta} > \varsigma_{\tilde{k}}$. $Z_{\gamma t}^K$ is the value of capital expenditures for product γ in period t. The distribution of future TFP values is specified in a similar fashion. The values of future TFP are allowed to depend on the firms' investment in R&D, their investment in intangibles (SGA), and the knowledge spillovers across firms. The dynamics of firms' TFP and fixed assets are independent. We choose the threshold paramaters and elasticities of the transition probabilities with respect to investment by matching the transition probabilities and investment rates that we observe in the data. In this case, we define the investment rates as a ratio of firms' investment of a specific to their variable surplus[14]. Markup values for each firm type are imputed as a cost-weighted averages for all the firms that belong to a corresponding type in the Compustat data. Firms' entry and exit rates are computed in a similar way based on the Business Dynamics Statistics data.

---

[14] In some cases, the investment rates that we observe in the data imply transition probability elasticities that violate Assumption 4. For such firm types, we assume that some share of the investment represents expenditures on variable inputs.

Table 2: Values of Calibrated Parameters

Parameter	Early Sub-Sample	Late Sub-Sample
Inter-Temporal EOS, ϑ	1.5	1.5
Disount Factr, ρ	4%	4%
Inter-Sectoral EOS, $\sigma_{\mathcal{V}}$	3.26	3.26
Output Growth Rate, g_Y	3.19%	1.95%
Population Growth Rate, g_L	1.09%	0.81%
Entry Cost Growth Rate, g_E	-0.59%	0.37%

Table 3: Inter-Sectoral Substitution Elasticities: 1985 and 2015.

Level of Aggregation	EOS Value	Standard Error	5% Lower Bound	5% Upper Bound
63 KLEMS sectors	3.259	0.070	3.122	3.396
4-digit NAICS	5.457	0.173	5.117	5.797
5-digit NAICS	8.227	0.201	7.833	8.622
6-digit NAICS	9.456	0.184	9.095	9.817

[Calibrated Parameters] For our counterfactual exercises, we also need to determine the values of several other parameters that are difficult to infer from the firm-level financial statements data. These parameters include the following

(i) parameters of consumer preferences: the intertemporal substitution elasticity $1/\vartheta$ and discount factor ρ; in addition, since our calibration contains two levels of aggregation, we also need to set the value of intersectoral substitution elasticity;

(ii) aggregate growth rates of population (g_L), entry costs (g_E), and aggregate productivity (g_A);

The values of ϑ and ρ are set to 1.5 and 4% respectively, consistently with the literature[15]. The values of US population growth rates are taken from the BLS data. The growth rates of real output match the values reported by BEA. The growth rate of entry costs is calibrated to match the population growth rate and the growth rate of the total number of firms in the US economy documented in the BDS.

[Upper-Tier Sectoral Preferences] The intersectoral elasticities of substitution are estimated using the following standard equation:

$$\log \lambda_{\nu t} = \log \bar{\lambda}_{\nu} + \left(1 - \frac{1}{\hat{\sigma}_{\mathcal{V}}}\right) (\log Y_{\nu t} - \log Y_t) + e_{\nu t}. \tag{64}$$

$\log \bar{\lambda}_{\nu}$ denotes a sector fixed effect. In practice, we estimate this equation using BEA data on sectoral

---

[15] Many studies, including e.g. Aghion et al. [2019], Akcigit and Ates [2019], and [Cavenaile et al., 2020], rely on log preferences with $\vartheta = 1$; Acemoglu et al. [2018] set ϑ to 2. Regarding the calibration of consumer discount factor, most studies either infer it from the dynamics of the real interest rates (Aghion et al. [2019] sets $\rho = 5.3\%$,, based on a 6.1% interest rate), or set it to a predetermined value close to zero: Acemoglu et al. [2018] set ρ to 2%, Akcigit and Ates [2019] set ρ to 5%, Cavenaile et al. [2020] – to 4%.

Table 4: Unmatched Aggregate Statistics: Calibration and Data

	Early Sub-Sample		Late Sub-Sample	
Variable	Calibration Value	Data Value	Calibration Value	Data Value
Aggregate Markup	1.345	1.342	1.418	1.427
Aggregate Entry Rate	0.098	0.112	0.089	0.087
Aggregate Exit Rate	0.097	0.095	0.080	0.081

sales shares and quantity indices. The resulting value of substitution elasticity is shown in Table 3. For comparison, we also list the estimates of substitution elasticities at other levels of aggregation[16].

6.4. Estimation Results and Model Fit

Our calibration matches the aggregate growth rates of output, population and aggregate producer masses by construction. We also match the surplus, investment, entry and exit rates, average marginal markups and profits at the firm-type level. We can also show that the aggregate values of these firm-type level statistics are consistent with the data. In Table 4 we list the values of aggregate markups, exit and entry rates that are implied by our calibrations. Notably the average markup value raises only by 8% between the early and late samples. The magnitude of the markup trend in Table 4 is not equal to the increase in average markups in Compustat data from 1980 to 2017 because we average markup values across years for each sub-period to calibrate the model.

Notably, the values of the aggregate productivity growth rates are in line with the BLS productivity estimates[17]. The TFP growth in the late sub-sample calibration is equal to 1.07%. This value is quite close to the BLS labor productivity[18] estimate of 1.4% for the period from 2005 to 2018. Similarly, for the early sub-period our model calibration suggests the TFP growth rate value of 1.9%. The corresponding BLS estimate is equal to 1.7%.

Table 5 displays the sales-weighted averages of estimated parameters for the early and late sub-samples. In both periods the estimated production functions are close to the constant returns to scale. The output elasticity with respect to variable inputs is around 90% throughout the data sample. The capital elasticity varies between 5 and 9 percent. While the elasticities of output with respect to variable inputs and capital are stable across years, the SGA elasticity increases significantly between 1980 and 2017. In our calibration, the SGA elasticity is set to 15% in the early sub-period, and the corresponding late sub-period value is 22%. Finally, we observe a weak upward trend in the substitution elasticity between product varieties. Notably, the estimates of σ_Γ that we derive from the production function estimation routine are rather close to the values that we obtain from reduced form estimation.

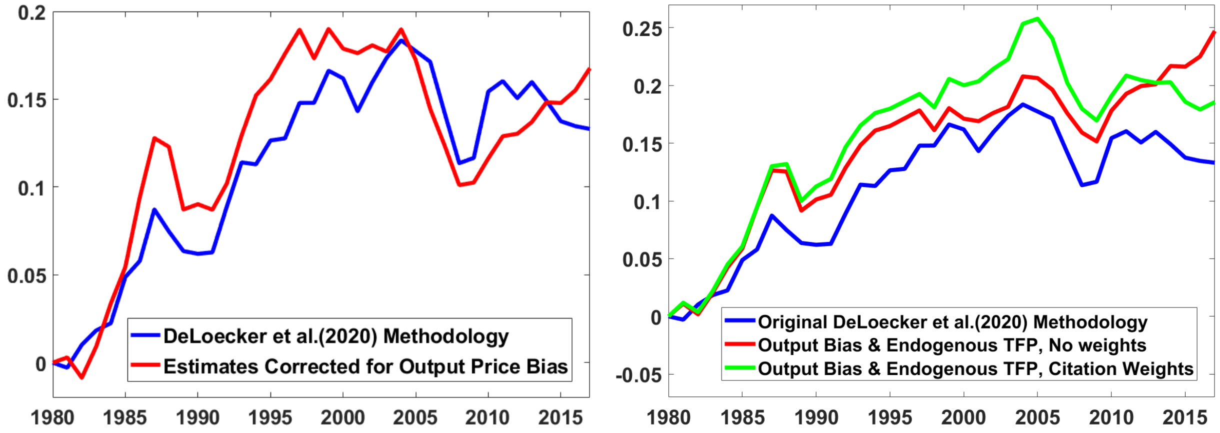
---

[16] We have used the BLS PPI indices and Census SUSB sales shares at the respective levels of aggregation to estimate elasticities at the 4, 5 and 6-digit levels.

[17] E.g., see Sprague(2021).

[18] Labor productivity (as opposed to TFP) represents an appropriate benchmark in this case, because capital intensity and labor composition of workers do not change along the balanced growth path.

Figure 4: Benchmark Markup Estimates



(a) Output Bias Correction

(b) Output Bias and Endogenous Growth

Table 5: Estimated Parameters: Production and Demand

Sample	Labor Elasticity (%)	Capital Elasticity (%)	Returns to Scale (%)	Demand EOS	SGA Elasticity (%)
Early	91.02	6.08	97.11	8.87	14.81
Late	89.51	8.16	97.67	9.03	21.92

Figure 4 displays the evolution of average cost-weighted markups for the US economy. In the left panel, we plot the estimates computed using the standard De Loecker et al. [2020] methodology and the estimates that are corrected for the output price bias. Overall, these time series display similar dynamics, although the modified estimates generate higher markup levels for most years. The right panel of Figure 4 plots the markup estimates that are computed using our benchmark methodology. In this case, we correct the markup estimates for output price endogeneity and allow for the endogenous accumulation of firm TFP. Our results suggest that the estimates of output elasticity with respect to variable inputs obtained using the conventional methodology are on average biased downward by 5-7% because the standard methods omit R&D expenditures from the estimation.

[Discussion: Markup Estimates vs. Market Power] Our theoretical analysis treats markups as exogenously fixed frictions that determine firms' variable surplus rates. Consistently with this assumption, we use structural estimates of marginal markups to calibrate firm surplus rates in our counterfactual exercises. A potential drawback of this approach is that the marginal markup estimates could capture distortions not generated by producer market power. Thus, our model calibration could either over or underestimate the true extent of market power and welfare losses due to sub-optimal markup levels. Our response to this criticism is twofold. First, in our calibration, we only use averages of markup estimates computed at the firm-type level, i.e., conditional on firm TFP and capital stock values. Thus, as long as other distortions average out to zero for each firm type, misallocation unrelated to market power should not affect our results. Second, our calibration strategy is justified by the fact that strategic interactions between firms in the dynamic oligopoly settings can generate markup values that do not depend on the consumers'

Table 6: Counterfactual Results I, Late Sub-Sample

% Changes in Macro Outcomes Relative to Benchmark Calibration

Shock	Welfare	Output	Growth Rate	Mass of Entrants	Entry rate
1% Increase in Markups	-0.507	-0.397	-2.731	5.030	0.018
1% Investment Subsidy	0.173	0.001	1.539	-0.694	-0.051
Zero Markup Variance, $\mu = 1.276$	13.75	25.28	31.89	-79.08	6.311
CES Markups, $\mu = 1.44$	11.24	18.31	31.89	-71.67	6.311
Marginal Cost Pricing, $\mu = 1$	10.67	16.79	31.89	-95.19	6.311

demand structure or production sets of firms. E.g., it follows from folk theorem literature[19] that any dynamic oligopoly game generically has multiple equilibria. Hence, there always exist multiple equilibrium sales and markup distributions in such settings, even if we hold demand and production structure constant. In addition, in Appendix, we describe the generalized oligopoly setting that, depending on the structure of strategic interactions between producers, can generate virtually any markup distribution, including the first-best and second-best markup values.

7. Counterfactual Exercises

We start our analysis by conducting several simple exercises that provide intuition for the first and second best comparisons. For both sub-samples, we document the reaction of the economy to the following markup shocks:

- **[1% Increase in Markups]** Markups are increased by 1%.
- **[Zero Markup Variance]** Variance of markups is reduced to zero, and all markups are reset to their average (unweighted) level.
- **[CES Markups]** All markups are equal to $\frac{\sigma_\mathcal{Y}}{\sigma_\mathcal{Y}-1}$ – this is the level of markups that implements optimal allocation in static models with free entry.
- **[Marginal Cost Pricing]** All prices are reset to the producers' marginal costs (marginal cost pricing benchmark).

Table 6 documents the results of these counterfactual exercises for the late sub-sample. Our model predicts that the decline in markup levels or the decline in variance of markups would benefit the US economy. In fact, an increase in markups leads to a decline in both, the growth rate and one period output. This occurs primarily because higher markup levels over-stimulate entry. The subsequent increase in labor demand leads to a decline in per-firm production employment and investment. These results also suggest that the uniform increase in markups would lead to a slight increase in business dynamism, since due to lower investment incumbent firms on average exit more often.

---

[19] E.g., see Fudenberg and Maskin [1986], and Fudenberg and Maskin [1990].

Table 7: Counterfactual Results I, Early Sub-Sample

% Changes in Macro Outcomes Relative to Benchmark Calibration

Shock	Welfare	Output	Growth Rate	Mass of Entrants	Entry rate
1% Increase in Markups	-0.218	-0.352	-0.233	2.764	-0.016
1% Investment Subsidy	0.143	0.008	0.780	-0.824	- 0.042
Constant Markups, $\mu = 1.258$	20.46	19.43	84.02	-61.93	-0.565
CES Markups, $\mu = 1.44$	18.42	13.78	84.02	-50.63	-0.565
Marginal Cost Pricing, $\mu = 1$	17.87	12.02	84.02	-89.91	-0.565

The increase in investment subsidy has the opposing effect on the economy. The difference between labor costs of incumbents and entrants leads to a decline in entry. An increase in investment of incumbents also lowers the entry rate and raises the aggregate productivity growth. At the same time, the increase in static output is negligible: production employment is not affected by the investment subsidies directly, and since the subsidy is uniform across firms and investment types the composition of sectors does not change significantly. Notably, the results presented in Table 6 suggest that changes in markup levels (or markup variance) have a more profound effect on social welfare relative to changes in investment subsidies of similar magnitude.

Due to the properties of CD-CES calibration and due to the fact that we have assumed the entrant type distribution that does not depend on equilibrium investment, the growth rate value, and the value of entry rate do not depend on the level of markups – if markups are uniform across firms and products. For the late sum-sample calibration, the transition to the balanced growth path with zero markup variance is associated with 32% increase in the aggregate productivity growth rate, and a 6% increase in the entry rate. Reduction in the variance of markups across products leads to an decline in the surplus rate and profits of larger producers, and the decline in future profits induced large firms to invest less. In turn, the entry rate increases because large incumbents exit more often whenever they invest less in intangibles and capital stock. On the aggregate level, these effects are stronger than the impact of markup reduction on the small producers, who experience an increase in their profits and investment.

The counterfactual results for the early sub-sample are overall similar. A 1% percent increase in markups again lowers the productivity growth rate, and static output. These effects are larger in magnitude relative to the increase in output and productivity growth associated with the investment subsidy. One notable distinction between the two sub-samples is that both the rise of market power and the reduction in markup variance lower the entry rate in the early data sample. Markups shocks that we consider depress entry in the early sample due to two reasons. First, the correlation between markups and productivity, or alternatively, markups and firm size, is stronger in the later period. Thus, whenever the markup variance declines, the increase in investment of smaller companies has a stronger effect on the entry rate relative to a decline in investment of large firms. Second, as Table 11 suggests, markup estimates in the early sample are positively correlated with entry rate. Thus, an increase in markups generates a larger increase in profits of entrants.

[First-Best Comparisons] Table 8 contains the estimates of welfare, output and growth rate gains

Table 8: Welfare Analysis: First and Second Best

% Changes in Macro Outcomes Relative to Benchmark Calibration

1. Late Sub-Sample			
BGP	Welfare	Output	Growth Rate
First Best	34.45	38.01	223.3
Second Best I (Markups)	19.71	19.47	124.44
Second Best II (Investment)	20.21	-5.485	258.9

2. Early Sub-Sample			
BGP	Welfare	Output	Growth Rate
First Best	31.06	39.98	127.5
Second Best I (Markups)	20.88	20.17	86.13
Second Best II (Investment)	18.24	-6.747	150.5

Table 9: First-Best: Policy decompositions

Welfare changes in % of Decentralized BGP Welfare.

Sample	Total Welfare	Market Power	Investment Subsidies	R&D	Capital	SGA
Late Sub-Sample	36.09	22.07	14.02	-0.686	3.212	11.49
Early Sub-Sample	36.70	6.146	30.55	5.602	0.996	23.95

generated by the first and second-best balanced growth paths. The first-best equilibria feature higher output levels and faster productivity growth for the early and late sub-periods, and the total effect of misallocation on welfare is significant in both samples. Moreover, the first-best estimates indicate that the distance to first best has increased by 3% between the early and late sub-periods. Notably, while static output gains remain stable across the time periods, the increase in productivity growth rate is much larger for the late sample. This suggests that the increase in misallocation implied by our calibrations is due to an increase in under-investment.

Table 9 contains the results of the first-order policy decompositions for the early and late sub-samples. The first-order changes in welfare levels are overall similar to the non-linear estimates of the distance to frontier. Still, the counterfactual results suggest that the nature of misallocation differs significantly across two data samples. In the earlier sample, investment subsidies play a much bigger role relative to the inefficient markup distribution. In fact, the latter accounts only for 17 % of the welfare differential. In contrast, in the later sample, the ”elimination” of firm market power is responsible for 61% of the distance to the first best. The takeaways of these exercises are in line with predictions of the literature on the rise of market power. If the social planner can alter producer markups and investment costs, the difference between the markup level that they pick and the decentralized markup values is larger whenever the decentralized markups are higher. To be more specific, in these exercises, the social planner chooses the aggregate markup equal to 1.06 in the late sub-sample. The first-best markup value for the early sub-sample is equal to 1.14.

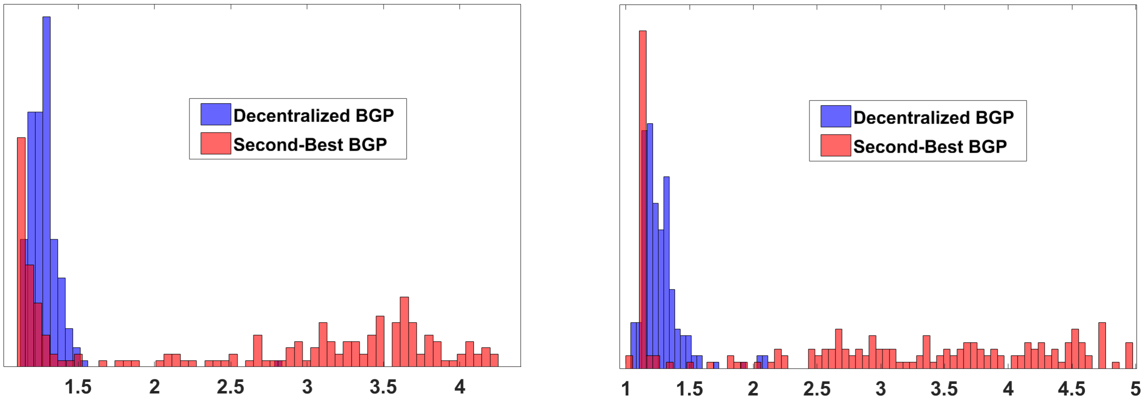
Finally, to conclude our description of the first-best allocations, Table 10 presents the detailed first-order

Table 10: First-Best BGP: Detailed First-Order Decomposition

Welfare changes in % of Decentralized BGP Welfare.

Sub-Sample	Total	Growth Rate	Capital Stock	Reallocation + Labor Costs	Love-of-Variety
Early	36.70	12.80	1.766	13.28	8.850
Late	36.09	26.62	2.070	5.019	2.379

Figure 5: Markup Distributions: Decentralized and Second-Best BGPs



(a) Early Sub-Sample

(b) Late Sub-Sample

decompositions of the distance to the first best. These decompositions rely on Equation 40 that determines the social return on investment, and the corresponding first-order condition for production employment. In the early sub-period, the social planner uses investment subsidies and markups to readjust inefficient entry level, and induce the incumbents to invest more in intangibles. In the late sub-period, the growth rate rate channel is much stronger. This again suggests that investment – and under-investment – has a large impact on welfare in the late sub-sample.

[Second-Best Comparisons: Markups] The estimates in Table 8 indicate that the distance to the second-best optimum, in which the social planner sets markup values, declines insignificantly between the sub-periods. This result is seemingly at odds with the first-best policy decompositions that we have discussed above. To understand why our primary measure of the welfare loss due to markups declines over time, we need to examine the distribution of the "socially-optimal" markups at the micro-level. Figure 5 plots the histograms of the decentralized markup values and the corresponding second-best policies. Table 11 records several descriptive statistics for the markup distributions in the second-best optimum and decentralized equilibrium. In both data samples, the second best markups are more dispersed, and have higher average values. Still, the distributions of markups across producer types differ significantly between samples. In the early sample, the markup values picked by the social planner are negatively correlated with products' capital stock and TFP. At the same time, the correlation between second-best markups and entry rates is positive. These properties of the socially-optimal markup distribution suggest that in the early sub-period the social planner uses markups to readjust the entry rate, and increase the equilibrium mass of firms in the economy. The first-order decomposition of the distance to second-best presented in Table 12 implies the same. In response to the second best policy shocks, the labor is reallocated from incumbents to entrants, and the increase in the mass of firms in the economy has a positive effect on welfare.

Table 11: Second-Best Markups: Descriptive Statistics

Statistic	Early Sample, DE	Early Sample, SB	Late Sample, DE	Late Sample, SB
1. Moments of Markup Distributions				
Mean Markup Value	1.258	2.554	1.276	4.0123
Standard deviation	0.019	1.334	0.167	2.362
2. Correlations of Markups with State Variables, %				
TFP	39.12	-35.97	57.16	7.858
Capital	10.99	-8.692	44.11	-8.604
3. Correlations of Markups with Exit/Entry rates, %				
Entry Rates	9.419	7.496	1.627	13.53
Exit Rates	-8.401	-5.424	-0.151	10.57
4. Correlations of Markups with Investment, %				
Total Investment	13.69	-27.49	51.79	7.047
TFP Investment Share	25.14	-8.830	19.59	1.545

Table 12: Second-Best BGP: Detailed First-Order Welfare Decomposition

Welfare changes in % of Decentralized BGP Welfare.

Sub-Sample	Total	Growth Rate	Capital Stock	Reallocation + Labor Costs	Love-of-Variety
Early	2.168	3.033	0.260	0.890	-2.016
Late	36.72	66.19	-7.012	-33.03	10.57

The picture is different during the late sub-period. Second-best markups are positively correlated with firm productivity, total investment, and the share of investment devoted to R&D and SGA. Consistently with these properties of the second-best markups, the first-order decomposition indicates that the increase in productivity growth acts as a leading source of welfare gains in the late sub-sample. Changes in output and allocation of labor between entrants and incumbents that are induced by the second-best policies are harmful to social welfare. Finally, the comparison of second-best markup distribution reveals why the welfare losses due to market power fall (marginally) between time periods. In the late sub-sample, the social planner uses markups to encourage investment in productivity, and thus they allocate higher surplus rates to larger companies that invest intensely in intangibles. This implies that overall the decentralized distribution of markups is closer to the second-best in the later sample. This informal observation is supported by Table 13 that records the correlation between the decentralized and second-best markup values for the two sub-samples.

Table 13: Correlation of Decentralized and Second-Best Markup Levels, %

Early Sample	Late Sample
-38.34	-11.04

8. Conclusions

In this project, we examine the effects of market power on social welfare in settings with endogenous technological progress and free entry. We propose several ways to assess welfare losses due to the sub-optimal distribution of markups. Our primary measure is the distance between the decentralized allocation and the second-best equilibrium in which the social planner is only allowed to set product prices or markups. We show that in this equilibrium, the social planner uses markups to balance out the impact of investment and production labor misallocation on welfare. We also decompose the distance to the socially-optimal balanced growth path allocation into terms that separately evaluate misallocation due to market power and inefficient investment. This exercise allows us to indirectly evaluate the effects of markups on welfare when the investment is at the social optimum. Finally, we compare the second-best and first-best equilibria to the allocation the social planner chooses if they can only subsidize or tax investment. We argue that the distance between this alternative second-best balanced growth path and the first best can also be interpreted as a measure of the social costs of market power.

In our applications, we compute the welfare losses due to sub-optimal markup distribution for the US economy from 1980 to 2017. We rely on Compustat firm-level data and BDS to calibrate our model. Our results indicate that producer market power generates significant social welfare losses throughout the sample period. The welfare level at the second-best equilibrium with socially-optimal markup distribution is on average 20% higher relative to the decentralized BGP. Somewhat surprisingly, if we use the distance to the second-best as a measure of welfare losses due to markups, market power costs do not increase over time despite the rise in market power. The detailed analysis reveals that the upward trend in markups is not costly for society because the markup distribution is closer to the social optimum in later periods. In the late sub-sample, the social planner assigns higher surplus rates to companies that invest in R&D and intangible capital in order to encourage economic growth. This channel is much weaker in the earlier periods. In contrast, the first-best decompositions suggest that, once the social planner is allowed to manage investment and markups, the costs of market power increase from 7% to 22% from 1980 to 2017.

References

D. Acemoglu, U. Akcigit, H. Alp, N. Bloom, and W. Kerr. Innovation, reallocation, and growth. *American Economic Review*, 108(11):3450–91, 2018.

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

P. Aghion and P. Howitt. A model of growth through creative destruction. *Econometrica*, 60:323–351, 1992.

P. Aghion, N. Bloom, R. Blundell, R. Griffith, and P. Howitt. Competition and innovation: An inverted-u relationship. *The quarterly journal of economics*, 120(2):701–728, 2005.

P. Aghion, A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li. A theory of falling growth and rising rents. Technical report, National Bureau of Economic Research, 2019.

U. Akcigit and S. T. Ates. What happened to us business dynamism? Technical report, National Bureau of Economic Research, 2019.

U. Akcigit, D. Hanley, and N. Serrano-Velarde. Back to basics: Basic research spillovers, innovation policy, and growth. *The Review of Economic Studies*, 88(1):1–43, 2021.

E. Atalay. How important are sectoral shocks? *American Economic Journal: Macroeconomics*, 9(4):254–80, 2017.

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

A. Atkeson and A. Burstein. Aggregate implications of innovation policy. *Journal of Political Economy*, 127 (6):2625–2683, 2019.

A. Atkeson and A. T. Burstein. Innovation, firm dynamics, and international trade. *Journal of political economy*, 118(3):433–484, 2010.

M. N. Baily, C. Hulten, D. Campbell, T. Bresnahan, and R. E. Caves. Productivity dynamics in manufacturing plants. *Brookings papers on economic activity. Microeconomics*, 1992:187–267, 1992.

D. Baqaee and E. Farhi. Entry vs. rents. Technical report, National Bureau of Economic Research, 2020a.

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

J.-P. Benassy. Taste for variety and optimum production patterns in monopolistic competition. *Economics Letters*, 52(1):41–47, 1996.

A. B. Bernard, S. J. Redding, and P. K. Schott. Multiple-product firms and product switching. *American Economic Review*, 100(1):70–97, 2010.

J. Bessen. Information technology and industry concentration. 2017.

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

N. Bloom, M. Schankerman, and J. Van Reenen. Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393, 2013.

S. Bond, A. Hashemi, G. Kaplan, and P. Zoch. Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*, 2021.

A. L. Bowley. *The Mathematical Groundwork of Economics: An Introductory Treatise, by AL Bowley.* Oxford: Clarendon Press, 1924.

C. Broda and D. E. Weinstein. Globalization and the gains from variety. *The Quarterly journal of economics*, 121(2):541–585, 2006.

T. Buettner. *The dynamics of firm profitability, growth, and exit.* London School of Economics and Political Science (United Kingdom), 2004.

L. Cavenaile, M. A. Celik, and X. Tian. Are markups too high? competition, strategic innovation, and industry dynamics. 2020.

C. Corrado, J. Haltiwanger, and D. Sichel. *Measuring Capital in the New Economy.* University of Chicago Press, 2005.

N. Crouzet and J. C. Eberly. Understanding weak capital investment: The role of market concentration and intangibles. Technical report, National Bureau of Economic Research, 2019.

S. J. Davis, R. J. Faberman, and J. Haltiwanger. Labor market flows in the cross section and over time. *Journal of Monetary Economics*, 59(1):1–18, 2012.

J. De Loecker. Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, 79(5):1407–1451, 2011.

J. De Loecker, P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik. Prices, markups, and trade reform. *Econometrica*, 84(2):445–510, 2016.

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

J. De Loecker, J. Eeckhout, and S. Mongey. Quantifying market power and business dynamism in the macroeconomy. Technical report, National Bureau of Economic Research, 2021.

M. De Ridder, B. Grassi, G. Morzenti, et al. The hitchhiker's guide to markup estimation. Technical report, 2021.

M. de Ridder et al. Market power and innovation in the intangible economy. Technical report, 2019.

B. Deaner. Proxy controls and panel data. *arXiv preprint arXiv:1810.00283*, 2018.

R. Decker, J. Haltiwanger, R. Jarmin, and J. Miranda. The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3):3–24, 2014.

X. Ding. Industry linkages from joint production. *Work. Pap., Georgetown Univ., Washington, DC*, 2020.

A. K. Dixit and J. E. Stiglitz. Monopolistic competition and optimum product diversity. *The American economic review*, 67(3):297–308, 1977.

U. Doraszelski and J. Jaumandreu. R & d and productivity: Estimating endogenous productivity. *Review of Economic Studies*, 80(4):1338–1383, 2013.

C. Edmond, V. Midrigan, and D. Y. Xu. How costly are markups? Technical report, 2021.

N. Engbom et al. Firm and worker dynamics in an aging labor market. Technical report.

M. Ewens, R. H. Peters, and S. Wang. *Acquisition prices and the measurement of intangible capital.* National Bureau of Economic Research, 2019.

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

D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, 1986.

D. Fudenberg and E. Maskin. Nash and perfect equilibria of discounted repeated games. *Journal of Economic Theory*, 51(1):194–206, 1990.

A. Gandhi, S. Navarro, and D. A. Rivers. On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016, 2020.

N. Giocoli. The escape from conjectural variations: the consistency condition in duopoly theory from bowley to fellner. *Cambridge Journal of Economics*, 29(4):601–618, 2005.

B. Grassi et al. Io in io: Competition and volatility in input-output networks. *Unpublished Manuscript, Bocconi University*, 2017.

Z. Griliches and H. Regev. Firm productivity in israeli industry 1979–1988. *Journal of econometrics*, 65(1): 175–203, 1995.

G. M. Grossman and E. Helpman. *Innovation and growth in the global economy.* MIT press, 1991a.

G. M. Grossman and E. Helpman. Quality ladders in the theory of growth. *The review of economic studies*, 58(1):43–61, 1991b.

G. M. Grossman, E. Helpman, E. Oberfield, and T. Sampson. Balanced growth despite uzawa. *American Economic Review*, 107(4):1293–1312, 2017.

G. Grullon, Y. Larkin, and R. Michaely. Are us industries becoming more concentrated? *Review of Finance*, 23(4):697–743, 2019.

G. Gutiérrez and T. Philippon. Investment-less growth: An empirical investigation. Technical report, National Bureau of Economic Research, 2016.

G. Gutiérrez and T. Philippon. Declining competition and investment in the us. Technical report, National Bureau of Economic Research, 2017.

R. E. Hall. New evidence on the markup of prices over marginal costs and the role of mega-firms in the us economy. Technical report, National Bureau of Economic Research, 2018.

A. C. Harberger. Monopoly and resource allocation. *The American Economic Review*, 44(2):77–87, 1954.

I. Hathaway and R. E. Litan. What's driving the decline in the firm formation rate? a partial explanation. *The Brookings Institution*, 2014.

B. Herrendorf, R. Rogerson, and A. Valentinyi. Two perspectives on preferences and structural transformation. *American Economic Review*, 103(7):2752–89, 2013.

H. Hopenhayn, J. Neira, and R. Singhania. From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share. Technical report, National Bureau of Economic Research, 2018.

H. A. Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, pages 1127–1150, 1992.

C.-T. Hsieh and P. J. Klenow. Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448, 2009.

A. B. Jaffe. Technological opportunity and spillovers of r&d: Evidence from firms' patents, profits, and market value. *The American Economic Review*, 76(5):984–1001, 1986.

K. L. Judd. On the performance of patents. *Econometrica: Journal of the Econometric Society*, pages 567–585, 1985.

R. F. Kahn. The problem of duopoly. *The Economic Journal*, 47(185):1–20, 1937.

F. Karahan, B. Pugsley, and A. Şahin. Demographic origins of the startup deficit. Technical report, National Bureau of Economic Research, 2019.

M. S. Kimball. The quantitative analytics of the basic neomonetarist model, 1995.

P. J. Klenow and J. L. Willis. Real rigidities and nominal price changes. *Economica*, 83(331):443–472, 2016.

T. J. Klette and Z. Griliches. The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of applied econometrics*, 11(4):343–361, 1996.

T. J. Klette and S. Kortum. Innovating firms and aggregate innovation. *Journal of political economy*, 112(5): 986–1018, 2004.

D. Koh, R. Santaeulàlia-Llopis, and Y. Zheng. Labor share decline and intellectual property products capital. *Econometrica*, 88(6):2609–2628, 2020.

M. Kugler and E. Verhoogen. Prices, plant size, and product quality. *The Review of Economic Studies*, 79(1): 307–339, 2012.

A. P. Lerner. The concept of monopoly and the measurement of monopoly power. *The review of economic studies*, 1(3):157–175, 1934.

J. Levinsohn and A. Petrin. Estimating production functions using inputs to control for unobservables. *The review of economic studies*, 70(2):317–341, 2003.

E. Liu, A. Mian, and A. Sufi. Low interest rates, market power, and productivity growth. Technical report, National Bureau of Economic Research, 2019.

E. R. McGrattan and E. C. Prescott. Unmeasured investment and the puzzling us boom in the 1990s. *American Economic Journal: Macroeconomics*, 2(4):88–123, 2010.

M. J. Melitz and S. Polanec. Dynamic olley-pakes productivity decomposition with entry and exit. *The Rand journal of economics*, 46(2):362–375, 2015.

L. R. Ngai and C. A. Pissarides. Structural change in a multisector model of growth. *American economic review*, 97(1):429–443, 2007.

E. Oberfield and D. Raval. Micro data and macro technology. *Econometrica*, 89(2):703–732, 2021.

U. D. of Labor Staff. *BLS handbook of methods*, volume 2490. US Government Printing Office, 1997.

G. S. Olley and A. Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica: Journal of the Econometric Society*, pages 1263–1297, 1996.

M. Peters and C. Walsh. Declining dynamism, increasing markups and missing growth: The role of the labor force. Technical report, Society for Economic Dynamics, 2019.

B. W. Pugsley and A. Sahin. Grown-up business cycles. *The Review of Financial Studies*, 32(3):1102–1147, 2019.

D. Restuccia and R. Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of economic dynamics*, 11(4):707–720, 2008. ISSN 1094-2025.

P. M. Romer. Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):S71–S102, 1990.

T. Sampson. Dynamic selection: an idea flows theory of entry, trade, and growth. *The Quarterly Journal of Economics*, 131(1):315–380, 2016.

A. Smith. *The wealth of nations*, volume 11937. na, 1776.

C. Syverson. Market structure and productivity: A concrete example. *Journal of political Economy*, 112(6): 1181–1222, 2004.

C. Syverson. Macroeconomics and market power: Facts, potential explanations and open questions, brookings economic studies. *Brookings Institution, Washington DC*, 2019.

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

H. Uzawa. Neutral inventions and the stability of growth equilibrium. *The Review of Economic Studies*, 28 (2):117–124, 1961.

H. Uzawa. Optimum technical change in an aggregative model of economic growth. *International economic review*, 6(1):18–31, 1965.

J. Weiss. Intangible investment and market concentration. Technical report, Working paper, 2019.

Appendix

A. Toy Model: Proofs

[First Best] Throughout this section, aggregate nominal output is normalized to 1. To get the first-best optimum conditions, we solve:

$$\begin{aligned} \max_{\mathcal{M},l,z} \quad & \mathcal{W} = \frac{(l)^{(1-\vartheta)} (\mathcal{M})^{(1-\vartheta)\frac{\sigma}{\sigma-1}}}{1-\exp^{-\rho+(1-\vartheta)(1+z)^{\omega}}}, \\ \text{s.t.} \quad & (\mathcal{M})^{-1} = l + z + \delta \mathcal{L}_E. \end{aligned} \tag{65}$$

Note that $(1+z)^{\omega}$ is the value of the output and productivity growth.

The FOCs are as follows: Δ denotes the Lagrange multiplier on the labor market clearing constraint

$$\begin{aligned} \text{FOC}_{\mathcal{M}}: \quad & \mathcal{W}(1-\vartheta)\frac{\sigma}{\sigma-1} = \Delta, \\ \text{FOC}_{l}: \quad & \mathcal{W}(1-\vartheta) = \Lambda^{Y}\Delta, \\ \text{FOC}_{z}: \quad & \mathcal{W}(1-\vartheta)\omega\frac{z}{1+z}\Lambda^{F} = \Lambda^{Z}\Delta. \end{aligned} \tag{66}$$

From here, we can derive the optimal allocation by dividing the FOCs by one another.

[Fixed Markup Equilibria] In a **decentralized equilibrium**, producers decide on investment by solving the following Bellman equation:

$$\begin{aligned} V(a_t, \mathcal{A}_t, \mathcal{M}_t) = \max_{z} \{ & S_t(l_t, a_t, \mathcal{A}_t, \mathcal{M}_t) - w_t z_t + \\ & + (1-\delta)\exp(-r) V(a_{t+1}, \mathcal{A}_{t+1}, \mathcal{M}_t) \}. \end{aligned} \tag{67}$$

where $\mathcal{A}_t$ denotes the aggregate productivity in period t, r – the interest rate, and $S_t = p_t y_t - w_t l_t = p_t y_t \left(1 - \frac{1}{\mu}\right)$ is the individual firm's surplus at time t.

To get to the decentralized allocation under markup μ, we need to solve the Bellman. Now, we rewrite the firm value function in a following way

$$\begin{aligned} V_0(a, \mathcal{A}, \mathcal{M}) = \max_{(z_t)_{t=0}^{\infty}} & \left\{ \sum_{t=0}^{\infty} \left(\frac{1}{\mathcal{M}} \frac{(a_t l_t)^{1-\frac{1}{\sigma}}}{(\mathcal{A}_t \bar{l}_t)^{1-\frac{1}{\sigma}}} \left(1 - \frac{1}{\mu}\right) - w_t z_t \right) (1-\delta)^t \exp^{-rt} \right\}, \\ \text{s.t.} \quad & l = (w\mu)^{-1} (\mathcal{M})^{-\sigma} \frac{(a)^{\sigma-1}}{(\mathcal{A}_t \bar{l}_t)^{\sigma-1}}, \\ & \bar{l} = (\mathcal{M} w \mu)^{-1}. \end{aligned} \tag{68}$$

The constraints simplify to $l = (Mw\mu)^{-1} (a/\mathcal{A})^{\sigma-1}$.

Note that firm sales in each period are equal to $1/\mathcal{M}$, and thus, the surplus is equal to $1/\mathcal{M}\left(1 - \frac{1}{\mu}\right)$.

We also know that, due to the BGP properties, investment in equal across time periods, and wages are equal to $\left(\mu L^{Y}\right)^{-1}$. The problem above simplifies to

$$V_0\left(a,\mathcal{A},\mathcal{M}\right)=\left(\mathcal{M}\right)^{-1}\max_{\left(z_t\right)_{t=0}^{\infty}}\left\{\sum_{t=0}^{\infty}\left(\left(\frac{a_t}{\mathcal{A}_t}\right)^{\sigma-1}\left(1-\frac{1}{\mu}\right)-\frac{z_t}{\mu\bar{l}_t}\right)\left(1-\delta\right)^t\exp^{-rt}\right\}. \tag{69}$$

Thus, the FOC is

$$\text{FOC}_z:\quad \frac{z}{\mu l}=\omega\left(\sigma-1\right)\frac{z}{1+z}\Lambda^{F\pi}\left(1-\frac{1}{\mu}\right). \tag{70}$$

$$\text{FOC}_z:\quad \frac{z}{l}=\omega\left(\sigma-1\right)\left(\mu-1\right)\Lambda^{F\pi}\frac{z}{1+z}. \tag{71}$$

This gives us the expression for the firms' value in this setting:

$$V\left(\mathcal{A},\mathcal{A},\mathcal{M}\right)=\left(\Lambda^{F\pi}+1\right)\left(1-\frac{1}{\mu}\right)\left(\mathcal{M}\right)^{-1}\left(1-\Lambda^{F\pi}\omega\left(\sigma-1\right)\frac{z}{1+z}\right). \tag{72}$$

From here, we derive the condition for the entry:

$$w\mathcal{L}_E=V\left(\mathcal{A},\mathcal{A},\mathcal{M}\right) \tag{73}$$

$$\frac{\mathcal{L}_E}{l}=\mu\mathcal{M}V\left(\mathcal{A},\mathcal{A},\mathcal{M}\right) \tag{74}$$

From here, we derive the ratio of investment to entry labor:

$$\frac{\mathcal{L}_E}{z}=\frac{1}{1-\delta}\frac{\Lambda^{F\pi}}{\omega(\sigma-1)\Lambda^{F\pi}}\left(1-\Lambda^{F\pi}\omega\left(\sigma-1\right)\frac{z}{1+z}\right)\left(1+\frac{1}{z}\right) \tag{75}$$

Given that the equilibrium entry rate is always equal to δ,

$$\frac{\Lambda^{E,\mu}}{\Lambda^{Z,\mu}}=\frac{\delta\mathcal{L}_E}{z}=\frac{\delta}{1-\delta}\left(\frac{1}{\omega\left(\sigma-1\right)}\left(1+\frac{1}{z}\right)-\Lambda^{F\pi}\right). \tag{76}$$

Note also that the (nominal) interest rate is equal to $r=\rho+\left(\vartheta-1\right)\left(1+z\right)^{\omega}$ – this is independent of the markups, production labor, or producer mass.

[Second-Best] To derive the optimal second-best markup value, note that the following conditions hold:

$$\frac{\mathrm{d}\log l}{\mathrm{d}\log\mu}=-\frac{\mu}{\mu-1}. \tag{77}$$

$$\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\mu}=\Lambda^{Y,\mu}\frac{\mu}{\mu-1}. \tag{78}$$

Then, the planner's optimality condition is reduced to

$$1=\Lambda^{F,\pi}\frac{\delta}{1-\delta}\frac{\Lambda^{Y}}{\Lambda^{E}}\left(\frac{1}{\sigma-1}-\omega\frac{z}{1+z}\Lambda^{F}\right). \tag{79}$$

B. General Model Setting: Miscellaneous Comments

[Discussion: Stochasticity of TFP and Capital] Equation 14 implicitly suggests that *both TFP and capital* evolve stochastically at the product and firm level: producers can influence the shape of distributions of their state variables in the next period, but they cannot fully get rid of variation in either capital or TFP. TFP is stochastic because of the differences in success rates of inventors, scope of innovations or the magnitude of productivity improvements associated with the introduction of new technologies. Uncertainty in future fixed asset stocks could be caused, e.g., by variability of depreciation rates: conditional on the value of capital stock and the age of equipment, some pieces of machinery can still wear off at different rates, and break down at random. The assumption that the process for firm TFP growth is random is standard for the growth literature, however the same is not true to for capital accumulation, which is assumed to be deterministic in most growth frameworks that feature capital[20]. It is thus useful to note that although our setting does not allow capital to grow deterministically, it allows us to consider the limit cases in which the variance of future stock of fixed assets tends to zero, conditional on investment, and the distribution of future capital stock approaches the Dirac delta function.

[Discussion: Creative Destruction via Passive Selection] In our setting there are no fixed costs, and thus there is no active selection: exit of products and firms occurs exogenously whenever producers are hit by negative ”exit” shocks; in the absence of such a shock, firms will always decide to produce non-zero quantities in all of their product lines. Thus, exit of products in our setting should be interpreted as products becoming completely obsolete – due to exogenous demand or technology factors. The absence of active selection does not mean that the forces of creative destruction are not present in our setting: instead of inducing product exit, competition affect firm profits via the intensive margin, by either increasing or decreasing producers' market shares.

Here we also would like to highlight the fact that the exit probability distributions implicitly specified in Equation 14 allow us to replicate allocations that are generated under Bertrand competition with perfect substitutes – this industry structure is traditional for the frameworks with creative destruction based on Aghion and Howitt [1992], [Grossman and Helpman, 1991b] and/or Klette and Kortum [2004]. Indeed, for such an economy the probability of exit for an incumbent product line is equal to the probability that a new good enters the sub-sector in the next period, and that this new good is assigned a higher TFP level. In turn, this probability is determined by the entry rates, and the frequency of product innovations done by incumbents. Similarly, the exit rates in our setting could be conditional on the values of relative productivity and capital, as in the models with active selection margin. In general, our setting can match any pattern of exit rates across markets and producer types. Given an appropriately chosen specification for exit probability distribution, many of our results would continue to hold in the settings with endogenous selection.

[Discussion: Uzawa and Capital Growth] Uzawa [1961] showed that, in the frameworks with aggregate production and investment, the balanced growth under non-CD production is only possible if technological progress is labor-augmenting. This result rests on the fact that aggregate capital stock has to

---

[20] E.g., see Weiss [2019].

grow at the same rate as the aggregate output and consumption for the economy's budget constraint to hold. In contrast to the classical setting of Uzawa [1961], we distinguish between the nominal value of the investment, equal to wz_θ^K for a firm θ, and the production value of the investment, proportional to z_θ^K. The nominal value of the aggregate investment grows at the same rate as the aggregate output, but the prices of capital goods determine the real value of capital stock and its growth. In other words, in our framework, the interest rate is the price of borrowed funds, not capital goods. In its spirit, our solution to the Uzawa problem is similar to the method suggested by Grossman et al. [2017], who assume that prices of capital goods trend downwards due to the "investment-specific" technological progress.

C. A More General Model

In a more general version of a model, we allow consumers to have Kimball [1995] preferences across sectors and products. The production function has a general structure with returns to scale parameter (sum of labor and capital elasticities) equal to ξ.

C.1. Setting

[Consumer Preferences] Preferences of consumers across final good varieties are given by a Kimball [1995] aggregator $\Upsilon_\mathcal{V}$:

$$1 = \int_{\nu \in \mathcal{V}_t} \Upsilon_\mathcal{V}\left(\frac{Y_{\nu t}}{Y_t}\left(|\mathcal{V}_t|\right)^{\eta_\mathcal{V}}\right) f_{\nu t} \mathrm{d}\nu. \tag{80}$$

Here $Y_{\nu t}$ is the consumption index for goods that belong to variety ν, and $|\mathcal{V}_t|$ is the total mass of varieties available to consumers at time t; $f_{\nu t}$ denotes the density of variety type ν among all sectors at time t.

Similarly, the variety-level output is a Kimball [1995] aggregate of outputs produced by individual firms:

$$1 = \sum_{\gamma \in \Gamma_\nu} \Upsilon_\Gamma\left(\frac{y_{t\gamma}}{Y_{\nu t}}\left(|\Gamma_\nu|\right)^{\eta_\Gamma}\right) \frac{f_{t\gamma}}{f_{\nu t}}. \tag{81}$$

$f_{t\gamma}$ denotes the density of product type γ among all products in Γ_t. Functions $\Upsilon_\mathcal{V}$ and Υ_Γ are smooth, increasing and concave. We also assume that $\Upsilon_\jmath(0) = 0$ and $\lim_{y \to 0^+} \Upsilon_\jmath(y) = \infty$ for $\jmath \in \{\mathcal{V}, \Gamma\}$. Parameter $\eta_\mathcal{V}$ regulates the strength of the love-of-variety effects on aggregate output. By analogy with CES preferences, we would typically expect $\eta_\mathcal{V}$ to be above 1.

The specification of Kimball preferences in Equation 80 preserves the model's tractability and ensures the existence of the balanced growth path (henceforth abbreviated as "BGP") under non-zero growth in the number of varieties $|\mathcal{V}_t|$. Importantly, we can evaluate the importance of productivity growth within firms and the variety mass expansion for social welfare using this setup. In most endogenous growth settings, the aggregate output growth is driven either by increases in firm productivities or the expansion of the product variety mass[21]. We want to capture both channels within our framework, and the preference specification

---

[21] In most models of Schumpeterian growth, starting with Aghion and Howitt [1992], the growth of output is based entirely on the

Table 14: Steady State Growth Rates, General Model

Variable	Growth Rate Value
1. Real Variables	
Mass of Producers, $\mathcal{M}_t$	$g_L - g_E$
Mass of Final Good Varieties, $\mathcal{V}_t$	$g_L - g_E$
Mass of Products, Γ_t	$g_L - g_E$
Product-Level Employment and Capital, $l_{t\gamma}$ and $\tilde{k}_{t\gamma}$	g_E
Product-Level Output, $y_{t\gamma}$	$g_A + \xi g_E$
Aggregate Output, Y_t	$g_A + \eta_{\mathcal{V}} g_L + (\xi - \eta_{\mathcal{V}})\, g_E$
2. Prices	
Wages, w_t	$-g_E$
Firm-Level Prices, p_t	$-g_A - \xi g^E$
CPI, P_t	$-g_A + (1 - \eta_{\mathcal{V}})\, g^L + (\eta_{\mathcal{V}} - \xi - 1)\, g_E$

Notes: In the expressions listed in this table, ξ corresponds to the returns-to-scale parameter of firms' production function, and $\eta_{\mathcal{V}}$ – to the love-of-variety parameter in consumer preferences.

in Equations 80 and 81 allows us to do this.

[Production] For each product $\gamma \in \Gamma_t$, the production function maps productivity, labor and capital stock to physical output:

$$y_{t\gamma} = a_\gamma \mathcal{A}_t g\left(k_\gamma \mathcal{K}_t, l_{t\gamma}\right) = \mathcal{A}_t \left(\mathcal{K}_t\right)^\xi a_\gamma \left(k_\gamma\right)^\xi \tilde{g}\left(\frac{l_{t\gamma}}{k_\gamma \mathcal{K}_t}\right), \tag{82}$$

where a_γ is the relative good-specific productivity, and $k_\gamma \mathcal{K}_t$ and $l_{t\gamma}$ represent the amounts of capital and labor used in production at time t. Function $g(\cdot)$ is homogeneous of degree ξ. We also assume that the re-scaled production function $\tilde{g}(\cdot)$ is smooth and satisfies the Inada conditions.

C.2. Analogues of Main Results

The growth rates for the general version of the model are shown in Table 14.

The first-best allocation can be characterized as follows:

Proposition .1. [First Best] *The allocations of production labor and investment that implement first best*

---

within-product productivity growth, while in Judd [1985] and Romer [1990] the output growth is driven by an increase in the number of varieties.

solve the following system of equations:

$$\begin{aligned} FOC_{l_\gamma}: \quad & \lambda_\gamma \omega_\gamma^L = \Lambda^Y \lambda_\gamma^l \eta_{\mathcal{V}}, \\ FOC_{z_\theta}: \quad & \lambda_\theta^Z d\log z_\theta + \Lambda^Z \mathbb{E}_{\lambda_\Theta^Z}\left[d\log \Psi^{\mathcal{M}}\right] + \Lambda^Y \mathbb{E}_{\lambda_\Gamma^l}\left[d\log \Psi^{\Gamma}\right] = \\ & \Lambda^F \frac{1}{\eta_{\mathcal{V}}} dg_A + \frac{\mathbb{E}_{\lambda_\Gamma}\left[\omega_\gamma^K\right]}{\eta_{\mathcal{V}}} d\log \mathcal{K} + \\ & + \mathbb{E}_{\Psi^{\mathcal{V}}}\left[d\log \Psi^{\mathcal{V}}\right] + \frac{\delta}{\eta_{\mathcal{V}}}\left(\mathbb{E}_{\lambda_\nu^\Upsilon}\left[d\log \Psi^{\mathcal{V}}\right] - \mathbb{E}_{\Psi^{\mathcal{V}}}\left[d\log \Psi^{\mathcal{V}}\right]\right), \end{aligned} \tag{83}$$

where λ_θ^Z is the share of investment type z_θ in all investment labor, and, similarly, λ_γ^l is the share of labor employed in production of good γ in all production labor. λ_ν^Υ[22] *is the weight of sector ν in the Kimball aggregator.*

This Proposition is overall similar to Proposition 4.3 in the main text. The main difference is that under Kimball demand the sectoral reallocation effects are proportional to the demand index. Also, labor and capital elasticities are product-specific, and thus the effect of aggregate capital stock is proportional to the sales-share-weighted output elasticity with respect to capital.

The FOCs that determine the second-best allocation have the following form:

Proposition .2. [Second Best, General Model] *The second-best markup levels solve the following system of equations:*

$$\begin{aligned} \mathrm{FOC}_{\mu_\Gamma}: \quad & \Lambda^Y \mathbb{E}_{\lambda_\Gamma^l}\left[\left(\frac{\lambda_\gamma}{\lambda_\gamma^l}\omega_\gamma^L - \Lambda^Y \eta_{\mathcal{V}}\right) \mathrm{d}\log l_\gamma\right] + \\ & + \Lambda^Z \eta_{\mathcal{V}} \mathbb{E}_{\bar{\lambda}_\Theta^Z}\left[\left(1 - \left(\lambda_\Theta^Z\right)^{-1} \Psi_\theta^Z\right) \mathrm{d}\log z_\theta\right] = 0. \end{aligned} \tag{85}$$

Ψ_θ^Z denotes the social return to investment, as in the less general version of the model.

Again, this Proposition is basically identical to the corresponding Proposition in the main text.

The main difference between this general version of the model, and the CD-CES version is in propagation of shocks. The propagation equations for the general version of the model are described by the following proposition:

Proposition .3. [Propagation Equations: Second Best, General Model] *At the second best optimum, the allocation function differentials $d\log l_\gamma$ and $d\log z_\theta\,(\iota)$ solve the system of differential equations described below, for all possible values of $d\log \zeta_\theta$. Let $\Sigma_{\Psi_\Gamma^E \boldsymbol{S}_\Gamma}$ denote the operator that computes the deviation of the*

---

[22] Formally, we have

$$\lambda_\nu^\Upsilon \propto \Upsilon_{\mathcal{V}}\left(\frac{Y_\nu}{Y}\left(|\mathcal{V}|\right)^{\eta_{\mathcal{V}}}\right) f_\nu. \tag{84}$$

function value from its $\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}$-weighted average, then

$$\begin{aligned}\left(Id_{\Gamma}-\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\Phi^{Y}\boldsymbol{\omega}_{\gamma}^{L}\right)d\log l_{\gamma}=&\,\Phi_{LZ}d\log z_{\Theta}+\\ &-\left(\boldsymbol{\zeta}_{\gamma}-\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\right)\left(\boldsymbol{\zeta}_{\gamma}-Id\right)^{-1}d\log\zeta_{\gamma},\\ \left(Id_{\Theta^{\mathcal{I}}}-\Phi_{ZZ}\right)d\log z_{\Theta}=&\,\mathbb{E}_{\Psi_{\Gamma}^{ZOwn}\boldsymbol{S}_{\Gamma}}\left[\frac{\zeta_{\gamma}}{\zeta_{\gamma}-1}d\log\zeta_{\gamma}\right]+\mathbb{E}_{\Psi_{\Gamma}^{ZOwn}\boldsymbol{S}_{\Gamma}}\left[d\log l_{\gamma}\right].\end{aligned}\tag{86}$$

In these equations, Φ^{Y} is the Hessian operator for the aggregate output with respect to the product outputs y_{Γ}, Ψ_{Γ}^{E} and Ψ_{Γ}^{ZOwn} are the maps of operators Ψ^{E} and Ψ^{ZOwn} on the product space, and the operators Φ_{LZ} and Φ_{ZZ} are defined as follows:

$$\begin{aligned}\Phi_{LZ}=&-\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\Phi^{Y}\omega_{\gamma}^{L}\frac{\mathrm{d}\log\mathcal{K}}{\mathrm{d}\log z_{\Theta}}-\mathbb{1}_{\Gamma}\mathbb{E}_{\Psi^{E}\boldsymbol{S}_{\Theta}}\left[\frac{\mathrm{d}\log\Psi^{E}}{\mathrm{d}\log z_{\Theta}}\right]\\ &+\Sigma_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\boldsymbol{\varepsilon}_{\nu}\left(\mathbb{E}_{\lambda_{\mathcal{V}}^{\Upsilon}}\left[\frac{\mathrm{d}\log\Psi^{\mathcal{V}}}{\mathrm{d}\log z_{\Theta}}\right]-\mathbb{E}_{\Psi^{\mathcal{V}}}\left[\frac{\mathrm{d}\log\Psi^{\mathcal{V}}}{\mathrm{d}\log z_{\Theta}}\right]\right),\\ \Phi_{ZZ}=&\,\mathbb{E}_{\Psi^{\mathrm{ZOwn}}\boldsymbol{S}_{\Theta}}\left[\frac{\mathrm{d}\log\Psi^{\mathrm{ZOwn}}}{\mathrm{d}\log z_{\Theta}}\right].\end{aligned}\tag{87}$$

The operator Φ_{LZ} in Equation 47 describes the feedback loop between changes in investment and changes in production employment under the markup shocks. Changes in investment affect production labor allocation via three channels. First, investment rates affect the aggregate capital stock. Since we have assumed that capital and labor act as complements in production, higher (capital) investment should lead to the reallocation of employment from labor-intensive towards capital-intensive products. Note that the product operator $\Phi^{Y}\omega_{\gamma}^{L}$ measures the elasticity of product sales shares with respect to aggregate capital, and thus the effect of $\mathrm{d}\log\mathcal{K}$ on employment is proportional to the deviation the sales share elasticity with respect to capital from its weighted average $\mathbb{E}_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\left[\Phi^{Y}\omega_{\gamma}^{L}\right]$. The term $\mathbb{E}_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\left[\Phi^{Y}\omega_{\gamma}^{L}\right]$ represents the positive effect of differential $\mathrm{d}\log\mathcal{K}$ on wages: higher stocks of fixed assets incentivize producers to employ more workers in the short run, since labor and capital act as complements. The term the sign of this term is ambiguous since higher levels of incumbents' investment could both increase and decrease the expected value of a firm for entrants, conditional on surplus levels. The last term in the expression for Φ_{LZ} corresponds to sectoral composition effects that affect the product-level employment via the changes aggregate output Y. This term is proportional to the deviation of the pseudo-demand elasticity of sector ν's output from the weighted average of pseudo-demand elasticities $\mathbb{E}_{\Psi_{\Gamma}^{E}\boldsymbol{S}_{\Gamma}}\left[\varepsilon_{\nu}\right]$. The weighted average term again represents the sectoral composition effect for wages. The sign of the term $\Phi_{LZ}\mathrm{d}\log z_{\Theta}$ on $\mathrm{d}\log l_{\Gamma}$ is ambiguous since all the effects of investment on production employment can be classified as "composition effects." Changes in z_{Θ} rarely if never lead to uniform changes in l_{Γ}, rather investment affects the relative employment levels across the sector and product types.

D. Generalized Oligopolistic Competition

As an alternative to marginal cost pricing setting, we consider a setup in which firms set prices and outputs strategically, by engaging in a game of generalized oligopolistic competition. In this setting, the amount of physical output that a firm manages to sell is a function of firm's own *action*, and the *actions* of its competitors:

Assumption 5. ***[Generalized Oligopolistic Competition]*** *In each period, firms set prices of their goods and the corresponding output levels by optimizing over actions* $(\chi_{\gamma t})_{\gamma\in\Gamma_\theta}$*. Firms take the actions of their competitors as given. The mapping between product output* $y_{\gamma t}$ *and the actions of firms that own competing products (*$\gamma' \in \Gamma_\nu$*) is as follows:*

$$y_{\gamma t} = \bar{y}_\gamma \chi_{\gamma t} \cdot \left(\prod_{\gamma'\in\Gamma_\nu, \gamma'\neq\gamma} \left(\chi_{\gamma' t}\right)^{\varsigma_{\gamma'}} \right). \tag{88}$$

Parameter $\varsigma_{\gamma'}$ regulates the sensitivity of sold output with respect to the actions of firm's competitor who owns good γ'; $\bar{y}_\gamma$ is a re-normalizing constant. Importantly, the equation above *does not* represent either a technological constraint on firm output or the demand constraint. Rather, it describes producers' perception of the behavior of their competitors[23]. Thus, it plays the same role in the determination of a short-run equilibrium as the Bertrand assumption that states that firms optimize over prices, or Cournot assumption on that firms choose quantities. Under generalized oligopolistic competition, the heterogeneity in "market power" is generated by the differences in modes of competition, i.e., payment functions and action spaces that firms take as given when they interact with their competitors.

The industry structure outlined above is sufficiently flexible to allow us to take comparative statics with respect to firm markups and evaluate welfare losses due to market power. For any strictly positive vector of markup values $\bar{\mu}$ we can find vectors of parameters $\left(\varsigma_{\gamma'}\right)_{\gamma'\in\Gamma_\nu}$, such that $\bar{\mu}$ will be generated in a decentralized equilibrium of our model. One corollary of this statement is that both Cournot and Bertrand markup distributions can be replicated in our setting provided that the values of elasticities ς are chosen appropriately[24]. The same holds for the socially optimal markup distribution, conditional on that all markup values are above one.

[Optimization Under Generalized Oligopoly] Under the generalized oligopolistic competition, the producer optimization is more involved. To note, in our setting the set of firms that own more than one product within the same variety market is zero mass, and thus it is w.l.o.g. to consider optimization for

---

[23] The industry structure that we consider here is to some extent similar to the conjectural variation model of Bowley [1924], as it generates the same patterns of firm behavior and the same range of equilibrium markup and sales share distributions. However, we would like to emphasize that the solution concept that we use is Nash equilibrium instead of conjectural equilibrium, and thus our setting is immune to some criticisms that concern the nature of the conjectural variations.

[24] For example, the case in which $\varsigma_{\gamma'} = 0$ and $\bar{y}_{\gamma'} = 1, \forall\gamma' \in \Gamma_\nu$ moves us back to the setting with Cournot competition.

each product separately. Then, a producer that owns product γ in market ν solves the following:

$$\begin{aligned} \max_{\chi_\gamma} \quad & S_{\gamma t} = p_{\gamma t} y_{\gamma t} - w_t l_{\gamma t}, \\ \text{s.t.} \quad & y_{\hat{\gamma} t} = \bar{y}_\gamma \chi_{\hat{\gamma} t} \cdot \left(\prod_{\gamma' \in \Gamma_\nu, \gamma' \neq \gamma} \left(\chi_{\gamma' t}\right)^{\varsigma_{\gamma'}} \right), \quad \forall \hat{\gamma} \in \Gamma_\nu, \\ & y_{\gamma t} = \tilde{a}_{\gamma t} g\left(l_{\gamma t}, \tilde{k}_{\gamma t}\right), \\ & |\Gamma_t| p_{\gamma t} = P_t Y_t \delta_t \delta_{\nu t} \Upsilon'_{\mathcal{V}} \left(\frac{Y_{\nu t}}{Y_t} |\mathcal{V}_t|^{\eta_{\mathcal{V}}} \right) \Upsilon'_{\Gamma} \left(\frac{y_{\gamma t}}{Y_{\nu t}} \left(|\Gamma_\nu|\right)^{\eta_\Gamma} \right) \frac{|\mathcal{V}_t|^{\eta_{\mathcal{V}}}}{Y_t} \left(|\Gamma_\nu|\right)^{\eta_\Gamma} . \end{aligned} \tag{89}$$

[Flexibility and Scope] The exogenous variation in producer ”market power” that is necessary for the comparative statics can be generated by the shocks to Cobb-Douglas elasticities ς_γ. Unlike the standard frameworks that are built on oligopoly or monopolistic competition, this variation in markups does not require us to change either consumer preferences or production constraints that the firms face. Thus, we can separate the differences in welfare that are generated by changes in strategic interactions between producers from the effects of other shocks. This setup also enables us to evaluate and interpret the first-order losses from sub-optimal producer market power – such an exercise is feasible in particular because we know that, as long as the ”optimal” distribution of market power prescribes positive values to all product-level markups, the ”optimal” allocation is generated in the equilibrium by some combination of the parameter values.

One drawback of this setting is that the values of higher-order moments of the demand functions generated by such an industry structure are fully determined by the same set of parameters as markups, and thus, e.g., it is not possible for us to match both Bertrand markups and Bertrand pass-through rates. We can further augment equation 88 and add more parameters to it if we want to match both markup and pass-through distributions. In the benchmark setting, we use the Cobb-Douglas aggregation for the producer actions for the sake of simplicity.

[Discussion: Other Industry Structures] In our setting, firms' actions affect producer sales shares and profits via influencing the quantities of products sold. As an alternative, we could have assumed that instead of quantities firm's actions affect prices of goods. E.g., by analogy with equation 88, we could assume that product prices are set to

$$p_{\gamma t} = \bar{p}_\gamma \chi_{\gamma t} \cdot \left(\prod_{\gamma' \in \Gamma_\nu} \left(\chi_{\gamma'}\right)^{\varsigma^p_{\gamma' t}} \right), \tag{90}$$

where $\chi_{\gamma t}$ is the action of a firm that sells product γ at time t, as before. Similarly to the game in which firm actions influence quantities, the setting that is build on price-based competition can generate any positive markup distribution. In particular, there will sets of parameters that implement Cournot and Bertrand allocation, as well as the socially optimal distribution of markups. Under the benchmark values of parameters $\varsigma_{\gamma'} = 0$ and $\bar{y}_{\gamma'} = 1$, $\forall \gamma' \in \Gamma_\nu$, the game based on equation 90 defaults to Bertrand oligopoly. Overall, the game specification with the price-based competition will have the same properties as the game based on equation 88 – and we chose quantity-based competition primarily because it is marginally more convenient under the Kimball setup.

There are a couple of other industry structure specifications that the reader might find appealing. First, by analogy with Cournot and Bertrand competition models, we could assume that, instead of keeping *either* quantities *or* prices constant, producers aim to keep constant the value of sine function $\varsigma\left(\cdot,\cdot\right)$ that takes both their output and their price as arguments. The action of the producer θ is such a setting would be defined as

$$\chi_{\gamma t} = \hat{\varsigma}\left(p_{\gamma t}, y_{\gamma t}\right). \tag{91}$$

Although this specification is simple and appealing, it generically does not generate the socially optimal distribution of markups and product-level outputs. Thus, it is not well suited for the welfare analysis exercises that we implement in later sections of this paper.

Another option would be to consider an analogue of Equation 88 that actually incorporates both Cournot and Bertrand as special cases. An example of such industry structure is as follows:

$$\aleph \in [0,1], \qquad \left(y_{\gamma t}\right)^{\aleph}\left(p_{\gamma t}\right)^{1-\aleph} = \bar{y}_{\gamma}\chi_{\gamma t}\cdot\left(\prod_{\gamma'\in\Gamma_{\nu},\gamma'\neq\gamma}\left(\chi_{\gamma' t}\right)^{\varsigma_{\gamma'}}\right). \tag{92}$$

Unlike Equation 88, this industry structure can match both Cournot and Bertrand markups and pass-through rates for specific values of $\aleph$. Since it includes both price-based and quantity-based industry structures (in Equations 88 and 90) as special cases, it is also flexible enough to generate any positive markup values. Once again, we stick to Cournot-like specification for the sake of simplicity.

[(Marginal) Markup Determination] In the oligopolistic competition setting that we consider, markup values are determined by the firms' actions. The markup values are characterized as follows. Let ε_{γ} denote the elasticity of function Υ'_{Γ}, evaluated at the equilibrium value of relative output y_{γ}/Y_{ν}. We refer to ε_{γ} as the quasi-elasticity of demand, because it is precisely equal to the elasticity of product's price with respect to its quantity under monopolistic competition. By analogy, ε_{ν} denotes the quasi-elasticity of demand on sectoral level. Then, for a *non-monopolistic* variety market ν, we have:

$$\begin{aligned}\left(\mu_{\gamma}\right)^{-1} = \quad & \underbrace{\left(1+\varepsilon_{\gamma}\right)}_{\text{Own Output Effect}} \quad - \quad \underbrace{\mathrm{Cov}_{\lambda_{\nu\gamma'}}\left(\varsigma_{\gamma},\varepsilon_{\gamma'}\right)}_{\text{Reallocation within Variety Sub-Sector}} \quad + \\ & + \underbrace{\mathbb{E}_{\lambda_{\nu\gamma'}}\left[\varsigma_{\gamma}\right]\left(\varepsilon_{\nu}-\varepsilon_{\gamma}\right)}_{\text{Variety-level Output Effect}}.\end{aligned} \tag{93}$$

Here $\lambda_{\nu\gamma'}$ is the market share of a single product of type γ'[25], and $\mathrm{Cov}_{\lambda_{\nu\gamma'}}\left(\varsigma_{\gamma},\varepsilon_{\gamma'}\right)$ is the $\lambda_{\nu\gamma'}$-weighted covariance of industry structure parameters ς_{γ} and quasi-elasticities of demand $\varepsilon_{\gamma'}$, evaluated within sector ν.

---

[25] We will use λ to denote sales or cost shares of individual products and firms, and $\bar{\lambda}$ to denote the type-level shares at product, variety and firm level. The following identities hold for λ and $\bar{\lambda}$: $f_{\gamma}|\Gamma|\lambda_{\gamma} = \bar{\lambda}_{\gamma}$, $f_{\gamma}|\mathcal{V}|\lambda_{\nu} = \bar{\lambda}_{\nu}$, and similarly, within variety sectors we have $f_{\nu\gamma}|\Gamma_{\nu}|\lambda_{\nu\gamma} = \bar{\lambda}_{\nu\gamma}$.

Equation 93 tells us that the sales elasticity with respect to firm's action χ_γ is comprised of three terms. First, χ_γ directly alters the quantity of good γ that is sold to consumers – this effect is summarized by the term $(1+\varepsilon_\gamma)$. As was implied above, the first term $(1+\varepsilon_\gamma)$ is precisely equal to product markup under monopolistic competition, and thus, the remaining two terms summarize the effect of oligopolistic competition on firm market power. The second term represents reallocation of sales and outputs within the variety sub-sector. The sign of the covariance term depends on the output of product γ relative to the variety-level output Y_ν: for a firm whose output is above average, the absolute values of the quasi-elasticities are positively correlated with output elasticities ς_γ, thus, the covariance term overall should have a positive sign. The opposite should be true for goods with lower outputs. Finally, firm's action χ_γ also affects producer's output elasticity by altering the sectoral output, this effect is summarized by a third term. If varieties of products are (at least locally) less substitutable than good types within varieties, it should be the case that $|\varepsilon_\nu| \geq |\varepsilon_\gamma|$ and the sign of the term should be proportional to $-\mathbb{E}_{\lambda_{\nu\gamma'}}[\varsigma_\gamma]$. Importantly, we would expect the sum of the oligopoly terms to be negative for most firms in the market: producer's should be able to benefit from their ability to impact the variety-level output and outputs of their competitors, and on average they should charge higher markups relative to the setting with monopolistic competition.

For the monopolistic sectors, note that, if a variety market only contains one firm, all the terms associated with the effects of firm's action χ_γ on outputs of other firms are reset to zero, and thus the inverse markup is equal to $(1+\varepsilon_\nu)$.

[Discussion: Generalized Oligopoly Markups Under CES] To provide more intuition for Equation 93, here we also derive the expressions for markups under CES demand. Suppose σ_Γ is the elasticity of substitution within varieties, and $\sigma_{\mathcal{V}}$ is the elasticity across varieties, then

$$\left(\mu_\gamma^{\text{CES}}\right)^{-1} = \left(1 - \frac{1}{\sigma_\Gamma}\right) + \mathbb{E}_{\lambda_{\nu\gamma'}}\left[\varsigma_\gamma\right]\left(\frac{1}{\sigma_\Gamma} - \frac{1}{\sigma_{\mathcal{V}}}\right). \tag{94}$$

Under CES demand the within-variety reallocation effects summarized by the covariance term in Equation 93 are always zero. Similarly to the discussion above, if $\sigma_{\mathcal{V}} < \sigma_{\mathcal{V}}$, and, as long as $\mathbb{E}_{\lambda_{\nu\gamma'}}[\varsigma_\gamma] > 0$, the oligopoly markups are higher than the markups under monopolistic competition.

Consistently with the discussion in Section 3, if we reset, $\bar{\varsigma}_\gamma$ to zero, the equation above is reduced to the standard expression for CES-Cournot markups, as in Atkeson and Burstein [2008]:

$$\left(\mu_\gamma^{\text{CES, Cournot}}\right)^{-1} = \left(1 - \frac{1}{\sigma_\Gamma}\right) + \lambda_{\nu\gamma}\left(\frac{1}{\sigma_\Gamma} - \frac{1}{\sigma_{\mathcal{V}}}\right). \tag{95}$$

We can also compute the values of $\bar{\varsigma}_\gamma$ that generate Bertrand markups: given the equilibrium Bertrand sales shares $\lambda_{\nu\gamma}^{\text{Bertrand}}$, we have:

$$\begin{aligned} \bar{\varsigma}_\gamma^{\text{Bertrand}} &= 1 - \frac{\sigma_\Gamma}{\sigma_\Gamma\left(1 - \lambda_\gamma^{\text{CES, Bertrand}}\right) + \lambda_\gamma^{\text{CES, Bertrand}}\sigma_{\mathcal{V}}} < 0, \\ \left(\mu_\gamma^{\text{CES, Bertrand}}\right)^{-1} &= \left(1 - \frac{1}{\sigma_\Gamma}\right) + \mathbb{E}_{\lambda_{\nu\gamma'}}\left[\varsigma_\gamma^{\text{Bertrand}}\right]\left(\frac{1}{\sigma_\Gamma} - \frac{1}{\sigma_{\mathcal{V}}}\right). \end{aligned} \tag{96}$$

We can note that the values of cross-elasticities $\varsigma_{\gamma\gamma'}$ are negative for Bertrand competition, conditional on $\sigma_\Gamma - \sigma_\mathcal{V} > 0$: lowering the price of good γ allows the owner of this good to sell more, and in addition the competitors of good γ are forced to reduce their outputs.

E. BGP Calculations

Note: All the proofs and derivations in this section are presented for the case of a general model with Kimball demand and general production structure.

[Firms: Short-term Optimization]In the short run, firms choose employment levels for each of their products by maximizing their variable surpluses, subject to the production, demand and industry structure constraints. Under the marginal cost pricing with fixed markups, the short run optimization is reduced to Equation 13, provided that the product prices and outputs are determined by consumer inverse demand functions:

$$|\Gamma_t| p_{\gamma t} = P_t Y_t \delta_t \delta_{\nu t} \Upsilon'_\mathcal{V} \left(\frac{Y_{\nu t}}{Y_t} |\mathcal{V}_t|^{\eta_\mathcal{V}} \right) \Upsilon'_\Gamma \left(\frac{y_{\gamma t}}{Y_{\nu t}} \left(|\Gamma_\nu|\right)^{\eta_\Gamma} \right) \frac{|\mathcal{V}_t|^{\eta_\mathcal{V}}}{Y_t} \left(|\Gamma_\nu|\right)^{\eta_\Gamma}. \tag{97}$$

δ_t and $\delta_{\nu t}$ are the *demand indices* for the aggregate output (δ_t) and variety-level output ($\delta_{\nu t}$), defined as follows

$$\frac{1}{\delta_t} = \int_{\nu \in \mathcal{V}_t} \Upsilon'_\mathcal{V} \left(\frac{Y_{\nu t}}{Y_t} |\mathcal{V}_t|^{\eta_\mathcal{V}} \right) \frac{Y_{\nu t}}{Y_t} |\mathcal{V}_t|^{\eta_\mathcal{V}} f_{\nu t} \mathrm{d}\nu. \tag{98}$$

$$\frac{1}{\delta_{\nu t}} = \sum_{\gamma \in \Gamma_\nu} \Upsilon'_\Gamma \left(\frac{y_{\gamma t}}{Y_{\nu t}} \left(|\Gamma_\nu|\right)^{\eta_\Gamma} \right) \frac{y_{\gamma t}}{Y_{\nu t}} \left(|\Gamma_\nu|\right)^{\eta_\Gamma} f_{\nu \gamma}. \tag{99}$$

The aggregate demand index δ_t is proportional to the marginal utility of aggregate output Y_t, and similarly, the product $\lambda_{\nu t}\delta_{\nu t}$ tracks the marginal utility of sectoral output $Y_{\nu t}$.

Dynamic Optimization We are solving for one-period growth rates, subject to constraints, and the relative steady state allocation; capital and productivity levels are fixed, so $\mathrm{d}\log a_\theta = 0$, and $\mathrm{d}\log k_\theta = 0$. The probability transition matrix notation: initial types are indexing the rows, and the future types are indexing the columns. Throughout this section, we will omit time (t) subscripts to keep the notation concise.

We will use the following notation for the "discount factors" in the Neumann series: let $\gamma = \exp\left(g_E - g_L\right) = \exp\left(-g_M\right)$, $\beta = \exp(-r)$, and ρ will denote the original discount rate for the consumers:

$$\mathcal{P} = \mathcal{P}\left(Z, \left\{\zeta_\iota\left(z(\iota)\right)\right\}_{\iota \in \mathcal{I}_A}, g_A, g_K\right). \tag{100}$$

$$\Psi\left(r, Z, g_A, g_K\right) = \Psi = \left(I - \beta\mathcal{P}\right)^{-1}, \quad \Psi^M\left(g_M, Z, g_A, g_K\right) = \Psi^M = \left(I - \gamma\mathcal{P}\right)^{-1}. \tag{101}$$

$$\Omega_\theta^{\mathrm{Own}}\left(\kappa\right) = \frac{\partial \mathcal{P}\left(Z, \left\{\zeta_\iota\left(z(\iota)\right)\right\}_{\iota \in \mathcal{I}_A}, g_A, g_K\right)}{\partial \log z_\theta\left(\kappa\right)}, \tag{102}$$

$$\Omega_{\theta}^{\text{All}}(\kappa) = \frac{\partial \mathcal{P}\left(Z, \{\zeta_{\iota}(z(\iota))\}_{\iota \in \mathcal{I}}, g_A, g_K\right)}{\partial \log z_{\theta}(\kappa)} + \frac{\partial \mathcal{P}\left(Z, \{\zeta_{\iota}(z(\iota))\}_{\iota \in \mathcal{I}}, g_A, g_K\right)}{\partial \log \zeta_{\iota}(Z)} \Omega_{\theta}^{\zeta \kappa} + \\ + \frac{\partial \mathcal{P}\left(Z, \{\zeta_{\iota}(z(\iota))\}_{\iota \in \mathcal{I}}, g_A, g_K\right)}{\partial g_A} \frac{\mathrm{d} g_A}{\mathrm{d} \log z_{\theta}(\kappa)}, \tag{103}$$

$$\Phi_{\theta}^{\text{Own}}(\kappa, k) = \frac{\mathrm{d}\Omega_{\theta}^{\text{Own}}(\kappa)}{\mathrm{d} \log z_{\theta}(k)}. \tag{104}$$

$\Omega^{\text{Own}}(\kappa)$ will denote the "stacked" non-zero lines of the corresponding matrices. Operator $\Omega_{\theta}^{\text{Ext}}(\kappa)$ is the difference between $\Omega_{\theta}^{\text{All}}(\kappa)$ and $\Omega_{\theta}^{\text{Own}}(\kappa)$. Note that given these definitions the value function and the equilibrium investment can be expressed as

$$w z_{\theta}(\kappa) = \beta \Omega_{\theta}^{\text{Own}}(\kappa) V, \qquad V = \Psi \pi. \tag{105}$$

Here π is the producer profits, equal to $\pi_{\theta} = S_{\theta} - w z_{\theta}$, where S is the surplus received from production.

Consumer optimization

$$\frac{f_{\gamma}}{f_{\nu}} = \frac{1}{|\Gamma_{\nu}|}. \tag{106}$$

Thus, at the product type level,

$$\lambda_{\gamma} = \delta_{\nu} \delta \left(\Upsilon_{\mathcal{V}}' \left(\frac{Y_{\nu}}{Y_t} (|\mathcal{V}_t|)^{\eta_{\mathcal{V}}} \right) \frac{Y_{\nu}}{Y_t} (|\mathcal{V}_t|)^{\eta_{\mathcal{V}}} \right) \cdot \\ \left(\Upsilon_{\Gamma}' \left(\frac{y_{\gamma}}{Y_{\nu}} \right) \frac{y_{\gamma}}{Y_{\nu}} f_{\gamma} \right). \tag{107}$$

$$\lambda_{\gamma} = \frac{p_{\gamma} y_{\gamma} f_{\gamma}}{PY}. \tag{108}$$

Also for a firm of type θ, we have

$$\gamma \in \Gamma_{\theta \nu}, \quad \lambda_{\theta \nu} = \lambda_{\nu \gamma}. \tag{109}$$

$$\gamma \in \Gamma_{\theta \nu}, \quad \lambda_{\theta \nu} = \lambda_{\nu \gamma} = \delta_{\nu} \delta \left(\Upsilon_{\mathcal{V}}' \left(\frac{Y_{\nu}}{Y} (|\mathcal{V}|)^{\eta_{\mathcal{V}}} \right) \frac{Y_{\nu}}{Y} (|\mathcal{V}|)^{\eta_{\mathcal{V}} - 1} \right) \cdot \\ \left(\Upsilon_{\Gamma}' \left(\frac{y_{\gamma}}{Y_{\nu}} \right) \frac{y_{\gamma}}{Y_{\nu}} (|\Gamma_{\nu}|)^{-1} \right). \tag{110}$$

$$\lambda_{\theta} = \sum_{\nu \in \mathcal{V}_{\theta}} \lambda_{\theta \nu}. \tag{111}$$

Then, the sales at the variety-product level are equal to $p_{\nu\gamma} y_{\nu\gamma} = \lambda_{\gamma\nu} \left(f_{\nu\gamma} |\Gamma_{\nu}| |\mathcal{V}| \right)^{-1}$.

Inter-Temporal Consumer Optimization At the steady state, we have the standard optimization condition

$$\text{FOC}_{a_t}: \quad (1 - \vartheta) P_{t+1}^{Y} \left(\frac{Y_t}{L_t} \right)^{-\vartheta} = \exp(r - \rho) (1 - \vartheta) P_t^{Y} \left(\frac{Y_{t+1}}{L_{t+1}} \right)^{-\vartheta}. \tag{112}$$

$$\text{FOC}_{a_t}: \quad P^Y_{t+1}\left(\frac{Y_t}{L_t}\right)^{-\vartheta} = \exp\left(r-\rho\right) P^Y_t \left(\frac{Y_{t+1}}{L_{t+1}}\right)^{-\vartheta}. \tag{113}$$

Note that the growth rate of nominal income is $g_M = g_L - g_E$, and thus the growth rate of CPI is equal to $g_L - g_E - g_Y$.

$$r - \left(g_L - g_E - g_Y\right) = \rho + \vartheta\left(g_Y - g_L\right). \tag{114}$$

Aggregate profit rate is defined as

$$\pi^{\text{Agg}} = \frac{\pi}{Lw + \pi}. \tag{115}$$

F. Proofs: Aggregate Growth Rates along BGP

From the type space identities it follows that the variety mass $\mathcal{V}$ grows at the same rate as the mass of producers $\mathcal{M}$.

For the growth rate of the nominal output, it holds:

$$\textbf{Agg. GDP growth:} \quad g_Y + g_P = \eta_{\mathcal{V}} g_M + \left(g_A + \xi g_K\right) + g_P = g_p + g_y + g_M = g_M, \tag{116}$$

From price normalization, we have:

$$g_p = -g_y. \tag{117}$$

$$g_M\left(\eta_{\mathcal{V}} - 1\right) = -g_A - \xi g_K - g_P \tag{118}$$

The dynamics of CPI depends on the growth in the number of varieties and firm-level output (or firm-level prices).

Static factor prices: for the constant endowment shares,

$$\textbf{Static factor prices:} \quad g_w = g_M - g_L = -g_E. \tag{119}$$

Also, from producer optimization, we have:

$$\textbf{Firm-Level Prices:} \quad g_p = g_{\text{mc}} = g_w + g_L - g_M - g_y - g_{\omega_L} = -g_y \tag{120}$$

Let g_A denote the productivity growth at firm level: Then, we have:

$$g_y = g_A + \xi g_K. \tag{121}$$

For profits, we have:

$$\textbf{Profits and Surplus:} \quad S_{\nu\gamma t} = \left(1 - \frac{\omega^L_{\nu\gamma t}}{\mu_{\nu\gamma t}}\right) p_{\nu\gamma t} y_{\nu\gamma t}, \quad \Leftrightarrow g_S = 0. \tag{122}$$

$$\pi_{\nu\gamma t} = S_{\nu\gamma t} - \mathcal{L}_I w, \quad \Leftrightarrow g_\pi = 0. \tag{123}$$

From this it follows that the growth rate of overhead costs should be equal to the growth rate of entry costs.

Also, the definition of the value function for firms and free entry condition imply that the following identity holds

$$g_V = 0. \tag{124}$$

Note that the fact that labor elasticities are constant means that the ratio of capital to labor at the firm level should remain constant, and thus,

$$g_K = g_L - g_M = g_E. \tag{125}$$

Note that this is capital per firm (or per product).

Thus, the real output growth rate can be expressed only in terms of the TFP growth

$$g_Y = \eta_{\mathcal{V}} g_M + (g_A + \xi g_K), \tag{126}$$

$$g_Y = g_A + \eta_{\mathcal{V}} g_L + (\xi - \eta_{\mathcal{V}}) g_E. \tag{127}$$

G. Proofs: Allocative Efficiency, Welfare Elasticities and Welfare Decompositions

[Proposition 4.3: Proof] Let Δ_1 denote the Lagrange multiplier on the aggregate demand constraint, $\Delta_{2\nu}$ – the multiplier on the sectoral demand constraint for sector ν, and Δ_3 – the multiplier on the labor market clearing constraint. Here we consider the balanced growth paths, and the labor supply in the initial period is normalized to 1.

[FOCs: Output Levels] :

$$\text{FOC}_Y: \quad (Y)^{1-\vartheta} = \Delta_1 \frac{1}{\delta}, \tag{128}$$

$$\text{FOC}_{Y_\nu}: \quad \Delta_1 \frac{\delta_\nu}{\delta} \lambda_\nu = \Delta_{2\nu}, \tag{129}$$

$$\text{FOC}_{Y_\nu}: \quad (Y)^{1-\vartheta} \delta_\nu \lambda_\nu = \Delta_{2\nu}, \tag{130}$$

[Labor] The FOCs yield the following conditions: relative levels of l^Y,

$$\text{FOC}_{l \text{ or } L^Y}: \quad (Y)^{1-\vartheta} \lambda_\gamma \omega^L_\gamma = L^Y \lambda^l_\gamma \Delta_3. \tag{131}$$

$$\frac{(Y)^{(1-\vartheta)}}{L^Y}\mathbb{E}_{\lambda}\left[\omega_{\gamma}^{L}\right] = \Delta_3. \tag{132}$$

$$\Delta_3 = \Delta_1 \frac{1}{\delta}\frac{\mathbb{E}_{\lambda}\left[\omega_{\gamma}^{L}\right]}{L^Y}. \tag{133}$$

Relative allocation:

$$\lambda_{\gamma}\omega_{\gamma}^{L} = \lambda_{\gamma}^{l}\mathbb{E}_{\lambda}\left[\omega_{\gamma}^{L}\right]. \tag{134}$$

[Investment] FOC for $z_{\theta}(\iota)$

$$\begin{aligned}
\mathrm{FOC}_{z_{\theta}(\iota)}: \quad & \Delta_3\Lambda^{Z}\lambda_{\theta}^{Z}\,\mathrm{d}\log z_{\theta} = \Lambda^{F}(1-\vartheta)\frac{1}{1-\exp^{-\delta+(1-\vartheta)g_Y}}(Y)^{(1-\vartheta)}\,\mathrm{d}g_A + \\
& + \left(1-\exp^{-\delta+(1-\vartheta)g_Y}\right)^{-1}(1-\vartheta)(Y)^{(1-\vartheta)}\,\delta\mathbb{E}_{\lambda_{\nu}^{\Upsilon}}\left[\mathrm{d}\log\Psi^{\mathcal{V}}\right] + \\
& + \left(1-\exp^{-\delta+(1-\vartheta)g_Y}\right)^{-1}(1-\vartheta)(Y)^{(1-\vartheta)}\left(\eta_{\mathcal{V}}-\delta\right)\mathrm{d}\log\mathcal{V} + \\
& + \left(1-\exp^{-\delta+(1-\vartheta)g_Y}\right)^{-1}(1-\vartheta)(Y)^{(1-\vartheta)}\,\mathbb{E}_{\lambda_{\Gamma}}\left[\omega_{\gamma}^{K}\right]\mathrm{d}\log K + \\
& - \left(1-\exp^{-\delta+(1-\vartheta)g_Y}\right)^{-1}(1-\vartheta)(Y)^{(1-\vartheta)}\,\eta_{\mathcal{V}}\Lambda^{Z}\mathbb{E}_{\lambda_{\Theta}^{Z}}\left[\mathrm{d}\log\Psi^{\Theta}\right] + \\
& - \left(1-\exp^{-\delta+(1-\vartheta)g_Y}\right)^{-1}(1-\vartheta)(Y)^{(1-\vartheta)}\,\eta_{\mathcal{V}}\Lambda^{Y}\mathbb{E}_{\lambda_{\Gamma}^{l}}\left[\mathrm{d}\log\Psi^{\Gamma}\right];
\end{aligned} \tag{135}$$

Re-normalizing:

$$\begin{aligned}
\mathrm{FOC}_{z_{\theta}(\iota)}: \quad & \Lambda^{Z}\lambda_{\theta}^{Z}(\iota)\,\mathrm{d}\log z_{\theta}(\iota) = \Lambda^{F}\frac{1}{\eta_{\mathcal{V}}}\mathrm{d}g_A + \\
& + \frac{\delta}{\eta_{\mathcal{V}}}\mathbb{E}_{\lambda_{\nu}^{\Upsilon}}\left[\mathrm{d}\log\Psi^{\mathcal{V}}\right] + \frac{\eta_{\mathcal{V}}-\delta}{\eta_{\mathcal{V}}}\mathbb{E}\left[\mathrm{d}\log\Psi^{\mathcal{V}}\right] + \frac{\mathbb{E}_{\lambda_{\Gamma}}\left[\omega_{\gamma}^{K}\right]}{\eta_{\mathcal{V}}}\mathrm{d}\log K + \\
& - \Lambda^{Z}\mathbb{E}_{\lambda_{\Theta}^{Z}}\left[\mathrm{d}\log\Psi^{\Theta}\right] - \Lambda^{Y}\mathbb{E}_{\lambda_{\Gamma}^{l}}\left[\mathrm{d}\log\Psi^{\Gamma}\right];
\end{aligned} \tag{136}$$

$$\mathrm{d}\log\mathcal{E} = -\sum_{\iota\in\mathcal{I}}\mathbb{E}_{f_{\Theta}}\left[\Phi^{\mathcal{M}}(\iota)\,\mathrm{d}\log z_{\Theta}(\iota)\right]. \tag{137}$$

$$\mathrm{d}\log f_{\Theta} = \sum_{\iota\in\mathcal{I}}\Phi^{\mathcal{M}}(\iota)\,\mathrm{d}\log z_{\Theta}(\iota) - \mathbb{E}_{f_{\Theta}}\left[\Phi^{\mathcal{M}}(\iota)\,\mathrm{d}\log z_{\Theta}(\iota)\right]. \tag{138}$$

$$\mathrm{d}\log\mathcal{K} = -\sum_{\iota\in\mathcal{I}}\mathbb{E}_{k_{\Gamma}\circ\mathcal{P}_{E}^{\Gamma}}\left[\Omega_{E}^{\Gamma}(\iota)\,\mathrm{d}\log z_{\Theta}(\iota)\right]. \tag{139}$$

$$\mathrm{d}g_A = -\sum_{\iota\in\mathcal{I}}\mathbb{E}_{a_{\Gamma}\circ\mathcal{P}_{E}^{\Gamma}}\left[\Omega_{E}^{\Gamma}(\iota)\,\mathrm{d}\log z_{\Theta}(\iota)\right]. \tag{140}$$

[Mass of Producers]The FOC with respect to the mass of producers $\mathcal{M}$:

$$\mathrm{FOC}_{\mathcal{M}}: \quad \Delta_2 \frac{\eta_{\mathcal{V}}}{\delta} = -\Delta_5. \tag{141}$$

$$\eta_{\mathcal{V}} = \frac{\mathbb{E}_{\lambda}\left[\omega^{L}_{\nu\gamma}\right]}{L^{Y}}. \tag{142}$$

[Capital Intensity] We can check that the solution for social planner's problem generates a finite K. In the limit $K \to \infty$, we have: (since all the other terms contain elasticities and converse to zero)

$$\mathrm{FOC}_{z_{\theta}(\kappa), K\to\infty}: \quad 0 = \eta_{\mathcal{V}} L^{Z} \lambda^{Z}_{\theta}\left(\kappa\right). \tag{143}$$

This implies that the investment labor should be equal to zero (assuming that the limit of δ is above zero). Zero investment under $K \to \infty$ would lead to a decline in capital stock per firm, which is inconsistent with $K \to \infty$.

On the other hand, we know that whenever $K \to 0$, the aggregate output also will approach zero. The social planner can do better by allocating the non-zero share of labor to capital investment, and preserving non-zero capital stock. Thus, $K \to 0$ does not generate a valid equilibrium. Formally, we know that the term $\mathbb{E}_{\lambda_{\nu\gamma}}\left[\omega^{K}_{\nu\gamma}\Omega^{Z}_{K}\right]$ goes to infinity as K approaches zero ($\omega^{K}_{\nu\gamma}$ is always between 0 and ξ for all varieties, but Ω^{Z}_{K} is proportional to Z/K, which grows large as $K \to 0$). All other terms are bounded.

$$\mathrm{FOC}_{z_{\theta}(\kappa), K\to 0}: \quad 0 = \mathbb{E}_{\lambda_{\nu\gamma}}\left[\omega^{K}_{\nu\gamma}\Omega^{Z}_{K}\right]. \tag{144}$$

[Proposition 4.] Log-linearizing the second best constraints, we get: here note that the operators Ψ^{E} and $\Psi^{Z\mathrm{Own}}\left(\iota\right)$ only depend on investment and exogenously fixed parameters

$$\begin{aligned} & z_{\Theta}\left(\iota\right) = \Psi^{Z\mathrm{Own}}\left(\iota\right)\frac{S_{\Theta}}{w}, \\ & \frac{|\Gamma|}{|\Gamma|}\frac{1}{f_{\gamma}}\bar{\lambda}_{\gamma}\omega_{L} = \mu_{\gamma} l_{\gamma} w, \\ & w = \frac{1}{\mathcal{L}_{E}}\Psi^{E} S_{\Theta}. \end{aligned} \tag{145}$$

Note that these equations imply that given the price normalization that we use, the per-firm allocation functions l_{Γ} and z_{Θ} are independent of of the mass of producers (only the per-firm surplus levels matter). Let us start with the investment equation. We have:

$$\mathrm{d}\log z_{\Theta}\left(\iota\right) = \sum_{\tilde{\iota}\in\mathcal{I}} \mathbb{E}_{\Psi^{Z\mathrm{Own}}(\iota)\boldsymbol{S}_{\Theta}}\left[\frac{\mathrm{d}\log\Psi^{Z\mathrm{Own}}\left(\iota\right)}{\mathrm{d}\log z_{\Theta}\left(\tilde{\iota}\right)}\right]\mathrm{d}\log z_{\Theta}\left(\tilde{\iota}\right) + \mathbb{E}_{\Psi^{Z\mathrm{Own}}(\iota)\boldsymbol{S}_{\Theta}}\left[\mathrm{d}\log S_{\theta}/w\right]. \tag{146}$$

From here, we can already get: stacking the FOCs for investment,

$$\mathrm{Id}\mathbb{1}_{\iota=\tilde{\iota}} - \Phi_{ZZ}\left(\iota, \tilde{\iota}\right) = \mathrm{Id}\mathbb{1}_{\iota=\tilde{\iota}} - \mathbb{E}_{\Psi^{Z\mathrm{Own}}(\iota)\boldsymbol{S}_{\Theta}}\left[\frac{\mathrm{d}\log\Psi^{Z\mathrm{Own}}\left(\iota\right)}{\mathrm{d}\log z_{\Theta}\left(\tilde{\iota}\right)}\right]. \tag{147}$$

For the surplus differentials, we have:

$$\mathrm{d}\log S_{\gamma} - \mathrm{d}\log w = \mathrm{d}\log l_{\gamma} + \frac{\zeta_{\gamma}}{\zeta_{\gamma}-1}\left(\mathrm{d}\log\mu_{\gamma} - \sigma_{\Gamma}\left(\mathrm{d}\log l_{\gamma} - \mathrm{d}\log\mathcal{K}\right)\right). \tag{148}$$

$$\mathrm{d}\log S_{\theta} - \mathrm{d}\log w = \mathbb{E}_{S_{\gamma},\gamma\in\Gamma_{\theta}}\left[\mathrm{d}\log l_{\gamma}\right] + \frac{\zeta_{\theta}}{\zeta_{\theta}-1}\mathbb{E}_{\lambda_{\gamma},\gamma\in\Gamma_{\theta}}\left[\mathrm{d}\log\zeta_{\gamma}\right]. \tag{149}$$

Note that here $\mathrm{d}\log\zeta$ solves:

$$\mathrm{d}\log\zeta_{\gamma} = \mathrm{d}\log\mu_{\gamma} - \sigma_{\Gamma}\mathrm{d}\log l_{\gamma} + \sigma_{\Gamma}\mathrm{d}\log\mathcal{K}. \tag{150}$$

We also set

$$\mathrm{d}\log\zeta_{\theta} = \mathbb{E}_{\lambda_{\gamma},\gamma\in\Gamma_{\theta}}\left[\mathrm{d}\log\zeta_{\gamma}\right]. \tag{151}$$

σ_{θ} is defined as a sales-share weighted average of product-level elasticities for firm θ:

$$\sigma_{\theta} = \mathbb{E}_{\lambda_{\gamma},\gamma\in\Gamma_{\theta}}\left[\sigma_{\gamma}\right], \tag{152}$$

and ζ_{θ} is the firm-level price-cost margin defined as a ratio of sales to variable costs.
Simplifying, we get

$$\left(\mathrm{Id} - \Phi_{ZZ}\right)\mathrm{d}\log z_{\Theta} = \mathbb{E}_{\Psi^{Z\mathrm{Own}}\Gamma\boldsymbol{S}_{\Gamma}}\left[\mathrm{d}\log S_{\gamma}/w\right]. \tag{153}$$

$$\begin{aligned}&\mathbb{E}_{\Psi_{\Gamma}^{Z\mathrm{Own}}\boldsymbol{S}_{\Gamma}}\left[\mathrm{d}\log S_{\gamma}/w\right] = \\ &= \mathbb{E}_{\Psi_{\Gamma}^{Z\mathrm{Own}}\boldsymbol{S}_{\Gamma}}\left[\frac{\zeta_{\gamma}}{\zeta_{\gamma}-1}\mathrm{d}\log\zeta_{\gamma}\right] + \mathbb{E}_{\Psi_{\Gamma}^{Z\mathrm{Own}}\boldsymbol{S}_{\Gamma}}\left[\mathrm{d}\log l_{\gamma}\right].\end{aligned} \tag{154}$$

Combining the terms:

$$\begin{aligned}\left(\mathrm{Id} - \Phi_{ZZ}\right)\mathrm{d}\log z_{\Theta} = {}&\mathbb{E}_{\Psi_{\Gamma}^{Z\mathrm{Own}}\boldsymbol{S}_{\Gamma}}\left[\frac{\zeta_{\gamma}}{\zeta_{\gamma}-1}\mathrm{d}\log\zeta_{\gamma}\right] + \\ &+ \mathbb{E}_{\Psi_{\Gamma}^{Z\mathrm{Own}}\boldsymbol{S}_{\Gamma}}\left[\mathrm{d}\log l_{\gamma}\right].\end{aligned} \tag{155}$$

Now, let us move to log-linearizing the production labor allocation. First, note that the labor costs can be expressed as

$$l_{\gamma}w = \frac{\bar{\lambda}_{\gamma}}{\zeta_{\gamma}f_{\gamma}}. \tag{156}$$

Then, the entry condition takes the form:

$$w = \frac{1}{\mathcal{L}_{E}}\Psi^{E}\left(\mathrm{Id} - \left(\boldsymbol{\zeta}_{\Gamma}\right)^{-1}\right)\frac{\bar{\lambda}_{\Gamma}}{f_{\Gamma}}. \tag{157}$$

Thus, since product type shares are equal to the elasticities of output with respect to product outputs (at a type level), we have

$$\mathrm{d}\log\zeta_\Gamma + \mathrm{d}\log l_\Gamma + \mathbb{1}_\Gamma \mathrm{d}\log w = \Phi^Y \boldsymbol{\omega}_\gamma^L \left(\mathrm{d}\log l_\Gamma - \mathbb{1}_\Gamma \mathrm{d}\log\mathcal{K}\right) + \frac{\partial\log\bar{\lambda}_\Gamma/f_\Gamma}{\partial\log z_\Theta}. \tag{158}$$

$$\begin{aligned}\left(\mathrm{Id}_\Gamma - \Phi^Y \boldsymbol{\omega}_\gamma^L\right)\left(\mathrm{d}\log l_\gamma - \mathbb{1}_\Gamma \mathrm{d}\log\mathcal{K}\right) + \mathbb{1}_\Gamma d\log w = -\mathrm{d}\log\mathcal{K}\\ - \mathrm{d}\log\zeta_\gamma - \mathrm{d}\log f_\gamma + \frac{\partial\log\bar{\lambda}_\Gamma}{\partial\log z_\Theta}.\end{aligned} \tag{159}$$

From the log-linearization of the investment equation, it follows that the elasticity of wages satisfies: rescaling the weights

$$\begin{aligned}\mathrm{d}\log w =& \mathbb{E}_{\Psi^E \boldsymbol{S}_\Gamma}\left[\mathrm{d}\log\Psi^E\right] + \mathbb{E}_{\Psi_\Gamma^E \boldsymbol{S}_\Gamma}\left[\frac{1}{\zeta_\gamma - 1}\mathrm{d}\log\zeta_\gamma\right] + \\ &+ \mathbb{E}_{\Psi_\Gamma^E \boldsymbol{S}_\Gamma}\left[\Phi^Y \boldsymbol{\omega}_\gamma^L \left(\mathrm{d}\log l_\Gamma - \mathbb{1}_\Gamma \mathrm{d}\log\mathcal{K}\right) + \frac{\partial\log\bar{\lambda}_\Gamma/f_\Gamma}{\partial\log z_\Theta}\right].\end{aligned} \tag{160}$$

Elasticities of sales shares $\bar{\lambda}_\gamma$ with respect to investment are as follows:

$$\frac{\partial\log\lambda_\nu}{\partial\log Y_\nu} = \varepsilon_\nu\left(1 - \lambda_\nu\right) - \left(\varepsilon_\nu - \mathbb{E}_{\lambda_{\nu'}}\left[\varepsilon_{\nu'}\right]\right). \tag{161}$$

$$\frac{\mathrm{d}\log\bar{\lambda}_\Gamma/f_\gamma}{\mathrm{d}\log z_\Theta\left(\iota\right)} = -\mathbb{E}_{\lambda_\nu}\left[\mathrm{d}\log f_\mathcal{V}\right] + \left(\Phi_\mathcal{V}^Y - \boldsymbol{\varepsilon}_\nu\left(\mathrm{Id} - \boldsymbol{\lambda}_\mathcal{V}\right)\right)\mathbb{E}_{\lambda_\mathcal{V}^\Upsilon}\left[\mathrm{d}\log f_\mathcal{V}\right]. \tag{162}$$

$$\frac{\mathrm{d}\log\bar{\lambda}_\Gamma/f_\gamma}{\mathrm{d}\log z_\Theta\left(\iota\right)} = -\mathbb{E}_{\lambda_\nu}\left[\mathrm{d}\log f_\mathcal{V}\right] + \left(\varepsilon_\nu - \mathbb{E}_{\lambda_{\nu'}}\left[\varepsilon_{\nu'}\right]\right)\mathbb{E}_{\lambda_\mathcal{V}^\Upsilon}\left[\mathrm{d}\log f_\mathcal{V}\right]. \tag{163}$$

The part that does not cancel out:

$$\frac{\mathrm{d}\log\bar{\lambda}_\Gamma/f_\gamma}{\mathrm{d}\log z_\Theta\left(\iota\right)} \propto \varepsilon_\nu \mathbb{E}_{\lambda_\mathcal{V}^\Upsilon}\left[\mathrm{d}\log f_\mathcal{V}\right] \propto \varepsilon_\nu\left(\mathbb{E}_{\lambda_\mathcal{V}^\Upsilon}\left[\mathrm{d}\log\Psi^\mathcal{V}\right] - \mathbb{E}_{\Psi^\mathcal{V}}\left[\mathrm{d}\log\Psi^\mathcal{V}\right]\right). \tag{164}$$

Assembling all the terms together,

$$\begin{aligned}&\left(\mathrm{Id}_\Gamma - \Phi^Y\omega_L\right)\mathrm{d}\log l_\gamma + \mathbb{1}_\Gamma \mathbb{E}_{\Psi_\Gamma^E \boldsymbol{S}_\Gamma}\left[\Phi^Y\omega_L \mathrm{d}\log l_\gamma\right] = \left(\mathbb{1}_\Gamma \mathbb{E}_{\Psi_\Gamma^E \boldsymbol{S}_\Gamma}\left[\Phi^Y\omega_L\right] - \Phi^Y\omega_L\right)\mathrm{d}\log\mathcal{K}\\ &- \mathrm{d}\log\zeta_\gamma - \mathbb{E}_{\Psi_\Gamma^E \boldsymbol{S}_\Gamma}\left[\frac{1}{\zeta_\gamma - 1}\mathrm{d}\log\zeta_\gamma\right] - \mathbb{E}_{\Psi^E \boldsymbol{S}_\Theta}\left[\mathrm{d}\log\Psi^E\right]\\ &+ \left(\boldsymbol{\varepsilon}_\nu - \mathbb{E}_{\Psi_\Gamma^E \boldsymbol{S}_\Gamma}\left[\varepsilon_\nu\right]\right)\left(\mathbb{E}_{\lambda_\mathcal{V}^\Upsilon}\left[\mathrm{d}\log\Psi^\mathcal{V}\right] - \mathbb{E}_{\Psi^\mathcal{V}}\left[\mathrm{d}\log\Psi^\mathcal{V}\right]\right).\end{aligned} \tag{165}$$

Under the CD-CES specification, several terms drop out from the labor equation:

$$
\begin{aligned}
&\left(\mathrm{Id}_\Gamma - \Phi^Y \boldsymbol{\omega}^L_\gamma\right) \mathrm{d}\log l_\gamma + \mathbb{1}_\Gamma \mathbb{E}_{\Psi^E_\Gamma \boldsymbol{S}_\Gamma}\left[\Phi^Y \boldsymbol{\omega}^L_\gamma \mathrm{d}\log l_\gamma\right] = -\mathrm{d}\log\zeta_\gamma \\
&- \mathbb{E}_{\Psi^E_\Gamma \boldsymbol{S}_\Gamma}\left[\frac{1}{\zeta_\gamma - 1}\mathrm{d}\log\zeta_\gamma\right] - \mathbb{E}_{\Psi^E \boldsymbol{S}_\Theta}\left[\mathrm{d}\log\Psi^E\right].
\end{aligned} \tag{166}
$$

Notably, the elasticities of aggregate variables g_A and $\mathcal{K}$ can also be expressed in terms of the elasticities. From equations 25 and 26, it follows that the elasticities of g_A and $\mathcal{K}$ are proportional to the weighted expectations of elasticities of the entry type distribution $\mathcal{P}^\Gamma_E$:

$$
\mathrm{d}\log\mathcal{K} = -\mathbb{E}_{k_\Gamma \boldsymbol{\mathcal{P}}^\Gamma_E}\left[\mathrm{d}\log\mathcal{P}^\Gamma_E\right]. \tag{167}
$$

$$
\begin{aligned}
\mathrm{d}g_A \frac{1-\mathcal{E}_\Gamma \bar{A}_E}{1-\mathcal{E}_\Gamma} &= -\mathbb{E}_{\Psi^\Gamma}\left[\mathrm{d}\log\Psi^\Gamma\right]\frac{\mathcal{E}_\Gamma\left(\bar{A}_E\right)}{\left(1-\mathcal{E}_\Gamma\right)^2} + \\
&+ \exp^{-g_A}\mathbb{E}_{\Psi^\Gamma}\left[a_\Gamma \boldsymbol{\mathcal{P}}^{\Delta\mathcal{A}}\left(\mathrm{d}\log\mathcal{P}^{\Delta\mathcal{A}} + \mathrm{d}\log\Psi^\Gamma\right)\right].
\end{aligned} \tag{168}
$$

To interpret the equation that describes the elasticity of capital stock, note that the dynamics of incumbent product and firm types is determined by the transition kernel $\mathcal{P}$, and thus, the dynamics of aggregate capital intensity is pinned down by the assumption on the dynamics of entrants' fixed assets. Thus, the effect of investment on capital stock and productivity growth is proportional to the effect of investment on the distribution of relative productivity and relative capital among entrants: an economy with higher capital intensity will prescribe lower relative fixed assets values to the entrants. As before, we can interpret the equation for the elasticity of productivity growth rate via statistical TFP decompositions. The first term in the right-hand side of Equation 168 is proportional to the elasticity of the entry rate, and thus, it measures the change in productivity growth rate g_A that is due to changes in entry. The second term in the right-hand side of Equation 168 evaluates the changes in productivity growth that are due to either composition effects in the distribution of incumbents (part of the term proportional to $\propto \mathrm{d}\log\Psi^\Gamma$) or changes in the improvements of TFP within-products (part of the term proportional to $\propto \mathrm{d}\log\mathcal{P}^{\Delta\mathcal{A}}$).

H. Data and Estimation

H.1. Discussion: Markup Estimation

[Discussion: Proxy Functions] The equation above is based on the Olley and Pakes [1996] identification strategy, in a sense that we rely on capital expenditures and R&D in order to pin down the variation in productivity. In contrast to the original Olley and Pakes [1996] methodology, we use more variables to proxy for TFP differences across producers, and thus, the scalar unobservability assumption is much weaker in our setting: we allow for multiple additional sources of variation in firms' investment, and these sources include additional heterogeneity in firm types, as well as differences in the market environment that firms face.

Depending on the assumptions that we are willing to make about the industry structure, the estimation strategy proposed by Levinsohn and Petrin [2003] might or might not be feasible in our setting. If we assume that there is no firm-level variation in markups, the expenditures on variable inputs should depend only on the capital stock, TFP and possibly, other fixed inputs. In such a case, the LP proxy strategy is valid. On the other hand, if markups vary across firms, the markup differences drive firms' decisions to hire workers and purchase materials. This means that the scalar unobservability assumption is violated whenever expenditure on variable inputs is used as a TFP proxy, and whenever firms have different beliefs about their competitors' strategies. Thus, we opt for using the investment-based proxy function: not much is known about the nature of competition in the US economy, and ultimately it is hard to make an argument for or against homogeneity of producers' beliefs about their competitors behavior. Importantly, investment-based proxy function also allows us to estimate the industry structure parameters independently from the demand and production functions.

[Discussion: Input Price Heterogeneity] In our estimation we implicitly assume that there is no heterogeneity in the input prices, including labor wages, prices of materials and capital goods. This restriction is imposed, first, because Compustat data does not allow us to reliably control for the input price variation across firms, except for including controls for the location of firms' plants and/or producers' sectoral identities. Second, despite the assumption on homogeneity of input prices, there are still some types of input price variation that can be accommodated by our empirical setting. So, whenever differences in input prices are due to the variation in quality of purchased goods or efficiency of workers, and whenever quality of inputs has an effect on the amount of output produced, or/and output's quality, the prices of inputs have to be included in the production function estimation, consistently with the arguments of Kugler and Verhoogen [2012] and De Loecker et al. [2016]. Moreover, suppose that, instead of paying for hours worked, or physical units of goods, firms buy ”efficiency units” of labor, materials and capital goods, where an ”efficiency unit” is a combination of the product's quality, quantity, and possibly, its other features. In such a setting, variation in the prices of physical units of goods does not necessarily imply differences in prices of efficiency units, e.g. wages per worker might differ across firms, but only to the extent that workers with different wages are supplying different amounts of ”efficiency units of labor” to their employers. Whenever prices of input efficiency units are constant across firms, input expenditure is an acceptable measure of input usage, regardless of the variation in wages or prices of physical units of materials and/or capital goods. Furthermore, we could argue that, as long as producers use similar technologies and have access to the same set of options in input markets, prices of efficiency units should not vary significantly.

[Discussion: Output Price Bias and Markup Estimation] One of the drawbacks of production function estimation implemented on Compustat data is that, by nature of financial statements data, we have to use sales as a primary measure of firms' output. Bond et al. [2021] and De Ridder et al. [2021] argue that output price variation, in the absence of proper controls, can render the marginal markup estimates meaningless. In turn, the literature on markup and production function estimation has developed a couple of methods that can be used to correct the elasticity estimates for the output price bias. In this paragraph, we discuss drawbacks and advantages of these methods, as well as their applications in our setting. Importantly, we view the production function estimation primarily as a source of parameter estimates for our counterfactual exercises. Thus, our main task is to evaluate validity of different estimation strategies under the assumptions of our theoretical model. In this discussion, we are not going to consider the properties of the estimation procedures in *other* settings.

First, De Loecker et al. [2020] suggest using firm sales shares ”*measured at various levels of aggregation (two, three, and four digit)*” to control for the difference between output and input prices. Specifically, the authors suggest using this method in case of ACF-corrected LP estimation[26]. Following the literature on proxy functions[27], this estimation strategy is viable, i.e., sales shares can act as an exact control for the price differentials, if the variable input usage is independent of the difference between output and input prices, conditional on firm sales shares. Formally, the conditional independence requirement implies the following: for any triple of values A, B and C, it has to be the case that

$$\begin{aligned} &\Delta p_{\gamma t} = \log p_{\gamma t} - \omega^L_{\nu t} \log w_t - \omega^K_{\nu t} \log p^K_t, \\ &\mathbb{P}\left[l_{\gamma t} = A \middle| \Delta p_{\gamma t} = B, \lambda^{\text{3-digit}}_{\gamma t} = C \right] = \mathbb{P}\left[wl_{\gamma t} = A | \lambda^{\text{3-digit}}_{\gamma t} = C \right]. \end{aligned} \tag{169}$$

Here we assume that the output-input price difference is proxied by the sales shares within a 3-digit industry only, all our arguments go through even if we use sales shares at more than one level of aggregation. We can then show that the above statement does not hold in our setting under the standard functional form specifications. For a moment, let us also assume that the sales shares that are used as proxies do not contain any measurement error terms. The price difference $\Delta p_{\gamma t}$ then can be expressed as follows:

$$\begin{aligned} \log p_{\gamma t} &= \frac{\mu_{\gamma t} w_t l_{\gamma t}}{y_{\gamma t} \omega^L_{\nu t}}, \\ \Delta p_{\gamma t} &= \log \mu_{\gamma t} + \left(1 - \omega^L_{\nu t}\right) \log w_t l_{\gamma t} - \omega^K_{\nu t} \log p^K_t \\ &\quad - \log a_{\gamma t} - \omega^K_{\nu t} \log k_{\gamma t} - \log \omega^L_{\nu t}. \end{aligned} \tag{170}$$

This equation is derived from the De Loecker et al. [2020] formula for the marginal markup estimates, we use it here because firm FOC that is used to derive the markup estimates also describes the optimal pricing strategy of producers. Here we also have to note that under CES demand, the sales shares of producers are an exact control for markups $\mu_{\gamma t}$. Thus, the probabilities in the condition 169 can be rewritten in a following way: assuming that the Cobb-Douglas parameters are constant,

$$\begin{aligned} \mathbb{P}\left[l_{\gamma t} = A \middle| \Delta p_{\gamma t} = B, \lambda^{\text{3-digit}}_{\gamma t} = C \right] &= \mathbb{P}\Bigg[A = B - \log \hat{\mu}\left(C\right) \\ &\qquad - \log w_t + \log Y_{\nu t} + \frac{\sigma_{\gamma t}}{\sigma_{\gamma t} - 1} C + \log \omega^L_{\nu t} \Bigg]. \\ \mathbb{P}\left[l_{\gamma t} = A | \lambda^{\text{3-digit}}_{\gamma t} = C \right] &= \mathbb{P}\left[A = \frac{1}{\omega^L_{\nu t}} \left(\log Y_{\nu t} + \frac{\sigma_{\gamma t}}{\sigma_{\gamma t} - 1} C - \log a_{\gamma t} - \omega^K_{\nu t} \log k_{\gamma t} \right) \right]. \end{aligned} \tag{171}$$

From the lines above, it is clear that the events $A|B, C$ and $A|C$ are generically distinct. The probability of

---

[26] Note that our proof of invalidity of De Loecker et al. [2020] price proxy strategy does not depend on the nature of TFP proxy variable. Regardless of whether we use ACF-corrected Olley and Pakes [1996] or ACF-corrected Levinsohn and Petrin [2003] methodology, it has to be the case that the output price variation is independent of variable input variation, conditional on the price proxy. Unfortunately, the control function that is suggested by De Loecker et al. [2020] for the non-ACF-corrected Olley and Pakes [1996] estimation is also invalid in our setting, primarily because the OP estimation only identifies the variable input elasticity if there exists variation in variable input usage conditional on firm state variables and the TFP proxy. In turn, output prices still depend on the value of variable input expenditures, and we cannot use only productivity proxies and sectoral demand shifters to fully control for price differences.

[27] E.g., see Deaner [2018] – this definition of a perfect control is also in line with the estimation strategies of Olley and Pakes [1996] and [Levinsohn and Petrin, 2003].

the first event only depends on the distribution of sectoral output $Y_{\nu t}$, and variable input price w_t, while the probability of a latter event is also affected by the distribution of firm-level capital stock $\log k_{\gamma t}$ and TFP $\log a_{\gamma t}$. In addition, assuming that the distributions of all the data variables are non-degenerate, if the conditional independence restriction holds for the triplet of values $A = \tilde{A}$, $B = \tilde{B}$, and $C = \tilde{C}$, it cannot hold for the triplet $A = \tilde{A}$, $B = \tilde{B} + 1$, and $C = \tilde{C}$. A similar argument applies if we assume that the price proxy function is formed using the sales share data that contain measurement error(s). It also does not matter whether or not we separately condition on productivity differences across firms.

Finally, let us briefly address the claim of De Loecker et al. [2016] and De Loecker et al. [2020] on that variation in the output prices can sometimes absorb variation in the input prices. This argument relies on the fact that regardless of the structure of demand, output prices are always higher for companies that face higher input prices. However, the issue here is that prices are proportional to the marginal costs, and the price differential term $\Delta p_{\gamma t}$ is equal to the difference between output prices and the CD price index that takes into account both capital and labor. As the equations above illustrate, in our setting marginal costs of production are not collinear with the Cobb-Douglas price index, and thus the price differential term will be correlated with the variable input usage, as well as sectoral demand shifters and physical output. Output prices absorb some part of the variation in input prices but *not all of it*, and this is not particularly helpful if we want to obtain unbiased elasticity estimates.

Also, both De Loecker et al. [2020] and Bond et al. [2021] cite markup inflexibility as an argument against including the assumptions on the demand system in the production function estimation. Our theoretical results demonstrate that, provided that the industry in question is oligopolistic, assumptions on the shape of the demand system do not restrict the set of possible markup distributions. Even CES demand is compatible with an arbitrary markup distribution, as long as the producer beliefs about the market environment are allowed to vary.

H.2. Markup Estimation Under Non-Linear Production and Demand

[Non-CD Production] The estimation process for the CES production functions is quite similar to the CD-CES benchmark. The only difference is that, under CES, the expression for firm-level output is modified to

$$\begin{aligned} \log y_{\nu\gamma t} = & \frac{\xi}{\sigma} \log\left(\omega^K_{\nu t} + \left(1 - \omega^K_{\nu t}\right)\left(\frac{l_{\gamma t}}{k_{\gamma t}}\right)^{\sigma}\right) + \xi_{\nu t} \log k_{\gamma t} \\ & + \omega^{\text{SGA}}_{\nu t} \log \text{SGA}_{\gamma t} + \mathfrak{A}\left(z^K_{\gamma t}, z^A_{\gamma t}, \tilde{k}_{\gamma}, \nu, Z\right). \end{aligned} \tag{172}$$

Thus, under CES production we need to estimate one additional parameter that regulates the substitution between capital and inputs in the COGS bundle. We prefer to explicitly restrict the production function to be CES rather than use a quadratic translog specification, because translog estimates often generate negative estimates of output elasticities $\omega^L_{\gamma t}$, at least for some share of firms in the data sample. Negative variable input elasticities in turn imply negative marginal markup estimates, and neither negative markups nor negative input elasticities are consistent with the short term producer optimization – and common sense.

[Non-CES Demand] We also implement the production function estimation exercise for the non-CES demand system. To do so, we assume that the consumer preferences at the product level satisfy the following restrictions:

$$\log \lambda_{\nu\gamma t} = \tau_\Gamma \log \left(\left(\frac{y_{\gamma t}}{Y_{\nu t}} \right)^{\sigma_\Gamma} + \varpi_\Gamma \right) + \log \hat{\delta}_{\nu t}, \tag{173}$$

Whenever $\varpi_\Gamma = 0$ or $\sigma_\Gamma \to \infty$ and $\sigma_\Gamma \tau_\Gamma \to \bar{\sigma} < \infty$, the demand is reset back to CES, and the corresponding elasticity of substitution in these cases would depend on the product of parameters σ_Γ and τ_Γ. The term $\log \hat{\delta}_{\nu t}$ in the equation above represents the demand index $\delta_{\nu t}$, and by construction it functions as an industry-year fixed effect. Importantly, $\lambda_{\nu\gamma t}$ is represented by monotone function of the relative product-level output, and this function can be inverted in a closed form – this property is necessary for our estimation strategy.

A useful property of this demand system is that it generates a schedule of markups that is concave and increasing in relative output $\frac{Y_{\gamma t}}{Y_{\nu t}}$. Moreover, unlike other types of Kimball preferences that are used in the literature, the quasi-elasticities of demand in this case are always finite and bounded between $\tau_\Gamma \sigma_\Gamma \left(1 + \frac{\varpi}{\underline{y}^{\sigma_\Gamma}}\right)^{-1}$ and $\tau_\Gamma \sigma_\Gamma$, where $\underline{y}$ is the lowest value of relative firm-level output in the sector. Formally, we have

$$1 + \varepsilon_{\gamma t} = \tau_\Gamma \sigma_\Gamma \left(1 + \varpi_\Gamma \left(\frac{Y_{\nu t}}{y_{\gamma t}} \right)^{\sigma_\Gamma} \right)^{-1}. \tag{174}$$

Finite quasi-elasticities generate finite markup values for all producers, in all variety markets. This further means that, unlike Klenow and Willis [2016] preferences, the physical output distribution and the sales distribution under this demand specification do not have to have upper bounds. This property is certainly desirable given that in the data both sales and output distributions have fat tails – in addition, the estimates of marginal markup are relatively low even for the largest firms, and they usually do not approach infinity as the firm output increases.

Under the demand specification described above, the sales function is non-linear in TFP error term $e^A_{\gamma t}$. This means that we cannot estimate Equation 175 as we did it under the CES demand. Instead, we invert the sales function in equation 173 to obtain the expression for relative product output. The second step of the modified ACF estimation is based on the following equation: assuming the production function is Cobb-Douglas,

$$\begin{aligned}
\mathfrak{a}_{\gamma t} &= \frac{1}{\sigma_\Gamma} \log \left(\left(\frac{\lambda_{\gamma t}}{\hat{\delta}_{\nu t}} \right)^{\frac{1}{\tau_\Gamma}} - \varpi_\Gamma \right) - \left(\omega^L_{\nu t} \log l_{\gamma t} + \omega^{\text{SGA}}_{\nu t} \log \text{SGA}_{\gamma t} + \omega^K_{\nu t} \log k_{\gamma t} - \log Y_{\nu t} \right), \\
\mathfrak{a}_{\gamma t} &= \phi^{\text{Own}}_{\nu t} \log z^A_{\gamma(t-1)} + \phi^{\text{Own, 0}}_{\nu(t-1)} \mathbb{1}_{z^A_{\gamma(t-1)}=0} + \sum_{\tau=t-5}^{t-1} \phi^{\text{Ext},\tau}_{\nu t} \bar{z}^A_{\nu\tau} + \\
&\quad + \sum_{\tau=t-5}^{t-1} \phi^{\text{Ext},\tau}_{t} \bar{z}^A_{t} + F\left(\mathfrak{a}_{\gamma(t-1)}\right) + e^A_{\gamma t}.
\end{aligned} \tag{175}$$

Here $\lambda_{\gamma t}$ is the value of product sales share net of the measurement error, as before. This specification also can be adapted for the case of CES production. To note, the terms $\hat{\delta}_{\nu t}$ can be identified up to a constant in the first step of ACF estimation, together with the measurement error $e_{\gamma t}$.

Figure 6: Change in Average Markups from 1980

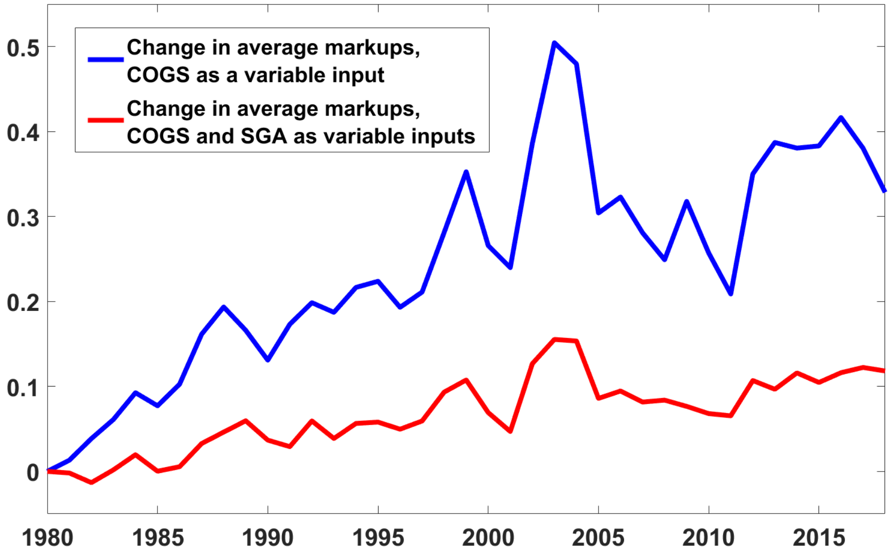
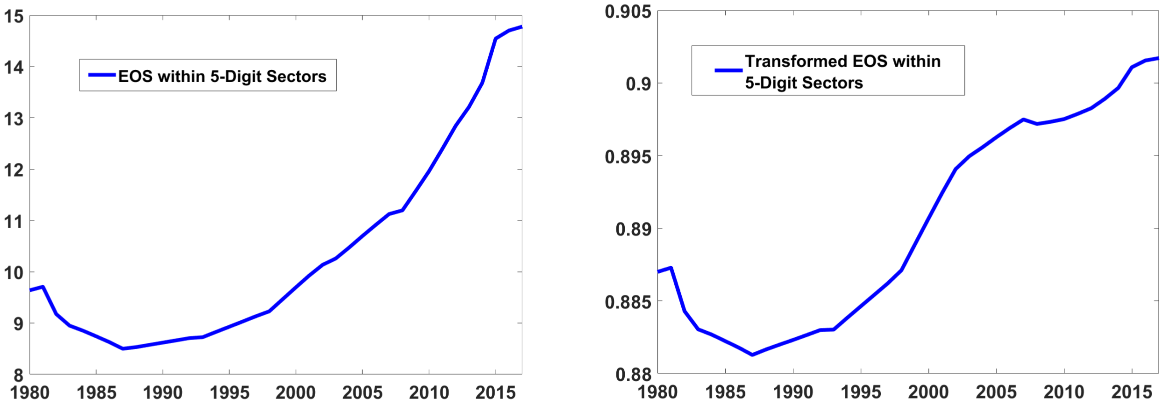


Figure 7: Elasticities of Substitution within 5-Digit Sectors



(a) EOS Values

(b) Transformed EOS Values ($(1 - 1/\sigma)$)

H.3. Miscellaneous Graphs and Tables

Table 16: Sectors for Production Function and Demand Estimation

BEA KLEMS Sector	**NAICS 2-digit #**	**Estimation Sector #**
1. Farms	11	1. Agriculture
2. Forestry, fishing, and related activities	11	1. Agriculture
3. Oil and gas extraction	21	2. Natural Resources
4. Mining, except oil and gas	21	2. Natural Resources
5. Support activities for mining	21	2. Natural Resources
6. Utilities	22	3. Utilities and Construction
7. Construction	23	3. Utilities and Construction
8. Wood products	32	5. Manufacturing #2

Table 16 – continued from previous page

BEA KLEMS Sector	NAICS 2-digit Sector #	Estimation Sector #
9. Nonmetallic mineral products	32	5. Manufacturing #2
10. Primary metals	33	6. Manufacturing #3
11. Fabricated metal products	33	6. Manufacturing #3
12. Machinery	33	6. Manufacturing #3
13. Computer and electronic products	33	6. Manufacturing #3
14. Electrical equipment	33	6. Manufacturing #3
15. Motor vehicles and parts	33	6. Manufacturing #3
16. Other transportation equipment	33	6. Manufacturing #3
17. Furniture and related products	33	6. Manufacturing #3
18. Miscellaneous manufacturing	33	6. Manufacturing #3
19. Food and beverage and tobacco products	31	4. Manufacturing #1
20. Textile mills and textile product mills	31	4. Manufacturing #1
21. Apparel and leather and allied products	31	4. Manufacturing #1
22. Paper products	32	5. Manufacturing #2
23. Printing and related support activities	32	5. Manufacturing #2
24. Petroleum and coal products	32	5. Manufacturing #2
25. Chemical products	32	5. Manufacturing #2
26. Plastics and rubber products	32	5. Manufacturing #2
27. Wholesale trade	42	7. Wholesale and Retail
28. Retail trade	44-45	7. Wholesale and Retail
29. Air transportation	48	8. Transportation
30. Rail transportation	48	8. Transportation
31. Water transportation	48	8. Transportation
32. Truck transportation	48	8. Transportation
33. Transit and ground passenger transportation	48	8. Transportation
34. Pipeline transportation	48	8. Transportation
35. Other transportation	48-49	8. Transportation
36. Warehousing and storage	49	8. Transportation
37. Publishing industries	51	9. Information
38. Motion picture and sound recording	51	9. Information
39. Broadcasting and telecommunications	51	9. Information
40. Information and data processing services	51	9. Information
41. Federal Reserve banks, credit intermediation, etc.	52	10. Finance
42. Securities, commodity contracts, and investments	52	10. Finance
43. Insurance carriers and related activities	52	10. Finance
44. Funds, trusts, and other financial vehicles	52	10. Finance

Table 16 – continued from previous page

BEA KLEMS Sector	NAICS 2-digit Sector #	Estimation Sector #
45. Real estate	53	11. Real Estate
46. Rental and leasing services	53	11. Real Estate
47. Legal services	54	12. Professional and Technical Services
48. Computer systems design and related services	54	12. Professional and Technical Services
49. Miscellaneous professional services	54	12. Professional and Technical Services
50. Management of companies and enterprises	55	13. Administrative Services, etc.
51. Administrative and support services	56	13. Administrative Services, etc.
52. Waste management and remediation services	56	13. Administrative Services, etc.
53. Educational services	61	14. Miscallaneous Services #1
54. Ambulatory health care services	62	14. Miscallaneous Services #1
55. Hospitals, nursing and residential care facilities	62	14. Miscallaneous Services #1
56. Social assistance	62	14. Miscallaneous Services #1
57. Performing arts, spectator sports, museums, etc.	71	14. Miscallaneous Services #1
58. Amusements, gambling, and recreation industries	71	14. Miscallaneous Services #1
59. Accommodation	72	15. Miscallaneous Services #2
60. Food services and drinking places	72	15. Miscallaneous Services #2
61. Other services, except government	81	15. Miscallaneous Services #2

Table 15: Aggregate Average Markup Values in 1980 and 2015

Year	COGS as the only variable input	COGS and SGA as variable inputs
1980	1.410	1.145
2015	1.793	1.249