DISCUSSION PAPER SERIES

DP17092
(v. 3)

SELECTION AND SORTING OF HETEROGENEOUS FIRMS THROUGH COMPETITIVE PRESSURES

Kiminori Matsuyama and Philip Ushchev

INDUSTRIAL ORGANIZATION, INTERNATIONAL TRADE AND REGIONAL ECONOMICS AND MACROECONOMICS AND GROWTH

CEPR

ISSN 0265-8003

SELECTION AND SORTING OF HETEROGENEOUS FIRMS THROUGH COMPETITIVE PRESSURES

Kiminori Matsuyama and Philip Ushchev

Discussion Paper DP17092
First Published 07 March 2022
This Revision 27 March 2025

Centre for Economic Policy Research
187 boulevard Saint-Germain, 75007 Paris, France
2 Coldbath Square, London EC1R 5HL
Tel: +44 (0)20 7183 8801
www.cepr.org

SELECTION AND SORTING OF HETEROGENEOUS FIRMS THROUGH COMPETITIVE PRESSURES

Abstract

We study how heterogenous firms differ in their responses to increased competitive pressures caused by a lower entry cost, an increase in market size, and globalization, and how these changes affect selection and sorting of heterogeneous firms. To this end, we model a monopolistic competitive sector with heterogenous firms and free-entry using H.S.A. (Homothetic Single Aggregator) class of demand systems, which contains CES and translog as special cases. H.S.A. is tractable due to its homotheticity and to its single aggregator that serves as a sufficient statistic for competitive pressures. It is also flexible enough to allow for the choke price, the 2nd and 3rd laws of demand. We prove the existence and uniqueness of equilibrium and conduct comparative static analysis on the distributions of the markup and pass-through rates across firms, as well as the firm size distribution, measured in the revenue, profit, and employment with sharp analytical results, often just by using simple diagrams. A decline in the entry and overhead costs, and an increase in market size create more competitive pressures, which lead to a decline in the markup rate (under the 2nd law) and an increase in the pass-through rate (under the 3rd law) in all firms. At the same time, they cause reallocation to more productive firms with higher markup and lower pass-through rates. Due to such a composition effect, the average markup rate and the aggregate profit share may go up and the average pass-through rate may go down due to (not in spite of) more competitive pressures. We also characterize how heterogenous firms sort themselves across markets with different sizes and how the markup and pass-through rates of heterogenous firms in the domestic and export markets respond to globalization caused by a reduction in the iceberg cost.

JEL Classification: D4, E2, L1, O4

Keywords: Heterogeneous firms, HSA, 2nd and 3rd laws of demand, Markup and pass-through rates, Selection, Sorting, Iceberg trade costs

Kiminori Matsuyama - k-matsuyama@northwestern.edu
Northwestern University and CEPR

Philip Ushchev - ph.ushchev@gmail.com
ECARES, Université Libre De Bruxelles and CEPR

Acknowledgements
We are grateful to the participants at numerous conferences, seminars, and workshops for their feedback, many of whom followed up with their comments and suggestions through personal communications. They are so numerous to thank them individually. Our special thanks go to Elhanan Helpman and Gianmarco Ottaviano, who acted as formal discussants. The usual disclaimer applies.

Selection and Sorting of Heterogeneous Firms through Competitive Pressures

Kiminori Matsuyama
Northwestern University
k-matsuyama@northwestern.edu

Philip Ushchev
ECARES, Université Libre de Bruxelles
ph.ushchev@gmail.com

First Version: March 7, 2022
Second Version: August 5, 2023
This Version: March 27, 2025

2025-03-27: 4:22 PM

Abstract: We study how heterogenous firms differ in their responses to increased competitive pressures caused by a lower entry cost, an increase in market size, and globalization, and how these changes affect selection and sorting of heterogeneous firms. To this end, we model a monopolistic competitive sector with heterogenous firms and free-entry using H.S.A. (*Homothetic Single Aggregator*) class of demand systems, which contains CES and translog as special cases. H.S.A. is tractable due to its homotheticity and to its single aggregator that serves as a sufficient statistic for competitive pressures. It is also flexible enough to allow for the choke price, the 2nd and 3rd laws of demand. We prove the existence and uniqueness of equilibrium and conduct comparative static analysis on the distributions of the markup and pass-through rates across firms, as well as the firm size distribution, measured in the revenue, profit, and employment with sharp analytical results, often just by using simple diagrams. A decline in the entry and overhead costs, and an increase in market size create more competitive pressures, which lead to a decline in the markup rate (under the 2nd law) and an increase in the pass-through rate (under the 3rd law) in all firms. At the same time, they cause reallocation to more productive firms with higher markup and lower pass-through rates. Due to such a composition effect, the average markup rate and the aggregate profit share may go up and the average pass-through rate may go down *due to* (not in spite of) more competitive pressures. We also characterize how heterogenous firms sort themselves across markets with different sizes and how the markup and pass-through rates of heterogenous firms in the domestic and export markets respond to globalization caused by a reduction in the iceberg cost.

Keywords: Heterogeneous firms, H.S.A., competitive pressures, the 2nd and 3rd laws, markup and pass-through rates, selection, sorting, iceberg trade cost, the composition effect, log-supermodularity.

Acknowledgement: We are grateful to the participants at numerous conferences, seminars, and workshops for their feedback, many of whom followed up with their comments and suggestions through personal communications. They are so numerous to thank them individually. Our special thanks go to Elhanan Helpman and Gianmarco Ottaviano, who acted as formal discussants. The usual disclaimer applies.

1. Introduction.

How do heterogenous firms differ in their responses to increased competitive pressures caused by a lower entry cost, an increase in market size, and/or globalization? How do these changes affect selection of heterogeneous firms? Or sorting of heterogeneous firms across different markets? And what are the impacts on the distribution of markup rates and pass-through rates, as well as the distribution of firm size, measured in revenue, profit, and employment? In the Melitz (2003) model of monopolistic competition with firm heterogeneity and free-entry, all firms sell their products at an exogenous and common markup rate with the pass-through rate equal to one, due to its assumption of CES demand system. Thus, the markup and pass-through rates are not only common across firms but also unresponsive to competitive pressures. Furthermore, a change in market size has no effect on selection and sorting of heterogenous firms and their behaviors, with all adjustments taking place at the extensive margin.

In this paper, we model a monopolistic competitive (MC) sector with firm heterogeneity and free entry, using the H.S.A. (*Homothetic Single Aggregator*) class of demand systems, thereby allowing for different firms to set different markup rates with different pass-through rates, both of which are responsive to a change in competitive pressures.[1] H.S.A. has many features that make it suitable for analyzing monopolistic competition with heterogenous firms and free entry.

First, H.S.A. is homothetic, unlike most non-CES demand systems that have been applied to monopolistic competition.[2] Introducing variable elasticity of demand without giving up homotheticity helps to isolate the roles of endogenous markup rates from those of nonhomotheticity. Moreover, the composition of market demand does not matter under homotheticity, which allows us to define a single measure of market size, even though market size of a sector may change for a variety of reasons, such as

---

[1]H.S.A. was first introduced by Matsuyama and Ushchev (2017, section 3). Matsuyama and Ushchev (2020a, 2022) restrict it further by defining it over a continuum of varieties and imposing gross substitutability to make it suitable for monopolistic competition settings. Other applications of H.S.A. to monopolistic competition include Baqaee, Farhi, and Sangani (2024), Fujiwara and Matsuyama (2022), and Grossman, Helpman, and Lhuiller (2023). Among these, Baqaee, Farhi and Sangani (2024) complements the theory in the present paper, as they calibrate H.S.A. using the firm-level data from Belgium.

[2]For example, Dixit and Stiglitz (1977, Section II) extended their monopolistic competition model to the directly explicitly additive (DEA) demand systems, which have been further explored by Krugman (1979), Behrens and Murata (2007), Zhelobodko, et.al. (2012), Melitz (2018), Dhingra and Morrow (2019), Latzer, Matsuyama, and Parenti (2019), Behrens et.al. (2020), Mayer et.al. (2021), Kokovin et. al. (2024), among many others. This class can also rationalize the reduced-form profit functions assumed in Mrázová-Neary (2017; 2019) and Nocke (2006). Though Dixit and Stiglitz called this class, “Variable Elasticity Case,” the well-known Bergson’s Law states that, within this class of demand systems, they are homothetic if and only if they are CES. In other words, any departure from CES within DEA introduces nonhomotheticity. Moreover, DEA is subject to Pigou’s Law, positive correlation with the income and price elasticities of demand, which has been rejected empirically by Deaton (1974) and many others. The linear-quadratic demand system introduced by Ottaviano, Tabuchi, and Thisse, (2002) and applied to monopolistic competition with heterogenous firms by Melitz and Ottaviano (2008) is also nonhomothetic. See Thisse and Ushchev (2018) for a survey of monopolistic competition with non-CES demand systems. Parenti, Ushchev and Thisse (2017) provides a unified treatment of this literature. Matsuyama (2023) offers a broader overview of non-CES demand systems.

productivity growth, globalization, a sectoral shift in demand, a change in the population size, etc.[3] Furthermore, a MC sector with homothetic demand systems remains tractable when it is embedded in a multi-sector model, because assuming homotheticity in every level of aggregation (except possibly at the highest level) allows for solving a model by using multi-stage budgeting procedure.[4]

Second, H.S.A. is flexible. It can accommodate the choke price, as well as the so-called Marshall's 2nd law of demand, "a higher price leads to a higher price elasticity," which implies incomplete pass-through--less productive firms have lower markup rates--, and what we call the 3rd law of demand, "a higher price leads to a smaller rate of change in the price elasticity," which implies that less productive firms have higher pass-through rates,[5] for which there is some supporting empirical evidence.[6] Furthermore, since this class contains CES (as well as translog) as a special case, H.S.A. can be used to perform the robustness check; it helps us understand which properties of the Melitz model, which assumes CES, carry over to a broader class of homothetic demand systems.

Third, H.S.A. retains much of the tractability of CES. This is due to its single aggregator property; that is the market share of each firm is a function of a single variable, its own price normalized by the single price aggregator, which serves as a sufficient statistic for competitive pressures, as it captures all the cross-product interactions in the demand systems, whether due to a change in the mass of active firms or a change in the prices of competing products. Furthermore, due to its homotheticity, the single aggregator enters all firm-specific variables (the markup and pass-through rates, the profit, the revenue and the employment) proportionately with the firm's marginal cost, so that competitive pressures act as a magnifier of firm heterogeneity. This allows us to take advantage of log-supermodularity[7] to

[3] In contrast, under nonhomotheticity, how firms would respond to a change in market size depends on, for example, whether it is caused by a change in the population size or by a change in per capita income. Moreover, allowing consumers to differ in their income would cause significant complications, which have not been successfully addressed.

[4] In contrast, most MC models with nonhomothetic demand systems assume that there is only one sector. This is because solving a MC sector with nonhomothetic demand systems in a multi-sector setting requires some additional restriction, such as assuming that there is only one outside sector that produces a homogeneous good competitively, or every sector has the same parametric family of nonhomothetic demand systems with identical parameter values. Perhaps this is the reason why MC models with nonhomothetic non-CES have not been used in the macro literature. Matsuyama (2025) offers more discussion on the advantages of homotheticity.

[5] About the terminology: Marshall's 1st law of demand states that a higher price reduces demand; it imposes the restriction on the 1st derivative of the demand curve. Marshall's 2nd law states that a higher price increases the price elasticity; it imposes the restriction on the 2nd derivative. We call the law stating that a higher price reduces the rate of change in the price elasticity as the 3rd law because it imposes the restriction on the 3rd derivative.

[6] For the empirical evidence on the 2nd law and incomplete pass-through, as well as the closely related but distinct concepts of the procompetitive effect and strategic complementarity in pricing, see Campbell and Hopenhayn (2005); Burstein-Gopinath (2014), DeLoecker and Goldberg (2014), Feenstra and Weinstein (2017), and Amiti, Itskhoki, and Konings (2019); For the empirical evidence on the 3rd law, see Berman, Martin, and Mayer (2012) and Amiti, Itskhoki, and Konings (2014). Recently, , Farhi, and Sangani (2024) nonparametrically calibrated H.S.A., with the results in support of the 2nd and the 3rd laws. In contrast, homothetic translog, applied to monopolistic competition by Feenstra (2003) and popular in the trade literature, satisfies the 2nd law but violates the 3rd law. It is also an isolated example and hence cannot be used as a tool for the robustness check for CES. (This motivated Matsuyama and Ushchev (2020a, 2022) to develop Generalized Translog, a family within H.S.A. that nests both CES and translog. See Appendix D.1.)

[7] See, for example, Costinot (2009) and Costinot and Vogel (2010; 2015).

study the differential impacts of competitive pressures on heterogeneous firms. It also enables us to use simple diagrams to prove the existence and uniqueness of free-entry equilibrium with firm heterogeneity and to conduct most comparative statics, which generate sharp analytical results without imposing any *parametric* restrictions on the demand system and productivity distribution.[8] Moreover, there is no need to assume zero overhead cost for tractability, unlike Melitz and Ottaviano (2008) and Arkolakis et.al. (2019) and most others that extend the Melitz model by introducing the procompetitive effect (Mayer et.al. 2021 being the exception). This is important not only because it makes our model applicable to the sectors characterized by high overhead costs and allows us to study the effects of the recent rise in overhead costs. Indeed, a combination of firm heterogeneity and the 2nd and 3rd laws of demand generates some new insights when the overhead cost is sufficiently high.[9]

Here's the summary of what we find.[10] Though a subset of these results can be shown under some other demand systems, the advantage of H.S.A. is that it enables us to put all of them together within a single framework, due to its flexibility and tractability.

- More productive firms, which always have higher profits and revenues, have higher markup rates under the 2nd law and lower pass-through rates under the 3rd law. Employments are not monotone in firm productivity; they are *hump-shaped* under the 2nd and 3rd laws. The 2nd law also implies the procompetitive effect and strategic complementarity in pricing.
- *A lower entry cost* leads to more competitive pressures, which reduces the markup rates of all firms under the 2nd law and raises the pass-through rates of all firms under the 3rd law. The profits of all firms decline (at faster rates among less productive firms under the 2nd law), which leads to a tougher selection. The revenues of all firms also decline (at faster rates among less productive firms under the 3rd law). *A lower overhead cost* has similar effects when the employment is decreasing in firm productivity, which occurs under the 2nd and the 3rd laws for a sufficiently high overhead cost.
- *Larger market size* also leads to more competitive pressures, reducing the markup rates of all firms under the 2nd law and raises the pass-through rates of all firms under the 3rd law. The profits among

[8] In contrast, under the two other classes of demand systems studied in Matsuyama and Ushchev (2020a), HDIA, which contains the Kimball (1995) demand system as a special case, and HIIA, we need the two aggregators, one for competitive pressures due to a change in the pricing of competing firms, and another for competitive pressures due to a change in the mass of firms. This poses a challenge for ensuring the existence and the uniqueness of the free-entry equilibrium and for conducting comparative statics exercises even in a single-market setting, since it would require further restrictions on the firm productivity distribution and the demand system. (Matsuyama and Ushchev (2020a) found the condition of the existence and the uniqueness under HDIA and HIIA only for the case of homogeneous firms.) The problem of ensuring the existence and the uniqueness under HDIA and HIIA would be even more challenging in a multi-market setting, which we develop in section 6 to study sorting of firms across markets. Generally, H.S.A. is more analytically tractable than HDIA and HIIA, when one needs to compare across the equilibriums in which different sets of firms are active. Matsuyama (2025) offers more discussion on the advantages of H.S.A., relative to other classes of homothetic demand systems.

[9] Another advantage of H.S.A., pointed out by Kasahara and Sugita (2020), is that the market share (in revenue) functions are the primitive of H.S.A., hence it can be readily identified with the typical firm-level data, which contain revenue, but not the output.

[10] The reader is also referred to Matsuyama (2025, Section 7), which offers an executive summary of these results.

more productive firms increase, while those among less productive decline under the 2nd law, which leads to a tougher selection. The revenues among more productive firms also increase, while those among less productive decline under the 3rd law at least when the overhead cost is not too large.

- An increase in competitive pressures due to a lower entry cost, a lower overhead cost, and a larger market size may lead to an *increase* in the (revenue-, profit- or employment-) weighted generalized (including arithmetic, geometric, and harmonic) mean of the firm-level markup rates under the 2nd law, despite that each surviving firm reduces its markup rate. This also means that the aggregate profit share increases due to more competitive pressures. These shocks may also lead to a *decline* in the weighted generalized mean of the firm-level pass-through rate under the 3rd law, despite that each surviving firm increases its pass-through rate. This is because they cause less productive firms with lower markup rates and higher pass-through rates to shrink and to exit, changing the composition of firms. For example, in response to a change in the entry cost, this composition effect dominates the effect on individual firms when the elasticity of marginal cost density is an increasing function, as found empirically in the calibration by Baqaee, Farhi and Sangani (2024), but not when it is a decreasing function (as in Fréchet, Weibull, and Lognormal), with the Pareto distribution being the knife-edge case. This suggests that a rise of the markup and a decline in the pass-through rate may occur due to *more* competitive pressures through reallocation from less productive firms to more productive firms. Hence they should not be interpreted as the *prima-facie* evidence for reduced competitive pressures.
- The impact on the mass of active firms depends, often critically, on whether the elasticity of the distribution of the marginal cost is increasing or decreasing with Pareto-distributed productivity being the knife-edge case.
- In a multi-market setting, competitive pressures are stronger in larger markets. And more productive firms sort themselves into larger markets under the 2nd Law. Due to this *composition effect*, the weighted-generalized mean of the markup (pass-through) rates can be *higher* (*lower* under the 3rd Law) in larger (thus more competitive) markets. This result suggests a caution when interpreting the evidence that compares the average markup and pass-through rates across markets with different sizes.
- When the firms have access to the export market (subject to the additional overhead cost and the iceberg trade cost), which also means that they face competition from foreign firms in their home market, globalization caused by a reduction in the iceberg cost leads to a decline in the share of firms operating only in the home market and an increase in the share of exporting firms. Moreover, exporting firms reduce (raise) their markup rates in their home (export) market under the 2nd law and raise (reduce) their pass-through rates in their home (export) market under the 3rd law.

Here's the roadmap. In section 2, we formally introduce the H.S.A. class of demand systems and apply it to a monopolistic competitive sector with heterogenous firms and free entry. We show that both the markup and pass-through rates of firms with the marginal cost ψ can be expressed as $\mu(\psi/A)$ and $\rho(\psi/A)$, functions of a single variable, ψ/A, the firm's "normalized cost", where A is the inverse measure of competitive pressures, the equilibrium value of the single aggregator, which serves a sufficient statistic capturing all the interactions across firms. Hence, higher competitive pressures, a lower A, act as a magnifier of firm heterogeneity. We also show that the profit, the revenue, and the employment of a ψ-firm can all be expressed as functions of ψ/A, multiplied by market size of this sector, E. Then, we derive the equilibrium conditions in terms of A and the cutoff marginal cost, ψ_c and show that the equilibrium is uniquely determined (Figure 1).

In section 3, we consider the CES benchmark to revisit the Melitz model, which implies constant markup rate and complete pass-through. We offer a simpler proof of the existence of the unique equilibrium (Figure 2) and a reproduction of the well-known results; We also show that the sign of the elasticity of the marginal cost distribution determines comparative statics on the masses of the entrance and active firms, with Pareto-distributed firm productivity being the knife-edge case (Proposition 1).

Then, we depart from CES. In section 4, we consider the cross-sectional implications of more competitive pressures (a lower A) under the 2nd law, i.e., when $\mu(\psi/A)$ is strictly decreasing (Proposition 2), and under the weak or strong 3rd law, i.e., when $\rho(\psi/A)$ is weakly or strictly increasing (Propositions 3, 4, and 5). These results are summarized in Figure 3. In section 5, we conduct comparative statics to study the impacts of changes in the entry cost, market size, and the overhead cost on competitive pressures, A, and selection, ψ_c (Proposition 6; Figure 4). We look at the market size effect on the profit and the revenue (Proposition 7). Figure 5 in Appendix puts together these results. Then, we study how the average markup and pass-through rates, measured by the weighted generalized mean, change through the composition effect, as well as the impact on TFP (Proposition 8) and the effects on the mass of the active firms (Proposition 9). At the end of section 5, we look at the limit case of no overhead cost, where the cutoff firms operate at the choke price (Figure 6). In this case, the impact of an increase in market size is isomorphic to that of a decline in the entry cost.

Then, in section 6, we consider a multi-market extension, in which each firm, after learning its productivity, decides whether to stay or exit and, if it stays, chooses among markets with different sizes. We show that larger markets are more competitive and that, under the 2nd law, there is a positive assortative matching between firm productivity and market size (Proposition 10; Figure 7). Then, we show the cross-sectional implications across markets (Figure 8). Due to the composition effect, the average markup (pass-through) rate, measured by the weighted-generalized mean, may be higher (lower) in larger markets, and a shock that increases competitive pressures in all markets may lead to higher average markup rates and lower average pass-through rates in all markets in spite of the 2nd law and the 3rd law (Proposition 11). In Section 7, we study the effects of globalization, caused by a reduction in the iceberg trade cost between two symmetric markets (Proposition 12). Section 8 concludes. Appendices A through C contain some technical materials, including the proofs of some lemmas and propositions. Appendix D discuss three parametric families of H.S.A. and discuss their key properties. Appendix E shows some additional figures.

2. Selection of Heterogeneous Firms

2.1. The Environment

Consider a sector, where the single final good is produced competitively by assembling a set of differentiated intermediate inputs using CRS technology, represented by a linear homogenous, monotone, and quasi-concave, symmetric production function, $X = X(\mathbf{x})$. Here, $\mathbf{x} = \{x_\omega; \omega \in \Omega\}$ is a quantity vector of intermediate inputs, where Ω denotes a continuum set of intermediate input varieties available. Alternatively, the CRS technology can also be represented by a linear homogenous, monotone, and quasi-concave, symmetric unit cost function, $P = P(\mathbf{p})$, where $\mathbf{p} = \{p_\omega; \omega \in \Omega\}$ is a price vector of intermediate inputs. From the duality theorem, the production function, $X(\mathbf{x})$, and the unit cost function, $P(\mathbf{p})$, are related to each other as:

$$X(\mathbf{x}) \equiv \min_{\mathbf{p}} \left\{ \mathbf{px} \equiv \int_\Omega p_\omega x_\omega d\omega \,\middle|\, P(\mathbf{p}) \geq 1 \right\};\ P(\mathbf{p}) \equiv \min_{\mathbf{x}} \left\{ \mathbf{px} \equiv \int_\Omega p_\omega x_\omega d\omega \,\middle|\, X(\mathbf{x}) \geq 1 \right\}.$$

Hence, we could use either $P(\mathbf{p})$ or $X(\mathbf{x})$ as a primitive of the CRS technology. The solutions to the above minimization problems yield the demand curve and the inverse demand curve for ω:

$$x_\omega = X(\mathbf{x}) \frac{\partial P(\mathbf{p})}{\partial p_\omega};\ p_\omega = P(\mathbf{p}) \frac{\partial X(\mathbf{x})}{\partial x_\omega}.$$

From either of these, we obtain, by using the Euler's theorem of linear homogenous functions,

$$\mathbf{px} = P(\mathbf{p})X(\mathbf{x}) = PX \equiv E,$$

which says that the total value of the differentiated intermediate inputs used in this sector is equal to $PX \equiv E$, its market size. The market share of each variety, s_ω, can be expressed as

$$s_\omega \equiv \frac{p_\omega x_\omega}{\mathbf{px}} = \frac{p_\omega x_\omega}{P(\mathbf{p})X(\mathbf{x})} = \frac{\partial \ln P(\mathbf{p})}{\partial \ln p_\omega} = \frac{\partial \ln X(\mathbf{x})}{\partial \ln x_\omega}. \tag{1}$$

Each intermediate input variety $\omega \in \Omega$ is produced from the single primary factor of production, "labor," taken as numeraire. and sold exclusively by a single monopolistically competitive (MC) firm, also indexed by $\omega \in \Omega$. These MC firms are ex-ante homogenous before entering the market, but they become ex-post heterogenous in their marginal cost of production. More specifically, each firm pays F_e units of "labor" to enter the market, which is the sunk cost of entry. Upon entry, each firm draws its constant (quality-adjusted) marginal cost of production, ψ_ω, paid in "labor" from the common cdf, $G(\psi)$, with the density function, $g(\psi) = G'(\psi) > 0$ over the support, $\left(\underline{\psi}, \overline{\psi}\right) \subseteq (0, \infty)$.[11,12] Then, firm ω needs to hire $F + \psi_\omega x_\omega$ units of "labor" to produce x_ω units of its own product, where F is the overhead cost, the fixed cost of production, which is not sunk. Thus, upon discovering its marginal cost, ψ_ω, firm ω calculates its gross profit, $\Pi(\psi_\omega)$, and chooses to stay in the market if $\Pi(\psi_\omega) \geq F$ and to exit if $\Pi(\psi_\omega) < F$. Finally, there is free-entry to the market. Ex-ante homogenous firms enter until their expected gross profit is equal to the entry cost; $F_e = \int_{\underline{\psi}}^{\overline{\psi}} \max\{\Pi(\psi) - F, 0\} dG(\psi)$. This ensures no excess profit in equilibrium, so that the total demand for "labor" in this sector is equal to $L = \mathbf{px} = P(\mathbf{p})X(\mathbf{x}) = E$.[13]

2.2. Symmetric H.S.A. Demand System with Gross Substitutes

A symmetric CRS technology belongs to H.S.A. with gross substitutes if there is a function, $s: \mathbb{R}_{++} \to \mathbb{R}_+$, which is strictly decreasing for $s(z) > 0$ with $\lim_{z\to 0} s(z) = \infty$ and $\lim_{z\to \bar{z}} s(z) = 0$, where $\bar{z} \equiv \inf\{z > 0 | s(z) = 0\}$,[14] such that the market share of $\omega \in \Omega$, eq.(1), can be written as

$$s_\omega = \frac{p_\omega x_\omega}{P(\mathbf{p})X(\mathbf{x})} = \frac{\partial \ln P(\mathbf{p})}{\partial \ln p_\omega} = s\left(\frac{p_\omega}{A(\mathbf{p})}\right), \tag{2}$$

where $A(\mathbf{p})$ is defined implicitly by **the adding-up constraint:**

$$\int_\Omega s\left(\frac{p_\omega}{A(\mathbf{p})}\right) d\omega \equiv 1. \tag{3}$$

If $\bar{z} < \infty$, $s_\omega = 0$ for $p_\omega \geq \bar{z}A(\mathbf{p})$, hence $\bar{z}A(\mathbf{p})$ is **the choke price**. For $p_\omega < \bar{z}A(\mathbf{p})$, $s_\omega > 0$. The assumption that $s(\cdot)$ is strictly decreasing in $z < \bar{z}$ means that inputs are gross substitutes. Symmetric CES with gross substitutes is a special case of H.S.A, with $s(z) = \gamma z^{1-\sigma}$ ($\sigma > 1$). Symmetric translog is another special case, with $s(z) = \max\{-\gamma \ln(z/\bar{z}), 0\}$.[15] Appendix D offers more parametric examples of symmetric H.S.A.

---

[11] We assume $G(\cdot) \in C^3(\underline{\psi}, \overline{\psi})$. All the parametric cdfs in this paper satisfy $G(\cdot) \in C^\infty\left(\underline{\psi}, \overline{\psi}\right) \subset C^3\left(\underline{\psi}, \overline{\psi}\right)$.

[12] Equivalently, each entrant draws its (quality-adjusted) productivity, $\varphi = 1/\psi$, from its cdf, $F(\varphi) = 1 - G(1/\varphi)$, whose support is $\varphi \in \left(\underline{\varphi}, \overline{\varphi}\right) \subseteq (0, \infty)$, with $\underline{\varphi} = 1/\overline{\psi}$ and $\overline{\varphi} = 1/\underline{\psi}$. See Appendix A for more detail on the relations between the two cdfs, $F(\cdot)$ and $G(\cdot)$, and between their densities.

[13] Notice that no assumption is made on how this sector interacts with the rest of the economy, except that market size of this sector, E, determined out of this sector, leads to this sector's "labor" demand, $L = E$. One could, of course, assume that the economy has only one sector and the representative household, endowed with L units of "labor", consumes only the final good produced in this sector, so that its budget constraint leads to $L = E$. However, there is no need to make such an assumption in our sector-level analysis.

[14] We assume $s(\cdot) \in C^3(0, \bar{z})$. All the parametric examples in this paper satisfy $s(\cdot) \in C^\infty(0, \bar{z}) \subset C^3(0, \bar{z})$.

[15] For $s: \mathbb{R}_{++} \to \mathbb{R}_+$, satisfying the above conditions, a class of the market share functions, $s_\gamma(z) \equiv \gamma s(z)$ for $\gamma > 0$, generate the same demand system with the same common price aggregator. We just need to renormalize the indices of varieties, as $\omega' = \gamma\omega$, so that $\int_\Omega s_\gamma(p_\omega/A(\mathbf{p}))d\omega = \int_\Omega s(p_{\omega'}/A(\mathbf{p}))d\omega' = 1$. In this sense, $s_\gamma(z) \equiv \gamma s(z)$ for $\gamma > 0$ are all equivalent. Note also that a class of the market share functions, $s_\lambda(z) \equiv s(\lambda z)$ for $\lambda > 0$, generate the same demand system, with $A_\lambda(\mathbf{p}) = \lambda A(\mathbf{p})$, because $s_\lambda(p_\omega/A_\lambda(\mathbf{p})) = s(\lambda p_\omega/A_\lambda(\mathbf{p})) = s(p_\omega/A(\mathbf{p}))$. In this sense, $s_\lambda(z) \equiv s(\lambda z)$ for $\lambda > 0$ are all equivalent. Using these equivalences, for example, one could obtain the CES case with $s(z) = z^{1-\sigma}$ ($\sigma > 1$) by setting $\gamma = 1$ and the translog case, with $s(z) = \max\{-\ln(z), 0\}$ by setting $\gamma = 1$ and $\lambda = 1/\bar{z} = 1$, without loss of generality.

The price elasticity of demand for $\omega \in \Omega$ is,

$$\zeta_\omega \equiv -\frac{\partial \ln x_\omega}{\partial \ln p_\omega} = 1 - \frac{z_\omega s'(z_\omega)}{s(z_\omega)} \equiv 1 - \mathcal{E}_s(z_\omega) \equiv \zeta(z_\omega) \equiv \zeta\left(\frac{p_\omega}{A(\mathbf{p})}\right) > 1,$$

with $\lim_{z\to\bar{z}} \zeta(z) = \infty$, if $\bar{z} < \infty$, where $\mathcal{E}_f(x) \equiv d\ln f(x)/d\ln x = xf'(x)/f(x)$ is "the elasticity operator," defined for a differentiable function $f(x) > 0$ of a single variable $x > 0$.

The market share function, $s(\cdot)$, is the primitive of the H.S.A.[16] On the other hand, $A(\mathbf{p})$, cannot be a primitive, as it needs to be derived from $s(\cdot)$ using eq.(3).[17] Clearly, $A(\mathbf{p})$ *must* satisfy linear homogeneity in $\mathbf{p}$ for any fixed Ω, in order for the market share $s(z_\omega) = s(p_\omega/A(\mathbf{p}))$ to add up to one.

Note that, under H.S.A., both the market share of $\omega \in \Omega$, $s(z_\omega)$, and its price elasticity, $\zeta(z_\omega)$, depend solely on *its normalized price,* $z_\omega \equiv p_\omega/A(\mathbf{p})$, defined as its own price, p_ω, divided by the *common price aggregator,* $A(\mathbf{p})$.[18] It is "the average input price" against which the prices of *all* inputs are measured. In other words, $A(\mathbf{p})$ is a sufficient statistic, which captures all the cross-variety effect, allowing us to keep track of all the interactions in the demand system, which is the key feature of H.S.A.[19]

After deriving $A(\mathbf{p})$ from $s(\cdot)$, using eq.(3), the unit cost function, $P(\mathbf{p})$, behind this H.S.A. demand system can be obtained by integrating eq.(2), which yields

$$cP(\mathbf{p}) = A(\mathbf{p}) \exp\left[-\int_\Omega s\left(\frac{p_\omega}{A(\mathbf{p})}\right)\Phi\left(\frac{p_\omega}{A(\mathbf{p})}\right)d\omega\right], \text{where } \Phi(z) \equiv \frac{1}{s(z)}\int_z^{\bar{z}} \frac{s(\xi)}{\xi} \mathrm{d}\xi > 0 \quad (4)$$

where $c > 0$ is an integral constant, proportional to TFP [20] and $\Phi(z)$ is the productivity gain created by the product sold at the normalized price, z, and thus can be interpreted as the measure of love-for-variety (see Matsuyama and Ushchev 2023b). Clearly, $P(\mathbf{p})$ satisfies linear homogenous and monotonic. Moreover, Matsuyama and Ushchev (2017; Proposition 1-i)) showed that it is strictly quasi-concave, thereby proving the integrability of H.S.A. in the sense of Samuelson (1950) and Hurwicz & Uzawa (1971); that is, the existence of the underlying CRS technology, $X(\mathbf{x})$ or $P(\mathbf{p})$, that generates this H.S.A.

---

[16] Any price elasticity function, satisfying $\zeta(\cdot) > 1$, with $\lim_{z\to\bar{z}} \zeta(z) = \infty$, if $\bar{z} < \infty$, can also be used as the primitive, because $s(\cdot)$ could be derived from $\zeta(\cdot)$, as $s(z) = \gamma \exp\left[\int_z^{\bar{z}} [\zeta(\xi) - 1]\, d\xi/\xi\right]$.

[17] For $A(\mathbf{p})$ to be well-defined for all $\mathbf{p} = \{p_\omega; \omega \in \Omega\}$ for any Lebesgue measure of Ω in eq.(3), it is necessary to assume $\lim_{z\to 0} s(z) = \infty$. Though satisfied by CES and translog, this assumption would rule out some demand system we want to explore. Instead, we assume that E is not too small to ensure that there will be enough firms to enter in equilibrium so that $A(\mathbf{p})$ will be well-defined, as will be seen later.

[18] Under general homothetic symmetric demand systems, the market share of $\omega \in \Omega$, and its price elasticity depend on $\mathbf{p}/p_\omega$, the price distribution normalized by its own price, an infinite dimensional object.

[19] The assumption that $s(\cdot)$ is independent of ω is not a defining feature of H.S.A.; it is due to the symmetry of the underlying production function that generates this demand system.

[20] This integral constant, $c > 0$, cannot be pinned down. First, $A(\mathbf{p})$, the "average input price", depends on the unit of measurement of inputs, but not on the unit of measurement of the final good. In contrast, $P(\mathbf{p})$ is the cost of producing one unit of the final good, when the input prices are given by $\mathbf{p}$. Hence, it depends not only on the unit of measurement of inputs but also on that of the final good. Second, a change in TFP, though it affects $P(\mathbf{p})$, leaves the market share, and hence $A(\mathbf{p})$, unaffected.

demand system. Note that, with the sole exception of CES, $A(\mathbf{p})/P(\mathbf{p})$ is not constant and depends on $\mathbf{p}$. This can be verified by differentiating eq.(3) to obtain

$$\frac{\partial \ln A(\mathbf{p})}{\partial \ln p_\omega} = \frac{z_\omega s'(z_\omega)}{\int_\Omega s'(z_{\omega'}) z_{\omega'} d\omega'} = \frac{[\zeta(z_\omega) - 1] s(z_\omega)}{\int_\Omega [\zeta(z_{\omega'}) - 1] s(z_{\omega'}) d\omega'},$$

which differs from $\partial \ln P(\mathbf{p})/\partial \ln p_\omega = s(z_\omega)$, unless $\zeta(z)$ is independent of z; i.e., $\zeta(z) = \sigma > 1 \Leftrightarrow s(z) = \gamma z^{1-\sigma} \Leftrightarrow \Phi(z) = 1/(\sigma - 1)$.[21] This should not come as a surprise. *A priori,* there is no reason to think that $A(\mathbf{p})$ and $P(\mathbf{p})$ should move together. After all, $A(\mathbf{p})$ is the "average input price", the inverse measure of competitive pressures for each input, which captures the *cross-price effects* in the demand system, while $P(\mathbf{p})$ is the inverse measure of TFP, which captures the *productivity (or welfare) effects* of price changes. And eq.(4) shows that the ratio of the two, $A(\mathbf{p})/P(\mathbf{p})$, depends on the weighted sum of $\Phi(z_\omega)$, a measure of love-for-variety, which is not constant unless CES.

2.3. Behaviors and Performances of Monopolistically Competitive Firms

2.3.1. Markup and Pass-Through Rates

After drawing its marginal cost, ψ_ω, firm ω would set its price p_ω to maximize its operating profit, if it would stay, as follows:

$$\Pi_\omega = \max_{p_\omega} (p_\omega - \psi_\omega) x_\omega = \max_{\psi_\omega < p_\omega < \bar{z}A} \left(1 - \frac{\psi_\omega}{p_\omega}\right) s\left(\frac{p_\omega}{A}\right) E,$$

for its *normalized cost*, $\psi_\omega/A \in (0, \bar{z})$, by taking E and $A = A(\mathbf{p})$ as given.[22] Or equivalently, it chooses its normalized price, $z_\omega \equiv p_\omega/A < \bar{z}$, to solve

$$\Pi_\omega = \max_{\psi_\omega/A < z_\omega < \bar{z}} \left(1 - \frac{\psi_\omega/A}{z_\omega}\right) s(z_\omega) E \equiv \pi\left(\frac{\psi_\omega}{A}\right) E > 0.$$

Its FOC of this profit maximization yields

Lerner Formula:
$$z_\omega \left[1 - \frac{1}{\zeta(z_\omega)}\right] = \frac{\psi_\omega}{A},$$

with $\psi_\omega/A < z_\omega < \bar{z}$, where we recall $\zeta(z) \equiv 1 - \mathcal{E}_s(z) > 1$ is the price elasticity function, satisfying $\lim_{z \to \bar{z}} \zeta(z) = \infty$, if $\bar{z} < \infty$.

In what follows, we impose the following regularity condition for the ease of exposition:

A1: For all $z \in (0, \bar{z})$,

$$\mathcal{E}_{z(\zeta-1)/\zeta}(z) = 1 + \mathcal{E}_{1-1/\zeta}(z) = 1 + \frac{z\zeta'(z)}{[\zeta(z) - 1]\zeta(z)} = 1 - \frac{\mathcal{E}_\zeta(z)}{\mathcal{E}_s(z)} > 0.$$

[21]See Matsuyama and Ushchev (2020a; Corollary 2 of Lemma 2). This holds more generally, that is, for asymmetric H.S.A., as well as H.S.A. with gross complements, as shown in Matsuyama and Ushchev (2017; Proposition 1-iii).
[22]For $\bar{z} < \infty$, no firm that draws $\psi_\omega > \bar{z}A$ would stay.

This condition, **A1,** states that the marginal revenue, LHS of FOC, is strictly increasing in $z_\omega \equiv p_\omega/A$ (hence strictly decreasing in x_ω) along the demand curve. Thus, the Inverted Function Theorem implies that, under **A1,** the Lerner formula can be inverted as $p_\omega/A = z_\omega = Z(\psi_\omega/A)$, $Z'(\cdot) > 0$. Thus, all the firms that share the same ψ_ω set the same price. This allows us to reindex firms by their marginal cost, ψ, and to write their profit-maximizing **normalized price** as an increasing function of their **normalized marginal cost**, ψ/A, as

Normalized Price:
$$z_\psi \equiv \frac{p_\psi}{A} = Z\left(\frac{\psi}{A}\right), Z'(\cdot) > 0.$$

satisfying $\psi/A < Z(\psi/A) < \bar{z}$ and $\lim_{\psi/A\to\bar{z}} Z(\psi/A) = \bar{z}$. From this, the price elasticity at the point of the demand curve where ψ-firms choose to operate and their markup rate can both be written as function of $\psi/A \in (0, \bar{z})$:

Price Elasticity:
$$\zeta(z_\psi) = \zeta\left(Z\left(\frac{\psi}{A}\right)\right) \equiv \sigma\left(\frac{\psi}{A}\right) > 1,$$

Markup Rate:
$$\mu_\psi \equiv \frac{p_\psi}{\psi} = \frac{\zeta\big(Z(\psi/A)\big)}{\zeta\big(Z(\psi/A)\big) - 1} = \frac{\sigma(\psi/A)}{\sigma(\psi/A) - 1} \equiv \mu\left(\frac{\psi}{A}\right) > 1,$$

which are related to each other as:
$$\frac{1}{\sigma(\psi/A)} + \frac{1}{\mu(\psi/A)} = 1 \iff \left[\sigma\left(\frac{\psi}{A}\right) - 1\right]\left[\mu\left(\frac{\psi}{A}\right) - 1\right] = 1$$

and that their elasticities are related as:
$$\mathcal{E}_\sigma\left(\frac{\psi}{A}\right) = -\frac{\mathcal{E}_\mu(\psi/A)}{\mu(\psi/A) - 1} \Leftrightarrow \mathcal{E}_\mu\left(\frac{\psi}{A}\right) = -\frac{\mathcal{E}_\sigma(\psi/A)}{\sigma(\psi/A) - 1}.$$

By log-differentiating the Lerner formula, we obtain the pass-through rate also as a function of $\psi/A \in (0, \bar{z})$:

Pass-Through Rate:
$$\rho_\psi \equiv \frac{\partial \ln p_\psi}{\partial \ln \psi} = \mathcal{E}_Z\left(\frac{\psi}{A}\right) = \frac{1}{1 + \mathcal{E}_{1-1/\zeta}\big(Z(\psi/A)\big)} \equiv \rho\left(\frac{\psi}{A}\right) = 1 + \mathcal{E}_\mu\left(\frac{\psi}{A}\right) > 0,$$

where $\rho(\psi/A) > 0$ is ensured by **A1**.

Note that, although $Z(\psi/A)$ is always strictly increasing in ψ/A, $\mu(\psi/A)$ and $\rho(\psi/A)$ can be increasing, decreasing, or nonmonotonic in general. Note also that market size, E, does not enter directly in $\mu(\psi/A)$ and $\rho(\psi/A)$, which means that market size may affect the markup and pass-through rates only indirectly through its effect on $A = A(\mathbf{p})$. Moreover, $A = A(\mathbf{p})$ enters only as the divisor of ψ. Thus, more competitive pressures, a higher $1/A$, acts like a uniform decline in productivity across firms.

2.3.2. Profit, Revenue and Employment

The revenue of a ψ-firm is simply its market share multiplied by market size:

Revenue:
$$R_\psi \equiv s(z_\psi)E = s\left(Z\left(\frac{\psi}{A}\right)\right)E \equiv r\left(\frac{\psi}{A}\right)E,$$

From the Lerner formula, the firm level (gross) profit share and the (variable) labor share in the revenue are its inverse price elasticity and the inverse markup rate, respectively, so that:

(Gross) Profit:
$$\Pi_\psi = \frac{s\big(Z(\psi/A)\big)}{\zeta\big(Z(\psi/A)\big)}E = \frac{r(\psi/A)}{\sigma(\psi/A)}E \equiv \pi\left(\frac{\psi}{A}\right)E,$$

(Variable) Employment:
$$\psi x_\psi = \frac{r(\psi/A)}{\mu(\psi/A)}E \equiv \ell\left(\frac{\psi}{A}\right)E.$$

Thus, the revenue, the (gross) profit and the (variable) employment are all expressed as functions of a single variable, ψ/A, multiplied by market size, E.[23] Furthermore, they vary with ψ/A as:

$$\frac{\partial \ln R_\psi}{\partial \ln \psi} = \frac{\partial \ln R_\psi}{\partial \ln(1/A)} = \mathcal{E}_r\left(\frac{\psi}{A}\right) = \mathcal{E}_s\left(Z\left(\frac{\psi}{A}\right)\right)\mathcal{E}_Z\left(\frac{\psi}{A}\right) = \left[1-\sigma\left(\frac{\psi}{A}\right)\right]\rho\left(\frac{\psi}{A}\right) < 0;$$
$$\frac{\partial \ln \Pi_\psi}{\partial \ln \psi} = \frac{\partial \ln \Pi_\psi}{\partial \ln(1/A)} = \mathcal{E}_\pi\left(\frac{\psi}{A}\right) = \mathcal{E}_r\left(\frac{\psi}{A}\right) - \mathcal{E}_\sigma\left(\frac{\psi}{A}\right) = 1-\sigma\left(\frac{\psi}{A}\right) < 0;$$
$$\frac{\partial \ln(\psi x_\psi)}{\partial \ln \psi} = \frac{\partial \ln(\psi x_\psi)}{\partial \ln(1/A)} = \mathcal{E}_\ell\left(\frac{\psi}{A}\right) = \mathcal{E}_r\left(\frac{\psi}{A}\right) - \mathcal{E}_\mu\left(\frac{\psi}{A}\right) = 1-\rho\left(\frac{\psi}{A}\right)\sigma\left(\frac{\psi}{A}\right) \gtreqless 0.$$

all of which are independent of market size E, and depend solely on ψ/A, through $\sigma(\cdot)$ and $\rho(\cdot)$. This means that, for non-CES H.S.A., market size E affects the relative firm size in revenue, gross profit, and employment only through its effects on $A = A(\mathbf{p})$. Since $\sigma(\cdot) > 1, \mathcal{E}_r(\cdot) < 0$ and $\mathcal{E}_\pi(\cdot) < 0$. Thus, both the revenue and the profit are always strictly decreasing in ψ/A. In contrast, $\mathcal{E}_\ell(\cdot)$ can change its sign, and hence the employment can be nonmonotonic in ψ/A.

2.4 Equilibrium Conditions:

Let us assume $F + F_e < \pi(0)E$. This ensures that a positive measure of firms always enter, because otherwise $A = A(\mathbf{p}) \to \infty$, and firms could earn enough gross profit to cover both the entry cost and the overhead cost, regardless of their marginal costs. Then, for a given market size, E, an equilibrium in this sector is characterized by the following three conditions:

Cutoff Rule: Firms choose to stay and produce if $\psi \in \left(\underline{\psi}, \psi_c\right)$ and exit without producing if $\psi \in (\psi_c, \overline{\psi})$, where ψ_c is the cutoff level of the marginal cost, determined by:

$$\pi\left(\frac{\psi_c}{A}\right)E = F \Leftrightarrow \frac{\psi_c}{A} = \pi^{-1}\left(\frac{F}{E}\right) < \bar{z}, \tag{5}$$

[23]This is one of the major advantages of using H.S.A. If we had used HDIA or HIIA instead, two aggregators would be needed to express the revenue, profit, and employment of each firm.

assuming the interior solution, $0 < G(\psi_c) < 1$. In Figure 1, the cutoff rule, eq.(5), is depicted as the ray whose slope, $\pi^{-1}(F/E)$, is decreasing in F/E. A smaller market size/overhead cost ratio thus causes a tougher selection, a smaller ψ_c, causing more firms to exit for a given A.

Free-Entry Condition: Expected gross profit is equal to the entry cost,

$$F_e = \int_{\underline{\psi}}^{\psi_c} \left[\pi\left(\frac{\psi}{A}\right)E - F\right] dG(\psi). \tag{6}$$

Figure 1 depicts the free-entry condition, eq.(6), as the C-shape curve, downward-sloping below the cutoff rule, upward-sloping above the cutoff, and vertical at the intersection.[24] A lower entry cost shifts the curve to the left, which causes more competitive pressures, a lower A.

Clearly, these two conditions jointly determine the equilibrium values of $A = A(\mathbf{p})$ and ψ_c uniquely as functions of F_e/E and F/E. The interior solution, $0 < G(\psi_c) < 1$, is ensured under:

$$0 < \frac{F_e}{E} = \int_{\underline{\psi}}^{\psi_c} \left[\pi\left(\pi^{-1}\left(\frac{F}{E}\right)\frac{\psi}{\psi_c}\right) - \frac{F}{E}\right] dG(\psi) < \int_{\underline{\psi}}^{\bar{\psi}} \left[\pi\left(\pi^{-1}\left(\frac{F}{E}\right)\frac{\psi}{\bar{\psi}}\right) - \frac{F}{E}\right] dG(\psi),$$

which is assumed to hold throughout the paper.[25] Note that this condition holds for a sufficiently small $F_e > 0$ with no further restrictions on $G(\cdot)$ or $s(\cdot)$.

Having $A = A(\mathbf{p})$ and ψ_c pinned down uniquely by eqs.(5)-(6), we can obtain the mass of the entrants, M, that pay the entry cost F_e, from[26]

Adding-up (Resource) Constraint: By rewriting eq.(3) as:

$$1 \equiv \int_{\Omega} s\left(\frac{p_\omega}{A}\right) d\omega = M \int_{\underline{\psi}}^{\psi_c} s\left(Z\left(\frac{\psi}{A}\right)\right) dG(\psi) = M \int_{\underline{\psi}}^{\psi_c} r\left(\frac{\psi}{A}\right) dG(\psi),$$

$$M = \left[\int_{\underline{\psi}}^{\psi_c} r\left(\frac{\psi}{A}\right) dG(\psi)\right]^{-1} = \left[\int_{\underline{\xi}}^{1} r\left(\pi^{-1}\left(\frac{F}{L}\right)\xi\right) dG(\psi_c\xi)\right]^{-1}. \tag{7}$$

Eqs.(5)-(7) fully determine the equilibrium.[27] For the equilibrium value of $MG(\psi_c)$, the mass of active firms, equal to the Lebesgue measure of Ω, we can use eq.(7) to obtain

---

[24]As $A \to \infty$, the free entry condition curve is asymptotic to the horizontal line defined by $G(\psi_c) = F_e/[\pi(0)L - F]$, which is bounded away from the lower bound, $\psi_c = \underline{\psi}$, if and only if $\pi(0) < \infty$.

[25]For $\bar{\psi} = \infty$, this condition is reduced to $\pi(0)E > F_e + F > F_e > 0$, which is already assumed. For $\bar{\psi} < \infty$, the upper bound on F_e is less than $\pi(0)E - F$, and simple algebra can show that this upper bound is independent of E under CES, while increasing in E under **A2** introduced later.

[26]What makes H.S.A. particularly tractable is this recursive structure. Under HDIA and HIIA, the two other classes of the demand system studied in Matsuyama and Ushchev (2020a), the market share of each firm depends on the two aggregators, one affecting the pricing decision of the firm and the other its entry decision. As a result, the free-entry equilibrium is determined jointly by the three conditions. This complicates not only comparative statics, but also requires further assumptions on the firm distribution and the demand system to ensure the existence and the uniqueness of the equilibrium.

[27]Of course, for these equilibrium conditions to be well-defined, the integrals in eq.(6) and eq.(7) must be finite, which is clearly the case if $\underline{\psi} > 0$. For $\underline{\psi} = 0$, Lemma 4 of Appendix B shows that $1 \le \lim_{z\to 0} \zeta(z) < 2 + \lim_{\psi \to 0} \mathcal{E}_g(\psi) < \infty$ is a sufficient condition.

$$MG(\psi_c) = \left[\int_{\underline{\psi}}^{\psi_c} r\left(\frac{\psi}{A}\right)\frac{dG(\psi)}{G(\psi_c)}\right]^{-1} = \left[\int_{\underline{\xi}}^{1} r\left(\pi^{-1}\left(\frac{F}{E}\right)\xi\right) d\tilde{G}(\xi;\psi_c)\right]^{-1}, \tag{8}$$

where the second equality is obtained by changing variables as $\xi \equiv \psi/\psi_c$ with $\underline{\xi} \equiv \underline{\psi}/\psi_c$, and

$$\tilde{G}(\xi;\psi_c) \equiv \frac{G(\psi_c\xi)}{G(\psi_c)}$$

is the cdf of the marginal cost relative to the cutoff marginal cost among the firms that stay. Lemma 2 of Appendix A shows that a lower ψ_c (tougher selection) shifts $\tilde{G}(\xi;\psi_c)$ to the right (left) in the MLR ordering if $\mathcal{E}_g'(\psi) < (>)0$, and to the right (left) in the FSD ordering if $\mathcal{E}_G'(\psi) < (>)0$.[28] Pareto distributed productivity, $G(\psi) = (\psi/\bar{\psi})^{\kappa}$, is the knife-edge case, where $\tilde{G}(\xi;\psi_c)$ is independent of ψ_c, because $\mathcal{E}_g'(\psi) = \mathcal{E}_G'(\psi) = 0$.

Another feature of the equilibrium is worth noting. From the equilibrium conditions, it is easy to verify that an industry-wide productivity shock of the form, $G(\psi) \to G(\psi/\lambda)$, causes the cutoff and competitive pressures to shift as $\psi_c \to \lambda\psi_c$ and $A \to \lambda A$, keeping ψ_c/A unchanged. Thus, the distribution of ψ/A across active firms remains unchanged, and hence the distributions of the normalized prices, of the markup and pass-through rates, and of the revenues, the profits, and the employments, as well as the masses of entrants and active firms, M and $MG(\psi_c)$, all remain unchanged. The distribution of the (unnormalized) prices shifts to the right and that of the quantities shifts to the left by the factor λ, and $P \to \lambda P$.

2.5. Aggregate Labor Cost and Profit Shares and TFP

For any two functions of ψ/A, $w(\cdot)$ and $f(\cdot)$, we denote the $w(\cdot)$-weighted average of $f(\cdot)$ among the active firms, $\psi \in (\underline{\psi}, \psi_c)$, by

$$\mathbb{E}_w(f) \equiv \frac{\int_{\underline{\psi}}^{\psi_c} f(\psi/A)w(\psi/A)dG(\psi)}{\int_{\underline{\psi}}^{\psi_c} w(\psi/A)dG(\psi)} = \frac{\int_{\underline{\psi}}^{\psi_c} f(\psi/A)w(\psi/A)\frac{dG(\psi)}{G(\psi_c)}}{\int_{\underline{\psi}}^{\psi_c} w(\psi/A)\frac{dG(\psi)}{G(\psi_c)}}.$$

Likewise, we denote the unweighted average of $f(\cdot)$ among the active firms, $\psi \in (\underline{\psi}, \psi_c)$ by

$$\mathbb{E}_1(f) \equiv \frac{\int_{\underline{\psi}}^{\psi_c} f(\psi/A)dG(\psi)}{\int_{\underline{\psi}}^{\psi_c} dG(\psi)} = \int_{\underline{\psi}}^{\psi_c} f\left(\frac{\psi}{A}\right)\frac{dG(\psi)}{G(\psi_c)}.$$

From these definitions, one can immediately derive the following identity:

[28] Lemma 1 of Appendix A shows that $\mathcal{E}_g'(\psi) < 0$ always implies $\mathcal{E}_G'(\psi) < 0$, while $\mathcal{E}_g'(\psi) \geq 0$ implies $\mathcal{E}_G'(\psi) \geq 0$ only with some boundary conditions. In Generalized Pareto (Example 2 of Appendix A), $\mathcal{E}_g'(\psi) \gtreqless 0$, depending on the parameters. Lognormal (Example 3) and Fréchet/Weibull (Example 4) satisfy $\mathcal{E}_g'(\psi) < 0$ hence $\mathcal{E}_G'(\psi) < 0$.

$$\mathbb{E}_w\left(\frac{f}{w}\right)=\frac{\mathbb{E}_1(f)}{\mathbb{E}_1(w)}=\left[\frac{\mathbb{E}_1(w)}{\mathbb{E}_1(f)}\right]^{-1}=\left[\mathbb{E}_f\left(\frac{w}{f}\right)\right]^{-1}.$$

By applying this identity to $\pi(\cdot)/r(\cdot)=1-\ell(\cdot)/r(\cdot)=1/\sigma(\cdot)=1-1/\mu(\cdot)$, the sector-level labor cost share can be expressed as:

$$\frac{\mathbb{E}_1(\ell)}{\mathbb{E}_1(r)}=\mathbb{E}_r\left(\frac{1}{\mu}\right)=1-\left[\mathbb{E}_\pi\left(\frac{\mu}{\mu-1}\right)\right]^{-1}=\frac{1}{\mathbb{E}_\ell(\mu)}.$$

Since the firm-level labor cost share is equal to its inverse markup rate, this expression shows that the average labor cost share in the sector should be measured by the revenue-weighted arithmetic mean of firm-level labor cost shares, or by the employment-weighted harmonic mean.[29] Likewise, the sector-level profit share can be expressed as:

$$\frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(r)}=\mathbb{E}_r\left(\frac{1}{\sigma}\right)=\frac{1}{\mathbb{E}_\pi(\sigma)}=1-\left[\mathbb{E}_\ell\left(\frac{\sigma}{\sigma-1}\right)\right]^{-1}.$$

Since the firm-level profit share is equal to its inverse price elasticity, this expression shows that the average profit share in the sector should be measured by the revenue-weighted arithmetic mean of firm-level profit shares or the profit-weighted harmonic mean.

For TFP, $X/L=X(\mathbf{x})/L=1/P(\mathbf{p})$, which is equal to the aggregate consumption per unit of labor, and the welfare measure, can be obtained from eq.(4) and eq.(7) as

$$\frac{X}{L}=\frac{1}{P}=\frac{c}{A}\exp\left[\mathbb{E}_r[\Phi\circ Z]\right].$$

3. CES Benchmark

First, consider the case of CES, a special case of H.S.A., $\zeta(z)=\sigma>1\Leftrightarrow s(z)=\gamma z^{1-\sigma}$ for all $z\in(0,\infty)$. This corresponds to the Melitz (2003) model. Though the Melitz model is well-known, it is instructive to obtain its properties as a special case of ours, because his analysis and its countless reproduction by others -- see a survey by Melitz and Redding (2014)--make heavy use of CES from the very beginning. This makes it hard to see which properties of the Melitz model are specific to CES or which ones can be generalized under H.S.A.

The markup rate is $\mu(\psi/A)=\sigma/(\sigma-1)$, and the pass-through rate is $\rho(\psi/A)=1$. Hence, they are both uniform across all firms, unaffected by $E, F_e, F, G(\cdot), A, \psi_c$, and thus never change across equilibriums. The profit is $\pi(\psi/A)E=c_0E(\psi/A)^{1-\sigma}$, where $c_0\equiv(\gamma/\sigma)(1-1/\sigma)^{\sigma-1}$. Thus, the cutoff rule, eq.(5), and free entry condition, eq.(6), become:

---

[29]This also suggests that the average markup rate should be measured by the revenue-weighted harmonic mean of firm-level markup rate or by the employment-weighted arithmetic mean, as pointed out by Baqaee, Farhi, and Sangani (2024) and Edmond, Midrigan, and Xu (2023).

Cutoff Rule:

$$c_0 E\left(\frac{\psi_c}{A}\right)^{1-\sigma} = F.$$

Free Entry Condition:

$$\int_{\underline{\psi}}^{\psi_c}\left[c_0 E\left(\frac{\psi}{A}\right)^{1-\sigma} - F\right] dG(\psi) = F_e;$$

As shown in Figure 2, the cutoff rule and the free-entry condition have the unique intersection.[30] An increase in E shifts the cutoff rule counter-clockwise, and the free-entry condition to the left, from the dashed curves to the solid ones. To see how the intersection moves, eliminate E from these two conditions to obtain

$$\int_{\underline{\psi}}^{\psi_c}\left(\left(\frac{\psi}{\psi_c}\right)^{1-\sigma} - 1\right) dG(\psi) = \frac{F_e}{F}. \tag{9}$$

As E increases, the intersection moves to the left along the locus given by eq.(9), which is independent of A, as depicted by the horizontal dotted line in Figure 2.[31] The equilibrium cutoff, ψ_c, is thus independent of E. Eq.(9) also shows that the equilibrium cutoff, ψ_c, declines in response to a lower F_e/F and to an improvement in productivity distribution, captured by a first-order stochastic dominant (FSD) shift of $\psi \sim G(\cdot)$ *to the left*. Furthermore, A can be expressed as

$$A = \psi_c\left(\frac{c_0 E}{F}\right)^{\frac{1}{1-\sigma}} = \left(\frac{c_0 E}{F_e}\int_{\underline{\psi}}^{\psi_c}[(\psi)^{1-\sigma} - (\psi_c)^{1-\sigma}]dG(\psi)\right)^{\frac{1}{1-\sigma}}.$$

Thus, a higher E, a lower F_e, a lower F, and a FSD shift of $\psi \sim G(\cdot)$ to the left all lead to more competitive pressures, a lower A. Since A/P is constant under CES, the effect on P is the same, and the effect on TFP, $X/L = 1/P$, goes the opposite direction.

The revenue, the (gross) profit and the (variable) employment of a ψ-firm are:

Revenue:

$$r\left(\frac{\psi}{A}\right)E = \sigma c_0 E\left(\frac{\psi}{A}\right)^{1-\sigma} = \sigma F\left(\frac{\psi}{\psi_c}\right)^{1-\sigma} \geq \sigma F$$

Profit:

$$\pi\left(\frac{\psi}{A}\right)E = c_0 E\left(\frac{\psi}{A}\right)^{1-\sigma} = F\left(\frac{\psi}{\psi_c}\right)^{1-\sigma} \geq F$$

Employment:

$$\ell\left(\frac{\psi}{A}\right)E = (\sigma-1)c_0 E\left(\frac{\psi}{A}\right)^{1-\sigma} = (\sigma-1)F\left(\frac{\psi}{\psi_c}\right)^{1-\sigma} \geq (\sigma-1)F$$

which are all decreasing power functions in ψ with the exponent, $1-\sigma < 0$. Thus, their ratios across two firms with $\psi, \psi' \in (\underline{\psi}, \psi_c)$, given by $(\psi/\psi')^{1-\sigma} > 1$, are independent of E, F_e, F and $G(\cdot)$, as well as A and ψ_c. Hence, the relative size of two firms, whether measured in the revenue, profit, or variable employment, never changes across different equilibriums.

[30] This proof of the existence and uniqueness of the equilibrium is simpler than Melitz (2003; Appendix B).
[31] The expression analogous to eq.(9) has been known; see, e.g., eq.(13) of Bernard, Redding and Schott (2007).

From the free entry condition and the adding-up constraint, $M[F_e + G(\psi_c)F] = E/\sigma$, which states that the aggregate entry cost plus the aggregate expected fixed cost is equal to the aggregate profit. Using eq.(9), this can be further rewritten to obtain:

$$M = \frac{E/\sigma}{F_e + G(\psi_c)F} = \frac{E}{\sigma F_e}\left[1 - \frac{1}{H(\psi_c)}\right];\ MG(\psi_c) = \frac{E/\sigma}{F_e/G(\psi_c) + F} = \frac{E}{H(\psi_c)\sigma F},$$

where $H(\psi_c) \equiv \int_{\underline{\xi}}^{1} (\xi)^{1-\sigma} d\tilde{G}(\xi;\psi_c)$. Since $(\xi)^{1-\sigma}$ is decreasing, Lemma 2 implies

$$\mathcal{E}_G'(\cdot) \gtreqless 0 \Rightarrow H'(\psi_c) \lesseqgtr 0,$$

from which it is straightforward to verify the following:

Proposition 1: Under CES,

1a: A higher E keeps ψ_c unaffected and increases both M and $MG(\psi_c)$ *proportionately;*

1b: A lower F_e decreases ψ_c and increases M; It increases $MG(\psi_c)$ if $\mathcal{E}_G'(\psi) < 0$, decreases $MG(\psi_c)$ if $\mathcal{E}_G'(\psi) > 0$ and keeps $MG(\psi_c)$ unaffected if $\mathcal{E}_G'(\psi) = 0$;

1c: A lower F increases ψ_c and increases $MG(\psi_c)$; It increases M if $\mathcal{E}_G'(\psi) < 0$, decreases M if $\mathcal{E}_G'(\psi) > 0$ and keeps M unaffected if $\mathcal{E}_G'(\psi) = 0$.

Although most of these results are known, the result that the sign of $d[MG(\psi_c)]/dF_e$ and the sign of dM/dF are the same with the sign of $\mathcal{E}_G'(\psi)$ seems new.[32] A FSD shift of $G(\cdot)$ to the left reduces ψ_c. However, its effects on M and $MG(\psi_c)$ are ambiguous in general.[33]

4. Heterogenous Firms under H.S.A.: Cross-Sectional Implications

We now depart from CES. Even though the 2nd and the 3rd laws may not be the universal laws, satisfied in every single sector in every single country, there seems to be ample evidence in their support, as cited in the introduction, so that we will primarily focus on the implications of the 2nd and the 3rd laws. In this section, we explore how the impacts of more competitive pressures (a lower A) vary across heterogeneous firms, first under the 2nd law and then under the 3rd law. Of course, A is an endogenous variable, whose change must be triggered by a change in E, F_e, and/or F. Nevertheless, we postpone such comparative statics analysis to the next section.

---

[32]We inquired Melitz about this, to which he replied that he had not seen these results. Appendix A shows that, $\mathcal{E}_g'(\cdot) < 0$ and $\mathcal{E}_G'(\cdot) < 0$ for Fréchet, Weibull, and Lognormal, which suggests, among others, that the results obtained by some studies on the Melitz model under Lognormal, e.g., Head, Mayer, and Theonig (2014), are qualitatively robust to any distribution with $\mathcal{E}_g'(\cdot) < 0$.

[33]To see this, consider the case of power-distributed marginal cost (i.e., Pareto-distributed productivity), $G(\psi) = \left(\psi/\overline{\psi}\right)^{\kappa}, 0 < \psi < \overline{\psi}, \kappa > \sigma - 1$, so that $\mathcal{E}_G'(\cdot) = 0$ and $\tilde{G}(\xi;\psi_c) = \xi^{\kappa}$, and $H(\psi_c) = \int_0^1 \kappa(\xi)^{\kappa-\sigma} d\xi = \frac{\kappa}{\kappa-\sigma+1} > 1$ is independent of ψ_c. Under the condition that ensures the interior solution, $G(\psi_c) = \frac{\kappa-\sigma+1}{\sigma-1}\left(\frac{F_e}{F}\right) < 1$, we have $M = \frac{\sigma-1}{\kappa}\left(\frac{L}{\sigma F_e}\right) > MG(\psi_c) = \frac{\kappa-\sigma+1}{\kappa}\left(\frac{L}{\sigma F}\right)$. Thus, a FSD shift in G, due to a change in $\overline{\psi}$, has no effect on $G(\psi_c), M$ nor $MG(\psi_c)$, while a FSD shift in G, due to a change in κ, affects $G(\psi_c), M$ and $MG(\psi_c)$.

4.1. Cross-Sectional Implications of the 2nd Law of Demand

A2: $\zeta'(z) > 0$ for all $z \in (0, \bar{z}) \Leftrightarrow \sigma'(\psi/A) = \zeta'\big(Z(\psi/A)\big)Z'(\psi/A) > 0$ for all $\psi/A \in (0, \bar{z})$

Under **A2,** $\mathcal{E}_\zeta(z) > 0 > \mathcal{E}_s(z)$ for all $z \in (0, \bar{z})$. Hence, **A2** implies **A1.** This assumption means that the price elasticity of demand, $\zeta(p_\psi/A)$, is strictly increasing in its price, p_ψ for a fixed A, which each firm takes as given. It is thus equivalent to Marshall's 2nd Law of demand. Under **A2**, $\zeta\big(Z(\psi/A)\big) \equiv \sigma(\psi/A)$ is strictly increasing in ψ/A and $\mu(\psi/A)$ is strictly decreasing in ψ/A. Hence,

$$\rho\left(\frac{\psi}{A}\right) \equiv \frac{\partial \ln p_\psi}{\partial \ln \psi} = \frac{\partial \ln(Z(\psi/A)A)}{\partial \ln \psi} = \mathcal{E}_Z\left(\frac{\psi}{A}\right) = 1 + \mathcal{E}_\mu\left(\frac{\psi}{A}\right) < 1,$$

so that less productive firms have lower markup rates and that the price responds less than proportionately to a change in the marginal cost (**Incomplete Pass-Through).** Furthermore,

$$\frac{\partial \ln p_\psi}{\partial \ln(1/A)} = \frac{\partial \ln(Z(\psi/A)A)}{\partial \ln(1/A)} = \frac{d \ln\big(Z(\psi/A)\big)}{d \ln(\psi/A)} - 1 = \rho\left(\frac{\psi}{A}\right) - 1 = \mathcal{E}_\mu\left(\frac{\psi}{A}\right) < 0.$$

Thus, the firm reduces its price (and its markup rate) in response to more competitive pressures, a higher $1/A$, which occurs either when other firms reduce their prices (**Strategic complementarity in pricing**) or when more firms enter **(Procompetitive entry**).[34]

For further exploration, let us reformulate the definitions of log-super(sub)modularity specifically for our context. A positive-valued function of a single variable, $\psi/A > 0$, $f(\psi/A) > 0$, when viewed as a function of the two variables, ψ and $1/A$, is strictly *log-super(sub)modular* in ψ and $1/A$ if $\partial^2 \ln f(\psi/A)/\partial\psi\partial(1/A) > (<)0$. Or, we sometimes say, more simply, that $f(\psi/A)$ is strictly *log-super(sub)modular,* when this condition holds. The log-super(sub)modularity of a decreasing function $f(\psi/A)$ thus means that more competitive pressures, a lower A, causes a disproportionately larger (smaller) decline in $f(\psi/A)$ for a higher ψ. The next lemma offers a simple way of verifying the log-super(sub)modularity of $f(\psi/A)$.

Lemma 5: For any positive-valued C^2-function f of a single variable, $\psi/A > 0$,

$$sgn\left\{\frac{\partial^2 \ln f(\psi/A)}{\partial\psi\partial(1/A)}\right\} = sgn\left\{\mathcal{E}_f'\left(\frac{\psi}{A}\right)\right\} = sgn\left\{\frac{d^2 \ln f\big(e^{\ln(\psi/A)}\big)}{(d \ln(\psi/A))^2}\right\}.$$

The proof is straightforward and hence omitted. This lemma, which is known,[35] states that $f(\psi/A)$ is strictly *log-super(sub)modular* in ψ and $1/A$ if and only if $\mathcal{E}_f(\cdot)$ is strictly increasing(decreasing), that is,

---

[34]As pointed out in Matsuyama and Ushchev (2020a), the 2nd law of demand (or incomplete pass-through) is in general neither sufficient nor necessary for procompetitive entry (or strategic complementarity in price), since the former is about how the price elasticity of demand for a firm's product responds to its own price, while the latter is about how it responds to the behaviors of other prices. Under H.S.A., the 2nd law of demand (or incomplete pass-through) is equivalent to procompetitive entry (or strategic complementarity in price), because the aggregator A, which captures all the interaction across firms, enters in the price elasticity function only as ψ/A, so that a change in A is isomorphic to a change in ψ.

[35] See, e.g., Sampson (2016; Lemma 1) and Davis and Dingel (2020; Lemma 8).

if and only if $\ln f(e^x) = \ln f(\psi/A)$ is strictly convex (concave) in $x \equiv \ln(\psi/A)$. Since $\mathcal{E}_\pi(\psi/A) = 1 - \sigma(\psi/A) < 0$ is strictly decreasing in ψ/A under **A2**, Lemma 5 tells us that the profit, $\pi(\psi/A)E$, is strictly log-submodular in ψ and $1/A$.

The next proposition summarizes these implications of **A2,**

Proposition 2 (Cross-Sectional Implications of 2nd Law): Under **A2,**

2a **(Incomplete pass-through)**:

$$0 < \frac{\partial \ln p_\psi}{\partial \ln \psi} = \rho\left(\frac{\psi}{A}\right) = 1 + \mathcal{E}_\mu\left(\frac{\psi}{A}\right) < 1.$$

2b **(Procompetitive effect/strategic complementarity):**

$$\frac{\partial \ln p_\psi}{\partial \ln(1/A)} = \rho\left(\frac{\psi}{A}\right) - 1 = \mathcal{E}_\mu\left(\frac{\psi}{A}\right) < 0.$$

2c **(Strictly log-submodular profit)**:

$$\mathcal{E}_\pi'\left(\frac{\psi}{A}\right) = -\sigma'\left(\frac{\psi}{A}\right) < 0 \Leftrightarrow \frac{\partial^2 \ln \pi(\psi/A)E}{\partial\psi\partial(1/A)} < 0.$$

Because $\pi(\psi/A)$ is strictly log-submodular, more competitive pressures, a higher $1/A$, causes a proportionately larger decline in the profit among higher-ψ firms. Because higher-ψ firms have lower profits, this implies that more competitive pressures lead to a larger dispersion of profits across firms with the profit density shifting toward lower-ψ firms. Figure 3a illustrates this by plotting the graphs of log-profit, $\ln \Pi_\psi$, as a function of log-marginal cost, $\ln \psi$. The graph is always downward-sloping. And it is strictly concave under A2. For a fixed E, a higher $1/A$, causes a parallel leftward shift of the graph. Due to the concavity, this means a larger downward shift for higher-ψ firms, that is, a proportionately larger decline in their profit.[36]

4.2. Cross-Sectional Implications of the 3rd Law of Demand

A2 implies that the markup rate function, $\sigma(\cdot)$, is strictly increasing but not the pass-through rate function, $\rho(\cdot)$, can be increasing, decreasing or nonmonotonic. Thus, **A2** alone does not imply the monotonicity of $\mathcal{E}_Z(\cdot) = \rho(\cdot)$; $\mathcal{E}_r(\cdot) = [1 - \sigma(\cdot)]\rho(\cdot)$; and $\mathcal{E}_\ell(\cdot) = 1 - \rho(\cdot)\sigma(\cdot)$. In other words, **A2** alone ensures neither log-supermodularity nor log-submodularity of $Z(\psi/A), r(\psi/A)E$ or $\ell(\psi/A)E$. Partially motivated by the evidence cited in the introduction, we now consider the following assumption.

A3: For all $z \in (0, \bar{z})$,

$$\mathcal{E}_{\zeta/(\zeta-1)}'(z) = -\frac{d}{dz}\left(\frac{z\zeta'(z)}{[\zeta(z)-1]\zeta(z)}\right) \geq 0 \Leftrightarrow \rho'\left(\frac{\psi}{A}\right) \geq 0$$

[36]Figure 3a also depicts the effect of a higher E for a fixed A as a parallel upward shift of the graph. In Proposition 6, it will be shown that a higher E always leads to a lower A. Thus, if A declines due to a higher E, the total impact of a higher E on the profit is captured by a combination of the parallel upward shift (the positive direct effect) and the parallel leftward shift (the indirect effect due to a lower A). We will look at this in more detail in Proposition 7a.

A3 means that the pass-through rate is weakly increasing in ψ, which we shall call **the 3rd Law of demand.** In particular, we call it the weak 3rd Law of demand or simply **the weak A3** when the inequality in **A3** holds weakly, and the strong 3rd Law of demand or simply **the strong A3**, when the inequality in **A3** holds strictly and hence the pass-through rate is strictly increasing in ψ. Of the three parametric families of H.S.A. discussed in Appendix D, Generalized Translog satisfies A2 but violates even the weak A3; Constant Pass-Through (CoPaTh) satisfies A2 and the weak A3, but violates the strong A3; and Power Elasticity of Markup Rates (PEM) satisfies both A2 and the strong A3.

Then, using Lemma 5, we have the following proposition:

Proposition 3 (Cross-Sectional Implications of 3rd Law):

3a (Weak (strict) log-supermodular price and markup rate): Under **the weak (strong) A3,**

$$\mathcal{E}_Z'\left(\frac{\psi}{A}\right) = \rho'\left(\frac{\psi}{A}\right) \geq (>) 0 \Leftrightarrow \frac{\partial^2 \ln(Z(\psi/A)A)}{\partial\psi\partial(1/A)} = \frac{\partial^2 \ln\mu(\psi/A)}{\partial\psi\partial(1/A)} \geq (>) 0,$$

3b (Strict log-submodular revenue): Under **A2** and **the weak A3,**

$$\mathcal{E}_r'\left(\frac{\psi}{A}\right) = \left[1 - \sigma\left(\frac{\psi}{A}\right)\right]\rho'\left(\frac{\psi}{A}\right) - \sigma'\left(\frac{\psi}{A}\right)\rho\left(\frac{\psi}{A}\right) < 0 \Leftrightarrow \frac{\partial^2 \ln r(\psi/A)}{\partial\psi\partial(1/A)} < 0$$

3c (Strict log-submodular employment): Under **A2** and **the weak A3,**

$$\mathcal{E}_\ell'\left(\frac{\psi}{A}\right) = -\sigma'\left(\frac{\psi}{A}\right)\rho\left(\frac{\psi}{A}\right) - \sigma\left(\frac{\psi}{A}\right)\rho'\left(\frac{\psi}{A}\right) < 0 \Leftrightarrow \frac{\partial^2 \ln\ell(\psi/A)}{\partial\psi\partial(1/A)} < 0.$$

Proposition 3a states that the price, $p_\psi = Z(\psi/A)A$, the markup rate, $\mu_\psi = \mu(\psi/A)$, and the normalized price, $Z(\psi/A)$, are all weakly (strictly) log-supermodular in ψ and $1/A$ under the weak (strong) A3. More competitive pressures thus cause a markup rate decline, proportionately no larger (strictly smaller) among higher-ψ firms. Since their markup rates are lower under A2, this also implies no larger (strictly smaller) dispersion of the markup rate across firms. Figure 3b illustrates this by plotting the graphs of log-markup rate, $\ln\mu_\psi$, as a function of log-marginal cost, $\ln\psi$. The graph is downward-sloping under A2. It is strictly convex under strong A3. A higher $1/A$ causes a parallel leftward shift of the graph. Due to the convexity, this means a larger downward shift among lower-ψ firms experience proportionately larger decline in the markup rate. This suggests that more competitive pressures reduce the distortion due to the markup rate heterogeneity (i.e., high-ψ firms produce too much relative to low-ψ firms).

Proposition 3b states that the revenue, $r(\psi/A)E$, is strictly log-submodular in ψ and $1/A$ under A2 and the weak A3. This means that a higher $1/A$ causes a proportionately larger decline in the revenue among higher-ψ firms. Since their revenues are lower, this also implies that more competitive pressures lead to a larger dispersion of revenues across firms with the profit density shifting toward lower-ψ firms.

Thus, $R_\psi = r(\psi/A)E$ under A2 and the weak A3 share the same properties with $\Pi_\psi = \pi(\psi/A)E$ under A2, as depicted in Figure 3a.[37] This finding, a shift of the revenue density from the less productive/smaller firms with lower markup rates to the more productive/larger firms with higher markup rates, is also confirmed by the calibration results by Baqaee, Farhi, and Sangani (2024).

Proposition 3c states that the employment, $\ell(\psi/A)E$, is also strictly log-submodular in ψ and $1/A$ under A2 and the weak A3. However, its strict log-submodularity has different implications from that of the profit $\pi(\psi/A)E$ and the revenue $r(\psi/A)E$. This is because the employment $\ell(\psi/A)E$ is hump-shaped in ψ/A under A2 and the weak A3. To see this, we first prove in Appendix C.1:

Lemma 6: Under A2 and the weak A3, $\lim_{\psi/A\to 0} \rho(\psi/A)\sigma(\psi/A) < 1 < \lim_{\psi/A\to \bar{z}} \rho(\psi/A)\sigma(\psi/A)$.

Since $\mathcal{E}_\ell(\psi/A) = 1 - \rho(\psi/A)\sigma(\psi/A)$ is globally decreasing, Lemma 6 implies that there exists a unique $\hat{\psi} > 0$, such that $\mathcal{E}_\ell(\psi/A) > 0$ for $\psi < \hat{\psi}$ and $\mathcal{E}_\ell(\psi/A) < 0$ for $\psi > \hat{\psi}$. Thus,

Proposition 4: Under **A2** and **the weak A3,** the employment function, $\ell(\psi/A) = r(\psi/A)/\mu(\psi/A)$ is hump-shaped, with its unique peak is reached at, $\hat{z} \equiv Z(\hat{\psi}/A) < \bar{z}$, where

$$\mathcal{E}_{s(\zeta-1)/\zeta}(\hat{z}) = 0 \Leftrightarrow \frac{\hat{z}\zeta'(\hat{z})}{\zeta(\hat{z})} = [\zeta(\hat{z}) - 1]^2 \Leftrightarrow \mathcal{E}_\ell\left(\frac{\hat{\psi}}{A}\right) = 0 \Leftrightarrow \rho\left(\frac{\hat{\psi}}{A}\right)\sigma\left(\frac{\hat{\psi}}{A}\right) = 1.$$

Figure 3c illustrates Propositions 3c and 4 by plotting the log-employment as a function of the log-marginal cost, which is not only strictly concave (Proposition 3c) but also hump-shaped (Proposition 4). Thus, depending on the location of the peak, $\hat{\psi}$, relative to the lower bound of the marginal cost, $\underline{\psi}$, and to the cutoff, ψ_c, there are three generic equilibrium configurations, as shown in the following corollary, along with the underlying condition for each of the three cases. The proof is straightforward and hence omitted.

Corollary of Proposition 4: Suppose **A2** and **the weak A3** hold. Then, among the active firms, $\psi \in (\underline{\psi}, \psi_c)$, the employments are:

- decreasing in ψ ($\hat{\psi} < \underline{\psi}$), iff $A < \underline{\psi}/Z^{-1}(\hat{z})$;
- increasing in ψ ($\psi_c < \hat{\psi}$), iff $F/E = \pi(\psi_c/A) > \pi(\hat{\psi}/A) = \pi(Z^{-1}(\hat{z}))$;
- hump-shaped in ψ ($\underline{\psi} < \hat{\psi} < \psi_c$), iff $F/E < \pi(\hat{\psi}/A) = \pi(Z^{-1}(\hat{z}))$ & $A > \underline{\psi}/Z^{-1}(\hat{z})$.

In the first case, the employments are positively related to firm productivity. This case prevails when $\underline{\psi}$ is sufficient high and competitive pressures are sufficiently strong. In the second case, the employments are

[37] Similar to the case of the profit, if A declines due to a higher E, the total impact of a higher E on the revenue profit is captured by a combination of the parallel upward shift (the positive direct effect) and the parallel leftward shift (the indirect effect due to a lower A) in Figure 3a. We will look at the total impact in more detail in Proposition 7b.

negatively related to productivity across all active firms. This occurs iff $F/E > \pi\big(Z^{-1}(\hat{z})\big)$, i.e., when the overhead is sufficiently high relative to market size, so that only the very productive firms stay. In the third case, the employments are negatively (positively) related to productivity among the very (not so) productive firms.[38]

Figure 3c also depicts a higher $1/A$ causing a parallel leftward shift of the graph and a higher E causing a parallel upward shift of the graph. Due to its hump-shape, a higher $1/A$ alone causes a crossing of the graphs before and after the change. Thus, the employments of low-ψ firms go up due to more competitive pressures even if market size is unchanged.[39] This never happens for the profit and revenue; more competitive pressures always reduce the profit and revenue for all firms, unless caused by an increase in market size.

For the pass-through rate function, we prove in Appendix C.2.,

Proposition 5: Suppose that A2 and the strong A3 hold, so that $0 < \rho(\psi/A) < 1$ and $\rho(\psi/A)$ is strictly increasing. Then, $\rho(\psi/A)$ is strictly log-supermodular for all $\psi/A < \overline{z}$ with a sufficiently small $\overline{z}$.

Figure 3d illustrates Proposition 5. It states that, under the strong A3, more competitive pressures causes a proportionately smaller increase in the pass-through rate for lower-ψ firms for a sufficiently small $\overline{z} > 0$.

5. Heterogenous Firms under H.S.A.: Comparative Statics

In Section 4, we studied how a change in competitive pressures, A, an endogenous variable, has differential effects on heterogeneous firms without specifying underlying exogenous shocks that cause it. We now study the effects of exogenous shocks to the entry cost F_e, the overhead F, and market size E. The recursive structure of the model allows us to proceed in two steps. First, we study the effects on competitive pressures, A and the cutoff, ψ_c, in section 5.1. and explore some of the implications in sections 5.2 and 5.3. Then, we study the effects on M and $MG(\psi_c)$ in section 5.4. Finally, we consider the limit case, $F \to 0$, where the cutoff firms are those that charge the choke price.

5.1. Effects of F_e, F, and E on ψ_c, ψ_c/A and A

Recall that the equilibrium values of $A = A(\mathbf{p})$ and ψ_c are uniquely determined by eq.(5) and eq.(6), as functions of F_e/E and F/E. By totally differentiating eq.(5) and eq.(6),

Proposition 6:

$$\begin{bmatrix} d\ln A \\ \\ d\ln\psi_c \end{bmatrix} = \frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} \begin{bmatrix} 1-f_x & & f_x \\ & & \\ 1-f_x & & f_x-\delta \end{bmatrix} \begin{bmatrix} d\ln(F_e/E) \\ \\ d\ln(F/E) \end{bmatrix},$$

---

[38] Even though Proposition 4 implies that the prediction under CES --the employment is a globally increasing function in firm productivity--, is not robust, the corollary suggests that, to find the evidence that the employment is decreasing in firm productivity, we need to look for an industry with a high overhead cost and a low labor cost.

[39] This occurs whenever $\ell(\cdot)$ is hump-shaped, for which A2 +the weak A3 is sufficient but not necessary. For example, $\ell(\cdot)$ is hump-shaped under Generalized Tranlog (Appendix D.1) for $\eta < 1$, though it violates the weak A3.

where

$$\frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} = \frac{1}{\mathbb{E}_\pi(\sigma) - 1} = \{\mathbb{E}_r[\mu^{-1}]\}^{-1} - 1 = \mathbb{E}_\ell(\mu) - 1 > 0$$

is the average profit/the average labor cost ratio among the active firms;

$$f_x \equiv \frac{FG(\psi_c)}{F_e + FG(\psi_c)} = \frac{\pi(\psi_c/A)}{\mathbb{E}_1(\pi)} < 1$$

is the share of the overhead in the total expected fixed cost, which is equal to the profit of the cut-off firm relative to the average profit among the active firms; and

$$\delta \equiv \frac{\mathbb{E}_\pi(\sigma) - 1}{\sigma(\psi_c/A) - 1} = \frac{\pi(\psi_c/A)}{\ell(\psi_c/A)} \frac{\mathbb{E}_1(\ell)}{\mathbb{E}_1(\pi)} \equiv f_x \frac{\mathbb{E}_1(\ell)}{\ell(\psi_c/A)} > 0$$

is the profit/labor cost ratio of the cut-off firm to the average profit/the average labor cost ratio among the active firms.

The derivation is straightforward and hence omitted. To summarize the effects of each shock separately.

Corollary of Proposition 6:

a) Entry Cost: $\frac{d\ln A}{d\ln F_e} = \frac{d\ln \psi_c}{d\ln F_e} = \frac{(1-f_x)\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} > 0$; $\frac{d\ln(\psi_c/A)}{d\ln F_e} = 0$

b) Market Size: $\frac{d\ln A}{d\ln E} = -\frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} < 0$; $\frac{d\ln(\psi_c/A)}{d\ln E} = \frac{\delta\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} > 0$; $\frac{d\ln \psi_c}{d\ln E} = \frac{(\delta-1)\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} \gtreqless 0 \Leftrightarrow \mathbb{E}_\pi(\sigma) \gtreqless \sigma(\psi_c/A)$. In particular, $\frac{d\ln \psi_c}{d\ln E} < 0$ holds globally if $\sigma'(\cdot) > 0$, i.e., under A2.

c) Overhead Cost: $\frac{d\ln A}{d\ln F} = \frac{f_x\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} > 0$; $\frac{d\ln(\psi_c/A)}{d\ln F} = -\frac{\delta\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} < 0$; $\frac{d\ln \psi_c}{d\ln F} = \frac{(f_x-\delta)\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} \gtreqless 0 \Leftrightarrow \ell(\psi_c/A) \gtreqless \mathbb{E}_1(\ell)$. In particular, $\frac{d\ln \psi_c}{d\ln F} > 0$ holds globally if $\ell'(\cdot) > 0$.

Figures 4a-4c illustrate Corollary of Proposition 6. Figure 4a shows **a decline in F_e.** A smaller entry cost makes the entry more attractive, while keeping an incentive to stay in the market after the entry unaffected. Thus, it shifts the free entry condition down and to the left, while keeping the cutoff rule unchanged. Hence, it leads to a decline in both ψ_c and A at the same rate, resulting in more competitive pressures and a tougher selection. Figure 4b shows **an increase in E.** A larger market size has two different effects. On one hand, it makes the entry more attractive, thus shifting the free entry condition down and to the left. On the other hand, it gives more incentive to stay in the market after the entry at each level of competitive pressures, thus rotating the cutoff rule counter-clockwise. The intersection thus unambiguously moves to the left, a decline in A. The impact on ψ_c depends on the relative magnitudes of the two effects. Eliminating E from eq.(5) and eq.(6) yields:

$$\int_{\underline{\psi}}^{\psi_c} \left[\frac{\pi(\psi/A)}{\pi(\psi_c/A)} - 1\right] dG(\psi) = \frac{F_e}{F},$$

which shows the locus along which the intersection moves as E changes. Its LHS is globally strictly increasing in ψ_c. It is also strictly decreasing in A, wherever $\mathbb{E}_\pi(\sigma) < \sigma(\psi_c/A)$ holds:[40] that is, whenever the profit-weighted average price elasticity across the active firms is lower than the price elasticity at the cutoff firm. This condition holds globally, if $\sigma(\cdot)$ is strictly increasing, i.e., **A2,** in which case the locus is globally upward-sloping, as depicted by the dotted line in Figure 4b. Thus, under **A2**, a higher E always causes a decline in both ψ_c and A, with ψ_c/A going up.[41]

Figure 4c shows **a decline in F**. Similar to a higher E, a smaller overhead cost has two different effects. It not only makes the entry more attractive, thus shifting the free entry condition down and to the left, but also gives more incentive to stay in the market after the entry, thus rotating the cutoff rule counter-clockwise. The intersection thus unambiguously moves to the left, causing a decline in A. To see the impact on ψ_c, eliminating F from eq.(5) and eq.(6) yields:

$$\int_{\underline{\psi}}^{\psi_c} \left[\pi\left(\frac{\psi}{A}\right) - \pi\left(\frac{\psi_c}{A}\right)\right] dG(\psi) = \frac{F_e}{E}.$$

As F changes, the intersection moves along this locus. Its LHS is globally strictly increasing in ψ_c. It is also strictly decreasing in A, wherever $f_x > \delta$, or equivalently $\ell(\psi_c/A) > \mathbb{E}_1(\ell)$ holds.[42] that is, whenever the average employment across the active firms is lower than the employment by the cutoff firm. This condition holds globally if $\ell(\cdot)$ is strictly increasing. As shown in Corollary of Proposition 4, this occurs under A2 and the weak A3 when the overhead cost is sufficiently large relative to market size. In this case, the locus is globally upward-sloping, as depicted by the dotted curve in Figure 4c. Hence a lower F always causes a decline in both ψ_c and A, with ψ_c/A going up.

5.2. Market Size Effect on Profit, $\boldsymbol{\Pi_\psi \equiv \pi(\psi/A)E}$ and Revenue, $\boldsymbol{R_\psi \equiv r(\psi/A)E}$

As we suggested in section 4, the full impacts of a higher E on the profit (under A2) and of the revenue (under A2 and the weak A3) are captured by a combination of the parallel upward shift (the direct effect) and the parallel leftward shift (the indirect effect due to a lower A) of the graph in Figure 3a. Because the positive direct effect is uniform across firms, while the negative indirect effect is smaller for low-ψ firms, the combined effect could result in a clockwise rotation of the graph, such that a higher E, accompanied by a lower A, leads to an increase in the profit and the revenue among low-ψ firms. We are now ready to state this result formally in Propositions 7a and 7b, whose proof is in Appendix C.3.

[40] This can be verified by differentiating the LHS with respect to A and making use of $\mathcal{E}_\pi(\psi/A) = 1 - \sigma(\psi/A)$.
[41] Since A2 implies the log-submodularity of $\pi(\psi/A)$, as shown in Proposition 2, $\pi(\psi/A)/\pi(\psi_c/A)$ is strictly decreasing in A for $\psi < \psi_c$, and so is the integrand of the LHS. Under the opposite of A2, $\sigma'(\cdot) < 0$, the locus would be negatively-sloped and a higher E would lead to an increase in ψ_c. CES is the borderline case, with the horizontal locus, hence a change in E has no effect on ψ_c.
[42] This can be verified by differentiating the LHS with respect to A and making use of $(\psi/A)\pi'(\psi/A) = \pi(\psi/A)\mathcal{E}_\pi(\psi/A) = \pi(\psi/A)[1 - \sigma(\psi/A)] = \pi(\psi/A) - r(\psi/A) = -\ell(\psi/A)$

Proposition 7a: Under **A2,** there exists a unique $\psi_0 \in (\underline{\psi}, \psi_c)$ such that $\sigma\left(\frac{\psi_0}{A}\right) = \mathbb{E}_\pi(\sigma)$ with

$$\frac{d \ln \Pi_\psi}{d \ln E} > 0 \Leftrightarrow \sigma\left(\frac{\psi}{A}\right) < \mathbb{E}_\pi(\sigma) \text{ for } \psi \in \left(\underline{\psi}, \psi_0\right),$$

and

$$\frac{d \ln \Pi_\psi}{d \ln E} < 0 \Leftrightarrow \sigma\left(\frac{\psi}{A}\right) > \mathbb{E}_\pi(\sigma) \text{ for } \psi \in (\psi_0, \psi_c).$$

Proposition 7b: Under **A2** and the weak **A3,** there exists $\psi_1 > \psi_0$, such that

$$\frac{d \ln R_\psi}{d \ln E} > 0 \text{ for } \psi \in \left(\underline{\psi}, \psi_1\right).$$

Furthermore, $\psi_1 \in (\psi_0, \psi_c)$ and

$$\frac{d \ln R_\psi}{d \ln E} < 0 \text{ for } \psi \in (\psi_1, \psi_c),$$

for a sufficiently small F.

Notice one critical difference between the impact on the profit under A2 (Proposition 7a) and that on the revenue under A2 plus the weak A3 (Proposition 7b). For the former, a clockwise rotation of the graph, as depicted in Figure 3a, occurs around the pivot point, ψ_0, which is always below the cutoff ψ_c. This is because the profit at the cutoff is always equal to the overhead cost, F. This means that a higher E always creates both winners among low-ψ firms and losers among high-ψ firms, generating what Mrázová-Neary (2017; 2019) dubbed as The Matthew Effect." For the latter, the pivot point, ψ_1, *may* be above the cutoff ψ_c. If so, all firms would experience an increase in their revenue. In Proposition 7b, we rule out this possibility only for a sufficiently small F.

In Appendix E, Figures 5a-5c graphically put together the main implications of Propositions 2, 3, 6, and 7 under A2 and the weak A3 for the effects on the log-markup rates, the log-profits, and the log-revenues, of more competitive pressures (a lower A) and a tougher selection (a lower ψ_c), when they are caused by a decline in F_e, an increase in E and a decline in F (with $\ell'(\cdot) > 0$). In all three cases, the log-profit is decreasing, and concave in the log-marginal cost due to A2 (Proposition 2) and the log-markup rate (log-revenue) is decreasing, and convex (concave) in the log-marginal cost due to A2 and the weak A3 (Proposition 3).

5.3. The Composition Effect: Average Markup and Pass-Through Rates and P/A.

In all three cases illustrated in Figures 4a-4c and Figures 5a-5c, the shocks that cause a decline in A, more competitive pressures, also cause a decline in ψ_c, a tougher selection. This creates non-trivial composition effects.

- Under A2, a lower A causes all surviving firms to reduce their markup rate $\mu(\psi/A)$. But it also causes the distribution to shift toward low-ψ firms with higher $\mu(\psi/A)$.
- Under strong A3, a lower A causes all surviving firms to increase their pass-through rates $\rho(\psi/A)$. But it also causes the distribution to shift toward low-ψ firms with lower $\rho(\psi/A)$.

Due to this composition effect, the average markup (and/or pass-through) rate may go in the opposite direction from the firm-level markup (and/or pass-through) rate. The next proposition is useful to answer the question under which conditions this happens.

Proposition 8: Assume that $\mathcal{E}_g'(\cdot)$ does not change its sign and $\underline{\psi} = 0$. Consider a shock to F_e, E, and/or F, which affects competitive pressures, i.e., $dA \neq 0$. Then, the response of any weighted generalized mean of any monotone function, $f(\psi/A) > 0$, defined by

$$I \equiv \mathcal{M}^{-1}\left(\mathbb{E}_w\big(\mathcal{M}(f)\big)\right)$$

with a monotone transformation $\mathcal{M}: \mathbb{R}_+ \to \mathbb{R}$ and a weighting function, $w(\psi/A) > 0$, satisfies:

	$f'(\cdot) > 0$	$f'(\cdot) = 0$	$f'(\cdot) < 0$
$\mathcal{E}_g'(\cdot) > 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Longrightarrow \frac{d\ln I}{d\ln A} > 0$	$\frac{d\ln I}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Longrightarrow \frac{d\ln I}{d\ln A} < 0$
$\mathcal{E}_g'(\cdot) = 0$ (Pareto)	$\frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln I}{d\ln A} \gtreqless 0$	$\frac{d\ln I}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln I}{d\ln A} \lesseqgtr 0$
$\mathcal{E}_g'(\cdot) < 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Longrightarrow \frac{d\ln I}{d\ln A} < 0$	$\frac{d\ln I}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Longrightarrow \frac{d\ln I}{d\ln A} > 0$

Moreover, if $\mathcal{E}_g'(\cdot) = \frac{d\ln(\psi_c/A)}{d\ln A} = 0$, $d\ln I/d\ln A = 0$ for any $f(\psi/A)$, monotonic or not. Furthermore, $\mathcal{E}_g'(\cdot)$ can be replaced with $\mathcal{E}_G'(\cdot)$ in all the above statements for $w(\psi/A) = 1$, i.e., the unweighted averages.

The proof is in Appendix C.4. Proposition 8 states that the impact on the weighted average of a monotone function $f(\cdot)$ depends not only on the sign of $f'(\cdot)$ but also on the signs of $\frac{d\ln(\psi_c/A)}{d\ln A}$ and of $\mathcal{E}_g'(\cdot)$. Here, the average can be any generalized mean of $f(\cdot) > 0$, including the arithmetic mean, $I \equiv \mathbb{E}_w(f)$ with $\mathcal{M}(f) = f$, the geometric mean, $I \equiv \exp[\mathbb{E}_w(\ln f)]$ with $\mathcal{M}(f) = \ln f$, and the harmonic mean $I \equiv [\mathbb{E}_w(f^{-1})]^{-1}$ with $\mathcal{M}(f) = f^{-1}$. Moreover, the weight $w(\cdot)$ can be any function of ψ/A, including the distribution of the revenue $r(\cdot)$, the profit $\pi(\cdot)$, or even the employment $\ell(\cdot)$, which may not be monotone in ψ/A.[43] Proposition 8 also states that a decline in F_e under Pareto, $G(\psi) = \left(\psi/\overline{\psi}\right)^{\kappa}$, offers a knife-edge case, where any $w(\cdot)$-weighted generalized mean of even a nonmonotonic $f(\cdot)$ remain

[43] Of course, which weighted generalized mean is used matters both conceptually and quantitatively; as stressed by Edmond, Midrigan, and Xu (2023).

unchanged. Proposition 8 also states that, for the *unweighted* generalized mean, the condition on the sign of $\mathcal{E}_g'(\cdot)$ can be replaced with the weaker condition on the sign of $\mathcal{E}_G'(\cdot)$ (Lemma 1 of Appendix A).[44]

Considering shocks to F_e, E, and F separately,

Corollary 1 of Proposition 8

a) Entry Cost: $f'(\cdot)\mathcal{E}_g'(\cdot) \gtreqless 0 \Leftrightarrow \frac{d\ln I}{d\ln F_e} = \frac{d\ln I}{d\ln A}\frac{d\ln A}{d\ln F_e} \gtreqless 0.$

b) Market Size: If $\mathcal{E}_g'(\cdot) \leq 0$, then, $f'(\cdot) \gtreqless 0 \Rightarrow \frac{d\ln I}{d\ln E} = \frac{d\ln I}{d\ln A}\frac{d\ln A}{d\ln E} \gtreqless 0.$

c) Overhead Cost: If $\mathcal{E}_g'(\cdot) \leq 0$, then, $f'(\cdot) \gtreqless 0 \Rightarrow \frac{d\ln I}{d\ln F} = \frac{d\ln I}{d\ln A}\frac{d\ln A}{d\ln F} \lesseqgtr 0.$

Furthermore, $\mathcal{E}_g'(\cdot)$ can be replaced with $\mathcal{E}_G'(\cdot)$ for $w(\psi/A) = 1$, i.e., the unweighted averages.

To interpret Corollary 1a) of Proposition 8, let $\mu(\cdot) = f(\cdot)$ under A2, $\mu'(\cdot) < 0$. Then, this result states that a lower A, due to a decline in F_e, causes any $w(\cdot)$-weighted generalized mean of the markup rate to increase if $\mathcal{E}_g'(\cdot) > 0$, and to decline if $\mathcal{E}_g'(\cdot) < 0$, with the Pareto case, $\mathcal{E}_g'(\cdot) = 0$, being the knife-edge. Likewise, for $\rho(\cdot) = f(\cdot)$ under the strong A3, $\rho'(\cdot) > 0$, a lower A, due to a decline in F_e, causes any $w(\cdot)$-weighted generalized mean of the pass-through rate to decline if $\mathcal{E}_g'(\cdot) > 0$, and to increase if $\mathcal{E}_g'(\cdot) < 0$. Thus, $\mathcal{E}_g'(\cdot) > 0$ is sufficient and necessary under which the composition effect dominates such that the average markup and pass-through rates move in the opposite direction from the firm-level markup and pass-through rates, while $\mathcal{E}_g'(\cdot) < 0$ is sufficient and necessary for the average rates to move in the same direction with the firm-level rates. To grasp the intuition, recall Lemma 2, which states that, when $\mathcal{E}_g'(\cdot) > 0$, a lower ψ_c (a tougher selection) shifts the distribution of $\xi \equiv \psi/\psi_c$ to the left in the MLR ordering. Thus, among the surviving firms, the distribution becomes more skewed towards low-ψ firms, which have higher markup and lower pass-through rates. This makes the composition effect dominate, causing the average markup rate to go up and the average pass-through rate to go down under more competitive pressures, despite that firm-level markup rates are down and firm-level pass-through rates are up. Interestingly, according to the calibration by Baqaee, Farhi, and Sangani (2024), which showed the evidence for A2 and strong A3, $\mathcal{E}_g'(\psi) > 0$ holds with a Pareto tail, $\lim_{\psi\to 0}\mathcal{E}_g'(\psi) = 0$.[45] This suggests that, more competitive pressures, through the composition effect, might have caused the recent rise in the average markup rate and the decline in the average pass-through rate.[46] At least, such empirical findings should not be interpreted as the *prima-facie* evidence for less competitive pressures. It

[44]In an earlier version of the paper, Matsuyama and Ushchev (2023a; Proposition 8b), we also showed that the direction of the change in the employment-weighted arithmetic mean of the markup rate in response to a change in the entry depends on the sign of $\mathcal{E}_G'(\cdot)$.

[45]This calibration finding is not in their paper. But, in response to our inquiry, the authors kindly computed and sent us the plot that shows that the elasticity of the productivity density function is positive with a Pareto tail, thereby confirming this finding.

[46]Indeed, Autor et.al. (2020) and De Loecker, Eeckhout, and Unger (2020) pointed out that much of the recent rise in the average markup is due to reallocation from the low markup firms to the high markup firms.

should also be pointed out that, as discussed in Section 2.5, the aggregate labor cost share is the reciprocal of the revenue-weighted harmonic mean of the markup rates and their employment-weighted arithmetic mean. The above result thus implies that, under A2, a lower A, due to a decline in F_e, causes the aggregate labor cost share to decline and the aggregate profit share to increase if $\mathcal{E}_g'(\cdot) > 0$. It has the opposite effect if $\mathcal{E}_g'(\cdot) < 0$ and no effect if $\mathcal{E}_g'(\cdot) = 0$.

Pareto, $\mathcal{E}_g'(\cdot) = 0$, is the knife-edge case in which the average rate does not move in response to a change in F_e, because it implies $\frac{d\ln(\psi_c/A)}{d\ln A} = 0$, as seen in Corollary a) of Proposition 6. In contrast, $\frac{d\ln(\psi_c/A)}{d\ln A} > 0$ for a change in E or in F, as seen in Corollary b) and c) of Proposition 6. This weakens the composition effect. As a result, $\mathcal{E}_g'(\cdot) \leq 0$ is sufficient for the average markup and pass-through rates to move in the same direction with the firm-level rates, as seen in Corollary b) and c) of Proposition 8. In other words, $\mathcal{E}_g'(\cdot) > 0$ is necessary (but not sufficient) for the average markup and pass-through rates to move in the opposite direction from the firm-level markup and pass-through rates.

Proposition 8 is also useful for finding the impact of more competitive pressures on P/A.

Corollary 2 of Proposition 8: Assume $\underline{\psi} = 0$, and neither $\zeta'(\cdot)$ nor $\mathcal{E}_g'(\cdot)$ change the signs. Consider a shock to F_e, E, and/or F, which affects competitive pressures, i.e., $dA \neq 0$. Then, the response of P/A satisfies:

	$\zeta'(\cdot) > 0$ (A2)	$\zeta'(\cdot) = 0$ (CES)	$\zeta'(\cdot) < 0$
$\mathcal{E}_g'(\cdot) > 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Rightarrow \frac{d\ln(P/A)}{d\ln A}$	$\frac{d\ln(P/A)}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Rightarrow \frac{d\ln(P/A)}{d\ln A}$
$\mathcal{E}_g'(\cdot) = 0$ (Pareto)	$\frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln(P/A)}{d\ln A}$	$\frac{d\ln(P/A)}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln(P/A)}{d\ln A}$
$\mathcal{E}_g'(\cdot) < 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Rightarrow \frac{d\ln(P/A)}{d\ln A}$	$\frac{d\ln(P/A)}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Rightarrow \frac{d\ln(P/A)}{d\ln A}$

Again, the proof is in Appendix C.4.

5.4. Comparative Statics on M, $MG(\psi_c)$ and TFP

The impact on the mass of entrants, M, is simple. From eq.(7), it immediately follows that it always increases under shocks that lead to more competitive pressures, $dA < 0$, and a tough selection, $d\psi_c < 0$, including all three cases illustrated in Figures 4a-4c and Figures 5a-5c.[47]

Let us now turn to the effects on the mass of active firms. The proof is in Appendix C.5.

Proposition 9: Assume that $\mathcal{E}_G'(\cdot)$ does not change its sign and $\underline{\psi} = 0$. Consider a shock to F_e, F, and/or E, which affects competitive pressures, i.e., $dA \neq 0$. Then, the response of the mass of active

[47] The question remains how M/E changes in response to a change in E. This turns out be a difficult question, and we were able to show only under A2 and under Pareto, M/E goes up in response to an increase in E; see Matsuyama and Ushchev (2023a; the first part of Proposition 9c).

firms, $MG(\psi_c)$, is as follows:

$$\text{If } \mathcal{E}_G'(\cdot) > 0, \qquad \frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Rightarrow \frac{d\ln[MG(\psi_c)]}{d\ln A} > 0;$$

$$\text{If } \mathcal{E}_G'(\cdot) = 0, \qquad \frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln[MG(\psi_c)]}{d\ln A} \gtreqless 0;$$

$$\text{If } \mathcal{E}_G'(\cdot) < 0, \qquad \frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Rightarrow \frac{d\ln[MG(\psi_c)]}{d\ln A} < 0.$$

Corollary 1 of Proposition 9

a) Entry Cost: $\mathcal{E}_G'(\cdot) \gtreqless 0 \Leftrightarrow \frac{d\ln[MG(\psi_c)]}{d\ln F_e} = \frac{d\ln[MG(\psi_c)]}{d\ln A}\frac{d\ln A}{d\ln F_e} \gtreqless 0.$

b) Market Size: $\mathcal{E}_G'(\cdot) \leq 0 \Rightarrow \frac{d\ln[MG(\psi_c)]}{d\ln E} = \frac{d\ln[MG(\psi_c)]}{d\ln A}\frac{d\ln A}{d\ln E} > 0.$

c) Overhead Cost: $\mathcal{E}_G'(\cdot) \leq 0 \Rightarrow \frac{d\ln[MG(\psi_c)]}{d\ln F} = \frac{d\ln[MG(\psi_c)]}{d\ln A}\frac{d\ln A}{d\ln F} < 0.$

Proposition 9 states that the impact on the mass of active firms depends on the signs of $\mathcal{E}_G'(\cdot)$ and $\frac{d\ln(\psi_c/A)}{d\ln A}$. In particular, its Corollary states that a decline in F_e causes the masses of active firms to go down if and only if $\mathcal{E}_G'(\cdot) > 0$, go up if and only if $\mathcal{E}_G'(\cdot) < 0$, with Pareto $\mathcal{E}_G'(\cdot) = 0$ being the knife-edge case, and that $\mathcal{E}_G'(\cdot) \leq 0$ is sufficient for $MG(\psi_c)$ to go up and $\mathcal{E}_G'(\cdot) > 0$ necessary for $MG(\psi_c)$ to go down in response to an increase in E or a decline in F.[48] It is worth noting that what matters here is the sign of $\mathcal{E}_G'(\cdot)$, which are weaker conditions than the sign of $\mathcal{E}_g'(\cdot)$; see Lemma 1 of Appendix A.

By combining Corollary 2 of Proposition 8 and Corollary 1 of Proposition 9, we now summarize the sufficient conditions under which a decline in A leads to a decline in P , i.e., a higher TFP, in the following corollary. Most of these results follow from Corollary 2 of Proposition 8, except $\frac{d\ln P}{d\ln A} > 0$ in the case of $\mathcal{E}_g'(\cdot) \leq 0$ and $\zeta'(\cdot) > 0$. This follows from Corollary 1 of Proposition 9, which shows that, under $\mathcal{E}_g'(\cdot) \leq 0$ and hence under $\mathcal{E}_G'(\cdot) \leq 0$, a lower F_e reduces A and increases $MG(\psi_c)$ weakly, and both a higher E and a lower F reduce A and increase $MG(\psi_c)$ strictly. This means that these three shocks lead to $dP/P < 0$ under $\mathcal{E}_g'(\cdot) \leq 0$ and $\zeta'(\cdot) > 0$.[49]

Corollary 2 of Proposition 9: Assume $\underline{\psi} = 0$, and neither $\zeta'(\cdot)$ nor $\mathcal{E}_g'(\cdot)$ change the signs. Consider a shock to F_e, E, and/or F, which affects competitive pressures, i.e., $dA \neq 0$. Then, the response of P satisfies:

[48]In Matsuyama and Ushchev (2023a; the second part of Proposition 9c), we also showed under A2 that $\mathcal{E}_G'(\cdot) \geq 0$ implies $MG(\psi_c)/E$ goes down in response to an increase in E.

[49]We have not been able to rule out the possibility $d\ln P/d\ln A < 0$ under $\mathcal{E}_g'(\cdot) > 0$ and $\zeta'(\cdot) < 0$ as well as under $\mathcal{E}_g'(\cdot) > 0$ and $\zeta'(\cdot) > 0$, for shocks to E or F, which would mean that a decline in A could cause an increase in P, i.e., a decline in TFP. However, the calibration by Baqaee, Farhi, and Sangani (2024) shows $\mathcal{E}_g'(\cdot) > 0$ and $\zeta'(\cdot) > 0$ and that a higher E leads to higher TFP, much of which is due to what they call the Darwinian effect, the reallocation from high-ψ firms to low-ψ firms.

	$\zeta'(\cdot) > 0$ (A2)	$\zeta'(\cdot) = 0$ (CES)	$\zeta'(\cdot) < 0$
$\mathcal{E}'_g(\cdot) > 0$	$\frac{d\ln P}{d\ln A} > 1 \; for \; F_e$	$\frac{d\ln P}{d\ln A} = 1$	?
$\mathcal{E}'_g(\cdot) = 0$ (Pareto)	$\frac{d\ln P}{d\ln A} = 1 \; for \; F_e$ $0 < \frac{d\ln P}{d\ln A} < 1 \; for \; F \; or \; E$	$\frac{d\ln P}{d\ln A} = 1$	$\frac{d\ln P}{d\ln A} = 1 \; for \; F_e$ $\frac{d\ln P}{d\ln A} > 1 \; for \; F \; or \; E$
$\mathcal{E}'_g(\cdot) < 0$	$0 < \frac{d\ln P}{d\ln A} < 1$	$\frac{d\ln P}{d\ln A} = 1$	$\frac{d\ln P}{d\ln A} > 1$

5.5. The Limit Case of $F \to 0$ with $\bar{z} < \infty$.

Before proceeding to a multi-market extension, we briefly look at a limit case, $F \to 0$, with $\bar{z} < \infty$. In this limit case, there is no overhead cost and the cutoff firms supply with zero markup, i.e., at that marginal cost equal to the choke price, $\psi_c = \bar{z}A$.[50] The equilibrium can be described by eq.(5) and eq.(6), which now become simply:

Cutoff Rule:

$$\pi\left(\frac{\psi_c}{A}\right) = 0 \Leftrightarrow \frac{\psi_c}{A} = Z\left(\frac{\psi_c}{A}\right) = \bar{z} = \pi^{-1}(0)$$

Free Entry Condition:

$$\frac{F_e}{E} = \int_{\underline{\psi}}^{\psi_c} \pi\left(\bar{z}\frac{\psi}{\psi_c}\right) dG(\psi) = \int_{\underline{\psi}}^{\bar{z}A} \pi\left(\frac{\psi}{A}\right) dG(\psi).$$

Notice that the cutoff rule alone determines $\psi_c/A = \bar{z}$. And the free-entry condition uniquely determines $\psi_c = \bar{z}A$ as functions of F_e/E with the interior solution, $0 < G(\psi_c) < 1$, ensured for

$$0 < \frac{F_e}{E} < \int_{\underline{\psi}}^{\bar{\psi}} \pi\left(\bar{z}\frac{\psi}{\bar{\psi}}\right) dG(\psi).$$

Simple algebra can verify that

$$\frac{d\psi_c}{\psi_c} = \frac{dA}{A} = \frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)}\left(\frac{dF_e}{F_e} - \frac{dE}{E}\right),$$

which can be also obtained from Proposition 6 by setting $f_x = \delta = 0$. Thus, a decline in F_e/E causes both ψ_c and A to decline at the same rate, with ψ_c/A unchanged, as shown in Figure 6a of Appendix E. Thus, $\frac{d\ln M}{d\ln(F_e/E)} < 0$. Moreover, Propositions 8 and 9 and their corollaries can be applied with $\frac{d\ln(\psi_c/A)}{d\ln A} = 0$. Thus, for any weighted generalized mean of $f(\cdot), I$,

$$f'(\cdot)\mathcal{E}'_g(\cdot) \gtreqless 0 \Leftrightarrow \frac{d\ln I}{d\ln(F_e/E)} = \frac{d\ln I}{d\ln A}\frac{d\ln A}{d\ln(F_e/E)} \gtreqless 0;$$

[50] Although one of the advantages of the Melitz model under H.S.A. is that it is tractable with $F > 0$, we look at this case because some existing studies, e.g., Melitz and Ottaviano (2008) and Arkolakis et.al. (2019), assume the choke price and $F = 0$.

and for the mass of active firms,

$$\mathcal{E}_g'(\cdot) \gtreqless 0 \Leftrightarrow \frac{d\ln[MG(\psi_c)]}{d\ln(F_e/E)} = \frac{d\ln[MG(\psi_c)]}{d\ln A}\frac{d\ln A}{d\ln(F_e/E)} \gtreqless 0.$$

Figure 6b of Appendix E illustrates the impacts on the markup rate, the profit and the revenue. While a decline in F_e causes the profit and the revenue of all surviving firms to decline with proportionately larger impacts on low-ψ firms, an increase in E causes the profit and revenue to go up among low-ψ firms. The profit and revenue always go down among high-ψ firms, with the clockwise rotation of the profit and revenue schedule, whose pivot point (ψ_0 for the profit; $\psi_1$for the revenue) is always located below the cutoff ψ_c, because the cutoff firms always earn zero revenue and profit.

6. Sorting of Heterogeneous Firms Across Multiple Markets

6.1. A Multi-Market Setting

We now extend the model such that this sector has $J \geq 2$ separate markets, indexed as $j = 1,2,\ldots,J$, from which firms need to choose. The structure of each market is as before; it produces a single final good with the H.S.A. technology to assemble market-specific differentiated intermediate inputs supplied by MC competitive producers. The only source of the heterogeneity across markets is market size for the final goods produced in each market, E_j.[51] This allows us to index the markets such that $E_1 > E_2 > \cdots > E_J > 0$ without further loss of generality.

As before, each entrant must pay the entry cost, $F_e > 0$, to draw its marginal cost, ψ. Then, after learning its marginal cost, they decide which market to enter and produce with an overhead cost, $F > 0$, or exit without producing. If ψ-firms choose not to exit, they would enter the market that gives the highest profit to earn

$$\Pi_\psi = \max\{\Pi_{1\psi},\ldots,\Pi_{J\psi}\},$$

where

$$\Pi_{j\psi} \equiv \frac{s\left(Z(\psi/A_j)\right)}{\zeta\left(Z(\psi/A_j)\right)}E_j \equiv \frac{r(\psi/A_j)}{\sigma(\psi/A_j)}E_j = \pi\left(\frac{\psi}{A_j}\right)E_j$$

is the profit earned by ψ-firms by entering market-j and A_j is the inverse measure of competitive pressures in market-j. The free entry condition is then

[51] In particular, we keep it simple by assuming that the factor price is common across the markets so that it can be taken as the numeraire. This poses no problem if the J markets are not spatially separated. When they are spatially separated, this assumption may not be realistic, and yet it is innocuous because the factor price difference across markets affects the market choice of all firms equally regardless of their productivity level, so that it does not affect the relative incentive for heterogenous firms to sort themselves across different markets.

$$\int_{\underline{\psi}}^{\overline{\psi}} \max\{\Pi_\psi - F, 0\}\, dG(\psi) = F_e.$$

6.2. Positive Assortative Matching Between Firms and Markets under A2

We now show a positive assortative matching between firms and markets under **A2** in the sense that more productive firms self-select into larger markets. Specifically, we now show that there is a sequence of monotonically increasing cutoffs, $\underline{\psi} = \psi_0 < \psi_1 < \psi_2 < \cdots < \psi_J < \overline{\psi}$, such that firms with $\psi \in (\psi_{j-1}, \psi_j)$ enter market-j, and those with $\psi \in (\psi_J, \overline{\psi})$ do not enter any market.

First, we prove that A_j is strictly monotone in j. Suppose the contrary, so that, for some j, $E_j > E_{j+1}$ and $A_j \geq A_{j+1}$. Because $\pi(\cdot)$ is strictly decreasing, this would mean that, for all ψ,

$$\pi(\psi/A_j) \geq \pi(\psi/A_{j+1}) \Longrightarrow \Pi_{j\psi} = \pi(\psi/A_j)E_j > \pi(\psi/A_{j+1})E_{j+1} = \Pi_{(j+1)\psi}$$

which would imply that no firm would enter market-$(j+1)$, and hence $A_{j+1} = \infty$, a contradiction. Thus, $0 < A_1 < A_2 < \cdots < A_J < \infty$, and $\pi(\psi/A_1) < \pi(\psi/A_2) < \cdots < \pi(\psi/A_J)$ for all ψ.

Second, for $j = 1,2,\ldots,J-1$, consider the following ratio:

$$\frac{\Pi_{j\psi}}{\Pi_{(j+1)\psi}} = \frac{\pi(\psi/A_j)E_j}{\pi(\psi/A_{j+1})E_{j+1}}.$$

As a function of ψ, this ratio has to be greater than one for some ψ and less than one for other ψ, to ensure that a positive measure of firms would enter both market-j and market-$(j+1)$. Since **A2** implies that $\pi(\psi/A)$ is strictly log-submodular in ψ and $1/A$ (Proposition 2c), this ratio is strictly decreasing in ψ because $A_j < A_{j+1}$. Thus, there exists ψ_j such that

$$\psi \lesseqgtr \psi_j \Leftrightarrow \frac{\Pi_{j\psi}}{\Pi_{(j+1)\psi}} = \frac{\pi(\psi/A_j)E_j}{\pi(\psi/A_{j+1})E_{j+1}} \gtreqless \frac{\pi(\psi_j/A_j)E_j}{\pi(\psi_j/A_{j+1})E_{j+1}} \equiv 1.$$

In other words, all firms with $\psi < \psi_j$ strictly prefer entering market-j to entering market-$(j+1)$, all firms with $\psi > \psi_j$ strictly prefer entering market-$(j+1)$ to entering market-j, and all firms with $\psi = \psi_j$ are indifferent between the two. For $j = J$, let ψ_J be defined by $\pi(\psi_J/A_J)E_J \equiv F$. Then, only the firms with $\psi \in [\psi_{j-1}, \psi_j]$ enter market-j. This also means that ψ_j is strictly monotone in j, because $\psi_{j-1} \geq \psi_j$ would imply that a positive measure of firms would not enter market-j, a contradiction. See Figure 7. Thus, the mass of the active firms in market-j is equal to $M[G(\psi_j) - G(\psi_{j-1})]$, and the mass of the firms that enter and choose not to stay in any market is $M[1 - G(\psi_J)]$.

The free entry condition can now be rewritten as:

$$\sum_{j=1}^{J} \int_{\psi_{j-1}}^{\psi_j} \left\{\pi\left(\frac{\psi}{A_j}\right)E_j - F\right\} dG(\psi) = F_e \tag{10}$$

The adding-up constraint in market-j is given by:

$$M \int_{\psi_{j-1}}^{\psi_j} r\left(\frac{\psi}{A_j}\right) dG(\psi) = 1, \tag{11}$$

where the cutoff rules are:

$$\frac{\pi(\psi_j/A_j)E_j}{\pi(\psi_j/A_{j+1})E_{j+1}} = 1, \tag{12}$$

for $j = 1,2, \ldots, J - 1$, and

$$\pi\left(\frac{\psi_J}{A_J}\right) E_J \equiv F, \tag{13}$$

for $j = J$. Altogether, these $2J + 1$ conditions in eqs.(10)-(13) determine $2J + 1$ endogenous variables, which are M, $\{A_j\}_{j=1}^J$ and $\{\psi_j\}_{j=1}^J$, $0 < A_1 < A_2 < \cdots < A_J < \infty$; $\underline{\psi} = \psi_0 < \psi_1 < \psi_2 < \cdots < \psi_J < \overline{\psi}$. To summarize,

Proposition 10: Positive Assortative Matching between Firm Productivity and Market Size
Suppose that J markets differ only in market size, as $E_1 > E_2 > \cdots > E_J > 0$. In equilibrium, large markets are characterized by more competitive pressures, $0 < A_1 < A_2 < \cdots < A_J < \infty$. And under **A2,** firms with $\psi \in (\psi_{j-1}, \psi_j)$ enter market-j for $j = 1, 2, \ldots, J$, and firms with $\psi \in (\psi_J, \overline{\psi})$ exit, with $\underline{\psi} = \psi_0 < \psi_1 < \psi_2 < \cdots < \psi_J < \overline{\psi}$, where the two strictly increasing sequences, $\{\psi_j\}_{j=1}^J$ and $\{A_j\}_{j=1}^J$, and M, the mass of entrant, are given by eqs.(10)-(13).

A2 is crucial for this result. Under the opposite of **A2**, $\pi(\psi/A)$ would be strictly log-supermodular in ψ and $1/A$, so that $\pi(\psi/A_j)E_j/\pi(\psi/A_{j+1})E_{j+1}$ would be strictly *increasing* in ψ. Thus, there would be a *negative* assortative matching with more productive firms self-selecting into smaller markets. Under CES, $\pi(\psi/A_j)E_j/\pi(\psi/A_{j+1})E_{j+1}$ is *independent* of ψ, hence, in equilibrium, it has to be equal to one so that all active firms are indifferent across all markets, and the equilibrium distribution is *indeterminate*. [52]
Thus, under A2, this model offers a demand-side mechanism for the positive assortative matching between firm productivity and market size.[53, 54,]

[52]Baldwin and Okubo (2006) considered sorting of heterogeneous firms under the CES demand. The positive assortative matching in their model is due to their equilibrium selection criterion based on the protocol that larger firms choose their markets earlier, which they argue is plausible because larger firms gain more (but not proportionately) from choosing the larger market. Some criticize this protocol as ad hoc, because smaller firms may move faster since they are more agile. Our analysis suggests that such a criticism is unwarranted because, if we consider their CES demand as a limit of the H.S.A. under A2, the same equilibrium will be selected.
[53]Kokovin et. al. (2024) also generates a positive assortative matching through a demand-side mechanism under the 2nd law. In contrast to our approach, they use a quasi-linear utility defined over the outside good and the (nonhomothetic) DEA aggregator of differentiated consumer goods, and they needed to impose the condition on the market size distribution to ensure the uniqueness of the equilibrium. DEA can also be used to offer a micro foundation for the reduced form profit function used by Nocke (2006) to generate the positive assortative matching.

6.3. Cross-Sectional, Cross-Market Patterns

Figures 8a-8d illustrate the patterns of the profit, the revenue, the markup and pass-through rates across firms that emerge in equilibrium as more productive firms sort themselves into larger markets.

The profit schedule, $\Pi_\psi = \max_j\{\pi(\psi/A_j)E_j\}$, shown in Figure 8a, is obtained by the upper envelope of $\pi(\psi/A_j)E_j$. It is globally continuous and strictly decreasing in ψ, with the kink at the cutoff point, ψ_j. It is continuous at each cutoff, ψ_j, because the lower markup rate in market-j cancels out its larger market size, keeping ψ_j-firms indiffierent btw market-j & market-$(j+1)$.

The revenue schedule, R_ψ, shown in Figure 8b, is continuously decreasing in ψ within each market. However, it exhibits a downward jump at the cutoff ψ_j $(j = 1, 2, \dots, J-1)$, as

$$\frac{r(\psi_j/A_j)E_j}{r(\psi_j/A_{j+1})E_{j+1}} = \frac{\sigma(\psi_j/A_j)\pi(\psi_j/A_j)E_j}{\sigma(\psi_j/A_{j+1})\pi(\psi_j/A_{j+1})E_{j+1}} = \frac{\sigma(\psi_j/A_j)}{\sigma(\psi_j/A_{j+1})} > 1.$$

This is because, if ψ_j-firms switch from market-$(j+1)$ to larger-but-more-competitive market-j, they need to lower the markup rate, so that they need to earn higher revenue in market-j than in market-$(j+1)$ to keep them indiffierent between the two markets. In spite of these discontinuities, R_ψ, is globally strictly decreasing in ψ.

On the other hand, the markup rate schedule, μ_ψ, shown in Figure 8c, is not globally monotonic in ψ. It is continuously decreasing in ψ within each market. At the cutoff ψ_j $(j = 1, 2, \dots, J-1)$, however, it jumps upward. This is because $A_j < A_{j+1}$ so that switching from market-j to smaller-but-less-competitive market-$(j+1)$ allows ψ_j-firms to increase the markup rates from $\mu(\psi_j/A_j)$ to $\mu(\psi_j/A_{j+1})$. The markup rate, μ_ψ, thus exhibits a sawtooth pattern.

Likewise, the pass-through rate schedule, ρ_ψ, is not generally monotonic. Figure 8d shows the schedule under the strong A3. It is continuously increasing in ψ within each market. At the cutoff ψ_j $(j = 1, 2, \dots, J-1)$, however, it jumps downward. This is because $A_j < A_{j+1}$ so that switching from market-j to smaller-but-less-competitive market-$(j+1)$ allows ψ_j-firms to reduce the pass-through rates from $\rho(\psi_j/A_j)$ to $\rho(\psi_j/A_{j+1})$. The pass-through rate, ρ_ψ, thus exhibits a sawtooth pattern.

6.4. The Composition Effect: Average Markup and Pass-Through Rates Across Markets

---

[54] This demand-side mechanism is in sharp contrast to the supply-side mechanisms studied in the spatial economics literature. For example, what generates the positive assortative matching in Behrens, Duranton, and Robert-Nicoud (2014) and Gaubert (2018), both of which use CES, is the assumption on the firm technology that more productive firms are better at leveraging local agglomeration externalities in larger cities, similar to what Davis and Dingel (2019) assumed in the context of sorting of workers across the cities.

Under A2, more productive firms have higher markup rates than less productive firms if they face the same level of competitive pressures. However, more productive firms sort themselves into large and hence more competitive markets. This generates the sawtooth pattern in Figure 8c. Due to this composition effect, the average markup rates in large and hence more competitive markets be higher. Likewise, under A2 and the strong A3, more productive firms have lower pass-through rates than less productive firms if they face the same level of competitive pressures. However, more productive firms also sort themselves into large and hence more competitive markets, which generates the sawtooth pattern in Figure 8d. Due to this composition effect, the average pass-through rates in larger and hence more competitive markets might be higher, as demonstrated in Proposition 11a. Proposition 11b also demonstrates the possiblity that, due to an exogenous shock that causes all markets to become more competitive, the average markup rates to go up and the average pass-through rates to go down in all markets due to the shift in the composition. The proofs of these propositions are in Appendix C.6.

Proposition 11a: Suppose A2 and $G(\psi) = \left(\psi/\overline{\psi}\right)^{\kappa}$. There exists a sequence, $E_1 > E_2 > \cdots > E_J > 0$, such that, in equilibrium, any weighted generalized mean of $f\left(\psi/A_j\right)$ across firms operating at market-j are increasing (decreasing) in j even though $f(\cdot)$ is increasing (decreasing) and hence $f\left(\psi/A_j\right)$ is decreasing (increasing) in j.

Proposition 11a suggests an example with $G(\psi) = \left(\psi/\overline{\psi}\right)^{\kappa}$, in which the average markup rates are *higher* under A2 (and the average pass-through rates are *lower* under Strong A3) in larger markets. And recall that, as discussed in Section 2.5, the aggregate labor cost share is the reciprocal of the revenue-weighted harmonic mean of the markup rates and their employment-weighted arithmetic mean. Thus Proposition 11a also suggest that, under A2, the aggregate labor cost share can be smaller and the aggregate profit share can be higher in larger markets with more competitive pressures.

Proposition 11b: Suppose A2 and $G(\psi) = \left(\psi/\overline{\psi}\right)^{\kappa}$. Then, a change in F_e keeps

i) the ratios $a_j \equiv \psi_{j-1}/\psi_j$ and $b_j \equiv \psi_j/A_j$

and

ii) any weighted generalized mean of $f\left(\psi/A_j\right)$ across firms operating at market-j, for any weighting function $w\left(\psi/A_j\right)$,

unchanged for all $j = 1,2, \dots, J$.

Proposition 11b suggests that a decline in F_e under $G(\psi) = \left(\psi/\overline{\psi}\right)^{\kappa}$ offers a knife-edge case, where the average markup and pass-through rates of all markets remain unchanged, which also means that the aggregate labor cost and profit shares across all markets remain unchanged.

Propositions 11a and 11b thus suggest a caution when testing A2 and A3 by comparing the average markup & pass-through rates across space and time.

7. International/Interregional Trade with Differential Market Access

Up to now, all MC firms are assumed to be able to sell its product only in the market they enter. Of course, one could interpret the effect of a market size increase as the effect of international/interregional trade when different markets change from complete autarky to complete integration, where firms gain equal access to all markets, regardless of they produce. locations. However, what are the effects if firms have to pay additional trade costs for selling to remote markets and market integration takes the form of a trade cost reduction?

To address this question, imagine that this sector has two symmetric markets in two countries/regions.[55] Both markets are characterized by market size E, and by "labor" supplied at the price equal to one. This ensures that the same level of competitive pressures prevail in both markets, which is denoted by A. After paying the entry cost, F_e, and learning its marginal cost of production ψ_ω, firm ω can produce its product and sell it to both markets, but this requires the overhead cost $F > 0$ in each market. Selling it in its home market requires no additional cost, while selling it in the other market (the export market) requires additional iceberg cost, $\tau > 1$. That is, only $1/\tau$ fraction of the product shipped arrives to the export market. This implies that the marginal cost of exporting is $\tau\psi_\omega$, greater than the marginal cost of selling at home, ψ_ω.

Then, the equilibrium is characterized by the following three conditions.

Cutoff Rules: Firm ω sells to both markets iff $\psi_\omega \leq \psi_{xc} = \psi_c/\tau < \psi_c$, and sells only to the home market iff $\psi_{xc} = \psi_c/\tau < \psi_\omega \leq \psi_c$ where

$$F \equiv \pi\left(\frac{\psi_c}{A}\right)E \equiv \pi\left(\frac{\tau\psi_{xc}}{A}\right)E.$$

Thus, a fraction $G(\psi_{xc})$ of firms sell to both; a fraction $G(\psi_c) - G(\psi_{xc})$ sells only to their home market; a fraction $1 - G(\psi_c)$ exits.

Free-Entry Condition: The expected profit from both markets is equal to the entry cost.

$$F_e = \int_{\underline{\psi}}^{\psi_c}\left[\pi\left(\frac{\psi}{A}\right)E - F\right]dG(\psi) + \int_{\underline{\psi}}^{\psi_{xc}}\left[\pi\left(\frac{\tau\psi}{A}\right)E - F\right]dG(\psi),$$

where the first (second) term of the RHS is the expected profit from selling at home (abroad).

[55] It is straightforward to extend the analysis to many symmetric markets, but not to two or asymmetric markets, which require some functional form assumptions on the market share functions and the productivity distributions.

Adding-Up (Resource) Constraint: Let M denote the mass of the firms that pay the entry cost in each market. Then,

$$M\left[\int_{\underline{\psi}}^{\psi_c} r\left(\frac{\psi}{A}\right) dG(\psi) + \int_{\underline{\psi}}^{\psi_{xc}} r\left(\frac{\tau\psi}{A}\right) dG(\psi)\right] = 1.$$

from which $MG(\psi_c)$, the mass of the domestic firms, and $MG(\psi_{xc})$, that of the foreign firms operating in each market.

Comparative Statics: By combining the cutoff rules and free-entry condition,

$$\frac{F_e}{E} = \int_{\underline{\psi}}^{\psi_c}\left[\pi\left(\frac{\psi}{\psi_c}\pi^{-1}\left(\frac{F}{E}\right)\right) - \frac{F}{E}\right] dG(\psi) + \int_{\underline{\psi}}^{\psi_c/\tau}\left[\pi\left(\frac{\tau\psi}{\psi_c}\pi^{-1}\left(\frac{F}{E}\right)\right) - \frac{F}{E}\right] dG(\psi).$$

This equation pins down uniquely the equilibrium value of $\psi_c \equiv \tau\psi_{xc} \equiv \pi^{-1}(F/E)A$. In what follows, assume that F_e is not too large to ensure the interior solution, $0 < G(\psi_c) < 1$. Then, the RHS of this condition is strictly increasing in $\psi_c \in \left(\underline{\psi}, \bar{\psi}\right)$, so that it is easy to verify:

Proposition 12 (Globalization): A decline in τ, the iceberg trade cost, causes a decline in $\psi_c = \tau\psi_{xc} = \pi^{-1}(F/E)A$ and an increase in $\psi_c/\tau = \psi_{xc} = \pi^{-1}(F/E)(A/\tau)$. Hence,

- The share of active domestic firms, $G(\psi_c)$, falls, while the share of foreign firms, $G(\psi_{xc})$ rises in both markets;
- The revenue and the profit in the domestic market, $r(\psi_\omega/A)E$ and $\pi(\psi_\omega/A)E$, decline, while the revenue and the profit in the export market, $r(\tau\psi_\omega/A)E$ and $\pi(\tau\psi_\omega/A)E$, rise;

from which the revenue and profit shares of the foreign firms rise in both markets. Moreover,

- The markup rates of the domestic firms, $\mu(\psi_\omega/A)$, declines, while those of the exporting foreign firms, $\mu(\tau\psi_\omega/A)$ rises **under the 2nd law;**
- The pass-through rates of the domestic firms, $\rho(\psi_\omega/A)$, rises, while those of the exporting foreign firms, $\rho(\tau\psi_\omega/A)$ declines **under the strong 3rd law.**

8. Concluding Remarks

In this paper, we develop a model of monopolistically competitive sector with heterogeneous firms and free entry under H.S.A., which contains CES and translog as special cases. H.S.A. is tractable due to its homotheticity and to its single aggregator that serves as a sufficient statistic for competitive pressures. It is also flexible enough to allow for the choke price, the 2nd law of demand, and what we call the 3rd law of demand. The single aggregator property makes it possible to prove the existence and uniqueness of the free-entry equilibrium and to conduct comparative static analysis, often using just simple diagrams. Furthermore, because the single aggregator enters all firm-specific variables proportionately with the firm-specific marginal cost, and hence acts as a magnifier of firm heterogeneity, we are able to characterize, by taking advantage of log-supermodularity, how a change in competitive pressures, whether due to a change in the entry cost, market size, the overhead cost, or the iceberg trade cost, affects heterogeneous firms differently under the 2nd and the 3rd laws of demand and thereby causing reallocation across firms, and hence selection of firms, and sorting of firms across different markets. Furthermore, we are able to show that, due to such a composition effect, the average markup (pass-through) rate may move in the opposite direction of the firm-level markup (pass-through) rate, which also means that a higher average markup rate and a higher aggregate profit share may be *due to* (not in spite of) more competitive pressures.

It is our hope that this model of a monopolistic competitive sector with heterogeneous firms and free enry under H.S.A. proves to be a useful building block in general equilibrium models, thereby opening up for the possiblity of addressing a wide range of issues, where markup rate and pass-through rate heterogeneity would play central roles.

References:

Amiti, M., O. Itskhoki, and J. Konings (2014), "Importers, Exporters, and Exchange Rate Disconnect. American Economic Review, 104(7), 1942–1978.

Amiti, M., O. Itskhoki, and J. Konings (2019), "International Shocks, Variable Markups, and Domestic Prices," The Review of Economic Studies. 86, 2356-2402.

Arkolakis, C., A. Costinot, D. Donaldson, and Rodríguez-Clare (2019), "The Elusive Pro-Competitive Effects of Trade, The Review of Economic Studies, 86 (1), 46–80.

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

Baldwin, RE., and T. Okubo (2006). "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting." Journal of Economic Geography 6 (3): 323–46.

Bagnoli, M., and T. Bergstrom (2005), "Log-concave probability and its applications," Economic Theory, 26, 445-469.

Baqaee, D., E. Farhi, and K. Sangani (2024), "The Darwinian Returns to Scale," Review of Economic Studies, 91(3): 1373-1405.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014), "Productive Cities: Sorting, Selection, and Agglomeration." Journal of Political Economy 122 (3): 507–53.

Behrens, K., G. Mion, Y. Murata, and J. Südekum (2020), "Quantifying the gap between equilibrium and optimum under monopolistic competition", Quarterly Journal of Economics, 135(4), 2299-2360.

Behrens, K. and Y. Murata, (2007), "General Equilibrium Models of Monopolistic Competition: A New Approach," *Journal of Economic Theory* 136, 776-787.

Berman, N., P. Martin, and T. Mayer (2012), "How Do Different Exporters React to Exchange Rate Changes? The Quarterly Journal of Economics, 127(1), 437–492.

Bernard, AB., SJ. Redding and PK. Schott (2007), "Comparative Advantage and Heterogeneous Firms," Review of Economic Studies, 74, 31-66.

Burstein, A., and G. Gopinath, G. (2014), "International Prices and Exchange Rates," in G. Gopinath, E. Helpman, and K. Rogoff (eds.), Handbook of International Economics , vol.4, 391-451, Amsterdam: Elsevier.

Campbell, JR., and HA. Hopenhayn, "Market size matters," The Journal of Industrial Economics, March 2005, Vol.53, 1-25.

Costinot, A. (2009): An Elementary Theory of Comparative Advantage, Econometrica, 77, 1165-1192.

Costinot, A., and J. Vogel (2010): Matching and Inequality in the World Economy, Journal of Political Economy, 118, 747-786.

Costinot, A., and J. Vogel (2015): Beyond Ricardo: Assignment Models in International Trade, Annual Review of Economics, 7, 31-62.

Davis, D., and J. Dingel (2019). "A Spatial Knowledge Economy," American Economic Review. 109(1): 153-170.

Davis, D., and J. Dingel (2020), The Comparative Advantage of Cities, Journal of International Economics, 123, 103291.

De Loecker, J., and P.K. Goldberg (2014), "Firm Performance in a Global Market," Annual Review of Economics, 6(1), 201–227.

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

Deaton A. (1974). "A Reconsideration of the Empirical Implications of Additive Preferences." Economic Journal, 84:338-48.

Dhingra, S., and J. Morrow (2019), "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity," Journal of Political Economy, vol.127, no.1, 196-232.

Dixit, AK., JE. Stiglitz, "Monopolistic Competition and Optimal Product Diversity," American Economic Review, June 1977, 67(3), 297-308.

Edmond, C., V. Midrigan, and D. Y. Xu, (2023), "How Costly are Markups?" Journal of Political Economy, 131(7): 1676-1730.

Feenstra, RC., "A Homothetic utility function for monopolistic competition models, without constant price elasticity," Economics Letters 78 (2003), 79-86.

Feenstra, RC. and DE. Weinstein, "Globalization, Markups, and US Welfare," Journal of Political Economy, 2017, 125(4), 1040-1074.

Fujiwara, I., and K. Matsuyama (2022), "Competition and the Phillips Curve," Keio and Northwestern.

Gaubert, C. (2018), "Firm Sorting and Agglomeration," American Economic Review, 108(11): 3117-3153.

Grossman, GM., E. Helpman, and H. Lhuillier (2023), "Supply Chain Resilience: Should Policy Promote International Diversification or Reshoring," Journal of Political Economy, 131(12):3462-3496.

Head, K., T. Mayer, M. Theonig, "Welfare and Trade without Pareto," American Economic Review Papers & Proceedings, May 2014, 104(5):310-316.

Hurwicz L., and H. Uzawa, (1971), "On the integrability of demand functions". In Chipman et al. *Preferences, utility, and demand: A Minnesota symposium*. New York: Harcourt, Brace, Jovanovich. Ch.6, 114–148.

Kasahara, H., and Y. Sugita, (2020), "Nonparametric Identification of Production Function, Total Factor Productivity, and Markup from Revenue Data," UBC and Hitotsubashi.

Kimball, M. (1995), "The Quantitative Analytics of the Basic Neomonetarist Model," Journal of Money, Credit and Banking, 27 (4), 1241–77.

Kokovin, S., A. Ozhegova, S. Sharapudinov, A. Tarasov, and P. Ushchev (2024). "A Theory of Monopolistic Competition with Horizontally Heterogeneous Consumers," AEJ Micro.

Krugman, PR. (1979). Increasing returns, monopolistic competition, and international trade, Journal of International Economics 9 (4), 469{479.

Latzer, H., K. Matsuyama, and M. Parenti, (2019) "Reconsidering the Market Size Effect in Innovation and Growth," CEPR DP #14250.

Matsuyama, K. (2023), "Non-CES Aggregators: A Guided Tour," Annual Review of Economics, vol.15.

Matsuyama, K. (2025), "Homothetic non-CES Demand Systems with Applications to Monopolistic Competition, Annual Review of Economics, vol. 17, forthcoming.

Matsuyama, K., and P. Ushchev (2017), "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems," CEPR DP #12210.

Matsuyama, K., and P. Ushchev (2020a), "When Does Procompetitive Entry Imply Excessive Entry?" CEPR DP #14991.

Matsuyama, K., and P. Ushchev (2020b), "Constant Pass-Through," CEPR DP #15475.

Matsuyama, K., and P. Ushchev (2022), "Destabilizing Effects of Market Size in the Dynamics of Innovation," Journal of Economic Theory, 200, 105415.

Matsuyama, K., and P. Ushchev (2023a), "Selection and Sorting of Heterogeneous Firms under Competitive Pressures," Version 2023.07.08, http://faculty.wcas.northwestern.edu/~kmatsu/

Matsuyama, K., and P. Ushchev (2023b), "Love-for-Variety," CEPR DP #18184.

Mayer, T., M.J. Melitz, G.I.P. Ottaviano (2021), "Product Mix and Firm Productivity Responses to Trade Competition," Review of Economics and Statistics, 103 (5): 874-891.

Melitz, M.J. (2003), "The Impact of Trade on Intra-Industry Reallocations, and Aggregate Industry Productivity," Econometrica, 71, 6, 1695–1725.

Melitz, M.J., (2018), "Competitive Effects of Trade: Theory and Measurement," Review of World Economics, 154 (1), 1–13.

Melitz, M.J., and G.I.P. Ottaviano (2008), "Market size, trade, and productivity," The Review of Economic Studies, 75 (1), 295–316.

Melitz, M.J., and S. J. Redding (2014), "Heterogeneous Firms and Trade." In *Handbook of International Economics*. Vol. 4, edited by Elhanan Helpman, Kenneth Rogoff, and Gita Gopinath, 1–54. Amsterdam: Elsevier North Holland.

Mrázová, M., and J. P. Neary (2017), "Not So Demanding: Demand Structure and Firm Behavior," American Economic Review, 107(12): 3835–3874

Mrázová, M., and J. P. Neary (2019), "Selection Effects with Heterogeneous Firms," Journal of the European Economic Association 17(4):1294–1334.

Nocke, V. (2006), "A Gap for Me: Entrepreneurs and Entry." Journal of the European Economic Association 4 (5): 929–56.

Ottaviano, G.I.P., T.Tabuchi, and J.-F. Thisse, (2002). "Agglomeration and trade revisited," International Economic Review, 43, 409–435.

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

Sampson, T., (2016), "Assignment Reversals: Trade, Skill Allocation and Wage Inequaity," Journal of Economic Theory, 163, 365-409.

Samuelson, P.A.,(1950), "The Problem of Integrability in Utility Theory." Economia 17 (68): 355–385.

Thisse, J.-F., and P. Ushchev (2018), Monopolistic Competition without Apology, Ch 5 in *Handbook of Game Theory and Industrial Organization*, Vol. I, ed by L. Cochon and M. Marini, Edgar Publication.

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-Francois Thisse, "Monopolistic Competition: Beyond the Constant Elasticity of Substitution," Econometrica, November 2012, Vol.80, No.6, 2765-2784.

Figures For

Selection and Sorting of Heterogeneous Firms through Competitive Pressures

Kiminori Matsuyama and Philip Ushchev

Date: 2025-03-27, Time: 4:22 PM

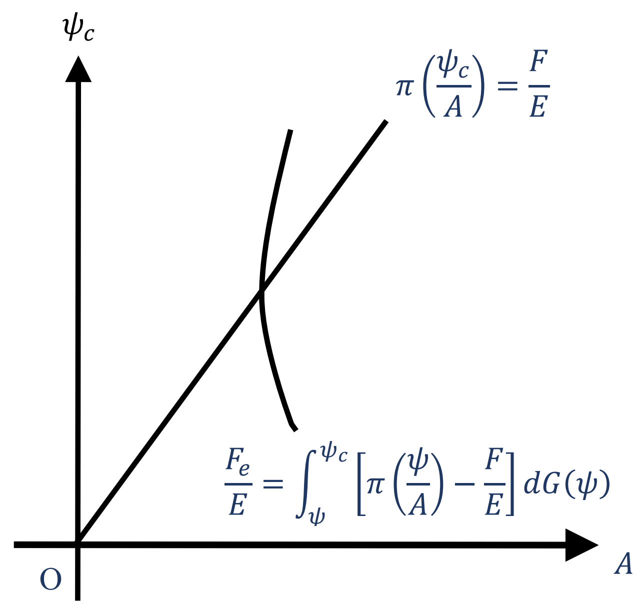
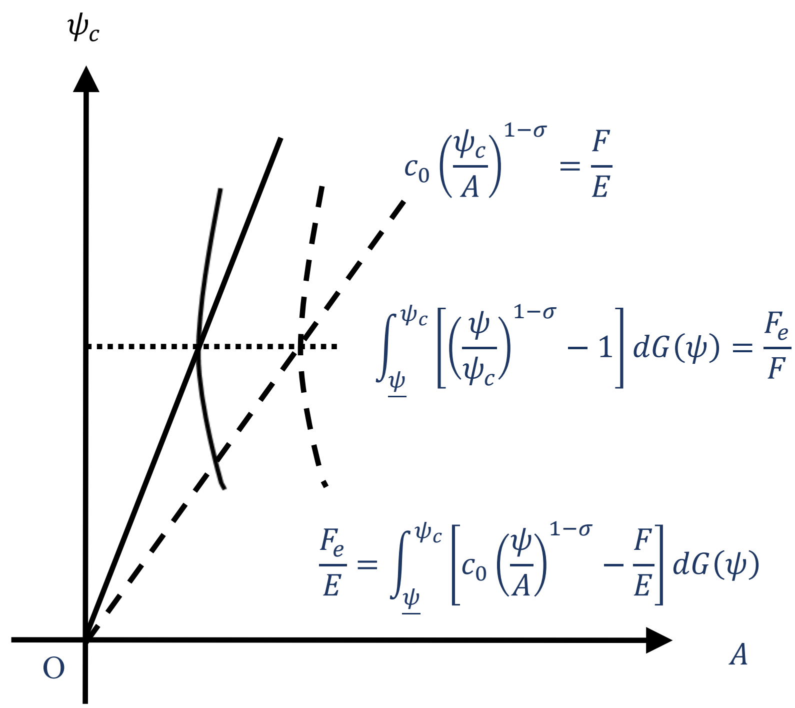
Figure 1: **Existence and Uniqueness** Cutoff Rule and Free Entry Condition jointly determine ψ_c and $A = A(\mathbf{p})$ uniquely.	
Figure 2: CES Benchmark	

Figure 3: Cross-Sectional Implications of A2 and A3

Figure 3a: Log-Submodular Profit under A2

Log-profit always downward-sloping and strictly concave under A2. A lower A causes a parallel leftward shift; A higher E causes a parallel upward shift.

[Under the weak A3, the graph of log-revenue has the same properties.]

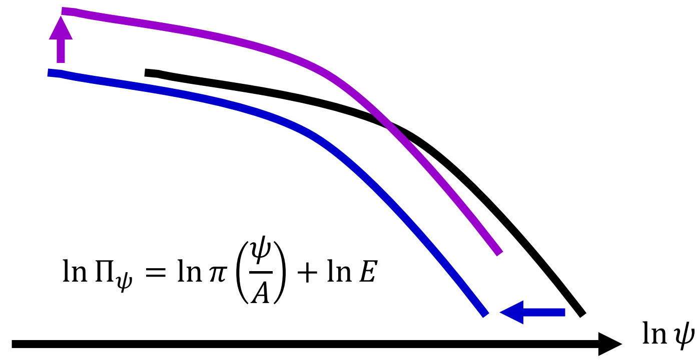


Figure 3b: A2 & A3 and Log-Supermodular Markup Rate

Downward-sloping under A2 and strict(weak)ly convex under strong(weak) A3. A lower A (more competitive pressures) causes a parallel leftward shift.

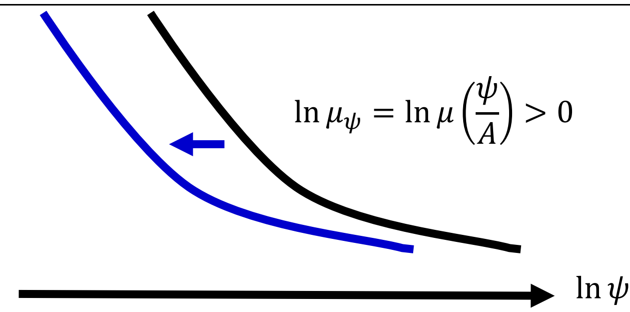


Figure 3c: A2 & the weak A3 and Log-Submodular Employment

Hump-shaped and strictly concave under A2 and the weak A3. A lower A (more competitive pressures) causes a parallel leftward shift; A higher E (larger market size) causes a parallel upward shift.

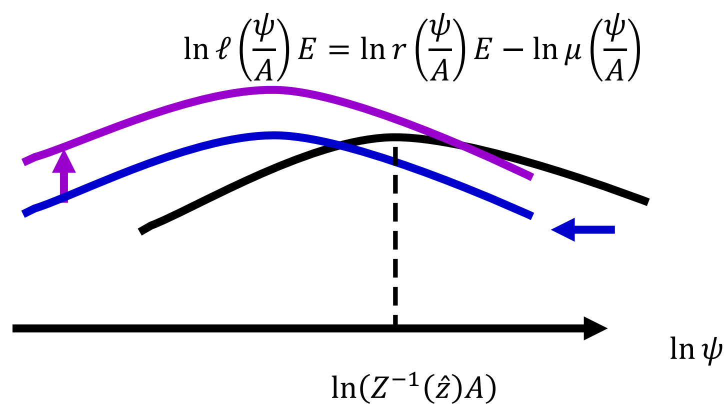


Figure 3d:

A2 and strong A3 and Pass-Through Rate

Under A2, $\ln \rho(\psi/A) < 0$;

Under strong A3, strictly increasing;

Under A2 and strong A3, globally strictly convex for a sufficiently small $\bar{z}$:

A lower A (more competitive pressures) causes a parallel leftward shift.

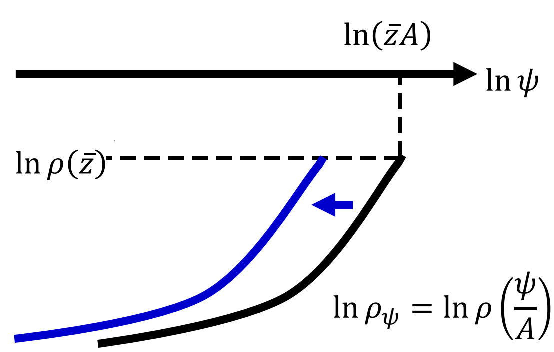


Figure 4: Comparative Statics on ψ_c and A

Figure 4a:

Effects of $F_e\downarrow$

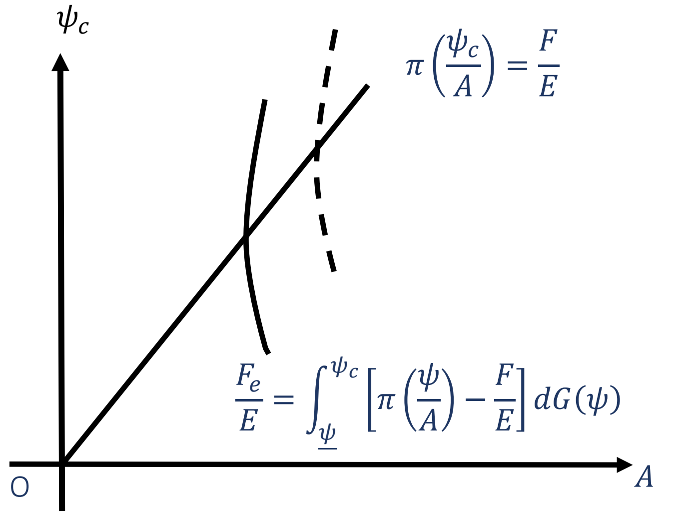


Figure 4b:

Effects of $E\uparrow$

under A2 $\Leftrightarrow \mathbb{E}_\pi(\sigma)<\sigma(\psi_c/A)$

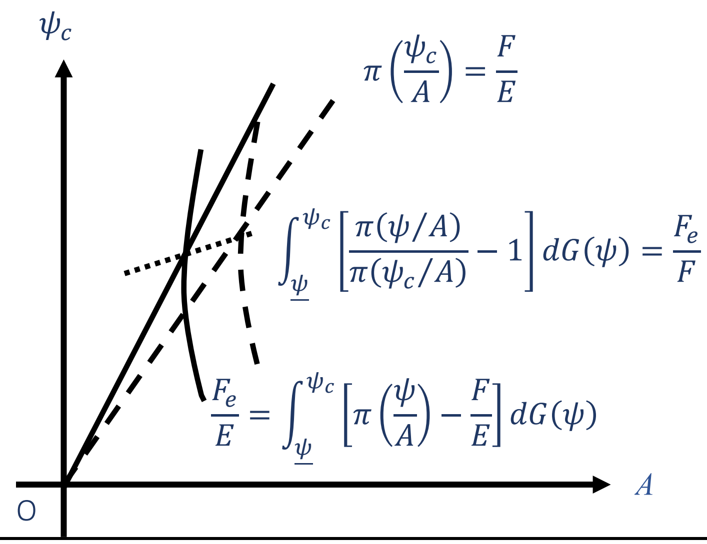


Figure 4c:

Effects of $F\downarrow$

under $\mathbb{E}_1(\ell)<\ell(\psi_c/A)$

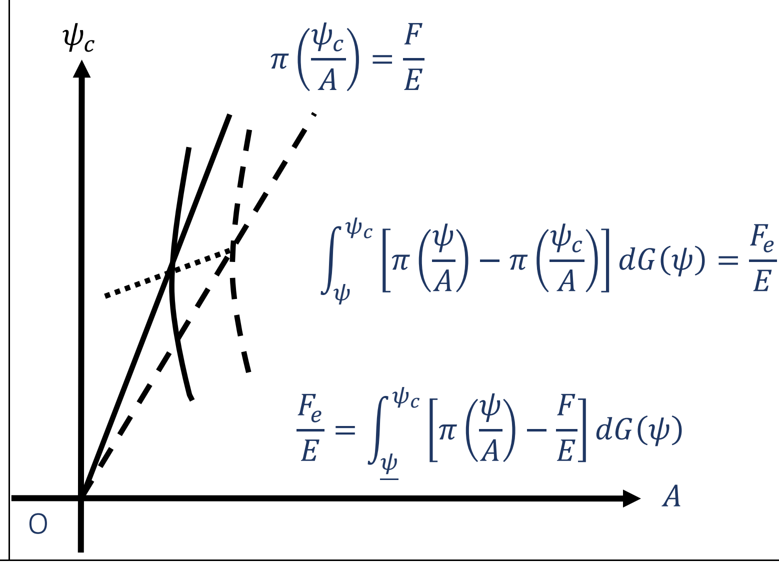


Figure 7:

Logic behind Sorting

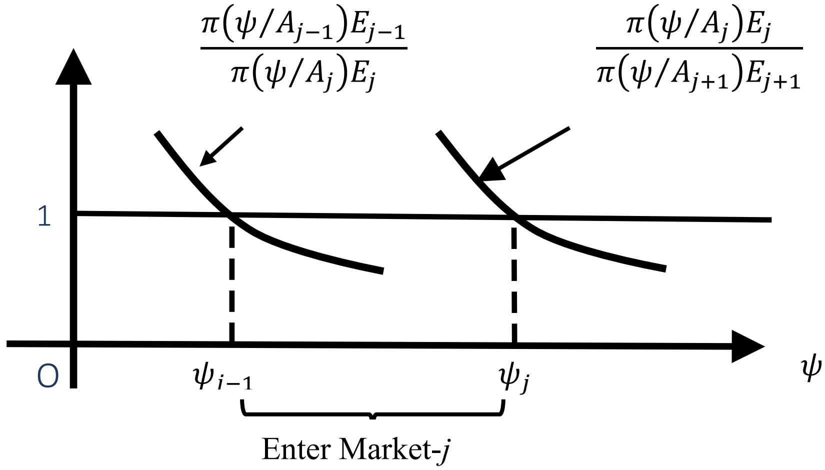
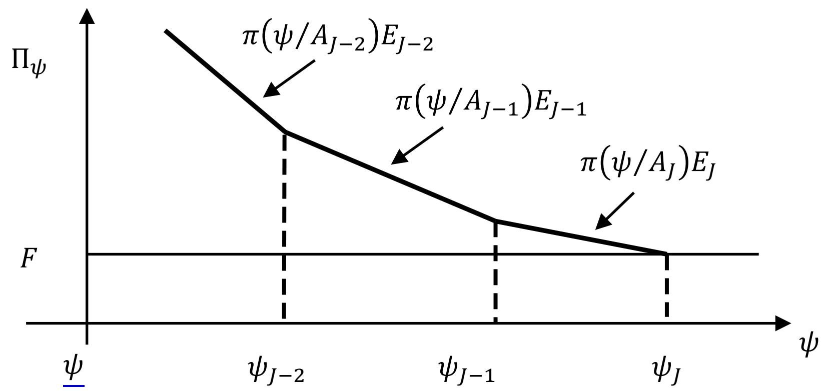
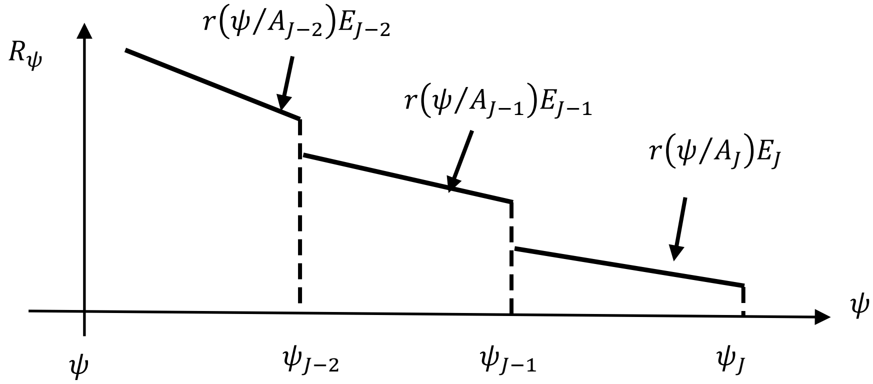
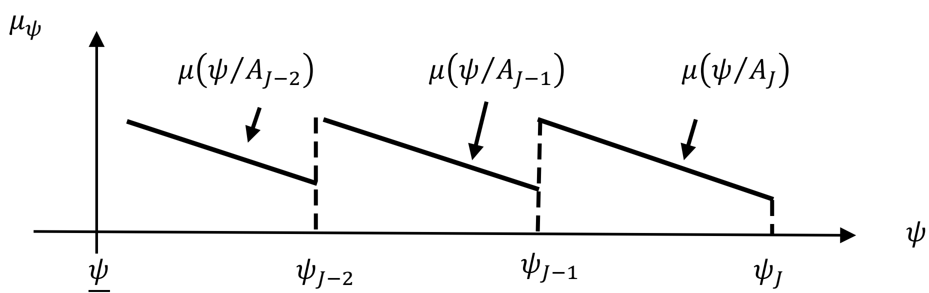
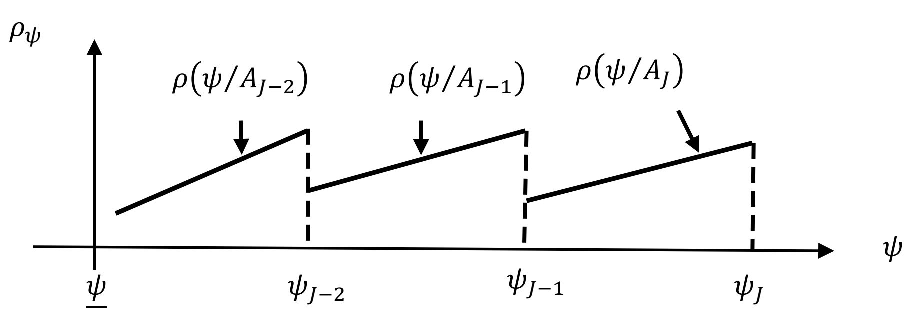


Figure 8: Profit, Revenue, Markup, and Pass-through Schedules across Firms and Markets

Figure 8a: **Profits: Under A2**	
Figure 8b: **Revenues under A2**	
Figure 8c: **Markup rates under A2**	
Figure 8d: **Pass-through rates under A2 and the strong A3**	

Online Appendices for

Selection and Sorting of Heterogeneous Firms through Competitive Pressures

Kiminori Matsuyama and Philip Ushchev

Date: 2025-03-27, Time: 4:22 PM

Appendix A: Firm type distributions and their elasticities

Let the distribution of the marginal cost, ψ, be given by its cdf, $G(\psi)$, with the support, $\left(\underline{\psi}, \overline{\psi}\right) \subseteq (0, \infty)$, and hence that of productivity, $\varphi = 1/\psi$, be given by its cdf, $F(\varphi) = 1 - G(1/\varphi)$, with the support, $\left(\underline{\varphi}, \overline{\varphi}\right) = \left(1/\overline{\psi}, 1/\underline{\psi}\right) \subseteq (0, \infty)$. We assume that these cdfs are thrice continuously differentiable, C^3, and hence that their pdfs satisfy, $G'(\psi) = g(\psi) > 0$ on $\left(\underline{\psi}, \overline{\psi}\right)$ and $F'(\varphi) = f(\varphi) > 0$ on $\left(\underline{\varphi}, \overline{\varphi}\right)$ and are twice continuously differentiable, C^2, so that $\mathcal{E}_G(\psi) \equiv \psi g(\psi)/G(\psi) \in C^2$, $\mathcal{E}_g(\psi) \equiv \psi g'(\psi)/g(\psi) \in C^1$ and $\mathcal{E}_F(\varphi) \equiv \varphi f(\varphi)/F(\varphi) \in C^2$, $\mathcal{E}_f(\varphi) \equiv \varphi f'(\varphi)/f(\varphi) \in C^1$ It is straightforward to show that:

$$\varphi f(\varphi) = \psi g(\psi);$$

$$\mathcal{E}_f(\varphi) + \mathcal{E}_g(\psi) = -2;$$

and

$$\varphi \mathcal{E}_f'(\varphi) = \psi \mathcal{E}_g'(\psi).$$

We also assume that the mean productivity is finite:

$$\int_{\underline{\varphi}}^{\overline{\varphi}} \varphi f(\varphi) d\varphi = \int_{\underline{\psi}}^{\overline{\psi}} \psi^{-1} g(\psi) d\psi < \infty.$$

This is guaranteed if $\underline{\psi} > 0 \Leftrightarrow \overline{\varphi} < \infty$. If $\underline{\psi} = 0 \Leftrightarrow \overline{\varphi} = \infty$, a sufficient condition for the finite mean productivity is given by:

$$-\lim_{\psi \to 0} \mathcal{E}_g(\psi) = \lim_{\varphi \to \infty} \mathcal{E}_f(\varphi) + 2 < 0.$$

To see this, note that $\lim_{\varphi \to \infty} \mathcal{E}_f(\varphi) + 2 < 0 \Leftrightarrow \lim_{\varphi \to \infty} \mathcal{E}_f(\varphi) + 1 < -1$ implies that $\varphi f(\varphi)$ decreases faster than $1/\varphi$ as $\varphi \to \infty$, $\int_{\underline{\varphi}}^{\infty} \varphi f(\varphi) d\varphi < \infty$.[56]

Lemma 1:

[56]Equivalently, $-\lim_{\psi \to 0} \mathcal{E}_g(\psi) < 0 \Leftrightarrow \lim_{\psi \to 0} \mathcal{E}_g(\psi) - 1 > -1$, implies that $\psi^{-1} g(\psi)$ increases slower than ψ^{-1} as $\psi \searrow 0$, hence $\int_0^{\overline{\psi}} \psi^{-1} g(\psi) d\psi < \infty$. Even though this condition for the finite mean productivity is sufficient but not necessary, it is close to being necessary in the sense that the mean productivity is infinite if $-\lim_{\psi \to 0} \mathcal{E}_g(\psi) = \lim_{\varphi \to \infty} \mathcal{E}_f(\varphi) + 2 > 0$. The case of $-\lim_{\psi \to 0} \mathcal{E}_g(\psi) = \lim_{\varphi \to \infty} \mathcal{E}_f(\varphi) + 2 = 0$ would require case-by-case scrutiny.

$$\mathcal{E}_g'(\psi) < 0, \forall \psi \in \left(\underline{\psi}, \overline{\psi}\right) \Longrightarrow \mathcal{E}_G'(\psi) < 0, \forall \psi \in \left(\underline{\psi}, \overline{\psi}\right).$$

Furthermore, if $\underline{\psi} = 0$ and $\lim_{\psi \to 0} \psi g(\psi) = 0$,

$$\mathcal{E}_g'(\psi) \geq 0, \forall \psi \in \left(0, \overline{\psi}\right) \Longrightarrow \mathcal{E}_G'(\psi) \geq 0, \forall \psi \in \left(0, \overline{\psi}\right).$$

Proof:[57] $\mathcal{E}_g'(\psi) \gtreqless 0, \forall \psi \in \left(\underline{\psi}, \overline{\psi}\right)$ implies

$$\left[\mathcal{E}_g(\psi) + 1\right]G(\psi) = \left[\mathcal{E}_g(\psi) + 1\right]\int_{\underline{\psi}}^{\psi} g(\xi)d\xi \gtreqless \int_{\underline{\psi}}^{\psi}\left[\mathcal{E}_g(\xi) + 1\right]g(\xi)d\xi$$

$$= \int_{\underline{\psi}}^{\psi}[\xi g'(\xi) + g(\xi)]d\xi = \int_{\underline{\psi}}^{\psi} d[\xi g(\xi)] = \psi g(\psi) - \lim_{\psi \to \underline{\psi}} \psi g(\psi),$$

which in turn implies

$$\mathcal{E}_G'(\psi) = \frac{d}{d\psi}\left[\frac{\psi g(\psi)}{G(\psi)}\right] = \frac{[\psi g'(\psi) + g(\psi)]G(\psi) - \psi[g(\psi)]^2}{[G(\psi)]^2}$$

$$= \frac{g(\psi)}{[G(\psi)]^2}\left\{\left[\mathcal{E}_g(\psi) + 1\right]G(\psi) - \psi g(\psi)\right\} \gtreqless -\frac{g(\psi)}{[G(\psi)]^2}\left[\lim_{\psi \to \underline{\psi}} \psi g(\psi)\right].$$

Hence, the first part always holds, while the second part holds because $\lim_{\psi \to 0} \psi g(\psi) = 0$.

This completes the proof. ■

The following lemma states how a change in ψ_c shifts the distribution of $\xi \equiv \psi/\psi_c$, the marginal cost relative to the cutoff marginal cost, ψ_c, among surviving firms. It shows that, if $\mathcal{E}_g(\cdot)$ is increasing (decreasing), an increase in ψ_c causes a shift to the right (left) in the sense of the monotone likelihood ratio ordering; and that, if $\mathcal{E}_G(\cdot)$ is increasing (decreasing), an increase in ψ_c causes a shift to the right (left) in the sense of the first-order stochastic dominance.

Lemma 2: Define $\xi \equiv \psi/\psi_c \in \left(\underline{\xi}, 1\right)$, where $\underline{\xi} \equiv \underline{\psi}/\psi_c$. Consider a cdf,

[57] For the second part of Lemma 1, we consider only the case of $\underline{\psi} = 0$, since $\underline{\psi} > 0$ and $\lim_{\psi \to \underline{\psi}} \psi g(\psi) = 0$ would imply $\lim_{\psi \to \underline{\psi}} g(\psi) = 0$, so that $\lim_{\psi \to \underline{\psi}} \mathcal{E}_g(\psi) = \infty$, hence $\mathcal{E}_g'(\psi) < 0$ for ψ close to $\underline{\psi} > 0$. Thus, it would be impossible to satisfy $\mathcal{E}_g'(\psi) \geq 0, \forall \psi \in \left(\underline{\psi}, \overline{\psi}\right)$. It is also worth noting that the second part would fail if $\underline{\psi} > 0$ and $\lim_{\psi \to \underline{\psi}} \psi g(\psi) > 0$. [An example is a truncated power, for which $\mathcal{E}_g'(\cdot) = 0$ but $\mathcal{E}_G'(\cdot) \neq 0$.] Lemma 1 can also be obtained as a corollary of Theorem 1 and Theorem 2 of Bagnoli and Bergstrom (2005) by noting that $\mathcal{E}_G'(\cdot) < (>) 0$ if and only if the cdf of $\theta \equiv \ln \psi$, $G\left(e^{\theta}\right)$, is log-concave (log-convex) and that $\mathcal{E}_g'(\cdot) < (>) 0$ if and only if the density of θ, $e^{\theta} g\left(e^{\theta}\right)$, is log-concave (log-convex).

$$\tilde{G}(\xi;\psi_c) \equiv \frac{G(\psi_c\xi)}{G(\psi_c)},$$

and its density function,

$$\tilde{g}(\xi;\psi_c) \equiv \frac{d\tilde{G}(\xi;\psi_c)}{d\xi} = \frac{\psi_c g(\psi_c\xi)}{G(\psi_c)},$$

whose support is $\left(\underline{\xi},1\right)$ with $\tilde{G}\left(\underline{\xi};\psi_c\right) = 0$ and $\tilde{G}(1;\psi_c) = 1$. Then,

$$\mathcal{E}_g'(\xi) \gtreqless 0\,, \forall\xi \in \left(\underline{\xi},1\right) \Rightarrow \frac{\partial^2 \ln \tilde{g}(\xi;\psi_c)}{\partial\xi\partial\psi_c} \gtreqless 0, \forall\xi \in (\underline{\xi},1)$$

and

$$\mathcal{E}_G'(\xi) \gtreqless 0\,, \forall\xi \in \left(\underline{\xi},1\right) \Rightarrow \frac{\partial\tilde{G}(\xi;\psi_c)}{\partial\psi_c} \lesseqgtr 0, \forall\xi \in \left(\underline{\xi},1\right).$$

Proof: The first statement follows from

$$\frac{\partial^2 \ln \tilde{g}(\xi;\psi_c)}{\partial\xi\partial\psi_c} = \frac{\partial^2 \ln g(\psi_c\xi)}{\partial\xi\partial\psi_c} = \mathcal{E}_g'(\psi_c\xi) \gtreqless 0, \qquad \forall\xi \in \left(\underline{\xi},1\right).$$

The second statement follows from

$$\frac{\partial \ln \tilde{G}(\xi;\psi_c)}{\partial \ln \psi_c} = \frac{\partial \ln[G(\psi_c\xi)/G(\psi_c)]}{\partial \ln \psi_c} = \mathcal{E}_G(\psi_c\xi) - \mathcal{E}_G(\psi_c) \lesseqgtr 0, \forall\xi \in (\underline{\xi},1),$$

if $\mathcal{E}_G'(\xi) \gtreqless 0$. This completes the proof. ■

The signs of $\mathcal{E}_g'(\cdot)$ and of $\mathcal{E}_G'(\cdot)$ play critical roles for some of the comparative statics results. Thus, we now list some parametric families of distributions (widely used in the literature), for which the sign of $\mathcal{E}_g'(\cdot)$ never changes over the support, which also means, from Lemma 1, that the sign of $\mathcal{E}_G'(\cdot)$ never changes over the support, either.

Example 1: Pareto (or power) distribution. The cdfs are given by

$$F(\varphi) = 1 - \left(\varphi/\underline{\varphi}\right)^{-\kappa} \Leftrightarrow G(\psi) = \left(\psi/\overline{\psi}\right)^{\kappa};$$

for $\varphi > \underline{\varphi} > 0 \iff 0 < \psi < \overline{\psi} < \infty$. The pdfs satisfy:

$$\varphi f(\varphi) = \kappa\left(\varphi/\underline{\varphi}\right)^{-\kappa} = \kappa\left(\psi/\overline{\psi}\right)^{\kappa} = \psi g(\psi)$$

Hence, $\mathcal{E}_f(\varphi) = -\kappa - 1$ and $\mathcal{E}_g(\psi) = \kappa - 1$, so that $\mathcal{E}_f'(\varphi) = \mathcal{E}_g'(\psi) = 0$. The condition for the finite mean productivity is given by $\kappa > 1$.

Example 2: Generalized Pareto (Power) distribution. The generalized Pareto (Power) family nests Pareto (Power) as a special case and allows all the three possibilities for $sgn\{\mathcal{E}_f'(\cdot)\} = sgn\{\mathcal{E}_g'(\cdot)\}$ to depend on the parameter values. The cdfs are given by

$$F(\varphi) = 1 - \left(1 + \frac{\varphi - \underline{\varphi}}{\lambda}\right)^{-\kappa}, \qquad \varphi > \underline{\varphi} > 0, \qquad \lambda > 0.$$

$$G(\psi) = \left(1 + \frac{1/\psi - 1/\overline{\psi}}{\lambda}\right)^{-\kappa}, \qquad 0 < \psi < \overline{\psi} < \infty, \qquad \lambda > 0.$$

Hence, the pdfs satisfy:

$$\varphi f(\varphi) = \frac{\varphi\kappa}{\lambda}\left(1 + \frac{\varphi - \underline{\varphi}}{\lambda}\right)^{-\kappa-1} = \frac{\kappa}{\psi\lambda}\left(1 + \frac{1/\psi - 1/\overline{\psi}}{\lambda}\right)^{-\kappa-1} = \psi g(\psi)$$

from which

$$\mathcal{E}_f(\varphi) = -(1+\kappa)\left(\frac{\varphi}{\lambda - \underline{\varphi} + \varphi}\right) = -(1+\kappa)\left(\frac{1/\psi}{\lambda - 1/\overline{\psi} + 1/\psi}\right) = -\mathcal{E}_g(\psi) - 2.$$

Clearly, the standard Pareto (Power) distribution is a special case with $\lambda = \underline{\varphi} = 1/\overline{\psi}$. More generally, one can readily verify that:

$$\psi\mathcal{E}_g'(\psi) = \varphi\mathcal{E}_f'(\varphi) = -(1+\kappa)\frac{\varphi\left(\lambda - \underline{\varphi}\right)}{\left(\lambda - \underline{\varphi} + \varphi\right)^2} \gtreqless 0 \iff \lambda \lesseqgtr \underline{\varphi} = 1/\overline{\psi}.$$

Example 3: Lognormal distribution. Since $\ln\varphi = -\ln\psi$, productivity is distributed lognormally if and only if the marginal cost is distributed lognormally. In this case, the support is $(0,\infty)$. For all $\varphi > 0$ and for all $\psi > 0$, the pdfs can be represented by

$$f(\varphi) = \frac{1}{\varphi\tilde{\sigma}\sqrt{2\pi}}\exp\left\{-\frac{(\log\varphi - \mu)^2}{2\tilde{\sigma}^2}\right\},$$

$$g(\psi) = \frac{1}{\psi\tilde{\sigma}\sqrt{2\pi}}\exp\left\{-\frac{(\log\psi + \mu)^2}{2\tilde{\sigma}^2}\right\},$$

where $\mu \in \mathbb{R}$ and $\tilde{\sigma} > 0$. The mean productivity is:

$$\int_0^\infty \varphi f(\varphi)d\varphi = \int_0^\infty \psi^{-1}g(\psi)d\psi = \exp\left\{\mu + \frac{\tilde{\sigma}^2}{2}\right\} < \infty.$$

The elasticities of the pdfs are strictly decreasing, because

$$\mathcal{E}_f(\varphi) = \frac{\mu - \log\varphi}{\tilde{\sigma}^2} - 1 = \frac{\mu + \log\psi}{\tilde{\sigma}^2} - 1 = -\mathcal{E}_g(\psi) - 2$$

$$\Rightarrow \varphi\mathcal{E}_f'(\varphi) = \psi\mathcal{E}_g'(\psi) = -\frac{1}{\tilde{\sigma}^2} < 0.$$

Hence, from Lemma 1, the elasticities of the cdfs are also strictly decreasing.

Example 4: Fréchet and Weibull distributions. The parametric families of Fréchet and Weibull distributions both belong to the class of extreme-value distributions.[58] When the distribution of φ is Fréchet (respectively, Weibull) if and only if that of $\psi = 1/\varphi$ is Weibull (respectively, Fréchet). Therefore, we consider the case of φ being Fréchet and omit the case of φ being Weibull.

For all $\varphi > 0$ and for all $\psi > 0$, the cdf of the Fréchet productivity distribution F and the corresponding Weibull cost distribution G are given, respectively, by

$$F(\varphi) = \exp\{-\varphi^{-\alpha}\}, \qquad G(\psi) = 1 - \exp\{-\psi^{\alpha}\},$$

where $\alpha > 0$. The pdfs are given by

$$f(\varphi) = \alpha\varphi^{-(1+\alpha)}\exp\{-\varphi^{-\alpha}\}, \qquad g(\psi) = \alpha\psi^{\alpha-1}\exp\{-\psi^{\alpha}\}.$$

Hence,

$$\mathcal{E}_f(\varphi) = -(1+\alpha) + \alpha\varphi^{-\alpha},\ \ \mathcal{E}_g(\psi) = \alpha - 1 - \alpha\psi^{\alpha}$$

$$\Rightarrow \varphi\mathcal{E}_f'(\varphi) = -\alpha^2\varphi^{-\alpha} = -\alpha^2\psi^{\alpha} = \psi\mathcal{E}_g'(\psi) < 0,$$

so that the elasticities of the pdfs are strictly decreasing, and so are the elasticities of the cdfs from Lemma 1. The mean productivity is finite if and only if

$$-\lim_{\psi\to 0}\mathcal{E}_g(\psi) = \lim_{\varphi\to\infty}\mathcal{E}_f(\varphi) + 2 = \alpha - 1 < 0 \Leftrightarrow \alpha > 1.$$

and given by:

$$\int_0^{\infty}\varphi f(\varphi)d\varphi = \int_0^{\infty}\psi^{-1}g(\psi)d\psi = \Gamma\left(1 - \frac{1}{\alpha}\right) < \infty,$$

where $\Gamma(x)$ is the Gamma function.

$$\Gamma(x) \equiv \int_0^{\infty} y^{x-1}\exp\{-y\}\,dy.$$

[58] The third parametric family belonging to the class of extreme-value distributions is the Gumbel distribution. However, without any modification (e.g., truncation), it is not a legitimate distribution for φ or ψ since its support includes negative real numbers.

Appendix B: A Sufficient Condition under which the equilibrium is well-defined.

For the equilibrium discussed in the main text to be well-defined, the integrals in the free entry condition and the adding-up constraint must be both well-defined. Since

$$\pi\left(\frac{\psi}{A}\right) = \frac{r(\psi/A)}{\sigma(\psi/A)} < r\left(\frac{\psi}{A}\right),$$

it suffices to show that

$$\int_{\underline{\psi}}^{\psi_c} r\left(\frac{\psi}{A}\right) dG(\psi) < \infty.$$

First, we introduce the following lemma.

Lemma 3: If $\zeta(0) < \infty,\ \lim_{z\to 0}\frac{z\zeta'(z)}{\zeta(z)} = \lim_{z\to 0}\mathcal{E}_\zeta(z) = 0.$

Proof: This follows from $1 < \zeta(z) = \zeta(0)\exp\left[\int_0^z \frac{\xi\zeta'(\xi)}{\zeta(\xi)}\frac{d\xi}{\xi}\right] = \zeta(0)\exp\left[\int_0^z \mathcal{E}_\zeta(\xi)\frac{d\xi}{\xi}\right] < \infty$. ■

Lemma 4. The above integral is finite and hence well-defined, either if $\underline{\psi} > 0 \Leftrightarrow \overline{\varphi} < \infty$ or

$$1 \le \lim_{z\to 0}\zeta(z) < 2 + \lim_{\psi\to 0}\mathcal{E}_g(\psi) = -\lim_{\varphi\to\infty}\mathcal{E}_f(\varphi) < \infty,$$

for $\underline{\psi} = 0 \Leftrightarrow \overline{\varphi} = \infty$.

Proof. Clearly, the integral is well-defined if $\underline{\psi} > 0$. Now suppose $\underline{\psi} = 0$, and $1 \le \lim_{z\to 0}\zeta(z) \equiv \zeta(0) < 2 + \lim_{\psi\to 0}\mathcal{E}_g(\psi) < \infty$. First, $1 \le \zeta(0) < \infty$ implies $\lim_{z\to 0}\mathcal{E}_\zeta(z) = 0$ from Lemma 3.

Second, because

$$\frac{\partial \ln\left[r\left(\frac{\psi}{A}\right)g(\psi)\right]}{\partial \ln\psi} = \mathcal{E}_g(\psi) + \frac{\partial \ln\left[\pi\left(\frac{\psi}{A}\right)\right]}{\partial \ln\psi} + \frac{\partial \ln\left[\sigma\left(\frac{\psi}{A}\right)\right]}{\partial \ln\psi} = \mathcal{E}_g(\psi) - \frac{\left[\sigma\left(\frac{\psi}{A}\right) - 1\right]^2}{\sigma\left(\frac{\psi}{A}\right) - 1 + \mathcal{E}_\zeta\left(Z\left(\frac{\psi}{A}\right)\right)},$$

$$\lim_{\psi\to 0}\frac{\partial \ln\left[r\left(\frac{\psi}{A}\right)g(\psi)\right]}{\partial \ln\psi} = \lim_{\psi\to 0}\mathcal{E}_g(\psi) - \zeta(0) + 1 > -1,$$

where use has been made of $\lim_{z\to 0}\mathcal{E}_\zeta(z) = 0$ and $\zeta(0) < 2 + \lim_{\psi\to 0}\mathcal{E}_g(\psi)$. This inequality means that, for every finite $\psi_c > 0$, there exist $\Lambda(\psi_c) > 0$ and $\delta > 0$ such that,

$$\int_0^{\psi_c} r\left(\frac{\psi}{A}\right) g(\psi) d\psi < \int_0^{\psi_c} \Lambda(\psi_c)\psi^{\delta-1} d\psi = \Lambda(\psi_c)\frac{{\psi_c}^{\delta}}{\delta} < \infty.$$

This completes the proof. ■

It should be noted that the finite mean productivity is neither sufficient nor necessary for the existence of equilibrium. The equilibrium exists even when the mean productivity is infinite, if

$$1 \le \lim_{z\to 0} \zeta(z) < 2 + \lim_{\psi\to 0} \mathcal{E}_g(\psi) = -\lim_{\varphi\to\infty} \mathcal{E}_f(\varphi) < 2,$$

while the equilibrium fails to exist even when the mean productivity is finite if

$$\lim_{z\to 0} \zeta(z) > 2 + \lim_{\psi\to 0} \mathcal{E}_g(\psi) = -\lim_{\varphi\to\infty} \mathcal{E}_f(\varphi) > 2.$$

For example, $\zeta(z) = \sigma > 1$ under CES, and $\mathcal{E}_g(\psi) = \kappa - 1$ under a Power (Pareto), so that the equilibrium exists if $1 < \sigma < \kappa + 1$, and the mean productivity is finite if $\kappa > 1$. Hence, the equilibrium exists even when the mean productivity is infinite, if $1 < \sigma < \kappa + 1 < 2$, while the equilibrium fails to exist even when the mean productivity is finite, if $\sigma > \kappa + 1 > 2$.

Appendix C: Technical Proofs

C.1. Proof of Lemma 6

Lemma 6: Under A2 and the weak A3, $\lim_{\psi/A \to 0} \rho(\psi/A)\sigma(\psi/A) < 1 < \lim_{\psi/A \to \bar{z}} \rho(\psi/A)\sigma(\psi/A)$.

Proof: The proof proceeds in two steps.
Step 1: **A2** and the weak **A3** jointly imply

$$\lim_{\psi/A \to 0} \rho\left(\frac{\psi}{A}\right) < 1 \Leftrightarrow \lim_{z \to 0} \frac{z\zeta'(z)/\zeta(z)}{\zeta(z) - 1} > 0.$$

From Lemma 3, the numerator goes to zero, hence, $\lim_{z \to 0} \zeta(z) = \lim_{\psi/A \to 0} \sigma(\psi/A) = 1$, which proves $\lim_{\psi/A \to 0} \rho(\psi/A)\sigma(\psi/A) < 1$.

Step 2: For $\overline{z} < \infty$,

$$\lim_{z \to \bar{z}} \zeta(z) = \lim_{\psi/A \to \bar{z}} \sigma(\psi/A) = \infty \Longrightarrow \lim_{\psi/A \to \bar{z}} \rho(\psi/A)\sigma(\psi/A) = \infty.$$

For $\overline{z} = \infty$, if $\lim_{\psi/A \to \infty} \rho(\psi/A) = 1$,

$$\lim_{\psi/A \to \bar{z}} \rho(\psi/A)\sigma(\psi/A) = \lim_{\psi/A \to \bar{z}} \sigma(\psi/A) > 1.$$

On the other hand, if $\lim_{\psi/A \to \infty} \rho(\psi/A) < 1 \Leftrightarrow \lim_{z \to \infty} \frac{z\zeta'(z)/\zeta(z)}{\zeta(z)-1} > 0 \Leftrightarrow \lim_{z \to \infty} \frac{z\zeta'(z)}{\zeta(z)} > 0$,

$$\lim_{\psi/A \to \infty} \sigma\left(\frac{\psi}{A}\right) = \lim_{z \to \infty} \zeta(z) = \zeta(z') \exp\left[\int_{z'}^{\infty} \frac{\xi\zeta'(\xi)}{\zeta(\xi)} \frac{d\xi}{\xi}\right] = \infty \Longrightarrow \lim_{\psi/A \to \bar{z}} \rho\left(\frac{\psi}{A}\right) \sigma\left(\frac{\psi}{A}\right) = \infty.$$

Thus, in all of these cases,

$$\lim_{\psi/A \to \bar{z}} \rho(\psi/A)\sigma(\psi/A) > 1.$$

This completes the proof. ■

C.2. Proof of Proposition 5

To prove Proposition 5, we first need the following two lemmas. For this purpose, let us denote $\theta(z) \equiv \mathcal{E}_{1-1/\zeta}(z)$ so that $\rho(\psi/A) = \mathcal{E}_Z(\psi/A) = 1/\left[1 + \theta\left(Z(\psi/A)\right)\right]$.

Lemma 7:

$$\mathcal{E}_\rho\left(\frac{\psi}{A}\right) = \epsilon\left(Z\left(\frac{\psi}{A}\right)\right), \text{where } \epsilon(z) \equiv -\frac{z\theta'(z)}{[1 + \theta(z)]^2}.$$

Proof: Straightforward from the definition. ■

Lemma 8: For $0 \le \rho(0) < \infty, \lim_{z\to 0} \epsilon(z) = 0$.

Proof: From $\rho(\psi/A) = \frac{1}{1+\theta(Z(\psi/A))}$,

$$\rho\left(\frac{\psi}{A}\right) - \rho\left(\frac{\psi_0}{A}\right) = \frac{1}{1+\theta\big(Z(\psi/A)\big)} - \frac{1}{1+\theta\big(Z(\psi_0/A)\big)} = \int_{Z(\psi_0/A)}^{Z(\psi/A)} \frac{d}{d\xi}\left[\frac{1}{1+\theta(\xi)}\right] d\xi$$

$$= \int_{Z(\psi_0/A)}^{Z(\psi/A)} \left[-\frac{\theta'(\xi)}{[1+\theta(\xi)]^2}\right] d\xi \equiv \int_{Z(\psi_0/A)}^{Z(\psi/A)} \frac{\epsilon(\xi)}{\xi} d\xi,$$

for any $\psi_0 > 0$. From $0 \le \rho(0) < \infty$, the RHS remains bounded as $z_0 = Z(\psi_0/A) \to 0$. Hence,

$$\int_0^z \frac{\epsilon(\xi)}{\xi} d\xi < \infty,$$

which implies $\lim_{z\to 0} \epsilon(z) = 0$. This completes the proof. ■

Proposition 5: Suppose that A2 and the strong A3 hold, so that $0 < \rho(\psi/A) < 1$ and $\rho(\psi/A)$ is strictly increasing. Then, $\rho(\psi/A)$ is strictly log-submodular for all $\psi/A < \bar{z}$ with a sufficiently small $\bar{z}$.

Proof: Under **A2**, $\rho(\psi/A) < 1$ for all $\psi/A < \bar{z}$, hence the condition for Lemma 8 holds and $\lim_{z\to 0} \epsilon(z) = 0$. Under the strong **A3**, $\epsilon(z) \equiv -z\theta'(z)/[1+\theta(z)]^2 > 0$ for all $z > 0$. Thus, $\epsilon(\cdot) > 0$ is increasing for a sufficiently small $z > 0$. Hence, from Lemma 7, $\mathcal{E}_\rho(\psi/A)$ is strictly increasing in ψ/A for $\psi/A < Z(\psi/A) < \bar{z}$, with a sufficiently small $\bar{z}$. Hence, from Lemma 5, $\rho(\psi/A)$ is strictly log-submodular for any $\psi/A < Z(\psi/A) < \bar{z}$. This completes the proof. ■

C.3. Proof of Proposition 7a and 7b

Proposition 7a (Market Size Effect on Profit, $\Pi_\psi \equiv \pi(\psi/A)E$): Under **A2,** there exists a unique $\psi_0 \in (\underline{\psi}, \psi_c)$ such that $\sigma\left(\frac{\psi_0}{A}\right) = \mathbb{E}_\pi(\sigma)$ with

$$\frac{d \ln \Pi_\psi}{d \ln E} > 0 \Leftrightarrow \sigma\left(\frac{\psi}{A}\right) < \mathbb{E}_\pi(\sigma) \text{ for } \psi \in (\underline{\psi}, \psi_0),$$

and

$$\frac{d \ln \Pi_\psi}{d \ln E} < 0 \Leftrightarrow \sigma\left(\frac{\psi}{A}\right) > \mathbb{E}_\pi(\sigma) \text{ for } \psi \in (\psi_0, \psi_c).$$

Proof:

From Proposition 6, $\frac{d\ln A}{d\ln E} = \frac{1}{1-\mathbb{E}_\pi(\sigma)}$. Hence, using $\mathcal{E}_\pi\left(\frac{\psi}{A}\right) = 1 - \sigma\left(\frac{\psi}{A}\right)$,

$$\frac{d\ln\Pi_\psi}{d\ln E} = 1 + \frac{\partial\ln\pi(\psi/A)}{\partial\ln A}\frac{d\ln A}{d\ln E} = 1 - \mathcal{E}_\pi\left(\frac{\psi}{A}\right)\frac{d\ln A}{d\ln E} = \frac{\mathbb{E}_\pi(\sigma) - \sigma(\psi/A)}{\mathbb{E}_\pi(\sigma) - 1}.$$

Thus,

$$\frac{d\ln\Pi_\psi}{d\ln E} \gtreqless 0 \Leftrightarrow \sigma\left(\frac{\psi}{A}\right) \lesseqgtr \mathbb{E}_\pi(\sigma).$$

Since $\mathbb{E}_\pi(\sigma)$ is the (profit-weighted) average of $\sigma(\psi/A)$ over $\left(\underline{\psi}, \psi_c\right)$ and $\sigma(\psi/A)$ is strictly increasing under A2, there exists a unique $\psi_0 \in (\underline{\psi}, \psi_c)$ such that $\sigma(\psi_0/A) = \mathbb{E}_\pi(\sigma)$, and $\sigma(\psi/A) < \mathbb{E}_\pi(\sigma)$ for $\psi \in \left(\underline{\psi}, \psi_0\right)$ and $\sigma(\psi/A) > \mathbb{E}_\pi(\sigma)$ for $\psi \in (\psi_0, \psi_c)$. This completes the proof. ■

Proposition 7b (Market Size Effect on Revenue, $\boldsymbol{R_\psi \equiv r(\psi/A)E}$): Under **A2** and the weak **A3,** there exists $\psi_1 > \psi_0$, such that

$$\frac{d\ln R_\psi}{d\ln E} > 0 \text{ for } \psi \in \left(\underline{\psi}, \psi_1\right).$$

Furthermore, $\psi_1 \in (\psi_0, \psi_c)$ and

$$\frac{d\ln R_\psi}{d\ln E} < 0 \text{ for } \psi \in (\psi_1, \psi_c),$$

for a sufficiently small F.[59]

Proof:

From Proposition 6, $\frac{d\ln A}{d\ln E} = \frac{1}{1-\mathbb{E}_\pi(\sigma)}$. Hence, using $\mathcal{E}_r\left(\frac{\psi}{A}\right) = \rho\left(\frac{\psi}{A}\right)\left[1 - \sigma\left(\frac{\psi}{A}\right)\right]$,

$$\frac{d\ln R_\psi}{d\ln E} = 1 + \frac{\partial\ln r(\psi/A)}{\partial\ln A}\frac{d\ln A}{d\ln E} = 1 - \mathcal{E}_r\left(\frac{\psi}{A}\right)\frac{d\ln A}{d\ln E} = 1 - \rho\left(\frac{\psi}{A}\right)\left[\frac{\sigma(\psi/A) - 1}{\mathbb{E}_\pi(\sigma) - 1}\right].$$

Thus,

$$\frac{d\ln R_\psi}{d\ln E} \gtreqless 0 \Leftrightarrow \rho\left(\frac{\psi}{A}\right) \lesseqgtr \frac{\mathbb{E}_\pi(\sigma) - 1}{\sigma(\psi/A) - 1}.$$

Since $\sigma(\psi/A)$ is strictly increasing under A2 and $\rho(\psi/A)$ is non-decreasing under the weak A3, the above inequality changes the sign at most once at $\psi_1 \leq \bar{\psi}$, so that

$$\frac{d\ln R_\psi}{d\ln E} > 0 \text{ for all } \psi \in \left(\underline{\psi}, \psi_1\right)$$

and $\psi_1 > \psi_0 > \underline{\psi}$, because A2 implies

[59] We conjecture whether $\psi_c < \psi_1 \leq \bar{\psi}$ and $\frac{d\ln R_\psi}{d\ln E} > 0$ for all $\psi \in \left(\underline{\psi}, \psi_c\right)$ for a sufficiently large F.

$$\frac{d\ln R_\psi}{d\ln E} = \frac{d\ln \zeta_\psi}{d\ln E} + \frac{d\ln \Pi_\psi}{d\ln E} > \frac{d\ln \Pi_\psi}{d\ln E} \geq 0 \text{ for all } \psi \in \left(\underline{\psi}, \psi_0\right].$$

We now prove $\rho\left(\frac{\psi_c}{A}\right) > \frac{\mathbb{E}_\pi(\sigma)-1}{\sigma(\psi_c/A)-1}$ and hence $\psi_1 < \psi_c$ for a sufficiently small F by showing

$$\lim_{F\to 0} \rho\left(\frac{\psi_c}{A}\right) = \lim_{\psi_c/A\to \bar{z}} \rho\left(\frac{\psi_c}{A}\right) > \lim_{\psi_c/A\to \bar{z}} \left[\frac{\mathbb{E}_\pi(\sigma)-1}{\sigma(\psi_c/A)-1}\right] = \lim_{F\to 0} \left[\frac{\mathbb{E}_\pi(\sigma)-1}{\sigma(\psi_c/A)-1}\right]$$

We divide the proof of this inequality into the following three cases.

Case 1: $0 < \lim_{\psi_c/A\to \bar{z}} \rho\left(\frac{\psi_c}{A}\right) < 1$ and $\bar{z} < \infty$. Then, $\lim_{\psi/A\to \bar{z}} \sigma\left(\frac{\psi}{A}\right) = \infty \Rightarrow \lim_{\psi_c/A\to \bar{z}} \left[\frac{\mathbb{E}_\pi(\sigma)-1}{\sigma(\psi_c/A)-1}\right] = 0.$

Case 2: $0 < \lim_{\psi_c/A\to \bar{z}} \rho\left(\frac{\psi_c}{A}\right) < 1$ and $\bar{z} = \infty$. Then, $\lim_{\psi/A\to \infty} \rho\left(\frac{\psi}{A}\right) < 1 \Leftrightarrow \lim_{z\to\infty} \frac{z\zeta'(z)/\zeta(z)}{\zeta(z)-1} > 0 \Leftrightarrow$ $\lim_{z\to\infty} \frac{z\zeta'(z)}{\zeta(z)} > 0,$ so that $\lim_{\psi/A\to\infty} \sigma\left(\frac{\psi}{A}\right) = \lim_{z\to\infty} \zeta(z) = \zeta(z') \exp\left[\int_{z'}^{\infty} \frac{\xi\zeta'(\xi)}{\zeta(\xi)} \frac{d\xi}{\xi}\right] = \infty \Rightarrow \lim_{\psi_c/A\to \bar{z}} \left[\frac{\mathbb{E}_\pi(\sigma)-1}{\sigma(\psi_c/A)-1}\right] =$ $0.$

Case 3: $\lim_{\psi_c/A\to \bar{z}} \rho\left(\frac{\psi_c}{A}\right) = 1.$ Then, $\lim_{\psi_c/A\to \bar{z}} \left[\frac{\mathbb{E}_\pi(\sigma)-1}{\sigma(\psi_c/A)-1}\right] = \lim_{F\to 0} \left[\frac{\mathbb{E}_\pi(\sigma)-1}{\sigma(\psi_c/A)-1}\right] < 1.$

This completes the proof. ■

C.4. Proof of Proposition 8 and Its Corollaries

Proposition 8: Assume that $\mathcal{E}_g'(\cdot)$ does not change its sign and $\underline{\psi} = 0$. Consider a shock to F_e, E, and/or F, which affects competitive pressures, i.e., $dA \neq 0$. Then, the response of any weighted generalized mean of any monotone function, $f(\psi/A) > 0$, defined by

$$I \equiv \mathcal{M}^{-1}\left(\mathbb{E}_w\big(\mathcal{M}(f)\big)\right)$$

with a monotone transformation $\mathcal{M}: \mathbb{R}_+ \to \mathbb{R}$ and a weighting function, $w(\psi/A) > 0$, satisfies

	$f'(\cdot) > 0$	$f'(\cdot) = 0$	$f'(\cdot) < 0$
$\mathcal{E}_g'(\cdot) > 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Rightarrow \frac{d\ln I}{d\ln A} > 0$	$\frac{d\ln I}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Rightarrow \frac{d\ln I}{d\ln A} < 0$
$\mathcal{E}_g'(\cdot) = 0$ (Pareto)	$\frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln I}{d\ln A} \gtreqless 0$	$\frac{d\ln I}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln I}{d\ln A} \lesseqgtr 0$
$\mathcal{E}_g'(\cdot) < 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Rightarrow \frac{d\ln I}{d\ln A} < 0$	$\frac{d\ln I}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Rightarrow \frac{d\ln I}{d\ln A} > 0$

Moreover, if $\mathcal{E}_g'(\cdot) = \frac{d\ln(\psi_c/A)}{d\ln A} = 0$, $d\ln I/d\ln A = 0$ for any $f(\psi/A)$, monotonic or not. Furthermore, $\mathcal{E}_g'(\cdot)$ can be replaced with $\mathcal{E}_G'(\cdot)$ in all the above statements for $w(\psi/A) = 1$, i.e., the unweighted averages.

Proof: First, by setting $\xi \equiv \psi/A$, and $\psi_c/A = b > 0$,

$$\mathcal{M}(I) \equiv \mathbb{E}_w\big(\mathcal{M}(f)\big) = \frac{\int_0^b \mathcal{M}\big(f(\xi)\big)w(\xi)g(A\xi)d\xi}{\int_0^b w(\xi)g(A\xi)d\xi} \equiv \widehat{\mathcal{M}}(A,b).$$

Hence,

$$\frac{d\ln I}{d\ln A} = \frac{1}{I\mathcal{M}'(I)}\frac{d\mathcal{M}(I)}{d\ln A} = \frac{1}{I\mathcal{M}'(I)}\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial \ln b}\frac{d\ln b}{d\ln A} + \frac{1}{I\mathcal{M}'(I)}\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial \ln A}.$$

The first of the two partial derivatives of $\widehat{\mathcal{M}}(A,b)$ can be expressed as:

$$\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial \ln b} = b\frac{\mathcal{M}\big(f(b)\big)w(b)g(Ab)}{\int_0^b w(\xi)g(A\xi)d\xi} - b\widehat{\mathcal{M}}(A,b)\frac{w(b)g(Ab)}{\int_0^b w(\xi)g(A\xi)d\xi}$$

$$= \big[\mathcal{M}\big(f(b)\big) - \widehat{\mathcal{M}}(A,b)\big]\frac{bw(b)g(Ab)}{\int_0^b w(\xi)g(A\xi)d\xi}.$$

Hence,

$$sgn\left\{\frac{1}{I\mathcal{M}'(I)}\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial \ln b}\frac{d\ln b}{d\ln A}\right\} = sgn\left\{\frac{\big[\mathcal{M}\big(f(b)\big) - \widehat{\mathcal{M}}(A,b)\big]}{I\mathcal{M}'(I)}\frac{bw(b)g(Ab)}{\int_0^b w(\xi)g(A\xi)d\xi}\frac{d\ln b}{d\ln A}\right\}$$

$$= sgn\left\{f'(\cdot)\frac{d\ln b}{d\ln A}\right\}$$

Likewise, the second of the two partial derivatives of $\widehat{\mathcal{M}}(A,b)$ is given by

$$\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial \ln A} = \frac{\int_0^b \mathcal{M}\big(f(\xi)\big)\mathcal{E}_g(\xi A)w(\xi)g(\xi A)d\xi}{\int_0^b w(\xi)g(A\xi)d\xi} - \mathcal{M}(I)\frac{\int_0^b \mathcal{E}_g(\xi A)w(\xi)g(\xi A)d\xi}{\int_0^b w(\xi)g(A\xi)d\xi}$$

$$= \mathbb{E}_{w^0}\Big(\mathcal{M}(f(x))\mathcal{E}_g(xA)\Big) - \mathbb{E}_{w^0}\Big(\mathcal{M}\big(f(x)\big)\Big)\mathbb{E}_{w^0}\Big(\mathcal{E}_g(xA)\Big) = Cov_{w^0}\big[\mathcal{E}_g(xA),\mathcal{M}\big(f(x)\big)\big],$$

where the expectations and the covariance are taken with respect to the random variable whose density function is

$$w^0(x) \equiv \frac{w(x)g(xA)}{\int_0^b w(\xi)g(\xi A)d\xi}.$$

Hence,

$$sgn\left\{\frac{1}{I\mathcal{M}'(I)}\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial \ln A}\right\} = sgn\left\{\frac{Cov_{w^0}\big[\mathcal{E}_g(xA),\mathcal{M}\big(f(x)\big)\big]}{I\mathcal{M}'(I)}\right\} = sgn\{f'(\cdot)\mathcal{E}_g'(\cdot)\}.$$

Therefore,

$$\frac{d\ln I}{d\ln A} = \frac{1}{I\mathcal{M}'(I)}\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial \ln b}\frac{d\ln b}{d\ln A} + \frac{1}{I\mathcal{M}'(I)}\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial \ln A},$$

whose first term has the sign equal to $sgn\left\{f'(\cdot)\frac{d\ln(b)}{d\ln A}\right\}$ and whose second term has the sign equal to $sgn\{f'(\cdot)\mathcal{E}_g'(\cdot)\}$, from which all the results on the weighted generalized mean follows.

For the unweighted generalized mean, we could express

$$\widehat{\mathcal{M}}(A,b) \equiv \frac{\int_0^b \mathcal{M}(f(\xi))g(A\xi)d\xi}{\int_0^b g(A\xi)d\xi} = \int_0^b \mathcal{M}(f(\xi))d\left[\frac{G(A\xi)}{G(Ab)}\right]$$

so that

$$sgn\left\{\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial\ln A}\right\} = -sgn\{\mathcal{M}'(\cdot)f'(\cdot)\mathcal{E}_G'(\cdot)\}.$$

Hence,

$$\frac{A}{I}\frac{dI}{dA} = \frac{1}{I\mathcal{M}'(I)}\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial\ln b}\frac{d\ln b}{d\ln A} + \frac{1}{I\mathcal{M}'(I)}\frac{\partial\widehat{\mathcal{M}}(A,b)}{\partial\ln A},$$

whose first term has the sign equal to $sgn\left\{f'(\cdot)\frac{d\ln(b)}{d\ln A}\right\}$ and whose second term has the sign equal to $sgn\{f'(\cdot)\mathcal{E}_G'(\cdot)\}$, from which all the results on the unweighted generalized mean follows. ■

Corollary 1 of Proposition 8
a) Entry Cost: $f'(\cdot)\mathcal{E}_g'(\cdot) \gtreqless 0 \Leftrightarrow \frac{d\ln I}{d\ln F_e} = \frac{d\ln I}{d\ln A}\frac{d\ln A}{d\ln F_e} \gtreqless 0.$
b) Market Size: *If* $\mathcal{E}_g'(\cdot) \leq 0$, then, $f'(\cdot) \gtreqless 0 \Rightarrow \frac{d\ln I}{d\ln E} = \frac{d\ln I}{d\ln A}\frac{d\ln A}{d\ln E} \gtreqless 0.$
c) Overhead Cost: *If* $\mathcal{E}_g'(\cdot) \leq 0$, then, $f'(\cdot) \gtreqless 0 \Rightarrow \frac{d\ln I}{d\ln F} = \frac{d\ln I}{d\ln A}\frac{d\ln A}{d\ln F} \lesseqgtr 0.$
Furthermore, $\mathcal{E}_g'(\cdot)$ can be replaced with $\mathcal{E}_G'(\cdot)$ for $w(\psi/A) = 1$, i.e., the unweighted averages.

Proof: Corollary 1a) follows from $\frac{d\ln A}{d\ln F_e} > 0$ and $\frac{d\ln(\psi_c/A)}{d\ln F_e} = 0$, and hence $\frac{d\ln(\psi_c/A)}{d\ln A} = 0$.

Corollary 1b) follows from $\frac{d\ln A}{d\ln E} < 0$ and $\frac{d\ln(\psi_c/A)}{d\ln E} > 0$, and hence $\frac{d\ln(\psi_c/A)}{d\ln A} < 0$. Finally,

Corollary 1c) follows from $\frac{d\ln A}{d\ln F} > 0$ and $\frac{d\ln(\psi_c/A)}{d\ln F} < 0$, and hence $\frac{d\ln(\psi_c/A)}{d\ln A} < 0$. ■

Corollary 2 of Proposition 8: Assume $\underline{\psi} = 0$, and neither $\zeta'(\cdot)$ nor $\mathcal{E}_g'(\cdot)$ change the signs.

Consider a shock to F_e, E, and/or F, which affects competitive pressures, i.e., $dA \neq 0$. Then, the response of P/A satisfies

	$\zeta'(\cdot) > 0$	$\zeta'(\cdot) = 0$ (CES)	$\zeta'(\cdot) < 0$
$\mathcal{E}_g'(\cdot) > 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Rightarrow \frac{d\ln(P/A)}{d\ln A} > 0$	$\frac{d\ln(P/A)}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Rightarrow \frac{d\ln(P/A)}{d\ln A} < 0$

$\mathcal{E}_g'(\cdot) = 0$ (Pareto)	$\frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln(P/A)}{d\ln A} \gtreqless 0$	$\frac{d\ln(P/A)}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln(P/A)}{d\ln A} \lesseqgtr 0$
$\mathcal{E}_g'(\cdot) < 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Rightarrow \frac{d\ln(P/A)}{d\ln A} < 0$	$\frac{d\ln(P/A)}{d\ln A} = 0$	$\frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Rightarrow \frac{d\ln(P/A)}{d\ln A} > 0$

Proof: Relationship between A and P:

$$\ln\left(\frac{A}{cP}\right) = \int_{\Omega} \left[\int_{z(\omega)}^{\bar{z}} \frac{s(\xi)}{\xi} d\xi\right] d\omega = M \int_0^{\psi_c} \left[\int_{Z(\psi/A)}^{\bar{z}} \frac{s(\xi)}{\xi} d\xi\right] dG(\psi)$$

From the adding-up constraint,

$$\ln\left(\frac{A}{cP}\right) = \frac{\int_0^{\psi_c} \Phi\big(Z(\psi/A)\big)\, r(\psi/A) dG(\psi)}{\int_0^{\psi_c} r(\psi/A)\, dG(\psi)} = \mathbb{E}_r(\Phi \circ Z),$$

where

$$\Phi(z) \equiv \frac{1}{s(z)} \int_z^{\bar{z}} \frac{s(\xi)}{\xi} d\xi,$$

which satisfies Lemma 1 of Matsuyama and Ushchev (2023):

$$\zeta'(\cdot) \gtreqless 0 \;\Rightarrow \Phi'(\cdot) \lesseqgtr 0.$$

The results in the table follow from Proposition 8a for $I = A/P$, $\mathcal{M}(f) = f$, $f = \Phi \circ Z$ and $w = r$. Hence, $\frac{d\ln(P/A)}{d\ln A}$ is the sum of the two terms, one of which has the sign equal to $sgn\{\mathcal{E}_g'(\cdot)\zeta'(\cdot)\}$ and the other has the sign equal to $sgn\left\{\frac{d\ln\left(\frac{\psi_c}{A}\right)}{d\ln A}\zeta'(\cdot)\right\}$, from which the result follows. ■

C.5. Proof of Proposition 9 and Corollary 1 of Proposition

Proposition 9: Assume that $\mathcal{E}_G'(\cdot)$ does not change its sign and $\underline{\psi} = 0$. Consider a shock to F_e, F, and/or E, which affects competitive pressures, i.e., $dA \neq 0$. Then, the response of the mass of active firms, $MG(\psi_c)$, is as follows:

$$If\ \mathcal{E}_G'(\cdot) > 0, \qquad \frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Rightarrow \frac{d\ln[MG(\psi_c)]}{d\ln A} > 0;$$

$$If\ \mathcal{E}_G'(\cdot) = 0, \qquad \frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln[MG(\psi_c)]}{d\ln A} \gtreqless 0;$$

$$If\ \mathcal{E}_G'(\cdot) < 0, \qquad \frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Rightarrow \frac{d\ln[MG(\psi_c)]}{d\ln A} < 0.$$

Proof. From the adding-up constraint, $M\int_0^{\psi_c} r(\psi/A)\, g(\psi) d\psi = 1$,

$$\mathbb{E}_1(r) = \frac{\int_0^{\psi_c} r(\psi/A) g(\psi)\, d\psi}{\int_0^{\psi_c} g(\psi)\, d\psi} = \frac{1}{MG(\psi_c)}.$$

By applying Proposition 8 for $f(\cdot) = r(\cdot), w(\cdot) = 1$, and $\mathcal{M}(f) = f$, so that $I = \mathbb{E}_1(r)$, which is an unweighted generalized mean, and noting $r'(\cdot) < 0$ and

$$If\ \mathcal{E}_G'(\cdot) > 0, \qquad \frac{d\ln(\psi_c/A)}{d\ln A} \geq 0 \Rightarrow \frac{d\ln \mathbb{E}_1(r)}{d\ln A} = -\frac{d\ln[MG(\psi_c)]}{d\ln A} < 0;$$

$$If\ \mathcal{E}_G'(\cdot) = 0, \qquad \frac{d\ln(\psi_c/A)}{d\ln A} \gtreqless 0 \Leftrightarrow \frac{d\ln \mathbb{E}_1(r)}{d\ln A} = -\frac{d\ln[MG(\psi_c)]}{d\ln A} \lesseqgtr 0;$$

$$If\ \mathcal{E}_G'(\cdot) < 0, \qquad \frac{d\ln(\psi_c/A)}{d\ln A} \leq 0 \Rightarrow \frac{d\ln \mathbb{E}_1(r)}{d\ln A} = -\frac{d\ln[MG(\psi_c)]}{d\ln A} > 0.$$

This completes the proof. ■

Corollary 1 of Proposition 9

a) Entry Cost: $\mathcal{E}_G'(\cdot) \gtreqless 0 \Leftrightarrow \frac{d\ln[MG(\psi_c)]}{d\ln F_e} = \frac{d\ln[MG(\psi_c)]}{d\ln A}\frac{d\ln A}{d\ln F_e} \gtreqless 0.$

b) Market Size: $\mathcal{E}_G'(\cdot) \leq 0 \Rightarrow \frac{d\ln[MG(\psi_c)]}{d\ln E} = \frac{d\ln[MG(\psi_c)]}{d\ln A}\frac{d\ln A}{d\ln E} > 0.$

c) Overhead Cost: $\mathcal{E}_G'(\cdot) \leq 0 \Rightarrow \frac{d\ln[MG(\psi_c)]}{d\ln F} = \frac{d\ln[MG(\psi_c)]}{d\ln A}\frac{d\ln A}{d\ln F} < 0.$

Proof: Corollary a) follows from $\frac{d\ln A}{d\ln F_e} > 0$ and $\frac{d\ln(\psi_c/A)}{d\ln F_e} = 0$, and hence $\frac{d\ln(\psi_c/A)}{d\ln A} = 0$.

Corollary b) follows from $\frac{d\ln A}{d\ln E} < 0$ and $\frac{d\ln(\psi_c/A)}{d\ln E} > 0$, and hence $\frac{d\ln(\psi_c/A)}{d\ln A} < 0$. Finally,

Corollary c) follows from $\frac{d\ln A}{d\ln F} > 0$ and $\frac{d\ln(\psi_c/A)}{d\ln F} < 0$, and hence $\frac{d\ln(\psi_c/A)}{d\ln A} < 0$. ■

C.6. Proof of Propositions 11a and 11b

To prove Proposition 11, we will need the following lemma.

Lemma 9: Suppose $G(\psi) = (\psi/\overline{\psi})^{\kappa}$. Then, the equilibrium conditions can be stated as

$$\int_{a_j}^1 r(b_j\xi)\xi^{\kappa-1}d\xi = a_{j+1}^{-\kappa}\int_{a_{j+1}}^1 r(b_{j+1}\xi)\xi^{\kappa-1}d\psi;\ a_0 = 0$$

$$E_j\pi(b_j) = E_{j+1}\pi(a_j b_{j+1});\ \ E_J\pi(b_J) = F.$$

$$\sum_{j=1}^{J}(a_2 \dots a_{j-1})^{-\kappa}\int_{a_{j-1}}^1 [E_j\pi(b_j\xi) - F]\xi^{\kappa-1}d\xi = \left(\frac{\overline{\psi}}{\psi_1}\right)^{\kappa}\frac{F_e}{\kappa},$$

where $a_j \equiv \psi_{j-1}/\psi_j$ and $b_j \equiv \psi_j/A_j$.

Proof: First, from the adding-up constraints,

$$\int_{\psi_{j-1}}^{\psi_j} r\left(\frac{\psi}{A_j}\right)\psi^{\kappa-1}d\psi = \int_{\psi_j}^{\psi_{j+1}} r\left(\frac{\psi}{A_{j+1}}\right)\psi^{\kappa-1}d\psi.$$

for $j = 1,2,\dots,J-1$. By setting $\xi \equiv \psi/\psi_j$ in the LHS and $\xi \equiv \psi/\psi_{j+1}$ in the RHS, this can be written as:

$$\int_{\psi_{j-1}/\psi_j}^{1} r\left(\frac{\psi_j}{A_j}\xi\right)\xi^{\kappa-1}d\xi = \left(\frac{\psi_j}{\psi_{j+1}}\right)^{-\kappa}\int_{\psi_j/\psi_{j+1}}^{1} r\left(\frac{\psi_{j+1}}{A_{j+1}}\xi\right)\xi^{\kappa-1}d\psi.$$

Second, the cutoff conditions *for* $j = 1,2,\dots,J-1$ can rewritten as:

$$E_j\pi\left(\frac{\psi_j}{A_j}\right) = E_{j+1}\pi\left(\frac{\psi_j}{A_{j+1}}\right);$$

and

$$E_J\pi\left(\frac{\psi_J}{A_J}\right) = F.$$

Third, the free-entry condition can be written as

$$\sum_{j=1}^{J}\left(\frac{\psi_j}{\psi_1}\right)^{\kappa}\int_{\psi_{j-1}/\psi_j}^{1}\left[E_j\pi\left(\frac{\psi_j}{A_j}\xi\right) - F\right]\xi^{\kappa-1}d\xi = \left(\frac{\overline{\psi}}{\psi_1}\right)^{\kappa}\frac{F_e}{\kappa}.$$

Using $a_j \equiv \psi_{j-1}/\psi_j < 1$ and $b_j \equiv \psi_j/A_j$ for $j = 1,2\dots,J$, the three conditions can be written as:

$$\int_{a_j}^{1} r(b_j\xi)\xi^{\kappa-1}d\xi = a_{j+1}^{-\kappa}\int_{a_{j+1}}^{1} r(b_{j+1}\xi)\xi^{\kappa-1}d\psi;\ a_0 = 0$$

$$E_j\pi(b_j) = E_{j+1}\pi(a_j b_{j+1});\ \ E_J\pi(b_J) = F.$$

$$\sum_{j=1}^{J}(a_2\dots a_{j-1})^{-\kappa}\int_{a_{j-1}}^{1}[E_j\pi(b_j\xi) - F]\xi^{\kappa-1}d\xi = \left(\frac{\overline{\psi}}{\psi_1}\right)^{\kappa}\frac{F_e}{\kappa}.$$

This completes the proof. ■

Proposition 11a: Suppose A2 and $G(\psi) = (\psi/\overline{\psi})^{\kappa}$. There exists a sequence, $E_1 > E_2 > \cdots > E_J > 0$, such that, in equilibrium, any weighted generalized mean of $f(\psi/A_j)$ across firms operating at market-j are increasing (decreasing) in j even though $f(\cdot)$ is increasing (decreasing) and hence $f(\psi/A_j)$ is decreasing (increasing) in j.

Proof: First, consider an equilibrium along which

$$b_j = b = \pi^{-1}\left(\frac{F}{E_J}\right)$$

is constant. Then, the first condition implies that a_j solves the following difference equation, $a_{j+1} = D(a_j)$, defined by:

$$\int_{a_j}^{1} r(b\xi)\xi^{\kappa-1}d\xi \equiv a_{j+1}^{-\kappa}\int_{a_{j+1}}^{1} r(b\xi\,)\xi^{\kappa-1}d\psi.$$

with the initial condition, $a_0 = 0$. The LHS is strictly positive and strictly decreasing in $0 < a_j < 1$ and goes to zero as $a_j \to 1$, while the RHS is positive and strictly decreasing in $0 < a_{j+1} < 1$ and goes to infinity as $a_{j+1} \to 0$ and goes to zero as $a_{j+1} \to 1$. Hence, it has a unique solution, $a_{j+1} = D(a_j)$, which satisfies, for $0 \le a_j < 1, a_j < D(a_j) = a_{j+1} < 1$. Thus, $0 = a_0 < a_1 < \cdots < a_J < 1$. From A2, the second condition is satisfied with

$$\frac{L_j}{L_{j+1}} = \frac{\pi(a_j b)}{\pi(b)} > 1.$$

Furthermore, a_j is monotone increasing in j implies that any weighted generalized mean of $f(\psi/A_j) = f(b\,\psi/\psi_j)$,

$$\mathcal{M}^{-1}\left(\frac{\int_{a_{j-1}}^{1}\mathcal{M}\big(f(b\xi)\big)w(b\xi)\xi^{\kappa-1}d\xi}{\int_{a_{j-1}}^{1} w(b\xi)\xi^{\kappa-1}d\xi}\right),$$

is increasing (decreasing) in j if and only if $f(\cdot)$ is increasing (decreasing).

This completes the proof. ■

Proposition 11b: Suppose $G(\psi) = \left(\psi/\overline{\psi}\right)^{\kappa}$. Then, a change in F_e keeps

iii) the ratios $a_j \equiv \psi_{j-1}/\psi_j$ and $b_j \equiv \psi_j/A_j$

and

iv) any weighted generalized mean of $f(\psi/A_j)$ across firms operating at market-j, for any weighting function $w(\psi/A_j)$,

unchanged for all $j = 1,2,\dots,J$.

Proof:

i) The first two equilibrium conditions of Lemma 9 jointly pin down $(a_0, a_1, a_2, \dots, a_{J-1}; b_1, b_2, \dots, b_J)$ and hence the LHS of the third condition pins down the RHS. Thus, for all $j = 1,2, \dots, J$,

$$\frac{d\psi_j}{\psi_j} = \frac{dA_j}{A_j} = \frac{1}{\kappa}\frac{dF_e}{F_e}.$$

ii) Take any firm-specific variable that can be written as a function of ψ/A_j, $f(\psi/A_j)$, for firms operating at market-j, and let $w(\psi/A_j) > 0$ be a weighting function, such as the revenue, profit, or employment within market-j. A weighted generalized mean of $f(\psi/A_j)$ for market-j is given by

$$\mathcal{M}^{-1}\left(\frac{\int_{\psi_{j-1}}^{\psi_j} \mathcal{M}\left(f(\psi/A_j)\right) w(\psi/A_j) dG(\psi)}{\int_{\psi_{j-1}}^{\psi_j} w(\psi/A_j) dG(\psi)}\right).$$

Setting $\xi \equiv \psi/\psi_j$, the weighted average of $f(\psi/A_j)$ across firms operating at market-j becomes:

$$\mathcal{M}^{-1}\left(\frac{\int_{\psi_{j-1}/\psi_j}^{1} \mathcal{M}\left(f\left(\frac{\psi_j}{A_j}\xi\right)\right) w\left(\frac{\psi_j}{A_j}\xi\right) \xi^{\kappa-1} d\xi}{\int_{\psi_{j-1}/\psi_j}^{1} w\left(\frac{\psi_j}{A_j}\xi\right) \xi^{\kappa-1} d\xi}\right) = \mathcal{M}^{-1}\left(\frac{\int_{a_j}^{1} \mathcal{M}\left(f(b_j\xi)\right) w(b_j\xi) \xi^{\kappa-1} d\xi}{\int_{a_j}^{1} w(b_j\xi) \xi^{\kappa-1} d\xi}\right),$$

where $a_j \equiv \psi_{j-1}/\psi_j < 1$ and $b_j \equiv \psi_j/A_j$. Since a_j and b_j remain unchanged in response to a change in F_e by part i), any weighted generalized mean of $f(\psi/A_j)$ also remain unchanged in response to a reduction in F_e. This completes the proof. ■

Appendix D: Three Parametric Families of H.S.A.

D.1. Generalized Translog: Matsuyama and Ushchev (2020a, 2022).

For $\sigma > 1$ and $\beta, \eta, \gamma > 0$,

$$s(z) = \gamma\left(1 - \frac{\sigma - 1}{\eta}\ln\left(\frac{z}{\beta}\right)\right)^{\eta} = \gamma\left(-\frac{\sigma - 1}{\eta}\ln\left(\frac{z}{\bar{z}}\right)\right)^{\eta};\ z < \bar{z} \equiv \beta e^{\frac{\eta}{\sigma-1}}$$

$$\Rightarrow \zeta(z) = 1 + \frac{\sigma - 1}{1 - \frac{\sigma - 1}{\eta}\ln\left(\frac{z}{\beta}\right)} = 1 - \frac{\eta}{\ln(z/\bar{z})} > 1,$$

which is strictly increasing in z for all $z \in (0, \bar{z})$, hence satisfying **A2**. In contrast,

$$\frac{z\zeta'(z)}{[\zeta(z) - 1]\zeta(z)} = \frac{1}{\eta}\left[1 - \frac{1}{\zeta(z)}\right] = \frac{1}{\eta - \ln(z/\bar{z})}$$

is strictly increasing in z for all $z \in (0, \bar{z})$. Thus, the weak **A3** is violated.[60]

Notes:

- CES is the limit case, as $\eta \to \infty$, while holding $\beta > 0$ and $\sigma > 1$ fixed.

$$z < \bar{z} \equiv \beta e^{\frac{\eta}{\sigma-1}} \to \infty$$

$$\zeta(z) = 1 + \frac{\sigma - 1}{1 - \frac{\sigma - 1}{\eta}\ln\left(\frac{z}{\beta}\right)} \to \sigma;\ \ s(z) = \gamma\left(1 - \frac{\sigma - 1}{\eta}\ln\left(\frac{z}{\beta}\right)\right)^{\eta} \to \gamma\left(\frac{z}{\beta}\right)^{1-\sigma};$$

- Translog is the special case where $\eta = 1$.
- $z = Z\left(\frac{\psi}{A}\right)$ is given as the inverse of $\frac{\eta z}{\eta - \ln(z/\bar{z})} = \frac{\psi}{A}$;
- If $\eta \geq 1$, $\frac{z\zeta'(z)}{\zeta(z)} < \eta z\zeta'(z) = [\zeta(z) - 1]^2$; and employment is globally decreasing in z;
- If $\eta < 1$, employment is hump-shaped with the peak, given by $\eta\zeta(\hat{z}) = 1 \Leftrightarrow \hat{z}/\bar{z} = \frac{\hat{\psi}}{(1-\eta)\bar{z}A} = \exp\left[-\frac{\eta^2}{1-\eta}\right] < 1$, decreasing in η.

D.2. Constant Pass-Through (CoPaTh): Matsuyama and Ushchev (2020a, 2020b)

[60] Indeed, any H.S.A. satisfying A2 and $\lim_{z\to 0} s(z) = \infty$ violates the weak A3. To see this, under A2, $1 \leq \zeta(0) < \zeta(z) < \infty$ for any $\bar{z} > z_0 > z > 0$, hence, $0 < \int_z^{z_0} \frac{\zeta'(\xi)}{\zeta(\xi)} d\xi = \ln\zeta(z_0) - \ln\zeta(z) < \infty$. Moreover, under the weak A3, $\theta(z) \equiv \frac{z\zeta'(z)}{[\zeta(z)-1]\zeta(z)} > 0$ is non-increasing because $\theta\big(Z(\psi/A)\big) = \frac{1}{\rho(\psi/A)} - 1$. Thus, $\ln s(z) - \ln s(z_0) = \int_z^{z_0} \frac{\zeta(\xi)-1}{\xi} d\xi = \int_z^{z_0} \frac{1}{\theta(\xi)}\frac{\zeta'(\xi)}{\zeta(\xi)} d\xi \leq \frac{1}{\theta(z_0)}\int_z^{z_0} \frac{\zeta'(\xi)}{\zeta(\xi)} d\xi$, from which $\lim_{z\to 0} s(z) \leq \ln s(z_0) + \frac{1}{\theta(z_0)}\int_0^{z_0} \frac{\zeta'(\xi)}{\zeta(\xi)} d\xi < \infty$.

For $0 < \rho < 1, \sigma > 1, \beta > 0,$ and $\gamma > 0,$

$$s(z) = \gamma \left[\sigma - (\sigma - 1)\left(\frac{z}{\beta}\right)^{\frac{1-\rho}{\rho}}\right]^{\frac{\rho}{1-\rho}} = \gamma\sigma^{\frac{\rho}{1-\rho}}\left[1 - \left(\frac{z}{\bar{z}}\right)^{\frac{1-\rho}{\rho}}\right]^{\frac{\rho}{1-\rho}} \text{ for } z < \bar{z} \equiv \beta\left(\frac{\sigma}{\sigma - 1}\right)^{\frac{\rho}{1-\rho}}$$

$$\Rightarrow 1 - \frac{1}{\zeta(z)} = \left(\frac{z}{\bar{z}}\right)^{\frac{1-\rho}{\rho}} < 1 \text{ for } z < \bar{z} \equiv \beta\left(\frac{\sigma}{\sigma - 1}\right)^{\frac{\rho}{1-\rho}}$$

$$\Rightarrow \mathcal{E}_{1-1/\zeta}(z) = -\mathcal{E}_{\zeta/(\zeta-1)}(z) = \frac{1-\rho}{\rho} > 0.$$

satisfying **A2** and the weak form of A3 (but not the strong form).

Note: CES is the limit case, as $\rho \to 1$, while holding $\beta > 0$ and $\sigma > 1$ fixed:

$$z < \bar{z} \equiv \beta\left(\frac{\sigma}{\sigma - 1}\right)^{\frac{\rho}{1-\rho}} \to \infty;$$

$$\zeta(z) = \frac{\sigma}{\sigma - (\sigma - 1)\left(\frac{z}{\beta}\right)^{\frac{1-\rho}{\rho}}} \to \sigma;$$

$$s(z) = \gamma\left[\sigma - (\sigma - 1)\left(\frac{z}{\beta}\right)^{\frac{1-\rho}{\rho}}\right]^{\frac{\rho}{1-\rho}} \to \gamma\left(\frac{z}{\beta}\right)^{1-\sigma};$$

because, by applying l'Hôpital's rule for $\Delta = \frac{1-\rho}{\rho}$,

$$\lim_{\rho \nearrow 1} \ln \frac{s(z)}{\gamma} = \lim_{\Delta \searrow 0} \frac{\ln\left[\sigma - (\sigma - 1)\left(\frac{z}{\beta}\right)^{\Delta}\right]}{\Delta} = \lim_{\Delta \searrow 0} \frac{(1-\sigma)\left(\frac{z}{\beta}\right)^{\Delta} \ln\left(\frac{z}{\beta}\right)}{\sigma - (\sigma - 1)\left(\frac{z}{\beta}\right)^{\Delta}} = (1-\sigma)\ln\left(\frac{z}{\beta}\right).$$

Monopoly Pricing: From the firm's FOC:

$$z_\psi\left[1 - \frac{1}{\zeta(z_\psi)}\right] = \frac{\psi}{A}.$$

$$z_\psi \equiv Z\left(\frac{\psi}{A}\right) = (\bar{z})^{1-\rho}\left(\frac{\psi}{A}\right)^{\rho}$$

which features a constant (incomplete) pass-through rate, $0 < \rho < 1$. Hence, the weak form of **A3** holds, but not the strong form of **A3**. Furthermore,

$$\sigma\left(\frac{\psi}{A}\right) = \zeta\left(Z\left(\frac{\psi}{A}\right)\right) = \frac{1}{1 - \left(\frac{\psi}{\bar{z}A}\right)^{1-\rho}} = \frac{1}{1 - \left(1 - \frac{1}{\sigma}\right)^{\rho}\left(\frac{\psi}{\beta A}\right)^{1-\rho}} > \sigma$$

increasing in ψ/A for $\psi/A < \bar{z}$, while

$$r\left(\frac{\psi}{A}\right) = s\left(Z\left(\frac{\psi}{A}\right)\right) = \gamma\sigma^{\frac{\rho}{1-\rho}}\left[1-\left(\frac{\psi}{\bar{z}A}\right)^{1-\rho}\right]^{\frac{\rho}{1-\rho}} = \gamma\sigma^{\frac{\rho}{1-\rho}}\left[1-\left(1-\frac{1}{\sigma}\right)^{\rho}\left(\frac{\psi}{\beta A}\right)^{1-\rho}\right]^{\frac{\rho}{1-\rho}}$$

$$\pi\left(\frac{\psi}{A}\right) = \frac{r(\psi/A)}{\sigma(\psi/A)} = \gamma\sigma^{\frac{\rho}{1-\rho}}\left[1-\left(\frac{\psi}{\bar{z}A}\right)^{1-\rho}\right]^{\frac{1}{1-\rho}} = \gamma\sigma^{\frac{\rho}{1-\rho}}\left[1-\left(1-\frac{1}{\sigma}\right)^{\rho}\left(\frac{\psi}{\beta A}\right)^{1-\rho}\right]^{\frac{1}{1-\rho}}$$

are decreasing in ψ/A for $\psi/A < \bar{z}$. In contrast,

$$\ell\left(\frac{\psi}{A}\right) = r\left(\frac{\psi}{A}\right) - \pi\left(\frac{\psi}{A}\right) = \gamma\sigma^{\frac{\rho}{1-\rho}}\left(\frac{\psi}{\bar{z}A}\right)^{1-\rho}\left[1-\left(\frac{\psi}{\bar{z}A}\right)^{1-\rho}\right]^{\frac{\rho}{1-\rho}}$$

$$= \gamma\sigma^{\frac{\rho}{1-\rho}}\left(1-\frac{1}{\sigma}\right)^{\rho}\left(\frac{\psi}{\beta A}\right)^{1-\rho}\left[1-\left(1-\frac{1}{\sigma}\right)^{\rho}\left(\frac{\psi}{\beta A}\right)^{1-\rho}\right]^{\frac{\rho}{1-\rho}}$$

increasing in ψ/A for $\psi/A < \hat{\psi}/A \equiv \bar{z}(1-\rho)^{\frac{1}{1-\rho}}$ and decreasing in ψ/A for $\hat{\psi}/A < \psi/A < \bar{z}$.

Equivalently, employment is increasing in z for $z < \hat{z} \equiv (\bar{z})^{1-\rho}\left(\hat{\psi}/A\right)^{\rho} = \bar{z}(1-\rho)^{\frac{\rho}{1-\rho}}$ and decreasing in z for $\hat{z} < z < \bar{z}$. Note also that

$$\hat{z}/\bar{z} = (1-\rho)^{\frac{\rho}{1-\rho}} > \hat{\psi}/\bar{z}A = (1-\rho)^{\frac{1}{1-\rho}},$$

which is monotonically decreasing in ρ with $\hat{z}/\bar{z} \to 1$ and $\hat{\psi}/\bar{z}A \to 1$, as $\rho \to 0$, and $\hat{z}/\bar{z} \to 0$ and $\hat{\psi}/\bar{z}A \to 0$, as $\rho \to 1$.

D.3. Power Elasticity of Markup Rate (a.k.a. Fréchet Inverse Markup Rate): For $\kappa \geq 0$ and $\lambda > 0$

$$s(z) = \exp\left[\int_{z_0}^{z}\frac{c}{c-\exp\left[-\frac{\kappa\bar{z}^{-\lambda}}{\lambda}\right]\exp\left[\frac{\kappa\xi^{-\lambda}}{\lambda}\right]}\frac{d\xi}{\xi}\right],$$

with either $\bar{z} = \infty$ and $c \leq 1$ or $\bar{z} < \infty$ and $c = 1$. Then,

$$1-\frac{1}{\zeta(z)} = c\exp\left[\frac{\kappa\bar{z}^{-\lambda}}{\lambda}\right]\exp\left[-\frac{\kappa z^{-\lambda}}{\lambda}\right] < 1$$

$$\Longrightarrow \mathcal{E}_{1-1/\zeta}(z) = -\mathcal{E}_{\zeta/(\zeta-1)}(z) = \kappa z^{-\lambda};$$

satisfying **A2** and the strong **A3** for $\kappa > 0$ and $\lambda > 0$.

CES for $\kappa = 0$; $\bar{z} = \infty$; $c = 1-\frac{1}{\sigma}$; CoPaTh for $\bar{z} < \infty$; $c = 1$; $\kappa = \frac{1-\rho}{\rho} > 0$, and $\lambda \to 0$**.**

With $z = Z\left(\frac{\psi}{A}\right)$ given implicitly by $c\exp\left[\frac{\kappa\bar{z}^{-\lambda}}{\lambda}\right]z\exp\left[-\frac{\kappa z^{-\lambda}}{\lambda}\right] \equiv \frac{\psi}{A}$,

$$\rho\left(\frac{\psi}{A}\right)=\frac{1}{1+\kappa z^{-\lambda}}\Leftrightarrow\mathcal{E}_{\rho}\left(\frac{\psi}{A}\right)=\frac{\lambda\kappa z^{-\lambda}}{[1+\kappa z^{-\lambda}]^{2}}>0.$$

Hence,

$$\frac{\partial^{2}\ln\rho\left(\frac{\psi}{A}\right)}{\partial A\partial\psi}\lesseqgtr 0\Leftrightarrow\mathcal{E}_{\rho}'\left(\frac{\psi}{A}\right)\gtreqless 0\Leftrightarrow\kappa z^{-\lambda}\gtreqless 1\Leftrightarrow\frac{\psi}{A}\lesseqgtr(\kappa)^{\frac{1}{\lambda}}zc\exp\left[\frac{\kappa\bar{z}^{-\lambda}-1}{\lambda}\right].$$

Thus, the pass-through rate is log-submodular among more efficient firms, while log-supermodular among less efficient firms. In particular, if $\bar{z}<(\kappa)^{\frac{1}{\lambda}}$, $\frac{\partial^{2}\ln\rho(\psi/A)}{\partial A\partial\psi}<0$ for all $\psi/A<Z(\psi/A)<\bar{z}<\infty$.

Employment is hump-shaped with the peak at $\hat{z}=Z\left(\frac{\hat{\psi}}{A}\right)$, satisfying $\frac{\hat{z}\zeta'(\hat{z})}{\zeta(\hat{z})}\equiv[\zeta(\hat{z})-1]^{2}\Leftrightarrow\rho\left(\frac{\hat{\psi}}{A}\right)\sigma\left(\frac{\hat{\psi}}{A}\right)=1$. This is given by

$$c\left(1+\frac{\hat{z}^{\lambda}}{\kappa}\right)\exp\left[-\frac{\kappa\hat{z}^{-\lambda}}{\lambda}\right]\exp\left[\frac{\kappa\bar{z}^{-\lambda}}{\lambda}\right]=1\Leftrightarrow\left(1+\frac{\hat{z}^{\lambda}}{\kappa}\right)\hat{z}=\frac{\hat{\psi}}{A}.$$

Appendix E: Some Additional Figures

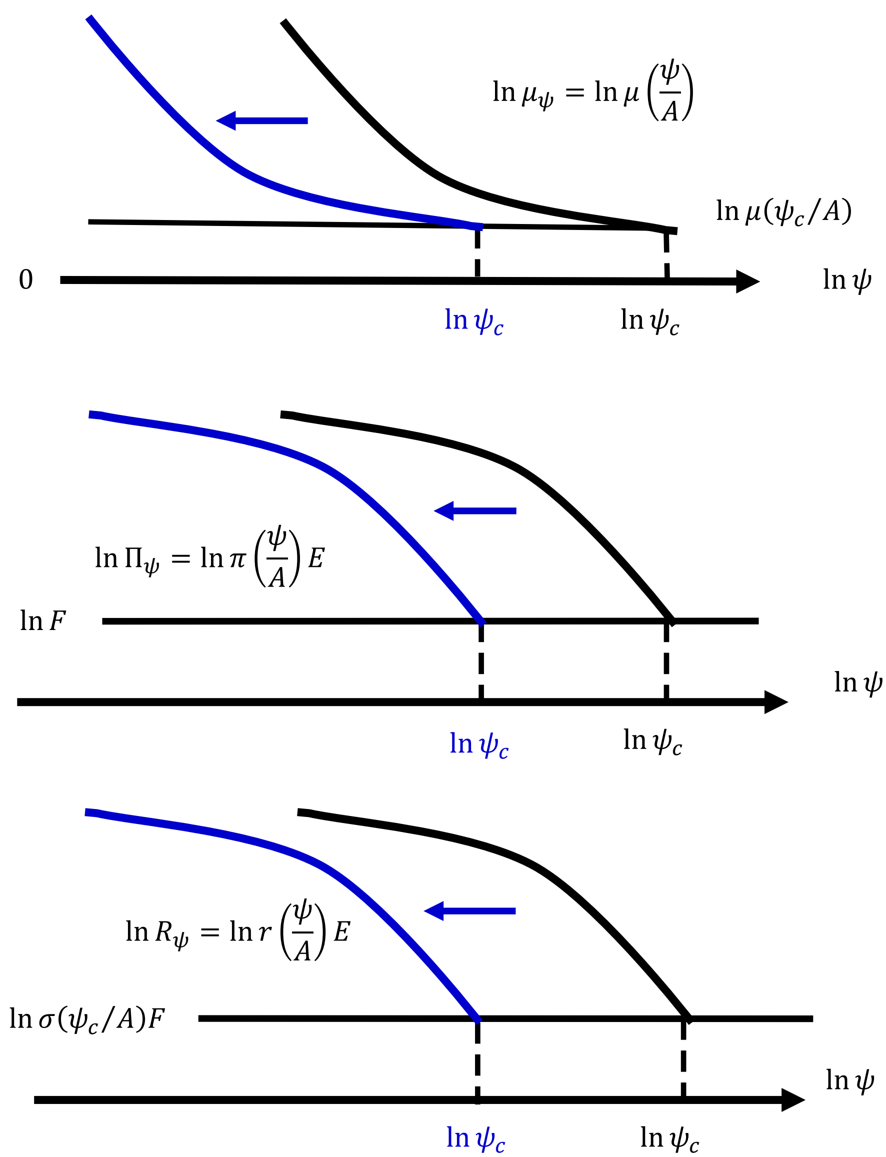


Figure 5a: $F_e \downarrow$ under A2 and the weak A3

From Corollary 6a of Proposition 6, $A \downarrow$, $\psi_c \downarrow$ with ψ_c/A unchanged. Hence, the cutoff firms before the change and those after the change have the same markup rate $\mu(\psi_c/A)$, the same profit $\pi(\psi_c/A)E = F$, and the same revenue, $r(\psi_c/A)E = \sigma(\psi_c/A)\pi(\psi_c/A)E = \sigma(\psi_c/A)F$.

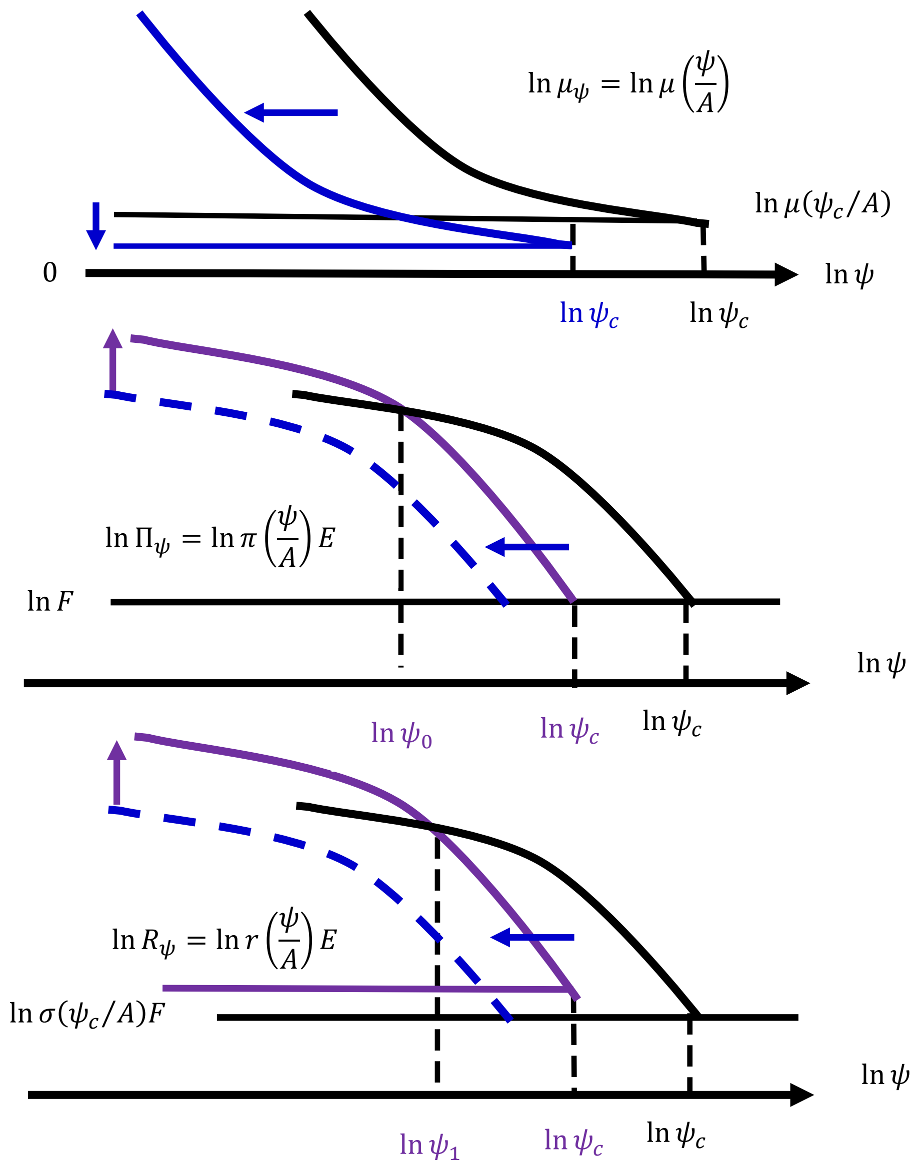


Figure 5b: An increase in E under A2 and the weak A3
From Corollary 6b of Proposition 6, $A \downarrow$, $\psi_c \downarrow$ with $\psi_c/A \uparrow$ and $\sigma(\psi_c/A) \uparrow$. Hence, compared to the cutoff firms before the change, the cutoff firms after the change have a lower markup rate, $\mu(\psi_c/A) \downarrow$, the same profit, $\pi(\psi_c/A)E = F$, and a higher revenue, $r(\psi_c/A)E = \sigma(\psi_c/A)F \uparrow$

From Proposition 7a, the profits are up (down) for $\psi < (>)\psi_0$.
From Proposition 7b, the revenues are up (down) for $\psi < (>)\psi_1$ for a sufficiently small F.

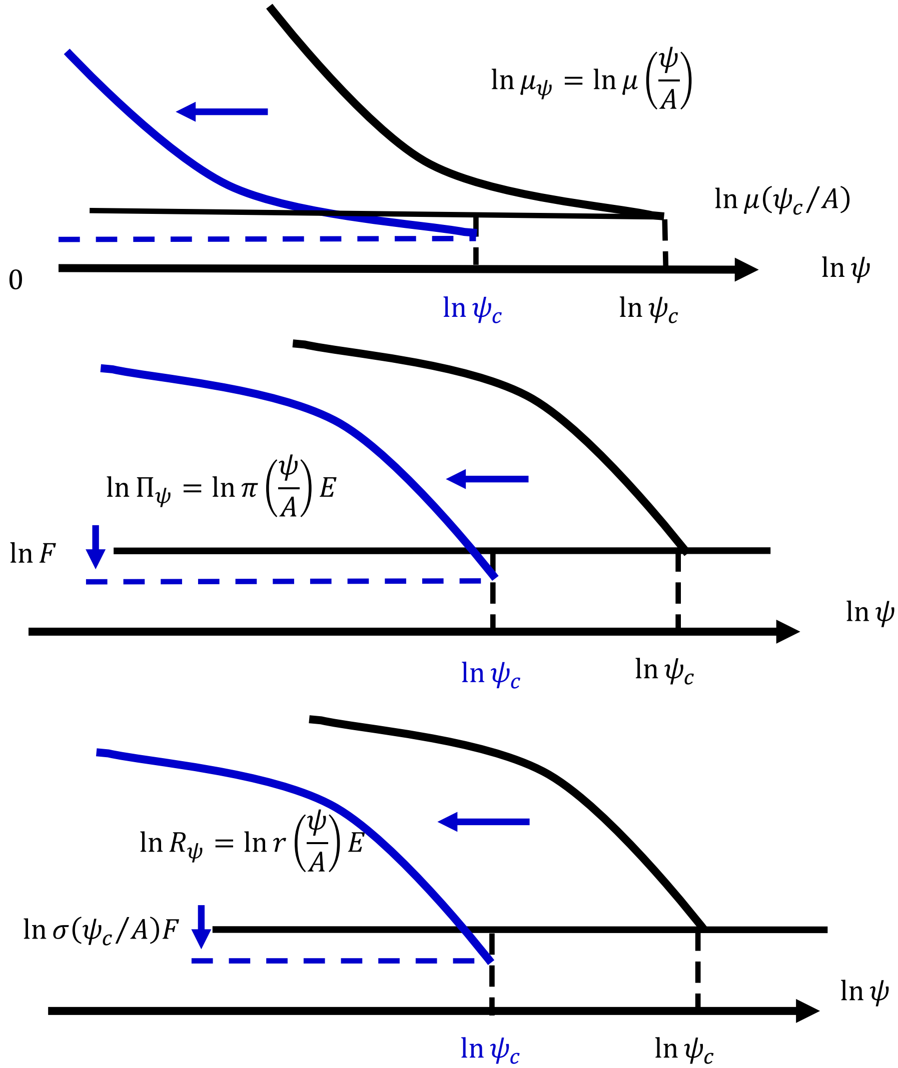


Figure 5c: $F\downarrow$ under A2 and the weak A3 with $\ell'(\cdot) > 0$
From Corollary 6c of Proposition 6, $A\downarrow$, $\psi_c\downarrow$ with $\psi_c/A\uparrow$ and $\sigma(\psi_c/A)\uparrow$. Hence, compared to the cutoff firms before the change, the cutoff firms after the change have a lower markup rate, $\mu(\psi_c/A)\downarrow$, a lower profit, $\pi(\psi_c/A)E = F\downarrow$, and a lower revenue, $r(\psi_c/A)E = \sigma(\psi_c/A)F\downarrow$.

Figure 6: The Limit Case: for $F \to 0$ with $\bar{z} < \infty$.

Figure 6a: $F_e/E \downarrow$ for $F \to 0$ with $\bar{z} < \infty$

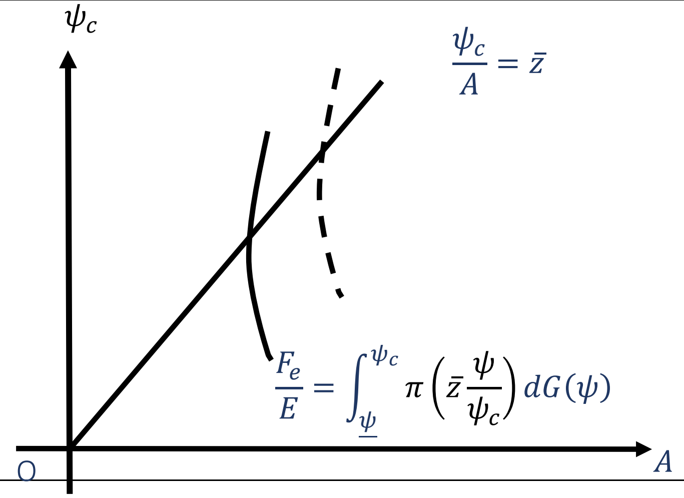


Figure 6b: $F_e/E \downarrow$ for $F \to 0$ with $\bar{z} < \infty$ under A2 and the weak A3

$A \downarrow, \psi_c \downarrow$ with $\psi_c/A = \bar{z}$ unchanged. Hence, the cutoff firms always (i.e., both before and after the change) have $\mu(\psi_c/A) = 1$ and $\pi(\psi_c/A)E = r(\psi_c/A)E = 0$.

In the middle and bottom panels, Blue indicates the effects of $F_e/E \downarrow$ due to $F_e \downarrow$ and Purple indicates the effects of $F_e/E \downarrow$ due to $E \uparrow$