Specialization vs Competition: An Anatomy of Increasing Returns to Scale

April 28, 2020

Abstract

We develop a model of monopolistic competition with a differentiated intermediate good and variable elasticity of technological substitution. The model allows to study the nature and origins of external increasing returns. We single out two sources of scale economies: specialization and competition. The former depends only on how TFP varies with input diversity, while the latter is fully captured by the behavior of the elasticity of substitution across inputs. This distinction gives rise to a full characterization of the rich array of competition regimes in our model. The necessary and sufficient conditions for each regime to occur are expressed in terms of the relationships between TFP and the elasticity of substitution as functions of the input diversity. Moreover, we demonstrate that, despite the folk wisdom resting on CES models, specialization economies are in general neither necessary nor sufficient for external increasing returns to emerge. This highlights the profound and non-trivial role of market competition in generating agglomeration economies and other phenomena driven by scale economies.

Keywords: External Increasing Returns; Variable Elasticity of Substitution; Specialization Effect; Competition Effect.

JEL Classification: D24, D43, F12, L13

1 Introduction

Consider the standard thought experiment in urban and regional economics: what if two economies (e.g., cities, regions) have a similar production structure, but one is larger than the other? The answer to this question is key for understanding some empirical patterns the New Economic Geography seeks to explain. For example, Rosenthal and Strange (2004) provide strong empirical evidence that wages in larger cities/regions are higher. Whether the same holds for prices is debatable. Indeed, Handbury and Weinstein (2015) describe higher prices in larger markets as a "common finding", but show that this relationship can be reversed after controlling for several measurement biases. In addition to that, Bellone et al. (2016) provide evidence that markups are lower at larger markets.

A plausible mechanism behind agglomeration economies relies on the assumption of *external increasing returns* (EIR henceforth).[1] Basically, the presence of EIR implies that ex post an increase in the size of the economy results in a more than proportional increase in aggregate output, even if the individual firm-level technologies have ex ante constant returns to scale (CRS henceforth). EIR emerge from various sources and play a big role in shaping market outcomes and spatial patterns.

One source of EIR is *specialization*: deeper division of labor boosts aggregate productivity. Another driving force of EIR is *market competition*: a larger market invites more firms, which eventually results in lower prices, lower markups, and higher firm-level output. To the best of our knowledge, the interactions between specialization and competition in shaping EIR have never been studied within a unified setting. Instead, these two forces have been mainly analyzed separately from each other within two different, although related, families of models.

In this paper, we look closer at the sources of EIR, and study how EIR channel the impact of city size on key equilibrium variables: wages, prices, and markups. To achieve our goal, we develop a two-sector model in the spirit of Ethier (1982) with a non-specified CRS production function in the final-good sector. The relationship between input diversity and TFP captures the *specialization/complexity* effect, while the behavior of the elasticity of substitution describes the *competition effect*.

Our main contribution is to demonstrate how using a non-specified production function with variable elasticity of substitution (VES henceforth) allows to study jointly specialization economies and competition effects. More precisely, we derive a *necessary and sufficient condition* for EIR in the final good sector to occur. This condition is expressed in terms of the TFP and the elasticity of substitution as functions of input diversity. We demonstrate that specialization economies per se are, in general, neither necessary nor sufficient for EIR to emerge. What matters is the *interplay* between the specialization/complexity effect and the competition effect. This unexpected result stands in a sharp

[1] See, e.g., Ch. 2 in Duranton and Puga (2004) or Ch. 3 in Fujita and Thisse (2013).

contrast to what happens when the world is CES: the competition effect is not present due to zero impact of entry on profit-maximizing markups. To sum up, in the CES case, "specialization economies = EIR". This explains why specialization economies have long been viewed as the dominant factor of scale economies, while the impact of market competition was, in this regard, definitely underestimated. On the contrary, our result highlights the non-trivial role of market competition in generating agglomeration economies and other phenomena driven by EIR (including, potentially, endogenous long-run economic growth).

Our other findings are as follows. First, we fully describe the rich array of equilibrium regimes in our model, and characterize the impact of horizontal innovation on prices, markups, and wages.[2] More precisely, we find necessary and sufficient conditions for competition to be (i) either price-decreasing or price-increasing, (ii) either markup-decreasing or markup-increasing, and (iii) either wage-increasing or wage-decreasing. The first condition involves only TFP as a function of input diversity, the second – only the elasticity of substitution, while the third blends both.

Second, we find that the competition effect may either reinforce or weaken the impact of the specialization effect on aggregate output. In our analysis, how the competition effect interacts with the specialization effect depends on whether the elasticity of substitution across the intermediate inputs (evaluated at a symmetric outcome) decreases or increases as a function of input diversity.

Third, our approach allows for complexity externalities, which may lead to a reduction of TFP in the final good sector in response to expanding variety of intermediate inputs.[3]

Finally, our main results hold for any production function which satisfies the properties of symmetry, strict quasi-concavity, and CRS, as well as having well-defined marginal products of inputs.

We believe that our contribution makes an important theoretical advancement compared to recent work on monopolistic competition with VES on the consumer's side. As an example, Zhelobodko et al. (2012) only distinguish between price-increasing and price-decreasing competition, as in their model prices and markups always move in the same direction in response to market size shocks. The reasons behind the deep differences in the results of the two settings, despite their formal similarity, are as follows. The counterpart of our TFP function in Zhelobodko et al. (2012) would be the representative consumer's utility level as a function of product variety. Due to the love for variety, which is a standard assumption in this family of models, the utility level would always increase in product diversity. Hence, no consumption-side counterpart of complexity diseconomies would occur. Moreover, even if we choose to introduce (e.g. á la Benassy, 1996) a negative consumption externality in order to capture a cost of processing information about more varieties, doing so will not affect the equilibrium pattern. Indeed, as shown by Dhingra and Morrow (2019), the

[2] Following the literature, what we understand by "horizontal innovation" is entry of new intermediate input producers.

[3] Examples of how this externality may work in growth theory can be found in Howitt (1999), Dalgaard and Kreiner (2001), and Bucci (2013).

behavior of the utility level in this type of models is crucial for welfare analysis, but is totally unrelated to the properties of free-entry equilibrium, while the elasticity of substitution yields a sufficient statistic for equilibrium behavior. On the contrary, in our model, the market outcome is determined by the interaction between the specialization/complexity effect and the competition effect. Having two primitives – the TFP function and the elasticity of substitution – instead of just one is the key property of our model which allows us to study jointly specialization/complexity and competition effects and results in a richer array of equilibrium patterns.

Literature review. The "specialization vs competition" dichotomy has been conceptualized by Adam Smith (1776), who was a prominent spokesman in favor of both a deeper division of labor and freer competition (Sandmo, 2011, Ch. 3).

Specialization effects are unequivocally captured by *Ethier-type models* (Ethier, 1982), i.e. two-sector monopolistic competition models in which the final good sector technology displays constant elasticity of substitution (CES) across differentiated inputs. The treatment of competition effects in the modern economic literature is, instead, at least twofold. On the one hand, in differentiated oligopoly models used in industrial organization, competition effect is often synonymous to a pro-competitive effect: markups fall with the the number of competitors due to strategic interactions. A notable exception is Chen and Riordan (2008), who study *price-increasing competition* using a discrete-choice model of product differentiation. On the other hand, *Dixit-Stiglitz-type models* (Dixit and Stiglitz, 1977), i.e. monopolistically competitive environments in which the final consumption good is differentiated, have been recently revisited and extended to the VES case, both on the consumer's side (Behrens and Murata, 2007; Bilbiie et al., 2012; Bertoletti and Etro, 2016, 2017) and on the producer's side (Kimball, 1995; Smets and Wouters, 2007). In these models, firms do not behave strategically because each separate firm is negligibly small compared to the whole market.[4] However, markups still vary in response to entry and exit of firms due to non-isoelastic demands. Just like the one based on differentiated oligopoly, the approach of monopolistic competition with VES is flexible enough to capture both price-decreasing and price-increasing competition (Zhelobodko et al., 2012; Parenti et al., 2017). Based on that, "competition effects" and "pro-competitive effects" are not synonymous to us in the rest of the paper. By competition effects we mean any changes in the market outcome channeled by the VES between differentiated varieties, without expecting a priori any particular direction of those changes. In contrast, specialization effects are those channeled through changes of TFP in response to entry/exit of firms.

It is worth stressing that the *joint* role of specialization and competition in generating EIR and shaping market outcomes has received, till now, very little attention in the literature. Studying these two forces is an intriguing task, as they may give rise to opposite effects on aggregate production. As

[4] In a recent survey, Thisse and Ushchev (2018) discuss the deep linkages between differentiated oligopoly models and non-atomic monopolistic competition with VES.

pointed out by Kremer (1993), more complex technologies involving a larger number of production tasks and/or more differentiated intermediate inputs may be detrimental to manufacturing activities, e.g. due to higher risks of failure. In other words, *complexity diseconomies*, as opposed to *specialization economies*, may occur.[5] Furthermore, as discussed above, competition may be either price-decreasing or price-increasing. These considerations suggest that the interaction between the two effects may lead to rich and unexpected market outcomes with non-trivial implications for urban and regional economics.

We believe that the main reason why the interaction between specialization and competition in generating EIR has definitely been understudied in the literature can mostly be found in the large popularity of the CES assumption. This assumption is appealing as it makes models tractable. The flip-side is that the equilibrium markup, which may serve as a reverse measure of the toughness of competition, remains unaffected by entry and/or by market-size shocks. As a consequence, *the CES assumption eliminates the competition effect*. The Marshallian externalities approach introduced by Abdel-Rahman and Fujita (1990) to study agglomeration economies at the city level widely use the CES assumption,[6] and so do many other strands of economic literature.[7] Hence, neither of these literatures allows distinguishing clearly between the impacts of specialization/complexity and toughness of competition on aggregate output and wages. This paper aims to fill this gap.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 characterizes the equilibrium for a given number of input-producing firms. We also suggest a classification of competitive regimes in the intermediate input sector, based on the impact of entry on prices, markups, and wages. Section 4 deals with a free-entry equilibrium, and studies how the interaction be-

[5] To measure complexity diseconomies empirically, Hidalgo et al. (2007) use the average product density as a proxy for the degree of complexity of a country's production structure: an economy with a higher average product density must possess more capabilities and technical competences to produce simultaneously within a denser space of activities. Using data on net trade flows to compute an average product density index, Ferrarini and Scaramozzino (2016) provide more recent empirical evidence for complexity diseconomies.

[6] Duranton and Puga (2004) and, more recently, Fujita and Thisse (2013) provide extensive surveys of this strand of literature.

[7] Just to make an example, the CES assumption is almost ubiquitous in endogenous growth models with horizontal innovation (Grossman and Helpman, 1990; Romer, 1990). These models generally highlight the positive effects of specialization, disregarding other possible effects which may stem from an increase in the toughness of market competition. Two notable exceptions are Bucci and Matveenko (2017), and Boucekkine et al. (2017). Bucci and Matveenko (2017) extend Romer (1990) to the case of a non-CES, non-homothetic aggregate production function in the final good sector, which leads to variable markups in the intermediate good sector. Each firm's markup is a function of the firm-level output, which in turn is affected by the TFP in the final good sector. In Boucekkine et al. (2017), variable markups emerge due to a non-CES preference structure in a horizontal innovation growth model à la Grossman and Helpman (1991, Ch. 3) . While in Boucekkine et al. (2017) firms' market power and long-run growth are negatively related, in Bucci and Matveenko (2017) the relation between markups and economic growth is non-monotonic, which seems to match better the empirical evidence (Aghion et al., 2005; Aghion and Griffith, 2005). See the concluding section of the present paper for a deeper discussion about these issues.

tween the specialization/complexity effect and the competition effect generates EIR. Section 5 discusses possible applications of our framework to some issues of urban and regional economics (including urban hierarchy, a micro-foundation of quantitative spatial models, and a justification for the use of flexible empirical strategies in studying the relationship between city size and wages). Section 6 concludes.

2 The Model

The economy is composed of two vertically related sectors. The intermediate inputs sector (sector $\mathcal{I}$ henceforth) produces a differentiated intermediate good under monopolistic competition. The number of firms in this sector ($\mathcal{I}$-firms henceforth) is endogenous due to free entry, while the only production factor is labor. Each worker inelastically supplies one unit of homogeneous labor. The labor market is perfectly competitive.

The final good sector (sector $\mathcal{F}$ henceforth) involves a unit mass of perfectly competitive firms ($\mathcal{F}$-firms henceforth) sharing the same CRS technology, which uses varieties of the intermediate good as inputs. The departure of our modeling strategy from Ethier (1982) is that we drop the standard CES assumption, working instead with a non-specified CRS production function.

2.1 Sector $\mathcal{F}$

The production of the homogeneous final good requires a continuum $[0, n]$ of inputs, each representing a specific variety of a horizontally differentiated intermediate good. All firms operating in sector $\mathcal{F}$ are endowed with the same production function F:

$$Y = F(\mathbf{q}, n), \tag{1}$$

where Y is the output of the final good, $\mathbf{q} = (q_i)_{i\in[0,n]}$ is the vector of inputs used in production, n stands for the number (or, more precisely, the *mass*) of intermediate inputs.[8]

We make the following assumptions. First, $F(\mathbf{q})$ is concave in $\mathbf{q}$, i.e., each input exhibits a diminishing marginal product (see Appendix A1 for a mathematical definition of the marginal product under a continuum of inputs). Second, $F(\mathbf{q})$ is positive homogeneous of degree 1. Third, $F(\mathbf{q})$ is *symmetric*, i.e., any permutation of intermediates does not change the output Y. The reason for imposing such a symmetry, which typically holds in monopolistic competition contexts, is to refrain from placing any ad hoc asymmetries on sector $\mathcal{I}$. Finally, we do not make any specific assumption about the sign of $\partial F/\partial n$: the production externality can be either positive or negative. A negative production externality could mean that, even if a firm eventually chooses not to use a newly developed type of input, collecting information about it still requires some cost.

[8] Note that the production function (1) involves n not only via $\mathbf{q}$, but also as a separate argument, which captures a *production externality*.

To show that our approach encompasses a broad range of technologies used in the literature, we provide a gallery of examples.

1. CES: variations on a theme. Our first example is the standard CES production function:

$$F(\mathbf{q}) \equiv \left(\int_0^n q_i^\rho \mathrm{d}i \right)^{1/\rho}, \quad 0 < \rho < 1. \tag{2}$$

One possible extension of (2) is to introduce a multiplicative TFP term varying with n (Ethier, 1982; Benassy, 1996):

$$F(\mathbf{q}) = n^\nu \left(\int_0^n q_i^\rho \mathrm{d}i \right)^{1/\rho}, \quad 0 < \rho < 1. \tag{3}$$

The TFP factor n^ν in (3) captures the production externality. When such externality is positive ($\nu > 0$), we deal with specialization economies, otherwise we deal with complexity diseconomies. See Appendix A2 for more details.

Second, the constant ρ in (2) may be replaced by a function $\rho(n)$, as in Gali (1995), who assumes $\rho'(n) > 0$, i.e. that varieties become better technological substitutes as their number increases.

2. Kimball's flexible aggregator. Kimball (1995) proposed a flexible class of production functions implicitly defined by:

$$\int_0^n \phi\left(\frac{q_i}{Y}\right) \mathrm{d}i = 1, \tag{4}$$

where $\phi(\cdot)$ is an increasing, strictly concave, and sufficiently differentiable function.[9] When $\phi(\cdot)$ in (4) is a power function, we fall back to the CES-case. Using Kimball's flexible aggregator is a well established way to generate variable markups in macroeconomic and trade models (Smets and Wouters, 2007; Amiti et al., 2019).

3. Translog technologies. Two other examples of tractable non-CES technologies are given by the translog production function (Kim, 1992) and translog price index (Feenstra, 2003), given, respectively, by:

$$\ln F(\mathbf{q}) = \frac{1}{n} \int_0^n \ln q_i \mathrm{d}i - \frac{\alpha}{2} \left[\int_0^n (\ln q_i)^2 \mathrm{d}i - \frac{1}{n} \left(\int_0^n \ln q_i \mathrm{d}i \right)^2 \right], \tag{5}$$

$$\ln P(\mathbf{p}) = \frac{1}{n} \int_0^n \ln p_i \mathrm{d}i - \frac{\beta}{2} \left[\int_0^n (\ln p_i)^2 \mathrm{d}i - \frac{1}{n} \left(\int_0^n \ln p_i \mathrm{d}i \right)^2 \right]. \tag{6}$$

[9] To guarantee that a solution to (4) does exist for any n, one should assume additionally that $\phi(0) \leq 0$, while $\phi(\infty) = \infty$.

The factor $1/n$ in (5) – (6) captures negative production externalities.[10]

We think that working with an arbitrary well-behaved symmetric CRS technology allows for more flexibility compared to focusing on Kimball's aggregator (4) or any other reasonably broad subclass of symmetric CRS production functions. In Section 2.3, we provide a mathematically more precise argument to this view.[11]

Specialization economies vs complexity diseconomies. We are now equipped to give precise definitions for the specialization economies and complexity diseconomies. Consider the behavior of F at a symmetric outcome, i.e. when $q_i = q$ for all $i \in [0, n]$. Denote by $\varphi(n)$ the level of output that can be produced when a firm uses one unit of each intermediate input. Formally, $\varphi(n) \equiv F\left(\mathbb{I}_{[0,n]}, n\right)$, where $\mathbb{I}_S$ is an indicator of $S \subseteq \mathbb{R}_+$. Given $\mathcal{F}$-firm's total expenditure E on intermediate inputs under unit prices, the *specialization economies* capture the idea that the division of labor generates productivity gains, namely a larger variety of intermediate inputs allows to produce a larger amount of final output. To put this in a more formal way, note that, because of CRS, output of the final good equals $q\varphi(n)$ when q units of each intermediate are employed. Hence, the specialization effect takes place if and only if

$$\frac{E}{n}\varphi(n) > \frac{E}{k}\varphi(k), \quad \text{where } k < n.$$

In other words, *specialization economies occur if and only if $\varphi(n)/n$ increases with n*, or, equivalently, when the elasticity of $\varphi(n)$ exceeds 1:

$$\frac{\varphi'(n)n}{\varphi(n)} > 1. \tag{7}$$

Otherwise output of the final good decreases with the intermediate inputs' range. In the latter case, we face *complexity diseconomies*.

To make these general definitions more intuitive, we illustrate them in Appendix A2 for the case of the augmented CES production technology (3).

TFP function. Since $\varphi(n)/n$ captures how total output of the final good varies with input diversity, the total quantity of the differentiated input employed being fixed, we find it reasonable to dub $\varphi(n)/n$ the *TFP function*. Since this function will play a crucial role in what follows, we choose to treat it as one of the two fundamental primitives of our model (the second one to be defined below). We may equivalently define *specialization economies as the situation when the TFP function increases in n*.

Specialization vs complexity: **a dual description**. Each $\mathcal{F}$-firm seeks to minimize production costs per unit of output:

$$\min_{\mathbf{q}} \int_0^n p_i q_i \mathrm{d}i \qquad \text{s.t.} \qquad F(\mathbf{q}) \geq 1. \tag{8}$$

[10] As shown by Matsuyama and Ushchev (2017), the translog technologies (5) – (6) are beyond Kimball's flexible aggregator (4).

[11] See also Table 1 and the discussion below (Section 3.2).

taking the total output Y as given. The value function $P(\mathbf{p})$ of the $\mathcal{F}$-firm's problem (8) is the *price index*, or the *unit cost function*.[12] By the duality principle, the production function $F(\mathbf{q})$ and the unit cost function $P(\mathbf{p})$ can be uniquely recovered from each other. Furthermore, the duality principle allows a convenient description of the trade-off between specialization and complexity in terms of the price index. Namely, when $p_i = p$ for all $i \in [0, n]$, then the final-good producer will purchase all inputs in equal volumes: $q = Y/\varphi(n)$. Consequently, the price index can be expressed as follows:

$$P = \frac{n}{\varphi(n)}\, p. \tag{9}$$

Combining (9) with our definition of specialization economies, we conclude that the price index decreases with the range n of inputs if and only if specialization economies take place.

Let us use examples to gain more intuition about the specialization/complexity trade-off. For the standard CES technology (2), the TFP function is a power function of the form $\varphi(n)/n = n^{1/(\sigma-1)}$. Since $\sigma > 1$, specialization economies take place. The same is true for any production function described by the Kimball's flexible aggregator (4) with $\phi(0) = 0$, for which $\varphi(n)/n = \phi^{-1}(1/n)$. On the contrary, the translog production function (5) exemplifies complexity diseconomies, as in this case the TFP function, $\varphi(n)/n = 1/n$, decreases with n. Finally, the dual translog (6) is a *borderline* case, since the TFP function is constant: $\varphi(n)/n = 1$. As a result, specialization and complexity fully balance each other.

The following result summarizes the properties of the above classes of production functions.

Proposition 1.

(i) *Kimball's flexible-aggregator technologies* (4) *satisfying* $\phi(0) = 0$ *exhibit specialization economies*;

(ii) *the translog production function* (5) *generates complexity diseconomies*;

(iii) *the translog cost function (6) strikes the exact balance between specialization economies and complexity diseconomies.*

Proof. See Appendix A3. □

Proposition 1 highlights the flexibility of our approach, which encompasses a wide variety of such technologies. In particular, our way of modeling production technology is more general than Kimball's flexible aggregator, as Kimball-type production functions include neither the augmented CES, nor the translog technologies.

[12] In the CES case, the price index is given by the well known formula:

$$P(\mathbf{p}) = \left(\int_0^n p_i^{1-\sigma} \mathrm{d}i \right)^{1/(1-\sigma)}.$$

2.2 Sector $\mathcal{I}$

There is a continuum of intermediate input producers sharing the same technology, which exhibits increasing returns to scale. We believe the assumption that technologies are identical across firms is not critical, since Combes et al. (2012) provide solid empirical evidence that productivity advantages of larger cities stem from agglomeration economies rather than selection effects.

Firm i's labor requirement for producing output q_i is given by $f + cq_i$, where $f > 0$ is the fixed cost and $c > 0$ is the constant marginal production cost. Thus, the profit π_i of firm i is defined by $\pi_i \equiv (p_i - cw)q_i - f$, where w is the wage rate. The inverse demand firm i faces stems from the first-order condition in the cost minimization problem in the $\mathcal{F}$-sector:

$$\frac{p_i}{P(\mathbf{p})} = \Phi(q_i, \mathbf{q}), \tag{10}$$

where $\Phi(q_i, \mathbf{q}) \equiv \partial F/\partial q_i$ is the marginal product of input i,[13] while $P(\mathbf{p})$ is the price index defined by (8), which now plays the role of a market aggregator, as it captures all the cross-price effects in the demand system for intermediate inputs. See Appendix A4 for derivation of (10). Note that, since $F(\mathbf{q})$ satisfies diminishing marginal returns, the inverse demand schedules (10) are downward-sloping.

Firm i seeks to maximize its profit:

$$\max_{q_i} \left[(P(\mathbf{p})\Phi(q_i, \mathbf{q}) - cw)\, q_i \right], \tag{11}$$

Because sector $\mathcal{I}$ is monopolistically competitive, each $\mathcal{I}$-firm takes the price index P as given. Hence, the first-order condition for (11) is given by

$$\Phi(q_i, \mathbf{q}) + q_i \frac{\partial \Phi}{\partial q_i} = \frac{cw}{P}. \tag{12}$$

Furthermore, given the mass n of $\mathcal{I}$-firms, the quantity profile $\mathbf{q}$ must satisfy the labor balance condition

$$c \int_0^n q_i \mathrm{d}i + fn = L, \tag{13}$$

which equates total labor supply to total labor demand.

The second-order condition, as well as technical details of possibly multiple solutions, are discussed in Appendix A5.

Elasticity of substitution and competition effect. The first-order condition (12) for profit maximization may be recast as

[13] Formally, the partial derivatives $\partial F/\partial q_i$ are not well-defined in the case of a continuum of inputs, which may seem to be an obstacle for working within a framework where the functional F is non-specified. It turns out, however, that putting slightly more structure on the space of input vectors $\mathbf{q}$ potentially available for the final good producers makes things work *as if* the marginal products were well-defined. See Appendix A1 for technical details.

$$\frac{p_i - cw}{p_i} = \eta(q_i, \mathbf{q}), \tag{14}$$

where η is the *marginal product elasticity*:

$$\eta(q_i, \mathbf{q}) \equiv -\frac{\partial \Phi}{\partial q_i} \frac{q_i}{\Phi(q_i, \mathbf{q})}. \tag{15}$$

At a symmetric outcome, when $p_i = p$ and $q_i = q$ for all $i \in [0, n]$, (14) boils down to

$$\frac{p - cw}{p} = \frac{1}{\sigma(n)}, \tag{16}$$

where $\sigma(n)$ is the elasticity of technological substitution[14] evaluated at the symmetric outcome:

$$\sigma(n) \equiv \left. \frac{1}{\eta(q_i, \mathbf{q})} \right|_{q_j = q_i \ \forall j \in [0,n]}. \tag{17}$$

We are now equipped to define the competition effect. The pricing rule (16) implies that the profit-maximizing markup equals $1/\sigma(n)$, hence $\sigma(n)$ represents toughness of competition among the $\mathcal{I}$-firms. In particular, $\sigma'(n) > 0$ means that competition gets tougher when more firms enter. We call such competition *markup-decreasing*. In the opposite case, when $\sigma'(n) < 0$, competition is *markup-increasing*. Hence, like the specialization/complexity effect, competition effect may be of either sign as well.

To grasp the intuition of the above definitions, consider some examples. Under the CES technology, profit-maximizing markups are unaffected by entry of new $\mathcal{I}$-firms. Under the translog technologies, the profit-maximizing markups are given by:

Translog production function	Translog expenditure function
$1 - \alpha n$	$\frac{1}{1+\beta n}$

Hence, both these technologies induce markup-decreasing competition. Finally, when the production function is given by Kimball's flexible aggregator (4), we have

$$\frac{1}{\sigma(n)} = \left. -\frac{\xi \phi''(\xi)}{\phi'(\xi)} \right|_{\xi = \phi^{-1}(1/n)}. \tag{18}$$

In this case, competition is markup-decreasing if and only if the elasticity of $\phi'(\xi)$ in ξ evaluated at $\xi = \phi^{-1}(1/n)$ is an increasing function. To illustrate how markup-increasing competition may arise, consider a Kimball-type production function whose aggregator function $\phi(\cdot)$ is given by $\phi(\xi) = \xi^a + \xi^b$, where $0 <$

[14] As defined by Nadiri (1982). Parenti et al. (2017) extend Nadiri's definition to the case of a continuum of inputs and prove that (17) holds true. Note that $\sigma(n)$ is independent of firm-level output q because, due to the CRS property of $F(\mathbf{q})$, $\eta(q_i, \mathbf{q})$ is homogeneous of degree zero in $(q_i, \mathbf{q})$.

$a < b < 1$. It is readily verified that this technology gives rise to a "bipower" inverse demand schedule (Mrázová and Neary, 2017). The elasticity of $\phi'(\xi)$ is given by

$$-\frac{\xi\phi''(\xi)}{\phi'(\xi)} = 1 - a - \frac{b(b-a)}{b + a\xi^{-(b-a)}},$$

and decreases in ξ, hence competition is markup-increasing.

2.3 The TFP and the elasticity of substitution as the primitives of the model

We have already noticed that the TFP function $\varphi(n)/n$ and the elasticity of substitution $\sigma(n)$ determine the key properties of, respectively, the $\mathcal{F}$-sector and the $\mathcal{I}$-sector behavior. For this reason, we choose to treat these two objects as the *primitives* of the model. We show below that both the taxonomy of competition regimes (Section 3) and the condition for EIR to emerge (Section 4) will be characterized in terms of properties of these two functions.

We now come back to the issue of justifying the level of generality we choose to work at. Indeed, at this level of generality, the TFP function and the elasticity of substitution may be viewed as two *independent ingredients* of our approach, in the sense that the information about one of them is generically insufficient to recover the other, which makes both of them the true primitives of our model. Focusing on a more specific class of technologies would imply a non-trivial relationship between the two. To illustrate this point, consider the family of production functions described by Kimball's flexible aggregator (4). In this case, the TFP function is given by

$$\frac{\varphi(n)}{n} = -\frac{\phi(\xi)}{\xi}\bigg|_{\xi=\phi^{-1}(1/n)}, \tag{19}$$

while the elasticity of substitution satisfies (18). As a consequence, the two fundamentals are linked via the aggregator function $\phi(\cdot)$, which allows unambiguously recovering one of them from the other. Therefore, focusing on certain classes of production functions may lead to a priori unsuspected restrictions on the primitives of the model.

3 Entry and the market outcomes

In this section, we study how horizontal innovation affects the market outcomes. We fully characterize the behavior of the economy in response to expanding input diversity. Our purpose here is to highlight the fundamental role of the relationship between the TFP function and the elasticity of substitution in shaping various competition regimes.

3.1 Equilibrium for a given number of $\mathcal{I}$-firms

We choose the final good to be the numeraire by normalizing its price to 1. Hence, the profit of an $\mathcal{F}$-firm is given by $(1-P)Y$. Perfect competition suggests free entry, which implies $P=1$. Combining this with (9) pins down the equilibrium input price at a symmetric outcome:

$$p^*(n) = \frac{\varphi(n)}{n}. \tag{20}$$

Equation (20) reveals that the equilibrium input price equals the TFP. Hence, *the input price increases with the number of firms n in sector $\mathcal{I}$ if and only if specialization economies occur* (see Section 2.1).

We now find the equilibrium magnitudes other than $p^*(n)$. First, combining (16) with (20) and $P=1$ yields the wage rate:

$$w^*(n) = \frac{1}{c}\frac{\sigma(n)-1}{\sigma(n)}\frac{\varphi(n)}{n}. \tag{21}$$

Second, plugging (21) into the product market balance $Y = Lw$, we obtain the aggregate output:

$$Y^*(n) = \frac{L}{c}\frac{\sigma(n)-1}{\sigma(n)}\frac{\varphi(n)}{n}, \tag{22}$$

Equations (21) and (22) provide decomposition of, respectively, equilibrium wage and output[15] into the *competition effect* captured by $[\sigma(n)-1]/\sigma(n)$, and the *specialization/complexity effect* captured by $\varphi(n)/n$. The former increases with n if and only if $\sigma'(n) > 0$, while the latter increases if specialization economies prevail over complexity diseconomies.

Finally, to determine the per-firm output $q^*(n)$ in sector $\mathcal{I}$, we observe that at a symmetric equilibrium the labor balance condition (13) takes the form:

$$(cq+f)n = L \tag{23}$$

Solving (23) with respect to q, we find:

$$q^*(n) = (L-fn)/(cn).$$

Clearly, $q^*(n)$ always decreases with the mass n of input-producing firms.

3.2 The impact of entry on prices, wages, and markups

Prices, wages, and markups are all endogenous in our model. Putting together (20), (16), and (21), we observe that the entry of new firms need not move these variables in the same direction. In what follows, we say that competition is (i) *price-decreasing* if $\mathrm{d}p^*(n)/\mathrm{d}n < 0$, and *price-increasing* otherwise; (ii)

[15] Up to, respectively, the coefficients $1/c$ and L/c, which are independent of the input diversity n.

markup-decreasing if $\mathrm{d}[(p^*(n) - cw^*(n))/p^*(n)]/\mathrm{d}n < 0$, and *markup-increasing* otherwise; (iii) *wage-decreasing* if $\mathrm{d}w^*(n)/\mathrm{d}n < 0$, and *wage-increasing* otherwise.

Proposition 2 summarizes the main results of Subsection 3.1 in terms of the above taxonomies.

Proposition 2. *Competition is*:

(*i*) *price-increasing if specialization economies prevail over complexity diseconomies, and price-decreasing otherwise*;

(*ii*) *markup-decreasing if* $\sigma'(n) > 0$, *and markup-increasing otherwise*;

(*iii*) *wage-increasing if and only if the inequality*

$$\frac{\varphi'(n)n}{\varphi(n)} + \frac{1}{\sigma(n)-1}\frac{\sigma'(n)n}{\sigma(n)} > 1 \tag{24}$$

holds, and wage-decreasing otherwise.

Proof. Parts (*i*) and (*ii*) follow, respectively, from (20) and (16). To prove part (*iii*), we differentiate both parts of (21) with respect to n, which yields:

$$\frac{\mathrm{d}w^*(n)}{\mathrm{d}n} = \frac{1}{cn}\frac{\sigma(n)-1}{\sigma(n)}\frac{\varphi(n)}{n}\left(\frac{\varphi'(n)n}{\varphi(n)} + \frac{1}{\sigma(n)-1}\frac{\sigma'(n)n}{\sigma(n)} - 1\right).$$

Hence, we have: $\mathrm{d}w^*(n)/\mathrm{d}n > 0$ if and only if (24) holds. This proves part (iii) and completes the proof. □

The intuition behind Proposition 2 is as follows. Whether competition is price-decreasing or price-increasing is determined solely by the properties of the TFP function $\varphi(n)/n$. In contrast, the behavior of markups in response to entry can be fully characterized in terms of the elasticity of substitution $\sigma(n)$ across input varieties. Finally, both $\varphi(n)$ and $\sigma(n)$ play a role in determining the impact of entry in sector $\mathcal{I}$ on the equilibrium wage. The condition (24) states that, for the market outcome to be wage-increasing, it must be that either TFP, or the elasticity of substitution, or both grow sufficiently fast in n.

Two more comments are in order. First, (24) yields a necessary and sufficient condition for EIR to emerge (see Proposition 3 in Section 4). Indeed, since this condition blends $\varphi(n)$ and $\sigma(n)$, it reveals that neither the TFP function alone nor the elasticity of substitution alone provide sufficient information to say whether external scale economies take place. Ultimately, what matters is the interplay between the two. Thus, both market interactions and non-market forces matter for EIR to emerge.

Second, Proposition 2 highlights the novelty of our results compared to Zhelobodko *et al.* (2012), who work with a one-sector Dixit-Stiglitz-type framework and find that entry leads to a drop or a hike in markups depending solely on how the elasticity of substitution varies with the individual consumption level. Moreover, the framework of Zhelobodko *et al.* (2012) neither allows to distinguish between markup-decreasing and price-decreasing competition, nor entails any effects of entry on the wage. Thus, our model provides a richer and subtler

taxonomy of market competition. To see this in more detail, Table 1 provides a compact summary of our results for different types of production functions.

	Translog cost function	CES production function	Translog production function
Price	No effect	$\uparrow$	$\downarrow$
Markup	$\downarrow$	No effect	$\downarrow$
Wage	$\uparrow$	$\uparrow$	No effect

Table 1: The impact of entry on prices, markups and wages for different types of production functions

Table 1 shows that under the translog cost function prices are neutral to entry, markups decrease with entry, and wages increase with entry. In the CES case, both prices and wages increase in response to entry, while the markups remain unchanged. Finally, under a translog production function wages remain unchanged when new firms enter, while prices and markups fall. These comparisons highlight the key role of the interaction between the specialization/complexity effect.

As revealed by part (i) of Proposition 2, our approach allows for both price-increasing and price-decreasing competition, depending on whether specialization economies prevail or not over complexity diseconomies. This result is important because it can be viewed as a theoretical answer to the inconclusive empirical findings discussed in the Introduction.

Most empirical work tends to suggest that larger markets exhibit *higher prices, lower markups, and higher wages.* Table 1 reveals that neither the CES production function, nor any of the two translog technologies can fully capture all these patterns alone. However, Proposition 2 suggests a qualified answer to the question of what kind of production function might ultimately work. According to (24), if competition is both price-increasing and markup-decreasing, then it is also wage-increasing. Hence, any production function that satisfies both $n\varphi'(n)/\varphi(n) > 1$ (specialization economies) and $\sigma'(n) > 0$ (increasing elasticity of substitution) generates price-increasing, markup-decreasing and wage-increasing competition.

4 External increasing returns to scale

This section provides the main result of this paper, as it describes how the mutual interaction between the specialization/complexity tradeoff and the competition effect generates EIR.

4.1 Free-entry equilibrium

Define a *symmetric free-entry equilibrium* as a vector (p^*, q^*, n^*, w^*, Y^*), which satisfies (20), (16), (21), the labor balance condition (23), and the zero profit condition:

$$(p - cw)q = wf. \tag{25}$$

Equilibrium number of firms. We first pin down the equilibrium number n^* of $\mathcal{I}$-firms. To do so, we restate (25) as follows:

$$\frac{p - cw}{p} = \frac{f}{f + cq}, \tag{26}$$

i.e., the markup equals the share of fixed cost in an $\mathcal{I}$-firm's total production cost. Combining (26) with the pricing rule (16) and the labor balance (23), we obtain:

$$\sigma(n) = \frac{L}{fn}. \tag{27}$$

The equilibrium number of firms n^* is uniquely pinned down by (27) if and only if the elasticity of $\sigma(n)$ exceeds -1, which holds when $\sigma(n)$ is either increasing or decreasing not too fast in n.

Specialization and competition under free entry. Given n^*, using (16) and (25) yields the equilibrium firm's size:

$$q^* = \frac{f}{c} \cdot [\sigma(n^*) - 1]. \tag{28}$$

According to (28), any (f/c)-preserving shock that generates additional entry in the intermediate sector would lead to a larger firm size q^* if and only if $\sigma'(n) > 0$.

Plugging (28) into the production function of sector $\mathcal{F}$, we obtain the aggregate output Y^*:

$$Y^* = \frac{L}{c} \cdot \frac{\sigma[n^*(L)] - 1}{\sigma[n^*(L)]} \cdot \frac{\varphi[n^*(L)]}{n^*(L)}, \tag{29}$$

while plugging n^* into (21) pins down the equilibrium wage w^*:

$$w^* = \frac{1}{c} \cdot \frac{\sigma[n^*(L)] - 1}{\sigma[n^*(L)]} \cdot \frac{\varphi\,[n^*(L)]}{n^*(L)}. \tag{30}$$

In equations (29) – (30), the term $[\sigma(n^*)-1]/\sigma(n^*)$ captures the competition effect, which stems from sector $\mathcal{I}$. This term increases with n, hence with the population size L, if and only if competiton is markup-decreasing. The term $\varphi(n^*(L))/n^*(L)$ represents the specialization/complexity effect and increases with population if and only if specialization economies take place, which is equivalent to price-increaing competition (Proposition 2).

In order to clarify how the degree of competitive toughness may impact the aggregate production function, we observe that total output $Q^* \equiv n^* q^*$ in sector $\mathcal{I}$ is given by

$$Q^* = \frac{L}{c}\left(1 - f\,\frac{n^*}{L}\right) = \frac{L}{c}\frac{\sigma(n^*) - 1}{\sigma(n^*)}. \tag{31}$$

Equation (31) follows from (23), (27), and (28). Using (31), the aggregate production function (29) may be restated as follows:

$$Y^*(L) = \frac{\varphi\left[n^*(L)\right]}{n^*(L)} Q\left[L,\, n^*(L)\right]. \tag{32}$$

The first term in (32) captures the specialization/complexity effect in sector $\mathcal{F}$, while the second term keeps track of the competition effect. In other words, in our framework *competition among input-producing firms affects total output of the final good through the total amount of the intermediate input.* More precisely, equations (27) and (31) imply that $Q\left[L,\, n^*(L)\right]$ increases more than proportionately with L if and only if competition is markup-decreasing. This, in turn, leads to competition generating a tendency toward external increasing (decreasing) returns to scale in sector $\mathcal{F}$. Compared to the standard CES model (where Q is readily verified to be exactly proportional to L, so that a competition effect cannot be taken into account), in the general case that we are analyzing *there are two sources of EIR: the specialization/complexity effect and the competition effect.*

Equations (29) – (30) help us understand more thoroughly where the limitations of the CES production function come from: the term $[\sigma(n^*)-1]/\sigma(n^*)$, which appears in (29) – (30), is constant under CES. Hence, the specialization effect would be the only source of EIR in the final good sector.

4.2 How EIR emerge

We are now equipped to characterize the comparative statics of the free-entry equilibrium with respect to the population size L. Our purpose here is to analyze how aggregate output varies with L and when EIR in the $\mathcal{F}$-sector emerge. The central result of the paper is as follows.

Proposition 3. *EIR take place if and only if (24) holds, or, equivalently, if and only if competition is wage-increasing.*

Proof. Under an increase in L, the left-hand side of equation (27) remains unchanged, while the right-hand side is shifted downwards. As a consequence, the equilibrium mass n^* of firms increases with L whenever the equilibrium is stable (see Section 4.1 for a discussion of stability). Combining this with (29), we find that the average product of labor $Y^*(L)/L$ increases with L if and only if $[\sigma(n)-1]\varphi(n)/n$ is an increasing function of n, which is equivalent to the necessary and sufficient condition (24) of wage-increasing competition. □

As discussed in Section 3, what renders competition wage-increasing or wage-decreasing is the *interplay* between the competition effect and the specialization/complexity effect. Thus, Proposition 3 stresses the importance of the interaction between these two effects in generating EIR. Indeed, by comparing (24) with the necessary and sufficient condition (7) for specialization economies to arise, we find that that the former contains an additional term, $n\sigma'(n)/\sigma(n)$, which captures the competition effect and is missing in (7). For the conditions (24) and (7) to coincide, this additional term must be zero, *which is only true in the CES case.* This explains why the previous studies have generally

explained the emergence of EIR by appealing solely to the presence of specialization economies. Meanwhile, the role of market interactions among firms in this process has been largely (and perhaps undeservedly) neglected.

As for the relationship between n^* and L, we have the following result.

Proposition 4.

(i) *The mass n^* of firms increases less than proportionately to L if and only if competition is markup-decreasing.*

(ii) *Compared to the CES case, markup-decreasing competition dampens the specialization effect, but simultaneously triggers a positive competition effect. Things get reversed under markup-increasing competition.*

Proof. Part (i) follows immediately from equation (27), while part (ii) follows from combining part (i) with (29). □

Table 2 summarizes our results about the roles that market-size and the interaction between the specialization and competition effects play in determining the equilibrium market-outcome under markup-decreasing and markup-increasing competition, respectively:

	$\sigma'(n) > 0$	$\sigma'(n) < 0$
n^*	increases less than proportionally in response to an increase in L	increases more than proportionally in response to an increase in L
Y^*, w^*	specialization effect weakened, positive competition effect	specialization effect reinforced, negative competition effect

Table 2: The impact of market-size and the interplay between the competition and the specialization effects in determining the equilibrium market-outcome: markup-decreasing vs markup-increasing competition

4.3 Examples

To illustrate the interactions between the specialization/complexity effect and the toughness of competition in generating EIR, consider the following examples.

CES production function. In this case, equation (27) is linear, i.e. the number of firms is proportional to total labor supply L. Hence, the competition effect is washed out, and the specialization effect is the only source of EIR. The aggregate production function is given by

$$Y^*(L) = AL^{1/(1-\rho)}, \quad A \equiv \frac{\rho}{c}\left(\frac{1-\rho}{f}\right)^{\rho/(1-\rho)}.$$

Translog cost function. Combining (6) with (27) yields $\beta n^2 + n = L/f$, which implies $n^* = \left(\sqrt{1+4L/f}-1\right)/(2\beta)$. In this case, the number of firms grows proportionally to $\sqrt{L}$. This is because, unlike the CES case, competition is tougher in a larger market. Furthermore, as stated by part (iii) of Proposition 1, complexity diseconomies and specialization economies *exactly offset each other*.

Therefore, the competition effect becomes the main force shaping the resulting aggregate production function which is given by

$$Y^*(L) = \frac{f}{4\beta c}\left(\sqrt{1+4L/f}-1\right)^2. \tag{33}$$

Equation (33) suggests that the average product of labor $Y^*(L)/L$ increases in L for all $L \geq 0$. In other words, EIR take place. However, the source of these increasing returns is radically different from that in the CES case. Namely, *agglomeration economies stem here solely from market interactions across firms,* while in the classical CES-based models they are generated entirely by technological externalities embodied in the specialization/complexity tradeoff.

Translog production function. In this case, the competition effect is even stronger. Indeed, as implied by (5), (27) takes the form: $1-\alpha n = fn/L$. Hence, $n^* = L/(\alpha L + f)$, which implies that the equilibrium mass of firms is bounded from above by $1/\alpha$. In other words, even when population L grows unboundedly, the number of firms the market invites to operate remains limited due to very tough competition. The aggregate production function is then given by

$$Y^*(L) = \frac{\alpha}{c}L. \tag{34}$$

Thus, in the case of the translog production function, the resulting technology exhibits *constant returns to scale.* This result is in line with Proposition 3: EIR arise only when competition is wage-decreasing, while under the translog production function entry has no impact on wages (see Table 1 in Section 3.2).

A micro-foundation for an S-shaped production function. Finally, we provide a simple micro-foundation for an S-shaped aggregate production function, which has been widely used in growth theory and development economics, especially in the analysis of poverty traps.[16] Consider a Kimball-type technology associated with the aggregator function $\phi(\xi) \equiv a\xi^\rho - b$, where a and b are positive constants, while $0 < \rho < 1$. Here, a can serve as a measure of "overall" TFP, while b shows the strength of the complexity externality. Solving in closed form for the production function, we obtain

$$F(\mathbf{q}) = A(n)\left(\int_0^n q_i^\rho \mathrm{d}i\right)^{1/\rho}, \quad A(n) \equiv \left(\frac{a}{1+bn}\right)^{1/\rho}. \tag{35}$$

The TFP function underlying (35) is given by

$$\frac{\varphi(n)}{n} = \frac{1}{n}\left(\frac{an}{1+bn}\right)^{1/\rho}, \tag{36}$$

[16] See Skiba (1978), and, more recently, Azariadis and Stachurski (2005), as well as Banerjee and Duflo (2005), for examples on the possible consequences of using S-shaped production functions within these two branches of the economic literature. The idea dates back at least to Shapley and Shubik (1967).

while the elasticity of substitution across inputs is constant and is given by $\sigma = 1/(1-\rho)$, just like in the standard CES case. Using (36), it is readily verified that $\varphi(n)/n$ increases in the input diversity n for all $n < 1/(\rho b)$ and decreases otherwise. Hence, the TFP function is *bell-shaped*, meaning that specialization economies occur when the intermediate input is *not too much differentiated*, otherwise complexity diseconomies prevail.

The resulting aggregate production function $Y^*(L)$ reads as

$$Y^*(L) = \frac{f}{1-\rho}\left(\frac{L}{L + af/(1-\rho)}\right)^{1/\rho}. \tag{37}$$

According to (37), increasing returns to scale arise when L is sufficiently small; otherwise, decreasing returns to scale occur.

5 Possible applications: a discussion

How our approach can be applied to specific issues in urban and regional economics remains an important question to ask. We believe it potentially has a variety of applications, including productivity and specialization of cities, the determinants of city structure and land prices, cross-sectoral linkages, and comparisons between equilibrium and optimal city sizes. We focus here on three possible applications, which we find most straightforward and interesting.

First, our VES framework may shed more light on the *urban hierarchy* phenomenon, i.e., the coexistence of cities of different sizes (Christaller, 1933; Lösch, 1940; Henderson, 1974; Tabuchi and Thisse, 2006; Hsu, 2012). The fact that city sizes form hierarchical patterns has strong empirical support (Gabaix and Ioannides, 2004). To show how our approach can be applied to explaining this phenomenon, consider the standard agglomeration model with sharing externalities (Duranton and Puga, 2004). In that model, no urban hierarchy can emerge, because all cities have the same size in a stable equilibrium. This strong prediction is a byproduct of the CES production function. Using instead a general production function of type (1), one may expect the consumption curve to have multiple peaks and troughs.[17] Consequently, multiple stable equilibria may emerge. This, in turn, implies a non-degenerate city size distribution, hence urban hierarchy. We find this way of modeling urban hierarchy appealing for the following reason. Recent models of urban systems (Behrens et al., 2015; Behrens and Robert-Nicoud, 2015; Davis and Dingel, 2019) assume that individuals are heterogeneous in various dimensions. Similarly, using a bare-bones model of the origin of cities, de Palma et al. (2019) show that ex ante heterogeneity in individual preferences is critical for urban hierarchy to emerge. In contrast, our approach may give rise to a non-degenerate city size distribution without assuming any ex ante heterogeneity across individuals.

[17] In contrast, under CES the consumption curve is bell-shaped, see Duranton and Puga (2004) p. 2075, Fig. 1.

Second, quantitative spatial models (Diamond, 2016; Gaubert, 2018) seek to reproduce real-world urban systems and to quantify the consequences of various shocks and/or counterfactuals. However, those models "remain agnostic on the source of agglomeration externalities and their specific functional form" (Gaubert, 2018). Our approach could help to replace this "agnosticism" with more solid microeconomic foundations.

Finally, international trade studies display growing interest to wage inequality (Amiti and Davis, 2012; Helpman et al., 2010). Our findings suggest that part of wage dispersion is due to cross-country differences in the way specialization and competition interact with each other. A closely related issue is how city-level wages vary with city size. Although it is widely acknowledged by urban and regional economists that larger cities pay, on average, higher wages, the exact form of that relationship is ambiguous. Typically a log-linear relationship implied by the CES model is estimated, with city-specific dummies used to improve the fit (Duranton, 2014). Our paper provides theoretical underpinnings for using flexible empirical strategies based on non-linear specifications and/or non-parametric estimation methods.[18]

6 Concluding remarks

We have developed an Ethier-type two-sector model with variable elasticity of technological substitution across intermediate inputs. The model suggests a clear decomposition of external increasing returns (EIR) into two sources: the *specialization/complexity* effect, and the *competition* effect. The former is generated within the final good sector and shows how employing more varieties of intermediate inputs fosters/deters the production of the final good, while the latter stems from the market interactions among firms within the intermediate input sector. The market outcome is determined by the joint behavior of the TFP and the elasticity of substitution, which are both functions of the input diversity. In other words, the interplay between the competition effect and the specialization/complexity effect plays a key role in shaping the equilibrium properties of our model economy.

We have fully characterized the market outcomes driven by the interplay between the two above effects. This characterization has been useful in clarifying the origins of EIR. In particular, we have shown that, due to the interference of a non-trivial competition effect, the presence of specialization economies is neither necessary nor sufficient for EIR to emerge. This result highlights the limitations of the CES monopolistic competition model in studying production externalities: this approach, indeed, overlooks the relevance of the competition effect, as the level of market power does not vary with the number of firms. Therefore, our analysis points to the need for more work on the role of market competition

[18] Needless to say, we acknowledge that factors other than specialization economies and market competition also play a significant role in determining the city size-wage gap. Moreover, this gap may be different across workers being heterogeneous in experience and/or ability (see, e.g., Baum-Snow and Pavan, 2012). These issues are out of the scope of our paper.

in shaping agglomeration economies and other economic phenomena driven by scale effects. In addition to that, we argue that our theoretical findings are in line with recent empirical evidence on the behavior of prices, markups and wages with respect to the size of the economy. Finally, we sketch a possible application of our modeling approach to explaining urban hierarchy, providing microeconomic foundations to quantitative spatial models, and justifying the use of flexible empirical strategies in studying the relationship between city sizes and wages.

Two lines of further research seem to be of considerable interest. First, our VES-approach could lead to a richer R&D-based endogenous growth theory. However, this is a truly ambitious task, as in traditional horizontal R&D-based growth models (featuring constant returns to scale to rival inputs in the aggregate production function, F), endogenous markups would ultimately depend on the number n of available differentiated intermediate inputs. With n infinitely growing over time due to firms' research efforts, and without any other serious modeling alteration, the existence of a long-run balanced growth path equilibrium, commonly characterized by constant (though endogenous) markups, would definitely be compromised, unless strong assumptions are made about the R&D process. Moreover, as highlighted by Boucekkine et al. (2017), Bucci and Matveenko (2017), and Matsuyama et al. (2018), these assumptions have crucially to do also with the specific choice of VES preferences/technologies being made. Hence, it seems to us that an interesting research project to be definitely pursued in the future could, as a matter of fact, well consist in finding new modeling-paths for adapting our approach (based on variable and endogenous markups) also to the framework provided by dynamic, general-equilibrium, R&D-based growth theory. This is clearly not an easy task (and, indeed, it is outside the scope of the present article), but it is probably worth putting it into the future research-agenda of any interested growth theorist, as the reward (in terms of new theoretical/modeling scenarios) can be really very high.

Second, our VES Ethier-type modeling approach could help better understand global value chains. Despite a number of recent prominent contributions in this area (including Costinot et al., 2013; Alfaro et al., 2019), we believe that decomposing the impact of production structure on the characteristics of global value chains would be a useful contribution. To accomplish a project like this, we would need more than two sectors and a richer pattern of vertical relationships between sectors, as well as a non-trivial distribution of various production activities in space. Also, it would be necessary to relax the assumption of perfect competition in the final-good sector, in order to take into account strategic interactions driving the organizational decisions of firms. Doing so would result in a model generating a much more complex bundle of general-equilibrium effects than the "specialization/complexity vs competition" dichotomy, which is our current focus. Therefore, the effective investigation of these possible new research questions is ultimately left to future work.[19]

[19]We are particularly grateful to the anonymous referees for bringing these issues to our attention.

References

[1] Abdel-Rahman, H., and M. Fujita (1990). Product variety, Marshallian externalities, and city sizes. *Journal of Regional Science* 30: 165-183.

[2] Alfaro, L., Chor, D., Antras, P., and P. Conconi (2019). Internalizing global value chains: A firm-level analysis. *Journal of Political Economy* 127: 508-559.

[3] Aghion P, Bloom N, Blundell R, Griffith R, Howitt P. (2005) Competition and innovation: an inverted-U relationship. *Quarterly Journal of Economics* 120:701–728

[4] Aghion P, Griffith R (2005) Competition and growth: reconciling theory and evidence. MIT Press, Cambridge

[5] Amiti, M., and D.R. Davis (2012). Trade, firms, and wages: Theory and evidence. *Review of Economic Studies* 79: 1-36.

[6] Amiti, M., Itskhoki, O., and J. Konings (2019). International shocks, variable markups, and domestic prices. *Review of Economic Studies* 86: 2356-2402.

[7] Azariadis, C., and J. Stachurski (2005). "Poverty Traps". In P. Aghion and S.N. Durlauf (Eds.), Handbook of Economic Growth. Amsterdam: Elsevier-North Holland, Chap. 5 (Volume 1A): 295-384.

[8] Baum-Snow, N., and R. Pavan (2012). Understanding the city size wage gap. *Review of Economic Studies* 79: 88-127.

[9] Banerjee, A.V., and E. Duflo (2005). *Growth Theory through the Lens of Development Economics.* In: P. Aghion and S.N. Durlauf (Eds.), Handbook of Economic Growth, Vol. 1A. Elsevier-North, Holland, pp. 473-552.

[10] Behrens, K. and Y. Murata (2007) General equilibrium models of monopolistic competition: A new approach. *Journal of Economic Theory* 136: 776 – 87.

[11] Behrens, K., Duranton, G., and F. Robert-Nicoud (2015). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy* 122: 507-553.

[12] Behrens, K., and F. L. Robert-Nicoud (2015). *Agglomeration Theory with Heterogeneous Agents.* In: Duranton, G., Henderson, J.V., and W. C. Strange. Handbook in Regional and Urban Economics, Vol 5. Elsevier, North-Holland, pp. 171–245.

[13] Bellone, F., Musso, P., Nesta, L., and F. Warzynski (2016). International trade and firm-level markups when location and quality matter. *Journal of Economic Geography* 16: 67-91.

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

[15] Bertoletti, P., and F. Etro (2016). Preferences, entry, and market structure. *RAND Journal of Economics* 47: 792-821.

[16] Bertoletti, P., and F. Etro (2017). Monopolistic competition when income matters. *Economic Journal*, in press.

[17] Bilbiie F., Ghironi, F., and M. Melitz (2012). Endogenous entry, product variety, and business cycles. *Journal of Political Economy* 120: 304-345.

[18] Boucekkine, R., Latzer, H., and M. Parenti (2017). Variable markups in the long-run: A generalization of preferences in growth models. *Journal of Mathematical Economics* 68: 80-86.

[19] Bucci, A. (2013). Returns to specialization, competition, population, and growth. *Journal of Economic Dynamics and Control* 37: 2023-2040.

[20] Bucci, A., and V. Matveenko (2017). Horizontal differentiation and economic growth under non-CES aggregate production function. *Journal of Economics* 120: 1-29.

[21] Chen, Y., and M. H. Riordan (2008). Price-increasing competition. *RAND Journal of Economics* 39: 1042-1058.

[22] Christaller, W. (1933). Die Zentralen Orte in Süddeutschland. Jena: Gustav Fischer Verlag. English translation: *The Central Places of Southern Germany*. Englewood Cliffs, NJ: Prentice-Hall (1966).

[23] Costinot, A, Vogel, J., and S. Wang (2013). An Elementary Theory of Global Supply Chains. *Review of Economic Studies* 80: 109-144.

[24] Combes, P. P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica* 80: 2543-2594.

[25] Combes, P.P., Mayer, T., and J.F. Thisse (2008). *Economic geography: The integration of regions and nations.* Princeton University Press.

[26] Dalgaard, C.-J., and C.T. Kreiner (2001). Is declining productivity inevitable? *Journal of Economic Growth* 6: 187-203.

[27] Davis, D.R., and J. Dingel (2019). A spatial knowledge economy. *American Economic Review* 109: 153–170.

[28] De Palma, A., Y. Y. Papageorgiou, J.-F. Thisse, and P. Ushchev (2019). About the Origin of Cities. *Journal of Urban Economics* 111: 1-13.

[29] Dhingra, S., and J. Morrow (2019). Monopolistic competition and optimum product diversity under firm heterogeneity. *Journal of Political Economy* 127: 196-232.

[30] Diamond, R. (2016). The determinants and welfare implications of US workers diverging location choices by skill: 1980-2000. *American Economic Review* 106: 479-524.

[31] Dixit, A.K., and J.E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *American Economic Review* 67: 297-308.

[32] Duranton, G., and D. Puga (2004). *Micro-foundations of urban agglomeration economies.* In: Henderson, J.V., and J.-F. Thisse (Eds.), Handbook of Regional and Urban Economics, Vol. 4. Elsevier-North Holland, pp. 2063-2117.

[33] Duranton, G. (2014). Growing through cities in developing countries. *The World Bank Research Observer* (first published online: April 15, 2014, doi:10.1093/wbro/lku006).

[34] Ethier, W. J. (1982). National and international returns to scale in the modern theory of international trade. *American Economic Review* 72: 389-405.

[35] Feenstra, R.C. (2003). A homothetic utility function for monopolistic competition models, without constant price elasticity. *Economics Letters* 78: 79-86.

[36] Ferrarini, B., and P. Scaramozzino (2016). Production Complexity, Adaptability and Economic Growth. *Structural Change and Economic Dynamics* 37: 52-61.

[37] Fujita, M., and J.-F. Thisse (2013). *Economics of Agglomeration: Cities, Industrial Location, and Globalization. Second Edition.* Cambridge, UK: Cambridge University Press.

[38] Gabaix, X., and Y. M. Ioannides (2004). *The evolution of city size distributions.* In: Henderson, J.V., and J.-F. Thisse (Eds.), Handbook of Regional and Urban Economics, Vol. 4. Elsevier, Amsterdam, pp. 2341-2378.

[39] Gali, J. (1995). Product diversity, endogenous markups, and development traps. *Journal of Monetary Economics* 36: 39-63.

[40] Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review* 108: 3117-3153.

[41] Grossman, G.M., and E. Helpman (1990). Comparative advantage and long run growth. *American Economic Review* 80: 796-815.

[42] Grossman, G.M., and E. Helpman (1991). Innovation and Growth in the Global Economy. Cambridge, MA: MIT Press.

[43] Handbury, J., and D. E. Weinstein (2015). Goods prices and availability in cities. *The Review of Economic Studies*, 82: 258-296.

[44] Helpman, E., O. Itskhoki, and S. Redding (2010). Inequality and unemployment in a global economy. *Econometrica* 78: 1239-1283.

[45] Henderson, J.V. (1974). The sizes and types of cities. *American Economic Review* 64: 640-656.

[46] Hidalgo, C.A., B. Klinger, A.L. Barabasi, and R. Hausmann (2007). The Product Space Conditions the Development of Nations. *Science* 317(5837): 482-487.

[47] Howitt, P. (1999). Steady endogenous growth with population and R&D inputs growing. *Journal of Political Economy* 107: 715–730.

[48] Hsu, W.-T. (2012). Central place theory and city size distribution. *Economic Journal* 122: 903-932.

[49] Kim, H. Y. (1992). The translog production function and variable returns to scale. *Review of Economics and Statistics* 74: 546-552.

[50] Kimball, M. S. (1995). The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit and Banking* 27: 1241-1277.

[51] Kremer, M. (1993). The O-Ring Theory of Economic Development. *Quarterly Journal of Economics* 108: 551-575.

[52] Lösch, A. 1940. *Die Räumliche Ordnung der Wirtschaft.* Jena: Gustav Fischer. English translation: *The Economics of Location.* New Haven, CN: Yale University Press (1954).

[53] Matsuyama, K., H. Latzer, and M. Parenti (2018). Reconsidering the Market Size Effect in Innovation and Growth. CEPR DP14250.

[54] Matsuyama, K., and P. Ushchev (2017) Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems . CEPR DP12210.

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

[56] Nadiri, M. I. (1982) *Producers theory.* In Arrow, K.J. and M.D. Intriligator (eds.) Handbook of Mathematical Economics. Volume II. Amsterdam: North-Holland, pp. 431 – 90.

[57] Parenti, M., P. Ushchev, and J.-F. Thisse (2017). Toward a theory of monopolistic competition. *Journal of Economic Theory* 167: 86-115.

[58] Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy* 98: S71–S102.

[59] Rosenthal, S. S., and W. C. Strange (2004). *Evidence on the nature and sources of agglomeration economies.* In: Henderson, J.V., and J.-F. Thisse (Eds.), Handbook of Regional and Urban Economics, Vol. 4. Elsevier, Amsterdam, pp. 2119-2171.

[60] Sandmo, A. (2011). *Economics evolving: A history of economic thought.* Princeton University Press.

[61] Shapley, L. S., and M. Shubik (1967). Ownership and the production function. *Quarterly Journal of Economics* 81: 88-111.

[62] Skiba, A. K. (1978). Optimal growth with a convex-concave production function. *Econometrica* 46: 527-539.

[63] Smets, F. and R. Wouters (2007) Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review* 97: 586 – 606.

[64] Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations.* London: W. Strahan and T. Cadell.

[65] Tabuchi, T., and J. F. Thisse (2006). Regional specialization, urban hierarchy, and commuting costs. *International Economic Review* 47: 1295-1317.

[66] Thisse, J. F., and P. Ushchev (2018). *Monopolistic competition without apology.* In: Corchón, L. C., and M. Marini (Eds.), Handbook of Game Theory and Industrial Organization, Vol. 1. Edward Elgar Publishing, pp. 93-136.

[67] Uzawa, H. (1964). Duality principles in the theory of cost and production. *International Economic Review* 5: 216-220.

[68] Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012). Monopolistic competition: beyond the constant elasticity of substitution. *Econometrica* 80: 2765–2784.

Appendix

A1. Marginal products under a continuum of inputs

We restrict our attention to such input vectors $\mathbf{q}$ that have a finite second moment, i.e. $\int_0^n q_i^2 \mathrm{d}i < \infty$. In other words, $\mathbf{q} \in L_2([0,n])$. Intuitively, this assumption allows *mean* and *variance* of the input vector to be well-defined.

Following Parenti et al. (2017), we also assume Fréchet-differentiability of the production function, i.e. we postulate that there exists a functional $\Phi : \mathbb{R}_+ \times L_2 \to \mathbb{R}_+$, such that

$$F(\mathbf{q}+\mathbf{h}) = F(\mathbf{q}) + \int_0^n \Phi(q_i, \mathbf{q}) h_i di + \circ(||\mathbf{h}||_2) \text{ for all } \mathbf{q}, \mathbf{h} \in L_2. \tag{38}$$

In (38), $||\cdot||_2$ stands for the L_2-norm, i.e. $||\mathbf{h}||_2 \equiv \sqrt{\int_0^n h_i^2 \mathrm{d}i}$, whereas $\Phi(q_i, \mathbf{q})$ is the *marginal product* of intermediate input i. Concavity of F implies that Φ is decreasing in q_i.

Lemma. *Let $F : L_2 \to \mathbb{R}_+$ be a Fréchet-differentiable functional, which is positive homogeneous of degree* 1. *Then* (i) $\Phi(q_i, \mathbf{q})$ *is positive homogeneous of degree zero in* $(q_i, \mathbf{q})$, *and* (ii) *the Euler's identity*

$$F(\mathbf{q}) = \int_0^n q_i \Phi(q_i, \mathbf{q}) \mathrm{d}i, \tag{39}$$

holds.

Proof. To prove part (i), we recast (38) as follows:

$$F(t\mathbf{q} + t\mathbf{h}) = F(t\mathbf{q}) + \int_0^n \Phi(tq_i, t\mathbf{q}) t h_i di + \circ(t||\mathbf{h}||_2) \text{ for all } \mathbf{q}, \mathbf{h} \in L_2,\ t \in \mathbb{R}_+. \tag{40}$$

Dividing both sides of (40) by t and using homogeneity of F, we obtain

$$F(\mathbf{q} + \mathbf{h}) = F(\mathbf{q}) + \int_0^n \Phi(tq_i, t\mathbf{q}) h_i di + \circ(||\mathbf{h}||_2). \tag{41}$$

Combining (38) with (41), we find that $\phi(tq_i, t\mathbf{q})$ is a Fréchet derivative of F computed at $\mathbf{q}$ for any $t > 0$. By uniqueness of Fréchet derivative, $\phi(tq_i, t\mathbf{q})$ must be independent of t, which proves part (i) of the Lemma.

To prove part (ii), note that (38) implies the following identity:

$$\frac{F((t+\tau)\mathbf{q}) - F(t\mathbf{q})}{\tau} = \int_0^n \Phi(tq_i, t\mathbf{q}) q_i di + \frac{\circ(\tau)}{\tau} \text{ for all } \tau \in \mathbb{R}. \tag{42}$$

Using homogeneity of F and Φ, we obtain (39) as the limiting case of (42) under $\tau \to 0$. □

A2. Specialization economies and complexity diseconomies under augmented CES

In this Appendix, we illustrate the general definitions of specialization economies and complexity diseconomies (see Subsection 2.1) by considering the special case of the augmented CES production technology given by (3). In equation (3), when sufficiently negative, ν is a measure of the magnitude of the *complexity effect*: a larger number of intermediate inputs being simultaneously combined within the same production process can lead to a reduction in aggregate output (we come back to this issue immediately below). To be more precise, *complexity diseconomies* are said to occur if and only if $\nu < 1 - 1/\rho$, otherwise *specialization economies* take place. The logic behind these definitions is as follows: evaluating the total output Y given by (3) at a symmetric vector of inputs,[20] we obtain

[20] i.e. such that $q_i = q$ for all $i \in [0, n]$, where $q > 0$ is given.

$Y = n^{\nu+1/\rho}q$. The above inequalities keep track of whether Y increases more or less than proportionately with n. The baseline case described by (2) corresponds to $\nu = 0$, hence the baseline CES technology always exhibits specialization economies.

It is worth noting that a negative production externality need not result in complexity diseconomies. Consider the augmented CES when production externality is negative but not too strong: $0 > \nu > -(1-\rho)/\rho$. In this case, we have

$$\frac{\varphi'(n)n}{\varphi(n)} = 1/\rho + \nu > 1,$$

whence specialization economies occur. We conclude that *only a sufficiently strongly negative production externality can generate complexity diseconomies.*

A3. Proof of Proposition 1.

(i) If a production function satisfies (4), we have

$$\frac{\varphi(n)}{n} = \frac{1/n}{\phi^{-1}(1/n)}. \tag{43}$$

Because $\phi(\cdot)$ is increasing and concave, it must be that $\phi^{-1}(\cdot)$ is increasing and convex. If $\phi(0) = 0$, then the elasticity of $\phi^{-1}(\cdot)$ always exceeds 1. As a consequence, $\varphi(n)/n$ decreases in $1/n$ and increases with n.

When $\phi(0) \neq 0$, the above argument is no longer valid. Indeed, as implied by (36), production function given by (35) provides a counterexample. This completes the proof of (i).

(ii)-(iii). As shown in Section 2.1, under (5) we have $\varphi(n) = 1$, while (6) yields $\varphi(n) = n$ for all $n > 0$. Combining this with the definition (7) of specialization economies completes the proof. □

A4. Deriving the inverse demands (10) for inputs.

The FOC for cost minimization in the $\mathcal{F}$-sector is given by

$$p_i = \lambda\Phi(q_i, \mathbf{q}), \tag{44}$$

where $\Phi(q_i, \mathbf{q})$ is the marginal product of input i (see Appendix A1), while λ is the Lagrange multiplier of the firm's program (8). Multiplying both sides of (44) by q_i, integrating both sides w.r.t. i across $[0, N]$, and using Euler's identity (39) (see Appendix A1), we get:

$$\int_0^N p_i q_i \mathrm{d}i = \lambda \int_0^N q_i \Phi(q_i, \mathbf{q}) = \lambda F(\mathbf{q}). \tag{45}$$

As implied by the $\mathcal{F}$-firm's profit maximization program (8), we have:

$$\int_0^N p_i q_i \mathrm{d}i = P(\mathbf{p}), \qquad F(\mathbf{q}) = 1. \tag{46}$$

Using (45) – (46), we find that $\lambda = P(\mathbf{p})$. Plugging $\lambda = P(\mathbf{p})$ back into (44), we obtain the inverse demand schedule (10) for input i. □

A5. Second-order conditions and no asymmetric equilibria

Observe that the left-hand side of (12) is positive homogeneous of degree zero. This implies that the solution of (12) cannot be unique. Indeed, multiplying a solution of (12) by a constant yields another solution. The "proper" equilibrium is pinned down by the labor balance condition (13).

To guarantee that equation (12) is compatible with profit-maximizing behavior by firms, the second-order condition must hold, which amounts to assuming that the real operating profit $[\Phi(q_i, \mathbf{q}) - cw/P]\, q_i$ of firm i is strictly quasi-concave in q_i for all $\mathbf{q}$.

To rule out a continuum of asymmetric equilibria in the quantity-setting game of firms, we introduce a stronger assumption: *the left-hand side of* (12) *is decreasing in* q_i *for any* $\mathbf{q}$. Imposing this condition is equivalent to assuming that the operating profit of each firm is strictly concave in its output. This assumption holds for the CES and, more generally, for any production function of the type (4) such that

$$-\frac{\phi'''(\xi)}{\phi''(\xi)}\xi < 2 \quad \text{for all} \quad \xi > 0.$$

This rules out the possibility of asymmetric equilibria because (12) has a unique solution $q_i^*(\mathbf{q})$, which is the same for all firms $i \in [0, n]$.