ANNUAL REVIEWS

Annual Review of Economics

Homothetic Non-CES Demand Systems with Applications to Monopolistic Competition

Kiminori Matsuyama

Department of Economics, Northwestern University, Evanston, Illinois, USA;
email: k-matsuyama@northwestern.edu

Annu. Rev. Econ. 2025. 17:261–92

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-081624-082324

Keywords

substitutability versus love-for-variety, equilibrium versus optimal, homothetic single aggregator, second and third laws of demand, firm heterogeneity

Abstract

This article reviews homothetic non-CES (constant elasticity of substitution) demand systems and their implications when applied to monopolistic competition, to offer guidance to those looking for flexible and yet tractable ways of departing from CES. Under general homothetic symmetric non-CES, two measures, substitutability and love-for-variety, are introduced to identify the condition under which the equilibrium product variety is excessive or insufficient. Because homotheticity and symmetry alone impose little restriction to make further progress, we turn to the homothetic single aggregator (HSA) class. HSA is more flexible than CES and translog, which are its special cases, and yet equally analytically tractable, because all cross-variety interactions are summarized by the single aggregator. Under HSA, substitutability is increasing in product variety if and only if Marshall's second law holds, which is a sufficient condition for love-for-variety to be diminishing in product variety and for the equilibrium product variety to be excessive. Monopolistic competition under HSA remains tractable even under various forms of firm heterogeneity and in multi-market settings.

1. INTRODUCTION

CES: constant elasticity of substitution

MC: monopolistically competitive/ monopolistic competition

We all know the constant elasticity of substitution (CES) demand system. It is ubiquitous in business cycle theory, economic growth, international trade, spatial economics, and other applied general equilibrium fields. We love using CES, because it has many knife-edge properties, which help to make it tractable. At the same time, these knife-edge properties make CES too restrictive for many applications. Of course, many researchers have tried some non-CES demand systems, but they typically look for alternatives, such as linear-quadratic or translog, which come with their own drawbacks and limitations. What is needed is to generalize CES by relaxing some of its knife-edge properties in order to have more flexibility without losing too much tractability of CES.

A previous review in the *Annual Review of Economics*, titled "Non-CES Aggregators: A Guided Tour" (Matsuyama 2023), reviewed several classes of non-CES demand systems and offered some guidance to those looking for flexible and yet tractable ways of departing from CES. Due to space limitations, however, it focused on non-CES that are suited for applying to intersectoral demand systems, with special emphasis on nonhomotheticity and gross complementarities across sectors and the essentiality of goods and factors.

This review focuses instead on applications of homothetic non-CES to demand systems for differentiated products within a monopolistically competitive (MC) industry with free entry and endogenous product variety. This necessitates some additional restrictions on the class of demand systems studied, listed below.

- Endogenous range of inessential products: To allow for firms to enter or exit with their own products, demand systems need to be well-defined even when some products are unavailable or not yet invented.
- Gross substitutability across products: That is, the price elasticity of demand for each product is greater than 1, or, equivalently, the market share of each product is decreasing in its own price. This ensures that MC firms face a positive marginal revenue curve.

I will further impose two restrictions:

- Continuum of products: This not only helps tractability by making product variety a continuous variable but also ensures that individual firms, unless they produce a positive measure of product varieties, cannot affect (and hence do not worry about the potential impacts of their actions on) the industry-level variables, one of the defining features of MC that distinguishes it from oligopoly.
- Symmetric demand systems: This helps to highlight the supply-side heterogeneity across firms, such as productivity difference à la Melitz (2003), differential market access across firms that are based in different locations (as in many models of international trade and spatial economics), the price dispersion created by the price setting mechanism à la Calvo (1983), and technology diffusion that causes some but not all MC firms to face competitive fringes à la Judd (1985).

The restriction of homotheticity and symmetry is imposed mostly due to page limitations.[1] Nevertheless, the reader should also note that homothetic and symmetric demand systems are not so

[1]There are at least two more reasons for focusing on homotheticity. First, most earlier studies of MC under non-CES make use of nonhomothetic symmetric demand systems. For example, Dixit & Stiglitz (1977, section II), Behrens & Murata (2007), Zhelobodko et al. (2012), Mrázová & Neary (2017), Latzer et al. (2020), and Matsuyama & Ushchev (2024c) use the directly explicitly additive (DEA) class of nonhomothetic symmetric demand systems. The indirectly explicitly additive (IEA) class used by Bertoletti & Etro (2017), Boucekkine et al. (2017), and Matsuyama & Ushchev (2024b), as well as the linear-quadratic demand system used by

restrictive as they may seem, because one can nest them into a nonhomothetic and/or asymmetric upper-tier demand system, as in a multi-sector model of Matsuyama (2019). In other words, homothetic symmetric non-CES can serve as building blocks to construct such nonhomothetic and/or asymmetric non-CES.

HSA: homothetic single aggregator

Here is the road map of this review. Section 2 offers a quick refresher on CES and its application to what I call the Dixit & Stiglitz (1977) environment, where MC firms are symmetric not only on the demand side but also on the supply side. Section 3 discusses general homothetic symmetric demand systems. Among others, this section introduces two measures, substitutability, $\sigma(V)$, and love-for-variety, $\mathcal{L}(V)$, both as functions of product variety V. These two measures help to characterize the demand system. Section 4 applies these general demand systems to the Dixit-Stiglitz environment. It characterizes the symmetric equilibrium under the assumption that it exists uniquely and conducts comparative statics that depend on $\sigma(V)$ but not on $\mathcal{L}(V)$. On the other hand, the optimal allocation depends on $\mathcal{L}(V)$ but not on $\sigma(V)$. By comparing the equilibrium and the optimum, this section identifies the sufficient and necessary condition under which the equilibrium product variety, V^{eq}, is excessive or insufficient relative to the optimal product variety, V^{op}. Yet, it is not possible to make further progress under general homothetic symmetric demand systems, because homotheticity and symmetry alone impose little restriction on the relation between $\sigma(V)$ and $\mathcal{L}(V)$, so that "almost anything goes."

Section 5 thus turns to the subclass of homothetic symmetric demand systems called homothetic single aggregator (HSA). This class of demand systems, which contains CES and translog (Feenstra 2003) as special cases, is characterized by the presence of a single price aggregator, a sufficient statistic that captures everything one needs to know to understand the cross-variety interactions. Due to such a significant reduction in dimensionality, HSA is highly tractable yet more flexible than CES and translog. Moreover, it imposes tighter relations between $\sigma(V)$ and $\mathcal{L}(V)$. Under HSA, Marshall's second law of demand (i.e., the price elasticity of demand is increasing in its own price) is equivalent to increasing substitutability, $\sigma'(V) > 0$, and both are sufficient for diminishing love-for-variety, $\mathcal{L}'(V) < 0$. Armed with these results, Section 6 applies HSA to the Dixit-Stiglitz environment. Under HSA, it is straightforward to show that the equilibrium is unique and symmetric. Moreover, the equilibrium product variety is excessive under diminishing love-for-variety. Therefore, Marshall's second law and, equivalently, increasing substitutability are sufficient for excessive product variety under HSA.[2]

Ottaviano et al. (2002) and Melitz & Ottaviano (2008), is also nonhomothetic. This literature has been reviewed by Parenti et al. (2017) and Thisse & Ushchev (2018). Melitz (2018) also reviewed the work using the DEA class. A second reason for focusing on homotheticity is that 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 procedures. 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 restrictions, such as assuming that there is only one outside sector that produces a homogeneous good competitively or that every sector has the same parametric family of nonhomothetic demand systems with identical parameter values.

[2]Matsuyama & Ushchev (2020a, 2023) showed that many results in Sections 5 and 6 below hold also in two other classes of homothetic symmetric demand systems: symmetric homothetic direct implicit additivity (HDIA) with gross substitutes, an extension of the Kimball (1995) aggregator with an endogenous product range, and symmetric homothetic indirect implicit additivity (HIIA) with gross substitutes. The three classes, HSA, HDIA, and HIIA, which were originally developed by Matsuyama & Ushchev (2017) without symmetry and gross substitutes restrictions, all share CES as a special case but are otherwise pairwise disjoint. HDIA and HIIA are less tractable than HSA. Some additional restrictions are needed just to ensure the uniqueness and

Then, Section 7 applies HSA to the Melitz (2003) environment, where ex-ante symmetric firms learn their marginal costs after entry, drawn from a common distribution, and become ex-post heterogeneous. Again, under HSA, it is straightforward to show that the equilibrium exists uniquely and to conduct comparative statics. Section 8 discusses how HSA can accommodate other types of firm heterogeneity. **Supplemental Appendix 1** explains why HSA is more tractable than homothetic direct implicit additivity (HDIA) and homothetic indirect implicit additivity (HIIA). **Supplemental Appendix 2** lists some parametric families of HSA for the quantitatively oriented reader who may want to use them for calibration and estimation.

Supplemental Material >

HDIA: homothetic direct implicit additivity

HIIA: homothetic indirect implicit additivity

TFP: total factor productivity

Before proceeding, some caveats should be mentioned. First, this is a review of non-CES and of their key implications when applied to MC. My goal is to offer guidance to those who are looking for tractable and yet flexible ways of departing from CES in their applications, and I hope that the readers will find useful building blocks for constructing their own models. However, this is not intended to be a review of applications of MC under non-CES to some topics in economics, whether they are in international trade, economic geography, economic growth, or Keynesian macroeconomics. Such a review needs a separate treatment for each topic, some of which I hope to write in the future. Second, because the materials reviewed here are theoretical in nature, I try not to sacrifice the logical rigor and yet not to be bogged down in technicalities. I offer the intuition and explain the logics behind the main results but skip many derivations. Furthermore, some regularity conditions, such as continuity and differentiability, are often not explicitly stated. Moreover, the space limitations prevent me from discussing any empirical evidence that motivates some assumptions. This review should thus be treated as a reading guide for the references cited, not as a substitute for reading them. Finally, I have encountered repeatedly through the years several false claims about non-CES demand systems. Often taken for granted, these false claims are not only found in published and discussion papers but also heard in seminars, coming both from the speakers and from those in the audience. I have also seen them in referee reports, both as an editor and as a submitting author. In this review, I explicitly discuss several common fallacies and explain why they are wrong but without citing any references. They are so widespread that I have no idea who should be given "credit" for starting each fallacy. Indeed, many of them are a kind of logical pitfall that anyone could fall into. (I confess that I used to believe Fallacies 3 and 4, discussed in Section 3, myself.) By flagging these fallacies without finger-pointing, I am hoping to prevent misinformation from spreading, particularly, to the new generations of researchers.

2. DIXIT-STIGLITZ UNDER CES: A QUICK REFRESHER

We discuss CES demand systems in terms of demand for differentiated intermediate inputs generated by a competitive industry that produces a single final good, using the symmetric CES production function

$$X = X(\mathbf{x}) = Z\left[\int_{\Omega} (x_\omega)^{1-\frac{1}{\sigma}}\, d\omega\right]^{\frac{\sigma}{\sigma-1}},$$

where $\mathbf{x} = \{x_\omega; \omega \in \Omega\}$ is an input quantity vector; Ω is the set of input varieties, indexed by ω, that are available in equilibrium, whose mass is denoted by $V \equiv |\Omega|$. Under CES, the elasticity of substitution across varieties is a parameter, $\sigma > 1$, and $Z > 0$ is total factor productivity (TFP).

the symmetry of the equilibrium in the Dixit-Stiglitz environment. Moreover, these two are not analytically tractable with firm heterogeneity. This is because the cross-variety interactions are captured by two aggregators under HDIA and HIIA, in contrast to one aggregator under HSA. For these reasons, I focus on HSA from Section 5 on.

2.1. CES Demand System

Facing $\mathbf{p} = \{p_\omega; \omega \in \Omega\}$, the input price vector, the competitive industry chooses $\mathbf{x}$ to minimize the production cost, which leads to the unit cost function

$$P(\mathbf{p}) \equiv \min_{\mathbf{x}} \left\{ \mathbf{px} \equiv \int_\Omega p_\omega x_\omega \mathrm{d}\omega \middle| X(\mathbf{x}) \geq 1 \right\} = \frac{1}{Z} \left[\int_\Omega (p_\omega)^{1-\sigma} \mathrm{d}\omega \right]^{\frac{1}{1-\sigma}},$$

with demand for ω

$$x_\omega = \left(\frac{p_\omega}{ZP(\mathbf{p})} \right)^{-\sigma} \frac{X(\mathbf{x})}{Z} = \frac{(p_\omega)^{-\sigma}}{(ZP(\mathbf{p}))^{1-\sigma}} E$$

and the budget share of ω

$$s_\omega \equiv \frac{p_\omega x_\omega}{P(\mathbf{p}) X(\mathbf{x})} = \frac{p_\omega x_\omega}{E} = \left(\frac{p_\omega}{ZP(\mathbf{p})} \right)^{1-\sigma} = \left(\frac{Z x_\omega}{X(\mathbf{x})} \right)^{1-\frac{1}{\sigma}},$$

where $E \equiv P(\mathbf{p})X(\mathbf{x}) = \mathbf{px}$ is the size of this industry and hence the market size for differentiated inputs, which we treat as given. From the well-known duality principle, $X(\mathbf{x})$ can be recovered from $P(\mathbf{p})$ as

$$X(\mathbf{x}) \equiv \min_{\mathbf{p}} \left\{ \mathbf{px} \equiv \int_\Omega p_\omega x_\omega \mathrm{d}\omega \middle| P(\mathbf{p}) \geq 1 \right\} = Z \left[\int_\Omega (x_\omega)^{1-\frac{1}{\sigma}} \mathrm{d}\omega \right]^{\frac{\sigma}{\sigma-1}}.$$

2.2. The Dixit-Stiglitz Environment

We now apply CES to what I shall call the Dixit-Stiglitz environment. There exists a single primary factor of production, labor, taken as numeraire. Each differentiated intermediate input, $\omega \in \Omega$, is produced from labor and sold exclusively by a single MC firm, also indexed by $\omega \in \Omega$. These MC firms are symmetric: Not only do their products enter symmetrically in the demand system, but they also share the same technology.[3] Each firm needs to hire $F + \psi x_\omega$ units of labor to supply x_ω units of its own product. Here F is the fixed cost, a combination of the entry/innovation cost required to develop its own product and to enter the market and of the overhead cost required to stay in the market; ψx_ω is the production cost, or employment, where ψ is a constant marginal cost of production, and the inverse of productivity. Finally, there is free entry to the market. Firms enter/exit until their gross profit is equalized to the fixed cost, $\Pi_\omega = F$. This ensures that there is no excess profit in equilibrium and that the total labor demand of this sector is $L = \mathbf{px} = P(\mathbf{p})X(\mathbf{x}) = E$.[4]

2.3. Equilibrium

As the sole producer of its own product, each MC firm sets its price, p_ω, to maximize its gross profit,

$$\Pi_\omega = (p_\omega - \psi) x_\omega = \frac{(p_\omega - \psi)(p_\omega)^{-\sigma}}{(ZP(\mathbf{p}))^{1-\sigma}} E,$$

[3]Being symmetric in both demand and supply sides, MC firms in the Dixit-Stiglitz environment are often referred to as homogeneous or representative firms, even though what they supply is product-differentiated.

[4]Notice that no assumption is made concerning how this sector interacts with the rest of the economy, except that E is the aggregate spending on this sector, which leads to this sector's labor demand, $L = E$. Of course, one could assume that 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, the sector-level analysis in this review does not need to make such an assumption.

holding the industry-wide variables, $P(\mathbf{p})$ and E, fixed. The first-order condition of the profit maximization leads to the familiar Lerner pricing formula and the markup rule

$$p_\omega\left(1-\frac{1}{\sigma}\right)=\psi \;\Leftrightarrow\; p_\omega\equiv p=\left(\frac{\sigma}{\sigma-1}\right)\psi\equiv\mu\psi,$$

where μ is the constant and common markup rate. Thus, all firms set the same price, and the equilibrium is symmetric. By dropping the index to denote the common values, we have $p_\omega = p$ and $x_\omega = x$, which implies $pxV = E$. Thus, the common gross profit is $\Pi = (p-\psi)x = px/\sigma = E/\sigma V$. Finally, the free-entry/exit condition implies that the common gross profit is equal to the fixed cost in equilibrium, $E/\sigma V^{\text{eq}} = F$, so that

$$V^{\text{eq}}=\frac{E}{\sigma F};\; p^{\text{eq}}=\left(\frac{\sigma}{\sigma-1}\right)\psi;\; x^{\text{eq}}=\frac{(\sigma-1)F}{\psi}.$$

Thus, in equilibrium, the revenue $p^{\text{eq}}x^{\text{eq}} = E/V^{\text{eq}} = \sigma F$ is divided into the (gross) profit, $(p^{eq}-\psi)x^{eq} = F$, and the production cost, $\psi x^{eq} = (\sigma-1)F$, in every firm. Notice that the firm's revenue, profit, and production cost are all independent of ψ. Moreover, the profit and production cost shares in revenue,

$$\frac{F}{p^{\text{eq}}x^{\text{eq}}}=\frac{1}{\sigma};\; \frac{\psi x^{\text{eq}}}{p^{\text{eq}}x^{\text{eq}}}=1-\frac{1}{\sigma}=\frac{1}{\mu},$$

and the profit/production cost ratio,

$$\frac{F}{\psi x^{\text{eq}}}=\frac{\mu}{\sigma}=\frac{1}{\sigma-1}=\mu-1,$$

are all constant and independent of E/F under CES.

2.4. Comparative Statics

By denoting the percentage change by $\hat{q}\equiv\partial\ln q=\partial q/q$, we can show that the three endogenous variables V^{eq}, p^{eq}, and x^{eq} respond to the three exogenous variables E, F, and ψ as

$$\widehat{V^{\text{eq}}}=\hat{E}-\hat{F},\widehat{p^{\text{eq}}}=\hat{\psi},\text{ and }\widehat{x^{\text{eq}}}=\hat{F}-\hat{\psi}.$$

Note that the firm behavior, $(p^{\text{eq}},\ x^{\text{eq}})$, is not affected by E, while the mass of firms, V^{eq}, responds proportionally to E. Thus, the adjustment to a market size change takes place only at the extensive margin under CES.

2.5. Optimality of the Equilibrium Allocation

Now imagine that this sector were fully integrated and could control all intermediate inputs production. Then, we would have

$$\max X(\mathbf{x})=\max Z\left[\int_\Omega (x_\omega)^{1-\frac{1}{\sigma}}\mathrm{d}\omega\right]^{\frac{\sigma}{\sigma-1}}\quad\text{such that}\int_\Omega \psi x_\omega \mathrm{d}\omega+VF\le E.$$

The optimal allocation is clearly symmetric, $x_\omega = x > 0$ for $\omega\in\Omega$, simplifying the problem to

$$\max_{(\psi x+F)V\le E} V^{\frac{\sigma}{\sigma-1}}(Zx)=\frac{ZF}{\psi}\max_V V^{\frac{1}{\sigma-1}}\left(\frac{E}{F}-V\right).$$

By solving this problem, the optimal allocation is given by

$$V^{\text{op}}=\frac{E}{\sigma F},\; x^{\text{op}}=\frac{(\sigma-1)F}{\psi},$$

which is identical with the equilibrium allocation.

The optimality result is surprising. A priori, one would expect that MC equilibrium would not be optimal due to the presence of externalities. First, there are negative externalities due to the business stealing effect: A firm, when paying the fixed cost to enter and stay with its own product, does not take into account that this action reduces demand for other products and their profits, which would suggest excessive product variety. On the other hand, there are positive externalities due to incomplete appropriability: A firm is motivated to produce and sell its own variety not by the social surplus but by the profit, which is a fraction of the social surplus. This would suggest insufficient product variety. As explained by Tirole (1988, chapter 7) and Matsuyama (1995, section 3E), these two sources of externalities cancel out each other under CES, which is why the equilibrium is optimal. I show in Section 4 how departing from CES in the Dixit-Stiglitz environment could break the optimality. This feature makes the Dixit-Stiglitz environment a useful benchmark against which the efficiency implications of non-CES can be evaluated.[5]

CRS: constant returns to scale

Unfortunately, the logic behind the optimality result under CES is poorly understood.

Fallacy 1. The equilibrium allocation is optimal because all the products are sold at the same markup rate, and hence the relative prices across products are not distorted.

It is easy to see why this is false. If this logic were correct, the equilibrium would be optimal as long as all products were sold at the same markup rate, and it would not have to be equal to $\sigma/(\sigma-1)$. Indeed, any symmetric equilibrium would be optimal, even if the demand system were non-CES and/or in the presence of a uniform taxation on differentiated products. The logic is incorrect, because the common markup rate merely ensures that the allocation across available products is not distorted; it does not ensure that the equilibrium incentive to make another product available is optimal.

Fallacy 2. The equilibrium allocation is optimal if and only if it is under CES.

This is the polar opposite of Fallacy 1. Of course, the optimality under CES is not robust because it must satisfy the knife-edge condition, the two sources of externalities canceling out each other. However, CES is not unique in this respect, as explained in Section 4.

3. GENERAL HOMOTHETIC SYMMETRIC DEMAND SYSTEMS[6]

Let us now assume that the industry uses symmetric production technologies, specified either as the CRS production function, $X(\mathbf{x})$, which satisfies linear homogeneity, monotonicity, and strict quasi-concavity, or as its corresponding unit cost function,

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

where $\mathbf{x} = \{x_\omega;\, \omega \in \overline{\Omega}\}$, the input quantity vector, and $\mathbf{p} = \{p_\omega;\, \omega \in \overline{\Omega}\}$, the input price vector, are now defined over $\overline{\Omega}$, the set of all potential input varieties, so that $\Omega \subset \overline{\Omega}$, the set of available input varieties, with $V \equiv |\Omega|$. Thus, $\overline{\Omega}\backslash\Omega$ is the set of unavailable varieties, with $x_\omega = 0$ and $p_\omega = \infty$ for $\omega \in \overline{\Omega}\backslash\Omega$. To ensure the feasibility of production, we need to assume that inputs are inessential,

[5] Of course, we can also break the optimality by changing the environment while keeping CES. For example, the equilibrium is no longer optimal if producing intermediate inputs needs not only labor but also the final good, if the taxation is added, or if labor is differentiated and MC firms face an upward-sloping labor supply curve, which gives them monopsony power, allows them to set markdown in the labor market, etc. One could also change the environment to ensure the optimality of the equilibrium allocation under any symmetric demand systems, if the resource used in F is in fixed supply and is not used in the production.

[6] This section draws heavily from Matsuyama & Ushchev (2023).

i.e., $\overline{\Omega}\backslash\Omega \neq \emptyset$ does not imply $X(\mathbf{x}) = 0 \Leftrightarrow P(\mathbf{p}) = \infty$. Recall that, from the duality principle, the production function $X(\mathbf{x})$ can be recovered from $P(\mathbf{p})$ as

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

so that we could use either $P(\mathbf{p})$ or $X(\mathbf{x})$ as the primitive of this CRS production technologies.

3.1. Demand Systems

The demand curve and the inverse demand curve for $\omega \in \Omega$ are

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

from either of which Euler's homogeneous function theorem implies

$$\mathbf{px} = \int_{\Omega} p_\omega x_\omega d\omega = \int_{\Omega} p_\omega \frac{\partial P(\mathbf{p})}{\partial p_\omega} X(\mathbf{x})\, d\omega = \int_{\Omega} P(\mathbf{p}) \frac{\partial X(\mathbf{x})}{\partial x_\omega} x_\omega d\omega = P(\mathbf{p}) X(\mathbf{x}) = E.$$

The budget share of $\omega \in \Omega$, $s_\omega \equiv p_\omega x_\omega / P(\mathbf{p}) X(\mathbf{x})$, can be thus written as a homogeneous function of degree zero both in price and in quantity:

$$s_\omega = \frac{\partial \ln P(\mathbf{p})}{\partial \ln p_\omega} \equiv s(p_\omega, \mathbf{p}) = s(1, \mathbf{p}/p_\omega)\,;\, s_\omega = \frac{\partial \ln X(\mathbf{x})}{\partial \ln x_\omega} \equiv s^*(x_\omega, \mathbf{x}) = s^*(1, \mathbf{x}/x_\omega).$$

From now on, we also impose gross substitutability,

$$\frac{\partial \ln s(p_\omega; \mathbf{p})}{\partial \ln p_\omega} < 0 \Leftrightarrow \frac{\partial \ln s^*(x_\omega;\; \mathbf{x})}{\partial \ln x_\omega} > 0,$$

which ensures that the firm selling $\omega \in \Omega$ faces the positive marginal revenue curve.[7]

The price elasticity of demand for $\omega \in \Omega$, $\zeta_\omega \equiv -\partial \ln x_\omega / \partial \ln p_\omega$, can be also written as a homogeneous function of degree zero in prices or in quantities:

$$\zeta_\omega = 1 - \frac{\partial \ln s(p_\omega; \mathbf{p})}{\partial \ln p_\omega} \equiv \zeta(p_\omega; \mathbf{p}) = \zeta(1, \mathbf{p}/p_\omega) > 1;$$

$$\zeta_\omega = \left[1 - \frac{\partial \ln s^*(x_\omega; \mathbf{x})}{\partial \ln x_\omega}\right]^{-1} \equiv \zeta^*(x_\omega; \mathbf{x}) = \zeta^*(1, \mathbf{x}/x_\omega) > 1.$$

Notice that the restriction of gross substitutability is equivalent to the restriction that the price elasticity should always be greater than 1. In general, the price elasticity can be increasing or decreasing in its own price. The literature typically focuses on the increasing case,

$$\frac{\partial \ln \zeta(p_\omega; \mathbf{p})}{\partial \ln p_\omega} > 0 \Leftrightarrow \frac{\partial \ln \zeta^*(x_\omega;\; \mathbf{x})}{\partial \ln x_\omega} < 0.$$

This is Marshall's second law of demand, or the second law for short. For the case where the price elasticity is decreasing, we say the anti-second law holds. Clearly, CES is the borderline case. All other examples listed in **Supplemental Appendix 2** satisfy the second law.

Supplemental Material >

[7] Under CES, $\sigma > 1$ ensures both the inessentiality and the gross substitutability of inputs. In general, the inessentiality and gross substitutability are different concepts and need to be assumed separately.

Note that the budget share of $\omega \in \Omega$, s_ω, and its price elasticity of demand, ζ_ω, are both functions of $\mathbf{p}/p_\omega$ or $\mathbf{x}/x_\omega$. Of course, symmetry implies that they are invariant of permutation, but they still depend on the entire distribution of the prices (or the quantities) relative to the price (or quantity) of $\omega \in \Omega$, which is infinite dimensional. This suggests that the cross-variety interactions could be complicated under general homothetic symmetric demand systems.

3.2. Substitutability and Love-for-Variety Measures

We now introduce two measures that help characterize general homothetic symmetric demand systems. First, define the unit quantity vector,

$$\mathbf{1}_\Omega \equiv \{(1_\Omega)_\omega; \omega \in \overline{\Omega}\}, \qquad \text{where} \qquad (1_\Omega)_\omega \equiv \begin{cases} 1 & \text{for} \quad \omega \in \Omega \\ 0 & \text{for} \quad \omega \in \overline{\Omega}\backslash\Omega \end{cases},$$

which is the indicator function of Ω, and the unit price vector,

$$\mathbf{1}_\Omega^{-1} \equiv \{(1_\Omega^{-1})_\omega; \omega \in \overline{\Omega}\}, \qquad \text{where} \qquad (1_\Omega^{-1})_\omega \equiv \begin{cases} 1 & \text{for} \quad \omega \in \Omega \\ \infty & \text{for} \quad \omega \in \overline{\Omega}\backslash\Omega \end{cases}.$$

Clearly, $\int_\Omega (1_\Omega)_\omega d\omega = \int_\Omega (1_\Omega^{-1})_\omega d\omega = |\Omega| \equiv V$. Moreover, at the symmetric patterns, $\mathbf{p} = p\mathbf{1}_\Omega^{-1}$ and $\mathbf{x} = x\mathbf{1}_\Omega$, the budget share of each variety is simply

$$s_\omega = s(1, \mathbf{p}/p_\omega) = s^*(1, \mathbf{x}/x_\omega) = s\left(1, \mathbf{1}_\Omega^{-1}\right) = s^*(1, \mathbf{1}_\Omega) = 1/V,$$

and the price elasticity of each variety,

$$\zeta_\omega = \zeta(1, \mathbf{p}/p_\omega) = \zeta^*(1, \mathbf{x}/x_\omega) = \zeta\left(1, \mathbf{1}_\Omega^{-1}\right) = \zeta^*(1, \mathbf{1}_\Omega) > 1,$$

is also a function of V only, hence can be denoted as $\sigma(V)$. Furthermore, as shown by Matsuyama & Ushchev (2023), $\zeta(1, \mathbf{1}_\Omega^{-1}) = \zeta^*(1, \mathbf{1}_\Omega)$ is equal to the Allen-Uzawa elasticity of substitution[8] between any pair, ω and $\omega' \in \Omega$, evaluated at $\mathbf{p} = p\mathbf{1}_\Omega^{-1}$. Hence, we have the following definition.

Definition. The substitutability measure across varieties is defined by

$$\sigma(V) \equiv \zeta\left(1; \mathbf{1}_\Omega^{-1}\right) = \zeta^*(1; \mathbf{1}_\Omega) > 1.$$

Note that $\sigma(V) > 1$ is guaranteed by gross substitutability. If $\sigma'(V) > (<)0$, we call this the case of increasing (decreasing) substitutability. In general, $\sigma(V)$ may be nonmonotonic in V.

Love-for-variety is commonly defined by the rate of productivity gain from a higher V, at $\mathbf{x} = x\mathbf{1}_\Omega$, holding xV constant:

$$\left.\frac{d\ln X(\mathbf{x})}{d\ln V}\right|_{\mathbf{x}=x\mathbf{1}_\Omega, xV=\text{const.}} = \left.\frac{d\ln xX(\mathbf{1}_\Omega)}{d\ln V}\right|_{xV=\text{const.}} = \frac{d\ln X(\mathbf{1}_\Omega)}{d\ln V} - 1.$$

Since $X(\mathbf{1}_\Omega)$ is a function of V only, so is this measure, and hence it can be denoted as $\mathcal{L}(V)$.An alternative (and my favorite) definition of love-for-variety is the rate of decline in $P(\mathbf{p})$ from a higher V, at $\mathbf{p} = p\mathbf{1}_\Omega^{-1}$, holding p constant:

$$-\left.\frac{d\ln P(\mathbf{p})}{d\ln V}\right|_{\mathbf{p}=p\mathbf{1}_\Omega^{-1}, p=\text{const.}} = -\frac{d\ln P\left(\mathbf{1}_\Omega^{-1}\right)}{d\ln V}.$$

Since $P(\mathbf{1}_\Omega^{-1})$ is a function of V only, so is this measure.

[8]Because there is a continuum of inputs, there is no point in looking into the Morishima elasticity of substitution.

The two definitions are indeed equivalent. This can be verified by setting $\mathbf{x} = x\mathbf{1}_\Omega$ and $\mathbf{p} = p\mathbf{1}_\Omega^{-1}$ to $\mathbf{px} = P(\mathbf{p})X(\mathbf{x})$:

$$pxV = pP\left(\mathbf{1}_\Omega^{-1}\right)xX(\mathbf{1}_\Omega) \Rightarrow -\frac{\mathrm{d}\ln P\left(\mathbf{1}_\Omega^{-1}\right)}{\mathrm{d}\ln V} = \frac{\mathrm{d}\ln X(\mathbf{1}_\Omega)}{\mathrm{d}\ln V} - 1.$$

Hence, we have the following definition.

Definition. The love-for-variety measure is defined by

$$\mathcal{L}(V) \equiv -\frac{\mathrm{d}\ln P\left(\mathbf{1}_\Omega^{-1}\right)}{\mathrm{d}\ln V} = \frac{\mathrm{d}\ln X(\mathbf{1}_\Omega)}{\mathrm{d}\ln V} - 1 > 0.$$

Note that $\mathcal{L}(V) > 0$ is guaranteed by the strict quasi-concavity of the production technologies. If $\mathcal{L}'(V) > (<)0$, we call this the case of increasing (diminishing) love-for-variety. In general, $\mathcal{L}(V)$ may be nonmonotonic in V.

Under CES, we have the following:

- The price elasticity of demand is constant: $(p_\omega; \mathbf{p}) = \zeta^*(x_\omega; \mathbf{x}) = \sigma$;
- Substitutability is constant: $\sigma(V) = \sigma$; and
- Love-for-variety is constant and $\mathcal{L}(V) = \mathcal{L} = 1/(\sigma - 1)$, which is inversely related to σ.[9]

Under general homothetic symmetric demand systems, however, we can say little about the relation between $\zeta(p_\omega; \mathbf{p}) = \zeta^*(x_\omega; \mathbf{x})$, $\sigma(V)$ and $\mathcal{L}(V)$, even though the following claims are often made.

Fallacy 3. $\sigma(V)$ is constant only under CES.

Fallacy 4. $\sigma'(V) > (<)0$ if and only if the second law (anti-second law) holds.

Fallacy 5. $\sigma(V)$ is an inverse measure of love-for-variety, $\mathcal{L}(V)$.[10]

I refer readers to Matsuyama & Ushchev (2023, 2024a) for some counterexamples. Symmetry and homotheticity alone are not strong enough to impose much restriction, because $\sigma(V)$ depends only on the local properties of the demand system, while $\mathcal{L}(V)$ depends on its global properties, and the budget share and the price elasticity of each variety can depend on the entire distribution of prices across different varieties. Nevertheless, one might find that the claims made in these fallacies are appealing features for a demand system to have. Even though these claims are false under general homothetic symmetric demand systems, they are true under HSA, as shown in Section 5.

[9]Benassy (1996) proposed to break the tight relation between $\sigma(V)$ and $\mathcal{L}(V)$ under the standard CES by making TFP a function of V as $Z(V)$, justified by some sorts of direct externalities from V to TFP (or affinity in the context of spatial economics). Such modified CES yields $\mathcal{L}(V) = \partial \ln Z(V)/\partial \ln V + 1/(\sigma - 1)$. This allows the gap between the observed love-for-variety and the love-for-variety implied by CES demand to be accounted for by $\partial \ln Z(V)/\partial \ln V$, the term I would call "the Benassy residual," in analogy with the Solow residual in the growth accounting. Moreover, he assumed that $\partial \ln Z(V)/\partial \ln V = \nu - 1/(\sigma - 1)$, so that $\mathcal{L}(V) = \nu$, which can be chosen independently from $\sigma(V) = \sigma$. If we assume instead that $\partial \ln Z(V)/\partial \ln V$ is another parameter independent of $\sigma(V) = \sigma$, $\mathcal{L}(V)$ is still inversely related to $\sigma(V) = \sigma$. Even if one believed in the presence of direct externalities from V to TFP or to affinity, any estimate of the Benassy residual hinges on the CES assumption. In any case, introducing the Benassy residual does not serve our goal of characterizing MC models under homothetic non-CES demand systems.

[10]Though many have derived $\sigma(V)$ for specific non-CES demand systems, I am unaware of any attempt prior to Matsuyama & Ushchev's (2023) to derive $\mathcal{L}(V)$ for any non-CES. I suspect that those who made this claim just take it for granted that $\mathcal{L} = 1/(\sigma - 1)$ under CES would be generalized to $\mathcal{L}(V) = 1/(\sigma(V) - 1)$ under non-CES.

4. DIXIT-STIGLITZ UNDER GENERAL HOMOTHETIC DEMAND SYSTEMS[11]

Let us now apply the general homothetic symmetric demand system to the Dixit-Stiglitz environment.

Fallacy 6. With the symmetric firms, the equilibrium is symmetric.

In general, the symmetry of the model environment only ensures the symmetry of the set of equilibria, not the symmetry of any equilibrium (this is called symmetry-breaking; see Matsuyama 2008). Even if the symmetric equilibrium exists, it may coexist with a symmetric set of asymmetric equilibriums. In an asymmetric equilibrium in the Dixit-Stiglitz environment, ex-ante symmetric firms pursue different pricing strategies, where some choose to have higher markup rates with smaller quantities while others choose to have lower markup rates with larger quantities, and the masses of firms choosing between the two strategies adjust in such a way that they are indifferent between the two, so that firms become endogenously asymmetric, giving rise to an endogenous price distribution.

4.1. Symmetric Equilibrium

Nevertheless, let us proceed under the assumption that a symmetric equilibrium exists. Each firm chooses p_ω to maximize its gross profit,

$$\Pi_\omega = (p_\omega - \psi)x_\omega = \left(1 - \frac{\psi}{p_\omega}\right)p_\omega x_\omega = \left(1 - \frac{\psi}{p_\omega}\right)s(p_\omega, \mathbf{p})E,$$

holding E and $\mathbf{p}$ as given. The first-order condition generates the Lerner pricing formula,

$$p_\omega\left(1 - \frac{1}{\zeta(p_\omega; \mathbf{p})}\right) = \psi.$$

In any symmetric equilibrium, $\mathbf{p} = p\mathbf{1}_\Omega^{-1}$, $\zeta(p_\omega; \mathbf{p}) = \zeta(1, \mathbf{1}_\Omega^{-1}) = \sigma(V)$. Hence, we have

$$p_\omega\left(1 - \frac{1}{\sigma(V)}\right) = \psi \;\Leftrightarrow\; p_\omega \equiv p = \frac{\sigma(V)}{\sigma(V) - 1}\psi \equiv \mu(V)\psi,$$

where the markup rate, $\mu(V)$, satisfies the following identities:

$$\frac{1}{\sigma(V)} + \frac{1}{\mu(V)} = 1;\; \frac{1}{\sigma(V) - 1} = \frac{\mu(V)}{\sigma(V)} = \mu(V) - 1$$

and[12]

$$\mathcal{E}_\sigma(V) = -\frac{\mathcal{E}_\mu(V)}{\mu(V) - 1};\; \mathcal{E}_\mu(V) = -\frac{\mathcal{E}_\sigma(V)}{\sigma(V) - 1}.$$

The common gross profit is $\Pi = (p - \psi)x = px/\sigma(V) = E/[V\sigma(V)]$, which must be equal to the fixed cost, F. Thus, in a symmetric equilibrium, we have

$$V^{\text{eq}}\sigma(V^{\text{eq}}) = \frac{E}{F}.$$

Notice that V^{eq} is independent of ψ, and so are the revenue, $p^{\text{eq}}x^{\text{eq}} = \sigma(V^{\text{eq}})F$, the gross profit, F, and the production cost, $[\sigma(V^{\text{eq}}) - 1]F$, of each firm. However, the price, $p^{\text{eq}} = \mu(V^{\text{eq}})\psi$, and the quantity, $x^{\text{eq}} = \sigma(V^{\text{eq}})F/[\mu(V^{\text{eq}})\psi]$, are not.

The symmetric equilibrium is unique for any $E/F > 0$ if and only if $V\sigma(V)$ is globally increasing in V. This condition can be written as $1 + \mathcal{E}_\sigma(V) > 0$ or, equivalently, $\mathcal{E}_\mu(V) < \mu(V) - 1$.

---

[11]This section and the next draw heavily from Matsuyama & Ushchev (2020a).

[12]Throughout this review, $\mathcal{E}_f(x) \equiv xf'(x)/f(x) = \partial \ln f(x)/\partial \ln x$ denotes the elasticity of a positive-valued function, $f(x) > 0$, defined over a positive real number $x > 0$.

Under the same condition, V^{eq} is globally increasing in E/F.[13] Clearly, $\sigma'(\cdot) > 0$, the case of increasing substitutability, or equivalently, $\mu'(\cdot) < 0$, the case of procompetitive entry, is sufficient but not necessary.

In the symmetric equilibrium, the profit and production cost shares are

$$\frac{1}{\sigma(V^{\mathrm{eq}})}, \frac{1}{\mu(V^{\mathrm{eq}})},$$

and the profit/production cost ratio is

$$\frac{\mu(V^{\mathrm{eq}})}{\sigma(V^{\mathrm{eq}})} = \frac{1}{\sigma(V^{\mathrm{eq}}) - 1} = \mu(V^{\mathrm{eq}}) - 1$$

in all firms. All of them generally vary with V^{eq} and hence with $E/F > 0$.

4.2. Comparative Statics

Under the condition that ensures the uniqueness of the symmetric equilibrium and its stability, $1 + \mathcal{E}_\sigma(V) > 0$, we have

$$\widehat{V^{\mathrm{eq}}} = \frac{\hat{E} - \hat{F}}{1 + \mathcal{E}_\sigma(V^{\mathrm{eq}})}, \ \widehat{p^{\mathrm{eq}}} = \frac{\mathcal{E}_\mu(V^{\mathrm{eq}})\,(\hat{E} - \hat{F})}{1 + \mathcal{E}_\sigma(V^{\mathrm{eq}})} + \hat{\psi}, \text{ and } \widehat{x^{\mathrm{eq}}} = \frac{\mu(V^{\mathrm{eq}})\,\mathcal{E}_\sigma(V^{\mathrm{eq}})(\hat{E} - \hat{F})}{1 + \mathcal{E}_\sigma(V^{\mathrm{eq}})} + \hat{F} - \hat{\psi}.$$

Thus, for $\mathcal{E}_\sigma(V) \gtreqless 0 \Leftrightarrow \mathcal{E}_\mu(V) \lesseqgtr 0$, the market size effect is

$$0 < \frac{\partial \ln V^{\mathrm{eq}}}{\partial \ln E} = 1 - \frac{\partial \ln(p^{\mathrm{eq}} x^{\mathrm{eq}})}{\partial \ln E} = \frac{1}{1 + \mathcal{E}_\sigma(V^{\mathrm{eq}})} \lesseqgtr 1,$$

$$\frac{\partial \ln p^{\mathrm{eq}}}{\partial \ln E} = \frac{\mathcal{E}_\mu(V^{\mathrm{eq}})}{1 + \mathcal{E}_\sigma(V^{\mathrm{eq}})} \lesseqgtr 0, \ \frac{\partial \ln x^{\mathrm{eq}}}{\partial \ln E} = \frac{\mu(V^{\mathrm{eq}})\,\mathcal{E}_\sigma(V^{\mathrm{eq}})}{1 + \mathcal{E}_\sigma(V^{\mathrm{eq}})} \gtreqless 0,$$

and the profit/production cost ratio changes as

$$\frac{\partial \ln(\mu(V^{\mathrm{eq}})\,/\sigma(V^{\mathrm{eq}}))}{\partial \ln E} = \frac{\mathcal{E}_\mu(V^{\mathrm{eq}}) - \mathcal{E}_\sigma(V^{\mathrm{eq}})}{1 + \mathcal{E}_\sigma(V^{\mathrm{eq}})} \lesseqgtr 0.$$

The intuition is easy to grasp. For example, consider the case of increasing substitutability, $\mathcal{E}_\sigma(V) > 0$, or, equivalently, the case of procompetitive entry, $\mathcal{E}_\mu(V) < 0$. In response to a market size increase, more firms enter and product variety goes up. When this makes the products more substitutable, $\mathcal{E}_\sigma(V) > 0$, the markup rate goes down, $\mathcal{E}_\mu(V) < 0$, necessitating each firm to increase the scale of operation and earn more revenue just to break even. Because each firm is larger, the masses of firms and product variety go up at a rate lower than the rate of market size increase. This also means a decline in the profit/production cost ratio. Note that these comparative statics results depend on $\mathrm{sgn}\{\mathcal{E}_\sigma(V)\} = -\mathrm{sgn}\{\mathcal{E}_\mu(V)\}$, i.e., on how the markup rate responds to entry, not on whether the second law hold or not. They are also unrelated to the property of $\mathcal{L}(V)$, which plays a crucial role in determining the optimal allocation.

4.3. Optimal Allocation

This can be obtained by solving the following problem:

$$\max X(\mathbf{x}) \text{ such that } \int_\Omega \psi x_\omega \mathrm{d}\omega + VF \le E.$$

[13] Locally increasing $V\sigma(V)$ in the neighborhood of a symmetric equilibrium also ensures its local stability in any adjustment process with the following property: $\dot{V}_t \gtreqless 0$ if and only if $\Pi_t = E/V_t\sigma(V_t) \gtreqless F$.

The solution satisfies $x_\omega = x > 0$ for $\omega \in \Omega$; $x_\omega = 0$ for $\omega \notin \Omega$, simplifying the problem to

$$\max X(\mathbf{x}) = \max_{V(\psi x+F)\le E} xX(\mathbf{1}_\Omega) = \frac{F}{\psi}\max_V \frac{X(\mathbf{1}_\Omega)}{V}\left(\frac{E}{F} - V\right).$$

From the first-order condition, and using $\mathcal{L}(V) = \frac{\mathrm{d}\ln X(\mathbf{1}_\Omega)}{\mathrm{d}\ln V} - 1$, the optimal variety V^{op} satisfies

$$\left[1 + \frac{1}{\mathcal{L}(V^{\mathrm{op}})}\right]V^{\mathrm{op}} = \frac{E}{F}.$$

This condition fully characterizes V^{op} if the left-hand side is strictly increasing, which also ensures that V^{op} is globally increasing in E/F. This is clearly satisfied for the case of diminishing love-for-variety, $\mathcal{L}'(V) < 0$, which is sufficient but not necessary.

4.4. Optimal Versus Equilibrium

By comparing the two conditions,

$$\left[1 + \frac{1}{\mathcal{L}(V^{\mathrm{op}})}\right]V^{\mathrm{op}} = \frac{E}{F} \text{ and } \sigma(V^{\mathrm{eq}})V^{\mathrm{eq}} = \frac{E}{F},$$

with the left-hand side of each condition strictly increasing in V^{op} and in V^{eq}, respectively, one could easily verify the following.

Proposition 1. Assume that the symmetric equilibrium exists uniquely in the Dixit-Stiglitz environment under general homothetic symmetric demand systems. Then, we have

$$\mathcal{L}(V) \gtreqless \frac{1}{\sigma(V) - 1} \text{ for all } V > 0 \Leftrightarrow V^{\mathrm{eq}} \lesseqgtr V^{\mathrm{op}} \text{ for all } E/F > 0.$$

The logic behind this result is simple: $\mathcal{L}(V)$ captures the social incentive to add product variety and positive externalities due to incomplete appropriability, while $[\sigma(V) - 1]^{-1} = \mu(V) - 1 = \mu(V)/\sigma(V)$, the profit/production cost ratio, captures the private incentive to add product variety and negative externalities due to the business stealing effect. In general, these two do not coincide. For some classes of demand systems, the former dominates the latter, $\mathcal{L}(V)[\sigma(V) - 1] > 1$, hence $V^{\mathrm{eq}} < V^{\mathrm{op}}$. For some other classes, the latter dominates the former, $\mathcal{L}(V)[\sigma(V) - 1] < 1$, hence $V^{\mathrm{eq}} > V^{\mathrm{op}}$. In-between, there are the borderline classes for which the two coincide, $\mathcal{L}(V)[\sigma(V) - 1] = 1$, hence $V^{\mathrm{eq}} = V^{\mathrm{op}}$. CES belongs to the borderline but is not the only one, and the optimality in any of the borderline classes is not robust.[14] Moreover, though $\mathcal{L}(V)[\sigma(V) - 1] = 1$ ensures the optimality of the unique symmetric equilibrium, it does not rule out the existence of a symmetric set of asymmetric equilibria, none of which is optimal.

Proposition 1 gives us the condition for evaluating the optimality of the symmetric equilibrium, if it exists uniquely. However, because homotheticity and symmetry alone impose little restriction on the relation between $\sigma(V)$ and $\mathcal{L}(V)$, "almost anything goes." It is thus necessary to restrict the demand systems to make further progress. The next section introduces such a restriction in the form of HSA demand systems.

5. HSA DEMAND SYSTEMS

A homothetic symmetric demand system belongs to the HSA class with gross substitutes if the budget share of $\omega \in \Omega$ is a strictly decreasing function of its normalized price, $z_\omega \equiv p_\omega/A(\mathbf{p})$, only

[14] Matsuyama & Ushchev (2024a) constructed a family of the demand system defined by the geometric mean of the generalized translog (see **Supplemental Appendix 2**, example 2), which contains generic cases of excessive and insufficient entry and a continuum of non-generic cases of optimal entry, to which CES belongs.

Supplemental Material >

where the normalized price is defined by its own price, p_ω, divided by a single price aggregator, $A(\mathbf{p})$, which is common across all varieties. That is, all the cross-price effects are summarized in a single number, $A(\mathbf{p})$, or a sufficient statistic.

5.1. Definition

Formally, a homothetic symmetric demand system belongs to HSA with gross substitutes if there exists a function of a single variable, $s : \mathbb{R}_{++} \to \mathbb{R}_+$, 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 $\bar{z} \equiv \inf\{z > 0 | s(z) = 0\}$,[15] such that the budget share of $\omega \in \Omega$ can be expressed as

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

where $A(\mathbf{p})$ is defined implicitly by

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

The price elasticity of demand for $\omega \in \Omega$ is, for $p_\omega < \bar{z}A(\mathbf{p})$,

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

with $\lim_{z\to\bar{z}} \zeta(z) = \infty$, if $\bar{z} < \infty$. The second law holds if and only if $\zeta'(\cdot) > 0$.[16] If $\bar{z} < \infty$, $s_\omega = 0$ for $p_\omega \geq \bar{z}A(\mathbf{p})$, and hence $\bar{z}A(\mathbf{p})$ is the choke price.

Note that the budget share function, $s(\cdot)$, is the primitive of the HSA demand system. The common price aggregator, $A(\mathbf{p})$, is not, because it needs to be derived from $s(\cdot)$ using the adding-up constraint, $\int_\Omega s(p_\omega/A(\mathbf{p}))\mathrm{d}\omega \equiv 1$. Clearly, $A(\mathbf{p})$ is linear homogeneous in $\mathbf{p}$ for any fixed Ω, and the budget share $s(z_\omega) = s(p_\omega/A(\mathbf{p}))$ adds up to 1 by construction. Note that $A(\mathbf{p})$ is common across varieties and that both the budget share of $\omega \in \Omega$, $s(z_\omega)$, and its price elasticity, $\zeta(z_\omega)$, are functions of $z_\omega \equiv p_\omega/A(\mathbf{p})$ only. Thus, all the cross-variety effects in HSA are summarized by the single price aggregator, $A(\mathbf{p})$.[17]

After deriving $A(\mathbf{p})$ from $s(\cdot)$, the unit cost function, $P(\mathbf{p})$, can be derived by integrating $s_\omega = \partial \ln P(\mathbf{p})/\partial \ln p_\omega = s(p_\omega/A(\mathbf{p}))$ as

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

where $\Phi(z)$ is the productivity gain created by the product sold at the normalized price, z, and $c > 0$ is an integral constant, which is proportional to TFP. Clearly, $P(\mathbf{p})$ is linear homogeneous

[15]For $s : \mathbb{R}_{++} \to \mathbb{R}_+$ satisfying these conditions, a class of the budget share functions, $s(z; \gamma) \equiv \gamma s(z)$ 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 $\int_\Omega s(p_\omega/A(\mathbf{p}); \gamma)\mathrm{d}\omega = \int_\Omega s(p_{\omega'}/A(\mathbf{p}))\mathrm{d}\omega' = 1$. In this sense, $s(z; \gamma) \equiv \gamma s(z)$ for $\gamma > 0$ are all equivalent. Also, a class of the budget share functions, $s(z; \beta) \equiv s(z/\beta)$ for $\beta > 0$, generate the same demand system, with $A(\mathbf{p}; \beta) = A(\mathbf{p})/\beta$, because $s(p_\omega/A(\mathbf{p}; \beta); \beta) = s(p_\omega/\beta A(\mathbf{p}; \beta)) = s(p_\omega/A(\mathbf{p}))$. In this sense, $s(z; \beta) \equiv s(z/\beta)$ for $\beta > 0$ are all equivalent.

[16]Conversely, one can obtain $s(\cdot)$ as $s(z) = \gamma \exp\left[\int_{z_0}^{z} \frac{1-\zeta(\xi)}{\xi} \mathrm{d}\xi\right]$ from any $\zeta(\cdot) > 1$, with $\lim_{z\to\bar{z}} \zeta(z) = \infty$, if $\bar{z} < \infty$.

[17]Recall that, under general homothetic symmetric demand systems, the budget 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.

and monotonic. Moreover, Matsuyama & Ushchev (2017) showed that it is strictly quasi-concave, thereby proving the integrability (in the sense of Samuelson 1950 and Hurwicz & Uzawa 1971) of HSA demand systems. It is worth emphasizing that $P(\mathbf{p})/A(\mathbf{p})$ is not constant, with the sole exception of CES.[18] $A(\mathbf{p})$ and $P(\mathbf{p})$ generally move differently in response to a change in $\mathbf{p}$. This should make sense, because $A(\mathbf{p})$ is the inverse measure of competitive pressures from other products, which captures the cross-variety interactions in the demand system, while $P(\mathbf{p})$ is the unit cost function, which captures the productivity consequences of price changes; there is no reason to expect them to move together in general.[19] In other words, $A(\mathbf{p})$ in the definition of HSA cannot be replaced by $P(\mathbf{p})$, though many have claimed to the contrary.

Fallacy 7. $s_\omega = f(p_\omega/P(\mathbf{p}))$, with $f'(\cdot) < 0$, defines the class of flexible homothetic demand systems, which contains CES as a special case where $s_\omega \propto (p_\omega/P(\mathbf{p}))^{1-\sigma}$.

This is false because $\partial \ln P(\mathbf{p})/\partial \ln p_\omega = s_\omega = f(p_\omega/P(\mathbf{p}))$ is a partial differential equation of $P(\mathbf{p})$, whose solution must take the form of $s_\omega \propto (p_\omega/P(\mathbf{p}))^{1-\sigma}$, which means that CES is the only demand system that satisfies this property.

5.2. Substitutability and Love-for-Variety Under HSA

For symmetric price patterns, $\mathbf{p} = p1_\Omega^{-1}$, $z_\omega = z$ satisfies $s(z)V = 1$ and $-\ln P(1_\Omega^{-1}) = \ln c + \ln z + \Phi(z)$, from which we can show the following.

Proposition 2. Under HSA,

$$\sigma(V) = \zeta\left(\frac{1}{A(1_\Omega^{-1})}\right) = \zeta\left(s^{-1}\left(\frac{1}{V}\right)\right) > 1,$$

$$\mathcal{L}(V) \equiv -\frac{\mathrm{d}\ln P(1_\Omega^{-1})}{\mathrm{d}\ln V} = \Phi\left(s^{-1}\left(\frac{1}{V}\right)\right) > 0.$$

Since $s^{-1}(1/V)$ is increasing in V, $\operatorname{sgn}\{\zeta'(\cdot)\} = \operatorname{sgn}\{\sigma'(\cdot)\}$ and $\operatorname{sgn}\{\Phi'(\cdot)\} = \operatorname{sgn}\{\mathcal{L}'(\cdot)\}$. In particular, increasing substitutability and procompetitive entry is equivalent to the second law under HSA. Moreover, Matsuyama & Ushchev (2020a, 2023) show that

$$\zeta'(\cdot) \gtreqless 0,\ \forall z \in (z_0, \bar{z}) \Rightarrow \Phi'(z) \lesseqgtr 0,\ \forall z \in (z_0, \bar{z}).$$

The reverse is not true in general, except that

$$\Phi'(z) = 0,\ \forall z \in (z_0, \bar{z}) \ \Rightarrow\ \zeta'(\cdot) = 0,\ \forall z \in (z_0, \bar{z}).$$

Hence, from Proposition 2, we have the following.

[18]To see this, differentiating the adding-up constraint $\int_\Omega s(p_\omega/A(\mathbf{p}))\mathrm{d}\omega \equiv 1$ yields

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

which differs from $\partial \ln P(\mathbf{p})/\partial \ln p_\omega = s(z_\omega)$, unless $\zeta(z)$ is constant, i.e., except in the case of CES.

[19]Moreover, $A(\mathbf{p})$, being linear homogeneous in $\mathbf{p}$, the input price vector, 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 $\mathbf{p}$. Hence, it depends not only on the unit of measurement of inputs but also on that of the final good. Furthermore, a change in TFP, while affecting $P(\mathbf{p})$, leaves the market share unaffected. This is why the HSA demand system and $A(\mathbf{p})$ are independent of the integral constant, $c > 0$, which cannot be determined.

Proposition 3. Under HSA,

$$\sigma'(V) \gtreqless 0,\ \forall V \in (1/s(z_0), \infty) \ \Rightarrow\ \mathcal{L}'(V) \lesseqgtr 0,\ \forall V \in (1/s(z_0), \infty).$$

The reverse is not true in general, except that

$$\mathcal{L}'(V) = 0,\ \forall V \in (1/s(z_0), \infty) \ \Rightarrow\ \sigma'(V) = 0,\ \forall V \in (1/s(z_0), \infty).$$

Thus, the second law, increasing substitutability, and procompetitive entry are all equivalent to each other under HSA. Moreover, if any of them holds globally, it is sufficient (but not necessary) for globally diminishing love-for variety under HSA.[20]

6. DIXIT-STIGLITZ UNDER HSA

Let us now apply HSA to the Dixit-Stiglitz environment.[21]

6.1. Equilibrium

Holding E and $A = A(\mathbf{p})$ fixed, each firm chooses p_ω (hence $z_\omega \equiv p_\omega/A$) to maximize its gross profit,

$$(p_\omega - \psi)x_\omega = \left(1 - \frac{\psi}{p_\omega}\right)p_\omega x_\omega = \left(1 - \frac{\psi}{p_\omega}\right)s\left(\frac{p_\omega}{A(\mathbf{p})}\right)E = \left(1 - \frac{\psi/A}{z_\omega}\right)s(z_\omega)E.$$

The first-order condition can be written as the Lerner pricing formula, normalized by A, as

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

In what follows, let us make the following assumption for expositional reasons.[22]

Assumption 1. For all $z \in (0, \bar{z})$,

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(z\left[1 - \frac{1}{\zeta(z)}\right]\right) > 0.$$

Clearly, the second law, $\zeta'(z) > 0$, is sufficient but not necessary for Assumption 1. This Assumption states that, for any $A = A(\mathbf{p})$, the marginal revenue of each firm is strictly increasing in p_ω (i.e., decreasing in x_ω). Under Assumption 1, the left-hand side of the normalized Lerner formula is strictly increasing, and it can be inverted to express the profit-maximizing z_ω as

$$z_\omega = \frac{p_\omega}{A} = \tilde{Z}\left(\frac{\psi}{A}\right);\ \tilde{Z}'(\cdot) > 0.$$

Thus, the equilibrium is symmetric, so that $p_\omega = p$ and $z_\omega = z$, satisfying

$$z = \frac{p}{A} = \frac{p}{A(\mathbf{p})} = \frac{1}{A(1_\Omega^{-1})} = s^{-1}\left(\frac{1}{V}\right).$$

[20]Note that $\sigma'(\cdot) > 0$ everywhere over $(V,\ \infty)$ is sufficient for $\mathcal{L}'(V) < 0$, but $\sigma'(V) > 0$ is not. This is because substitutability is a local property of the demand system, while love-for-variety depends on its global properties.
[21]In addition to the work of Matsuyama & Ushchev (2020a,b, 2022a,b, 2024a), recent applications of HSA to MC include work by Fujiwara & Matsuyama (2022), Grossman et al. (2023), and Baqaee et al. (2024). Trottner (2023) and Ren & Zhang (2025) apply HSA to both monopolistic and monopsonic competition among firms with two-sided market power.
[22]Even without Assumption 1, the profit-maximizing z_ω is strictly increasing and the maximized profit $\Pi_\omega = s(z_\omega)E/\zeta(z_\omega)$ is strictly decreasing in the normalized cost ψ/A, which is all we need to establish the symmetry and uniqueness of the equilibrium. Without Assumption 1, however, z_ω is piecewise-continuous (i.e., it jumps up at some values of ψ/A), and Π_ω is piecewise-differentiable, which complicates the exposition.

Moreover, Assumption 1 is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}z}\ln\left(\frac{s(z)}{\zeta(z)}\right) < 0 \Leftrightarrow \frac{\mathrm{d}}{\mathrm{d}V}\ln V\sigma(V) > 0,$$

so that the maximized gross profit of each firm,

$$\left(1-\frac{\psi}{p}\right)s(z)E = \left(1-\frac{\psi/A}{z}\right)s(z)E = \frac{s(z)}{\zeta(z)}E = \frac{E}{V\sigma(V)},$$

is strictly decreasing in z and in V. Hence, the free-entry condition uniquely pins down ψ/A, z, and V. Thus, the equilibrium is symmetric and unique under HSA. From Section 4, we have

$$V^{\text{eq}}\sigma(V^{\text{eq}}) = \frac{E}{F},\ p^{\text{eq}} = \mu(V^{\text{eq}})\,\psi,\ x^{\text{eq}} = \frac{(\sigma(V^{\text{eq}})-1)F}{\psi} = \frac{F}{(\mu(V^{\text{eq}})-1)\,\psi},$$

and the profit share and the production cost share in the revenue in all firms are equal to

$$\frac{1}{\sigma(V^{\text{eq}})},\ \frac{1}{\mu(V^{\text{eq}})},$$

and the ratio of the profit to the production cost is equal to

$$\frac{\mu(V^{\text{eq}})}{\sigma(V^{\text{eq}})} = \frac{1}{\sigma(V^{\text{eq}})-1} = \mu(V^{\text{eq}})-1$$

in all firms. They all vary with V^{eq}, and hence with E/F, under non-CES HSA.

6.2. Comparative Statics

Clearly, the results obtained in Section 4 for a general homothetic demand system under the assumption that the symmetric equilibrium exists uniquely carry over to this case. Moreover, the comparative statics results for $\mathcal{E}_\sigma(V) \gtreqless 0 \Leftrightarrow \mathcal{E}_\mu(V) \lesseqgtr 0$ carry over for $\zeta'(z) \gtreqless 0$, because they are equivalent under HSA.

6.3. Optimal Versus Equilibrium

Differentiating $\Phi(z) \equiv [\int_z^{\bar{z}} \frac{s(\xi)}{\xi}\mathrm{d}\xi]/s(z)$ yields

$$\frac{\partial \ln \Phi(z)}{\partial \ln z} = -\frac{zs'(z)}{s(z)} - \frac{1}{\Phi(z)} = \zeta(z) - 1 - \frac{1}{\Phi(z)},$$

so that

$$\Phi'(z) \lesseqgtr 0 \ \Leftrightarrow\ \zeta(z) - 1 \lesseqgtr \frac{1}{\Phi(z)}.$$

Because $\sigma(V) = \zeta(s^{-1}(1/V))$, $\mathcal{L}(V) = \Phi(s^{-1}(1/V))$, and $s^{-1}(1/V)$ is increasing in V, the above equivalence translates into

$$\mathcal{L}'(V) \lesseqgtr 0 \ \Leftrightarrow \mathcal{L}(V) \lesseqgtr \frac{1}{\sigma(V)-1}.$$

Hence, from Propositions 1, 2, and 3, we derive the following.

Proposition 4. In the Dixit-Stiglitz environment under HSA,

$$\zeta'(z) \lesseqgtr 0 \text{ for all } z>0 \Leftrightarrow \sigma'(V) \lesseqgtr 0 \text{ for all } V > 0$$

$$\Rightarrow$$

$$\mathcal{L}'(V) \gtreqless 0 \text{ for all } V > 0 \Leftrightarrow V^{\text{eq}} \lesseqgtr V^{\text{op}} \text{ for all } E/F>0.$$

Moreover,

$$\zeta(z) = \text{const.} \;\Leftrightarrow\; \sigma(V) = \text{const.} \Leftrightarrow \mathcal{L}(V) = \text{const.} \;\Leftrightarrow\; V^{\text{eq}} = V^{\text{op}}$$
$$\text{for all } E/F > 0.$$

Thus, under HSA, equilibrium variety is excessive (insufficient) if and only if love-for-variety is diminishing (increasing), for which globally increasing (decreasing) substitutability or, equivalently, the second law (the anti-second law) is sufficient. Moreover, CES is the only HSA demand system in which substitutability is constant, love-for-variety is constant, and the equilibrium is optimal.

7. MELITZ UNDER HSA[23]

Let us now depart from the Dixit-Stiglitz environment and introduce heterogeneity across firms and their differentiated inputs.

7.1. The Melitz Environment

Consider what I shall call the Melitz (2003) environment. As before, there exists a single primary factor of production, called simply labor and taken as numeraire. Each differentiated input variety, $\omega \in \Omega$, is produced from labor and sold exclusively by a single MC firm, also indexed by $\omega \in \Omega$, and the firms' products enter symmetrically in the demand system. Moreover, the firms are ex-ante identical before they enter the market. However, unlike in the Dixit-Stiglitz environment, they become ex-post heterogeneous 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 marginal cost of production, ψ_ω, from the common cumulative distribution function, $G(\psi)$, with the density function $g(\psi) = G'(\psi) > 0$ over the support $(\underline{\psi}, \bar{\psi}) \subseteq (0, \infty)$. 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 identical firms enter until their expected gross profit is equal to the entry cost, $F_e = \int_{\underline{\psi}}^{\bar{\psi}} \max\{\Pi(\psi) - F, 0\} \mathrm{d}G(\psi)$. This ensures no excess profit in equilibrium, so that the total demand for labor in this sector is equal to $L = \mathbf{px} = P(\mathbf{p})X(\mathbf{x}) = E$. Let us now apply HSA to the Melitz environment.[24]

7.2. Pricing Behavior: Markup and Pass-Through Rates Across Firms

Knowing its marginal cost, ψ_ω, and holding E and $A = A(\mathbf{p})$ fixed, firm ω chooses p_ω (hence $z_\omega \equiv p_\omega/A$) to maximize

$$(p_\omega - \psi_\omega)x_\omega = \left(1 - \frac{\psi_\omega/A}{z_\omega}\right) s(z_\omega)E,$$

whose first-order condition is given by

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

[23]This section draws heavily from Matsuyama & Ushchev (2022b).

[24]The original Melitz model is a special case of Melitz under HSA. Melitz under homothetic direct implicit additivity (HDIA) or homothetic indirect implicit additivity (HIIA) is not analytically tractable without some additional assumptions (e.g., the presence of the choke price combined with zero overhead cost, as in Arkolakis et al. 2019). One could say very little under general homothetic symmetric demand systems.

Under Assumption 1, this can be inverted as $p_\omega/A = z_\omega = \tilde{Z}(\psi_\omega/A)$, $\tilde{Z}'(\cdot) > 0$. Thus, all the firms that share the same ψ_ω set the same price. This means that we can identify firms only by their marginal cost, ψ, so that we reindex them by ψ. Their profit-maximizing normalized price satisfies

CoPaTh: constant pass-through

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

The price elasticity of demand at the point ψ-firms operate and their markup rate can both expressed as functions of ψ/A,[25]

$$\zeta(z_\psi) = \zeta\left(\tilde{Z}\left(\frac{\psi}{A}\right)\right) \equiv \sigma\left(\frac{\psi}{A}\right), \ \mu\left(\frac{\psi}{A}\right) \equiv \frac{\sigma(\psi/A)}{\sigma(\psi/A) - 1},$$

which are related with the following identities:

$$\frac{1}{\sigma(\psi/A)} + \frac{1}{\mu(\psi/A)} = 1, \ \mathcal{E}_\sigma\left(\frac{\psi}{A}\right) = -\frac{\mathcal{E}_\mu(\psi/A)}{\mu(\psi/A) - 1}, \text{ and } \mathcal{E}_\mu\left(\frac{\psi}{A}\right) = -\frac{\mathcal{E}_\sigma(\psi/A)}{\sigma(\psi/A) - 1}.$$

The pass-through rate is also a function of ψ/A:

$$\rho_\psi \equiv \frac{\partial \ln p_\psi}{\partial \ln \psi} = \mathcal{E}_{\tilde{Z}}\left(\frac{\psi}{A}\right) \equiv \rho\left(\frac{\psi}{A}\right) = \frac{1}{1 + \mathcal{E}_{1-1/\zeta}\left(\tilde{Z}(\psi/A)\right)} = 1 + \mathcal{E}_\mu\left(\frac{\psi}{A}\right) > 0.$$

Note that $\sigma(\cdot)$, $\mu(\cdot)$, and $\rho(\cdot)$ are all functions of the normalized cost, ψ/A, only. This means that, for non-CES HSA, market size E affects the pricing behaviors of firms only through its effects on $A = A(\mathbf{p})$. (They are constant under CES: $\sigma(\cdot) = \sigma$, $\mu(\cdot) = \sigma/(\sigma - 1) = \mu$, and $\rho(\cdot) = 1$.) Moreover, $A = A(\mathbf{p})$ enters only as the divisor of ψ. This means that a decline in A, i.e., more competitive pressures, acts like a uniform decline in productivity across firms.

Moreover, it is straightforward to verify that

$$\zeta'(\cdot) \gtreqless 0 \Leftrightarrow \mathcal{E}_\sigma(\cdot) \gtreqless 0 \Leftrightarrow \mathcal{E}_\mu(\cdot) \lesseqgtr 0 \Leftrightarrow \rho(\cdot) \lesseqgtr 1.$$

Under the second law, $\zeta'(\cdot) > 0$, high-ψ firms set lower markup rates, and their pass-through rates are less than 1 (incomplete pass-through). The equivalence of the second law and incomplete pass-through is general and not specific to HSA, though it hinges on the assumption that the MC firms are price takers on their input market. Under HSA, the second law, $\zeta'(\cdot) > 0$, also implies that more competitive pressures, i.e., a lower A, force all firms to lower their markup rates regardless of their marginal cost ψ.

The second law alone does not say how the pass-through rate varies across firms or how it responds to more competitive pressures. Motivated by some evidence that more productive firms have lower pass-through rates, let us introduce the strong (weak) third law,[26]

$$\mathcal{E}_{1-1/\zeta}{}'(\cdot) < (\leq) 0,$$

which implies $\rho'(\cdot) > (\geq) 0$ and $\mathcal{E}_\mu{}'(\cdot) > (\geq) 0$. Among the parametric families listed in **Supplemental Appendix 2**, generalized translog violates even the weak third law; CoPaTh,

Supplemental Material >

[25]Notice some abuse of notations here. Until the previous section, $\sigma(\cdot)$ and $\mu(\cdot)$ are both functions of V, denoting the common values across symmetric firms. In this section, $\sigma(\cdot)$ and $\mu(\cdot)$ are both functions of ψ/A, denoting the price elasticity and the markup rate of ψ-firms. This should not cause any confusion. They are clearly related. Since $s^{-1}(1/V)$ is increasing in V, and $\tilde{Z}(\psi/A)$ is increasing in ψ/A, $\sigma(V) \equiv \zeta(s^{-1}(1/V))$ and $\sigma(\psi/A) \equiv \zeta(\tilde{Z}(\psi/A))$ are both increasing and $\mu(V)$ and $\mu(\psi/A)$ are both decreasing if and only if $\zeta(\cdot)$ is increasing.

[26]The first law of demand states that a higher price reduces demand, restricting the first derivative of the demand curve. The second law states that a higher price increases the price elasticity, restricting the second derivative. We call this law—whereby a higher price reduces the rate of change in the price elasticity—the third law because it restricts the third derivative.

which features a constant pass-through rate, satisfies the weak (but not strong) third law; and PEM/FIM, which features power elasticity of markup rate/Fréchet inverse markup rate, satisfies the strong third law.

PEM: power elasticity of markup rate

FIM: Fréchet inverse markup rate

Fallacy 8. Translog is flexible, as it can approximate any homothetic symmetric demand system.

Some even claim that because translog is flexible, the results shown under translog hold under general homothetic demand systems. These claims are simply false. Symmetric translog (Feenstra 2003) belongs to the generalized translog, which in turn belongs to HSA. Its budget share function can be expressed as $s(z) = -\max\{\ln z, 0\}$ without any loss of generality. It has no parameter to fit the data. It is highly tractable, which explains its popularity, but it has no flexibility whatsoever.[27] Moreover, it violates even the weak third law, and it is thus inconsistent with the evidence that more productive firms have lower pass-through rates.[28]

Under the strong third law, high-ψ firms have higher pass-through rates, and more competitive pressures, a lower A, cause the pass-through rate to go up across all firms. The strong third law is also equivalent to

$$\frac{\partial^2 \ln \mu(\psi/A)}{\partial\psi\,\partial\,(1/A)} > 0.$$

That is, $\mu(\psi/A)$ is log-supermodular[29] in ψ and $1/A$. Because $\mu(\psi/A)$ is decreasing in ψ under the second law, this means that more competitive pressures cause a proportionately smaller decline in the markup rate for high-ψ firms, and thus a smaller dispersion of the markup rates across firms. 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).

7.3. Revenue, Gross Profit, and Employment Across Firms

Revenue, gross profit, and employment of ψ-firms can all be written as functions of ψ/A multiplied by market size E, because

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

[27]Some agree with me about the nonflexibility of symmetric translog. For example, Edmond et al. (2023, p. 1623) wrote that "Kimball…is more flexible than…symmetric translog…and is better able to match our calibration targets. But…translog…is more tractable than…Kimball…and leads to sharp analytic results." In this respect, I argue that HSA dominates both Kimball and symmetric translog, because it is as flexible as Kimball and as tractable as symmetric translog.

[28]I am not sure why some people believe that translog is flexible. Maybe it is because translog offers local second-order approximation to any unit cost function, which may be good enough for studying the impacts of small shocks in a competitive economy, where all firms are price takers. However, it is not good enough when firms make price-setting and entry decisions, because these decisions depend on the global properties and the third derivatives of the unit cost function. Perhaps this belief is analogous to the widespread use of the quadratic function in the early days of portfolio theory, purportedly because it offers local second-order approximation to any risk-averse utility function, in spite of its counterfactual implication that the rich invest a larger fraction of the wealth into the safe asset. This was at least until Arrow (1971) pointed out that how the household wealth affects its portfolio choice depends on how the Arrow-Pratt measures of absolute and relative risk aversion vary with consumption, which hinges on its third derivatives.

[29]A positive-value function, $f(x, y) > 0$, is log-supermodular in x and y if $\partial^2 \ln f(x, y)/\partial x\partial y > 0$ and log-submodular in x and y if $\partial^2 \ln f(x, y)/\partial x\partial y < 0$. Costinot & Vogel (2015) offer an accessible exposition of their properties.

$$\Pi_\psi = \frac{r(\psi/A)}{\sigma(\psi/A)}E \equiv \pi\left(\frac{\psi}{A}\right)E, \text{ and}$$
$$\psi x_\psi = \frac{r(\psi/A)}{\mu(\psi/A)}E \equiv \ell\left(\frac{\psi}{A}\right)E.$$

Moreover, they vary according to

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

all of which are independent of market size E and depend solely on ψ/A, through $\sigma(\cdot)$ and $\rho(\cdot)$. [Under CES, $\sigma(\cdot) = \sigma$ and $\rho(\cdot) = 1$, so that $\mathcal{E}_r(\cdot) = \mathcal{E}_\pi(\cdot) = \mathcal{E}_\ell(\cdot) = 1 - \sigma < 0$.] This means that, for non-CES HSA, market size E affects the relative firm size in revenue, gross profit, and employment only through its effects on $A = A(\mathbf{p})$. (Under CES, the relative firm size never changes.) Moreover, $A = A(\mathbf{p})$ enters only as the divisor of ψ; a decline in A thus acts as if firm productivity declined uniformly, not only in terms of its implications on the firm behavior but also in terms of its implications on the firm relative performance.

Note also that $R_\psi = r(\psi/A)E$ and $\Pi_\psi = \pi(\psi/A)E$ are both strictly decreasing in ψ/A, but $\ell(\psi/A)E$ may be nonmonotonic in ψ/A, because $1 - \sigma(\cdot)\rho(\cdot)$ may change its sign. Under the second and the weak third law, $\sigma(\cdot)\rho(\cdot)$ is strictly increasing, and one can show that $\ell(\psi/A)E$ is hump-shaped in ψ/A. Moreover, the profit is log-submodular in ψ and $1/A$ under the second law,

$$\frac{\partial^2 \ln \Pi_\psi}{\partial \psi \partial(1/A)} = -\sigma'\left(\frac{\psi}{A}\right) < 0,$$

while $R_\psi = r(\psi/A)E$ and $\ell(\psi/A)E$ are log-submodular in ψ and $1/A$ under the second and weak third laws:

$$\frac{\partial^2 \ln R_\psi}{\partial \psi \partial(1/A)} = \left[1-\sigma\left(\frac{\psi}{A}\right)\right]\rho'\left(\frac{\psi}{A}\right) - \sigma'\left(\frac{\psi}{A}\right)\rho\left(\frac{\psi}{A}\right) < 0,$$
$$\frac{\partial^2 \ln(\psi x_\psi)}{\partial \psi \partial(1/A)} = -\sigma'\left(\frac{\psi}{A}\right)\rho\left(\frac{\psi}{A}\right) - \sigma\left(\frac{\psi}{A}\right)\rho'\left(\frac{\psi}{A}\right) < 0.$$

Since R_ψ and Π_ψ are both decreasing in ψ, this implies that more competitive pressures cause a proportionately larger decline in the revenue and the profit among high-ψ firms, and hence a larger dispersion in revenue and profit across firms.

Up to now, we looked at how different firms respond to a change in competitive pressures, A. Of course, $A = A(\mathbf{p})$ is endogenous, so that it can change only in response to some changes in exogenous variables, such as the entry cost, F_e, the overhead cost, F, and market size, E. To understand this, let us now turn to the equilibrium analysis.

7.4. Equilibrium

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

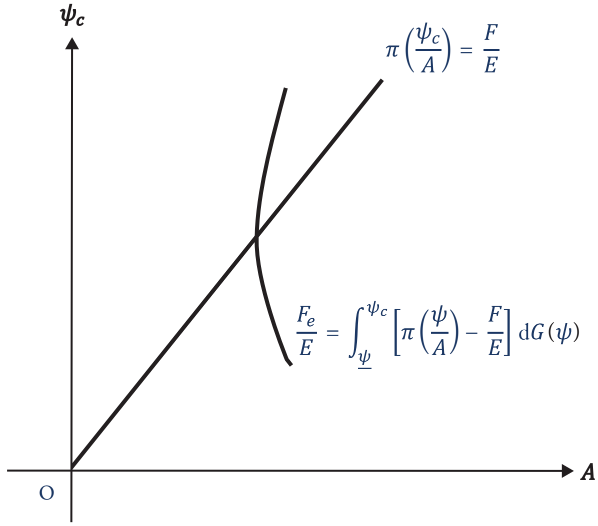

Figure 1

The cutoff rule and the free-entry condition under HSA.

and the overhead cost, regardless of their marginal costs. An equilibrium is characterized by the following three conditions.

Cutoff rule. Firms choose to stay if $\psi \leq \psi_c$ and to exit if $\psi > \psi_c$, where ψ_c is the cutoff level of the marginal cost, determined by

$$\pi\left(\frac{\psi_c}{A}\right)E = F.$$

Figure 1 depicts the cutoff rule as the ray from the origin, whose slope, $\psi_c/A = \pi^{-1}(F/E)$, 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,

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

Figure 1 depicts this condition as the C-shaped curve, downward-sloping below the cutoff rule, upward-sloping above it, and vertical at the intersection. The curve shifts to the left as the entry cost declines, which causes A to decline.

As **Figure 1** illustrates, the cutoff rule and the free-entry condition alone fully determine the equilibrium values of $A = A(\mathbf{p})$ and ψ_c uniquely as functions of F_e/E and F/E. In what follows, assume that F_e is not too large to ensure the interior solution, $0 < G(\psi_c) < 1$.

With A and ψ_c pinned down, we can calculate the mass of entering firms, M, and that of active firms, $V = MG(\psi_c)$, from the following.[30]

[30]One of the advantages of HSA is that the equilibrium is solved recursively. The mass of entrants and hence the mass of firms staying active are derived using the adding-up (resource) constraint after solving for the

Adding-up (resource) constraint. This can be written as

$$\int_{\Omega} s\left(\frac{p_\omega}{A}\right) \mathrm{d}\omega = M \int_{\underline{\psi}}^{\psi_c} r\left(\frac{\psi}{A}\right) \mathrm{d}G(\psi) = 1.$$

From this, the mass of active firms (hence the mass of product variety) is

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

Here $\tilde{G}(\xi;\psi_c) \equiv G(\psi_c\xi)/G(\psi_c)$ is the cumulative distribution function of $\xi \equiv \psi/\psi_c$, defined for $\underline{\xi} \equiv \underline{\psi}/\psi_c < \xi \leq 1$. The above expression shows that the selection ψ_c affects the equilibrium product variety V through its effect on $\tilde{G}(\xi;\psi_c)$. It turns out that a lower ψ_c (a tougher selection) shifts $\tilde{G}(\xi;\psi_c)$ to the right (left) in the sense of the monotone likelihood ratio if $\mathcal{E}_g'(\psi) < (>)0$. Pareto-distributed productivity, $G(\psi) = (\psi/\bar{\psi})^\kappa$, is the borderline case, $\mathcal{E}_g'(\psi) = 0$, in which $\tilde{G}(\xi;\psi_c)$ is independent of ψ_c. Since Fréchet, Weibull, and lognormal distributions all satisfy $\mathcal{E}_g'(\psi) < 0$, a lower ψ_c (a tougher selection) shifts $\tilde{G}(\xi;\psi_c)$ to the right. However, there is some evidence for $\mathcal{E}_g'(\psi) > 0$, which suggests that a lower ψ_c (a tougher selection) shifts $\tilde{G}(\xi;\psi_c)$ to the left.

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 V, 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 $P \to \lambda P$.

7.5. Average Markup Rate, Profit and Production Cost Measures Across Active Firms

Except under CES, heterogeneous firms differ in their markup rates, $\mu(\psi/A)$, so they differ also in the gross profit and the production cost shares in revenue,

$$\frac{\pi(\psi/A)}{r(\psi/A)} = \frac{1}{\sigma(\psi/A)}, \quad \frac{\ell(\psi/A)}{r(\psi/A)} = \frac{1}{\mu(\psi/A)},$$

and in the profit/production cost ratio,

$$\frac{\pi(\psi/A)}{\ell(\psi/A)} = \frac{\mu(\psi/A)}{\sigma(\psi/A)} = \frac{1}{\sigma(\psi/A) - 1} = \mu(\psi/A) - 1.$$

---

cutoff and a single measure of the competitive pressures. On the other hand, HDIA and HIIA both have two price aggregators, one capturing how competitive pressures affect firms' pricing behavior and the other capturing how competitive pressures from entry affect firms' profit without affecting their pricing behavior. For this reason, all the equilibrium conditions need to be solved simultaneously under HDIA and HIIA, with the feedback effect from the resource constraint. This makes ensuring the existence and uniqueness of the equilibrium and the comparative static exercises challenging. This shows up clearly in the work of Baqaee et al. (2024). Under HSA, the allocative efficiency term in the market size effect in their Theorem 1 has no denominator, while the corresponding expression under HDIA in their section 8 has a denominator that captures the feedback effect and could be negative or change the sign, indicating the possibility of the multiplicity and nonexistence of equilibriums.

It turns out that comparative statics requires comparing the profit, the revenue, and the employment of the firms at the cutoff with those of the industry average. Let

$$\mathbb{E}_1(f) \equiv \frac{\int_{\underline{\psi}}^{\psi_c} f\left(\frac{\psi}{A}\right) \mathrm{d}G(\psi)}{\int_{\underline{\psi}}^{\psi_c} \mathrm{d}G(\psi)} \text{ and } \mathbb{E}_w(f) \equiv \frac{\int_{\underline{\psi}}^{\psi_c} f\left(\frac{\psi}{A}\right) w\left(\frac{\psi}{A}\right) \mathrm{d}G(\psi)}{\int_{\underline{\psi}}^{\psi_c} w\left(\frac{\psi}{A}\right) \mathrm{d}G(\psi)}$$

denote, respectively, the unweighted average of $f(\cdot)$ and the $w(\cdot)$-weighted average of $f(\cdot)$ across the active firms, which are related as follows:

$$\frac{\mathbb{E}_1(f)}{\mathbb{E}_1(w)} = \mathbb{E}_w\left(\frac{f}{w}\right) = \frac{1}{\mathbb{E}_f(w/f)}.$$

For example, by applying this formula, we could have

$$\frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(r)} = \mathbb{E}_r\left(\frac{\pi}{r}\right) = \mathbb{E}_r\left(\frac{1}{\sigma}\right) = \frac{1}{\mathbb{E}_\pi(r/\pi)} = \frac{1}{\mathbb{E}_\pi(\sigma)},$$

that is, the sector-level profit share is equal to the revenue-weighted arithmetic mean and the profit-weighted harmonic mean of the profit shares across active firms. Likewise,

$$\frac{\mathbb{E}_1(\ell)}{\mathbb{E}_1(r)} = \mathbb{E}_r\left(\frac{\ell}{r}\right) = \mathbb{E}_r\left(\frac{1}{\mu}\right) = \frac{1}{\mathbb{E}_\ell(r/\ell)} = \frac{1}{\mathbb{E}_\ell(\mu)},$$

that is, the sector-level production cost share is equal to the revenue-weighted arithmetic mean and the employment-weighted harmonic mean of the production cost shares across active firms.

7.6. Comparative Statics: Competitive Pressures and Firm Selection

By totally differentiating the cutoff rule and the free-entry condition with respect to F_e, E, and F, their effects on $A = A(\mathbf{p})$ and ψ_c are

$$\hat{A} = \frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)}\left[(1-f_x)\left(\widehat{\frac{F_e}{E}}\right) + f_x\left(\widehat{\frac{F}{E}}\right)\right],\ \widehat{\psi_c} = \frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)}\left[(1-f_x)\left(\widehat{\frac{F_e}{E}}\right) + (f_x - \delta)\left(\widehat{\frac{F}{E}}\right)\right],$$

where

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

$$f_x \equiv \frac{FG(\psi_c)}{F_e + FG(\psi_c)} = \frac{\pi(\psi_c/A)}{\mathbb{E}_1(\pi)} < 1, \text{ and } \delta \equiv \frac{\mathbb{E}_\pi(\sigma)-1}{\sigma(\psi_c/A)-1} = \frac{\pi(\psi_c/A)}{\ell(\psi_c/A)}\frac{\mathbb{E}_1(\ell)}{\mathbb{E}_1(\pi)} \equiv f_x\frac{\mathbb{E}_1(\ell)}{\ell(\psi_c/A)} > 0.$$

Let us look at each shock separately. First, consider the entry cost F_e:

$$\frac{\partial \ln A}{\partial \ln F_e} = \frac{\partial \ln \psi_c}{\partial \ln F_e} = (1-f_x)\frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} > 0.$$

A decline in F_e shifts the C-shaped curve to the left in **Figure 1** and leads to a decline in A (more competitive pressures) and a decline in ψ_c (a tougher selection).

Next, consider market size E:

$$\frac{\partial \ln A}{\partial \ln E} = -\frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} < 0;\ \frac{\partial \ln \psi_c}{\partial \ln E} = -(1-\delta)\frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)}.$$

A higher E shifts the C-shaped curve to the left and the cutoff rule counter-clockwise in **Figure 1**.[31] This always leads to a lower A but to a lower ψ_c if and only if $\delta < 1$, i.e., $\mathbb{E}_\pi(\sigma) < \sigma(\psi_c/A)$, which holds under the second law.[32]

Finally, consider the overhead cost, F:

$$\frac{\partial \ln A}{\partial \ln F} = f_x \frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)} > 0, \quad \frac{\partial \ln \psi_c}{\partial \ln F} = (f_x - \delta)\frac{\mathbb{E}_1(\pi)}{\mathbb{E}_1(\ell)}.$$

A decline in F also shifts the C-shaped curve to the left and makes the cutoff rule steeper in **Figure 1**. This always leads to a lower A but to a lower ψ_c if and only if $\delta < f_x$, i.e., $\mathbb{E}_1(\ell) < \ell(\psi_c/A)$, which holds if more productive firms employ less, which occurs under the second and the weak third laws when F is sufficiently high but not under CES.

7.7. Comparative Statics: Firm Size Distributions in Revenue and Profit

A change in F_e and a change in F both affect the revenue, $R_\psi = r(\psi/A)E$, and the profit, $\Pi_\psi = \pi(\psi/A)E$, across firms only through their effect on A. Thus, we already know that a decline in F_e or in F causes the profit and the revenue of all firms to decline, and the negative effects are proportionately larger among high-ψ firms in the profit (under the second law) and in the revenue (under the third law), thereby causing a larger dispersion in the profit and in the revenue.

For an increase in market size, E, we also need to take into account the direct positive effect of E in addition to the indirect negative effect through the decline in A. The direct positive effect is uniform across all firms. Under CES, the indirect negative effect is also uniform, so that these two effects cancel out. Under the second law, however, the decline in A caused by an increase in E causes the profit distribution to skew more toward low-ψ firms. Because of this, the combined effect is that the profit is up among low-ψ firms, it is down among middle-ψ firms, and high-ψ firms are forced to exit (a decline in ψ_c). Under the second and the weak third laws, the combined effect on the revenue is similar, possibly except when the overhead cost F is large.[33]

7.8. Comparative Statics: Average Markup and Pass-Through Rates

Under the second law, more competitive pressures, a lower A, causes the procompetitive effect—i.e., the markup rate $\mu(\psi/A)$ declines for each firm—but it also causes the firm size distribution to shift toward low-ψ firms with higher markup rates. Likewise, under the strong third law, a lower A causes the procompetitive effect—i.e., the pass-through rate $\rho(\psi/A)$ increases for each firm—but it also causes the firm size distribution to shift toward low-ψ firms with lower pass-through rates. Due to the composition effect working against the procompetitive effect, how more competitive pressures affect the average rates in the industry depends on whether the elasticity of the density function, $\mathcal{E}_g(\cdot)$, is globally increasing or globally decreasing.[34] For a change in F_e that keeps ψ_c/A intact, the composition effect dominates the procompetitive effect and the average

[31] Since $F > 0$, the cutoff rule implies $\pi(\psi_c/A) > 0$, hence $\psi_c/A < \tilde{Z}(\psi_c/A) < \bar{z}$. If $F = 0$ and the choke price exists, $\bar{z} < \infty$, the cutoff rule is $\pi(\psi_c/A) = 0$, so that $\psi_c/A = \tilde{Z}(\psi_c/A) = \bar{z}$. Hence, a change in E does not affect the cutoff rule, and the result is the same with a change in F_e.

[32] Under CES, $\delta = 1$, hence the cutoff does not change. The profit at the cutoff is always equal to F, and with the constant markup rate, the revenue and the employment at the cutoff are also unaffected. Moreover, we know that firm size distribution is not affected by a change in A. Thus, all firms are unaffected. Thus, the only effect of a change in E under CES is a proportional change in $V = MG(\psi_c)$.

[33] This qualification is necessary because the markup rate goes down for all firms, so that the cutoff firms need to earn a higher revenue to earn enough profit to cover the overhead cost F. Hence, when F is large, the revenue of all the firms that stay may increase.

[34] The following results hold for any industry average of $f(\psi/A) = \mu(\psi/A)$ or $f(\psi/A) = \rho(\psi/A)$ of the form $I \equiv \mathcal{M}^{-1}(\mathbb{E}_w(\mathcal{M}(f)))$, where $\mathcal{M} : \mathbb{R}_+ \to \mathbb{R}$ is a monotone transformation and $w(\psi/A)$ is a weighted function.

rates move in the opposite direction from the firm-level rates if and only if $\mathcal{E}_g'(\cdot) > 0$. Thus, under the second law, an entry cost decline, which causes all firms to lower their markup rates, ends up increasing the average markup rate by shifting the firm size distribution toward more productive, high-markup firms. In contrast, the procompetitive effect dominates the composition effect and the average rates move in the same direction with the firm-level rates if and only if $\mathcal{E}_g'(\cdot) < 0$, which is satisfied, for example, by Fréchet, Weibull, and lognormal distributions. $\mathcal{E}_g'(\cdot) = 0$ (i.e., Pareto-distributed productivity) is the knife-edge case where there is no change in the average rates. For a change in E or in F for which $\mathrm{d}\ln \psi_c/\mathrm{d}\ln A < 1$ holds, $\mathcal{E}_g'(\cdot) > 0$ is a necessary condition for the composition effect to dominate, while $\mathcal{E}_g'(\cdot) \leq 0$ is a sufficient condition for the procompetitive effect to dominate.

7.9. Comparative Statics: Total Factor Productivity

The logic above can also be applied to the impact of more competitive pressures on TFP, because $\ln(A/cP) = \mathbb{E}_r[\Phi \circ \tilde{Z}]$ is the revenue-weighted average of $\Phi(\tilde{Z}(\psi/A))$ across active firms, and $\zeta'(\cdot) \gtreqless 0 \Rightarrow \Phi \circ \tilde{Z}'(\cdot) \lesseqgtr 0$. Under the second law, $\zeta'(\cdot) > 0$ implies $\Phi'(\cdot) < 0$, hence $\Phi(\tilde{Z}(\psi/A))$ is decreasing in ψ/A. This implies, for example, that a change in F_e that keeps ψ_c/A intact implies $\mathrm{d}\ln P/\mathrm{d}\ln A \gtreqless 1$ if and only if $\mathcal{E}_g'(\cdot) \gtreqless 0$.

7.10. Comparative Statics: Masses of Entrants and Active Firms

The effect of the mass of entrants, M, is simple. It immediately follows from the adding-up constraint that M increases when hit by any shock that causes a decline in A and a decline in ψ_c. For the mass of active firms (and hence product variety) $V = MG(\psi_c)$, M and $G(\psi_c)$ move in the opposite direction. The overall effect depends on whether the elasticity of the cumulative distribution function, $\mathcal{E}_G(\cdot)$, is globally increasing or globally decreasing.[35] A decline in F_e that keeps ψ_c/A intact causes $MG(\psi_c)$ to increase if and only if $\mathcal{E}_G'(\cdot) < 0$ and to decline if and only if $\mathcal{E}_G'(\cdot) > 0$. Again, $\mathcal{E}_G'(\cdot) = 0$ (i.e., Pareto-distributed productivity) is the knife-edge case where $MG(\psi_c)$ remains unchanged. For an increase in E or a decline in F, $\mathcal{E}_G'(\cdot) > 0$ is necessary for $MG(\psi_c)$ to go down and $\mathcal{E}_G'(\cdot) \leq 0$ is sufficient for $MG(\psi_c)$ to go up.

7.11. Sorting of Heterogeneous Firms Across Markets

As an application, let us consider a multi-market extension of Melitz under HSA. Imagine that there are $J \geq 2$ markets, indexed by $j = 1, 2, \ldots, J$, and their market sizes, $E_1 > \cdots > E_j > \cdots > E_J > 0$, are the only exogenous source of heterogeneity across markets. The primary factor of production, labor, is full mobile, equalizing its price across the markets, so that we can still use it as the numeraire. As before, each MC firm pays $F_e > 0$ to draw its marginal cost $\psi \sim G(\psi)$. After learning its ψ, each firm now decides which market to enter and to produce with an overhead cost, $F > 0$, or to exit without producing in any market. Firms sell their products at the profit-maximizing prices in the market they enter.

The unique equilibrium under the second law is characterized by $A_1 < A_2 < \cdots < A_J$ and $\underline{\psi} = \psi_0 < \psi_1 < \psi_2 < \cdots < \psi_J = \psi_c < \bar{\psi}$, with firms $\psi \in (\psi_{j-1}, \psi_j)$ entering market j. The intuition is simple. The ratio of the profit of ψ firms in market $j-1$ relative to that in market j is $[\pi(\psi/A_{j-1})/\pi(\psi/A_j)][E_{j-1}/E_j]$. For both markets to attract some firms, this ratio must be greater

All Hölder means are special cases, including the arithmetic mean $[I = \mathbb{E}_w(f)]$, the geometric mean $[\ln I = \mathbb{E}_w(\ln f)]$, and the harmonic mean $[1/I = \mathbb{E}_w(1/f)]$, and the weight function, $w(\psi/A)$, can be the profit, the revenue, and the employment.

[35] Globally increasing (decreasing) $\mathcal{E}_G(\cdot)$ is a weaker condition than globally increasing (decreasing) $\mathcal{E}_g(\cdot)$.

than 1 for some firms and less than 1 for others, which implies $A_j < A_{j+1}$, i.e., more competitive pressures in larger markets. The log-submodularity of $\pi(\psi/A)$ in ψ and $1/A$ under the second law means that this ratio is decreasing in ψ, which means that more productive firms sort themselves into larger markets.

Because of more competitive pressures in larger markets, firms are forced to set lower markup rates as they move to larger markets under the second law. However, larger markets attract more productive firms with high markup rates. Due to this composition effect, the average markup rate can be higher in larger markets. If the strong third law also holds, firms set higher pass-through rates in larger markets, but larger markets attract more productive firms with lower pass-through rates. Due to this composition effect, the average pass-through rate may be lower in larger markets. These results should provide a caution against testing the second and third laws by comparing the average markup/pass-through rates in a cross section of cities.

7.12. International/Interregional Trade with Differential Market Access

Up to now, it has been assumed that each firm can sell its product only in the market where it is located. 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, which allows firms to gain equal access to all markets, regardless of their locations. However, what are the effects if firms must 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 the MC industry has two symmetric markets in two countries/regions.[36] Both markets are characterized by market size E and by labor supplied at the price equal to 1. This ensures that the same level of competitive pressures prevails 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 an additional iceberg cost, $\tau > 1$. That is, only $1/\tau$ fraction of the product shipped arrives at the export market. This implies that the marginal cost of exporting is $\tau\psi_\omega$, which is 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 if and only if $\psi_\omega \leq \psi_{xc} = \psi_c/\tau < \psi_c$, and it sells only to the home market if and only if $\psi_{xc} = \psi_c/\tau < \psi_\omega \leq \psi_c$, 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, and a fraction $1 - G(\psi_c)$ exits.

Free-entry condition. The expected profit from both markets is equal to the entry cost,

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

where the first (second) term of the right-hand side is the expected profit from selling at home (abroad).

[36]Extending the analysis below to many symmetric markets is straightforward. However, extending it to two or more asymmetric markets requires additional assumptions. Matsuyama & Ottaviano (2024) conduct such analysis for the CoPaTh family of HSA.

Adding-up (resource) constraint. Let M denote the mass of the firms that pay the entry cost in each market. Then,

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

From this, we can determine $MG(\psi_c)$, the mass of the domestic firms, and $MG(\psi_{xc})$, that of the foreign firms operating in each market. By combining the cutoff rules and the free-entry condition, we have

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

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, $\psi_c < \bar{\psi}$ or $G(\psi_c) < 1$. Then, the right-hand side of this condition is strictly increasing in $\psi_c \in (\underline{\psi}, \bar{\psi})$, so that it is easy to verify that a decline in τ (globalization) causes the following.

- A decline in ψ_c and an increase in ψ_{xc}. Hence, $G(\psi_c)$ falls, $G(\psi_{xc})$ rises, and the share of exporting firms, $G(\psi_{xc})/G(\psi_c)$, rises.
- A decline in A and an increase in A/τ. Hence,
  - $r(\psi_\omega/A)$ and $\pi(\psi_\omega/A)$ decline and $r(\tau\psi_\omega/A)$ and $\pi(\tau\psi_\omega/A)$ rise. Thus, the shares of each domestic firm in revenue and profit are down, and those of each foreign firm are up. Moreover, there are more foreign firms relative to the domestic firms, so that the shares of all the domestic (foreign) firms are down (up) in each market.
  - $\mu(\psi_\omega/A)$ declines and $\mu(\tau\psi_\omega/A)$ rises under the second law, and $\rho(\psi_\omega/A)$ rises and $\rho(\tau\psi_\omega/A)$ declines under the strong third law. Thus, each exporting firm reduces its markup rate but raises its pass-through rate at home and simultaneously raises its markup rate and reduces its pass-through rate abroad. In each market, the markup (pass-through) rates set by the domestic firms are down (up), while those set by the foreign firms are up (down).

Erratum >

8. OTHER FORMS OF FIRM HETEROGENEITY UNDER HSA

One of the advantages of HSA is its analytical tractability when used in MC models with entry/exit and heterogeneous firms. The Melitz-type firm heterogeneity in productivity discussed in Section 7 is an example of such MC models in which different firms, despite the fact that their products enter symmetrically in the demand system, could set different prices with different markup rates. In Section 7.12, we also looked at the case where firms that differ in market access compete against each other in the same market.

However, even when the firms share the same productivity and the same market access, they may set different prices due to different pricing constraints. I discuss two examples.

8.1. Sticky Prices

In New Keynesian macroeconomics, sticky prices are often modeled by imposing some constraints on the pricing behaviors of MC firms (see Gali 2015). For example, under the Rotemberg (1982) pricing rule, symmetric firms always set the same price, as in the Dixit-Stiglitz environment, but they need to pay the adjustment cost that is increasing the price change, so that the price adjusts sluggishly. Under the Calvo (1983) pricing rule, only a fraction of firms are randomly given the opportunity to reset their prices at each moment, so that individual prices can jump infrequently but the average price adjusts sluggishly, and at any point of time the firms are heterogeneous in their prices. Most models in this literature assume a fixed set of firms with no entry and use CES demand systems. Exceptions include the work of Bilbiie et al. (2007, 2014), who considered entry/exit under CES and translog with the Rotemberg pricing. Recently, Fujiwara & Matsuyama (2022) replaced CES and translog with HSA. The authors found that a higher entry cost and resulting market concentration causes a flattening of the Phillips curve in two cases: under the second law with the Rotemberg pricing and under the third law with the Calvo pricing. Because translog violates the third law, the latter case implies that, under translog and Calvo pricing, a higher entry cost and resulting market concentration would make the Phillips curve steeper, contrary to the empirical evidence. Fujiwara & Matsuyama (2022) also considered HDIA and HIIA, but a full general equilibrium analysis was feasible only under HSA.

8.2. Technology Diffusion and Competitive Fringes

Up to now, it has been assumed that each MC firm is the sole producer of its own product, and its market power is constrained only by the price elasticity of the demand curve it faces. In some cases, however, firms may be constrained by the presence of competitive fringes. For example, in the dynamic MC models by Judd (1985) and Matsuyama (1999), each firm pays the innovation cost to enter with its own product, for which it enjoys monopoly power only temporarily due to technology diffusion. After the loss of its monopoly power, the innovating firm is forced to sell its product at the marginal cost due to the presence of competitive fringes. Thus, different products are priced differently, depending on how recently they were introduced. This causes synchronization and endogenous fluctuation of innovation activities under some conditions.[37] However, these conditions are independent of market size in the models of Judd (1985) and Matsuyama (1999), both of which use CES demand system and feature the exogenously constant markup rate by the innovators while they enjoy the monopoly power. Matsuyama & Ushchev (2022a) replaced CES by HSA in the Judd model to allow for the second law and procompetitive entry. The Judd model under HSA remains analytically tractable, and the authors were able to demonstrate how a large market size makes endogenous fluctuations of innovation activities more likely. Matsuyama & Ushchev (2022c) considered the Judd model under HDIA, an extension of the Kimball (1995) aggregator that allows its product range to vary endogenously. This "Judd meets Kimball" model is not analytically tractable and can be solved only numerically.

9. CONCLUDING REMARKS

Instead of recapitulating, let me flag two issues that have not attracted much attention in the literature of MC models under non-CES demand systems.

9.1. Two-Sided Market Power

It is commonly assumed that MC firms are price takers in the factor markets. In this review, I follow this convention by taking labor as the numeraire. While useful as a benchmark, this assumption

[37]This is because a potential innovator needs to enter when the market for its product is large enough to recover the innovation cost. If an innovator chooses to enter when others do, most of the competing products are monopolistically priced. If an innovator enters after an innovation wave, it competes against products that are mostly priced competitively. This generates an incentive to innovate when others innovate, which creates an innovation wave. This, in turn, cause the market to become too saturated and innovation stops for a while, until the growth of the economy or obsolescence of the existing products makes innovation profitable again.

is restrictive to the extent that the firms offer jobs that are differentiated in the eyes of workers, so that even atomistic MC firms face upward-sloping labor supply curves, which give them the market power in the labor market. With such two-sided market powers, many of the predictions need to be reexamined. For example, incomplete pass-through is no longer prima-facie evidence for the second law of demand, because how the firms respond to shocks reflects the endogeneity not only of their price markup rates but also of their wage markdown rates, as pointed out by Ren & Zhang (2025).

9.2. Demand Elasticity Heterogeneity

Virtually all existing studies assume that the demand system for differentiated products sold by MC firms is generated by the representative consumers/firms. This assumption obviates the need to deal with many issues that arise if the demand system is instead generated by heterogeneous buyers who differ in their price elasticities of demand. For example, if MC firms cannot price-discriminate across buyers, their markup rates depend on the composition of their buyers. This could lead to the anti-second law, because a higher price shifts the demand composition toward the buyers with lower price elasticities. Indeed, this could lead to the symmetric firms pursuing different pricing strategies, as some sell mostly to buyers with low price elasticities at higher markup rates and others sell also to buyers with high price elasticities at lower markup rates.[38] Matsuyama & Ushchev (2024a) take a small step toward understanding a MC model with demand elasticity heterogeneity.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This review is based on my graduate lectures at Northwestern University, the University of Tokyo, and the European University Institute. I would like to thank the students for their questions. I also benefited from the opportunity to test-drive this review at the Federal Reserve Bank of Chicago. Actual writing was done mostly during my visit to the Becker Friedman Institute at the University of Chicago, whose hospitality is gratefully acknowledged. Much of the content is drawn from my joint work with Philip Ushchev, and this review benefited greatly from his comments. I also acknowledge valuable feedback from Arnaud Costinot, Santiago Franco, Gianmarco Ottaviano, Thomas Sampson, Kunal Sangani, and Dalton Zhang. The usual disclaimer applies.

LITERATURE CITED

Arkolakis C, Costinot A, Donaldson D, Rodríguez-Clare A. 2019. The elusive pro-competitive effects of trade. *Rev. Econ. Stud.* 86(1):46–80

Arrow KJ. 1971. Theory of risk aversion. In *Essays in the Theory of Risk-Bearing*, chap. 3. Chicago: Markham Publ.

[38]This issue is more acute under a nonhomothetic non-CES demand system generated by consumers with heterogeneous income. Even if all consumers share the same preferences, the rich could have lower price elasticities than the poor. Depending on the income distribution, some MC firms may choose to sell only to the rich by setting high markup rates, while other MC firms may choose to sell to both the rich and the poor at lower markup rates. Surprisingly, virtually none of the existing studies on MC models under nonhomothetic non-CES have addressed this issue, with those of Foellmi & Zweimueller (2006) and Foellmi et al. (2018) being the sole exceptions that I am aware of.

Baqaee D, Farhi E, Sangani K. 2024. The Darwinian returns to scale. *Rev. Econ. Stud.* 91(3):1373–405
Behrens K, Murata Y. 2007. General equilibrium models of monopolistic competition: a new approach. *J. Econ. Theory* 136:776–87
Benassy JP. 1996. Taste for variety and optimum production patterns in monopolistic competition. *Econ. Lett.* 52:41–47
Bertoletti P, Etro F. 2017. Monopolistic competition when income matters. *Econ. J.* 127:1217–43
Bilbiie FO, Fujiwara I, Ghironi F. 2014. Optimal monetary policy with endogenous entry and product variety. *J. Monet. Econ.* 64:1–20
Bilbiie FO, Ghironi F, Melitz MJ. 2007. Monetary policy and business cycles with endogenous entry and product variety. *NBER Macroecon. Annu.* 22:299–353
Boucekkine R, Latzer H, Parenti M. 2017. Variable markups in the long-run: a generalization of preferences in growth models. *J. Math. Econ.* 68:80–86
Calvo GA. 1983. Staggered prices in a utility-maximizing framework. *J. Monet. Econ.* 12(3):383–98
Costinot A, Vogel J. 2015. Beyond Ricardo: assignment models in international trade. *Annu. Rev. Econ.* 7:31–62
Dixit AK, Stiglitz JE. 1977. Monopolistic competition and optimal product diversity. *Am. Econ. Rev.* 67(3):297–308
Edmond C, Midrigan V, Xu DY. 2023. How costly are markups? *J. Political Econ.* 131(7):1619–75
Feenstra RC. 2003. A homothetic utility function for monopolistic competition models without constant price elasticity. *Econ. Lett.* 78:79–86
Foellmi R, Henestrick C, Zweimueller J. 2018. International arbitrage and the extensive margin of trade between rich and poor countries. *Rev. Econ. Stud.* 85:475–510
Foellmi R, Zweimueller J. 2006. Income distribution and demand-induced innovations. *Rev. Econ. Stud.* 73:941–60
Fujiwara I, Matsuyama K. 2022. *Competition and the Phillips curve*. CEPR Discuss Pap. 17521-3, Cent. Econ. Policy Res., London
Gali J. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton, NJ: Princeton Univ. Press. 2nd ed.
Grossman GM, Helpman E, Lhuillier H. 2023. Supply chain resilience: Should policy promote international diversification or reshoring? *J. Political Econ.* 131(12):3462–96
Hurwicz L, Uzawa H. 1971. On the integrability of demand functions. In *Preferences, Utility, and Demand: A Minnesota Symposium*, ed. JS Chipman, L Hurwicz, MK Richter, HF Sonnenschein, pp. 114–48. New York: Harcourt Brace Jovanovich
Judd KL. 1985. On the performance of patents. *Econometrica* 53:567–86
Kimball M. 1995. The quantitative analytics of the basic neomonetarist model. *J. Money Credit Bank.* 27(4):1241–77
Latzer H, Matsuyama K, Parenti M. 2020. *Reconsidering the market size effect in innovation and growth*. CEPR Discuss. Pap. 14250, Cent. Econ. Policy Res., London
Matsuyama K. 1995. Complementarities and cumulative processes in models of monopolistic competition. *J. Econ. Lit.* 33:701–29
Matsuyama K. 1999. Growing through cycles. *Econometrica* 67:335–47
Matsuyama K. 2008. Symmetry-breaking. In *The New Palgrave Dictionary of Economics*, ed. S Durlauf, LE Blume, pp. 6484–87. London: Palgrave Macmillan. 2nd ed.
Matsuyama K. 2019. Engel's law in the global economy: demand-induced patterns of structural change, innovation and trade. *Econometrica* 87:497–528
Matsuyama K. 2023. Non-CES aggregators: a guided tour. *Annu. Rev. Econ.* 15:235–65
Matsuyama K. 2024. *Constant relative and absolute pass-through families under H.S.A.* Unpublished note
Matsuyama K, Ottaviano GIP. 2024. *Globalization and firm selection with constant pass-through*. Work in progress
Matsuyama K, Ushchev P. 2017. *Beyond CES: three alternative classes of flexible homothetic demand systems*. CEPR Discuss. Pap. 12210, Cent. Econ. Policy Res., London
Matsuyama K, Ushchev P. 2020a. *When does procompetitive entry imply excessive entry?* CEPR Discuss. Pap. 14991, Cent. Econ. Policy Res., London
Matsuyama K, Ushchev P. 2020b. *Constant pass-through*. CEPR Discuss. Pap. 15475, Cent. Econ. Policy Res., London

Matsuyama K, Ushchev P. 2022a. Destabilizing effects of market size in the dynamics of innovation. *J. Econ. Theory* 200:105415

Matsuyama K, Ushchev P. 2022b. *Selection and sorting of heterogeneous firms through competitive pressures*. CEPR Discuss. Pap. 17092-2, Cent. Econ. Policy Res., London

Matsuyama K, Ushchev P. 2022c. *Destabilizing effects of market size in the dynamics of innovation: Judd meets Kimball*. Paper presented at presented at the 11th Workshop Dynamic Models in Economic and Finance, Urbino, Italy, Sept. 8–10

Matsuyama K, Ushchev P. 2023. *Love-for-variety*. CEPR Discuss. Pap. 18184-2, Cent. Econ. Policy Res., London

Matsuyama K, Ushchev P. 2024a. *Anatomy of market failures in monopolistic competition*. Paper presented at the Hitotsubashi-Gakushuin Conference on International Trade and FDI, Gakushuin Univ., Tokyo, Jpn., Dec. 14–15

Matsuyama K, Ushchev P. 2024b. *Destabilizing effects of market size in the dynamics of innovation: the role of per capita expenditure*. Paper presented at the 12th Workshop Dynamic Models in Economic and Finance, Urbino, Italy, Sept. 12–14

Matsuyama K, Ushchev P. 2024c. *Destabilizing effects of market size in the dynamics of innovation: the role of population size*. Paper presented at the 12th Workshop Dynamic Models in Economic and Finance, Urbino, Italy, Sept. 12–14

Melitz MJ. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71(6):1695–725

Melitz MJ. 2018. Competitive effects of trade: theory and measurement. *Rev. World Econ.* 154(1):1–13

Melitz MJ, Ottaviano GIP. 2008. Market size, trade, and productivity. *Rev. Econ. Stud.* 75(1):295–316

Mrázová M, Neary JP. 2017. Not so demanding: demand structure and firm behavior. *Am. Econ. Rev.* 107(12):3835–74

Ottaviano GIP, Tabuchi T, Thisse J-F. 2002. Agglomeration and trade revisited. *Int. Econ. Rev.* 43:409–35

Parenti M, Ushchev P, Thisse J-F. 2017. Toward a theory of monopolistic competition. *J. Econ. Theory* 167:86–115

Ren K, Zhang DR. 2025. *Price markups or wage markdowns?* Unpublished manuscript

Rotemberg JJ. 1982. Sticky prices in the United States. *J. Political Econ.* 90(6):1187–211

Samuelson PA. 1950. The problem of integrability in utility theory. *Economia* 17(68):355–85

Thisse JF, Ushchev P. 2018. Monopolistic competition without apology. In *Handbook of Game Theory and Industrial Organization*, Vol. I, ed. LC Corchón, MA Marini, pp. 93–136. Cheltenham, UK: Edward Edgar Publ.

Tirole J. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press

Trottner F. 2023. *Unbundling market power*. Work. Pap., Univ. Calif. San Diego, La Jolla

Zhelobodko E, Kokovin S, Parenti M, Thisse J-F. 2012. Monopolistic competition: beyond the constant elasticity of substitution. *Econometrica* 80(6):2765–84