Innovation, Firm Size Distribution, and Gains from Trade*

Yi-Fan Chen Wen-Tai Hsu Shin-Kun Peng†

March 6, 2022

Abstract

Power laws in productivity and firm size are well-documented empirical regularities. As they are upper right-tail phenomena, this paper shows that assuming *asymptotic power functions* for various model primitives (such as demand and firm heterogeneity) are sufficient for matching these regularities. This greatly relaxes the functional-form restrictions in economic modeling and can be beneficial in certain contexts. We demonstrate this in a modified Melitz (2003) model which embeds an innovation mechanism in order to endogenize the productivity distribution and generate both of the above-mentioned power laws. We also investigate the model's welfare implications with regard to innovation by conducting a quantitative analysis of the welfare gains from trade.

JEL Codes: F12, F13, F41.

Keywords: Power law, firm heterogeneity, asymptotic power functions, regular variation, innovation, gains from trade

---

*This paper was previously circulated under the title "Productivity Investment, Power Laws, and Welfare Gains from Trade". We are grateful for the insightful and helpful comments by the Co-Editor, three anonymous referees, Mathieu Parenti and Mark Razhev. For helpful comments, we also thank Thomas Chaney, Pao-Li Chang, Davin Chor, Jen-Feng Lin, Peter Neary, Jacques-François Thisse, and Jonathan Vogel and the seminar participants at the Baptist University of Hong Kong, Chinese University of Hong Kong, National Taiwan University, University of Melbourne, University of Tokyo, the 2018 Midwest Trade and Theory Meetings, the 2018 SMU-NUS-Paris Joint Trade Workshop, and the 2019 SMU-INSEAD Joint Trade Workshop. The authors gratefully acknowledge the financial support from Academia Sinica (Academia Sinica Investigator Award #IA-105-H04) and the Ministry of Science and Technology (MOST 105-2410-H-001-001-MY3).

†Chen: Department of Applied Economics, National University of Kaohsiung. Email: yfchen@nuk.edu.tw; Hsu and Peng: Institute of Economics, Academia Sinica. Email: wthsu@econ.sinica.edu.tw and speng@econ.sinica.edu.tw, respectively.

1 Introduction

In the literature concerning agent heterogeneity, in particular firm heterogeneity, power functions are often assumed in various key components of model setups. For example, the CES utility function implies that demand is a power function of price; the firm productivity distribution is often assumed to be Pareto (e.g., Melitz and Ottaviano 2008 and Chaney 2008, who assumes this in a Melitz [2003] model) or Fréchet (Bernard et al. 2003, Eaton and Kortum 2002).[1] These assumptions are made for tractability, and they are often justified by enabling the models to generate power laws in productivity and firm size, which are widely-documented empirical regularities (see, for example, Axtell 2001, Luttmer 2007, and Nigai 2017). Formally, a distribution is said to exhibit a power law, and is then called a power-law distribution, if its tail probability at the upper tail is given by a power function, i.e., $\lim_{x\to\infty} \Pr(X \geq x) = \alpha x^{-\zeta}$, for some positive constants α and ζ.[2]

In principle, since power laws are in themselves upper right-tail phenomena, the functional-form restrictions on these key setups for matching the power laws could be relaxed, and in certain contexts, such a relaxation can be beneficial. For example, the CES utility function has been shown to be quite restrictive as it implies constant markups and homothetic demand (Zhelobodko et al. 2012), and the empirical firm size distribution is shown to follow a Pareto distribution only at the upper right tail (Nigai 2017). This paper aims to clarify what relaxations on these key setups can be made, and we show these in the canonical trade model of Melitz (2003). As the distribution of productivity draws in the Melitz model is exogenously given, we embed an innovation mechanism into Melitz (2003) in order to microfound both the power laws in productivity and firm size.[3]

[1] The Fréchet and Pareto distributions are tail-equivalent as their tail probabilities are asymptotically proportional. Hence, the Fréchet is also a power-law distribution and generates power laws in firm size.

[2] It has been shown that the power laws in productivity and firm size provide a microfoundation for the gravity equations (Arkolakis et al. 2019 and Chaney 2018) and that the few very large firms may be what matters the most for macroeconomic performance, i.e., granular economies (Gabaix 2011). Furthermore, the power-law coefficients are often tightly connected with welfare evaluation (as suggested by Arkolakis et al. 2012 and Arkolakis et al. 2019). Thus it is important to understand the circumstances under which these power laws may emerge.

[3] Empirical evidence has shown that trade may induce changes in the productivity of surviving/active firms; see, e.g., Pavcnik (2002), Lileeva and Trefler (2010), Bustos (2011), and Aghion et al. (2019).

Our model starts with a simple relation between productivity, innovation effort, and a firm's capability. More innovation leads to higher productivity, and the higher a firm's capability, the less innovation effort is needed to entail the same level of productivity. One feature of this relation is that productivity is determined by effort and capability jointly in a multiplicative manner; we microfound this feature by an R&D process in which firms decide the complexity of their production procedures and conduct Bernoulli trials (experiments) to improve the performance of each procedure. Firms differ in their probabilities of failure in these Bernoulli trials. We first provide an illustrative example that embeds this relation into a standard Melitz model (with a CES demand) as an innovation stage after entry and before production. As a first cut, we show that if the distribution of the probability of failure across firms has a *finite and positive* density near zero (i.e., those very capable firms), then the power laws emerge.

We then show that power laws continue to hold when all functions as the primitives of this model are generalized to *asymptotic power functions*, henceforth APFs. This class of functions is actually more general than it may seem at first glance. First, when the demand is generalized to an APF (i.e., an *asymptotic CES*), many widely-used non-CES and/or non-homothetic preferences are included. Second, the innovation cost function permits at least general polynomials. Third, the limit of density of the distribution of failure probability at zero can be either zero, a positive constant, or infinite, provided that the density is asymptotically a power function at the left tail. This actually subsumes the limit condition in the illustrative example. As we will explain in Section 2, this includes many well-known and widely-used distributions.

The key step involves mapping from model primitives as APFs via firms' optimal choices to power laws. Such a mapping is straightforward if the model primitives are exact power functions, as in Chaney (2008). However, this is not a trivial task when the primitives are not exact power functions because firms' optimal choices need to be properly taken care of when the underlying functional forms for demand and innovation cost are unknown. Our contribution here is to utilize the tool of *regular variation* to tackle this problem, as APFs are *regularly varying*.

All of the above-mentioned results are shown in a closed economy. We then go on to show that these results continue to hold in a very general open-economy environment where all model parameters are allowed to be country-specific. Interestingly, the tail indices of both the productivity and firm size distributions of each country depend on the market with the largest competitiveness (largest price elasticities). As a result, opening up to trade (weakly) fattens the right tails of both productivity and firm size distributions.

Our setup on innovation is relatively simple in the sense that we use a general functional relation to link productivity to innovation effort and firm capability. This approach is adopted instead of using a more sophisticated setup because we seek to demonstrate the techniques and the above-described results under a simple and standard setup for clarity. The approach can be applied to more sophisticated models of innovation. For example, in the immense literature on innovation and growth, many have assumed power functions for demand or underlying firm heterogeneity for tractability in order to focus on dynamics for various purposes; see, e.g., Sampson (2016), Desmet et al. (2018), and Hsieh et al. (2021). Our results can be used either to generalize their environments by means of APFs or as a justification for these more restrictive assumptions. Another strand of the literature focuses on the relations between trade and innovation in static settings, such as Lileeva and Trefler (2010), Bustos (2011), Bas and Ledezma (2015), Bonfiglioli et al. (2019), and Aghion et al. (2019), in which either the underlying firm heterogeneity is assumed to follow a Pareto distribution or the functional form is not specified for generality. In terms of matching the power laws, our approach can thus be applied to these models either as a generalization to those assuming the existence of a Pareto distribution or as a refinement to those keeping the functional forms general.

Such an approach can be applied generally to models of agent heterogeneity. An obvious example is indeed the model of Melitz (2003), as the power law in firm size emerges if the demand and the distribution of productivity draws are each an APF. This approach can also be applied to the model of Yeaple (2005), who shows how trade affects the wage distribution and skill premium. In light of our findings, it is a simple exercise to generate a power law in the wage distribution in his framework by assuming that worker skill is distributed in the form of an APF and the marginal product of labor as a function of worker skill is also an APF.

To further appreciate the role of APFs, it is important to distinguish our work from that of Mrázová et al. (2021), who show that if two firm variables of interest are related by a power function, then the distribution of one variable following a “generalized power function” (GPF) implies that the other variable also follows the *same* GPF family. They emphasize the role of the CREMR (Constant Revenue Elasticity of Marginal Revenue) demand in linking firm productivity and sales so that both variables are in the same GPF family. What concerns the current paper is different, as we focus on the environment under which power laws would emerge. Most importantly, neither of the two classes of functions, APFs and GPFs, subsumes the other; hence none of the theoretical results in either paper subsumes the other. It is permissible that the APFs for demand, innovation cost, and failure probability are all of different families, and there can also be applications of theorems by Mrázová et al. (2021) that do not conform to power laws.

To study the role of innovation in this modified Melitz model, the remainder of this paper examines the effect of trade cost on the productivity distribution and conducts a quantitative analysis of the welfare gains from trade. For this purpose and for tractability, we focus on a symmetric-country world with CES preferences. Section 3 analyzes how the productivity distribution is affected by trade liberalization. We show that a lower variable trade cost increases (decreases) exporters’ (non-exporters’) innovation effort. On the one hand, a lower trade cost implies a larger effective market size facing the exporters. Hence, the exporters’ marginal benefit of having a higher productivity increases, leading them to innovate more. On the other hand, non-exporters face more import competition and make less profit as the prices of imported goods are reduced, not only because of a lower variable trade cost, but also due to the fact that these foreign exporters become more productive. Consequently, a lower trade cost negatively affects the productivity of non-exporters.

Section 4 performs a quantitative analysis to clarify how innovation affects the welfare gains from trade. To highlight the role played by innovation, we compare the welfare gains from trade in this model with those in the Melitz (2003) model with the Pareto productivity distribution. When firms' R&D abilities are uniformly distributed, the resulting productivity follows the Pareto distribution (with a jump at the exporting cutoff); thus such a parameterization is adopted. We calibrate the model to match the same trade elasticity, domestic expenditure share, and the share of exporters. Following Melitz and Redding's (2015) approach of comparing across models by fixing common parameters, we compare the two models conditional on the same trade elasticity and values of the common parameters. Our quantitative analysis finds that the model with innovation entails larger welfare gains from trade than Melitz–Pareto by about 40%.

This paper is closely related to the literature on power laws in firm size. Our central contribution to this literature is showing the functional-form generality for various aspects of a rather generalized general equilibrium model of trade in generating the power laws. Thus, it is fundamentally different from the popular explanation of power laws based on firm-size dynamics that follow a random growth process; see, e.g., Luttmer (2007), Acemoglu and Cao (2015), and König et al. (2016).[4] Recently, Chaney (2014, 2018) and Geerolf (2017) have provided explanations for the power law in firm size via network and firm hierarchy, respectively. Our mathematical mechanism for power laws is different from Chaney (2014, 2018) which relies on "preferential attachments" that are similar to random growth mathematically but have interesting applications to trade. Our economic mechanism differs from Geerolf (2017), which builds on the canonical firm-hierarchy model of Garicano (2000) but does not feature innovation, selection, or trade. Moreover, in terms of the mathematical mechanism, Geerolf's (2017) application of the power law change of variable technique to his model is also used here in our illustrative example, but his study does not explore APFs or use regular variation.

[4]Also see Gabaix (2009) for a survey of this random-growth approach. Even though König et al. (2016) also incorporate innovation, how innovation is used to generate power laws is quite different from our study, as their approach is still primarily a random-growth process.

Since Arkolakis et al. (2012; henceforth ACR), there has been a revived and still growing literature on the welfare gains from trade (see Costinot and Rodríguez-Clare 2015 for a survey). The literature on the gains from trade with an innovation mechanism has built on dynamic models in which substantial dynamic-innovation gains are found.[5] By contrast, we show that gains from trade due to innovation can also be substantial in a static general equilibrium model.[6]

2 Power Laws in Productivity and Firm Size

We start with a closed economy model to illustrate the mechanism of innovation. We show how power laws for productivity and firm size emerge from such a model. Section 2.3 shows that such results can easily be extended to a general open-economy environment.

2.1 Model Setup

There are N individuals in the economy, each of which is endowed with one unit of labor. All individuals are identical in their income earned from wages w, and they spend their income on a continuum of varieties, each of which is indexed by υ. Assume that the aggregate inverse demand function is given by $p = D\left(q\left(\upsilon\right); A\right)$ on $\left[\underline{q}, \infty\right)$ with $\underline{q} \geq 0$.[7] That is, we assume that the inverse demand of a variety depends on all the other varieties only through an aggregate variable $A \in \mathbb{R}$. We assume that D is twice-differentiable and that the law of demand holds: $D' < 0$.[8]

[5] For example, Hsieh et al. (2021) find that the innovation mechanism may increase the gains from trade by 31$\sim$75%, while Impullitti and Licandro (2017) and Hsu et al. (2020b) find that it can amplify the gains from trade by 2 and 4.5 times, respectively.

[6] As our model entails the local ACR formula, the innovation mechanism here does not contribute to any extra gains from trade conditional on the domestic expenditure share and trade elasticity. However, our quantitative exercise is different because we adopt the approach of Melitz and Redding (2015) to compare the implications of different models conditional on the same parameters.

[7] We allow $\underline{q} > 0$ for generality. There are demands with minimal consumption, such as the CREMR demand given in Table 1.

[8] Such an inverse demand function can be generated by (but is not limited to) maximizing an additively separable utility function $U = \int_{\upsilon \in \Upsilon} u\left(x\left(\upsilon\right)\right) d\upsilon$ subject to the budget constraint $\int_{\upsilon \in \Upsilon} p\left(\upsilon\right) x\left(\upsilon\right) d\upsilon = w$. The sub-utility $u\left(.\right)$ is defined on $\left[\underline{x}, \infty\right)$ with $\underline{x} \geq 0$. Assume that $u' > 0$ and $u'' < 0$. The standard solution

On the production side, labor is the only input, and firms engage in monopolistic competition. To enter, each entrant hires a κ_e amount of labor, which allows the entrant to obtain a differentiated variety and a draw of an R&D parameter γ from a given distribution which we will explain shortly. For a firm to produce, κ_D units of labor as fixed input are required. The productivity of a firm is endogenously determined and denoted by φ. Denoting wages as w, the total cost of production as a function of output q is $w\left(q/\varphi+\kappa_D\right)$. A firm's profit from production is

$$\pi\left(\varphi\right)\equiv\max_{q}\widetilde{\pi}\left(q;\varphi\right)\equiv\max_{q}pq-w\varphi^{-1}q-w\kappa_D. \tag{1}$$

A solution to this problem is denoted by $q^*\left(\varphi\right)$. Note that if there are multiple optimal solutions, $q^*\left(\varphi\right)$ could refer to any one of them.[9] In addition, $\pi\left(\varphi\right)$ is a function regardless of whether there are multiple values of $q^*\left(\varphi\right)$.

Each firm determines its productivity by conducting R&D. That is, in addition to the entry and production stages in Melitz (2003), there is a second stage in which firms innovate to determine productivity. The R&D efforts are in terms of labor, and the labor requirement k for a γ-typed firm to acquire a productivity level φ is given by the function

$$k=\gamma V\left(\varphi\right)+\kappa_R, \tag{2}$$

where $V\left(\cdot\right)$ is twice-differentiable, strictly increasing, and convex on $\mathbb{R}_+$ with $\lim_{\varphi\to\infty}V\left(\varphi\right)=\infty$, and κ_R is the fixed cost of innovation. In this *innovation cost function*, γ is multiplicatively separable from V and serves as an inverse measure for a firm's R&D efficiency. This functional form can be microfounded by an R&D process in which firms decide the complexity of their production procedures and conduct Bernoulli trials (experiments) to

yields the inverse demand function $p=D\left(q\left(\upsilon\right);A\right)\equiv u'\left(\frac{q(\upsilon)}{N}\right)/A=u'\left(x\left(\upsilon\right)\right)/A$ on $\left[\underline{q},\infty\right)$, where $\underline{q}\equiv N\underline{x}$ and A is the Lagrange multiplier of an individual consumer's problem and is a general equilibrium object. Note that $u''<0$ implies that $D'<0$. For forms of $D\left(q\left(\upsilon\right);A\right)$ other than $u'\left(q\left(\upsilon\right)\right)/A$, see Table 1 and the discussion following Assumption 1.

[9]In a similar monopolistic-competition environment, Parenti et al. (2017) show that the profit-maximization problem entails a unique solution if and only if the marginal revenue is strictly decreasing. This requires the inverse demand function to be concave, linear, or not too convex. In general, multiple solutions are possible if no constraint is imposed on $D''(.)$.

improve the performance of each procedure. Firms differ in their probabilities of failure $\gamma \in (0,1]$ in these Bernoulli trials. The c.d.f. and p.d.f. of the distribution of γ are denoted by $F(\cdot)$ and $f(\cdot)$, respectively. Equation (2) suffices for our purposes; we provide a microfoundation of this R&D process in Appendix A.1.

A γ-typed firm chooses an optimal productivity level φ that maximizes its total profit

$$\Pi(\gamma) \equiv \max_{\varphi} \pi(\varphi) - w\gamma V(\varphi) - w\kappa_R, \tag{3}$$

and the resulting optimal choice of φ is denoted by $\varphi^* = \widetilde{\varphi}(\gamma)$. Again, if there are multiple optimal solutions, $\widetilde{\varphi}(\gamma)$ could refer to any one of them. At the beginning of the second stage, a firm chooses to innovate if and only if $\Pi(\gamma) \geq 0$, i.e., when the firm with γ finds that its operating profit $pq - w\varphi^{-1}q$ under $\widetilde{\varphi}(\gamma)$ and $q^*(\widetilde{\varphi}(\gamma))$ is sufficient to cover the innovation cost $w\gamma V(\widetilde{\varphi}(\gamma)) + w\kappa_R$ and fixed cost of production $w\kappa_D$. Any firm which chooses to innovate must subsequently operate in the third stage because $\Pi(\gamma) \geq 0$ implies that $\pi(\varphi) > 0$. As there may be firm selection, the set of surviving firms is denoted by Ω. The free-entry condition is

$$E(\Pi) \equiv \int_{\Omega} \Pi(\gamma)\, dF(\gamma) = w\kappa_e. \tag{4}$$

The goods market clearing condition is expressed as

$$M_e \int_{\Omega} q^*(\widetilde{\varphi}(\gamma))\, D(q^*(\widetilde{\varphi}(\gamma)); A)\, dF(\gamma) = wN, \tag{5}$$

where M_e is the number of entrants. The labor market clearing condition is $N = N_e + N_{in} + N_p$, where N_e, N_{in}, and N_p are the masses of labor employed for entry, innovation, and production, respectively: $N_e = M_e\kappa_e, N_{in} = M_e \int_{\Omega} (\gamma V(\widetilde{\varphi}(\gamma)) + \kappa_R)\, dF(\gamma)$, and $N_p = M_e \int_{\Omega} \left(\frac{q^*(\gamma)}{\widetilde{\varphi}(\gamma)} + \kappa_D \right) dF(\gamma)$.

In sum, the model contains the following three stages:

Stage 1. Entry Stage: Each potential entrant decides whether to enter the market. If an entrant decides to enter, it pays the fixed entry cost κ_e, obtains a differentiated variety, and draws its type γ randomly from the distribution $f(\gamma)$.

Stage 2. Innovation Stage: Given γ, each firm decides whether to innovate or not, and if it does decide to, determines its productivity level φ. Those that innovate proceed to the next stage, and those that do not innovate exit the market.

Stage 3. Production/Consumption Stage: Each firm pays κ_D and determines its output and price. Production and consumption take place and markets clear.

An equilibrium consists of $\{w, A, M_e, \Omega\}$ such that the set of surviving firms Ω is determined by $\Pi(\gamma) \geq 0$, and $\{w, A, M_e\}$ are jointly determined by (4–5) and the labor market clearing condition. In such a closed-economy setting, one of these conditions is made redundant by Walras' Law. We choose labor as the numéraire, and so $w = 1$.

Note that the free-entry condition implies that consumers' total expenditure is equal to the entry costs, innovation costs, and all of the production costs.[10] In other words, consumers are collectively the financiers for the firms in their entry and innovation stages. To ensure that all firms that are willing to innovate (and subsequently produce) get funded, we assume that financiers do not observe firm types. As entrants are ex-ante identical, the firm ownership is uniformly allocated to consumers.

2.2 Equilibrium and Power Laws

This subsection provides an exposition of how power laws for both productivity and firm size emerge. We first provide an illustrative example which differs from the Melitz model only by having an innovation stage in which the variable innovation cost function V is given by a power function. We show how a weak restriction on the underlying firm heterogeneity f allows the power laws in productivity and firm size to emerge. By using regular variation, we then show that the power laws continue to hold when inverse demand, innovation cost, and the underlying firm heterogeneity are all APFs.

[10] It is straightforward to verify this by plugging (1) and (3) into the free-entry condition (4).

2.2.1 Illustrative Example Using the Melitz Model

As in Melitz (2003), consider a CES demand: $p = \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{1}{\sigma}} q^{-\frac{1}{\sigma}}$ where $\sigma > 1$. For tractability, we use a simple power function for the innovation cost: $k = \gamma\varphi^{\beta} + \kappa_R$, where $\beta > 1$. For any φ, a firm's optimal output in Stage 3 is given by $q^*(\varphi) = \frac{N}{P^{1-\sigma}}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}\varphi^{\sigma}$. The operating profit is accordingly $\pi(\varphi) = \frac{N}{P^{1-\sigma}}\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\sigma-1}\varphi^{\sigma-1} - \kappa_D$.

In Stage 2, a γ-type firm decides its productivity level to maximize its total profit (3). The resulting productivity as a function of γ is

$$\widetilde{\varphi}(\gamma) \equiv \left[\frac{\frac{N}{P^{1-\sigma}}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{1}{\beta-\sigma+1}} \gamma^{-\frac{1}{\beta-\sigma+1}}. \tag{6}$$

It is readily verified that the second-order condition is satisfied if and only if $\beta > \sigma - 1$, i.e., the innovation cost function is sufficiently convex. This condition is imposed here. A firm's total profit becomes

$$\Pi(\gamma) = \frac{\beta-\sigma+1}{\sigma-1}\left(\frac{\sigma-1}{\sigma}\right)^{\frac{\beta\sigma}{\beta-\sigma+1}}\left(\frac{N}{\beta P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}}\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} - \kappa_D - \kappa_R, \tag{7}$$

and hence there is a unique cutoff

$$\gamma_D = \left[\frac{1}{\kappa_D+\kappa_R}\frac{\beta-\sigma+1}{\sigma-1}\left(\frac{\sigma-1}{\sigma}\right)^{\frac{\beta\sigma}{\beta-\sigma+1}}\left(\frac{N}{\beta P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}}\right]^{\frac{\beta-\sigma+1}{\sigma-1}} \tag{8}$$

such that $\Pi(\gamma) \geq 0$ if and only if $\gamma \leq \gamma_D$.

An equilibrium consists of $\{P, M_e, \gamma_D\}$ such that γ_D is given by (8), and $\{P, M_e\}$ are jointly determined by the price index formula $P^{1-\sigma} = M_e\left\{\int_0^{\gamma_D}\left[\frac{\sigma-1}{\sigma}\widetilde{\varphi}(\gamma)\right]^{\sigma-1} dF(\gamma)\right\}$ and the free-entry condition $\int_0^{\gamma_D}\Pi(\gamma)\,dF(\gamma) = \kappa_e$, where $\widetilde{\varphi}(\gamma)$ and $\Pi(\gamma)$ are given by (6–7).[11]

It is important to explore the implications of adding innovation to the Melitz model, as this is the distinctive feature of our model. For this purpose, we study the comparative statics of the innovation cost parameter β on the key equilibrium objects γ_D and M_e. Com-

[11]Note that the price index is indeed the A defined in Section 2.1, and the price index formula is derived from the goods market clearing condition.

bining (8) with the free-entry condition, we have $\frac{\kappa_e}{\kappa_D+\kappa_R} = \int_0^{\gamma_D} \left[\left(\frac{\gamma_D}{\gamma} \right)^{\frac{\sigma-1}{\beta-\sigma+1}} - 1 \right] dF(\gamma)$, which implies that γ_D increases in β because the right-hand side increases in γ_D and $\beta > \sigma - 1$. As innovation becomes more costly, equilibrium productivities are lowered and competition is softened. Thus, selection becomes more lenient. It is readily verified that $M_e = \frac{N}{\kappa_e + (\kappa_D+\kappa_R)F(\gamma_D)} \times \left(\frac{1}{\sigma} - \frac{\sigma-1}{\sigma}\frac{1}{\beta} \right)$. How M_e changes in β is generally ambiguous because the first multiplicative term decreases in β while the second term increases in β. This ambiguity reflects two counterveiling forces on the expected profit. First, higher innovation costs imply that an entrant obtains a lower productivity, which lowers the expected profit and discourages entry. Second, competition is softened because others also obtain lower productivities, increasing the expected profit. However, if one assumes F to be the uniform distribution (which we sometimes do in later parts of the paper because it generates the Pareto distribution), then $M_e = \frac{\sigma-1}{\beta\sigma}\frac{N}{\kappa_e}$, which means that the first above-mentioned effect dominates.

Next, we examine the productivity and firm size distributions. Firm size is defined by firm revenue $s \equiv pq$. It is readily verified that firm size as a function of γ is given by

$$\widetilde{s}(\gamma) \equiv \left[\frac{N}{P^{1-\sigma}} \left(\frac{\sigma-1}{\sigma} \right)^{\sigma} \right]^{\frac{\beta}{\beta-\sigma+1}} \beta^{-\frac{\sigma-1}{\beta-\sigma+1}} \frac{\sigma}{\sigma-1} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}. \tag{9}$$

Let G and G_s be the cumulative density functions of productivity φ and firm size s, respectively; let g and g_s denote the corresponding density functions. Clearly, a distribution exhibiting a power law with a tail index ζ is equivalent to its density following a power function with exponent $-\zeta - 1$ at the right tail. It is often more convenient to work with the equivalent definition in terms of density. By applying a change of variable, the density functions of productivity and firm size are

$$\begin{aligned} g(\varphi) =& \frac{f\left(\widetilde{\varphi}^{-1}(\varphi)\right)}{F(\gamma_D)} \frac{N}{P^{1-\sigma}} \frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}(\beta-\sigma+1)}{\beta} \varphi^{-(\beta-\sigma+1)-1}, \\ g_s(s) =& \frac{f\left(\widetilde{s}^{-1}(s)\right)}{F(\gamma_D)} \left(\frac{N}{P^{1-\sigma}} \right)^{\frac{\beta}{\sigma-1}} \frac{\beta-\sigma+1}{\beta\sigma} \left(\frac{\sigma-1}{\sigma} \right)^{\beta} s^{-\frac{\beta-\sigma+1}{\sigma-1}-1}. \end{aligned}$$

Observe from (6) and (9) that both φ and s become arbitrarily large as γ approaches 0.

Hence, $g(\varphi)/\varphi^{-(\beta-\sigma+1)-1}$ and $g_s(s)/s^{-\frac{\beta-\sigma+1}{\sigma-1}-1}$ approach constants if $\lim_{\gamma\to 0} f(\gamma) = K > 0$. In other words, power laws emerge if the density of γ has a finite and positive limit at zero. If γ is uniformly distributed, then the above distributions are both Pareto, a special case of power-law distributions. Note that the expected profit $E(\Pi)$ is finite if and only if $\int_0^{\gamma_D} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} f(\gamma)\, d\gamma < \infty$. It is readily verified that the condition $\lim_{\gamma\to 0} f(\gamma) = K > 0$ ensures that the expected profit is finite if $\beta > 2(\sigma-1)$.

The above mechanism is referred to as a "power law change of variable close to the origin": if the density of a random variable x has a finite and positive limit at the origin, and the variable of interest y is related to x in a reciprocal way, then y becomes arbitrarily large as x goes to 0 and the distribution of y exhibits a power law tail.[12] Since productivity φ is related to γ in a reciprocal way given by (6), the condition $\lim_{\gamma\to 0} f(\gamma) = K > 0$ entails a power law in the productivity distribution.

The above simple example illustrates how the addition of the innovation stage to the Melitz model and a weak restriction on firm heterogeneity f can give rise to power laws in both productivity and firm size. One naturally wonders how much the result depends on the power-function assumptions on demand D and innovation cost V, and what happens if the density f tends to infinity or zero when $\gamma \to 0$. We will show that the conditions on D, V, and f can all be generalized to asymptotic power functions using regular variation. In particular, the limit condition $\lim_{\gamma\to 0} f(\gamma) = K > 0$ is simply a special case of f being an APF, which can go to either zero, a positive constant, or infinity when $\gamma \to 0$.

2.2.2 Preliminaries: Regularly and smoothly varying functions

We first provide some preliminaries on regular variation. A function $v(x)$ is *regularly varying* if for some $\zeta \in \mathbb{R}$ and for all $t > 0$, $\lim_{x\to\infty} v(tx)/v(x) = t^{\zeta}$. This implies that one can write $v(x) = x^{\zeta} l(x)$, where $l(x)$ is referred to as a *slowly varying* function, i.e., a regularly varying function with $\zeta = 0$. The definition of a smoothly varying function is as follows (see, e.g., Bingham et al. 1989):

[12]This technique has already been used in physics; see Sornette (2006, Section 14.2.1).

Definition 1. A positive function v defined on some neighborhood of infinity *varies smoothly* with index $\zeta \in \mathbb{R}$ if for all $n \geq 1$

$$\lim_{x\to\infty} \frac{x^n v^{(n)}(x)}{v(x)} = \zeta(\zeta-1)...(\zeta-n+1),$$

where $v^{(n)}(x)$ denotes the n-th derivative of $v(x)$.

A smoothly varying function is a regularly varying function that does not oscillate too much. More importantly, any regularly varying function can be approximated by a smoothly varying function asymptotically (Theorem 1.8.2 of Bingham et al. 1989). Since we are concerned with the tail behavior of the productivity distribution and operationally smoothly varying functions will be used, this theorem ensures that our results also apply to regularly varying functions. Note that if $l(x)$ is a smoothly and slowly varying function, then Definition 1 implies that

$$\lim_{x\to\infty} x\frac{l'(x)}{l(x)} = \lim_{x\to\infty} x^2 \frac{l''(x)}{l(x)} = 0. \tag{10}$$

We now formally state our assumptions on the demand and innovation cost functions.

Assumption 1. *The inverse demand function of each variety can be expressed as* $p = D(q;A) \equiv q^{-\frac{1}{\sigma}} Q(q;A)$, *where* $\sigma > 1$ *and* $\lim_{q\to\infty} Q(q;A) = C_Q > 0$.

That Q has a positive limit is an equivalent way of stating that the inverse demand D is an asymptotic power function. This implies that Q is slowly varying and hence D is regularly varying. As mentioned, we work with the *smoothly varying representations* of these functions without loss of generality. As we will show shortly that there are one-to-one mappings at the tails between $\gamma \to 0$ and $\varphi \to \infty$ and between $\varphi \to \infty$ and $q \to \infty$, the requirement that $\sigma > 1$ is needed to ensure that the demand is consistent with monopoly pricing at these tails.

APF inverse-demand functions are actually more general than they may appear at first glance. Needless to say, they include the CES demand. As shown in Table 1, several important classes of demand functions with variable demand elasticity also satisfy this

Demand class	Functional form	Inverse demand
Bipower Direct	$q = \widehat{a}p^{-\nu} + ap^{-\sigma} \equiv \boldsymbol{q}(p)$ $\sigma > \nu \geq 1, a > 0$	$p = q^{-\frac{1}{\sigma}} \left(\widehat{a}\left[\boldsymbol{q}^{-1}(q)\right]^{\sigma-\nu} + a\right)^{\frac{1}{\sigma}}$
Pollak (HARA)	$q = \widehat{a} + ap^{-\sigma}$ $\sigma > 1, a > 0, \widehat{a} \leq 0$	$p = q^{-\frac{1}{\sigma}} a^{\frac{1}{\sigma}} \left(1 - \frac{\widehat{a}}{q}\right)^{-\frac{1}{\sigma}}$
PIGL	$q = \widehat{a}p^{-1} + ap^{-\sigma} \equiv \boldsymbol{q}(p)$ $\sigma > 1, a > 0$	$p = q^{-\frac{1}{\sigma}} \left(\widehat{a}\left[\boldsymbol{q}^{-1}(q)\right]^{\sigma-1} + a\right)^{\frac{1}{\sigma}}$
QMOR	$q = ap^{r-1} + \widehat{a}p^{\frac{r}{2}-1} \equiv \boldsymbol{q}(p)$ $\sigma \equiv 1 - r > 1, a > 0$	$p = q^{\frac{1}{r-1}} \left(a + \widehat{a}\left[\boldsymbol{q}^{-1}(q)\right]^{-\frac{r}{2}}\right)^{\frac{1}{1-r}}$
Bipower Inverse	$p = \widehat{a}q^{-\nu} + aq^{-\frac{1}{\sigma}}$ $\sigma \geq \sigma\nu > 1, a > 0$	$p = q^{-\frac{1}{\sigma}} \left(\widehat{a}q^{\frac{1}{\sigma}-\nu} + a\right)$
CEMR (Inverse PIGL)	$p = \widehat{a}q^{-1} + aq^{-\frac{1}{\sigma}}$ $\sigma > 1, a > 0$	$p = q^{-\frac{1}{\sigma}} \left(\widehat{a}q^{\frac{1-\sigma}{\sigma}} + a\right)$
CREMR	$p = \frac{a}{q}(q - \widehat{a})^{\frac{\sigma-1}{\sigma}}$ $\sigma > 1, a > 0, q > \widehat{a}\sigma$	$p = q^{-\frac{1}{\sigma}} a \left(1 - \frac{\widehat{a}}{q}\right)^{\frac{\sigma-1}{\sigma}}$

Table 1: Examples of demands satisfying Assumption 1

assumption.[13] Assumption 1 includes several demand classes that exhibit "manifold invariance" (Mrázová and Neary 2017),[14] including Bipower Direct demand, Bipower Inverse demand, Pollak family demand which is equivalent to the HARA (Hyperbolic Absolute Risk Aversion) preference, PIGL (Price-Independent Generalized Linear) demand, the QMOR (Quadratic Mean of Order r) expenditure function, and CEMR (Constant Elasticity of Marginal Revenue) demand. It also includes CREMR demand (Mrázová et al. 2021).

Note that the conditions $D' < 0$ and that $\lim_{q\to\infty} Q(q; A) = C_Q > 0$ imply that $p = D(q) > 0$ for all $q \in [\underline{q}, \infty)$. Thus, any demand with a quantity intercept (such as the linear demand) is excluded by Assumption 1. For a different reason, CARA (Constant Absolute Risk Aversion) demand is also excluded because its price elasticity tends to 0 when q goes to infinity, which is inconsistent with the requirement that $\sigma > 1$.[15]

[13]Details are provided in the online appendix, which is available at https://wthsu.com.

[14]A demand manifold depicts the relation between price elasticity and the curvature of the demand function, and the demand manifolds in these two classes are invariant to changes in general equilibrium objects, making them powerful tools for inferring demand/welfare by micro-level information such as firm sales and markups.

[15]To see this, observe that the CARA demand can be written as $q = a - b \ln p$, where $a, b > 0$. Its price

Assumption 2. *The innovation cost function can be written as* $k(\varphi) = \gamma V(\varphi) + \kappa_R \equiv \gamma\varphi^{\beta} L(\varphi) + \kappa_R$*, where* $\beta > 1$ *and* $\lim_{\varphi\to\infty} L(\varphi) = C_L > 0$*.*

The assumption on the innovation cost function parallels that on the inverse demand function. That is, V is an asymptotic power function; L is slowly varying and V is regularly varying. Obviously, simple power functions are included, but general polynomial functions are also included.

2.2.3 Equilibrium quantity and productivity

One key step in the illustrative example involves inverting (6). Such inversion is possible because $q^*(\varphi)$ and $\widetilde{\varphi}(\gamma)$ are monotonic functions in that example. Such a property is not guaranteed when the functional forms of the inverse demand D and variable innovation cost V are generally unknown. By dealing with the second- and third-stage problems in the limit with regular variation, we now show that $q^*(\varphi)$ and $\widetilde{\varphi}(\gamma)$ are indeed unique at least for firms with small γ.

For any given φ, the first- and second-order conditions for an interior solution q from (1) in Stage 3 are

$$p'q + p - \varphi^{-1} = 0 \tag{11}$$

$$p''q + 2p' < 0. \tag{12}$$

With the law of demand, these imply that $|\epsilon(q)| \equiv -p/(qp') > 1$, and $\mu(q) \equiv -(p''q)/p' < 2$. That is, at the interior solution q, the demand elasticity must be greater than 1 so as to be consistent with monopoly pricing, and the convexity of the demand curve must be sufficiently small in order to satisfy the second-order condition.

Note that Assumption 1 regulates the inverse demand $p = D(q; A)$ only for large values of q. As there is no guarantee that the profit function will be strictly concave or quasi-concave in the entire domain of q, there may exist corner solutions to the profit-maximization problem or multiple solutions satisfying (11) and (12). The lemma below

elasticity equals b/q.

shows that, for firms with high productivity levels, the solution is interior and unique. For these firms and with Assumption 1, (11) and (12) can be written as

$$\varphi = q^{\frac{1}{\sigma}}\left[Q \times \left(1-\frac{1}{\sigma}+q\frac{Q'}{Q}\right)\right]^{-1} \tag{13}$$

$$q^{-\frac{1}{\sigma}-1}Q\left[-\frac{1}{\sigma}\left(1-\frac{1}{\sigma}\right)+2\left(1-\frac{1}{\sigma}\right)q\frac{Q'}{Q}+q^2\frac{Q''}{Q}\right] \equiv \widetilde{\pi}_{qq}\left(q;\varphi\right)<0. \tag{14}$$

The following assumption is a restriction on demand around $\underline{q}$ that is needed for equilibrium consistency, as will be explained in the proof of Lemma 1. It turns out that this assumption also rules out the potential corner solution at $\underline{q}$.

Assumption 3. *The inverse demand function D is such that the revenues around $\underline{q} \geq 0$ remain finite. That is, $\lim_{q\to\underline{q}} s\left(q\right)<\infty$.*

We have the following lemma:

Lemma 1. *Suppose that Assumptions 1 and 3 hold, and consider those firms with sufficiently large φ. There exists a unique solution to the third-stage problem, i.e., $q^*\left(\varphi\right)$ is unique. Moreover, $q^*\left(\varphi\right)$ strictly increases in φ, $\lim_{\varphi\to\infty} q^*\left(\varphi\right)=\infty$, and $\lim_{\varphi\to\infty}\pi\left(\varphi\right)=\infty$.*

Proof. Applying (10), $q\frac{Q'}{Q}$ tends to zero and Q tends to a constant when $q\to\infty$. For a firm with an arbitrarily large φ, there exists a large q that satisfies (13) because the term in the brackets tends to a constant. However, there is a possibility that this firm with an arbitrarily large φ might choose a finite $q>\underline{q}$ such that the term $Q\cdot\left(1-\frac{1}{\sigma}+q\frac{Q'}{Q}\right)$ tends to zero. Nevertheless, plugging (13) into (1) entails $\pi\left(\varphi\right)=q^{1-\frac{1}{\sigma}}Q\left(\frac{1}{\sigma}-q\frac{Q'}{Q}\right)-\kappa_D$. Assumption 1 and (10) imply that when q becomes arbitrarily large as φ becomes arbitrarily large, then the profit also becomes arbitrarily large. However, if a finite $q>\underline{q}$ is chosen, then because this q is such that either $\frac{1}{\sigma}-q\frac{Q'}{Q}$ tends to one or Q tends to zero, the resulting profit must be finite. Let $\widehat{q}\left(\varphi\right)$ denote the solution that entails the largest profit among the solutions to (13) when φ is arbitrarily large. Thus, $\widehat{q}\left(\varphi\right)$ is unique and $\lim_{\varphi\to\infty}\widehat{q}\left(\varphi\right)=\infty$. As a result, when φ (and hence $\widehat{q}$) becomes arbitrarily large, the second-order condition (14) is satisfied because of (10). By applying the implicit function theorem on (13) and noting that $\widetilde{\pi}_{qq}\left(\widehat{q}\left(\varphi\right);\varphi\right)<0$, we have

$$\frac{d\widehat{q}}{d\varphi}=-\frac{\varphi^{-2}}{\widetilde{\pi}_{qq}\left(\widehat{q}\left(\varphi\right);\varphi\right)}>0. \tag{15}$$

Finally, the only concern with $\widehat{q}\left(\varphi\right)$ not being the profit-maximizing quantity is that it might be dominated by a corner solution at $\underline{q}$.[16] Suppose that $\underline{q}$ is an optimal solution that leads to an infinite profit. This means that $\lim_{q\to\underline{q}} s\left(q\right)=\infty$. Because this is implied by the demand structure and applies to all firms/varieties, this would be inconsistent with the equilibrium concept as it violates the consumers' budget constraint. This must be ruled out, and Assumption 3 serves this purpose. As a result, the possibility of a corner solution at $\underline{q}$ is ruled out. Hence, the optimal quantity $q^{*}\left(\varphi\right)$ is given by $\widehat{q}\left(\varphi\right)$ and is thus unique. □

As Assumption 3 is simply a restriction for equilibrium consistency, in this sense it is not restrictive. If $\underline{q}>0$ and Assumption 3 is violated, then $\lim_{q\to\underline{q}} D\left(q;A\right)=\infty$, i.e., $\underline{q}$ forms an asymptote of the demand curve and is indeed an optimal solution for firms. For example, for the Pollak demand given in Table 1, if $\widehat{a}\leq 0$ is changed to $\widehat{a}>0$, then $\underline{q}=\widehat{a}$ actually forms such an asymptote. When $\underline{q}=0$, it is possible that $\lim_{q\to 0} s\left(q\right)<\infty$ even when $\lim_{q\to 0} D\left(q;A\right)=\infty$ (such as the CES demand with $\sigma>1$). Hence, in this case, that $\underline{q}=0$ forms an asymptote is not an issue per se. This is why Assumption 3 is written in terms of sales rather than in terms of the inverse demand. For example, Assumption 3 rules out the CES demand with $\sigma\leq 1$. In sum, Assumption 3 rules out any demand defined on $\left[\underline{q},\infty\right)$ with $\underline{q}>0$ being an asymptote and any demand such that $\lim_{q\to 0} s\left(q\right)=\infty$.

In Stage 2, a firm chooses φ to maximize its profit. We consider the set of γ's which would choose sufficiently large φ. Then, applying Lemma 1 and the envelope theorem, the first- and second-order conditions of the second-stage problem can be written as

$$\gamma=\frac{q^{*}\left(\varphi\right)}{\varphi^{2}V'\left(\varphi\right)}. \tag{16}$$

[16] Note that it is impossible for a profit-maximizing quantity to be a finite $q_0>\underline{q}\geq 0$ because this would imply that $\lim_{q\to q_0} p\left(q\right)=\infty$, which violates the law of demand.

$$-2\varphi^{-3}q^*(\varphi)+\varphi^{-2}\frac{\partial q^*(\varphi)}{\partial\varphi}-\gamma V''(\varphi)<0. \tag{17}$$

Here, the innovation cost function must be sufficiently convex so that (17) holds. Lemma 2 below shows that the second-order condition holds for a large φ if and only if $\beta>\sigma-1$.

The question now becomes which γ's would choose large a φ. It is intuitive that a firm endowed with a higher R&D ability (smaller γ) innovates more and obtains a higher productivity; as γ tends to 0, the productivity tends to infinity. The following lemma formalizes this intuition:

Lemma 2. *Suppose that Assumptions 1–3 hold, and that $\beta>\sigma-1$. Consider those firms with sufficiently small γ. The optimal choice of φ exists and is unique. Such an optimal choice is denoted by $\varphi^*=\widetilde{\varphi}(\gamma)$. Moreover, φ^* is strictly decreasing in γ, and thus the inverse function exists and is denoted by $\widetilde{\gamma}(\varphi)$ and $\lim_{\varphi\to\infty}\widetilde{\gamma}(\varphi)=0$.*

Proof. Plugging (13) into (16), the first-order condition can be written as

$$\gamma=\frac{\left[Q(\varphi)\left(1-\frac{1}{\sigma}+q^*(\varphi)\frac{Q'(\varphi)}{Q(\varphi)}\right)\right]^\sigma}{L(\varphi)\left[\beta+\varphi\frac{L'(\varphi)}{L(\varphi)}\right]}\varphi^{-(\beta-\sigma+1)}. \tag{18}$$

Using (13–15), and (18), the left-hand side of (17) becomes

$$-Q^\sigma\left(1-\frac{1}{\sigma}+q\frac{Q'}{Q}\right)^\sigma\varphi^{\sigma-3}\left[2+\frac{1-\frac{1}{\sigma}+q\frac{Q'}{Q}}{-\frac{1}{\sigma}\left(1-\frac{1}{\sigma}\right)+2\left(1-\frac{1}{\sigma}\right)q\frac{Q'}{Q}+q^2\frac{Q''}{Q}}+\frac{\beta(\beta-1)+2\beta\varphi\frac{L'}{L}+\varphi^2\frac{L''}{L}}{\beta+\varphi\frac{L'}{L}}\right]. \tag{19}$$

Assumptions 1 and 3 imply that Lemma 1 holds. Assumptions 1–2, Lemma 1, and (10) imply that for large values of φ, (19) converges to $-C_Q^\sigma\left(\frac{\sigma-1}{\sigma}\right)^\sigma(\beta-\sigma+1)\lim_{\varphi\to\infty}\varphi^{\sigma-3}$, which is strictly negative if and only if $\beta>\sigma-1$. Thus, for a firm with an arbitrarily small γ there exists a large φ, denoted by φ^*, satisfying (18) and the second-order condition (17).

However, there is a possibility that this firm with an arbitrarily small γ might choose a finite φ such that either $Q(\varphi)\left(1-\frac{1}{\sigma}+q^*(\varphi)\frac{Q'(\varphi)}{Q(\varphi)}\right)$ tends to 0 or $\beta+\varphi\frac{L'(\varphi)}{L(\varphi)}$ tends to infinity so that (18) holds. Note that $V'>0$ would be violated if $\lim_{\varphi\to\varphi_0}L(\varphi)=\infty$ for some finite φ_0. With (13) and (18), (3) becomes

$$\Pi = Q^{\sigma} \cdot \left(1 - \frac{1}{\sigma} + q^{*}\frac{Q'}{Q}\right)^{\sigma} \left[\frac{\frac{1}{\sigma} - q^{*}\frac{Q'}{Q}}{1 - \frac{1}{\sigma} + q^{*}\frac{Q'}{Q}} - \frac{1}{\beta + \varphi\frac{L'}{L}}\right]\varphi^{\sigma-1} - \kappa_R - \kappa_D. \tag{20}$$

This implies that if $Q(\varphi)\left(1 - \frac{1}{\sigma} + q^{*}(\varphi)\frac{Q'(\varphi)}{Q(\varphi)}\right)$ tends to 0 or $\beta + \varphi\frac{L'(\varphi)}{L(\varphi)}$ tends to infinity at some finite φ, then the profit is finite. By contrast, the profit becomes arbitrarily large for an arbitrarily large φ. Thus, a finite φ would not be the solution to (18) when γ becomes arbitrarily small. Hence φ^{*} is the unique solution and is denoted by $\widetilde{\varphi}(\gamma)$.

For large values of φ, it is readily verified that (17) implies that the derivative of the right-hand side of (16) is negative. Hence, $\widetilde{\varphi}'(\gamma) < 0$ and the inverse function $\widetilde{\gamma}(\varphi)$ is well-defined. Obviously, $\lim_{\varphi\to\infty}\widetilde{\gamma}(\varphi) = 0$. □

As for the requirement for the inverse demand D for obtaining our main results, $\sigma > 1$ stated in Assumption 1 is not only sufficient but also necessary. To see this, first recall that $D(q;A) = q^{-\frac{1}{\sigma}}Q(q;A)$ and $\lim_{q\to\infty}Q(q;A) = C_Q > 0$. In this setting, $\sigma < 1$ is not permissible because this means that for firms with sufficiently small γ, the demand elasticity is less than 1, which is inconsistent with monopolistic competition. When $\sigma = 1$, it is possible that the demand elasticity remains greater than 1 for all q (or for the large q's that are relevant for our purposes) even though its limit is 1. However, this implies that firm sales approach a constant ($\lim_{q\to\infty} pq = \lim_{q\to\infty} q^{-\frac{1}{\sigma}}Q(q;A)q = C_Q$), which is inconsistent with the power law in firm size.

2.2.4 Power laws for productivity and firm size

We now show how the power laws for productivity and firm size arise. Note that Lemma 2 and (20) imply that $\lim_{\gamma\to 0}\Pi(\gamma) = \infty$. Thus, the set of γ's that survive must be non-empty, i.e., $\Pr(\gamma \in \Omega) > 0$. Observe that the p.d.f. of productivity is

$$g(\varphi) = \frac{f(\widetilde{\gamma}(\varphi))}{\Pr(\gamma \in \Omega)}\left|J(\varphi)\right|,$$

where $J(\varphi) = |\partial\widetilde{\gamma}(\varphi)/\partial\varphi|$ is the Jacobian. Appendix A.2 shows that

$$
\begin{aligned}
|J(\varphi)| = & \left|\frac{\partial}{\partial\varphi}\frac{q^*(\varphi)}{\varphi^2 V'(\varphi)}\right| \qquad (21)\\
= & \frac{Q^\sigma}{L}\frac{\left(1-\frac{1}{\sigma}+q^*\frac{Q'}{Q}\right)^\sigma}{\beta+\varphi\frac{L'}{L}} \cdot \left[2+\frac{\beta(\beta-1)+2\beta\varphi\frac{L'}{L}+\varphi^2\frac{L''}{L}}{\beta+\varphi\frac{L'}{L}}\right.\\
& \left. + \frac{1-\frac{1}{\sigma}+q^*\frac{Q'}{Q}}{-\frac{1}{\sigma}\left(1-\frac{1}{\sigma}\right)+2\left(1-\frac{1}{\sigma}\right)q^*\frac{Q'}{Q}+(q^*)^2\frac{Q''}{Q}}\right] \cdot \varphi^{-(\beta-\sigma+1)-1}.
\end{aligned}
$$

Proposition 1 is our main result.

Proposition 1. *Suppose that Assumptions 1–3 hold. Also suppose that* $f(\gamma) = \gamma^\alpha m(\gamma)$ *where* $\alpha > -1$ *and* $\lim_{\gamma\to 0} m(\gamma) = C_m$*, and that* $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$*. Then, in equilibrium both the productivity and firm size distributions exhibit power laws with tail indices* $(\alpha+1)(\beta-\sigma+1)$ *and* $\frac{(\alpha+1)(\beta-\sigma+1)}{\sigma-1}$*, respectively.*

Proof. We sketch the proof as follows; the detailed proof is relegated to Appendix A.2. Note that $\alpha > -1$ and $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$ ensure that $\beta > \sigma - 1$; hence, with Assumptions 1–3, Lemmas 1–2 hold. For the free-entry condition to hold, the expected profit must be finite, i.e., $\int_\Omega \Pi(\gamma)\, dF(\gamma) < \infty$. Whether this integral is finite depends on the firms with small γ, and what matters is essentially the orders of demand, σ, the innovation cost function, β, and the distribution of the failure probability around $\gamma = 0$. We show in Appendix A.2 that this is ensured when $\alpha > -1$ and $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$. Intuitively, the innovation cost function must be sufficiently convex. Observe (21). First note that Assumptions 1–2, Lemmas 1–2, and (10) imply that $Q(q; A)$ and $L(\varphi)$ converge to some constants C_Q and C_L, respectively, and that $q^*\frac{Q'}{Q}$, $\varphi\frac{L'}{L}$, $(q^*)^2\frac{Q''}{Q}$, and $\varphi^2\frac{L''}{L}$ all go to zero. Thus, $|J(\varphi)|$ converges to a power function of φ with exponent $-(\beta-\sigma+1)-1 < 0$. The assumption on $f(\widetilde{\gamma}(\varphi))$ allows us to write $g(\varphi) = \frac{\widetilde{\gamma}(\varphi)^\alpha m(\widetilde{\gamma}(\varphi))}{\Pr(\gamma\in\Omega)}|J(\varphi)|$, and (18) implies that $\widetilde{\gamma}(\varphi)^\alpha m(\widetilde{\gamma}(\varphi))$ converges to a power function of φ with exponent $-\alpha(\beta-\sigma+1)$. Thus, the productivity distribution exhibits a power law with a tail index $(\alpha+1)(\beta-\sigma+1)$. Following the same procedure, the firm size distribution also exhibits a power law with a tail index $\frac{(\alpha+1)(\beta-\sigma+1)}{\sigma-1}$. □

Proposition 1 establishes how power laws emerge from a general environment in a standard general-equilibrium model. It connects with the empirical regularity in firm size and provides a microfoundation for assuming power-law distributions in the theoretical literature, e.g., the Pareto, Fréchet, and the two-piece distribution in Nigai (2017). As mentioned, the class of APFs includes many widely-used non-CES and non-homothetic preferences for the inverse demand. For the innovation cost, it includes all polynomial functions that are increasing unboundedly when φ goes to infinity and are sufficiently convex so that $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$.

That the distribution $f(\gamma)$ is an asymptotic power function around 0 is also more general than it seems. This includes many well-known, widely-used distributions such as Beta (which subsumes the uniform), Gamma, F, and Weibull.[17] Table 2 provides a list of examples in this class.[18] Compared with Geerolf's (2017) power-law result, Proposition 1 is more general, as Geerolf's key condition is equivalent to $\lim_{\gamma\to 0} f(\gamma) = K > 0$ used in our illustrative example in Section 2.2.1 and is a special case here, i.e., $\alpha = 0$.[19]

Note that Proposition 1 as our main result and the main result in Mrázová et al. (2021) do not subsume each other. Mrázová et al. (2021) show that if two firm variables of interest are related by a power function, then the distribution of one variable following a "generalized power function" (GPF) implies that the other variable also follows the same GPF family with a transformation of parameters.[20] There is some overlap between GPF distributions and the distributions as APFs, but neither is a subset of the other. For example, the exponential distribution is an APF at the left tail as it is a special case of the Gamma

[17] For the distributions that are defined on $(0,\infty)$, proper truncation to the right is needed as the distribution of γ is on $(0,1]$.

[18] For 0 to be in the support of Generalized Pareto, we require $\mu \leq 0$ for $\xi \geq 0$ and $\mu \leq 0 \leq \mu - \frac{\sigma}{\xi}$ for $\xi < 0$. The parameters of other distributions are all positive.

[19] Proposition 1 shows that the environment for the power laws to emerge can be generalized by APFs, and this implies fewer constraints on the behaviors of firms that are not large. A relevant question is how well firm behaviors under a particular set of APFs approximate the behaviors of large and productive firms observed in the data. This depends on the speed at which the functions converge to power functions; certain empirical tests on the goodness of fit are warranted. This interesting empirical question is beyond the scope of the current paper and is left for future research.

[20] Note that their Proposition 1 actually allows the two variables y and z to be related by $y = y_0 h(z)^E$; y and z being related by a power function is a special case.

Distribution	$f(\gamma) \propto$	Distribution	$f(\gamma) \propto$
Beta	$\gamma^{\alpha}(1-\gamma)^{b-1}$	Kumaraswamy	$(1+\alpha)\, b\gamma^{\alpha}\left(1-\gamma^{1+\alpha}\right)^{b-1}$
Gamma	$\gamma^{\alpha}e^{-\frac{\gamma}{b}}$	Log-Logistic	$\left(\frac{1+\alpha}{b}\right)\left(\frac{\gamma}{b}\right)^{\alpha}\left[1+\left(\frac{\gamma}{b}\right)^{1+\alpha}\right]^{-2}$
F	$\gamma^{\alpha}\left[b+2(1+\alpha)\gamma\right]^{-\frac{2(1+\alpha)+b}{2}}$	Rayleigh	$\frac{\gamma}{a^2}e^{-\frac{\gamma^2}{2a^2}}$
Weibull	$\frac{(1+\alpha)}{b}\left(\frac{\gamma}{b}\right)^{\alpha}e^{-\left(\frac{\gamma}{b}\right)^{\alpha}}$	Generalized Pareto	$\frac{1}{\sigma}\left(1+\xi\frac{\gamma-\mu}{\sigma}\right)^{-\left(1+\frac{1}{\xi}\right)}$
Inverse Pareto	$b(1+\alpha)\gamma^{\alpha}$		

Table 2: Examples of distributions that are asymptotic power functions at the left tail

distribution with $\alpha = 0$. However, the exponential distribution is not a GPF, as mentioned in Appendix A.1 in their paper. On the other hand, the log-normal distribution is not an APF at the left tail, but it is a GPF. The Pareto distribution is a special case that is in both classes.

We now turn to a general open economy to investigate whether and how power laws hold in that environment.

2.3 Power Laws in an Open Economy

There are $n+1$ asymmetric countries indexed by $i \in \{0, 1, ..., n\}$ with the asymmetry in possibly every aspect of the model. Not only can all the trade cost, entry cost, and fixed cost of production parameters vary across countries, but the inverse demand function D_i, innovation cost function k_i, and the density function of failure probability f_i can all be country-specific (hence $\{\sigma_i, \beta_i, \alpha_i\}$ can also be country-specific). Similar to the closed-economy case, Assumptions 1–3 are assumed to hold with $C_{Q,i}$ and $C_{L,i}$ also allowed to be country-specific.

The timing is the same as in the closed economy, and in the production stage each firm can determine whether to export, and, if so, the price and quantity of exported goods. Let $\tau_{ij} \geq 1$ denote the iceberg trade cost. Assume strictly positive trade costs between any pair of countries, i.e., $\tau_{ij} > 1$ for all $i \neq j$. That is, countries are geographically segmented.

After paying the fixed cost of production $\kappa_{D,i}$, the profit of a firm located in country i obtained from selling to country j is

$$\pi_{ij}(\varphi) \equiv \max_{q_{ij}} \left[D_j(q_{ij}; A) - \frac{\tau_{ij} w_i}{\varphi} \right] q_{ij} - w_i \kappa_{ij}, \tag{22}$$

where κ_{ij} is the fixed selling cost, and w_i is the wage in country i. The second-stage problem is

$$\Pi_i(\gamma) \equiv \max_{\varphi, \{\mathbb{I}_{ij}\}} \sum_j \left[\pi_{ij}(\varphi) \mathbb{I}_{ij}(\gamma) \right] - w_i \kappa_{D,i} - w_i \gamma V_i(\varphi) - w_i \kappa_{R,i}, \tag{23}$$

where $\mathbb{I}_{ij} = \{0, 1\}$ is the indicator function that indicates whether the firm with γ in country i sells to country j.

Similar to the closed-economy case, let $q^*_{ij}(\varphi)$ and $\widetilde{\varphi}_i(\gamma)$ denote optimal solutions to (22) and (23), respectively. The set of surviving γ's is denoted by $\Omega_i \subseteq (0, 1]$. The free-entry condition is given by

$$E(\Pi_i) = \int_{\Omega_i} \Pi_i(\gamma) \, dF_i(\gamma) = w_i \kappa_{e,i}, \tag{24}$$

The goods market clearing condition can be written as

$$w_i N_i = \sum_j \left[M_{e,j} \int_{\Omega_j} D_i\left(q^*_{ji}(\widetilde{\varphi}_i(\gamma)); A \right) q^*_{ji}(\widetilde{\varphi}_i(\gamma)) \mathbb{I}_{ji}(\gamma) \, dF_j(\gamma) \right]. \tag{25}$$

The labor market clearing condition in country i is $N_i = N_{e,i} + N_{in,i} + N_{p,i}$,where $N_{e,i}$, $N_{in,i}$, and $N_{p,i}$ are the masses of labor employed for entry, innovation, and production, respectively:

$$\begin{aligned}
N_{e,i} =& M_{e,i} \kappa_{e,i} \\
N_{in,i} =& M_{e,i} \int_{\Omega_i} \gamma V_i(\widetilde{\varphi}_i(\gamma)) \, dF_i(\gamma) + M_{e,i} \kappa_{R,i} \Pr(\gamma \in \Omega_i) \\
N_{p,i} =& M_{e,i} \int_{\Omega_i} \sum_j \mathbb{I}_{ij}(\gamma) \left[\frac{\tau_{ij} q^*_{ij}(\widetilde{\varphi}_i(\gamma))}{\widetilde{\varphi}_i(\gamma)} + \kappa_{ij} \right] dF(\gamma) + M_{e,i} \kappa_{D,i} \Pr(\gamma \in \Omega_i).
\end{aligned}$$

An equilibrium consists of $\{w_i, A_i, M_{e,i}, \Omega_i\}_{i=0}^n$ such that $\Pi_i(\gamma) \geq 0$ if and only if $\gamma \in \Omega_i$, and $\{w_i, A_i, M_{e,i}\}$ are jointly determined by (24–25) and the labor market clearing condition for all $i \in \{0, 1, ..., n\}$.

The first-order condition for q_{ij} is similar to (13) and is given by

$$\varphi = w_i \tau_{ij} q_{ij}^{\frac{1}{\sigma_j}} \left[Q_j \times \left(1 - \frac{1}{\sigma_j} + q_{ij} \frac{Q_j'}{Q_j} \right) \right]^{-1}. \tag{26}$$

Assume that $\beta_i > \sigma_j - 1$ for all i and j. It is readily verified that parallels to Lemmas 1 and 2 both hold. Hence, for small γ and large φ, $q_{ij}^* (\varphi)$ and $\widetilde{\varphi}_i (\gamma)$ are monotonic functions exhibiting similar properties to those mentioned in the two lemmas. Let $\widetilde{\gamma}_i (\varphi)$ denote the inverse function of $\widetilde{\varphi}_i$. Similar to (16), the first-order condition on φ is

$$\gamma = \frac{\sum_j \mathbb{I}_{ij} \tau_{ij} q_{ij}^* (\varphi)}{\varphi^2 V_i' (\varphi)}. \tag{27}$$

Combining (26) with (27) yields

$$\gamma = \frac{\sum_j \mathbb{I}_{ij} \tau_{ij}^{1-\sigma_j} w_i^{-\sigma_j} Q_j^{\sigma_j} \cdot \left(1 - \frac{1}{\sigma_j} + q_{ij}^* \frac{Q_j'}{Q_j} \right)^{\sigma_j} \cdot \varphi^{\sigma_j - \beta_i - 1}}{L_i \cdot \left(\beta_i + \varphi \frac{L_i'}{L_i} \right)}. \tag{28}$$

By the parallels to Lemmas 1 and 2, when φ becomes arbitrarily large, the firm must sell to every market j because the fixed selling cost κ_{ij} is fixed while the gross profit also becomes arbitrarily large. Each component in the numerator of (28) is similar to those in the closed-economy case. Thus, for an arbitrarily small γ, there exists a corresponding large φ such that (28) holds with $\mathbb{I}_{ij} = 1$ for all j.

Appendix A.3.1 shows that if $\alpha_i > -1$ and $\beta_i > \frac{\alpha_i + 2}{\alpha_i + 1} (\max_j \sigma_j - 1)$, then the expected profit of entrants in each country remains finite. Since we are concerned with the tail behavior of the productivity distribution, it suffices to focus on the right-most piece of the productivity distribution. The corresponding Jacobian is obtained by differentiating (27):

$$|J_i (\varphi)| = \left| \underbrace{\frac{\partial \widetilde{\gamma}_i (\varphi)}{\partial \varphi}}_{-} \right| = - \sum_{j=0}^{n} \frac{\partial}{\partial \varphi} \frac{\tau_{ij} q_{ij}^* (\varphi)}{\varphi^2 V_i' (\varphi)}. \tag{29}$$

Obviously, each component of (29) is similar to (21), and converges to a power function of φ with exponent $-(\beta_i + 2 - \sigma_j)$. Following the same argument in Proposition 1, Appendix A.3.2 shows that the productivity distribution exhibits a power law with the tail index $(\alpha_i + 1)(\beta_i + 1 - \max_j \sigma_j)$.

Let s_{ij} denote a firm's sales from i to j; thus the size of a firm that exports to all countries is $s \equiv \sum_{j=0}^{n} s_{ij}$. Noting that $\frac{\partial s}{\partial \varphi} = \sum_{j=0}^{n} \frac{\partial s_{ij}}{\partial \varphi} = \sum_{j=0}^{n} \frac{\partial s_{ij}}{\partial q_{ij}^*} \frac{\partial q_{ij}^*}{\partial \varphi}$ and following a procedure similar to that in the proof of Proposition 1, Appendix A.3.3 shows that the firm size distribution also follows a power law with the tail index $\frac{(\alpha_i+1)(\beta_i+1-\max_j \sigma_j)}{\max_j \sigma_j - 1}$. Thus, we have the following proposition.

Proposition 2. *Suppose that Assumptions 1–3 hold. For all* $i \in \{0, 1, 2, ..., n\}$*, suppose that* $f_i(\gamma) = \gamma^{\alpha_i} m_i(\gamma)$ *where* $\alpha_i > -1$ *and* $\lim_{\gamma \to 0} m_i(\gamma) = C_{m,i}$*, and that* $\beta_i > \frac{\alpha_i+2}{\alpha_i+1}(\max_j \sigma_j - 1)$*. Then, the productivity distribution in each country* i *has a power law tail with a tail index of* $(\alpha_i + 1)(\beta_i + 1 - \max_j \sigma_j)$*, and the distribution of firm size has a power law tail with a tail index of* $\frac{(\alpha_i+1)(\beta_i+1-\max_j \sigma_j)}{\max_j \sigma_j - 1}$.[21]

The tail indices of both the productivity and firm size distributions in each country i are associated with its technology parameters α_i and β_i, as well as the largest price elasticity of demand among all destination countries $\max \sigma_j$. However, they are not affected by the level of trade cost or the size of the destination country. The key intuition hinges on distinguishing the scaling and shape parameters of the firm size distribution. The tail index is a shape parameter, which captures how firm sales change with underlying firm heterogeneity. The most productive firms of any country i export to all countries. Their sales in the most price-elastic country are most sensitive to changes in productivity, and thus these sales dominate the sales to other countries in the limit in determining the tail index. Obviously, the shape parameter α_i of the distribution $f_i(\gamma)$ matters in the tail index, as well as the innovation cost parameter β_i as it influences firms' productivity decisions. By contrast, the level of trade cost and the size of the destination country affect the scaling parameter of the firm size distribution but not the tail index.

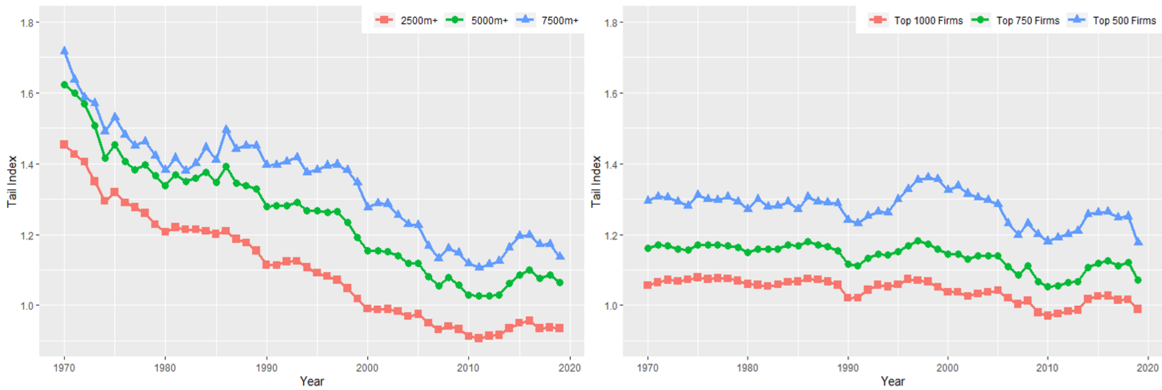
[21]The statement about tail indices here resembles the well-known theorem that the tail index of a sum of independent Pareto random variables is the minimum of the tail indices of these random variables. However, the different components of (29) are not literally independent random variables.

Proposition 2 implies that adding a country to the trade network (i.e., that country opens up to trade) causes the tails of both the productivity and firm-size distributions in each country to (weakly) become fatter. In the trade model by di Giovanni et al. (2011), productivity and firm size distributions are assumed to be Pareto and their tail indices are exogenous and are thus not affected by trade. They show that trade may cause the empirical estimates of tail indices to be lower than the true ones. However, Proposition 2 here implies a very different message from theirs because the true tail indices in our model can be affected by trade opening. Another closely related work is that by Bonfiglioli et al. (2019) who also show that trade fattens the tail of the firm size distribution. Both their model and ours predict that trade induces more innovation. However, despite the similarity in messages, the mechanisms for how trade fattens the tail index differ. In our model, trade opening fattens the tail index because of the addition of a more price-elastic market ($\max_j \sigma_j$). By contrast, their model assumes that innovating entrants make technological choices by choosing tail indices directly subject to a certain investment cost before drawing their productivity; trade amplifies the benefits of greater productivity and thus fatter tails are chosen.

2.4 Linking the Model to the Rise of Top Firms

To explore the links between our model and the firm size distribution in reality, we examine the changes in the tail index of firm size distribution over time using Compustat for publicly-listed firms in North America. Consistent with our model, a firm's size is measured by the real revenue of a firm;[22] the "right tail" is defined in terms of the size of real revenue and firm ranking. A tail index is estimated for each year during 1970–2019 and by the approach proposed by Gabaix and Ibragimov (2011). The left panel of Figure 1 shows the results for firms with real revenues above 2,500, 5,000, and 7,500 million dollars, whereas the right panel shows those for the top 500, 750, and 1,000 firms. Both panels show that the tail index generally declines over time, which is consistent with the recent

[22]Firms' revenues are deflated by the GDP deflator obtained from the U.S. Bureau of Economic Analysis.



Notes: The left panel depicts the tail index for each year during 1970–2019 for firms with real revenues above 2,500, 5,000, and 7,500 million dollars. The right panel depicts the tail index for the top 500, 750, and 1,000 firms.

Figure 1: Time Series of the Tail Index of the Firm Size Distribution

rise of top firms; see, e.g., Gabaix and Landier (2008), Autor et al. (2020) and De Loecker et al. (2020).[23]

Both Autor et al. (2020) and De Loecker et al. (2020) have shown that the rise of top firms is positively associated with these firms' innovation activities. However, they do not investigate the underlying causes for innovation. In the lens of this model, two potential reasons for the rise in innovation activities and top firms are as follows. First, the drastic improvements in information and communication technologies (ICT) since 1970 implies the reduction in the innovation-cost parameter β. As is well-known, improvements in ICT facilitate information flows and knowledge diffusion, which stimulate innovation activities and increase firm productivity and size. This effect is larger for larger firms. Second, globalization may cause firms to innovate more intensively. Proposition 2 predicts that the firm size distribution becomes fatter under trade compared with autarky as more competitive markets (namely, those with larger price elasticities) become accessible to top firms, prompting them to become more productive by innovating more intensively.[24]

[23]Our exercise of showing the declining trend complements these studies as they show the rise of top firms based on the markups, profit rates, market shares, or market values of these firms, but not the tail index.

[24]While our theory operates at the limit where the top firms export to every country regardless of the levels of trade costs, in reality not all of the top 1,000 firms sell to all countries. Our model can mimic such a more realistic scenario with a finite number of draws from the distribution of γ and with heterogeneous trade costs facing firms. Similarly, as trade costs are reduced, top firms export to more countries and are incentivized to

3 The Effect of Trade on the Productivity Distribution

This section analyzes the effects of trade on the productivity distribution. For tractability, we follow Melitz (2003) by assuming $n+1$ symmetric countries and CES demand in this and the next sections. In particular, for the welfare analysis in the next section, the CES demand is needed to be comparable with the ACR formula. Furthermore, for tractability and in both sections, we assume a simple function for the innovation cost: $k = \gamma\varphi^{\beta} + \kappa_R$. We allow the distribution of γ to be general until Section 4.2 where we need to generate a Pareto productivity distribution for comparison purposes. Given the functional-form assumptions regarding the inverse demand and innovation cost, Assumptions 1–2 are satisfied. The profit-maximizing solution of $q^*(\varphi)$ and $\widetilde{\varphi}(\gamma)$ must be interior and unique as given by the relevant first- and second-order conditions. Thus, Assumption 3 is no longer needed.

To solve the model, we start with the production stage. It is readily verified that

$$\begin{aligned}\pi_D(\varphi) =& \frac{N}{P^{1-\sigma}}\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\sigma-1}\varphi^{\sigma-1} - \kappa_D \\ \pi_X(\varphi) =& \tau^{1-\sigma}\frac{N}{P^{1-\sigma}}\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\sigma-1}\varphi^{\sigma-1} - \kappa_X.\end{aligned}$$

In the innovation stage, a firm decides its productivity level according to whether it serves the foreign market or not. The profits of a non-exporting firm and an exporting one are $\Pi_D = \pi_D(\varphi) - \gamma\varphi^{\beta} - \kappa_R$ and $\Pi_X = \pi_D(\varphi) + n\pi_X(\varphi) - \gamma\varphi^{\beta} - \kappa_R$, respectively, where the domestic market is denoted by subscript D and each foreign market is denoted by subscript X. The optimal productivities are

$$\widetilde{\varphi}(\gamma) = \begin{cases} \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{1}{\beta-\sigma+1}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{1}{\beta-\sigma+1}}\gamma^{-\frac{1}{\beta-\sigma+1}} & \text{for non-exporting firms} \\ \phi\left(\frac{N}{P^{1-\sigma}}\right)^{\frac{1}{\beta-\sigma+1}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{1}{\beta-\sigma+1}}\gamma^{-\frac{1}{\beta-\sigma+1}} & \text{for exporting firms}\end{cases}, \qquad (30)$$

where $\phi \equiv \left(1+n\tau^{1-\sigma}\right)^{\frac{1}{\beta-\sigma+1}}$.

become more productive by innovating more intensively.

Since exporting decisions are made after the firm has made its innovation decision, the firm chooses a higher productivity level if it plans to export afterward. The ratio ϕ can thus be interpreted as the *productivity advantages* of the exporting firms versus the non-exporting ones. As shown in the proof for Lemma 2, the second-order condition is satisfied if $\beta > \sigma - 1$, and the solution given by (30) is indeed optimal. The above leads to

$$\Pi_D(\gamma) = \left(\frac{N\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta P^{1-\sigma}} \right)^{\frac{\beta}{\beta-\sigma+1}} \frac{\beta-\sigma+1}{\sigma-1} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} - \kappa_D - \kappa_R \tag{31}$$

$$\Pi_X(\gamma) = \left(\frac{N\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta P^{1-\sigma}} \right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\beta\left(1+n\tau^{1-\sigma}\right)}{\sigma-1} \phi^{\sigma-1} - \phi^{\beta} \right] \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} - \kappa_D - \kappa_R - n\kappa_X. \tag{32}$$

Observe that the gross profits are proportional to $\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}$. The cutoffs are given by

$$\gamma_D = \left[\frac{1}{\kappa_D + \kappa_R} \left(\frac{N\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta P^{1-\sigma}} \right)^{\frac{\beta}{\beta-\sigma+1}} \frac{\beta-\sigma+1}{\sigma-1} \right]^{\frac{\beta-\sigma+1}{\sigma-1}} \tag{33}$$

$$\gamma_X = \left\{ \frac{1}{n\kappa_X} \left(\frac{N\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta P^{1-\sigma}} \right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\beta\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}}{\sigma-1} - \phi^{\beta} - \frac{\beta-\sigma+1}{\sigma-1} \right] \right\}^{\frac{\beta-\sigma+1}{\sigma-1}}, \tag{34}$$

such that $\Pi_D(\gamma) \geq 0$ if and only if $\gamma \leq \gamma_D$ and $\Pi_X(\gamma) \geq \Pi_D(\gamma)$ if and only if $\gamma \leq \gamma_X$. From (33) and (34), we have

$$\delta \equiv \frac{\gamma_X}{\gamma_D} = \left(\frac{\kappa_D + \kappa_R}{n\kappa_X} \right)^{\frac{\beta-\sigma+1}{\sigma-1}} \left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1 \right]^{\frac{\beta-\sigma+1}{\sigma-1}}. \tag{35}$$

If $\gamma_D \leq \gamma_X$, then all operating firms are exporters, which is counterfactual. Thus, similar to the literature, we consider only the case of $\gamma_X < \gamma_D$, i.e., $\delta < 1$, which requires trade frictions κ_X or τ to be sufficiently large relative to the fixed costs of production and innovation $\kappa_D + \kappa_R$.

The free-entry condition is $E(\Pi) = \int_0^{\gamma_X} \Pi_X(\gamma)\, dF(\gamma) + \int_{\gamma_X}^{\gamma_D} \Pi_D(\gamma)\, dF(\gamma) = \kappa_e$. An equilibrium is accordingly defined by (30), (33), (34), the free-entry condition, and the price index:

$$P^{1-\sigma} = M_e \left[\int_{\gamma_X}^{\gamma_D} \left(\frac{\sigma-1}{\sigma} \right)^{\sigma-1} \widetilde{\varphi}(\gamma)^{\sigma-1} dF(\gamma) + \int_0^{\gamma_X} \left(\frac{\sigma-1}{\sigma} \right)^{\sigma-1} \widetilde{\varphi}(\gamma)^{\sigma-1} dF(\gamma) \right] \quad (36)$$
$$+ nM_e \int_0^{\gamma_X} \tau^{1-\sigma} \left(\frac{\sigma-1}{\sigma} \right)^{\sigma-1} \widetilde{\varphi}(\gamma)^{\sigma-1} dF(\gamma),$$

where M_e denotes the mass of entrants paying the entry cost. The price index is composed of three terms. The first and second terms are associated with the prices charged by domestic non-exporting and exporting firms, respectively. The third term is associated with the imported goods. Note that by (30), there is a jump in the function $\widetilde{\varphi}(\gamma)$ at γ_X.

The following proposition establishes the unique existence of the equilibrium.

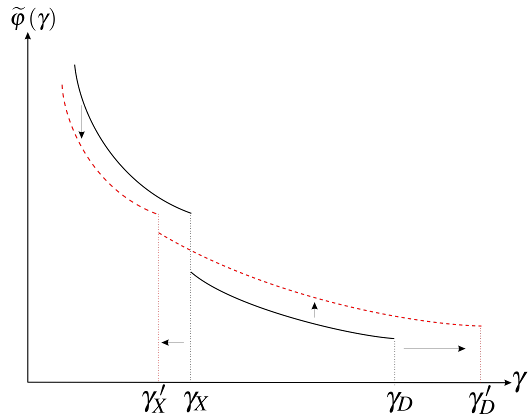
Proposition 3. *Suppose that $f(\gamma) = \gamma^\alpha m(\gamma)$ where $\alpha > -1$ and $m(\gamma)$ is slowly varying around the origin, $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$ and $\delta < 1$ where δ is defined by (35). Then, $E(\Pi)$ is a strictly increasing function of γ_D. If $\kappa_e \in (0, E(\Pi)|_{\gamma_D=1})$, then a unique equilibrium exists, and there are both exporters and non-exporters in the economy. In addition, the power laws in productivity and firm size hold.*

Proof. That the power laws hold follows immediately from Proposition 1. For the rest, see Appendix A.4. □

We explore how the iceberg cost τ affects the productivity distribution. The results are summarized in the following proposition, and are illustrated in Figure 2. The proof is relegated to Appendix A.5.

Proposition 4. *Assume that the conditions of Proposition 3 hold. An increase in τ results in a higher γ_D and a lower γ_X. Productivity φ increases (decreases) for any non-exporting (exporting) firm which remains non-exporting (exporting) after the shock. Productivity decreases for any firm which switches from exporting to non-exporting after the shock.*

To see the intuition behind how τ affects the selection and exporting cutoffs, first note that an increase in τ makes exporting more difficult so that firms must be more efficient



Notes: The black and solid (red and dashed) curve in this figure depicts the productivity level $\varphi(\gamma)$ before (after) the increase in trade cost τ. An increase in τ increases the selection cutoff γ_D and decreases the exporting cutoff γ_X.

Figure 2: Effect of Increasing Trade Cost τ

in innovation to become exporters. Therefore, the exporting cutoff γ_X decreases. Because having fewer exporters entails less import competition faced by the firms in the domestic market, the selection of firms becomes more lenient and the surviving cutoff γ_D increases.

Rearranging (33), we have $P \propto \gamma_D^{\frac{1}{\beta}}$. Thus, a higher γ_D induced by a higher τ raises the price index, which reflects the fact that differentiated goods are more expensive in units of labor when trade frictions are larger. Due to less import competition, for non-exporting firms which remain non-exporting, there is more incentive to acquire a higher productivity, as is evident by observing (30) and (31). For exporting firms which continue to export, their domestic profits also benefit from less import competition, but as their productivity advantage ϕ shrinks with greater trade friction, their effective market sizes may shrink (see [30] and [32]). The latter force dominates the former and hence their productivities are actually reduced. A lower γ_X implies that some firms switch from exporting to not exporting. For these firms, productivities decrease because of the loss of the foreign market.

In terms of the empirical evidence, as most empirical studies are on trade liberalization, it is easier to think of the case of a decreasing τ. That a decrease in τ results in tougher selection and an increase in the fraction of exporters is the same prediction as in the standard Melitz model; see Pavcnik (2002), Trefler (2004), and Bustos (2011) for related empirical evidence. Moreover, in response to trade liberalization, Bustos (2011) finds that both continuing and new exporters innovate more than non-exporters, and Lileeva and Trefler (2010) find that new exporters innovate more and experience higher productivity growth. These findings are consistent with the predictions in Proposition 4. However, whether non-exporters conduct less R&D in response to trade liberalization seems to be an open and interesting empirical question that could be an avenue for future research.

4 Welfare Gains from Trade

This section shows the properties of welfare gains from trade in our model and then carries out a corresponding quantitative analysis.

4.1 Welfare Formula and Trade Elasticity

Welfare in both our model and the ACR framework is measured by $W_j = w_j N_j / P_j$, and the trade elasticity is defined by $\varepsilon = \partial \ln (X_{ij}/X_{jj}) / \partial \ln \tau$ with $i \neq j$. In ACR, technology choice is also incorporated, and the choice is made simultaneously with production and sales. Our model is different from the ACR framework because innovation occurs after entry and before production and sales. Proposition 5 shows that the welfare gains from trade in our model still follow the local ACR formula with a variable trade elasticity.[25] The proof is relegated to Appendix A.6.

Proposition 5. *Suppose that the conditions of Proposition 3 hold. For a general distribution of* γ*,* $F(.)$*, the welfare gains from trade follow the local ACR formula:* $\frac{d \ln W}{d \ln \tau} = \frac{1}{\varepsilon} \frac{d \ln \lambda}{d \ln \tau} = -(1-\lambda)$.

It is readily verified with a numerical example that the trade elasticity, for which the formula is given in Appendix A.6, is variable in τ and depends on the distribution of γ. Section 3 shows that trade costs affect firm-level productivities, as well as the selection

[25] One can obtain the gains from trade under large changes in τ by integrating over the local formula.

and exporting cutoffs. In particular, the productivity schedule across firm types has a jump at the exporting cutoff γ_X, and trade costs affect the productivities of exporters and non-exporters in opposite ways. It can be shown that if $\kappa_X = 0$, in which case every surviving firm is an exporter, the trade elasticity equals $1-\sigma$. When all surviving firms are exporters, the productivity schedule no longer has a jump, and trade costs affect firm-level productivities in similar ways. As a result, trade costs affect trade flows only through the intensive margin in a multiplicative manner, and the trade elasticity becomes a constant.[26]

4.2 Quantitative Analysis of Welfare Gains from Trade

Here, we conduct a quantitative analysis of the welfare gains from trade. In particular, to assess the role of innovation quantitatively, we compare our model with the Melitz model with the Pareto productivity distribution (henceforth MP), as both our model and MP satisfy the (local) ACR formula, differing only in how the productivity distribution is generated. Formally, the density function of the productivity distribution in the MP model is denoted by $g^{MP}(\varphi) = \theta^{MP}\varphi^{-\theta^{MP}-1}$ where $\theta^{MP} > \sigma - 1$ is the tail index. The trade elasticity in MP is $\varepsilon^{MP} = -\theta^{MP}$. Note that in general the price index can be written as $P^{1-\sigma} = P_D^{1-\sigma} + nP_X^{1-\sigma}$, where $P_D^{1-\sigma}$ and $P_X^{1-\sigma}$ are the components of $P^{1-\sigma}$ in which the goods are from domestic and foreign firms, respectively. Thus, $\lambda \equiv P_D^{1-\sigma}/P^{1-\sigma}$. In MP,

$$\lambda^{MP} = \left[1 + n\tau^{1-\sigma}\left(\frac{\varphi_X}{\varphi_D}\right)^{\sigma-\theta^{MP}-1}\right]^{-1} = \left[1 + n\tau^{-\theta^{MP}}\left(\frac{\kappa_X}{\kappa_D}\right)^{\frac{\sigma-\theta^{MP}-1}{\sigma-1}}\right]^{-1}.$$

As mentioned, since $\varepsilon = d\ln\left(\frac{1-\lambda}{n\lambda}\right)/d\ln\tau$ under the symmetric country setting,

$$\frac{d\ln W^{MP}}{d\ln\tau} = \frac{1}{\varepsilon^{MP}}\frac{d\ln\lambda^{MP}}{d\ln\tau} = -\left(1-\lambda^{MP}\right).$$

We now turn to our model, which is hereafter referred to as IN (innovation). To isolate

[26] As is evident from ACR's proof, a constant trade elasticity is a strong restriction. There could be various reasons as to why trade elasticity becomes a variable. For example, Melitz and Redding (2015) show that trade elasticity becomes a variable in the Melitz (2003) model when the productivity distribution deviates from the Pareto; Hsu et al. (2020a) show that it becomes a variable in the Bertrand competition model in Bernard et al. (2003) when the productivity draw is from the log-normal distribution, instead of the Fréchet.

the effect of innovation, we assume that γ is uniformly distributed so that the resulting productivity distribution is similar to the Pareto distribution with the tail index $\theta \equiv \beta - \sigma + 1$, except that there is a jump at γ_X when $\kappa_X > 0$. The domestic expenditure share in our model is

$$\lambda = \frac{1 + \left[\phi^{\sigma-1} - 1\right]\left(\frac{\gamma_X}{\gamma_D}\right)^{1-\frac{\sigma-1}{\theta}}}{1 + \left[\left(1 + n\tau^{1-\sigma}\right)\phi^{\sigma-1} - 1\right]\left(\frac{\gamma_X}{\gamma_D}\right)^{1-\frac{\sigma-1}{\theta}}}, \tag{37}$$

where γ_X/γ_D is given by (35).

To quantify the model, we calibrate the values of σ, β, n, τ, and $\kappa_X/(\kappa_D + \kappa_R)$. We calibrate these parameters from the viewpoint of the US in 2002. Feenstra and Weinstein (2017) report that the median of markups in the US is 1.3. Taking this median as a representative value for our constant-markup model, $\sigma \approx 4.33$. Under the uniform distribution of γ, δ (see [35]) is the fraction of exporters among all (surviving) firms. As documented by Bernard et al. (2007), this fraction in the US in 2002 equals 0.18.

Denote domestic absorption and imports as DA and M. By definition, λ then equals $(DA - M)/DA$. Using data from Penn World Table 9.0 (PWT 9.0), λ is 0.853 in 2002 for the US.[27] To better fit our symmetric-country model, the number of countries, $n + 1$, is computed as the ratio of the world GDP to that of the US. Using PWT 9.0, this number equals 4.41. We therefore set $n = 3$. We adopt the estimate of the trade elasticity in Simonovska and Waugh (2014), which is 4.63.[28] Thus, $\theta^{MP} = 4.63$. We calibrate β, $\kappa_X/(\kappa_D + \kappa_R)$, and τ to match $\lambda = 0.853$, $\delta = 0.18$, and $\varepsilon = 4.63$ using (35), (37), and the ε formula given in Appendix A.6. The result is $\beta = 7.838$, $\kappa_X/(\kappa_D + \kappa_R) = 0.572$, and $\tau = 2.097$. This implies that the tail index $\theta = 4.51$, which is rather close to θ^{MP}.

Given the calibrated parameters, we compute the local welfare gains for both the IN and MP models.[29] We also compare the welfare gains by moving from autarky to the

[27] We also use the US's Input-Output Table (obtained from OECD-IOT) as our alternative data set to compute λ. We compute DA by subtracting the net exports from the total value-added across industries. With this alternative data set, λ equals 0.862 and is similar to that computed with PWT 9.0.

[28] See their Table 7.

[29] It should be clear at this point that the fixed costs of innovation and production are isomorphic in our

current level of trade cost τ for both models. As the global ACR formula applies to the MP model, $\frac{W^{MP}}{W^{MP}_{\tau\to\infty}} = \left(\lambda^{MP}\right)^{-\frac{1}{\theta^{MP}}}$. The global ACR formula does not apply to our IN model, but combining (33) and (35) into the free-entry condition yields

$$\frac{W}{W_{\tau\to\infty}} = \frac{P_{\tau\to\infty}}{P} = \left\{1 + n^{1-\frac{\theta}{\sigma-1}} \left(\frac{\kappa_X}{\kappa_D + \kappa_R}\right)^{1-\frac{\theta}{\sigma-1}} \left[\left(1 + n\tau^{1-\sigma}\right)^{\frac{\beta}{\theta}} - 1\right]^{\frac{\theta}{\sigma-1}}\right\}^{\frac{1}{\beta}}.$$

From autarky to the calibrated τ, IN and MP entail 3.5% and 2.5% of the welfare gains, respectively. Hence, the gains from trade in IN are 40% larger than those in MP. The welfare elasticities to trade cost, $d\ln W/d\ln\tau$, are -0.147 and -0.108, for IN and MP, respectively. This implies that for small changes in τ, the welfare gains in IN are 36.1% higher than those in MP; this is quite similar to the case that compares with autarky.

The welfare elasticity, which is given by $1-\lambda$ as in Proposition 5, increases when τ decreases in both the IN and MP models. The gains from trade (compared with autarky) are simply the integral of the welfare elasticity from $\tau\to\infty$ to the current τ. The IN model entails higher gains from trade than the MP model because $\lambda < \lambda^{MP}$ at every value of τ. To see this, recall from Proposition 4 that a reduction in trade cost induces exporters to innovate more and become more productive and non-exporters to innovate less and become less productive. Hence, the productivity advantage of exporters vs. non-exporters widens with trade liberalization at a higher rate than the MP model. In autarky, the domestic expenditure $\lambda = 1$ in both models; once a country opens to trade, the rate of decrease in λ is larger in the IN model, implying that $\lambda < \lambda^{MP}$ at every value of $\tau < \infty$.

5 Conclusion

This paper has demonstrated that with an innovation stage added to a standard general equilibrium model of trade, power laws for both productivity and firm size could emerge in a rather general environment. The conditions placed on the inverse demand, innova-

model, as $\kappa_D + \kappa_R$ is what determines the selection cutoff. Correspondingly, as there is no innovation stage in the MP model, κ_D is what determines the selection cutoff. Thus we interpret the calibrated $\kappa_X/(\kappa_D+\kappa_R)$ as the κ_X/κ_D in the MP model.

tion cost, and underlying firm heterogeneity being APFs are all more general than it may seem: the demand class includes various non-CES and non-homothetic preferences; the class of innovation cost functions includes all polynomial functions that are sufficiently convex and increasing unboundedly when productivity goes to infinity; the density of firm heterogeneity includes many well-known, widely-used distributions. All of these results hold in a very general open economy in which all parameters can be country-specific and all bilateral trade costs can be country-pair-specific. Our approach can generally be applied to other topics that concern agent heterogeneity and power laws, as discussed in the Introduction.

Conditional on the same trade elasticity and values of the common parameters, quantitatively our model yields 40% higher welfare gains from trade than the Melitz–Pareto model. This suggests the importance of incorporating innovation in a trade model because innovation naturally reacts to changes in trade cost. The economics is fundamentally a market-size effect that works differently for exporters and non-exporters.

Welfare gains from trade critically depend on the tail indices of the power laws, which depend on the price elasticities and how costly it is to engages in innovation. Interestingly, trade plays an important role because the market with the greatest competitiveness (largest price elasticities) dominates and determines the tail index. This provides an important angle to comprehend trade wars. For example, the Trump administration's sharp increase in tariffs against Chinese products, regardless of whether it benefits or hurts the US or global economy, will certainly have a strong negative impact on the Chinese aggregate economy and welfare because the US tends to be the largest and most competitive market, and thus affects the top Chinese firms the most.

References

Acemoglu, D., and D. Cao (2015), "Innovation by Entrants and Incumbents," *Journal of Economic Theory*, 157, 255–294.

Aghion, P, A. Bergeaud, M. Lequien, and M. J. Melitz (2019), "The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports," Harvard University manuscript.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012), "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2019), "The Elusive Pro-competitive Effects of Trade," *Review of Economic Studies*, 86, 46–80.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020), "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 135, 645–709.

Axtell, R. L. (2001), "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818–20.

Bas, M., and I. Ledezma (2015), "Trade Liberalization and Heterogeneous Technology Investments," *Review of International Economics*, 23, 738–781.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003), "Plants and Productivity in International Trade," *American Economic Review*, 93, 1268–90.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2007), "Firms in International Trade," *Journal of Economic Perspectives*, 21, 105–130.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989), *Regular Variation*, Cambridge University Press.

Bonfiglioli, A., R. Crinò, and G. Gancia (2019), "Trade, Finance, and Endogenous Firm Heterogeneity," *Journal of the European Economic Association*, 17, 79–130.

Bustos, P. (2011), "Multilateral Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms," *American Economic Review*, 101, 304–340.

Chaney, T. (2008), "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98, 1707–21.

Chaney, T. (2014), "The Network Structure of International Trade," *American Economic Review*, 104, 3600–34.

Chaney, T. (2018), "The Gravity Equation in International Trade: An Explanation," *Journal of Political Economy*, 126, 150–177.

Costinot, A., & A. Rodríguez-Clare (2015), "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in E. Helpman, K. Rogoff, & G. Gopinath (Eds.), *Handbook of International Economics* (Vol. 4). North-Holland: Elsevier.

De Loecker, J., J. Eeckhout, and G. Unger (2020), "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 135, 561–644.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018), "The Geography of Development," *Journal of Political Economy*, 126, 903–983.

di Giovanni, J., A. A. Levchenko, and R. Rancière (2011), "Power Laws in Firm Size and Openness to Trade: Measurement and Implications," *Journal of International Economics*, 85, 42–52.

Eaton, J., and S. Kortum (2002), "Technology, Geography, and Trade," *Econometrica*, 70, 1741–79.

Feenstra, R. C. and D. E. Weinstein (2017), "Globalization, Markups, and US Welfare," *Journal of Political Economy*, 125, 1040–74.

Gabaix, X. (2009), "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–294.

Gabaix, X. (2011), "The Granular Origins of Aggregate Fluctuation," *Econometrica*, 79, 733–772.

Gabaix, X., and R. Ibragimov (2011), "Rank - 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business and Economic Statistics*, 29, 24–39.

Gabaix, X., and A. Landier (2008), "Why Has CEO Pay Increased so Much?" *Quarterly Journal of Economics*, 123, 49–100.

Garicano, L. (2000), "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108, 874–904.

Geerolf, F. (2017), "A Theory of Pareto Distributions," UCLA manuscript.

Hsieh, C.-T., P. J. Klenow, and I. B. Nath (2021), "A Global View of Creative Destruction," University of Chicago manuscript.

Hsu, W.-T., Y. Lu, and G. L. Wu (2020a), "Competition, Markups, and Gains from Trade: A Quantitative Analysis of China Between 1995 and 2004," *Journal of International Economics*, 122, 103266.

Hsu, W.-T., R. G. Riezman, and P. Wang (2020b), "Innovation, Growth, and Dynamic Gains from Trade," NBER Working Paper No. 26470.

Impullitti, G., and O. Licandro (2018), "Trade, Firm Selection and Innovation: The Competition Channel," *Economic Journal*, 128(608), 189–229.

König, M. D., J. Lorenz, and F. Zilibotti (2016), "Innovation vs. Imitation and the Evolution of Productivity Distributions," *Theoretical Economics*, 11, 1053–1102.

Lileeva, A., and D. Trefler (2010), "Improved Access to Foreign Markets Raises Plant-level Productivity... For Some Plants," *Quarterly Journal of Economics*, 125, 1051–99.

Luttmer, E. G. (2007), "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, 122, 1103–44.

Melitz, M. J. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

Melitz, M. J. and G. I. Ottaviano (2008), "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75, 295–316.

Melitz, M. J., and S. J. Redding (2015), "New Trade Models, New Welfare Implications," *American Economic Review*, 105, 1105–46.

Mrázová, M., and J. P. Neary (2017), "Not So Demanding: Demand Structure and Firm Behavior," *American Economic Review*, 107, 3835–74.

Mrázová, M., J. P. Neary, and M. Parenti (2021), "Sales and Markup Dispersion: Theory and Empirics," *Econometrica*, 89, 1753–88.

Nigai, S. (2017), "A Tale of Two Tails: Productivity Distribution and the Gains from Trade," *Journal of International Economics*, 104, 44–62.

Parenti, M., P. Ushchev, and J. F. Thisse (2017), "Toward a Theory of Monopolistic Competition," *Journal of Economic Theory*, 167, 86–115.

Pavcnik, N. (2002), "Trade Liberalization, Exit and Productivity Improvement: Evidence from Chilean Plants," *Review of Economic Studies*, 69, 245–276.

Sampson, T. (2016), "Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth," *Quarterly Journal of Economics,* 131, 315–380.

Simonovska, I., and M. E. Waugh (2014), "Trade Models, Trade Elasticities, and the Gains from Trade," NBER Working Paper No. 20495.

Sornette, D. (2006), *Critical Phenomena in Natural Sciences - Chaos, Fractals, Self-Organization and Disorder*, Springer Verlag.

Trefler, D. (2004), "The Long and Short of the Canada-U.S. Free Trade Agreement," *American Economic Review*, 94, 870–895.

Yeaple, S. (2005), "A Simple Model of Firm Heterogeneity, International Trade, and Wages," *Journal of International Economics*, 65, 1–20.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012), "Monopolistic Competition in General Equilibrium: Beyond the Constant Elasticity of Substitution," *Econometrica*, 80, 2765–84.

A Appendix

A.1 Microfoundation for Innovation Cost Function

Each firm determines its productivity level by engaging in R&D activities in the following manner. The production process involves a continuum of procedures, and the firm chooses the size of the continuum, a. How well the firm performs in each procedure (which we term the quality of the procedure) depends on the outcome of a sequence of experiments that the firm conducts. For each procedure, every firm starts off with one quality unit. When the first experiment is successful, then the firm obtains one additional quality unit for this procedure, and can continue to conduct the second experiment. Recursively, every successful experiment results in one additional quality unit and the chance to conduct the next experiment. If the experiment fails, however, no more experiments will be performed and the quality of the procedure is finalized. Firms differ in their probabilities of failure, $\gamma \in (0,1]$. The probability of obtaining quality $y = 1, 2, ...$ for a procedure is therefore $(1-\gamma)^{y-1}\gamma$, i.e., y is geometrically distributed.

Each procedure requires a worker, say, a research assistant, to perform the experiments. Therefore, the mass of research assistants employed by the firm equals the size of the continuum of procedures, a. The productivity φ is a function of the total quality of all a procedures, $aE(y)$: $\varphi \equiv B(aE(y)) = B\left(a\sum_{y=1}^{\infty}(1-\gamma)^{y-1}\gamma y\right) = B\left(\frac{a}{\gamma}\right)$. The function $B(\cdot)$ is strictly increasing and concave on $\mathbb{R}_+$, and $\lim_{a\to\infty} B\left(\frac{a}{\gamma}\right) = \infty$. The concavity of $B(\cdot)$ reflects the management burden for the firm to manage these research assistants. Inverting the above equation yields $k = \gamma V(\varphi) + \kappa_R$, where $V \equiv B^{-1}$ is strictly increasing and convex in φ and $\lim_{\varphi\to\infty} V(\varphi) = \infty$. For a firm to perform all these experiments, a fixed cost, κ_R, such as setting up a laboratory is required. Thus, the total labor requirement for a firm to acquire productivity φ is $k = \gamma V(\varphi) + \kappa_R$.

A.2 Proof of Proposition 1

We prove this proposition in three steps. In the first step, we show that $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$ must hold for the expected profit to be finite to ensure equilibrium existence. In the second and third steps we show that the productivity and firm size distributions exhibit power laws.

Step 1: We require $\int_\Omega \Pi(\gamma)\,dF(\gamma) < \infty$ for the free-entry condition to be well-defined. Since $\Pi(\gamma)$ is finite for all $\gamma > 0$, the only concern for the expected profit to explode is when γ is close to 0. Note that using (13) and (18) we can write

$$\Pi(\gamma) = \left(\frac{1}{L}\frac{1}{\beta+\varphi\frac{L'}{L}}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}\left(1-\frac{1}{\sigma}+q^*\frac{Q'}{Q}\right)^{\frac{(1+\beta)(\sigma-1)}{\beta-\sigma+1}} Q^{\frac{\beta\sigma}{\beta-\sigma+1}}\left(\frac{1}{\sigma}-q^*\frac{Q'}{Q}-\frac{1-\frac{1}{\sigma}+q^*\frac{Q'}{Q}}{\beta+\varphi\frac{L'}{L}}\right)\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}-\kappa_R-\kappa_D.$$

The expected profit is finite if $\int_\Omega [\Pi(\gamma)+\kappa_R+\kappa_D]\,\gamma^\alpha m(\gamma)\,d\gamma < \infty$. Assumptions 1–2 and $\lim_{\gamma\to 0} m(\gamma) = C_m$ imply that

$$\lim_{\gamma\to 0} \frac{[\Pi(\gamma) + \kappa_R + \kappa_D]}{\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}} m(\gamma) = \frac{C_m C_Q^{\frac{\sigma\beta}{\beta-\sigma+1}} \left(\frac{\sigma-1}{\sigma}\right)^{\frac{\sigma\beta}{\beta-\sigma+1}}}{C_L^{\frac{\sigma-1}{\beta-\sigma+1}} \beta^{\frac{\beta}{\beta-\sigma+1}}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right);$$

hence for any $\omega > 0$ there exists a $\overline{\gamma} > 0$ such that for any $\gamma < \overline{\gamma}$,

$$\frac{[\Pi(\gamma) + \kappa_R + \kappa_D]}{\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}} < \frac{C_m C_Q^{\frac{\sigma\beta}{\beta-\sigma+1}} \left(\frac{\sigma-1}{\sigma}\right)^{\frac{\sigma\beta}{\beta-\sigma+1}}}{C_L^{\frac{\sigma-1}{\beta-\sigma+1}} \beta^{\frac{\beta}{\beta-\sigma+1}}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right) + \omega.$$

By picking a sufficiently small $\overline{\gamma}$ and noting that $\overline{\gamma} < 1$, we have

$$\int_0^{\overline{\gamma}} [\Pi(\gamma) + \kappa_R + \kappa_D] f(\gamma) d\gamma < \left[\frac{C_m C_Q^{\frac{\sigma\beta}{\beta-\sigma+1}} \left(\frac{\sigma-1}{\sigma}\right)^{\frac{\sigma\beta}{\beta-\sigma+1}}}{C_L^{\frac{\sigma-1}{\beta-\sigma+1}} \beta^{\frac{\beta}{\beta-\sigma+1}}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right) + \omega\right] \int_0^1 \gamma^{\alpha-\frac{\sigma-1}{\beta-\sigma+1}} d\gamma.$$

It follows that $\int_0^1 \gamma^{\alpha-\frac{\sigma-1}{\beta-\sigma+1}} d\gamma < \infty$ if $\alpha > -1$ and $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$.

Step 2: Starting from the definition of the Jacobian and using (15), we have

$$|J(\varphi)| = \left|\frac{\partial\widetilde{\gamma}(\varphi)}{\partial\varphi}\right| = \frac{q^*(\varphi)}{\varphi^2 V'(\varphi)} \left(2\varphi^{-1} + \frac{V''(\varphi)}{V'(\varphi)}\right) + \frac{1}{\varphi^2 V'(\varphi)} \frac{\varphi^{-2}}{\widetilde{\pi}_{qq}(q^*(\varphi);\varphi)}.$$

Then, by (13), (14), and Assumptions 1–2 we can replace $V'(\varphi)$, $V''(\varphi)$, $q^*(\varphi)$, and $\widetilde{\pi}_{qq}(q^*(\varphi);\varphi)$ to obtain (21) as

$$\begin{aligned} |J(\varphi)| =& \frac{Q^\sigma}{L} \frac{\left(1 - \frac{1}{\sigma} + q^* \frac{Q'}{Q}\right)^\sigma}{\beta + \varphi \frac{L'}{L}} \cdot \left[2 + \frac{\beta(\beta-1) + 2\beta\varphi\frac{L'}{L} + \varphi^2 \frac{L''}{L}}{\beta + \varphi\frac{L'}{L}}\right. \\ &\left. + \frac{1 - \frac{1}{\sigma} + q^* \frac{Q'}{Q}}{-\frac{1}{\sigma}\left(1 - \frac{1}{\sigma}\right) + 2\left(1 - \frac{1}{\sigma}\right) q^* \frac{Q'}{Q} + (q^*)^2 \frac{Q''}{Q}}\right] \cdot \varphi^{-(\beta-\sigma+1)-1}. \end{aligned}$$

By Assumptions 1–2, Lemmas 1 and 2, and (10), we have

$$\lim_{\varphi\to\infty} \frac{|J(\varphi)|}{\varphi^{-(\beta-\sigma+1)-1}} = \frac{C_Q^\sigma \left(\frac{\sigma-1}{\sigma}\right)^\sigma (\beta-\sigma+1)}{C_L \beta}.$$

The p.d.f. of productivity is therefore

$$g(\varphi) = \frac{f(\widetilde{\gamma}(\varphi))}{\Pr(\gamma\in\Omega)} |J(\varphi)| = \frac{f(\widetilde{\gamma}(\varphi))}{\Pr(\gamma\in\Omega)} \frac{|J(\varphi)|}{\varphi^{-(\beta-\sigma+1)-1}} \varphi^{-(\beta-\sigma+1)-1}$$

$$=\frac{m\left(\widetilde{\gamma}\left(\varphi\right)\right)}{\Pr\left(\gamma\in\Omega\right)}\left[\frac{Q^{\sigma}}{L}\frac{\left(1-\frac{1}{\sigma}+q\frac{Q'}{Q}\right)^{\sigma}}{\beta+\varphi\frac{L'}{L}}\right]^{\alpha}\frac{\left|J\left(\varphi\right)\right|}{\varphi^{-(\beta-\sigma+1)-1}}\varphi^{-(1+\alpha)(\beta-\sigma+1)-1}.$$

As φ becomes arbitrarily large, $m\left(\widetilde{\gamma}\left(\varphi\right)\right)$, the bracketed term, and $\frac{|J(\varphi)|}{\varphi^{-(\beta-\sigma+1)-1}}$ converge to constants. It thus follows that the productivity distribution exhibits a power law with a tail index $(1+\alpha)(\beta-\sigma+1)$.

Step 3: By (13) and Lemma 1, firm size in terms of sales s is a function of φ:

$$s=\varphi^{\sigma-1}Q^{\sigma}\left(1-\frac{1}{\sigma}+q^{*}\frac{Q'}{Q}\right)^{\sigma-1}. \tag{38}$$

By Assumptions 1–2 and Lemmas 1–2, there are one-to-one mappings at the tails between $s\to\infty$ and $\varphi\to\infty$, and between $\varphi\to\infty$ and $\gamma\to 0$, such that $\lim_{\varphi\to\infty}s=\infty$. Let $s\left(\varphi\right)$ denote the firm size with productivity φ as defined by (38); $\varphi\left(s\right)$ denotes its inverse function. Combining (18) and (38), we have

$$\widetilde{\gamma}\left(\varphi\left(s\right)\right)\equiv\widetilde{\gamma}\left(s\right)=\gamma=\frac{Q^{\sigma}}{L}\frac{\left(1-\frac{1}{\sigma}+q\frac{Q'}{Q}\right)^{\sigma}}{\beta+\varphi\frac{L'}{L}}\left[Q^{\sigma}\left(1-\frac{1}{\sigma}+q\frac{Q'}{Q}\right)^{\sigma-1}\right]^{\frac{\beta-\sigma+1}{\sigma-1}}s^{-\frac{\beta-\sigma+1}{\sigma-1}},$$

which converges to a power function of s with exponent $-\frac{\beta-\sigma+1}{\sigma-1}$ under Assumptions 1–2.

Using (13) and (14), we have

$$\frac{\partial s\left(\varphi\right)}{\partial\varphi}=\frac{\partial s}{\partial q^{*}}\frac{\partial q^{*}}{\partial\varphi}=\frac{1-\frac{1}{\sigma}+q^{*}\frac{Q'}{Q}}{\frac{1}{\sigma}\left(1-\frac{1}{\sigma}\right)-2\left(1-\frac{1}{\sigma}\right)q^{*}\frac{Q'}{Q}-\left(q^{*}\right)^{2}\frac{Q''}{Q}}q^{*}\varphi^{-2}>0. \tag{39}$$

Using (13), (38), and (39), we obtain the Jacobian $\left|J_{s}\left(s\right)\right|$:

$$\begin{aligned}\left|J_{s}\left(s\right)\right|=\left|\frac{\partial\widetilde{\gamma}\left(s\right)}{\partial s}\right|&=\left|\frac{\partial\widetilde{\gamma}\left(\varphi\right)}{\partial\varphi}\frac{\partial\varphi\left(s\right)}{\partial s}\right|=\frac{\left|J\left(\varphi\right)\right|}{\varphi^{-(\beta-\sigma+1)-1}}\varphi^{-(\beta-\sigma+1)-1}\left(\frac{\partial s\left(\varphi\right)}{\partial\varphi}\right)^{-1}\\&=\frac{\left|J\left(\varphi\right)\right|}{\varphi^{-(\beta-\sigma+1)-1}}\frac{\frac{1}{\sigma}\left(1-\frac{1}{\sigma}\right)-2\left(1-\frac{1}{\sigma}\right)q^{*}\frac{Q'}{Q}-\left(q^{*}\right)^{2}\frac{Q''}{Q}}{\left(1-\frac{1}{\sigma}+q^{*}\frac{Q'}{Q}\right)^{-(\beta-\sigma+1)}Q^{-\frac{\sigma(\beta-\sigma+1)}{\sigma-1}}}s^{-\frac{\beta-\sigma+1}{\sigma-1}-1}.\end{aligned}$$

The density of the firm size distribution $g_{s}\left(s\right)$ can be expressed as

$$g_{s}\left(s\right)=\frac{m\left(\widetilde{\gamma}\left(s\right)\right)\left[\widetilde{\gamma}\left(s\right)\right]^{\alpha}}{\Pr\left(\gamma\in\Omega\right)}\left|J_{s}\left(s\right)\right|=\frac{m\left(\widetilde{\gamma}\left(s\right)\right)}{\Pr\left(\gamma\in\Omega\right)}\frac{\left|J_{s}\left(s\right)\right|}{s^{-\frac{\beta-\sigma+1}{\sigma-1}-1}}\left(\frac{\widetilde{\gamma}\left(s\right)}{s^{-\frac{\beta-\sigma+1}{\sigma-1}}}\right)^{\alpha}s^{-\frac{(\alpha+1)(\beta-\sigma+1)}{\sigma-1}-1}.$$

By Assumptions 1–2, Lemmas 1–2, and (10), we know that $|J_s(s)|/s^{-\frac{\beta-\sigma+1}{\sigma-1}-1}$, $\widetilde{\gamma}(s)^{\alpha}/s^{-\alpha\frac{\beta-\sigma+1}{\sigma-1}}$, and $m(\widetilde{\gamma}(s))$ converge to constants as s tends to infinity. Therefore, the firm size distribution exhibits a power law with a tail index $\frac{(\alpha+1)(\beta-\sigma+1)}{\sigma-1}$. □

A.3 Derivation for Section 2.3

A.3.1 Finiteness of Expected Profit

Let $\bar{k} \equiv \arg\max_k \sigma_k$; we can rewrite (28) by extracting $\varphi^{\sigma_{\bar{k}}-\beta_i-1}$ as

$$\gamma = \left[\sum_j \frac{\mathbb{I}_{ij}\tau_{ij}^{1-\sigma_j} w_i^{-\sigma_j} Q_j^{\sigma_j} \cdot \left(1 - \frac{1}{\sigma_j} + q_{ij}^* \frac{Q_j'}{Q_j}\right)^{\sigma_j}}{L_i \cdot \left(\beta_i + \varphi \frac{L_i'}{L_i}\right)} \varphi^{\sigma_j - \sigma_{\bar{k}}}\right] \varphi^{\sigma_{\bar{k}}-\beta_i-1}. \tag{40}$$

We can also rewrite (26) as

$$q_{ij} = Q_j^{\sigma_j} \times \left(1 - \frac{1}{\sigma_j} + q_{ij}\frac{Q_j'}{Q_j}\right)^{\sigma_j} w_i^{-\sigma_j} \tau_{ij}^{-\sigma_j} \varphi^{\sigma_j}. \tag{41}$$

Using Assumptions 1–2, (40), and (41), the total profit becomes

$$\begin{aligned}
\Pi_i &= \sum_j \mathbb{I}_{ij}\left(p_{ij}q_{ij} - \tau_{ij}w_i\varphi^{-1}q_{ij}\right) - w_i\gamma\varphi^{\beta_i}L_i - w_i\left(\sum_j \mathbb{I}_{ij}\kappa_{ij} + \kappa_{D,i} + \kappa_{R,i}\right) \\
&= \frac{\sum_j \mathbb{I}_{ij}\left(\frac{\frac{1}{\sigma_j} - q_{ij}\frac{Q_j'}{Q_j}}{1 - \frac{1}{\sigma_j} + q_{ij}\frac{Q_j'}{Q_j}} - \frac{1}{\beta_i + \varphi\frac{L_i'}{L_i}}\right) Q_j^{\sigma_j}\left(1 - \frac{1}{\sigma_j} + q_{ij}\frac{Q_j'}{Q_j}\right)^{\sigma_j} w_i^{1-\sigma_j}\tau_{ij}^{1-\sigma_j}\varphi^{\sigma_j-\sigma_{\bar{k}}}}{\left[\sum_j \frac{\mathbb{I}_{ij}\tau_{ij}^{1-\sigma_j} w_i^{-\sigma_j} Q_j^{\sigma_j}\cdot\left(1 - \frac{1}{\sigma_j} + q_{ij}^*\frac{Q_j'}{Q_j}\right)^{\sigma_j}}{L_i\cdot\left(\beta_i + \varphi\frac{L_i'}{L_i}\right)}\varphi^{\sigma_j-\sigma_{\bar{k}}}\right]^{\frac{\sigma_{\bar{k}}-1}{\sigma_{\bar{k}}-\beta_i-1}}} \gamma^{\frac{\sigma_{\bar{k}}-1}{\sigma_{\bar{k}}-\beta_i-1}} \\
&\quad - w_i\left(\sum_j \mathbb{I}_{ij}\kappa_{ij} + \kappa_{D,i} + \kappa_{R,i}\right).
\end{aligned}$$

As $\widetilde{\varphi}_i(\gamma)$ and $q_{ij}^*(\varphi)$ exist and are unique, Assumptions 1–2 and (10) imply that $\frac{\Pi_i + w_i\left(\sum_j \mathbb{I}_{ij}\kappa_{ij} + \kappa_{D,i} + \kappa_{R,i}\right)}{\gamma^{\frac{\sigma_{\bar{k}}-1}{\sigma_{\bar{k}}-\beta_i-1}}}$ converges to a constant as γ becomes infinitesimal. Recall that $\alpha_i > -1$ for all i; the same procedure as in Appendix A.2 implies that $E(\Pi_i) < \infty$ if $\beta_i > \frac{\alpha_i+2}{\alpha_i+1}(\sigma_{\bar{k}} - 1)$. □

A.3.2 Productivity Distribution

Following the same procedure as in Appendix A.2, (26) and (27) yield

$$\begin{aligned}
\frac{\partial}{\partial\varphi}\frac{\tau_{ij}q_{ij}^{*}(\varphi)}{\varphi^{2}V_i'(\varphi)} =& -\frac{\tau_{ij}q_{ij}^{*}(\varphi)}{\varphi^{2}V_i'(\varphi)}\left[2\varphi^{-1}+\frac{V_i''(\varphi)}{V_i'(\varphi)}\right]+\frac{\tau_{ij}}{\varphi^{2}V_i'(\varphi)}\frac{\partial q_{ij}^{*}(\varphi)}{\partial\varphi}\\
=& -\frac{w_i^{-\sigma_j}\tau_{ij}^{1-\sigma_j}Q_j^{\sigma_j}}{L_i(\varphi)}\frac{\left(1-\frac{1}{\sigma_j}+q_{ij}\frac{Q_j'}{Q_j}\right)^{\sigma_j}}{\beta_i+\varphi\frac{L_i'(\varphi)}{L_i(\varphi)}}\cdot\left[2+\frac{\beta_i(\beta_i-1)+2\beta_i\varphi\frac{L_i'(\varphi)}{L_i(\varphi)}+\varphi^{2}\frac{L_i''(\varphi)}{L_i(\varphi)}}{\beta_i+\varphi\frac{L_i'(\varphi)}{L_i(\varphi)}}\right.\\
&\left.+\frac{1-\frac{1}{\sigma_j}+q_{ij}\frac{Q_j'}{Q_j}}{-\frac{1}{\sigma_j}\left(1-\frac{1}{\sigma_j}\right)+2\left(1-\frac{1}{\sigma_j}\right)q_{ij}\frac{Q_j'}{Q_j}+q_{ij}^{2}\frac{Q_j''}{Q_j}}\right]\varphi^{-(\beta_i-\sigma_j+1)-1}\\
\equiv& -J_{ij}\varphi^{-(\beta_i-\sigma_j+1)-1}. \qquad (42)
\end{aligned}$$

Using (29), we have

$$|J_i(\varphi)|=\left|\sum_{j=0}^{n}\frac{\partial}{\partial\varphi}\frac{\tau_{ij}q_{ij}^{*}(\varphi)}{\varphi^{2}V_i'(\varphi)}\right|=\left(\sum_j J_{ij}\varphi^{\sigma_j-\sigma_{\bar{k}}}\right)\varphi^{-(\beta_i-\sigma_{\bar{k}}+1)-1}. \qquad (43)$$

As J_{ij} converges to a constant, $\sum_j J_{ij}\varphi^{\sigma_j-\sigma_{\bar{k}}}$ converges to a constant because $\varphi^{\sigma_j-\sigma_{\bar{k}}}\to 0$ for all $j\neq\bar{k}$ and $\varphi^{\sigma_j-\sigma_{\bar{k}}}=1$ for $j=\bar{k}$.

As a result, the productivity distribution is given by

$$\begin{aligned}
g_i(\varphi)=&\frac{m_i(\widetilde{\gamma}_i(\varphi))}{\Pr(\gamma\in\Omega_i)}\widetilde{\gamma}_i(\varphi)^{\alpha_i}|J_i(\varphi)|\\
=&\frac{m_i(\widetilde{\gamma}_i(\varphi))}{\Pr(\gamma\in\Omega_i)}\left(\frac{\widetilde{\gamma}_i(\varphi)}{\varphi^{\sigma_{\bar{k}}-\beta_i-1}}\right)^{\alpha_i}\left(\sum_j J_{ij}\varphi^{\sigma_j-\sigma_{\bar{k}}}\right)\varphi^{-(1+\alpha_i)(\beta_i-\sigma_{\bar{k}}+1)-1}.
\end{aligned}$$

From Assumptions 1–2, (10), (40), (42), and $\lim_{\gamma\to 0}m_i(\gamma)=C_{m,i}$, the distribution of φ exhibits a power law with a tail index $(\alpha_i+1)(\beta_i+1-\max_j\sigma_j)$. □

A.3.3 Firm Size Distribution

For firms with sufficiently large φ, the firm size s is defined as the sum of export revenue $s\equiv\sum_j s_{ij}$. As $\widetilde{\varphi}_i(\gamma)$ and $q_{ij}^{*}(\varphi)$ exist and are unique, and since $s_{ij}=p_{ij}q_{ij}=q_{ij}^{1-\frac{1}{\sigma_j}}Q_j$, the functions $s_{ij}(\gamma)$, $s_{ij}(\varphi)$, $s_{ij}(q_{ij})$ and their inverse functions exist. Moreover, s_{ij} is decreasing in γ and increasing in both φ and q_{ij}. When γ becomes arbitrarily small, both

φ and s_{ij} become arbitrarily large. By (26) and $s=\sum_j q_{ij}^{1-\frac{1}{\sigma_j}}Q_j$, we have

$$\varphi=\frac{s^{\frac{1}{\sigma_{\bar{k}}-1}}}{\sum_j\left[w_i^{1-\sigma_j}\tau_{ij}^{1-\sigma_j}\left[Q_j\times\left(1-\frac{1}{\sigma_j}+q_{ij}\frac{Q_j'}{Q_j}\right)\right]^{\sigma_j-1}Q_j\varphi^{\sigma_j-\sigma_{\bar{k}}}\right]^{\frac{1}{\sigma_{\bar{k}}-1}}}. \tag{44}$$

Deriving $\partial q_{ij}^*/\partial\varphi$ using (26) and applying the product rule to $s_{ij}=q_{ij}^{1-\frac{1}{\sigma_j}}Q_j$ yields

$$\begin{aligned}\frac{\partial s_{ij}(\varphi)}{\partial\varphi}&=\frac{\partial s_{ij}}{\partial q_{ij}^*}\frac{\partial q_{ij}^*}{\partial\varphi}=\frac{1-\frac{1}{\sigma_j}+q_{ij}^*\frac{Q_j'}{Q_j}}{\frac{1}{\sigma_j}\left(1-\frac{1}{\sigma_j}\right)-2\left(1-\frac{1}{\sigma_j}\right)q_{ij}^*\frac{Q_j'}{Q_j}-\left(q_{ij}^*\right)^2\frac{Q_j''}{Q_j}}w_i\tau_{ij}\varphi^{-2}q_{ij}\\&=\frac{\left(1-\frac{1}{\sigma_j}+q_{ij}^*\frac{Q_j'}{Q_j}\right)\left[Q_j\times\left(1-\frac{1}{\sigma_j}+q_{ij}^*\frac{Q_j'}{Q_j}\right)\right]^{\sigma_j}}{\frac{1}{\sigma_j}\left(1-\frac{1}{\sigma_j}\right)-2\left(1-\frac{1}{\sigma_j}\right)q_{ij}^*\frac{Q_j'}{Q_j}-\left(q_{ij}^*\right)^2\frac{Q_j''}{Q_j}}w_i^{1-\sigma_j}\tau_{ij}^{1-\sigma_j}\varphi^{\sigma_j-2}.\end{aligned} \tag{45}$$

As a result,

$$\frac{\partial s(\varphi)}{\partial\varphi}=\sum_{j=0}^n\frac{\partial s_{ij}(\varphi)/\partial\varphi}{\varphi^{\sigma_j-2}}\varphi^{\sigma_j-2}=\left(\sum_{j=0}^n\frac{\partial s_{ij}(\varphi)/\partial\varphi}{\varphi^{\sigma_j-2}}\varphi^{\sigma_j-\sigma_{\bar{k}}}\right)\varphi^{\sigma_{\bar{k}}-2}. \tag{46}$$

Using (29), (43), (44), (45), and (46), the absolute value of the Jacobian term $|J_{s,i}(\varphi)|$ is thus

$$\begin{aligned}|J_{s,i}(\varphi)|\equiv\left|\frac{\partial\widetilde{\gamma}_i(s)}{\partial s}\right|&=\left|\frac{\partial\widetilde{\gamma}_i(\varphi)}{\partial\varphi}\frac{\partial\varphi(s)}{\partial s}\right|=|J_i(\varphi)|\left(\frac{\partial s(\varphi)}{\partial\varphi}\right)^{-1}\\&=\frac{\sum_j J_{ij}\varphi^{\sigma_j-\sigma_{\bar{k}}}}{\sum_{j=0}^n\frac{\partial s_{ij}(\varphi)/\partial\varphi}{\varphi^{\sigma_j-2}}\varphi^{\sigma_j-\sigma_{\bar{k}}}}\left(\frac{\varphi}{s^{\frac{1}{\sigma_{\bar{k}}-1}}}\right)^{-\beta_i}s^{-\frac{\left(\beta_i-\sigma_{\bar{k}}+1\right)}{\sigma_{\bar{k}}-1}-1}.\end{aligned}$$

As a result, the above equation along with (40) yields the firm size distribution $g_{s,i}(s)$:

$$\begin{aligned}g_{s,i}(s)=&\frac{m_i(\widetilde{\gamma}_i(s))}{\Pr(\gamma\in\Omega_i)}\widetilde{\gamma}_i(s)^{\alpha_i}|J_{s,i}(\varphi)|\\=&\frac{m_i(\widetilde{\gamma}_i(s))}{\Pr(\gamma\in\Omega_i)}\frac{\sum_j J_{ij}\varphi^{\sigma_j-\sigma_{\bar{k}}}}{\sum_{j=0}^n\frac{\partial s_{ij}(\varphi)/\partial\varphi}{\varphi^{\sigma_j-2}}\varphi^{\sigma_j-\sigma_{\bar{k}}}}\left(\frac{\varphi}{s^{\frac{1}{\sigma_{\bar{k}}-1}}}\right)^{-\beta_i+\alpha_i(\sigma_{\bar{k}}-\beta_i-1)}\left(\frac{\widetilde{\gamma}_i(\varphi)}{\varphi^{\sigma_{\bar{k}}-\beta_i-1}}\right)^{\alpha_i}s^{-\frac{(\alpha_i+1)(\beta_i-\sigma_{\bar{k}}+1)}{\sigma_{\bar{k}}-1}-1}.\end{aligned}$$

From (10), (40), (43), (44), (45), Assumptions 1–2, and $\lim_{\gamma\to 0}m_i(\gamma)=C_{m,i}$, each of the multiplicative terms besides $s^{-\frac{(\alpha_i+1)\left(\beta_i-\sigma_{\bar{k}}+1\right)}{\sigma_{\bar{k}}-1}-1}$ converges to a constant. Thus, the firm size distribution exhibits a power law with a tail index $\frac{(\alpha_i+1)(\beta_i+1-\max_j\sigma_j)}{\max_j\sigma_j-1}$. □

A.4 Proof of Proposition 3

Applying the symmetric-country assumption to Proposition 2 implies that $E(\Pi) < \infty$ under $\alpha > -1$ and $\beta > \frac{\alpha+2}{\alpha+1}(\sigma - 1)$. Using (31–33), the definition of ϕ, and recalling that $\gamma_X/\gamma_D = \delta < 1$, we can restate the expected profit as a function of γ_D:

$$E(\Pi) = (\kappa_D + \kappa_R)\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}}\left\{\Gamma_D + \left[\left(1 + n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\Gamma_X\right\} - (\kappa_D + \kappa_R)F(\gamma_D) - n\kappa_X F(\gamma_X),$$

where $\Gamma_z \equiv \int_0^{\gamma_z} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} dF(\gamma)$ for $z \in \{D, X\}$. In equilibrium, γ_D solves the free-entry condition. Using (35), it is then readily verified that

$$\frac{\partial E(\Pi)}{\partial \gamma_D} = (\kappa_D + \kappa_R)\frac{\sigma - 1}{\beta - \sigma + 1}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}-1}\left[(\Gamma_D - \Gamma_X) + \left(1 + n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\Gamma_X\right] > 0.$$

Note that both Γ_D and Γ_X are positive and increasing in γ_D; thus $\lim_{\gamma_D\to\infty}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}}\Gamma_D = \lim_{\gamma_D\to\infty}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}}\Gamma_X = \infty$. Since both $F(\gamma_D)$ and $F(\gamma_X)$ are less than 1, it follows that $\lim_{\gamma_D\to\infty} E(\Pi) = \infty$. Since $E(\Pi)$ is bounded from above by $E(\Pi)|_{\gamma_D=1}$, for any $\kappa_e \in (0, E(\Pi)|_{\gamma_D=1})$ there exists a unique γ_D such that the free-entry condition holds, hence establishing the uniqueness of equilibrium. □

A.5 Proof of Proposition 4

We first derive the effect of τ on the surviving cutoff γ_D. Total differentiating $E(\Pi)$ with respect to τ yields $\frac{d\gamma_D}{d\tau} = -\frac{\partial E(\Pi)/\partial\tau}{\partial E(\Pi)/\partial\gamma_D}$. We have obtained $\partial E(\Pi)/\partial\gamma_D$ in Appendix A.4. Partially differentiating the expected profit with respect to τ yields

$$\frac{\partial E(\Pi)}{\partial \tau} = -\frac{(\sigma - 1)\beta}{\beta - \sigma + 1}\frac{n\tau^{-\sigma}}{1 + n\tau^{1-\sigma}}(\kappa_D + \kappa_R)\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}}\left(1 + n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\Gamma_X.$$

This then leads to

$$\frac{d\gamma_D}{d\tau} = \beta\gamma_D\frac{n\tau^{-\sigma}}{1 + n\tau^{1-\sigma}}\frac{(1 + n\tau^{1-\sigma})^{\frac{\beta}{\beta-\sigma+1}}\Gamma_X}{(\Gamma_D - \Gamma_X) + (1 + n\tau^{1-\sigma})^{\frac{\beta}{\beta-\sigma+1}}\Gamma_X} > 0.$$

The effect of τ on the exporting cutoff γ_X is defined by $\frac{d\gamma_X}{d\tau} = \delta\frac{d\gamma_D}{d\tau} + \frac{d\delta}{d\tau}\gamma_D$; therefore

$$\frac{d\gamma_X}{d\tau} = \frac{\delta\beta\gamma_D n\tau^{-\sigma}}{1+n\tau^{1-\sigma}}\left[\frac{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\Gamma_X}{\left(\Gamma_D-\Gamma_X\right)+\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\Gamma_X} - \frac{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}}{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1}\right].$$

The first and second terms in the brackets are less and greater than 1, respectively. We thus conclude that $d\gamma_X/d\tau < 0$. Combining (33) and (30), we obtain the equilibrium productivity:

$$\widetilde{\varphi}(\gamma) = \begin{cases} \left(\kappa_D+\kappa_R\right)^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} & \text{if } \gamma\in(\gamma_X,\gamma_D] \\ \phi\left(\kappa_D+\kappa_R\right)^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} & \text{if } \gamma\in[0,\gamma_X] \end{cases}. \tag{47}$$

For the effect on productivity, taking derivatives of (47) yields

$$\frac{d\widetilde{\varphi}(\gamma)}{d\tau} = \begin{cases} \frac{\left(\kappa_D+\kappa_R\right)^{\frac{1}{\beta}}n\tau^{-\sigma}\left(1+n\tau^{1-\sigma}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}\Gamma_X\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}}{\left(\Gamma_D-\Gamma_X\right)+\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\Gamma_X}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{\beta+1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} > 0 & \text{for non-exporting firms} \\ -\phi\frac{\left(\kappa_D+\kappa_R\right)^{\frac{1}{\beta}}n\tau^{-\sigma}\left(1+n\tau^{1-\sigma}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}\Gamma_X\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}}{\left(\Gamma_D-\Gamma_X\right)+\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\Gamma_X}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{\beta+1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} < 0 & \text{for exporting firms.} \end{cases}$$

The claims on the comparative statics of φ thus follow. □

A.6 Proof of Proposition 5

We can write the price index as $P^{1-\sigma} = P_D^{1-\sigma} + nP_X^{1-\sigma}$, where $P_D^{1-\sigma}$ and $nP_X^{1-\sigma}$ are the components of $P^{1-\sigma}$ in which the goods are from domestic and foreign firms, respectively. Therefore, the share of expenditure on domestic products is defined by $\lambda \equiv P_D^{1-\sigma}/P^{1-\sigma}$, and the share of expenditure on goods from a foreign country is defined by $\lambda_X = P_X^{1-\sigma}/P^{1-\sigma} = (1-\lambda)/n$. Plugging (30) into (36) and using the above definitions yields

$$\lambda = \frac{\Gamma_D + \left(\phi^{\sigma-1}-1\right)\Gamma_X}{\Gamma_D + \left[\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}-1\right]\Gamma_X} \tag{48}$$

$$\lambda_X = \frac{\phi^{\sigma-1}\tau^{1-\sigma}\Gamma_X}{\Gamma_D + \left[\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}-1\right]\Gamma_X}. \tag{49}$$

From (33) and (34) we have

$$d\ln\gamma_D = \beta d\ln P \tag{50}$$

$$d\ln\gamma_X = \beta d\ln P + d\ln\delta. \tag{51}$$

Note that the assumptions of Proposition 3 ensure that $E(\Pi) < \infty$; hence Γ_D and Γ_X are both finite. We further define the short-hand notation $\eta_z \equiv \gamma_z^{1-\frac{\sigma-1}{\beta-\sigma+1}} f(\gamma_z)\Gamma_z^{-1}$, where $z \in \{D, X\}$. The welfare is defined as the real income $W \equiv N/P$. It thus follows from (50) that $\frac{d\ln W}{d\ln\tau} = -\frac{1}{\beta}\frac{d\ln\gamma_D}{d\ln\tau}$. With $E(\Pi)$ given in Appendix A.4, $E(\Pi) = \kappa_e$ yields

$$\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} = \frac{\kappa_e + (\kappa_D + \kappa_R)F(\gamma_D) + n\kappa_X F(\gamma_X)}{(\kappa_D + \kappa_R)\left\{\Gamma_D + \left[(1 + n\tau^{1-\sigma})^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\Gamma_X\right\}}.$$

By log-differentiating the above equation, combining the definitions of η_D and η_X with the free-entry condition, and noting that $(1 + n\tau^{1-\sigma})^{\frac{\beta}{\beta-\sigma+1}} = (1 + n\tau^{1-\sigma})\phi^{\sigma-1}$, one can obtain $d\ln\gamma_D = \beta(1-\lambda)d\ln\tau$ with a few algebraic manipulations. The welfare elasticity is accordingly $\frac{d\ln W}{d\ln\tau} = \lambda - 1$. With symmetric countries, the trade elasticity equals $\varepsilon = d\ln(\lambda_X/\lambda)/d\ln\tau$. The ACR formula is restated as $\frac{d\ln W}{d\ln\tau} = \frac{d\ln\lambda/d\ln\tau}{d\ln(\lambda_X/\lambda)/d\ln\tau} = \lambda - 1$. Thus our model entails the local ACR formula.

For the trade elasticity, recall that $d\ln\delta = d\ln(\gamma_X/\gamma_D)$ by (50) and (51), and $d\ln\gamma_D/d\ln\tau = \beta(1-\lambda)$. Using (48) and (49), log-differentiating λ_X/λ with respect to τ thus yields

$$\begin{aligned}\varepsilon = &(\sigma-1)\frac{\Gamma_D - \Gamma_X}{\Gamma_D + (\phi^{\sigma-1}-1)\Gamma_X}\frac{d\ln\phi}{d\ln\tau} + (1-\sigma)\\ &+ \frac{\Gamma_D}{\Gamma_D + (\phi^{\sigma-1}-1)\Gamma_X}\eta_X\frac{d\ln\frac{\gamma_X}{\gamma_D}}{d\ln\tau} + \frac{\Gamma_D}{\Gamma_D + (\phi^{\sigma-1}-1)\Gamma_X}(\eta_X - \eta_D)\beta(1-\lambda).\end{aligned}$$

It is readily verified with a numerical example that the trade elasticity is variable in τ and depends on the distribution of γ. □