Innovation, Firm Size Distribution, and Gains from Trade*

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

May 31, 2019

Abstract

We study a trade model with monopolistic competition *à la* Melitz (2003) that is standard except that firm heterogeneity is endogenously determined by firms innovating to enhance their productivities. We show that the equilibrium productivity and firm-size distributions exhibit power-law tails under rather general conditions on demand and innovating technology. Moreover, the emergence of the power laws is consistent with rather general underlying primitive heterogeneity among firms. We investigate the model's welfare implications, and conduct a quantitative analysis of welfare gains from trade. We find that, conditional on the same trade elasticity and values of the common parameters, our model yields 40% higher welfare gains from trade than a standard model with exogenously given productivity distribution.

Codes: F12, F13, F41.

Keywords: innovation, power law, welfare gains from trade, firm heterogeneity.

*This paper was previously circulated under the title "Productivity Investment, Power Laws, and Welfare Gains from Trade". We are grateful for the detailed and insightful comments by Mathieu Parenti and Mark Razhev. For helpful comments, we also thank Pao-Li Chang, Davin Chor, Jacques-Francois Thisse, and Jen-Feng Lin and the seminar participants at the University of Melbourne, the 2018 Asia-Pacific Trade Seminars, the 2018 SMU-NUS-Paris Joint Trade Workshop, the 2018 Taiwan Economic Research, and the 2018 Macroeconometric Modeling Workshop. The authors gratefully acknowledge the financial support from Academia Sinica (Academia Sinica Investigator Award #23-18) and Ministry of Science and Technology (MOST 105-2410-H-001-001-MY3).

†Institute of Economics, Academia Sinica, Taipei, Taiwan. E-mail: f97323011@ntu.edu.tw.

‡School of Economics, Singapore Management University, Singapore. E-mail: wentaihsu@smu.edu.sg.

§Institute of Economics, Academia Sinica, and Department of Economics, National Taiwan University, Taipei, Taiwan. E-mail: speng@econ.sinica.edu.tw.

1 Introduction

In the last two decades of the development of the trade literature on heterogeneous firms, the source of heterogeneity has been mostly exogenous, e.g., the exogenously given productivity distribution in Bernard, Eaton, Jensen, and Kortum (2003), Eaton and Kortum (2002), Melitz (2003), Melitz and Ottaviano (2008), and the large literature following these work-horse models. Trade affects which parts of the productivity distribution in each country are utilized via either firm selection or comparative advantage. Nevertheless, empirical evidence shows that trade affects productivity at the level of individual firms, making the distribution of productivity endogenous. For examples, see Pavcnik (2002), Fernandes (2007), Bustos (2011), and Aghion, Bergeaud, Lequien, and Melitz (2018). Furthermore, Bustos (2011) and Aghion et al. (2018) show that innovation is a plausible mechanism through which trade affects firms' productivity.[1] Both studies explain their empirical results using models that assume firm productivity as a function of innovating activities.

Following the above-mentioned studies, we study a variant of the Melitz (2003) model in which we embed an innovation mechanism. Surprisingly, we find that this mechanism allows us to microfound the power laws in firm productivity and firm size under a rather generalized general equilibrium model of international trade with minimal assumption on underlying firm heterogeneity. This paper also gauges the importance of this innovation mechanism by analyzing the welfare gains from trade via a comparison with a Melitz model via the lens of the framework proposed by Arkolakis, Costinot, and Rodríguez-Clare (2012; henceforth ACR).

Power laws in both productivity and firm size are widely documented empirical regularities (see, for examples, Axtell 2001, Luttmer 2007, and Nigai 2017). Moreover, it has been shown that these power laws provide microfoundation for the gravity equations (Arkolakis, Costinot, Donaldson, and Rodríguez-Clare 2018 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 ACR and Arkolakis et al. 2018). The validity of these power laws forms a strong justification for making power-function assumptions in economic models, which make complex and large-scale quantitative analysis possible.[2] Thus, it is important to understand what may be a plausible general explanation for these power laws.

Our model starts with a simple R&D process. An entrant firm decides the complexity of the production process and hence the number of procedures. For each procedure, the entrant firm

[1] Aghion et al. (2018) document that innovation activities are strongly and positively correlated with export. Export of innovating firms are approximately 10 times larger than that of non-innovating firms. Their empirical analysis also shows that export demand shocks have a significant positive impact on innovation activities for large firms, especially for firms at top deciles.

[2] See Chang and Hsu (2019) for a discussion on the significance of the power laws in international trade and economic geography.

conducts a sequence of experiments to enhance the performance of the procedure, and the entrant firms differ in their probabilities of failure in conducting the experiments. This process results in a neat relation between capability, innovation effort, and resulting productivity. We then embed this relation into a standard Melitz model (with a CES preference) as an innovation stage after entry and before production. As a first-cut, we show that if the distribution of failure probability has a *finite and positive* density near zero, then the power laws emerge. This is a rather weak condition, as it requires little on the functional form of underlying firm heterogeneity. This is achieved via a "Power Law Change of Variable Close to the Origin" technique which we explain in Section 2.[3]

We find that the model can actually be generalized on several fronts. First, the preference/demand side can be generalized from CES to a *regularly varying* demand, which is much broader as it allows for non-homothetic preference and variable markups. Several non-CES preferences studied by Mrázová and Neary (2017) in fact entail regularly varying demand functions. Second, both the innovation-productivity relation and the distribution of failure probability can also be generalized to be regularly varying functions. In particular, this implies that the power laws continue to hold even if the density of the distribution of failure probability near zero is zero or infinite, provided that the distribution is regularly varying at the left tail. As we will explain in Section 2, this includes many well-known and widely-used distributions.

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 with all model parameters allowed to be country-specific. Interestingly, the tail indices of both 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 for each country.

We also analyze how 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, the 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 productivities of non-exporters.

We conduct a quantitative analysis to clarify how innovation affects welfare gains from trade. Despite some slight differences from the class of models characterized by ACR, the welfare gains from trade still follow their formula, i.e., $d\ln W = \frac{1}{\varepsilon} d\ln\lambda$, where W is welfare, λ is the expen-

[3]This technique has already been used in physics; see Jan et al. (1999), Sornette (2002) and Newman (2005). The name of the technique was given by Sornette (2006, Section 14.2.1).

diture share on domestic goods, and ε is the trade elasticity. We refer to this formula as the local ACR formula as it deals with small changes in trade cost. However, the (global) ACR formula $W'/W = (\lambda'/\lambda)^{1/\varepsilon}$ for large changes in trade cost does not apply here because the trade elasticity ε in our model is a variable. Nevertheless, one can obtain the welfare changes for large changes in trade cost by integrating over the local ACR formula.

To highlight the role played by innovation, we compare the welfare gains from trade with the Melitz (2003) model with an exogenous Pareto distribution of productivity, in which both the global and local ACR formula hold. For this purpose, we focus on a symmetric-country world with CES preference and a power function for innovation-productivity relation. When firms' R&D abilities are uniformly distributed, the resulting productivity distribution has a Pareto right tail; thus such a parameterization is adopted. We calibrate the model to match the same trade elasticity, domestic expenditure share, and the share of exporters. 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%. The intuition is as follows. As mentioned, exporters innovate more and non-exporters innovate less when facing trade liberalization, thus creating a larger productivity advantage of exporters over non-exporters. Compared with the Melitz model with exogenous productivity distribution, the above-mentioned effect entails larger imports and exports, and by the ACR formula, this implies larger welfare gains from trade. That the model with innovation entails significantly larger welfare gains from trade confirms the importance of incorporating innovation and endogenous choice of productivity.

This paper is closely related to the literature on power laws in firm size. A popular explanation for such power laws is based on firm-size dynamics that follow a random growth process; see, for example, Luttmer (2007), Rossi-Hansberg and Wright (2007), and Acemoglu and Cao (2015).[4] Recently, Chaney (2014, 2018) and Geerolf (2017) have provided explanations for the power law in firm size via network and firm hierarchy, respectively. Note that no models of the above-mentioned studies are free of functional form assumption or restrictions; for examples, Luttmer (2007) and Acemoglu and Cao (2015) assume CES and constant-relative-risk-aversion (CRRA) preferences. Thus, our relaxation of demand from the CES to regularly varying functions should be viewed as an advantage rather than a strong restriction. Most importantly, the common theme of these studies and our work is that power laws emerge with minimal assumptions on the underlying firm heterogeneity. Our model differs from these studies in its economic mechanism, and is most closely related to Geerolf (2017) in terms of mathematical mechanism because both use the "power law change of variable close to the origin" technique.

This paper is also closely related to Yeaple (2005), Bustos (2011), Bas and Ledezma (2015),

[4] Also see Gabaix (2009) for a survey of the literature.

Aghion et al. (2018), and Bonfiglioli, Crinò, and Gancia (2018),[5] who also model how innovation effort affects productivity. Whereas the mechanism of our theory bears some similarity to these studies, our work differs at least in the two following aspects: (1) we show that the concentration of innovation effort among exporters and large firms results in power laws in both productivity and firm size under a rather general environment; (2) we investigate the welfare effect of such innovation effort.

As mentioned, our theoretical and quantitative analyses on the welfare gains from trade is closely related to ACR.[6] Our approach in modeling innovation is similar in spirit to the technological choice embedded in the ACR framework, but is different in form.[7] Nevertheless, we show that the ACR formula still holds in our model, despite a variable trade elasticity. Our work is also closely related to Melitz and Redding (2015) who conduct a welfare comparison between homogeneous-firm and heterogeneous-firm models by fixing common parameters. To highlight the role of innovation, the welfare comparison between our model and Melitz-Pareto is conducted in a similar fashion to Melitz and Redding as our quantitative comparison is done conditional on the same trade elasticity and common parameters. Whereas Melitz and Redding show that the heterogeneous-firm model adds additional gains from trade compared with homogeneous firms,[8] we show that innovation further adds gains from trade compared with the Melitz-Pareto model. Moreover, we show that such extra gains could be substantial.

The rest of the paper is organized as follows. Section 2 presents the model and shows how power laws emerge. Section 3 provides comparative statics of productivity distribution on trade costs and other parameters. Section 4 studies the properties pertaining to welfare gains from trade and conducts a quantitative analysis. Section 5 concludes.

2 Power Laws in Productivity and Firm Size

We first 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. Such results easily extend to a general open-economy environment, as we show in Section 2.3.

[5] A feature in many of these studies is that productivity or quality is affected by choices in some type of fixed costs. Also see Sutton (1991) for an early example of such modeling.

[6] Since ACR, the interests on welfare gains from trade have revitalized and the related literature has been rapidly growing. For a small set of examples, see Caliendo and Parro (2015) on the roles of intermediate goods and sectoral linkages; Melitz and Redding (2014) on how sequential production can amplify welfare gains from trade; and Hsieh and Ossa (2016) on the global welfare impact of China's trade integration and productivity growth. For pro-competitive effects, see Arkolakis et al. (2018), Edmond, Midrigan, and Xu (2015), Feenstra and Weinstein (2017), and Holmes, Hsu, and Lee (2014). Our work differs in that we focus on the effect of innovation.

[7] As is made clear in Section 2, innovation effort is determined in the stage before production and consumption, whereas ACR assumes they are simultaneous.

[8] See their Propositions 2 and 3.

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 individual inverse demand function is given by $p = D(q(\upsilon); A)$ on $[\underline{q}, \infty)$ with $\underline{q} \geq 0$. Namely, we assume that the inverse demand of a variety depends on all the other varieties only through aggregate variables $A \in \mathbb{R}^n$. Assume that D is twice differentiable and that the law of demand holds: $D' < 0$.[9]

On the production side, labor is the only input, and firms engage in monopolistic competition. To enter, each entrant hires κ_e amount of labor, which allows the entrant to obtain a distinct variety and a draw of a R&D parameter γ from a given distribution which we explain shortly. For a firm to produce, κ_D units of labor as fixed input is required. The productivity of a firm is endogenously determined and denoted as φ. Thus the unit labor requirement is φ^{-1}. By choosing labor as numeraire, the wage equals 1, and the total cost of production as a function of output q is $q/\varphi + \kappa_D$. As in Melitz (2003), a positive κ_D results in firm selection. As we will see, whether there is selection or not ($\kappa_D > 0$ or $\kappa_D = 0$) is immaterial for the results on power laws; we keep selection for generality and for the welfare comparison with the literature. A firm's profit from production is thus

$$\pi(\varphi) = pq - \varphi^{-1}q - \kappa_D. \tag{1}$$

Each entrant can determine its productivity level by engaging in R&D activities in the following manner. The production process involves a continuum of procedures, and the entrant can choose the size of the continuum, k. How well the firm can perform 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 is endowed with one quality unit to begin with. 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. But if the experiment fails, no more experiments will be performed and the quality of the procedure is finalized. Firms differ in their probabilities of failure, $\gamma \in (0, 1)$. In short, the probability of obtaining quality y for a procedure is $(1-\gamma)^{y-1}\gamma$, i.e., y is geometrically distributed on the continuum of size k. The process is illustrated as in Figure 1.

[9] Such an inverse demand function can be generated by (but not limited to) maximizing an additively separable utility $U = \int_{\upsilon\in\Upsilon} u(q(\upsilon))\, d\upsilon$ subject to the budget constraint $\int_{\upsilon\in\Upsilon} p(\upsilon)\, q(\upsilon)\, d\upsilon = w$. The sub-utility $u(.)$ is defined on $[\underline{q}, \infty)$ with $\underline{q} \geq 0$. Assume that $u' > 0$ and $u'' < 0$. The standard solution yields the inverse demand function $p = D(q(\upsilon); A) \equiv u'(q(\upsilon))/A$, where A is the Lagrange multiplier of the consumer's problem and is a general equilibrium object. Note that $u'' < 0$ implies that $D' < 0$. For other forms of $D(q(\upsilon); A)$ than $u'(q(\upsilon))/A$, see the discussion following Assumption 1 and Table 1.

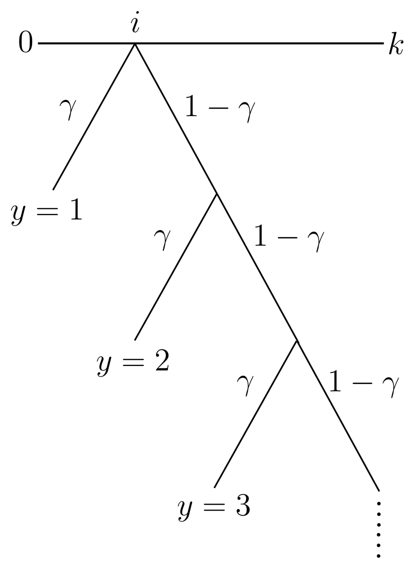


Figure 1: A sequence of Bernoulli trials

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 mass of procedures, k. The productivity φ is a function of the total quality of all k procedures, $kE(y)$. That is,

$$\varphi \equiv B(kE(y)) = B\left(k\sum_{y=1}^{\infty}(1-\gamma)^{y-1}\gamma y\right) = B\left(\frac{k}{\gamma}\right).$$

The function $B(\cdot)$ is strictly increasing and concave on $\mathbb{R}_+$, and $\lim_{k\to\infty} B\left(\frac{k}{\gamma}\right) = \infty$. The concavity of $B(\cdot)$ reflects the management burden for the firm to manage these research assistants. For operational convenience, we rewrite the above equation as

$$k = \gamma B^{-1}(\varphi) \equiv \gamma V(\varphi), \tag{2}$$

where $V \equiv B^{-1}$ is strictly increasing and convex in φ and $\lim_{\varphi\to\infty} V(\varphi) = \infty$. Here, k as a function in γ and φ given by (2) defines what we term an innovation cost function. The c.d.f. and p.d.f. of the distribution of γ are denoted as $F(\cdot)$ and $f(\cdot)$, respectively. We assume that $f(.)$ is continuous and positive on $(0, 1)$. The higher the γ, the more costly to obtain the same φ.[10]

[10]Note that above-described process entails a deterministic relation between firm heterogeneity γ and productivity φ by (2), and is unrelated to the random growth process used in the literature. First, the random walk or Brownian motion in a random growth process is idiosyncratic to firms with different capability. For the firms with the same γ, they may be struck by different shocks over time. Second, whereas the random walk or Brownian motion with a reflection barrier is the basis for entailing power laws in those models, the mechanism generating power laws here does not rely either on the random walk or Brownian motion or a reflection barrier. In particular, no central limit theorem is applied here.

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

$$\Pi(\varphi;\gamma) = \pi(\varphi) - \gamma V(\varphi), \tag{3}$$

and the resulting optimal choice of φ is denoted as $\varphi^* = \widetilde{\varphi}(\gamma)$.

Given a non-degenerate distribution of optimal φ, there may exist a cutoff $\varphi_D > 0$ below which firms decide not to produce and obtain $\pi(\varphi) = 0$. To justify paying $\gamma V(\varphi) > 0$, $\pi(\varphi) > \gamma V(\varphi)$ is needed. Suppose the optimal choice of φ is strictly decreasing in γ. Then, the fact that $\pi(\varphi) = 0$ for those firms with $\varphi < \varphi_D$ implies a corresponding cutoff $\gamma_D > 0$ such that $\pi(\widetilde{\varphi}(\gamma_D)) = \gamma_D V(\widetilde{\varphi}(\gamma_D))$. Thus, the free entry condition can be written as

$$E(\Pi) \equiv \int_0^{\gamma_D} \Pi(\widetilde{\varphi}(\gamma);\gamma)\, dF(\gamma) = \kappa_e. \tag{4}$$

In sum, the model contains three stages as follows:

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 and draws its type γ randomly from the distribution $f(\gamma)$.

Stage 2. Innovation Stage: Given γ, each firm decides whether to invest in productivity or not, and if yes, how much to invest to determine its productivity level φ.

Stage 3. Production/Consumption Stage: Each firm decides whether to produce or not. If yes, each firm pays κ_D and determines its output and price. Production and consumption take place and markets clear.

2.2 Equilibrium and Power Laws

This subsection provides an exposition of how power laws for both productivity and firm size emerge in a closed economy. We first provide an illustrative example that shows how a simple condition allows these power laws to emerge in a Melitz model. We then provide an exposition of a more general model and hence more general results.

2.2.1 Illustrative Example with Melitz Model

First follow Melitz (2003) to consider a CES demand: $p = \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{1}{\sigma}} q^{-\frac{1}{\sigma}}$ where $\sigma > 1$. Also for tractability, we use a simple power function for the innovation cost: $k = \gamma\varphi^{\beta}$, where $\beta > 1$.

We solve the model backwards. For any φ, a firm's optimal output in the production stage is given by

$$q(\varphi) = \frac{N}{P^{1-\sigma}} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \varphi^{\sigma}. \tag{5}$$

The operating profit is accordingly

$$\pi(\varphi) = \frac{N}{P^{1-\sigma}} \frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\sigma-1} \varphi^{\sigma-1} - \kappa_D. \tag{6}$$

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

$$\widetilde{\varphi}(\gamma) \equiv \left[\frac{\frac{N}{P^{1-\sigma}} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{1}{\beta-\sigma+1}} \gamma^{-\frac{1}{\beta-\sigma+1}}. \tag{7}$$

The second-order condition for innovation must hold so that (7) is optimal. 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 convex enough. By inserting (6) into (7), a firm's total profit becomes

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

hence there is an unique cutoff

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

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

An equilibrium is defined by (7), (8), the aggregate price

$$P^{1-\sigma} = M_e \left[\int_0^{\gamma_D} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \widetilde{\varphi}(\gamma)^{\sigma-1} dF(\gamma)\right]$$

where M_e denotes the mass of entrants paying the entry cost, and the free entry condition (4). For the free entry condition to entail a solution, we need a condition to ensure that the expected profit $\int_0^{\gamma_D} \Pi(\widetilde{\varphi}(\gamma); \gamma)\, dF(\gamma)$ is finite. The same condition will also ensure that price index P is well defined.

The firm size is characterized by the sales revenue $s \equiv pq$. Using (5) and (7), firm size as a function of γ is therefore

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

By applying change of variable, the productivity and firm size distributions are obtained as follows:

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

For the above distributions to exhibit power law, we require $\lim\limits_{\varphi\to\infty} \frac{g(\varphi)}{\varphi^{-(\beta-\sigma+1)-1}}$ and $\lim\limits_{s\to\infty} \frac{g_s(s)}{s^{-\frac{\beta-\sigma+1}{\sigma-1}-1}}$ to be a constant respectively. Observe from (7) and (9) that both φ and s become arbitrarily large as γ approaches 0, and note that both P and γ_D are independent of γ. Hence, *the above distributions exhibit power laws if* $\lim\limits_{\gamma\to 0} f(\gamma) = K > 0$:

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

In other words, we only require $f(\gamma)$ to have a positive probability density near the 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 $\lim\limits_{\gamma\to 0} f(\gamma) = K > 0$ ensures that the expected profit is finite if $\beta > 2(\sigma-1)$.

The mechanism above is referred to as "power law change of variable close to the origin": if the distribution of a variable x has a positive density near the origin, and the interested variable y relates to x in a reciprocal manner $y = x^{-1}$, then y becomes arbitrarily large as x goes to 0 and the distribution of y exhibits a power law tail (Sornette 2006, Section 14.2.1). Since the innovation efforts entail a reciprocal relationship between γ and φ given by (7), the condition that $f(\gamma)$ being positive near the origin thus entails a power law productivity distribution. In other words, the condition on $f(\gamma)$ implies that there is a sufficient mass of capable firms, resulting in a fat-tailed productivity distribution. Since firm size is proportional to productivity so it also takes a reciprocal relationship with γ, the resulting firm size distribution thus exhibits a power law under the same condition.

The restriction on $f(\gamma)$ is rather weak as it only asks for a positive density around the origin. Note that the result does not invoke any fat right tail as γ itself is a probability; thus $f(\gamma)$ is defined on $(0,1)$. The above simple example illustrates how an addition of the innovation stage to the Melitz model and a minimal requirement on firm heterogeneity $f(\gamma)$ can give rise to both power laws in productivity and firm size. In the remainder of this section, we turn to the general model to show that this result actually holds in much larger classes of demand and innovation cost functions than those assumed here. Moreover, we can further relax the restriction on $f(\gamma)$ and show that all of these results hold in an open economy where all parameters are allowed to be country specific.

2.2.2 Preliminaries: Regularly and smoothly varying functions

We first provide some preliminaries on regular variation that are applied to both the inverse demand and innovation cost functions. A function $v(x)$ is *regularly varying* if for some $\zeta \in \mathbb{R}$,

$$v(x) = x^{\zeta} l(x),$$

where $l(x)$ is such that for all $t > 0$

$$\lim_{x\to\infty} \frac{l(tx)}{l(x)} = 1.$$

The function $l(x)$ is referred to as a *slowly varying* function. If $l(x)$ is a constant, then the function $v(x)$ reduces to a power function. This implies $v(tx) \approx t^{\zeta} v(x)$ for large x; that is, a regularly varying function is a homogeneous function (of degree ζ) asymptotically. The definition of a smoothly varying function is as follows (see e.g. Bingham et al. 1989).

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

$$\lim_{x\to\infty} \frac{x^n v^{(n)}(x)}{v(x)} = \zeta(\zeta-1)\ldots(\zeta-n+1), \tag{10}$$

where $v^{(n)}(x)$ denotes the n-th derivative of $v(x)$. An equivalent definition is as follows: Consider a transformation to the infinitely differentiable regularly varying function $v(x)$: $\widehat{v}(x) \equiv \log v(e^{x})$. Then, $v(x)$ is a smoothly varying function if

$$\lim_{x\to\infty} \widehat{v}'(x) = \zeta, \text{ and } \lim_{x\to\infty} \widehat{v}^{(n)}(x) = 0 \; \forall\, n \geq 2. \tag{11}$$

Literally speaking, 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, this theorem ensures that our results also apply to all regularly varying inverse demand and innovation cost functions.

We now show a lemma that will prove useful throughout the paper.

Lemma 1. *If $v(x) = x^{\zeta} l(x)$ is a smoothly varying function, then*

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

Proof. By the definition of smoothly varying function, the following equations must hold:

$$\lim_{x\to\infty} \frac{xv'(x)}{v(x)} = \lim_{x\to\infty}\left(\zeta + x\frac{l'(x)}{l(x)}\right) = \zeta \tag{12}$$

$$\lim_{x\to\infty} \frac{x^2 v''(x)}{v(x)} = \lim_{x\to\infty}\left(\zeta(\zeta-1) + 2\zeta x\frac{l'(x)}{l(x)} + x^2\frac{l''(x)}{l(x)}\right) = \zeta(\zeta-1). \tag{13}$$

Equation (12) implies that $\lim_{x\to\infty} x\frac{l'(x)}{l(x)} = 0$, therefore Equation (13) implies that $\lim_{x\to\infty} x^2\frac{l''(x)}{l(x)} = 0$. □

We now formally state our assumption on the demand and innovation cost functions as follows.

Assumption 1. *The inverse demand function of each variety can be written as $p = D(q;A) \equiv q^{-\frac{1}{\sigma}}Q(q;A)$, where $\sigma > 1$ and $\lim_{q\to\infty} Q(q;A) = C_Q > 0$. The innovation cost function can be written as $k(\varphi) = \gamma V(\varphi) \equiv \gamma\varphi^{\beta}L(\varphi)$, where $\beta > 1$ and $\lim_{\varphi\to\infty} L(\varphi) = C_L > 0$.*

Both Q and L are slowly varying functions because they have positive limits at infinity. Assumption 1 thus implies that both the demand and the innovation cost functions are regularly varying. Without loss of generality, we work with the *smoothly varying representations* of these functions following Theorem 1.8.2 of Bingham et al. (1989).

Assumption 1 essentially requires the demand to be asymptotically CES, but the admissible class of demand is actually more general than it seems at first glance. Needless to say, this includes the CES demand. As shown in Table 1, several important classes of demand functions with variable demand elasticity also satisfy this assumption.[11] For example, Assumption 1 includes several demand classes that exhibit "manifold invariance" (Mrázová and Neary 2017),[12] including Bipower

[11] The details are provided in Appendix A.1.

[12] A demand manifold depicts a 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.

Demand class	Functional form	Inverse demand
Bipower Direct	$q=\widehat{a}p^{-\nu}+ap^{-\sigma}\equiv\boldsymbol{q}(p)$ $\sigma>1, \sigma>\nu, a>0$	$p=q^{-\frac{1}{\sigma}}\left(\widehat{a}\left[\boldsymbol{q}^{-1}(q)\right]^{\sigma-\nu}+a\right)^{\frac{1}{\sigma}}$
Pollak (HARA)	$q=\widehat{a}+ap^{-\sigma}$ $\sigma>1, a>0$	$p=q^{-\frac{1}{\sigma}}a^{\frac{1}{\sigma}}\left(1-\frac{\widehat{a}}{q}\right)^{-\frac{1}{\sigma}}$
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	$q=ap^{r-1}+\widehat{a}p^{\frac{r}{2}-1}\equiv\boldsymbol{q}(p)$ $\sigma\equiv1-r>1, a>0$	$p=q^{\frac{1}{r-1}}\left(a+\widehat{a}\left[\boldsymbol{q}^{-1}(q)\right]^{-\frac{r}{2}}\right)^{\frac{1}{1-r}}$
Bipower Inverse	$p=\widehat{a}q^{-\nu}+aq^{-\frac{1}{\sigma}}$ $\sigma>1, \nu>1/\sigma, a>0$	$p=q^{-\frac{1}{\sigma}}\left(\widehat{a}q^{\frac{1}{\sigma}-\nu}+a\right)$
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

Direct demand, Bipower Inverse demand, Pollak Family demand (Pollak 1971, which is equivalent to the HARA [Hyperbolic Absolute Risk Aversion] preference [Merton 1971; Zhelobodko et al. 2012]), PIGL (Price-Independent Generalized Linear) demand (Muellbauer 1975), QMOR (Quadratic Mean of Order r) expenditure function (Diewert 1976; Feenstra 2018), and CEMR (Constant Elasticity of Marginal Revenue) demand. It also includes CREMR (Constant Revenue Elasticity of Marginal Revenue) demand (Mrázová, Neary, and Parenti 2017).[13]

As we will show shortly that there are one-to-one mappings at the tails between $\gamma\to0$ and $\varphi\to\infty$ and between $\varphi\to\infty$ and $q\to\infty$, the requirement of $\sigma>1$ is needed to ensure that the demand is consistent with monopoly pricing at these tails. Note that the CARA (Constant Absolute Risk Aversion) demand is excluded because its price elasticity tends to 0 when q goes to infinity.[14] Linear demand is also excluded because q is a finite value when $p=0$. Put differently, the linear demand is inconsistent with power laws as it never generates unbounded firm sales.

[13] Mrázová, Neary, and Parenti (2017) have shown that CREMR is the only consistent demand class in a monopolistic competitive framework when both the productivity and sales distributions are required to be "general power functions". As will be shown shortly, Assumption 1 leads to power laws for both productivity and sales distribution. Nevertheless, it is worth noting that distributions with power-law tails are not necessarily "general power functions", whereas general power functions do not necessarily exhibit power laws in their tails. Thus, neither our framework nor Mrázová, Neary, and Parenti's (2017) is a subset of the other.

[14] To see this, observe that the CARA demand can be written as $q=a-b\ln p$, where $a>0$, $b>0$. Its price elasticity equals b/q. This implies that the monopoly for each variety chooses a finite q even when its productivity φ tends to infinity.

The assumption on the innovation cost function parallels that on the inverse demand function. Obviously, simple power functions are included, but general polynomial functions are also included.

2.2.3 Equilibrium productivity and firm-size distributions

We solve the model backwards. For any given φ, the first- and second-order conditions for an interior solution q from (1) in the production stage are

$$p'q + p - \varphi^{-1} = 0 \tag{14}$$

$$p''q + 2p' < 0. \tag{15}$$

These imply that $|\epsilon(q)| \equiv -p/(qp') > 1$, and $\mu(q) \equiv -(p''q)/p' < 2$. Namely, 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 only regulates the inverse demand $p = D(q; A)$ for large values of q. As there is no guarantee that the profit function will be concave in the entire domain of q, there may exist corner solutions to the profit-maximization problem or multiple local optima satisfying (14) and (15). As we are concerned with the right tail of the firm size distribution (in terms of sales revenue $s = D(q; A)q$), what is relevant is large values of q. This is because by Assumption 1, $\lim_{q\to\infty} s(q) = \lim_{q\to\infty} q^{1-\frac{1}{\sigma}} Q(q; A) = \infty$.

Using Assumption 1, we can rewrite (14) and (15) as

$$\varphi = q^{\frac{1}{\sigma}} \left[Q \times \left(1 - \frac{1}{\sigma} + q\frac{Q'}{Q} \right) \right]^{-1} \tag{16}$$

$$q^{-\frac{1}{\sigma}-1} Q \left[-\frac{1}{\sigma}\left(1 - \frac{1}{\sigma}\right) + 2\left(1 - \frac{1}{\sigma}\right) q\frac{Q'}{Q} + q^2 \frac{Q''}{Q} \right] \equiv \pi_{qq}(q, \varphi) < 0. \tag{17}$$

Let q^* be a solution to (16). For large values of φ, we can show that q^* exists and is unique. Moreover, q^* strictly increases in φ and $\lim_{\varphi\to\infty} q^*(\varphi) = \infty$. The following assumption rules out the corner solution.

Assumption 2. *The inverse demand function D is such that the revenues around $\underline{q}$ remain finite. Namely, $\lim_{q\to\underline{q}} s(q) < \infty$.*

We have the following lemma.

Lemma 2. *Suppose that Assumption 1 holds. For sufficiently large φ, the interior solution $q^*(\varphi)$ that satisfies (16) exists and is unique. Moreover, $q^*(\varphi)$ strictly increases in φ and $\lim_{\varphi\to\infty} q^*(\varphi) =$*

∞. *If, in addition, Assumption 2 holds, then $q^*(\varphi)$ is the unique profit-maximizing quantity and* $\lim_{\varphi \to \infty} \pi(\varphi) = \infty$.

Proof. Lemma 1 implies that $q\frac{Q'}{Q}$ 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 (16) because the term in the bracket tends to a constant. Thus, q^* exists. However, there is a possibility that this firm with an arbitrarily large φ might choose a finite q such that the term $Q \cdot \left(1 - \frac{1}{\sigma} + q\frac{Q'}{Q}\right)$ tends to zero. Nevertheless, note that by plugging (16) into (1), we have

$$\pi(\varphi) = q^{1-\frac{1}{\sigma}} Q \left(\frac{1}{\sigma} - q\frac{Q'}{Q}\right) - \kappa_D.$$

Assumption 1 and Lemma 1 imply that if q becomes arbitrarily large as φ becomes arbitrarily large, then the profit also becomes arbitrarily large. However, if a finite 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. Thus, q^* is unique and $\lim_{\varphi \to \infty} q^*(\varphi) = \infty$. As a result, when φ (and hence q) becomes arbitrarily large, the second-order condition (17) is satisfied because of Lemma 1. Applying the implicit function theorem on (16), we have

$$\frac{dq^*}{d\varphi} = -\frac{\varphi^{-2}}{\pi_{qq}(q^*, \varphi)} > 0 \tag{18}$$

as $\pi_{qq}(q^*, \varphi) < 0$. Finally, the only concern that q^* is not the profit-maximizing quantity is that it might be dominated by a corner solution at $\underline{q}$.[15] For this concern to be valid, it requires that the profit tends to infinity as $q \to \underline{q}$. This, in turn, requires that $\underline{q}$ forms an asymptote of the demand curve so that $\lim_{q \to \underline{q}} s(q) = \infty$.[16] This possibility is ruled out by Assumption 2, and thus q^* is the unique profit-maximizing quantity. □

In the innovation stage, a firm chooses φ to maximize its profit. By the envelope theorem, the first-order condition of φ is

$$\frac{d\Pi(\varphi; \gamma)}{d\varphi} = \varphi^{-2} q^*(\varphi) - \gamma V'(\varphi) = 0, \tag{19}$$

[15]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 $\lim_{q \to q_0} p(q) = \infty$, which violates the law of demand.

[16]The Pollak demand with $\sigma > 0$, $a > 0$, and $\widehat{a} > 0$ is such an example. Here, the demand requires that $q > \widehat{a}$, and $s(q)$ being increasing (concave) in q when $q > \frac{\sigma}{\sigma-1}\widehat{a}$ ($q > \frac{2\sigma}{\sigma-1}\widehat{a}$). However, the optimal output degenerates to $\widehat{a}$ for all φ because $\lim_{q \to \widehat{a}} \pi(q) = \lim_{q \to \widehat{a}} \left(s(q) - \varphi^{-1} q\right) = \infty$.

and thus the optimal productivity φ satisfies

$$\gamma = \frac{q^*(\varphi)}{\varphi^2 V'(\varphi)}. \tag{20}$$

The associated second-order condition is

$$-2\varphi^{-3} q^*(\varphi) + \varphi^{-2} \frac{\partial q^*(\varphi)}{\partial \varphi} - \gamma V''(\varphi) < 0. \tag{21}$$

Similar to Sutton (1991), the innovation cost function must be sufficiently convex so that (21) holds. Lemma 3 below shows that the second-order condition holds for large φ if and only if $\beta > \sigma - 1$.

It is intuitive that a firm endowed with a higher R&D ability (smaller γ) invests more and obtains a higher productivity; as γ tends to 0 then the productivity tends to infinity. The following lemma establishes this intuition.

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

Proof. By plugging (16) into (20), we obtain

$$\gamma = \frac{\left[Q(\varphi)\left(1 - \frac{1}{\sigma} + q^*(\varphi)\frac{Q'(\varphi)}{Q(\varphi)}\right)\right]^\sigma}{L(\varphi)\left[\beta + \varphi\frac{L'(\varphi)}{L(\varphi)}\right]} \varphi^{-(\beta - \sigma + 1)}. \tag{22}$$

Using (16), (17), (18), and (22), we can restate the left-hand side of (21) as

$$-Q^\sigma \left(1 - \frac{1}{\sigma} + q\frac{Q'}{Q}\right)^\sigma \varphi^{\sigma - 3} \left[2 + \frac{1 - \frac{1}{\sigma} + q\frac{Q'}{Q}}{-\frac{1}{\sigma}\left(1 - \frac{1}{\sigma}\right) + 2\left(1 - \frac{1}{\sigma}\right) q\frac{Q'}{Q} + q^2 \frac{Q''}{Q}} + \frac{\beta(\beta - 1) + 2\beta\varphi\frac{L'}{L} + \varphi^2\frac{L''}{L}}{\beta + \varphi\frac{L'}{L}}\right]. \tag{23}$$

Assumptions 1 and 2 imply that Lemma 2 holds. Lemmas 1 and 2 and Assumption 1 imply that for large values of φ, (23) converges to $-C_Q^\sigma \left(\frac{\sigma-1}{\sigma}\right)^\sigma (\beta - \sigma + 1) \lim_{\varphi \to \infty} \varphi^{\sigma - 3}$ and is strictly negative if and only if $\beta > \sigma - 1$. As a result, for a firm with an arbitrarily small γ there exists a large φ, denoted as φ^*, satisfying (22) and the second-order condition (21).

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 (22) holds. Note that $V' > 0$ would be violated if $\lim_{\varphi \to \varphi_0} L(\varphi) = \infty$ for some finite φ_0. Hence, $L(\varphi)$

must be finite for all finite values of φ. Observe that by plugging (16) and (22) into (3) we have

$$\begin{aligned}\Pi =&\pi(\varphi) - \gamma V(\varphi)\\ =&Q^{\sigma} \cdot \left(1 - \frac{1}{\sigma} + q^* \frac{Q'}{Q}\right)^{\sigma} \left[\frac{\frac{1}{\sigma} - q^* \frac{Q'}{Q}}{1 - \frac{1}{\sigma} + q^* \frac{Q'}{Q}} - \frac{1}{\beta + \varphi \frac{L'}{L}}\right] \varphi^{\sigma-1} - \kappa_D.\end{aligned}$$

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 also finite. In contrast, the profit becomes arbitrarily large for an arbitrarily large φ. Thus, a finite φ would not be the solution to (22) when γ becomes arbitrarily small, and hence φ^* is the unique solution and is denoted as $\widetilde{\varphi}(\gamma)$.

For large values of φ, the derivative of the right-hand side of Equation (20) is

$$\begin{aligned}&\left(\frac{1}{\varphi^2 V'(\varphi)}\right)^2 \left(\varphi^2 V'(\varphi) \frac{\partial q^*(\varphi)}{\partial \varphi} - 2V'(\varphi) q^*(\varphi) \varphi - V''(\varphi) q^*(\varphi) \varphi^2\right)\\ =&\frac{1}{V'(\varphi)} \left(-2\varphi^{-3} q^*(\varphi) + \varphi^{-2} \frac{\partial q^*(\varphi)}{\partial \varphi} - \gamma V''(\varphi)\right) \text{ by equation (20).}\\ <&0,\end{aligned}$$

where the last inequality holds by (21). 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 in Melitz (2003), the existence of a fixed cost of production $\kappa_D > 0$ gives rise to firm selection. This means that a successful entrant must be capable enough to obtain a high enough productivity to survive. As $\Pi(\widetilde{\varphi}(\gamma); \gamma) = \pi(\widetilde{\varphi}(\gamma)) - \gamma V(\widetilde{\varphi}(\gamma))$, $d\Pi/d\gamma = -V < 0$ by the envelope theorem. Thus, any firm produces if and only if $\gamma \leq \gamma_D$, where γ_D is defined by[17]

$$\Pi(\widetilde{\varphi}(\gamma_D), \gamma_D) = \pi(\widetilde{\varphi}(\gamma_D)) - \gamma_D V(\widetilde{\varphi}(\gamma_D)) = 0. \tag{24}$$

We now complete the description of the equilibrium conditions. An equilibrium is defined by first-order conditions for q and φ, (16) and (20), the cutoff condition (24), and the free entry condition (4).

Now we are ready to show how the power laws for productivity and firm size (in terms of sales revenue $s = pq$) arise. By a change of variables, the p.d.f. of productivity is

$$g(\varphi) = \frac{f(\widetilde{\gamma}(\varphi))}{F(\gamma_D)} J(\varphi),$$

[17]The following definition of γ_D implicitly assumes continuity of Π in γ. Note that smooth variation guarantees that all relevant functions are continuous for large values of q and φ and small values of γ. However, even when Π is discontinuous in some large values of γ, a cutoff γ_D can still be well-defined as long as Π strictly decreases in γ.

where the Jacobian $J(\varphi)$ is given by (See Appendix A.2)

$$\begin{aligned} J(\varphi) =& \left|\frac{\partial\widetilde{\gamma}(\varphi)}{\partial\varphi}\right| = \left|\frac{\partial}{\partial\varphi}\frac{q^*(\varphi)}{\varphi^2 V'(\varphi)}\right| \\ =& \frac{Q^\sigma}{L}\frac{\left(1-\frac{1}{\sigma}+q^*\frac{Q'}{Q}\right)^\sigma}{\beta+\varphi\frac{L'}{L}} \cdot \left[2+\frac{\beta(\beta-1)+2\beta\varphi\frac{L'}{L}+\varphi^2\frac{L''}{L}}{\beta+\varphi\frac{L'}{L}}\right. \\ &+ \left.\frac{1-\frac{1}{\sigma}+q^*\frac{Q'}{Q}}{-\frac{1}{\sigma}\left(1-\frac{1}{\sigma}\right)+2\left(1-\frac{1}{\sigma}\right)q^*\frac{Q'}{Q}+(q^*)^2\frac{Q''}{Q}}\right] \cdot \varphi^{-(\beta-\sigma+1)-1}. \end{aligned} \tag{25}$$

For power laws, we consider how the density g behaves as φ grows arbitrarily large. The following Proposition 1 shows the conditions under which power laws for productivity and firm size arise.

Proposition 1. *Suppose that Assumptions 1 and 2 hold. Also suppose that $f(\gamma) = \gamma^\alpha m(\gamma)$ where $\alpha > -1$ and $m(\gamma)$ is slowly varying around the origin, 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; for the detailed proof, see 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 and 2, Lemma 3 holds. For the free entry condition to hold, the expected profit must be finite, i.e., $\int_0^{\gamma_D}\Pi(\widetilde{\varphi}(\gamma);\gamma)\,dF(\gamma) < \infty$. Whether this integral is finite depends on small γ (that is, the high-capability firms), and what matters is essentially the orders of demand, σ, innovation cost function, β, and the distribution of 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 the Jacobian (25). First note that by Assumption 1 and Lemma 2, the slowly varying functions $Q(q;A)$ and $L(\varphi)$ converge to some constants C_Q and C_L, respectively. Also, by Lemma 1, smoothly varying demand and innovation cost imply 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, the Jacobian $J(\varphi)$ is regularly varying in φ with degree $-(\beta-\sigma+1)-1 < 0$. If $f(\gamma)$ is regularly varying with degree α around 0, then $f(\widetilde{\gamma}(\varphi))$ is regularly varying as φ goes to infinity with degree $-\alpha(\beta-\sigma+1)$. Thus, the productivity distribution exhibits a power law with a tail index $(\alpha+1)(\beta-\sigma+1)$. Following the same reasoning, 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 can emerge from a generalized environment of a standard general equilibrium model. Regularly varying demand and innovation cost functions together entail the reciprocal relationship between γ and φ (and hence s). As discussed in Section

2.2.2, regularly varying demand includes a large class of non-CES and non-homothetic preferences. The condition that $f(\gamma)$ being regularly varying around 0 further ensures a sufficient mass of highly capable firms (small γ). Accordingly, the power laws emerge via the power law change of variable mechanism. Note that the right tail of $f(\gamma)$ is immaterial to our finding.

The distribution $f(\gamma)$ being regularly varying around 0 is also more general than it seems. This distribution class includes, for examples, uniform and Beta, and with appropriate truncation to the right, Gamma and Weibull distributions. A longer list of this distribution class can be obtained in Table C1 in Hsu (2012). Compared with Geerolf's (2017) power-law result, Proposition 1 is more general as his condition $\lim_{\gamma \to 0} f(\gamma) = K > 0$ is a special case here, i.e., a particular case of slowly varying functions.

Proposition 1 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 distribution or the two-piece distribution by Nigai (2017). We now turn to a general open economy to investigate whether and how power laws hold in that environment. In particular, we will see how country-specific parameters affect the tail indices of the two distributions.

2.3 Power Laws in Open Economy

2.3.1 Model setup in open economy

There are $n+1$ asymmetric countries 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 (and hence $\{\sigma_i, \beta_i, \alpha_i\}$ can also be country-specific). Similar to the closed-economy case, Assumptions 1 and 2 are assumed to hold with $C_{Q,i}$ and $C_{L,i}$ also allowed to be country-specific.

The timing is identical to the closed economy case, except that in the production stage each firm can determine whether to export, and, if yes, the price and quantity of exported goods. 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

$$\pi_{ij}(\varphi) = p_{ij}q_{ij} - \tau_{ij}w_i\varphi^{-1}q_{ij} - \kappa_{ij}, \tag{26}$$

where $\tau_{ij} \geq 1$ denotes the variable trade cost, κ_{ij} denotes the fixed selling cost from i to j, and w_i denotes the wage in country i. Then, a firm produces if and only if

$$\Pi_i(\varphi) = \left[\sum_j \pi_{ij}(\varphi)\right] - \kappa_{D,i} \geq 0.$$

2.3.2 Equilibrium and power laws for productivity and firm size

Given φ, the first-order condition for q_{ij} is similar to (16) and is given as follows:

$$\varphi = w_i \tau_{ij} q_{ij}^{\frac{1}{\sigma_j}} \left[Q_j \times \left(1 - \frac{1}{\sigma_j} + q_{ij} \frac{Q_j'}{Q_j}\right)\right]^{-1}. \tag{27}$$

It is straightforward to see that Lemma 2 holds here. That is, we have $\lim_{\varphi\to\infty} q_{ij}^* (\varphi) = \infty$ and $\lim_{\varphi\to\infty} \pi_{ij} (\varphi) = \infty$. Note that 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.

Observe that for a given γ, the first-order condition is

$$\gamma = \frac{\sum_j \mathbb{I}_{ij} \tau_{ij} q_{ij}^* (\varphi)}{\varphi^2 V_i' (\varphi)}, \tag{28}$$

where $\mathbb{I}_{ij} = \{0, 1\}$ is the indicator function that indicates whether the firm with γ at country i sells to country j. By combining (27) with (28), we can rewrite (28) as

$$\gamma = \frac{\sum_j \mathbb{I}_{ij} \tau_{ij}^{1-\sigma_j} w_i^{-\sigma_j} Q_j^{\sigma_j} \cdot \left(1 - \frac{1}{\sigma_j} + q_{ij}^* \frac{Q_j'}{Q_j}\right)^{\sigma_j} \cdot \varphi^{\sigma_j - \beta_i - 1}}{L_i \cdot \left(\beta_i + \varphi \frac{L_i'}{L_i}\right)}. \tag{29}$$

Each component in the numerators of (29) is similar to those in the closed-economy case. Thus, for an arbitrarily small γ, there exists a corresponding large φ such that (29) holds with $\mathbb{I}_{ij} = 1$ for all j. The same proof in Lemma 3 rules out other potential solutions. Therefore, the conclusion of Lemma 3 also holds here. That is, if $\beta_i > \sigma_j - 1$ for all i and j, the optimal choice of φ exists and is unique for firms with sufficiently small γ.[18] Denote this optimal choice as $\varphi_i^* = \widetilde{\varphi}_i (\gamma)$; we have $\widetilde{\varphi}_i' (\gamma) < 0$ and $\lim_{\varphi\to\infty} \widetilde{\gamma}_i (\varphi) = 0$.

To ensure the expected profit of entrants in each country remains finite, $\alpha_i > -1$ and $\beta_i > \frac{\alpha_i+2}{\alpha_i+1} (\max_j \sigma_j - 1)$ are required.[19] Since we are concerned with the tail behavior of the producti-

[18] The second-order condition for productivity choice holds if $\beta_i > \sigma_j - 1$ holds for all i and j. To see this, note that the second-order condition is such that $\varphi^{-2} \sum_j \mathbb{I}_{ij} \tau_{ij} \frac{\partial q_{ij}^*(\varphi)}{\partial \varphi} - 2\varphi^{-3} \sum_j \mathbb{I}_{ij} \tau_{ij} q_{ij}^* (\varphi) - \gamma V_i'' (\varphi) < 0$. Following the same proof in Lemma 3, for large values of φ the left-hand side of the second-order condition converges to $-\sum_j (\beta_i - \sigma_j + 1) \left[\mathbb{I}_{ij} w_i^{-\sigma_j} \tau_{ij}^{1-\sigma_j} C_{Q_j}^{\sigma_j} \left(\frac{\sigma_j - 1}{\sigma_j}\right)^{\sigma_j} \lim_{\varphi\to\infty} \varphi^{\sigma_j - 3}\right]$, and is strictly negative if $\beta_i > \sigma_j - 1$ for all i and j. Note that this condition is sufficient but not necessary, but if all countries are symmetric in both β_i and σ_j, the condition becomes necessary.

[19] To see this, first extract $\varphi^{\max_k \sigma_k - \beta_i - 1}$ from the summation term in (29) to obtain $\gamma = \left(\sum_j T_{ij} \varphi^{\sigma_j - \max_k \sigma_k}\right) \times \varphi^{\max_k \sigma_k - \beta_i - 1}$, where $T_{ij} = \mathbb{I}_{ij} \tau_{ij}^{1-\sigma_j} w_i^{-\sigma_j} Q_j^{\sigma_j} \cdot \left(1 - \frac{1}{\sigma_j} + q_{ij}^* \frac{Q_j'}{Q_j}\right)^{\sigma_j} / (\beta_i L_i + \varphi L_i')$. Because each T_{ij} converges to some positive constant and $\sigma_j - \max_k \sigma_k \le 0$ with equality holding for at least one term, $\sum_j T_{ij} \varphi^{\sigma_j - \max_k \sigma_k}$ converges to a positive constant as φ goes to infinity. Therefore, we can write $\varphi \propto \gamma^{\frac{1}{\max_k \sigma_k - \beta_i - 1}}$. Rearranging (27)

vity distribution, it suffices to focus on the right-most piece of the productivity distribution. The corresponding Jacobian is obtained by differentiating Equation (28), i.e.,

$$J_i(\varphi) = \left| \underbrace{\frac{\partial \widetilde{\gamma}_i(\varphi)}{\partial \varphi}}_{-} \right| = -\sum_{j=0}^{n} \frac{\partial}{\partial \varphi} \frac{\tau_{ij} q_{ij}^*(\varphi)}{\varphi^2 V_i'(\varphi)}. \tag{30}$$

Obviously, each component of Equation (30) is similar to Equation (25), and is regularly varying with degree $-(\beta_i + 2 - \sigma_j)$. Following the same argument to Proposition 1, the productivity distribution exhibits a power law with the tail index $(\alpha_i + 1)(\beta_i + 1 - \max_j \sigma_j)$.[20]

We now turn to the firm size distribution. Denote s_{ij} as a firm's sales from i to j and thus the firm size of the firms that export 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 similar procedure to the proof of Proposition 1 and Footnote 20, firm size distribution also follows a power law with the tail index $\frac{(\alpha_i+1)(\beta_i+1-\max_j \sigma_j)}{\max_j \sigma_j - 1}$. The above derivation leads to the following proposition.

Proposition 2. *Suppose that Assumptions 1 and 2 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 $m_i(\gamma)$ is slowly varying around the origin, 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 the tail index of failure probability α_i, the innovation technology parameter β_i and the largest σ_j among all destination countries. As a larger σ_j generally implies a larger elasticity of substitution and larger price elasticity, the destination with the largest σ_j entails the largest responsiveness of firm sales to productivity changes. Thus, the destination with the largest σ_j plays the dominant role in determining the tail indices of every source country. The same logic applies

yields $q_{ij} \propto \varphi^{\sigma_j}$, and by (26) the profit gross of fixed cost $\pi_{ij} + \kappa_{ij}$ is proportional to $\varphi^{\sigma_j - 1}$. Plugging these results into the expected profit $\int_0^{\overline{\gamma}} \Pi_i(\gamma) dF(\gamma) = \int_0^{\overline{\gamma}} \sum_j (\pi_{ij} + \kappa_{ij}) dF(\gamma) - \int_0^{\overline{\gamma}} \sum_j \kappa_{ij} dF(\gamma) - \kappa_{D,i}$ and utilizing the relationship that $\varphi \propto \gamma^{\frac{1}{\max_k \sigma_k - \beta_i - 1}}$, the remainder of the proof follows that in Step 1 of Appendix A.2. The idea is that $\int_0^{\overline{\gamma}} \sum_j (\pi_{ij} + \kappa_{ij}) dF(\gamma)$ is an integral of a power function of γ, and hence some constraint on the exponent is needed to ensure convergence.

[20] To see this, extract $\varphi^{-(\beta_i+1-\max_k \sigma_k)}$ from the right-hand side of (30) so that $J_i(\varphi)$ can be written as the product of $\varphi^{-(\beta_i+1-\max_k \sigma_k)-1}$ and a summation term. Note that each element in the summation term is regularly varying in φ with degree $\sigma_j - \max_k \sigma_k \leq 0$, with equality holding for at least one term. Therefore, the summation term converges to a constant. Then we combine with $f_i(\gamma)$ to obtain the tail index $(\alpha_i + 1)(\beta_i + 1 - \max_k \sigma_k)$.

[21] Note that 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 (30) are not literally independent random variables.

analogously for the firm size distribution. Proposition 2 implies that opening up to trade causes the tails of both productivity and firm-size distributions in each country to (weakly) fatten. A similar theoretical prediction has been provided by di Giovanni, Levchenko, and Rancière (2011) and is also empirically tested in the same paper.

3 The Effects of Trade on Productivity Distribution

This section analyzes the effects of trade. In particular, we focus on how trade costs affect 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. Also for tractability, we assume a power function for the innovation cost: $k = \gamma\varphi^{\beta}$. 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.

3.1 Equilibrium

Given the functional-form assumptions on the inverse demand and innovation cost, Assumption 1 is satisfied. Moreover, the profit-maximizing solution of $q^*(\varphi)$ and $\widetilde{\varphi}(\gamma)$ must be interior and unique given by the relevant first- and second-order conditions. Thus, Assumption 2 is no longer needed.

To solve the model, we start with the production stage. The optimal quantity that a firm produces for the domestic market (denoted by subscript D) and the foreign market (denoted by subscript X) are respectively given by

$$\begin{aligned} q_D(\varphi) =& \frac{N}{P^{1-\sigma}}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}\varphi^{\sigma} \\ q_X(\varphi) =& \tau^{-\sigma}\frac{N}{P^{1-\sigma}}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}\varphi^{\sigma}. \end{aligned}$$

Accordingly, the operating profits in the domestic and each foreign markets are

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

In the innovation stage, a firm decides its productivity level according to whether it serves the foreign market or not. Respectively, the profits of a non-exporting firm and an exporting one are therefore

$$\Pi_D(\varphi) = \pi_D(\varphi) - \gamma\varphi^\beta \tag{31}$$

$$\Pi_X(\varphi) = \pi_D(\varphi) + n\pi_X(\varphi) - \gamma\varphi^\beta. \tag{32}$$

The productivity level is such that

$$\widetilde{\varphi}(\gamma) = \begin{cases} \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{1}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^\sigma}{\beta}\right]^{\frac{1}{\beta-\sigma+1}} \gamma^{-\frac{1}{\beta-\sigma+1}} & \text{for non-exporting firms} \\ \phi\left(\frac{N}{P^{1-\sigma}}\right)^{\frac{1}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^\sigma}{\beta}\right]^{\frac{1}{\beta-\sigma+1}} \gamma^{-\frac{1}{\beta-\sigma+1}} & \text{for exporting firms} \end{cases}, \tag{33}$$

where

$$\phi \equiv \left(1 + n\tau^{1-\sigma}\right)^{\frac{1}{\beta-\sigma+1}}. \tag{34}$$

Since exporting decisions are made after the firm has invested in its productivity, 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 nonexporting ones. Since countries are symmetric, the argument in Footnote 18 implies that (33) is optimal if and only if $\beta > \sigma - 1$.

By substituting Equation (33) into (31) and (32), the profits for both non-exporting and exporting firms become

$$\Pi_D(\gamma) = \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^\sigma}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right)\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} - \kappa_D \tag{35}$$

$$\Pi_X(\gamma) = \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^\sigma}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\beta}{\sigma-1}\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1} - \phi^\beta\right]\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} - \kappa_D - n\kappa_X. \tag{36}$$

Observe that the gross profits are proportional to $\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}$. The cutoff types are thus obtained as

$$\gamma_D = \left[\kappa_D^{-1}\left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^\sigma}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right)\right]^{\frac{\beta-\sigma+1}{\sigma-1}} \tag{37}$$

$$\gamma_X = \left\{n^{-1}\kappa_X^{-1}\left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^\sigma}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\beta}{\sigma-1}\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1} - \phi^\beta - \left(\frac{\beta-\sigma+1}{\sigma-1}\right)\right]\right\}^{\frac{\beta-\sigma+1}{\sigma-1}}, \tag{38}$$

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 $\gamma \leq \gamma_X$. Notice that, if $\gamma_D \leq \gamma_X$, then all operating firms choose high productivity levels and become exporters. Similar to the literature, we consider only the case of $\gamma_X < \gamma_D$ because all firms exporting is counter-factual.

From (37) and (38), we have

$$\delta \equiv \frac{\gamma_X}{\gamma_D} = \left(\frac{\kappa_D}{n\kappa_X}\right)^{\frac{\beta-\sigma+1}{\sigma-1}} \left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right]^{\frac{\beta-\sigma+1}{\sigma-1}}. \tag{39}$$

To ensure that $\gamma_X < \gamma_D$, so that there are both exporters and non-exporters in the economy, we assume that $\delta < 1$, which requires trade frictions κ_X or τ to be sufficiently large relative to the fixed cost of production κ_D.

In the entry stage, each firm decides whether to enter the market. The free entry condition implies that the equilibrium entry is such that the expected profit of entry equals the entry cost for each firm,

$$E(\Pi) = \int_0^{\gamma_X} \Pi_X(\gamma)\, dF(\gamma) + \int_{\gamma_X}^{\gamma_D} \Pi_D(\gamma)\, dF(\gamma) = \kappa_e. \tag{40}$$

An equilibrium is accordingly defined by (33), (37), (38), (40) and the aggregate price

$$\begin{aligned} P^{1-\sigma} =& M_e \left[\int_{\gamma_X}^{\gamma_D} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \widetilde{\varphi}(\gamma)^{\sigma-1} dF(\gamma) + \int_0^{\gamma_X} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \widetilde{\varphi}(\gamma)^{\sigma-1} dF(\gamma)\right] \\ &+ nM_e \int_0^{\gamma_X} \tau^{1-\sigma} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \widetilde{\varphi}(\gamma)^{\sigma-1} dF(\gamma) \end{aligned} \tag{41}$$

where M_e denotes the mass of entrants paying the entry cost.

The aggregate price 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 foreign exporters. Note that by (33), there is a jump in the function $\widetilde{\varphi}(\gamma)$ at γ_X.

In Appendix A.3, we provide the derivation of the following equilibrium outcome. First, an equilibrium exists and is unique. In equilibrium, γ_D is given by

$$\kappa_D \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} \left\{\Gamma_D + \left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\Gamma_X\right\} - \kappa_D F(\gamma_D) - n\kappa_X F(\gamma_X) = \kappa_e, \tag{42}$$

where $\Gamma_z \equiv \int_0^{\gamma_z} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} dF(\gamma)$ for $z \in \{D, X\}$. Note that $\frac{\Gamma_z}{F(\gamma_z)}$ is proportional to the average productivity of firms in $(0, \gamma_z)$. Therefore, Γ_z measures the contribution of the productivities in $(0, \gamma_z)$ to the expected profit of an entrant.

The price index and mass of entrant are

$$P^{1-\sigma} = N\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}\beta^{-1}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{\frac{\beta-\sigma+1}{\beta}}\kappa_D^{-\frac{\beta-\sigma+1}{\beta}}\gamma_D^{-\frac{\sigma-1}{\beta}} \tag{43}$$

$$M_e = \frac{N}{\kappa_e+\kappa_D F(\gamma_D)+n\kappa_X F(\gamma_X)}\frac{\sigma-1}{\beta\sigma}\left(\frac{\beta-\sigma+1}{\sigma-1}\right). \tag{44}$$

The equilibrium productivity is

$$\widetilde{\varphi}(\gamma) = \begin{cases} \kappa_D^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} & \text{if } \gamma\in(\gamma_X,\gamma_D] \\ \phi\kappa_D^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} & \text{if } \gamma\in[0,\gamma_X] \end{cases}, \tag{45}$$

and the associated density is

$$g(\varphi) = \begin{cases} \frac{\left[\frac{(\sigma-1)\kappa_D}{\beta-\sigma+1}\right]^{\frac{\beta-\sigma+1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta}}(\beta-\sigma+1)}{F(\gamma_D)} f\left(\left[\frac{(\sigma-1)\kappa_D}{\beta-\sigma+1}\right]^{\frac{\beta-\sigma+1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta}}\varphi^{-(\beta-\sigma+1)}\right)\varphi^{-(\beta-\sigma+1)-1} & \text{if } \varphi\in\left[\varphi_D,\varphi_X^-\right) \\ 0 & \text{if } \varphi\in\left[\varphi_X^-,\varphi_X^+\right) \\ \frac{\left[\frac{(\sigma-1)\kappa_D}{\beta-\sigma+1}\right]^{\frac{\beta-\sigma+1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta}}(\beta-\sigma+1)}{F(\gamma_D)} f\left(\phi^{\beta-\sigma+1}\left[\frac{(\sigma-1)\kappa_D}{\beta-\sigma+1}\right]^{\frac{\beta-\sigma+1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta}}\varphi^{-(\beta-\sigma+1)}\right)\phi^{\beta-\sigma+1}\varphi^{-(\beta-\sigma+1)-1} & \text{if } \varphi\in\left[\varphi_X^+,\infty\right) \end{cases}, \tag{46}$$

where $\varphi_D \equiv \left[\frac{(\sigma-1)\kappa_D}{\beta-\sigma+1}\right]^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\gamma_D^{-\frac{1}{\beta-\sigma+1}}$, $\varphi_X^- \equiv \lim_{\gamma\to\gamma_X+}\widetilde{\varphi}(\gamma) = \left[\frac{(\sigma-1)\kappa_D}{\beta-\sigma+1}\right]^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\gamma_X^{-\frac{1}{\beta-\sigma+1}}$, and $\varphi_X^+ \equiv \lim_{\gamma\to\gamma_X-}\widetilde{\varphi}(\gamma) = \phi\left[\frac{(\sigma-1)\kappa_D}{\beta-\sigma+1}\right]^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\gamma_X^{-\frac{1}{\beta-\sigma+1}}$.

Equation (45) shows that equilibrium productivity not only depends on γ, but also on firm selection; we analyze these in Section 3.2. The power-law results in the previous section hold here, as the functional-form assumptions here on D and V satisfy Assumption 1.[22]

The following proposition shows the conditions under which an equilibrium exists and is unique.

Proposition 3. *Suppose that $\alpha > -1$, $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$ and that $\delta < 1$ where δ is defined by (39). 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 nonexporters in the economy.*

Proof. Applying the symmetric-country assumption to Proposition 2 implies that $E(\Pi) < \infty$ under $\alpha > -1$ and $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$. When $\delta < 1$, $E(\Pi)$ can be expressed as the left-hand

[22] Moreover, the productivity distribution further belongs to a general functional class: the General Power Function (GPF) class (Mrázová, Neary, and Parenti 2017). A distribution of φ belongs to the GPF class if its c.d.f. takes a form such that $H\left(\theta_0+\theta_1\varphi^{\theta_2}\right)$, where θ_0, θ_1, and θ_2 are constants, and $H(\cdot)$ is a monotonic function. Several frequently used skewed distributions belong to this functional class, including Pareto, lognormal, Frechet, and Inverse-Weibull distributions. Our framework thus resonates well with Mrázová, Neary, and Parenti (2017) not only because it provides a microfoundation to the GPF class, but also because it narrows down to those with a power-law tail.

side of (42), and is increasing in γ_D as shown in Appendix A.3. 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 (42) holds. □

3.2 The Comparative Statics of Trade Costs

We explore some key comparative statics on the productivity distribution, and summarize the results in the following proposition.

Proposition 4. *Assume that the conditions of Proposition 3 hold. We have the following comparative statics.*

(1) An increase in κ_D results in a lower γ_D, a higher γ_X, and a higher φ for all γ. The new productivity distribution FOSD (First-Order Stochastically Dominates) the old distribution.

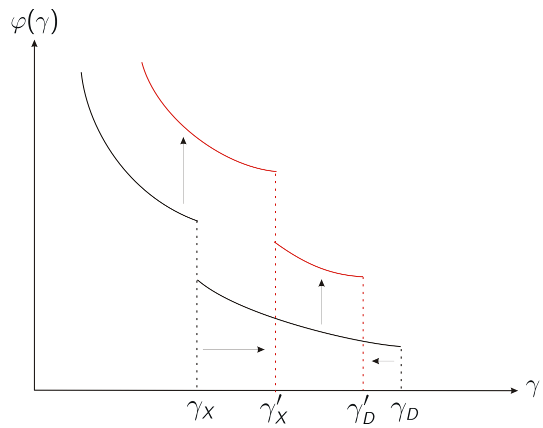
(2) An increase in κ_X results in a higher γ_D and a lower γ_X. Moreover, productivity φ increases for any exporting (non-exporting) firm which remains exporting (non-exporting) after the shock.

(3) 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.

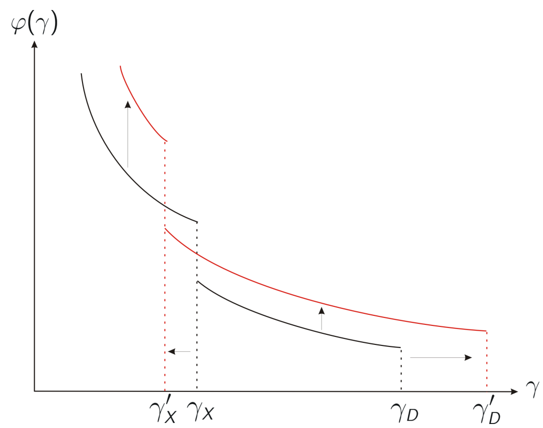
Proof. See Appendix A.4. □

In Figures 2a to 2c we illustrate the comparative statics of the three parameters in Proposition 4. Point 1 states that an increment of fixed production cost raises the average productivity by shifting the whole distribution rightwards. A higher fixed production cost means that firms are less likely to survive even in the domestic market. Therefore, on the one hand the surviving firms must be more efficient in innovation. On the other hand, because fewer firms operate in the market due to a higher fixed production cost, the foreign firms thus face less competition in the domestic market and have more incentive to export even if it is not so efficient. Therefore, γ_X increases. Moreover, a higher κ_D and γ_D together creates a substitution effect by raising the aggregate price; hence each surviving firm has more incentive to acquire a higher productivity. As a result, the productivity distribution shifts to the right from both the extensive and the intensive margins.

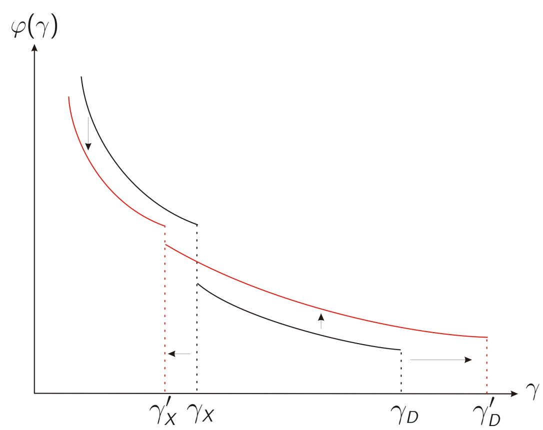
Regarding the intuitions behind Point 2, an increase in κ_X makes exporting more difficult. For nonexporting firms which remain nonexporting, they face less import competition, and thus have



(a) The effect of an increment of κ_D



(b) The effect of an increment of κ_X



(c) The effect of an increment of τ

Figure 2: Comparative statics

more incentive to invest in productivity to extract the gains from the effectively larger market size facing them. Similarly, for exporting firms which remain exporting, they invest more in productivity not only because they face less import competition in their home market, but also because they face less competition in their foreign markets. The fact that γ_D increases implies a more lenient selection, and so some entrants who were not able to survive before can now survive with positive innovation. There is a set of firms which switch from exporting to not exporting, but their changes in productivity are ambiguous because on the one hand, they seek to invest less due to the loss of the foreign market, but they also seek to invest more for the lesser import competition. In sum, if the change in κ_X is infinitesimal, then the overall productivity distribution FOSD the previous one almost surely.

For the mechanism behind Point 3, first note that an increase in τ adds friction to exporting, and its effect on selection and export cutoffs (γ_D and γ_X) is similar to the case of an increase in κ_X. For the nonexporting firms which remain nonexporting, they invest more due to less import competition. Similar to the case of an increase in κ_X, for exporting firms which remain exporting, they have incentives to invest more in productivity because of less competition in both domestic and foreign markets. But since their effective market size shrinks due to increased friction, this force dominates and so their productivities actually reduce. For the firms which switch from exporting to not exporting, their productivities decrease because the loss of the foreign market.

It is interesting and important to note that Proposition 4 actually implies the same effects of trade on the selection and export cutoffs as Melitz (2003) but *different implications on the average productivity*. As shown in the appendix in Melitz (2003), an increase in κ_X or τ implies lower average productivity because more lenient selection includes firms with lower productivities before the change. However, when productivity is an endogenous choice, Proposition 4 shows that average productivity actually increases when κ_X increases, either because the newly included firms due to more lenient selection increase their innovation from zero to positive, or because the incumbent firms invest more due to less competition in both home and foreign markets. When τ increases, the effect on the average productivity is generally ambiguous as exporting firms invest less but nonexporting or newly included firms invest more. If κ_X is sufficiently high so that the fraction of exporting firms is relatively small, then an increase in τ may increase average productivity.[23]

4 Welfare Gains from Trade

This section analyzes the properties of welfare gains from trade in our model and then carries out a corresponding quantitative analysis. As is standard, welfare in both our model and the ACR

[23]Even though the average productivity may increase with κ_X or τ, such increases unambiguously decrease welfare, which is readily verified by examining (43).

framework is measured by $W_j = w_j N_j / P_j$. We are concerned with the welfare gains from trade, $d \ln W / d \ln \tau$. ACR show that under CES demand and certain macro restrictions, the welfare change from a small change in the trade cost, $d \ln \tau$, is given by $\frac{1}{\varepsilon} d \ln \lambda$ (local formula), where λ is the expenditure share on domestic goods, and $\varepsilon = \partial \ln (X_{ij}/X_{jj}) / \partial \ln \tau$ with $i \neq j$ is the trade elasticity. If the trade elasticity is invariant in τ, then the welfare change from a large change in the trade cost can be expressed as $W^{new}/W^{old} = \left(\lambda^{new}/\lambda^{old}\right)^{1/\varepsilon}$ (global formula). As the trade elasticity depends on λ, the main message from ACR is that trade flows provide sufficient statistic to the welfare gains from trade.

Even though the ACR framework includes technology choices, our model is different from the ACR framework because of the sequential nature of innovation.[24] It turns out the welfare gains from trade in our model still follow the local ACR formula (that is, the formula for small changes in τ). However, the trade elasticity is a variable in τ, and hence the global ACR formula (for large changes in τ) is not applicable.[25] It is important to note that both of the above-mentioned points hold for arbitrary distributions of γ. In our quantitative analysis, to investigate the role played by (process) innovation, our main comparison is with the Melitz model with an exogenous Pareto distribution.

4.1 Welfare Formula and Trade Elasticity

The following proposition shows that, for any distribution of γ, the local ACR formula holds with a variable trade elasticity. Under the symmetric country setting with wages normalized to 1, the expenditure share on the product imported from a foreign country equals $(1-\lambda)/n$ and the trade elasticity equals $\varepsilon = d \ln \left(\frac{1-\lambda}{n\lambda}\right) / d \ln \tau$. Recall that $\Gamma_z = \int_0^{\gamma_z} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} dF(\gamma)$ measures the contribution of the productivities in $(0, \gamma_z)$ to the expected profit of an entrant. Define the elasticity of Γ_z to the cutoff γ_z as η_z.

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:*

$$\frac{d \ln W}{d \ln \tau} = \frac{1}{\varepsilon} \frac{d \ln \lambda}{d \ln \tau} = -(1-\lambda). \tag{47}$$

The trade elasticity is given by

$$\varepsilon = (\sigma - 1) \frac{\Gamma_D - \Gamma_X}{\Gamma_D + (\phi^{\sigma-1} - 1)\Gamma_X} \frac{d \ln \phi}{d \ln \tau} + (1 - \sigma) \tag{48}$$

[24] In ACR (2012), the technological decision that influences productivity is simultaneous with production and sales, and the technology innovation is multiplicative in the overall fixed cost of production and exporting.

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

$$+\frac{\Gamma_D}{\Gamma_D+(\phi^{\sigma-1}-1)\Gamma_X}\eta_X\frac{d\ln\frac{\gamma_X}{\gamma_D}}{d\ln\tau}+\frac{\Gamma_D}{\Gamma_D+(\phi^{\sigma-1}-1)\Gamma_X}(\eta_X-\eta_D)\beta(1-\lambda),$$

where the domestic expenditure share is given by

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

Proof. See Appendix A.5. □

It is readily verified with a numerical example that the elasticity is variable in τ. In the next proposition, we show that when there is no fixed exporting cost, the trade elasticity becomes a constant.

Proposition 6. *Suppose that $\kappa_X = 0$ and assume that the conditions of Proposition 3 hold, then all firms such that $\gamma \leq \gamma_D$ survive and export, where*

$$\gamma_D^{\kappa_X=0}=\left[\frac{1}{\kappa_D}\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}\left[\frac{N}{(P^{\kappa_X=0})^{1-\sigma}}\right]^{\frac{\beta}{\beta-\sigma+1}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)\right]^{\frac{\beta-\sigma+1}{\sigma-1}};$$

the welfare formula satisfies the local ACR formula and the trade elasticity is a constant

$$\varepsilon^{\kappa_X=0}=1-\sigma<0. \tag{49}$$

Moreover, $\lambda > \lambda^{\kappa_X=0}$, and hence the welfare gains from trade are higher than the case with selection to export.

Proof. See Appendix A.6. □

Recall that the productivity distribution is piecewise because the fact that exporters have an extra incentive to invest creates a jump around γ_X. When there is no fixed exporting cost, the resulting productivity distribution becomes continuous. A comparison between Propositions 5 and 6 shows that whether $\kappa_X > 0$ or not is the key between variable or constant elasticity. What is intriguing is that these results are independent of the distributional forms. In any case, it is worth noting that the ACR local formula holds regardless of the distribution of γ or productivity. As shown in the ACR paper, the ACR formula holds for the Melitz model with an (exogenous) Pareto productivity distribution under a general asymmetric-country setting. Here we show that under a symmetric-country setting, the distributional form assumption is dispensable.

The intuition as to why the trade elasticity is a constant is as follows. All firms are exporters, and trade costs affect trade flows only through the intensive margin, which is affected by productivity and trade cost in a multiplicative way. Thus, trade elasticity is a constant and not dependent on the distribution of γ. In contrast, when there is selection to export, trade costs affects the intensive margin of exporters and nonexporters in different ways, and the extensive margin also matters. Both of these effects make the trade elasticity a variable and dependent on the distribution of γ.

As for the welfare gains from trade, the case without fixed exporting costs always yields higher gains than the case with it. This result is intuitive because trade liberalization induces further innovation for all firms in the former case, whereas only some firms are induced to invest more in the later case.

4.2 Quantitative Analysis of Welfare Gains from Trade

In this subsection we conduct a quantitative analysis of welfare gains from trade. In particular, to assess the role of innovation quantitatively, we compare with the Melitz model with an exogenous Pareto productivity distribution (henceforth MP), as both our model and MP satisfy the (local) ACR formula but differ only in how the productivity distribution is generated. Formally, the density function of the productivity distribution in the MP model is denoted as $g^{MP}(\varphi) = \theta^{MP}\varphi^{-\theta^{MP}-1}$ where $\theta^{MP} > \sigma - 1$ is the tail index. The trade elasticity in MP is $\varepsilon^{MP} = -\theta^{MP}$. The domestic expenditure share is given by

$$\lambda^{MP} = \left[1 + n\tau^{1-\sigma}\left(\frac{\varphi_X}{\varphi_D}\right)^{\sigma-\theta^{MP}-1}\right]^{-1}; \tag{50}$$

we relegate the derivation to Appendix A.7. It is well-known that the MP model is in the ACR class, and under symmetric countries, we have

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

We now turn to our model, which is referred to as IN (innovation) from now on. To single out 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 $\kappa_X > 0$. Under this assumption, the fraction of exporters equals δ as defined in (39). The domestic expenditure share in our model is

$$\lambda = \frac{1 + \left[\phi^{\sigma-1} - 1\right]\left(\frac{\gamma_X}{\gamma_D}\right)^{1-\frac{\sigma-1}{\theta}}}{1 + \left[\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1} - 1\right]\left(\frac{\gamma_X}{\gamma_D}\right)^{1-\frac{\sigma-1}{\theta}}}. \tag{51}$$

To quantify the model, we calibrate the values of σ, β, n, τ, and κ_X/κ_D. We calibrate these

Model	$\frac{W}{W_{\tau\to\infty}}$	$\frac{d\ln W}{d\ln\tau}$
IN	1.035	-0.147
MP	1.025	-0.108

Table 2: Welfare gains from trade

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 for our constant-markup model, $\sigma \approx 4.33$. Under the uniform distribution of γ, δ (see [39]) is the fraction of exporters among all (surviving) firms. As documented by Bernard, Jensen, Redding, and Schott (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.[26] 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. Also using PWT 9.0, this number equals 4.41. We therefore set $n = 3$.

As our model IN is most similar to the MP model, we adopt the estimate of the trade elasticity in Simonovska and Waugh (2014), which is 4.63.[27] This implies that $\theta^{MP} = 4.63$. We calibrate β, κ_X/κ_D, and τ to match $\lambda = 0.853$, $\delta = 0.18$, and $\varepsilon = 4.63$ using (51), (39), and (48). The results are $\beta = 7.838$, $\kappa_X/\kappa_D = 0.572$, and $\tau = 2.097$. This, in turn, implies that $\theta = 4.505$, which is rather close to θ^{MP}.

Given the calibrated parameters, we compute the local welfare gains for both IN and MP models. We also compare the welfare gains by moving from autarky ($\tau \to \infty$) to the current level of trade cost τ for both models. In Appendix A.7 we show that the welfare gains relative to autarky in both models are given by

$$\frac{W}{W_{\tau\to\infty}} = \left\{1 + n^{1-\frac{\theta}{\sigma-1}} \left(\frac{\kappa_X}{\kappa_D}\right)^{1-\frac{\theta}{\sigma-1}} \left[\left(1 + n\tau^{1-\sigma}\right)^{\frac{\beta}{\theta}} - 1\right]^{\frac{\theta}{\sigma-1}}\right\}^{\frac{1}{\beta}} \tag{52}$$

$$\frac{W^{MP}}{W^{MP}_{\tau\to\infty}} = \left[1 + n\tau^{-\theta^{MP}} \left(\frac{\kappa_X}{\kappa_D}\right)^{1-\frac{\theta^{MP}}{\sigma-1}}\right]^{\frac{1}{\theta^{MP}}}. \tag{53}$$

The results are as shown in Table 2. From autarky to the calibrated τ, IN and MP entail 3.5% and 2.5% of welfare gains, respectively. Hence, the gains from trade in IN is 40% larger than those

[26] 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.

[27] See their Table 7.

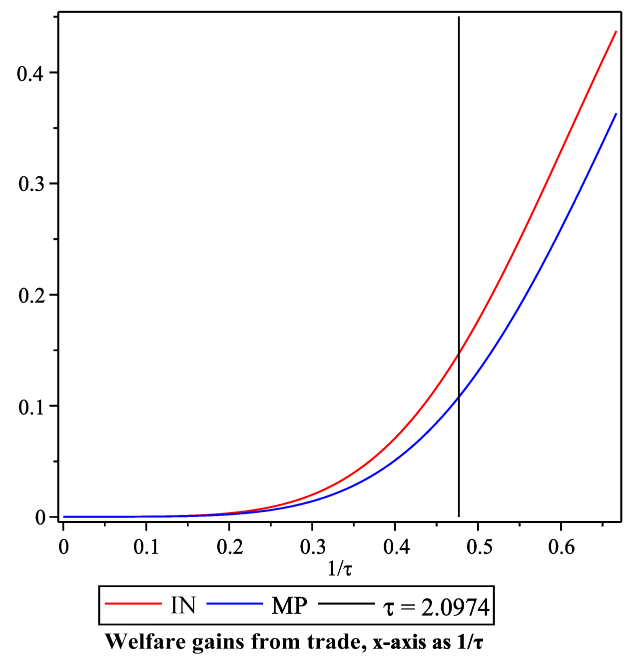


Figure 3: $\left|\frac{d\ln W}{d\ln\tau}\right|$

in MP. The welfare elasticities to trade cost are -0.147 and -0.108, for IN and MP, respectively. This implies that for small changes of τ, the welfare gains in IN is 36.1% higher than those in MP; this is quite similar to the above-mentioned comparison with autarky.

Figure 3 shows the welfare elasticity (in absolute value) under different values of τ given the calibrated β and κ_X/κ_D. The result is plotted with the horizontal axis being $1/\tau$, of which the lower bound, 0, corresponds to autarky and the upper bound corresponds to $\delta \to 1$ (recall that we restrict $\delta < 1$). The welfare elasticity (actually given by $1-\lambda$ as in Proposition 5) increases when there is more trade openness in both IN and MP models. The gains from trade (compared with autarky) is simply the area under the curve of welfare elasticity. It is clearly seen from Figure 3 that the IN model entails higher gains from trade than the MP model because $\lambda < \lambda^{MP}$ at every value of τ. To see why this is the case, recall from Proposition 4 that a reduction in trade cost induces exporters to invest more and become more productive and non-exporters to invest less and become less productive. This means that the productivity advantage of exporter vs. non-exporters widens with trade liberalization at a larger rate than the MP model. Thus, the rate of increase in $1-\lambda$ (the expenditure share on imports) is larger in the IN model than in the MP one.

5 Conclusion

This paper has demonstrated that with a process innovation stage added to a standard Melitz model, power laws for both productivity and firm size could emerge under a rather general environment. As highlighted by both Arkolakis et al. (2018) and Chaney (2018), the power law for productivity or firm size is instrumental for the gravity equation. Also evidenced is the fact that the performance of top firms is what matters the most for the aggregate economies (Gabaix 2011). Thus, understanding the root of these power laws is of first-order importance. Even though this paper is not the first to provide a microfoundation for these power laws, we add a new angle to this literature by applying the "power law change of variable close to the origin" technique in a general equilibrium model of trade and highlight the role of innovation. Both Geerolf (2017) and our work suggest that there may be more application of this technique in some suitable models of firm heterogeneity.

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. Exporters enjoy a larger market size with trade liberalization and tend to innovate more, whereas non-exporters do not enjoy such a benefit but suffer from stronger import competition that shrinks their effective market size. As a result, trade liberalization leads to further concentration of innovation and hence overall economic activities favor the large firms.

As shown by the welfare analysis, welfare gains from trade critically depend on the tail indices of these power laws, which reflects how granular the economy is. In this model, the tail indices depend on the price elasticities and how costly it is to conduct innovation. Interestingly, trade plays an important role because the market with the largest 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.

Reference

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 (2018), "Exports and Innovation: Theory and Evidence," 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 (2018), "The Elusive Pro-competitive Effects of Trade," *Review of Economic Studies*, forthcoming.

Axtell, R. L. (2001), "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818-1820.

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, 1268-1290.

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 (2018), "Betting on Exports: Trade and Endogenous Heterogeneity." *Economic Journal*, 128, 612-651.

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.

Caliendo, L. and F. Parro (2015), "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies*, 82, 1-44.

Chaney, T. (2014), "The Network Structure of International Trade," *American Economic Review*, 104, 3600-3634.

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

Chang, P. L., and W. T. Hsu (2019), "Geography, Trade and Power-law Phenomena," forthcoming, *Oxford Research Encyclopedia of Economics and Finance*.

Cooke, R. M., D. Nieboer, and J. Misiewicz (2014), *Fat-tailed Distributions: Data, Diagnostics, and Dependence*, John Wiley & Sons.

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.

Diewert, W. E. (1976), “Exact and Superlative Index Numbers,” *Journal of Econometrics*, 4, 115-145.

Eaton, J., and S. Kortum (2002), “Technology, Geography, and Trade,” *Econometrica*, 70, 1741-1779.

Edmond, C., V. Midrigan, and D. Y. Xu (2015), "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 105, 3183-3221.

Feenstra, R. C. and D. E. Weinstein (2017), “Globalization, Markups, and US Welfare,” *Journal of Political Economy*, 125, 1040-1074.

Feenstra, R. C. (2018), “Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity,” *Journal of International Economics*, forthcoming.

Fernandes, A. (2007), “Trade Policy, Trade Volumes and Plant-level Productivity in Colombian Manufacturing Industries,” *Journal of International Economics*, 71, 52-71.

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.

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

Holmes, T. J., W. T. Hsu, and S. Lee (2014), “Allocative Efficiency, Mark-ups, and the Welfare Gains from Trade,” *Journal of International Economics*, 94, 195-206.

Hsieh, C. T., and R. Ossa (2016), “A Global View of Productivity Growth in China,” *Journal of International Economics*, 102, 209-224.

Hsu, W. T. (2012), “Central Place Theory and City Size Distribution,” *Economic Journal*, 122, 903-932.

Jan, N., L. Moseley, T. Ray, and D. Stauffer (1999), “Is the Fossil Record Indicative of a Critical System,” *Advances in Complex Systems*, 2, 137-141.

Luttmer, E. G. (2007), “Selection, growth, and the size distribution of firms,” *Quarterly Journal of Economics*, 122, 1103-1144.

Merton, R. C. (1971), “Optimum Consumption and Portfolio Rules in a Continuous-Time Model,” *Journal of Economic Theory*, 3, 373-413.

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 (2014), “Missing Gains from Trade?” *American Economic Review: Papers and Proceedings*, 104, 317-21.

Melitz, M. J., and S. J. Redding (2015), “New Trade Models, New Welfare Implications,” *American Economic Review*, 105, 1105-1146.

Mrázová, M., and J. P. Neary (2017), “Not So Demanding: Demand Structure and Firm Behavior,” *American Economic Review*, 107, 3835-3874.

Mrázová, M., J. P. Neary, and M. Parenti (2017), “Sales and Markup Dispersion: Theory and Empirics,” CEPR Discussion Papers DP12044.

Muellbauer, J. (1975), “Aggregation, Income Distribution and Consumer Demand,” *Review of Economic Studies*, 42, 525-543.

Newman, M. E. J. (2005), “Power Laws, Pareto Distributions and Zipf’s Law,” *Contemporary Physics*, 46, 323-351.

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

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

Pollak, R. A. (1971), “Additive Utility Functions and Linear Engel Curves,” *Review of Economic Studies*, 38, 401-414.

Rossi-Hansberg, E, and M. L. J. Wright (2007), “Establishment Size Dynamics in the Aggregate Economy,” *American Economic Review*, 97, 1639-66.

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

Sornette, D. (2002), "Mechanism for Powerlaws without Self-Organization," *International Journal of Modern Physics C*, 13, 133-136.

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

Sutton, J. (1991), "The Analytical Framework II: Endogenous Sunk Costs," *Sunk Costs and Market Structure*, MIT Press, 45-82.

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, 2765-2784.

A Appendix

A.1 Derivations for Table 1

Bipower direct demand $q = \widehat{a}p^{-\nu} + ap^{-\sigma}$**:**

Assume that $\sigma > 1$, $a > 0$, and $\sigma > \nu$; we immediately obtain

$$\begin{aligned} q =& p^{-\sigma}\left(\widehat{a}p^{\sigma-\nu} + a\right) \equiv \boldsymbol{q}\left(p\right) \\ \lim_{p\to 0}\left(\widehat{a}p^{\sigma-\nu} + a\right) =& a \\ \lim_{p\to 0} q =& \infty \\ \lim_{p\to 0} pq =& \infty. \end{aligned}$$

Moreover, $\partial p/\partial q = -p^{-\nu-1}\left(\nu\widehat{a} + \sigma a p^{\nu-\sigma}\right)$ and is strictly negative for small enough p. Thus for large enough q the inverse demand is well-defined as $p = q^{-\frac{1}{\sigma}}\left(\widehat{a}\left[\boldsymbol{q}^{-1}\left(q\right)\right]^{\sigma-\nu} + a\right)^{\frac{1}{\sigma}} \equiv q^{-\frac{1}{\sigma}}Q\left(q\right)$ where $\lim_{q\to\infty} Q\left(q\right) = a^{\frac{1}{\sigma}}$.

Pollak demand $q = \widehat{a} + ap^{-\sigma}$**:**

Assume that $\sigma > 1$ and $a > 0$; we immediately obtain

$$\begin{aligned} q =& \widehat{a} + ap^{-\sigma} \\ \lim_{p \to 0} q =& \infty \\ \lim_{p \to 0} pq =& \infty. \end{aligned}$$

As a result, $p = q^{-\frac{1}{\sigma}} a^{\frac{1}{\sigma}} \left(1 - \frac{\widehat{a}}{q}\right)^{-\frac{1}{\sigma}} \equiv q^{-\frac{1}{\sigma}} Q(q)$ where $\lim_{q \to \infty} Q(q) = a^{\frac{1}{\sigma}}$.

PIGL demand $q = \widehat{a}p^{-1} + ap^{-\sigma}$**:**

Assume that $\sigma > 1$ and $a > 0$; we immediately obtain

$$\begin{aligned} q =& p^{-\sigma} \left(\widehat{a}p^{\sigma-1} + a\right) \equiv \boldsymbol{q}(p) \\ \lim_{p \to 0} \left(\widehat{a}p^{\sigma-1} + a\right) =& a \\ \lim_{p \to 0} q =& \infty \\ \lim_{p \to 0} pq =& \infty. \end{aligned}$$

Moreover, $\partial p / \partial q = -p^{-2} \left(\widehat{a} + \sigma a p^{1-\sigma}\right)$ and is strictly negative for small enough p. Thus for large enough q the inverse demand is well-defined as $p = q^{-\frac{1}{\sigma}} \left(\widehat{a} \left[\boldsymbol{q}^{-1}(q)\right]^{\sigma-1} + a\right)^{\frac{1}{\sigma}} \equiv q^{-\frac{1}{\sigma}} Q(q)$ where $\lim_{q \to \infty} Q(q) = a^{\frac{1}{\sigma}}$.

QMOR demand $q = ap^{r-1} + \widehat{a}p^{\frac{r}{2}-1}$**:**

Assume that $\sigma \equiv 1 - r > 1$ and $a > 0$. This implies that $r < 0$ and $1 - r > 1 - \frac{r}{2} > 0$; thus

$$\begin{aligned} q =& p^{r-1} \left(\widehat{a}p^{-\frac{r}{2}} + a\right) \equiv \boldsymbol{q}(p) \\ \lim_{p \to 0} \left(\widehat{a}p^{-\frac{r}{2}} + a\right) =& a \\ \lim_{p \to 0} q =& \infty \\ \lim_{p \to 0} pq =& \infty. \end{aligned}$$

Moreover, $\partial p / \partial q = p^{r-2} \left[\left(\frac{r}{2} - 1\right) \widehat{a} p^{-\frac{r}{2}} + (r-1) a\right]$ and is strictly negative for small enough p. Thus for large enough q the inverse demand is well-defined as $p = q^{\frac{1}{r-1}} \left(a + \widehat{a} \left[\boldsymbol{q}^{-1}(q)\right]^{-\frac{r}{2}}\right)^{\frac{1}{1-r}} \equiv$

$q^{\frac{1}{r-1}}Q(q)$ where $\lim_{q\to\infty} Q(q) = a^{\frac{1}{1-r}}$.

Bipower inverse demand $p = \widehat{a}q^{-\nu} + aq^{-\frac{1}{\sigma}}$**:**

Assume that $\sigma > 1$, $a > 0$, and $\nu > \frac{1}{\sigma}$; we immediately obtain

$$\begin{aligned}
p =& q^{-\frac{1}{\sigma}}\left(\widehat{a}q^{\frac{1}{\sigma}-\nu} + a\right) \equiv q^{-\frac{1}{\sigma}}Q(q) \\
\lim_{q\to\infty} Q(q) =& a \\
\lim_{q\to\infty} p =& 0 \\
\lim_{q\to\infty} pq =& \infty.
\end{aligned}$$

CEMR demand $p = \widehat{a}q^{-1} + aq^{-\frac{1}{\sigma}}$**:**

Assume that $\sigma > 1$ and $a > 0$; we immediately obtain

$$\begin{aligned}
p =& q^{-\frac{1}{\sigma}}\left(\widehat{a}q^{\frac{1}{\sigma}-1} + a\right) \equiv q^{-\frac{1}{\sigma}}Q(q) \\
\lim_{q\to\infty} Q(q) =& a \\
\lim_{q\to\infty} p =& 0 \\
\lim_{q\to\infty} pq =& \infty.
\end{aligned}$$

CREMR demand $p = \frac{a}{q}(q - \widehat{a})^{\frac{\sigma-1}{\sigma}}$**:**

Assume that $\sigma > 1$, $a > 0$, and $q > \widehat{a}\sigma$; we immediately obtain

$$\begin{aligned}
p =& q^{-\frac{1}{\sigma}}a\left(1 - \frac{\widehat{a}}{q}\right)^{\frac{\sigma-1}{\sigma}} \equiv q^{-\frac{1}{\sigma}}Q(q) \\
\lim_{q\to\infty} Q(q) =& a \\
\lim_{q\to\infty} p =& 0 \\
\lim_{q\to\infty} pq =& \infty.
\end{aligned}$$

$\square$

A.2 Proof of Proposition 1

We show this proposition in the following four 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 the existence of equilibrium. In the second and third steps we show that both productivity and firm size distributions exhibit power laws.

Step 1:

We require $\int_0^{\gamma_D} \Pi(\widetilde{\varphi}(\gamma);\gamma)\,dF(\gamma) < \infty$ for the free entry condition to be well-defined. Since $\Pi(\widetilde{\varphi}(\gamma);\gamma)$ is finite for all $\gamma > 0$, the only possibility for the expected profit to explode is when γ is close to 0. Note that using (16) and (22) we can write

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

Because κ_D is a constant, the expected profit is finite if $\int_0^{\gamma_D}[\Pi(\widetilde{\varphi}(\gamma);\gamma)+\kappa_D]\,f(\gamma)\,d\gamma < \infty$. By Assumption 1 it follows that

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

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

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

By picking a sufficiently small $\overline{\gamma}$ and noting that $f(\gamma)=\gamma^{\alpha}m(\gamma)$, we have

$$\begin{aligned}
\int_0^{\overline{\gamma}}[\Pi(\widetilde{\varphi}(\gamma);\gamma)+\kappa_D]\,f(\gamma)\,d\gamma &= \int_0^{\overline{\gamma}}\frac{[\Pi(\widetilde{\varphi}(\gamma);\gamma)+\kappa_D]}{\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}}\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}f(\gamma)\,d\gamma \\
&< \int_0^{\overline{\gamma}}\left[\frac{C_Q^{\frac{\sigma\beta}{\beta-\sigma+1}}\left(\frac{\sigma-1}{\sigma}\right)^{\frac{\sigma\beta}{\beta-\sigma+1}}}{C_L^{\frac{\sigma-1}{\beta-\sigma+1}}\beta^{\frac{\beta}{\beta-\sigma+1}}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)+\omega\right]\gamma^{\alpha-\frac{\sigma-1}{\beta-\sigma+1}}m(\gamma)\,d\gamma \\
&= \left[\frac{C_Q^{\frac{\sigma\beta}{\beta-\sigma+1}}\left(\frac{\sigma-1}{\sigma}\right)^{\frac{\sigma\beta}{\beta-\sigma+1}}}{C_L^{\frac{\sigma-1}{\beta-\sigma+1}}\beta^{\frac{\beta}{\beta-\sigma+1}}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)+\omega\right]\int_0^{\overline{\gamma}}\gamma^{\alpha-\frac{\sigma-1}{\beta-\sigma+1}}m(\gamma)\,d\gamma \\
&< \left[\frac{C_Q^{\frac{\sigma\beta}{\beta-\sigma+1}}\left(\frac{\sigma-1}{\sigma}\right)^{\frac{\sigma\beta}{\beta-\sigma+1}}}{C_L^{\frac{\sigma-1}{\beta-\sigma+1}}\beta^{\frac{\beta}{\beta-\sigma+1}}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)+\omega\right]\int_0^{1}\gamma^{\alpha-\frac{\sigma-1}{\beta-\sigma+1}}m(\gamma)\,d\gamma.
\end{aligned}$$

The expected profit is finite if $\int_0^1\gamma^{\alpha-\frac{\sigma-1}{\beta-\sigma+1}}m(\gamma)\,d\gamma = E\left(\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}\right) < \infty$. Let $y=\gamma^{-1}$;

we can restate $f(\gamma)$ as $f_y(y) = y^{-(\alpha+1)-1}m(y^{-1})$ so that $E\left(\gamma^{-\frac{\sigma-1}{\beta-\sigma+1}}\right) = E_y\left(y^{\frac{\sigma-1}{\beta-\sigma+1}}\right)$ where $\overline{y} \equiv \overline{\gamma}^{-1}$. Now we can recall Proposition 4.6 of Cooke, Nieboer, and Misiewicz (2014; pp. 53-54) that for a random variable x with a regularly varying tail probability with degree $-a$, i.e., $\Pr(X > x) \propto x^{-a}$ and $a > 0$, then for all $b > 0$ the moment $E\left(x^b\right)$ is finite if $a > b$. Because f_y has a regularly varying tail probability with degree $-(\alpha+1)$ and we require $\sigma > 1$, it follows that $E_y\left(y^{\frac{\sigma-1}{\beta-\sigma+1}}\right) < \infty$ if $\alpha > -1$ and $\beta > \frac{\alpha+2}{\alpha+1}(\sigma-1)$.

Step 2:

We derive Equation (25) in detail. Starting from the definition,

$$\begin{aligned}
J(\varphi) =& \left|\frac{\partial\widetilde{\gamma}(\varphi)}{\partial\varphi}\right| \\
=& \frac{q^*(\varphi)}{\varphi^2 V'(\varphi)}\left(2\varphi^{-1} + \frac{V''(\varphi)}{V'(\varphi)}\right) - \frac{1}{\varphi^2 V'(\varphi)}\frac{\partial q^*(\varphi)}{\partial\varphi} \\
=& \frac{q^*(\varphi)}{\varphi^2 V'(\varphi)}\left(2\varphi^{-1} + \frac{V''(\varphi)}{V'(\varphi)}\right) + \frac{1}{\varphi^2 V'(\varphi)}\frac{\varphi^{-2}}{\pi_{qq}(q^*(\varphi),\varphi)} \text{ by equation (18).}
\end{aligned}$$

Then, by Assumption 1, and Equations (16) and (17), we can replace $V'(\varphi)$, $V''(\varphi)$, $q^*(\varphi)$, and $\pi_{qq}(q^*(\varphi),\varphi)$ to obtain Equation (25) 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 Assumption 1, and Lemmas 1, 2, and 3, we immediately obtain the following results:

$$\lim_{\varphi\to\infty}\varphi\frac{L'}{L} = \lim_{\varphi\to\infty}\varphi^2\frac{L''}{L} = 0$$
$$\lim_{\varphi\to\infty}q^*\frac{Q'}{Q} = \lim_{\varphi\to\infty}(q^*)^2\frac{Q''}{Q} = 0$$
$$\lim_{\varphi\to\infty}L = C_L$$
$$\lim_{\varphi\to\infty}Q = C_Q.$$

Therefore,

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

By a change of variables and the results above, the p.d.f. of productivity is

$$\begin{aligned}
g(\varphi) =& \frac{f(\widetilde{\gamma}(\varphi))}{F(\gamma_D)} J(\varphi) \\
=& \frac{f(\widetilde{\gamma}(\varphi))}{F(\gamma_D)} \frac{J(\varphi)}{\varphi^{-(\beta-\sigma+1)-1}} \varphi^{-(\beta-\sigma+1)-1} \\
=& \frac{m(\widetilde{\gamma}(\varphi))}{F(\gamma_D)} \left[\frac{Q^\sigma}{L} \frac{\left(1 - \frac{1}{\sigma} + q\frac{Q'}{Q}\right)^\sigma}{\beta + \varphi\frac{L'}{L}} \right]^\alpha \frac{J(\varphi)}{\varphi^{-(\beta-\sigma+1)-1}} \varphi^{-(1+\alpha)(\beta-\sigma+1)-1}.
\end{aligned}$$

As φ becomes arbitrarily large, both the bracketed term and $\frac{J(\varphi)}{\varphi^{-(\beta-\sigma+1)-1}}$ converge to constants. Because $F(\gamma_D)$ is a constant, it follows that $g(\varphi)$ exhibits a power law to the right with a tail index $(1+\alpha)(\beta-\sigma+1)$ if $m(\widetilde{\gamma}(\varphi))$ is slowly varying as φ becomes arbitrarily large.

Recall Proposition 1.5.7 from Bingham et al. (1989) that, for two regularly varying functions $v_1(x) = x^{a_1} l_1(x)$ and $v_2(x) = x^{a_2} l_2(x)$ where $\lim_{x\to\infty} v_2(x) = \infty$, the composite function $v_c(x) \equiv v_1(v_2(x))$ is also regularly varying with degree $a_c \equiv a_1 a_2$. To show that $m(\widetilde{\gamma}(\varphi))$ is slowly varying, let $y = \gamma^{-1}$ so (22) becomes

$$y(\varphi) = \frac{L(\varphi)}{Q(\varphi)^\sigma} \frac{\beta + \varphi\frac{L'(\varphi)}{L(\varphi)}}{\left(1 - \frac{1}{\sigma} + q^*(\varphi)\frac{Q'(\varphi)}{Q(\varphi)}\right)^\sigma} \varphi^{\beta-\sigma+1}$$

and $\widetilde{m}(y) \equiv m(y^{-1}) = m(\gamma)$. Note that $\widetilde{m}$ is also slowly varying as y tends to infinity and that $y(\varphi)$ is regularly varying in φ of degree $\beta - \sigma + 1$ with $\lim_{\varphi\to\infty} y(\varphi) = \infty$. Taken together we have $m(\widetilde{\gamma}(\varphi)) = \widetilde{m}(y(\varphi))$ and Proposition 1.5.7-(ii) of Bingham et al. (1989) immediately implies that $\widetilde{m}(y(\varphi))$ is slowly varying in φ. As a result, $g(\varphi)$ is regularly varying in φ with a tail index $(1+\alpha)(\beta-\sigma+1)$.

Step 3:

By Equation (16) and Lemma 2, we find that firm size in terms of sales s is a function of φ:

$$s = \varphi^{\sigma-1} Q^\sigma \left(1 - \frac{1}{\sigma} + q^* \frac{Q'}{Q}\right)^{\sigma-1}. \tag{54}$$

By Assumption 1 and Lemmas 2 and 3, 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 $s(\varphi)$ denote the firm size with productivity φ as defined by (54); $\varphi(s)$ denotes its inverse function. By combining (22) and (54) we have

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

which is regularly varying in s with degree $-\frac{\beta-\sigma+1}{\sigma-1}$ under Assumption 1.

By the chain rule and using Equations (16) and (17), we further find that

$$\begin{aligned}\frac{\partial s(\varphi)}{\partial \varphi} =& \frac{\partial s}{\partial q^*} \frac{\partial q^*}{\partial \varphi} \\ =& \frac{1 - \frac{1}{\sigma} + q^* \frac{Q'}{Q}}{\frac{1}{\sigma}\left(1 - \frac{1}{\sigma}\right) - 2\left(1 - \frac{1}{\sigma}\right) q^* \frac{Q'}{Q} - (q^*)^2 \frac{Q''}{Q}} q^* \varphi^{-2} \\ >& 0.\end{aligned} \tag{55}$$

Using (54), (55), and (16), we can obtain the Jacobian $J_s(s)$ as

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

Note that $f(\gamma) = \gamma^\alpha m(\gamma)$ where $m(\gamma)$ is slowly varying; the firm size distribution $g_s(s)$ is given as

$$\begin{aligned}g_s(s) =& \frac{m(\widetilde{\gamma}(s)) [\widetilde{\gamma}(s)]^\alpha}{F(\gamma_D)} J_s(s) \\ =& \frac{m(\widetilde{\gamma}(s))}{F(\gamma_D)} \frac{J_s(s)}{s^{-\frac{\beta-\sigma+1}{\sigma-1}-1}} \left(\frac{\widetilde{\gamma}(s)}{s^{-\frac{\beta-\sigma+1}{\sigma-1}}}\right)^\alpha s^{-\frac{(\alpha+1)(\beta-\sigma+1)}{\sigma-1}-1}.\end{aligned}$$

By Assumption 1, Lemma 1, 2, and 3, we know that both $J_s(s) / s^{-\frac{\beta-\sigma+1}{\sigma-1}-1}$ and $\widehat{\gamma}(s)^\alpha / s^{-\alpha\frac{\beta-\sigma+1}{\sigma-1}}$ converge to constants as s tends to infinity. By the same argument applied in Step 2 we know that $m(\widehat{\gamma}(s))$ is slowly varying in s. Therefore, $g_s(s)$ is regularly varying in s with a tail index $\frac{(\alpha+1)(\beta-\sigma+1)}{\sigma-1}$. □

A.3 Derivation of Equilibrium in Section 3.1

The proof comes in two steps. First, we show the uniqueness of the equilibrium. Then, we derive the associated equilibrium outcome.

Step 1: The Uniqueness of Equilibrium

By Equations (35), (36), and (34) the expected profit is given by

$$\begin{aligned}
E(\Pi) =& \int_0^{\gamma_X} \Pi_X(\gamma)\, dF(\gamma) + \int_{\gamma_X}^{\gamma_D} \Pi_D(\gamma)\, dF(\gamma) \\
=& \left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right) \int_0^{\gamma_X} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} dF(\gamma) - \kappa_D \int_0^{\gamma_X} dF(\gamma) - n\kappa_X \int_0^{\gamma_X} dF(\gamma) \\
&+ \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right) \int_{\gamma_X}^{\gamma_D} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} dF(\gamma) - \kappa_D \int_{\gamma_X}^{\gamma_D} dF(\gamma) \\
=& \left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right) \int_0^{\gamma_X} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} dF(\gamma) \qquad (56) \\
&+ \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right) \left[\int_0^{\gamma_D} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} dF(\gamma) - \int_0^{\gamma_X} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} dF(\gamma)\right] \\
&- \kappa_D \int_0^{\gamma_D} dF(\gamma) - n\kappa_X \int_0^{\gamma_X} dF(\gamma).
\end{aligned}$$

Let $\Gamma_z \equiv \int_0^{\gamma_z} \gamma^{-\frac{\sigma-1}{\beta-\sigma+1}} dF(\gamma)$ where $z \in \{D, X\}$, and

$$\overline{R}_D \equiv \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right) (\Gamma_D - \Gamma_X)$$

$$\overline{R}_X \equiv \left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}} \left(\frac{\beta-\sigma+1}{\sigma-1}\right) \Gamma_X.$$

Using (37) and recalling that $\gamma_X/\gamma_D = \delta$, we further obtain

$$\begin{aligned}
\overline{R}_D =& \kappa_D \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} (\Gamma_D - \Gamma_X) \\
\overline{R}_X =& \left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} \kappa_D \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} \Gamma_X.
\end{aligned}$$

Hence we can restate (56) as

$$E(\Pi) = \kappa_D \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} \left\{\Gamma_D + \left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right] \Gamma_X\right\} - \kappa_D F(\gamma_D) - n\kappa_X F(\gamma_X). \qquad (57)$$

Differentiating (57) with respect to γ_D yields

$$\frac{\partial E(\Pi)}{\partial \gamma_D} = \kappa_D \left[f(\gamma_D) + \frac{\sigma-1}{\beta-\sigma+1} \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}-1} \Gamma_D\right]$$

$$
\begin{aligned}
&+\kappa_D\left\{\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1\right]\delta^{\frac{1-\sigma}{\beta-\sigma+1}}f\left(\gamma_X\right)\delta+\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1\right]\frac{\sigma-1}{\beta-\sigma+1}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}-1}\Gamma_X\right\}\\
&-\kappa_D f\left(\gamma_D\right)-n\kappa_X f\left(\gamma_X\right)\delta\\
=&\kappa_D\frac{\sigma-1}{\beta-\sigma+1}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}-1}\Gamma_D+\kappa_D\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1\right]\frac{\sigma-1}{\beta-\sigma+1}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}-1}\Gamma_X\\
&+\kappa_D\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1\right]\delta^{\frac{1-\sigma}{\beta-\sigma+1}}f\left(\gamma_X\right)\delta-n\kappa_X f\left(\gamma_X\right)\delta\\
=&\kappa_D\frac{\sigma-1}{\beta-\sigma+1}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}-1}\Gamma_D+\kappa_D\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1\right]\frac{\sigma-1}{\beta-\sigma+1}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}-1}\Gamma_X\\
&+n\kappa_X\delta^{\frac{\sigma-1}{\beta-\sigma+1}}\delta^{\frac{1-\sigma}{\beta-\sigma+1}}f\left(\gamma_X\right)\delta-n\kappa_X f\left(\gamma_X\right)\delta\\
=&\kappa_D\frac{\sigma-1}{\beta-\sigma+1}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}-1}\Gamma_D+\kappa_D\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1\right]\frac{\sigma-1}{\beta-\sigma+1}\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}-1}\Gamma_X\\
=&\frac{\sigma-1}{\beta-\sigma+1}\gamma_D^{-1}\left(\overline{R}_D+\overline{R}_X\right)>0.
\end{aligned}
$$

This implies that there is an unique γ_D such that the free entry condition (40) holds. This establishes the uniqueness of equilibrium.

Step 2: The Equilibrium Solutions

We start with the definition of the aggregate price. Using (33) and (41), the aggregate price can be restated as

$$
P^{1-\sigma}=M_e\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}\left(\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}\left\{\Gamma_D+\left[\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}-1\right]\Gamma_X\right\}.
$$

Rearranging the above equation, we have

$$
P^{(1-\sigma)\frac{\beta}{\beta-\sigma+1}}=M_e\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}N^{\frac{\sigma-1}{\beta-\sigma+1}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\sigma-1}{\beta-\sigma+1}}\left\{\Gamma_D+\left[\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}-1\right]\Gamma_X\right\}; \tag{58}
$$

thus

$$
\left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\beta-\sigma+1}}=\frac{\frac{N}{M_e}\left(\frac{\sigma-1}{\sigma}\right)^{1-\sigma}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{-\frac{\sigma-1}{\beta-\sigma+1}}}{\Gamma_D+\left[\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}-1\right]\Gamma_X}. \tag{59}
$$

Because $\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}=\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}$, we can then insert (59) into (56) to obtain

$$
E\left(\Pi\right)=\frac{N}{M_e}\frac{\sigma-1}{\beta\sigma}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)-\kappa_D F\left(\gamma_D\right)-n\kappa_X F\left(\gamma_X\right).
$$

Then, by the free entry condition we obtain Equation (44). By inserting Equation (44) into Equation (58) and rearranging along with Equation (42), we get Equation (43).

Next we derive the equilibrium productivity and its distribution. Inserting Equation (37) into Equation (33) we obtain (45). Then the threshold productivity levels φ_D, φ_X^-, and φ_X^+ respectively follow by inserting γ_D and γ_X into Equation (45). By a change of variables and minding the fact that the productivity distribution is conditional on $\gamma \leq \gamma_D$, we obtain Equation (46). □

A.4 Proof of Proposition 4

The proof comes in three steps. In the first step, we derive the comparative statics of γ_D. Then in the second step we derive the comparative statics for the other interested variables with the result from the first step. Finally, with the results from the previous two steps, we show that an increment in κ_D entails a new productivity distribution that FOSD (first order stochastic dominates) the old distribution.

Step 1: The Comparative Statics of γ_D

Let $x \in \{\kappa_D, \kappa_X, \tau\}$. Total differentiating $E(\Pi)$ with respect to x yields

$$\frac{d\gamma_D}{dx} = -\frac{\partial E(\Pi)/\partial x}{\partial E(\Pi)/\partial \gamma_D}.$$

Recall that

$$\frac{\partial E(\Pi)}{\partial \gamma_D} = \frac{\sigma - 1}{\beta - \sigma + 1}\gamma_D^{-1}\left(\overline{R}_D + \overline{R}_X\right) > 0;$$

what is left is to solve for $\partial E(\Pi)/\partial x$.

Observe that

$$\begin{aligned}\frac{dF(\gamma_X)}{d\delta} =& f(\delta\gamma_D)\gamma_D \\ \frac{d\Gamma_X}{d\delta} =& (\delta\gamma_D)^{-\frac{\sigma-1}{\beta-\sigma+1}} f(\delta\gamma_D)\gamma_D \\ =& (\delta\gamma_D)^{-\frac{\sigma-1}{\beta-\sigma+1}} \frac{dF(\gamma_X)}{d\delta}.\end{aligned}$$

We have the following results from Equation (57):

$$\begin{aligned}\frac{\partial E(\Pi)}{\partial \kappa_D} =& \frac{\overline{R}_D + \overline{R}_X - \kappa_D F(\gamma_D)}{\kappa_D} + \kappa_D \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}}\left[\left(1 + n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\frac{\partial \Gamma_X}{\partial \delta}\frac{\partial \delta}{\partial \kappa_D} - n\kappa_X\frac{\partial F(\gamma_X)}{\partial \delta}\frac{\partial \delta}{\partial \kappa_D} \\ =& \frac{\overline{R}_D + \overline{R}_X - \kappa_D F(\gamma_D)}{\kappa_D} + \kappa_D \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}}\left[\left(1 + n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right](\delta\gamma_D)^{-\frac{\sigma-1}{\beta-\sigma+1}}\frac{\partial F(\gamma_X)}{\partial \delta}\frac{\partial \delta}{\partial \kappa_D} \\ &- n\kappa_X\frac{\partial F(\gamma_X)}{\partial \delta}\frac{\partial \delta}{\partial \kappa_D} \\ =& \frac{\overline{R}_D + \overline{R}_X - \kappa_D F(\gamma_D)}{\kappa_D} + \frac{\partial F(\gamma_X)}{\partial \delta}\frac{\partial \delta}{\partial \kappa_D}\left[\gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} n\kappa_X \delta^{\frac{\sigma-1}{\beta-\sigma+1}}(\delta\gamma_D)^{-\frac{\sigma-1}{\beta-\sigma+1}} - n\kappa_X\right]\end{aligned}$$

$$
\begin{aligned}
&= \frac{\overline{R}_D + \overline{R}_X - \kappa_D F(\gamma_D)}{\kappa_D} \\
\frac{\partial E(\Pi)}{\partial \kappa_X} &= -nF(\gamma_X) + n\kappa_X \delta^{\frac{\sigma-1}{\beta-\sigma+1}} \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} (\delta\gamma_D)^{-\frac{\sigma-1}{\beta-\sigma+1}} \frac{\partial F(\gamma_X)}{\partial \delta}\frac{\partial \delta}{\partial \kappa_X} - n\kappa_X \frac{\partial F(\gamma_X)}{\partial \delta}\frac{\partial \delta}{\partial \kappa_X} \\
&= -nF(\gamma_X) \\
\frac{\partial E(\Pi)}{\partial \tau} &= \frac{\beta}{\beta-\sigma+1}\kappa_D \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} \left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}-1} (1-\sigma) n\tau^{-\sigma}\Gamma_X \\
&\quad + n\kappa_X \delta^{\frac{\sigma-1}{\beta-\sigma+1}} \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} (\delta\gamma_D)^{-\frac{\sigma-1}{\beta-\sigma+1}} \frac{\partial F(\gamma_X)}{\partial \delta}\frac{\partial \delta}{\partial \tau} - n\kappa_X \frac{\partial F(\gamma_X)}{\partial \delta}\frac{\partial \delta}{\partial \tau} \\
&= -\frac{(\sigma-1)\beta}{\beta-\sigma+1}\frac{n\tau^{-\sigma}}{1+n\tau^{1-\sigma}}\kappa_D \gamma_D^{\frac{\sigma-1}{\beta-\sigma+1}} \left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} \Gamma_X \\
&= -\frac{(\sigma-1)\beta}{\beta-\sigma+1}\frac{n\tau^{-\sigma}}{1+n\tau^{1-\sigma}}\overline{R}_X.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{d\gamma_D}{d\kappa_D} &= -\frac{\beta-\sigma+1}{\sigma-1}\frac{\gamma_D}{\kappa_D}\frac{\overline{R}_D + \overline{R}_X - \kappa_D F(\gamma_D)}{\overline{R}_D + \overline{R}_X} < 0 \\
\frac{d\gamma_D}{d\kappa_X} &= \frac{\beta-\sigma+1}{\sigma-1}\gamma_D \frac{nF(\gamma_X)}{\overline{R}_D + \overline{R}_X} > 0 \\
\frac{d\gamma_D}{d\tau} &= \beta\gamma_D \frac{n\tau^{-\sigma}}{1+n\tau^{1-\sigma}}\frac{\overline{R}_X}{\overline{R}_D + \overline{R}_X} > 0.
\end{aligned}
$$

Note that $\partial\gamma_D/\partial\kappa_D < 0$ holds because $\overline{R}_D + \overline{R}_X - \kappa_D F(\gamma_D) > E(\Pi) = \overline{R}_D + \overline{R}_X - \kappa_D F(\gamma_D) - n\kappa_X F(\gamma_X) > 0$.

Step 2: The Comparative Statics for the Rest of the Variables

We first examine the comparative statics of γ_X. Let $x \in \{\kappa_D, \kappa_X, \tau\}$. It follows that

$$
\frac{\partial \gamma_X}{\partial x} = \delta \frac{\partial \gamma_D}{\partial x} + \frac{\partial \delta}{\partial x}\gamma_D.
$$

Therefore,

$$
\begin{aligned}
\frac{\partial \gamma_X}{\partial \kappa_D} &= -\frac{\beta-\sigma+1}{\sigma-1}\frac{\gamma_D}{\kappa_D}\frac{\overline{R}_D + \overline{R}_X - f_D F(\gamma_D)}{\overline{R}}\delta + \frac{\beta-\sigma+1}{\sigma-1}\delta\delta^{-\frac{\sigma-1}{\beta-\sigma+1}}\frac{\delta^{\frac{\sigma-1}{\beta-\sigma+1}}}{\kappa_D}\gamma_D \\
&= \frac{\beta-\sigma+1}{\sigma-1}\gamma_D\delta\frac{F(\gamma_D)}{\overline{R}_D + \overline{R}_X} > 0 \\
\frac{\partial \gamma_X}{\partial \kappa_X} &= \frac{\beta-\sigma+1}{\sigma-1}\gamma_D\frac{nF(\gamma_X)}{\overline{R}_D + \overline{R}_X}\delta + \frac{\beta-\sigma+1}{\sigma-1}\delta\frac{-1}{\kappa_X}\gamma_D \\
&= \frac{\beta-\sigma+1}{\sigma-1}\gamma_D\delta\left(\frac{nF(\gamma_X)}{\overline{R}_D + \overline{R}_X} - \frac{1}{\kappa_X}\right) \\
&< 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \gamma_X}{\partial \tau} =& \beta \gamma_D \frac{n\tau^{-\sigma}}{1+n\tau^{1-\sigma}} \frac{\overline{R}_X}{\overline{R}_D+\overline{R}_X}\delta + \frac{\beta-\sigma+1}{\sigma-1}\delta \frac{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}}{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1} \frac{\beta}{\beta-\sigma+1} \frac{(1-\sigma)\, n\tau^{-\sigma}}{1+n\tau^{1-\sigma}}\gamma_D \\
=& \delta\beta\gamma_D \frac{n\tau^{-\sigma}}{1+n\tau^{1-\sigma}} \left[\frac{\overline{R}_X}{\overline{R}_D+\overline{R}_X} - \frac{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}}{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1} \right] \\
<& 0.
\end{aligned}
$$

Note that $\partial\gamma_X/\partial\kappa_X < 0$ holds because $\overline{R}_D + \overline{R}_X - \kappa_D F(\gamma_D) > E(\Pi) > 0$, and $\partial\gamma_X/\partial\tau < 0$ holds since $\frac{\overline{R}_X}{\overline{R}_D+\overline{R}_X} < 1$ and $\frac{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}}{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1} > 1$.

For the effect on productivity, by taking derivatives to (45) we obtain

$$
\frac{\partial\widetilde{\varphi}(\gamma)}{\partial\kappa_D} = \begin{cases} \frac{1}{\beta}\frac{F(\gamma_D)}{\overline{R}_D+\overline{R}_X}\kappa_D^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} > 0 & \text{for non-exporting firms} \\ \frac{1}{\beta}\frac{F(\gamma_D)}{\overline{R}_D+\overline{R}_X}\phi\kappa_D^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} > 0 & \text{for exporting firms} \end{cases}
$$

$$
\frac{\partial\widetilde{\varphi}(\gamma)}{\partial\kappa_X} = \begin{cases} \frac{1}{\beta}\frac{nF(\gamma_X)}{\overline{R}_D+\overline{R}_X}\kappa_D^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} > 0 & \text{for non-exporting firms} \\ \frac{1}{\beta}\frac{nF(\gamma_X)}{\overline{R}_D+\overline{R}_X}\phi\kappa_D^{\frac{1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta(\beta-\sigma+1)}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{1}{\beta}}\gamma^{-\frac{1}{\beta-\sigma+1}} > 0 & \text{for exporting firms} \end{cases}
$$

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

The claims on the comparative statics of φ thus follow.

Step 3: The FOSD Property

It is readily verified that the c.d.f. of the productivity distribution $G(\varphi)$ is given by

$$
G(\varphi) = \begin{cases} 1 - F\left(\kappa_D^{\frac{\beta-\sigma+1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{\beta-\sigma+1}{\beta}}\varphi^{-(\beta-\sigma+1)}\right)\frac{1}{F(\gamma_D)} \equiv G_1(\varphi) & \varphi \in \left[\varphi_D, \varphi_X^-\right) \\ 1 - \frac{F(\gamma_X)}{F(\gamma_D)} \equiv G_2 & \varphi \in \left[\varphi_X^-, \varphi_X^+\right) \\ 1 - F\left(\left(1+n\tau^{1-\sigma}\right)\kappa_D^{\frac{\beta-\sigma+1}{\beta}}\gamma_D^{\frac{\sigma-1}{\beta}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)^{-\frac{\beta-\sigma+1}{\beta}}\varphi^{-(\beta-\sigma+1)}\right)\frac{1}{F(\gamma_D)} \equiv G_3(\varphi) & \varphi \geq \varphi_X^+, \end{cases}
$$

where $G(\varphi)$ is nondecreasing in φ. By fixing at the same φ, we have $G_1(\varphi) > G_3(\varphi)$ for all φ. An increase in κ_D lowers γ_D, raises γ_X and raises $\widetilde{\gamma}(\varphi)$ for all φ. As a result, we find that $\partial G_1(\varphi)/\partial\kappa_D < 0$, $\partial G_2/\partial\kappa_D < 0$, and $\partial G_3(\varphi)/\partial\kappa_D < 0$ for all φ. Moreover,

$$
\frac{\partial\varphi_D}{\partial\kappa_D} = \frac{\kappa_D^{-1}\varphi_D}{\beta} + \underbrace{\frac{\sigma-1-\beta}{\beta}}_{-}\frac{\gamma_D^{-1}\varphi_D}{\beta-\sigma+1}\underbrace{\frac{\partial\gamma_D}{\partial\kappa_D}}_{-} > 0
$$

$$
\begin{aligned}
\frac{\partial \varphi_X^-}{\partial \kappa_D} =& \frac{\kappa_D^{-1}\varphi_X^-}{\beta} + \frac{\sigma-1-\beta}{\beta}\frac{\varphi_X^-\gamma_D^{-1}}{\beta-\sigma+1}\frac{\partial \gamma_D}{\partial \kappa_D} - \frac{\varphi_X^-\delta^{-1}}{\beta-\sigma+1}\frac{\partial \delta}{\partial \kappa_D} \\
=& \frac{\sigma-1-\beta}{\beta}\frac{\varphi_X^-\gamma_D^{-1}}{\beta-\sigma+1}\frac{\partial \gamma_D}{\partial \kappa_D} + \frac{\kappa_D^{-1}\varphi_X^-}{\beta}\frac{\sigma-1-\beta}{\sigma-1} \\
=& \frac{\kappa_D^{-1}\varphi_X^-}{\beta}\underbrace{\frac{\sigma-1-\beta}{\sigma-1}}_{-}\frac{\kappa_D F(\gamma_D)}{\overline{R}_D+\overline{R}_X} < 0 \\
\frac{\partial \varphi_X^+}{\partial \kappa_D} =& \frac{\kappa_D^{-1}\varphi_X^+}{\beta}\frac{\sigma-1-\beta}{\sigma-1}\frac{\kappa_D F(\gamma_D)}{\overline{R}_D+\overline{R}_X} < 0.
\end{aligned}
$$

Considering that κ_D increases from κ_D to some $\kappa_D^{new} > \kappa_D$, the resulting new productivity distribution is

$$
G^{new}(\varphi) = \begin{cases} G_1^{new}(\varphi) & \varphi \in \left[\varphi_D^{new}, \varphi_X^{-new}\right) \\ G_2^{new} & \varphi \in \left[\varphi_X^{-new}, \varphi_X^{+new}\right) \\ G_3^{new}(\varphi) & \varphi \geq \varphi_X^{+new}, \end{cases}
$$

where $\varphi_D < \varphi_D^{new}$, $\varphi_X^{-new} < \varphi_X^-$, and $\varphi_X^{+new} < \varphi_X^+$. There are two possible cases for the ranking of old and new cutoffs: $\varphi_D < \varphi_D^{new} < \varphi_X^{-new} < \varphi_X^- < \varphi_X^{+new} < \varphi_X^+$ and $\varphi_D < \varphi_D^{new} < \varphi_X^{-new} < \varphi_X^{+new} < \varphi_X^- < \varphi_X^+$.

For the first case, we first inspect the interval $\left[\varphi_D, \varphi_X^-\right)$. Because both $G_1(\varphi)$ and $G_1^{new}(\varphi)$ are increasing in φ, $\lim_{\varphi\to\varphi_X^-} G_1(\varphi) = G_2$, and $\lim_{\varphi\to\varphi_X^{-new}} G_1^{new}(\varphi) = G_2^{new}$, it follows that

$$
G_2 = \lim_{\varphi\to\varphi_X^-} G_1(\varphi) > G_1\left(\varphi_X^{-new}\right) > \lim_{\varphi\to\varphi_X^{-new}} G_1^{new}(\varphi) = G_2^{new}.
$$

Since $\varphi_D < \varphi_D^{new} < \varphi_X^{-new} < \varphi_X^-$, we conclude that $G^{new}(\varphi) < G(\varphi)$ within $\varphi \in \left[\varphi_D, \varphi_X^-\right)$. For the interval $\left[\varphi_X^-, \varphi_X^{+new}\right)$, we immediately conclude that $G(\varphi) > G^{new}(\varphi)$ since $G(\varphi) = G_2 > G_2^{new} = G^{new}(\varphi)$. For the remaining interval $\varphi \geq \varphi_X^{+new}$, note that $G_3(\varphi)$ and $G_3^{new}(\varphi)$ are increasing in φ, $\lim_{\varphi\to\varphi_X^+} G_3(\varphi) = G_2$, and $\lim_{\varphi\to\varphi_X^{+new}} G_3^{new}(\varphi) = G_2^{new}$. It follows that

$$
G_2 = \lim_{\varphi\to\varphi_X^+} G_3(\varphi) > G_3^{new}\left(\varphi_X^+\right) > \lim_{\varphi\to\varphi_X^{+new}} G_3^{new}(\varphi) = G_2^{new}.
$$

Hence we conclude that $G^{new}(\varphi) < G(\varphi)$ for $\varphi \geq \varphi_X^{+new}$. Together, we find that $G(\varphi) > G^{new}(\varphi)$ everywhere.

For the second case, we first focus on the interval $\left[\varphi_D, \varphi_X^{+new}\right)$. Note that

$$
\begin{aligned}
G(\varphi) &= G_1(\varphi) \text{ for } \varphi \in \left[\varphi_D, \varphi_X^{+new}\right) \\
G^{new}(\varphi) &= G_1^{new}(\varphi) \text{ for } \varphi \in \left[\varphi_D^{new}, \varphi_X^{-new}\right)
\end{aligned}
$$

$$G^{new}(\varphi) = G_2^{new} \text{ for } \varphi \in \left[\varphi_X^{-new}, \varphi_X^{+new}\right)$$

and $\varphi_D < \varphi_D^{new} < \varphi_X^{-new} < \varphi_X^{+new}$. Further note that $G_1\left(\varphi_X^{+new}\right) > G_1\left(\varphi_X^{-new}\right) > \lim_{\varphi \to \varphi_X^{-new}} G_1^{new}(\varphi) = G_2^{new}$ since $G(\varphi)$ is decreasing in κ_D. Accordingly we conclude that $G(\varphi) > G^{new}(\varphi)$ in the interval $\left[\varphi_D, \varphi_X^{+new}\right)$. For the interval $\left[\varphi_X^{+new}, \varphi_X^{+}\right)$, observe that

$$G(\varphi) = G_1(\varphi) \text{ for } \varphi \in \left[\varphi_X^{+new}, \varphi_X^{-}\right)$$
$$G^{new}(\varphi) = G_3^{new}(\varphi) \text{ for } \varphi \in \left[\varphi_X^{+new}, \varphi_X^{-}\right)$$
$$G(\varphi) = G_2 \text{ for } \varphi \in \left[\varphi_X^{-}, \varphi_X^{+}\right)$$

and $\varphi_X^{+new} < \varphi_X^{-} < \varphi_X^{+}$. Because $G_1(\varphi) > G_3(\varphi)$, $G_3(\varphi) > G_3^{new}(\varphi)$ and note that $G_2 > G_3(\varphi)$ for $\varphi \le \varphi_X^{+}$, we conclude that $G(\varphi) > G_3(\varphi) > G_3^{new}(\varphi)$ for the interval $\left[\varphi_X^{+new}, \varphi_X^{+}\right)$. For the remaining interval $\varphi \ge \varphi_X^{+}$, we conclude that $G(\varphi) > G^{new}(\varphi)$ because $G(\varphi) = G_3(\varphi) > G_3^{new}(\varphi) = G^{new}(\varphi)$. Together, we find that $G(\varphi) > G^{new}(\varphi)$ everywhere.

Since $G(\varphi) > G^{new}(\varphi)$ everywhere in both cases, the new distribution FOSDs the old one. □

A.5 Proof of Proposition 5

Let

$$P_D^{1-\sigma} \equiv M_e \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\sigma-1}{\beta-\sigma+1}} N^{\frac{\sigma-1}{\beta-\sigma+1}} P^{(\sigma-1)\frac{\sigma-1}{\beta-\sigma+1}} \left[(\Gamma_D - \Gamma_X) + \phi^{\sigma-1}\Gamma_X\right] \tag{60}$$

$$P_X^{1-\sigma} \equiv M_e \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\sigma-1}{\beta-\sigma+1}} N^{\frac{\sigma-1}{\beta-\sigma+1}} P^{(\sigma-1)\frac{\sigma-1}{\beta-\sigma+1}} \phi^{\sigma-1}\tau^{1-\sigma}\Gamma_X \tag{61}$$

denote the average prices of domestic and foreign products respectively. The aggregate price is $P^{1-\sigma} = P_D^{1-\sigma} + nP_X^{1-\sigma}$, and the expenditure share on domestic products is

$$\lambda \equiv \frac{P_D^{1-\sigma}}{P^{1-\sigma}} = \frac{1 + \left(\phi^{\sigma-1} - 1\right)\frac{\Gamma_X}{\Gamma_D}}{1 + \left[\left(1 + n\tau^{1-\sigma}\right)\phi^{\sigma-1} - 1\right]\frac{\Gamma_X}{\Gamma_D}}. \tag{62}$$

The expenditure share on the products imported from a symmetric foreign country is analogously given by

$$\lambda_X \equiv \frac{1-\lambda}{n} = \frac{\phi^{\sigma-1}\tau^{1-\sigma}\frac{\Gamma_X}{\Gamma_D}}{1 + \left[\left(1 + n\tau^{1-\sigma}\right)\phi^{\sigma-1} - 1\right]\frac{\Gamma_X}{\Gamma_D}}. \tag{63}$$

From (37) and (38) we have

$$d\ln\gamma_D = \beta d\ln P \tag{64}$$

$$d\ln\gamma_X = \beta d\ln P + d\ln\delta. \tag{65}$$

Note that the assumption that $E(\Pi) < \infty$ ensures that Γ_D and Γ_X are both finite. We further define the following short-hand notation

$$\eta_z \equiv \frac{\gamma_z^{1-\frac{\sigma-1}{\beta-\sigma+1}} f(\gamma_z)}{\Gamma_z} \; z \in \{D, X\}.$$

Now we show that our model entails the ACR formula. The welfare is defined as the real income $W \equiv N/P$. By (43), it follows that

$$\frac{d\ln W}{d\ln\tau} = -\frac{1}{\beta}\frac{d\ln\gamma_D}{d\ln\tau}.$$

Since γ_D is identified by (42) as

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

it follows that

$$\begin{aligned}
\frac{\sigma-1}{\beta-\sigma+1} d\ln\gamma_D =& d\ln\left(\kappa_e + \kappa_D F(\gamma_D) + n\kappa_X F(\gamma_X)\right) - d\ln\left\{\Gamma_D + \left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\Gamma_X\right\} \\
=& \frac{\kappa_D F(\gamma_D)}{\kappa_e + \kappa_D F(\gamma_D) + n\kappa_X F(\gamma_X)} \frac{\gamma_D f(\gamma_D)}{F(\gamma_D)} d\ln\gamma_D + \frac{n\kappa_X F(\gamma_X)}{\kappa_e + \kappa_D F(\gamma_D) + n\kappa_X F(\gamma_X)} \frac{\gamma_X f(\gamma_X)}{F(\gamma_X)} d\ln\gamma_X \\
& - \frac{1}{1 + \left[(1+n\tau^{1-\sigma})^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\frac{\Gamma_X}{\Gamma_D}} \eta_D d\ln\gamma_D - \frac{\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\frac{\Gamma_X}{\Gamma_D}}{1 + \left[(1+n\tau^{1-\sigma})^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\frac{\Gamma_X}{\Gamma_D}} \eta_X d\ln\gamma_X \\
& - \frac{\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\frac{\Gamma_X}{\Gamma_D}}{1 + \left[(1+n\tau^{1-\sigma})^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\frac{\Gamma_X}{\Gamma_D}} d\ln\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}} - 1\right].
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{\kappa_D F(\gamma_D)}{\kappa_e + \kappa_D F(\gamma_D) + n\kappa_X F(\gamma_X)} \frac{\gamma_D f(\gamma_D)}{F(\gamma_D)} =& \frac{\kappa_D \gamma_D^{1-\frac{\sigma-1}{\beta-\sigma+1}} f(\gamma_D)}{\kappa_D \Gamma_D\left\{1 + \left[(1+n\tau^{1-\sigma})^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\frac{\Gamma_X}{\Gamma_D}\right\}} \\
=& \frac{1}{1 + \left[(1+n\tau^{1-\sigma})^{\frac{\beta}{\beta-\sigma+1}} - 1\right]\frac{\Gamma_X}{\Gamma_D}} \eta_D,
\end{aligned}$$

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

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

Recall that $\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}=\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}$; we have

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

$$=\beta\frac{\sigma-1}{\beta-\sigma+1}\frac{n\tau^{1-\sigma}\phi^{\sigma-1}\frac{\Gamma_X}{\Gamma_D}}{1+\left[\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}-1\right]\frac{\Gamma_X}{\Gamma_D}}d\ln\tau$$

$$=\beta\frac{\sigma-1}{\beta-\sigma+1}(1-\lambda)\,d\ln\tau$$

$$\Rightarrow\frac{d\ln W}{d\ln\tau}=-\frac{1}{\beta}\frac{d\ln\gamma_D}{d\ln\tau}=\lambda-1. \tag{66}$$

Next, we show that (66) is consistent with the ACR formula. Note that trade elasticity is defined as $d\ln(\lambda_X/\lambda)/d\ln\tau$; thus under the symmetric country assumption we can restate the ACR formula as

$$\frac{d\ln W}{d\ln\tau}=\frac{1}{\varepsilon}\frac{d\ln\lambda}{d\ln\tau}$$

$$=\frac{d\ln\lambda/d\ln\tau}{d\ln(\lambda_X/\lambda)/d\ln\tau}$$

$$=\frac{d\ln\lambda/d\ln\tau}{d\ln\left(\frac{1-\lambda}{\lambda}\right)/d\ln\tau}$$

$$=(\lambda-1)\frac{d\ln\lambda/d\ln\tau}{d\ln\lambda/d\ln\tau}$$

$$=\lambda-1,$$

which is equivalent to (66).

For the trade elasticity, log-differentiating (62) and (63) with respect to τ yields

$$d\ln\frac{\lambda_X}{\lambda}=d\ln\left(\phi^{\sigma-1}\tau^{1-\sigma}\frac{\Gamma_X}{\Gamma_D}\right)-d\ln\left[1+\left(\phi^{\sigma-1}-1\right)\frac{\Gamma_X}{\Gamma_D}\right]$$

$$=(1-\sigma)\,d\ln\tau+(\sigma-1)\,d\ln\phi+d\ln\frac{\Gamma_X}{\Gamma_D}$$

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

where $d\ln\delta=d\ln\left(\gamma_X/\gamma_D\right)$ by (64) and (65), and $d\ln\gamma_D/d\ln\tau=\beta\left(1-\lambda\right)$ by (66). It is readily verified that $\frac{1-\frac{\Gamma_X}{\Gamma_D}}{1+\left(\phi^{\sigma-1}-1\right)\frac{\Gamma_X}{\Gamma_D}}>0$, $\frac{d\ln\phi}{d\ln\tau}<0$, and $\frac{d\ln\delta}{d\ln\tau}<0$. □

A.6 Proof of Proposition 6

Without selection to export, we always have $\Pi_X\left(\gamma\right)>\Pi_D\left(\gamma\right)$ so all firms serve the foreign markets on survival. As in Section 3.1, it is readily verified that

$$
\varphi^{\kappa_X=0}\left(\gamma\right)=\left(1+n\tau^{1-\sigma}\right)^{\frac{1}{\beta-\sigma+1}}\left[\frac{N}{\left(P^{\kappa_X=0}\right)^{1-\sigma}}\right]^{\frac{1}{\beta-\sigma+1}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{1}{\beta-\sigma+1}}\gamma^{-\frac{1}{\beta-\sigma+1}}
$$

$$
\left(\gamma_D^{\kappa_X=0}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}=\frac{1}{\kappa_D}\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\left[\frac{N}{\left(P^{\kappa_X=0}\right)^{1-\sigma}}\right]^{\frac{\beta}{\beta-\sigma+1}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right) \tag{67}
$$

$$
\kappa_e=\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\left[\frac{N}{\left(P^{\kappa_X=0}\right)^{1-\sigma}}\right]^{\frac{\beta}{\beta-\sigma+1}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)\Gamma_D^{\kappa_X=0} \tag{68}
$$

$$
-\kappa_D F\left(\gamma_D^{\kappa_X=0}\right) \tag{69}
$$

$$
\begin{aligned}
\left(P^{\kappa_X=0}\right)^{1-\sigma}=&M_e^{\kappa_X=0}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\sigma-1}{\beta-\sigma+1}}\left[\frac{N}{\left(P^{\kappa_X=0}\right)^{1-\sigma}}\right]^{\frac{\sigma-1}{\beta-\sigma+1}}\left(1+n\tau^{1-\sigma}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}\Gamma_D^{\kappa_X=0} \qquad (70)\\
&+nM_e^{\kappa_X=0}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\sigma-1}{\beta-\sigma+1}}\left[\frac{N}{\left(P^{\kappa_X=0}\right)^{1-\sigma}}\right]^{\frac{\sigma-1}{\beta-\sigma+1}}\tau^{1-\sigma}\left(1+n\tau^{1-\sigma}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}\Gamma_D^{\kappa_X=0}
\end{aligned}
$$

$$
\begin{aligned}
\lambda^{\kappa_X=0}=&\frac{M_e^{\kappa_X=0}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\sigma-1}{\beta-\sigma+1}}\left[\frac{N}{\left(P^{\kappa_X=0}\right)^{1-\sigma}}\right]^{\frac{\sigma-1}{\beta-\sigma+1}}\left(1+n\tau^{1-\sigma}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}\Gamma_D^{\kappa_X=0}}{\left(P^{\kappa_X=0}\right)^{1-\sigma}}\\
=&\frac{1}{1+n\tau^{1-\sigma}}.
\end{aligned}
$$

Since

$$
\begin{aligned}
d\ln\lambda^{\kappa_X=0} =& d\ln M_e^{\kappa_X=0} + \frac{\sigma-1}{\beta-\sigma+1}\left(\sigma-1\right)d\ln P^{\kappa_X=0} + \frac{\sigma-1}{\beta-\sigma+1}d\ln\left(1+n\tau^{1-\sigma}\right) \\
&+ \eta_D^{\kappa_X=0}d\ln\gamma_D^{\kappa_X=0} - \left(1-\sigma\right)d\ln P^{\kappa_X=0} \\
d\ln\lambda_X^{\kappa_X=0} =& d\ln M_e^{\kappa_X=0} + \frac{\sigma-1}{\beta-\sigma+1}\left(\sigma-1\right)d\ln P^{\kappa_X=0} + \frac{\sigma-1}{\beta-\sigma+1}d\ln\left(1+n\tau^{1-\sigma}\right) \\
&+ \eta_D^{\kappa_X=0}d\ln\gamma_D^{\kappa_X=0} + \left(1-\sigma\right)d\ln\tau - \left(1-\sigma\right)d\ln P^{\kappa_X=0},
\end{aligned}
$$

the trade elasticity is given by

$$
\varepsilon^{\kappa_X=0} = 1-\sigma < 0.
$$

By combining (67) and (68) we have

$$
\left(\gamma_D^{\kappa_X=0}\right)^{\frac{\sigma-1}{\beta-\sigma+1}} = \frac{\kappa_e + \kappa_D F\left(\gamma_D^{\kappa_X=0}\right)}{\kappa_D\Gamma_D^{\kappa_X=0}}. \tag{71}
$$

Using (67) and (70) we obtain

$$
\left(\gamma_D^{\kappa_X=0}\right)^{\frac{\sigma-1}{\beta-\sigma+1}} = \frac{N\frac{\sigma-1}{\beta\sigma}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)}{\kappa_D M_e^{\kappa_X=0}\Gamma_D}.
$$

By combining the above equation with (71), the mass of entrants is given by

$$
M_e^{\kappa_X=0} = \frac{N\frac{\sigma-1}{\beta\sigma}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)}{\kappa_e + \kappa_D F\left(\gamma_D^{\kappa_X=0}\right)}. \tag{72}
$$

Inserting (72) into (70) we have

$$
\left(P^{\kappa_X=0}\right)^{\frac{(1-\sigma)\beta}{\beta-\sigma+1}} = \frac{\left(\frac{\beta-\sigma+1}{\sigma-1}\right)N^{\frac{\beta}{\beta-\sigma+1}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\beta}{\beta-\sigma+1}}\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\Gamma_D^{\kappa_X=0}}{\kappa_e + \kappa_D F\left(\gamma_D^{\kappa_X=0}\right)}. \tag{73}
$$

Log-differentiating (73) yields

$$
\begin{aligned}
\frac{(1-\sigma)\beta}{\beta-\sigma+1}d\ln P^{\kappa_X=0} =& \frac{\beta}{\beta-\sigma+1}d\ln\left(1+n\tau^{1-\sigma}\right) + d\ln\Gamma_D^{\kappa_X=0} - \frac{\kappa_D}{\kappa_e+\kappa_D F\left(\gamma_D^{\kappa_X=0}\right)}f\left(\gamma_D^{\kappa_X=0}\right)\gamma_D^{\kappa_X=0}d\ln\gamma_D^{\kappa_X=0} \\
=& \frac{\beta}{\beta-\sigma+1}d\ln\left(1+n\tau^{1-\sigma}\right) + \eta_D^{\kappa_X=0}d\ln\gamma_D^{\kappa_X=0} - \left(\gamma_D^{\kappa_X=0}\right)^{1-\frac{\sigma-1}{\beta-\sigma+1}}\frac{f\left(\gamma_D^{\kappa_X=0}\right)}{\Gamma_D^{\kappa_X=0}}d\ln\gamma_D^{\kappa_X=0} \\
=& \frac{\beta}{\beta-\sigma+1}d\ln\left(1+n\tau^{1-\sigma}\right) + \eta_D^{\kappa_X=0}d\ln\gamma_D^{\kappa_X=0} - \eta_D^{\kappa_X=0}d\ln\gamma_D^{\kappa_X=0} \\
=& \frac{\beta}{\beta-\sigma+1}d\ln\left(1+n\tau^{1-\sigma}\right)
\end{aligned}
$$

$$\Rightarrow d\ln W^{\kappa_X=0} = -\frac{n\tau^{1-\sigma}}{1+n\tau^{1-\sigma}} d\ln\tau$$
$$= \left(\lambda^{\kappa_X=0}-1\right) d\ln\tau.$$

In Appendix A.5 we have shown that under the symmetric country assumption, the ACR formula implies that $d\ln W = (\lambda-1)\,d\ln\tau$. Therefore, the case without firm selection to export also belongs to the ACR class.

Finally, we show that the welfare gains from trade is lower than the case with selection to export. This is done by simply comparing the magnitude of $\lambda-1$. Suppose that $|d\ln W| > |d\ln W^{\kappa_X=0}|$. It follows that

$$\begin{aligned}
&1-\lambda > 1-\lambda^{\kappa_X=0}\\
\Rightarrow &\frac{n\tau^{1-\sigma}\left(1+n\tau^{1-\sigma}\right)^{\frac{\sigma-1}{\beta-\sigma+1}}\frac{\Gamma_X}{\Gamma_D}}{1+\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1\right]\frac{\Gamma_X}{\Gamma_D}} > \frac{n\tau^{1-\sigma}}{1+n\tau^{1-\sigma}}\\
\Rightarrow &\frac{\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}\frac{\Gamma_X}{\Gamma_D}}{1+\left[\left(1+n\tau^{1-\sigma}\right)^{\frac{\beta}{\beta-\sigma+1}}-1\right]\frac{\Gamma_X}{\Gamma_D}} > 1\\
\Rightarrow &\Gamma_X > \Gamma_D,
\end{aligned}$$

which contradicts the property that $\gamma_D > \gamma_X$. Therefore, it must be that $|d\ln W| < |d\ln W^{\kappa_X=0}|$.

□

A.7 Derivation of λ^{MP}, (52), and (53)

When φ follows a Pareto distribution, it is readily verified that

$$\varphi_D^{MP} = \left[\frac{\kappa_D}{N}\frac{\sigma-1}{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}\right]^{\frac{1}{\sigma-1}}\frac{1}{P^{MP}} \tag{74}$$

$$\varphi_X^{MP} = \tau\left[\frac{\kappa_X}{N}\frac{\sigma-1}{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}\right]^{\frac{1}{\sigma-1}}\frac{1}{P^{MP}} \tag{75}$$

$$\begin{aligned}
\left(P^{MP}\right)^{1-\sigma} &= M_e^{MP}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\theta^{MP}\left[\int_{\varphi_D^{MP}}^{\infty}\varphi^{\sigma-\theta^{MP}-2}d\varphi + n\tau^{1-\sigma}\int_{\varphi_X^{MP}}^{\infty}\varphi^{\sigma-\theta^{MP}-2}d\varphi\right]\\
&= M_e^{MP}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\frac{\theta^{MP}}{\theta^{MP}-\sigma+1}\left(\varphi_D^{MP}\right)^{\sigma-\theta^{MP}-1}\left[1+n\tau^{1-\sigma}\left(\frac{\varphi_X^{MP}}{\varphi_D^{MP}}\right)^{\sigma-\theta^{MP}-1}\right].
\end{aligned} \tag{76}$$

Therefore,

$$\lambda^{MP} = \frac{M_e^{MP}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\frac{\theta^{MP}}{\theta^{MP}-\sigma+1}\left(\varphi_D^{MP}\right)^{\sigma-\theta^{MP}-1}}{\left(P^{MP}\right)^{1-\sigma}} = \frac{1}{1+n\tau^{1-\sigma}\left(\frac{\varphi_X^{MP}}{\varphi_D^{MP}}\right)^{\sigma-\theta^{MP}-1}}.$$

Using (74), (75), and (76), the price index can be expressed as

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

The economy becomes autarky when $\tau\to\infty$, where the price index in this case is denoted as $P_{\tau\to\infty}^{MP}$. Since it is well-known that M_e^{MP} is independent of τ when productivity follows a Pareto distribution, it follows that

$$\frac{W^{MP}}{W_{\tau\to\infty}^{MP}} = \frac{P_{\tau\to\infty}^{MP}}{P^{MP}} = \left[1+n\tau^{1-\sigma}\left(\frac{\varphi_X^{MP}}{\varphi_D^{MP}}\right)^{\sigma-\theta^{MP}-1}\right]^{\frac{1}{\theta^{MP}}}.$$

Equation (53) thus follows by inserting (74) and (75) into the above equality.

As for (52), it is readily verified that

$$M_e = \frac{N}{\kappa_e}\frac{\sigma-1}{\theta}\frac{\sigma-1}{\beta\sigma}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)$$

$$\gamma_D = \left[\kappa_D^{-1}\left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\beta}{\theta}}\left(\frac{\sigma-1}{\sigma}\right)^{\frac{\sigma\beta}{\theta}}\beta^{-\frac{\beta}{\theta}}\left(\frac{\beta-\sigma+1}{\sigma-1}\right)\right]^{\frac{\theta}{\sigma-1}} \tag{77}$$

$$P^{1-\sigma} = M_e\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\frac{\theta}{\theta-\sigma+1}\left(\frac{N}{P^{1-\sigma}}\right)^{\frac{\sigma-1}{\theta}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\sigma-1}{\theta}}\gamma_D^{1-\frac{\sigma-1}{\theta}} \times\left[1-\left(\frac{\gamma_X}{\gamma_D}\right)^{1-\frac{\sigma-1}{\theta}}+\left(1+n\tau^{1-\sigma}\right)\phi^{\sigma-1}\left(\frac{\gamma_X}{\gamma_D}\right)^{1-\frac{\sigma-1}{\theta}}\right] \tag{78}$$

where $\theta\equiv\beta-\sigma+1$. Inserting (77) into (78) and rearranging, we obtain

$$P^{-\beta} = M_e\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\frac{\theta}{\theta-\sigma+1}N^{\frac{\theta}{\sigma-1}}\left[\frac{\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\beta}\right]^{\frac{\theta}{\sigma-1}}\left(\frac{\theta}{\kappa_D(\sigma-1)}\right)^{\frac{\theta}{\sigma-1}-1}$$

$$\times \left[1 - \left(\frac{\gamma_X}{\gamma_D} \right)^{1-\frac{\sigma-1}{\theta}} + \left(1 + n\tau^{1-\sigma} \right) \phi^{\sigma-1} \left(\frac{\gamma_X}{\gamma_D} \right)^{1-\frac{\sigma-1}{\theta}} \right].$$

We denote the price index under autarky as $P_{\tau \to \infty}$. Since M_e is independent of τ, it follows that

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

Using (34) and (39), (52) follows from the above equality. □