Survival of the Fittest in Cities: Urbanisation and Inequality

Behrens, Kristian; Robert-Nicoud, Frédéric

How to cite

BEHRENS, Kristian, ROBERT-NICOUD, Frédéric. Survival of the Fittest in Cities: Urbanisation and Inequality. In: Economic journal, 2014, vol. 124, n° 581, p. 1371–1400. doi: 10.1111/ecoj.12099

This publication URL: https://archive-ouverte.unige.ch/unige:40780
Publication DOI: 10.1111/ecoj.12099

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES ÉCONOMIQUES ET SOCIALES
Département des sciences économiques

WPS 13-07-4 Working Paper Series

Survival of the Fittest in Cities: Urbanisation and Inequality

Kristian Behrens
Frédéric Robert-Nicoud

July 2013

Survival of the Fittest in Cities: Urbanisation and Inequality

Kristian Behrens* Frédéric Robert-Nicoud†

July 16, 2013

Abstract

We develop a framework that integrates natural advantage, agglomeration economies, and firm selection to explain why large cities are both more productive and more unequal than small towns. Our model highlights interesting complementarities among those factors and it matches a number of key stylised facts about cities. A larger city size increases productivity via a selection process, and higher urban productivity provides incentives for rural-urban migration. Tougher selection increases both the returns to skills and earnings inequality in cities. We numerically illustrate a multi-city version of the model and explore the formation of new cities, the growth of existing cities, and changes in income inequality.

Keywords: city size; urbanisation; agglomeration; firm selection; income inequality

JEL Classification: D31; F12; R11; R12

*Canada Research Chair, Département des Sciences Economiques, Université du Québec à Montréal (UQÀM), Montréal, Canada; CIRPÉE, Canada; and CEPR, United Kingdom. E-mail: behrens.kristian@uqam.ca

†*Corresponding author:* Département des Sciences Economiques, Université de Genève; SERC, United Kingdom; and CEPR, United Kingdom. E-mail: Frederic.Robert-Nicoud@unige.ch.

1 Introduction

Large cities are more productive and more unequal than small towns. In a 2007 sample of 356 US cities, for instance, doubling city size is associated with an increase of value added per employee of 8.5% and with an increase of the Gini coefficient of city earnings of 1.2%. This paper introduces a multi-city, heterogeneous firms framework that integrates agglomeration economies, natural advantage, and selection to explain these facts. Agglomeration economies or indivisibilities – either internal or external to firms – are necessary for cities to exist. They lead to productivity that is increasing in city size. Natural advantage – which is especially important for early urban development – helps pin down city locations, raises productivity and earnings, and thereby attracts people and firms. Finally, larger cities also make for larger market places. Since selection is tougher in large markets, only the most productive firms may profitably operate there, and earnings inequality increases with city size.

The positive relationship between city size and productivity is better documented and better understood than the positive relationship between city size and urban inequality. Our aim is to study these facts jointly, and the novelty of our paper is to integrate agglomeration economies, natural advantage, and selection into a unified theoretical framework in which cities are determined endogenously. Three features of this exercise are noteworthy.

First, it highlights interesting complementarities among natural advantage, agglomeration, and selection. Natural advantage attracts firms and people to a city, and local increasing returns raise productivity: *natural advantage induces agglomeration*. The larger number of firms in larger cities implies tougher competition: *agglomeration induces selection*. In turn, the most productive firms get larger market shares and pay higher wages in big cities than in small towns and the opposite is true for the least productive surviving firms: *selection induces inequality*. The presence of more productive firms increases average productivity and lowers consumer prices, thereby attracting more people. This further strengthens agglomeration economies.

Second, our model matches a number of key stylised facts about cities. The effect of city population on productivity is causal (Rosenthal and Strange, 2004), even after controlling for sorting and selection (Combes, Duranton, Gobillon, Puga, and Roux, 2012). The returns to skills and income inequality increase with city population, even after controlling for the socioeconomic composition of cities (Wheeler, 2001; Glaeser, Resseger, and Tobio, 2008; Baum-Snow and Pavan, 2012). The distribution of firm productivity in any city is non-degenerate, with fewer low-productivity firms in larger cities (Combes *et al.*, 2012). Finally, goods are cheaper and come in more numerous varieties in large cities than in small towns (Glaeser, Kolko, and Saiz, 2001; Handbury and Weinstein, 2012).

Third, our model is amenable to comparative static exercises and to numerical simulations. We characterise all the properties of the symmetric equilibrium – as well as some of the asymmetric equilibria – analytically. We then use numerical simulations to explore the quantitative properties of the model for an equilibrium that matches some moments of our sample of US cities. We

find that everything that makes cities more productive and urban life more desirable makes the emergence of cities more likely and allows for larger cities in a way that is consistent with the historical analysis of Bairoch (1988) and others. We also stress the role of trade in increasing the relevant market size, which reinforces urban selection, sustains larger cities, and raises earnings inequality within cities.

1.1 Preview of analysis

We use the monopolistic competition framework of Ottaviano, Tabuchi, and Thisse (2002), as extended by Asplund and Nocke (2006) and Melitz and Ottaviano (2008) to heterogeneous entrepreneurs, borrowing from Lucas (1978). We then embed this production structure in a system of cities where urban costs increase in a standard 'Alonso-Mills-Muth' fashion with city population. All individuals are *ex ante* homogeneous and each individual initially chooses a location among several cities and a countryside. Cities form endogenously. In our model, individuals acquire the necessary skills to become entrepreneurs in cities only, since cities are places that favour learning and nurture innovation (Glaeser and Maré, 2001; Duranton and Puga, 2001). Firms are operated by entrepreneurs whose productivity is revealed after the irreversible location decision is made. Individuals select *ex post* into entrepreneurship or consume from their initial endowment. In our model, selection operates on the goods markets and it affects entrepreneurial earnings only. In other words, all workers are entrepreneurs. This is a convenient shortcut, for selection may also arise in labor markets. An alternative model, in which the variable component of production uses labor and in which fair wage considerations would result in rent sharing between management and production workers (as in Egger and Kreickemeier, 2009), would produce similar results: large cities would be more selective and more unequal.

Cities result from a tradeoff between agglomeration economies and urban costs as in Henderson (1974). Entrepreneurial profit increases with productivity and city population. Hence, more productive entrepreneurs benefit the most from being in larger cities. Since urban costs do not depend on entrepreneurial productivity but only on city size, less productive entrepreneurs enjoy lower profits and smaller market shares in larger cities. The complementarity between productivity and city population leads to a positive equilibrium relationship between city size and earnings inequality. Tougher selection in larger cities also increases firm productivity. This complementarity leads to a positive relationship between city population and average earnings.

The simple one-city version of our model is consistent with facts that have attracted attention in the literature. First, anything that makes cities more attractive – like lower commuting costs or a larger range of available products – makes their emergence more likely and also makes them bigger. Historically, more product differentiation makes urban production more valuable. Lower commuting costs brought about by the introduction of streetcars, cars, and elevators all allow for an increasing density and population of cities. Our model is well suited to the analysis of episodes

of urbanisation like the growth of manufacturing centres in 18th century England or the urban explosion in developing countries during the 20th century. In our framework, city formation and growth is driven by rural-urban migration that is especially relevant in those contexts.

The multi-city version of our model emphasises the role of transportation and trade in understanding where cities form and why they grow. Think about Yokohama, a village populated by fishermen before Japan reopened to international trade in the mid-19th century. As the major port on the Tokyo bay, it is now Japans' largest municipality and it hosts a population of almost 4 million. Buenos Aires, Hong Kong, Lagos, London or New York's predominant roles have also to do with them being – or having been – major ports. Many cities developed and are developing around river mouths or transportation network nodes (Bleakley and Lin, 2012). The location of Samarkand and Xi'an on the silk road and the designation of Kiakhta as the only trading point between Tsarist Russia and the Qing empire in the 18th and 19th centuries also underscore the role of transportation and trade for the fate of cities. More prosaically, Duranton and Turner (2012) estimate that the elasticity between a city's stock of interstate highway and its population is about 15% in their 1983–2003 panel of US cities.

Our model can also shed light on the growth of cities in the Third World, where the relatively high urbanisation rates coexists with severe urban poverty (the Harris-Todaro puzzle). Indeed, after entry, entrepreneurs who fail in the city have lower nominal and real incomes than workers in the countryside, while their consumer surplus exceeds that in the countryside. The reason is that they have access to urban consumption diversity even if they failed as entrepreneurs. This aspect is taken into account in the entry decision in our model.

Finally, cities that have unfavourable fundamentals are small and not very productive at equilibrium. In that case, urban migration is primarily motivated by urban wages (entrepreneurs' profits in the model) that are large relative to rural wages. Furthermore, the 'failure rate' is relatively low, i.e., the mass of unsuccessful entrepreneurs is small. These cities are 'producer cities'. The consumer surplus is rather small in this case. By contrast, cities with good underlying economic conditions are large, productive, and fiercely competitive: expected profits are no longer the primary driver of urban life – the failure rate is large – but the city's local and specific service and product mixes work like local amenities that attract consumers who display preference for diversity. These cities are 'consumer cities'.

1.2 Relation to extant literature

The main innovation of our paper is to integrate natural advantage, agglomeration, and selection in a model with endogenous cities.[1] Our model builds on and expands the large theoretical literature on agglomeration economies (see Duranton and Puga, 2004, for a review). The theoretical literature

[1] We ignore an additional reason: spatial selection, which is the focus of a complementary paper by Behrens, Duranton, and Robert-Nicoud (2012).

on selection in cities is much smaller. Baldwin and Okubo (2006) and Ottaviano (2012) develop models that are related to ours. In their models, there are only two locations that are dots on the map, there is no urban structure, and firms/entrepreneurs may relocate after learning their productivity. By contrast, the number of cities is endogenously determined and location choices are irreversible in our model.

The theoretical literature on sorting and inequality in cities is thinner still. Behrens, Duranton, and Robert-Nicoud (2012), Davis and Dingle (2012), and Eeckhout, Pinheiro, and Schmidheiny (2010) are three exceptions. The first analysis focuses on an equilibrium in which income inequality is invariant in city size; the second paper analytically establishes a positive equilibrium relationship between city size and inequality in a two-city variant of the model only; whereas the third analysis imposes some assumptions on the output elasticity of skills to generate fatter tails in the income distribution of big cities. Unlike ours, these papers focus mainly on spatial selection – also known as *sorting* – that is, situations in which more productive agents tend to disproportionately locate in large cities.

Extant empirical work on the existence and scope of agglomeration economies is abundant and consistent (see Rosenthal and Strange, 2004, for a review). Empirical evidence on selection effects is more mixed. Syverson (2004) finds that selection is increasing in market size in the US concrete industry, but Combes *et al.* (2012) find that productivity differences across French metropolitan areas are mostly explained by agglomeration economies. Our model allows for both mechanisms to play a role. Glaeser, Kahn, and Rappaport (2008) and Baum-Snow and Pavan (2012) document the positive relationship between urban population and earnings inequality for US cities. We provide a theoretical foundation for it. Finally, our model also stresses the role of intercity trade for urban size, urbanisation, and inequality, a rather neglected topic until now.

The remainder of the paper is organised as follows. Section 2 presents the model. Section 3 solves for its equilibrium considering a single city in isolation. Section 4 allows for multiple trading cities and studies the properties of the urban system. It also illustrates numerically some of the key features of our model. Section 5 concludes. We relegate most proofs and additional details to an extensive set of appendices.

2 The model

Consider a two-sector economy endowed with a large population, $\overline{L}$, of ex ante undifferentiated workers. There are several potential sites for cities, but in this section we study only a single city in isolation. This allows us to save on notation and to investigate the relationships between urbanisation and inequality in a parsimonious setting. We postpone the analysis of interactions among cities to Section 4.

Workers initially reside in the countryside, a location we label $l = 0$, and they can potentially move to the city, which we label $l = 1$. All workers are endowed with a large quantity of a homogeneous good and one unit of labor, which can be used either for producing that good or for acquiring skills. Skills are essential to the production of a differentiated good (more on this below). Acquiring them is costly and requires that workers move to the city, which provides an environment amenable to 'learning'. Learning in cities has an outcome that is uncertain: some workers will be more productive than others as a consequence of learning, and competition among skilled workers implies that only the most skilled can survive there.

There are two sectors in the economy. The first one produces a homogeneous good, whereas the second produces a large range of varieties of a differentiated good. Production of the homogeneous good is carried out by the workers who did not acquire skills and occurs under constant returns to scale. That good is traded in a competitive market that is perfectly integrated across locations. Hence, its price is the same everywhere, which makes it a natural choice for the numéraire. The differentiated good is only produced in the city by skilled workers using their 'entrepreneurial skills' as a fixed component and the numéraire as a variable component. The latter is obtained from the residents' endowments.[2]

2.1 Timing

The timing of the model is the following. First, workers decide whether to move to the city to acquire skills or to stay in the countryside and not acquire skills. We denote the mass of workers who chose to acquire skills in the city by H, which we refer to as the city size. To become skilled, they incur a sunk cost $f^E > 0$, paid in terms of the numéraire, which includes both the opportunity cost of foregoing the wage prevailing in the numéraire sector and the cost of moving to the city. The superscript 'E' is a mnemonic for 'education' or 'entrepreneur'. Living in a city gives rise to extra costs and benefits, which will be made precise below. Second, agents discover a variety and nature draws the marginal cost c at which they can produce this variety from some common and known distribution G. Third, upon observing their draws, agents chose whether to produce or not (and, in the multi-city extension of Section 4, which market(s) to serve).[3] Those who produce

[2]We assume that each worker's initial endowment of the homogeneous good is large enough so that: (i) this good is supplied in all cities in sufficient quantity for production of the differentiated good to occur; and (ii) all consumers have positive demand for the numéraire, which rules out income effects for the differentiated good in our quasi-linear specification below. As we explain later we assume, for simplicity, that migration from the countryside to the city is the only exchange between cities and the countryside. There is no trade between the two.

[3]We model selection from a product-market perspective. An alternative approach is to adopt a labor-market perspective by assuming that skilled workers supply differentiated labor but are not entrepreneurs. Workers move to cities to acquire education and demand for their skills is a priori uncertain. Having acquired education, workers above some threshold get the skilled jobs, whereas the others work in the numéraire sector. This approach leaves the product market and product variety aspects out of the picture, yet raises the question of how to model demand

maximise profits and all markets clear.

We assume that production occurs where agents live, i.e., agents who entered the city in the first stage produce and consume there. We also assume that agents are immobile after making their location choices. Hence, city dwellers cannot move back to the countryside if they get a bad draw. Our timing enables us to model the pervasiveness of mobility frictions in a parsimonious way: agents are freely mobile before observing their c, and fully immobile afterwards. The lack of ex post mobility after productivity is revealed implies that we do not allow for spatial sorting along skills. We do so mostly for analytical convenience as modelling selection and spatial sorting at the same time is difficult and beyond the scope of this paper (see Behrens *et al.*, 2012).[4] Furthermore, empirical evidence suggests that most entrepreneurs are 'local' and that entrepreneurship is a relatively immobile factor (Michelacci and Silva, 2007).

In the remainder of this section, we lay out explicit microeconomic building blocks – preferences, technology, urban structure – for the model and solve for all endogenous variables taking city size H as *given*. We endogenise city size and solve for the *spatial equilibrium* in the next section.

2.2 Preferences and urban structure

Following Asplund and Nocke (2006) and Melitz and Ottaviano (2008), all agents have identical quasi-linear preferences over the homogeneous good and a continuum $\mathcal{V}$ of varieties of horizontally differentiated urban goods and services. For simplicity, we assume that the latter are available exclusively in the city, while the homogeneous good is sold everywhere. Although this is a stark assumption, it fits with the observation that one of the key benefits of cities is indeed the access to a large range of mostly local goods and services they provide (Glaeser *et al.*, 2001; Lee, 2010).[5] Varieties of the differentiated good available in the city are indexed by ν. We denote by N the endogenously determined mass of varieties consumed in the city. The sub-utility over the differentiated varieties is assumed to be quadratic, so that utility for an urban resident is given by

$$U = \alpha \int_{\mathcal{V}} d(\nu) \mathrm{d}\nu - \frac{\gamma}{2} \int_{\mathcal{V}} d(\nu)^2 \mathrm{d}\nu - \frac{\eta}{2} \left[\int_{\mathcal{V}} d(\nu) \mathrm{d}\nu \right]^2 + d^0, \tag{1}$$

where $\alpha, \eta, \gamma > 0$ are preference parameters; and where d^0 and $d(\nu)$ stand for the consumption of the numéraire and of variety ν, respectively. Both depend, in general, on the skill level c. The

for skills. One possibility is to follow Ethier (1982) and to use a CES production function that aggregates skill types city-wide to produce a homogeneous good under increasing returns to scale.

[4] We could instead assume that, upon observing their draw c, agents may migrate back to the countryside at a cost f^M. Our main results would still hold as long as f^M is strictly positive. By contrast, assuming $f^M = 0$ would imply that a city's equilibrium income distribution is independent of its size (see Supplemental Appendix A.1 for details). This runs counter the empirical facts we know.

[5] We show in Supplemental Appendix A.2 that our qualitative results continue to hold true when the urban goods are available in the countryside at an extra cost. We show below that rural dwellers renounce to consume urban goods if this cost is larger than $\alpha^2/(2\eta)$. We henceforth impose this parameter restriction.

utility level of rural residents is simply equal to d^0 since they consume the numéraire good only.

To keep the analysis simple, we assume that urban costs, defined as the sum of commuting and housing costs, borne by a worker in a city of size H are given by θH, where $\theta > 0$ is a parameter. Linearly increasing per capita urban costs is a standard result in urban economics (see Appendix A.1 for microeconomic foundations). In sum, becoming an urban dweller involves two types of costs: the exogenous sunk cost f^E, and the endogenous urban costs θH.

Let $\Pi(c)$ denote the entrepreneurial profit of a city dweller with ability c, and let $\overline{d}^0$ be her initial endowment of the numéraire. The budget constraint of that agent is then given by

$$d^0 + \int_{\mathcal{V}} p(\nu) d(\nu) \mathrm{d}\nu + \theta H = \overline{d}^0 + \Pi(c) - (f^E - w). \tag{2}$$

The left-hand side of (2) consists in total spending on the numéraire, the differentiated good, and urban costs, respectively; the right-hand side consists in income from the endowment and entrepreneurial profits net of the sunk cost. Recall that the definition of f^E includes the opportunity cost of foregoing the wage w in the numéraire sector.

Maximising (1) subject to (2) yields the indirect utility of urban dwellers as $V(c) = \overline{d}^0 + w + \mathrm{CS} + \Pi(c) - \theta H - f^E$, where CS denotes the consumer surplus (see Appendix A.2 for computational details).

2.3 Production, profit maximisation, and selection

Production of the differentiated good requires the numéraire as an intermediate input. That input is available at unit cost everywhere. Let $p(c)$ and $q(c)$ denote the price set and the quantity sold by an entrepreneur with marginal cost c. Operating profits are equal to $\pi(c) \equiv [p(c) - c]\, q(c)$. At equilibrium, only entrepreneurs with c smaller than some cutoff c_1 are productive enough to sell in the city. Since agents are atomistic, they have a negligible impact on the equilibrium market aggregates. They therefore accurately take all other agents' decisions as given.

In order to keep the analysis tractable, we impose some assumptions on the distribution G of cost draws. We follow standard practice in the theoretical literature and assume that productivity draws $1/c$ in the city follow a Pareto distribution with a lower productivity bound $1/c_{\max}$ and shape parameter $k \geq 1$ (Helpman, Melitz, and Yeaple, 2004; Melitz and Ottaviano, 2008). This implies a distribution of cost draws given by $G(c) = (c/c_{\max})^k$ for $c \in [0, c_{\max}]$, with $\alpha > c_{\max}$.[6]

As shown in Appendix A.3, under the Pareto assumption the equilibrium prices that entrepreneurs set for their varieties can be expressed as

$$p(c) = \frac{c_1 + c}{2}, \quad \text{where} \quad c_1 \equiv \frac{2\alpha\gamma + \eta N \overline{c}}{2\gamma + \eta N} \quad \text{and} \quad \overline{c} = \frac{k}{k+1} c_1 \tag{3}$$

[6]Observe that Pareto is a good approximation of the upper tail of the productivity distribution, but a rather poor one of the lower tail (Combes *et al.*, 2012). However, analytical results using a distribution like the log-normal are unavailable in virtually any model in which locations have asymmetric sizes and productivities. Such distributions prove intractable in deriving the equilibrium expressions of our model.

denote the *domestic cost cutoff in the city* and the average marginal cost of entrepreneurs, respectively. The consumer price in the city is decreasing in the degree of competition, itself inversely related to c_1 as can be seen from expression (4) below. The cost cutoff c_1 is such that only entrepreneurs with marginal cost c lower than c_1 manage to serve the urban market. This cutoff satisfies the zero-profit condition $p(c_1) = c_1$. Using expressions (3), the mass of entrepreneurs actually selling in the city is given by:

$$N \equiv HG(c_1) = \frac{2\gamma(k+1)(\alpha - c_1)}{\eta c_1}. \tag{4}$$

Expression (4) establishes a positive equilibrium relationship between the mass of competitors N selling in the city and the toughness of selection prevailing there: the larger the mass of competitors, the smaller the share $G(c_1)$ of workers that can make it into entrepreneurship. We henceforth refer to $1 - G(c_1)$ as the 'failure rate' in the city.

2.4 Equilibrium payoffs

Using expressions (3) and (4), as well as the results in Appendix A.2, the consumer surplus can be expressed very compactly as follows:

$$\text{CS} \equiv \frac{\alpha - c_1}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_1\right). \tag{5}$$

The inverse productivity cutoff c_1 is proportional to the average inverse productivity $\overline{c}$ by (3); as such, it is a sufficient statistic to analyse the impact of any parameter change on consumer welfare. Clearly, consumer surplus is decreasing in the inverse productivity cutoff, namely, $\partial \text{CS}/\partial c_1 < 0$ since $c_1 \leq \alpha$ holds by (4). The consumer surplus is bounded above by $\alpha^2/(2\eta)$. Hence, imposing shopping costs for rural dwellers that exceed $\alpha^2/(2\eta)$ for urban goods implies that rural consumers *choose* to not consume the urban goods at equilibrium. We assume throughout this section that this condition holds.

We next use the profit-maximising prices (3) to express entrepreneurial profits as follows:

$$\Pi(c) = [p(c) - c]\, q(c) = \frac{H}{4\gamma}(c_1 - c)^2. \tag{6}$$

The masses of sellers obey the identity (4). The latter can be rewritten as:

$$\frac{\alpha - c_1}{A\eta c_1^{k+1}} \equiv H, \tag{7}$$

where $A \equiv 1/[2c_{\max}^k(k+1)\gamma]$ is a recurrent bundle of parameters that captures the natural advantage of location $l = 1$ to host a city. We can also think about A as being urban TFP. Note that A is decreasing in the upper bound $c_{\max}$ of the support of G and in the shape parameter k. As k rises, the mass of low-productivity entrepreneurs rises relative to the mass of highly productive ones,

and so a higher k implies a lower underlying productivity A. Observe that A also encapsulates the love-for-variety parameter γ: a higher γ means that consumers value variety more highly which, in equilibrium, implies that more firms operate at a smaller scale (see Ottaviano *et al.*, 2002; Peng, Thisse, and Wang, 2006).

The *indirect utility differential* between moving to the city or remaining in the countryside is

$$\Delta V(c) \equiv \Pi(c) + \text{CS} - \theta H - f^E, \tag{8}$$

and it depends on each agent's c, which is still unknown when the location decision is made. Ultimately, a worker decides to move to a city if her expected indirect utility is larger than the certain equivalent that she could secure in the numéraire sector in the countryside. Using (5) and as shown in Appendix A.4, the expected value of (8) is given by:

$$\mathbb{E}(\Delta V) = A\frac{Hc_1^{k+2}}{k+2} + \frac{\alpha - c_1}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_1\right) - \theta H - f^E. \tag{9}$$

Entry into the city takes place as long as it is profitable so that $\mathbb{E}(\Delta V) \leq 0$ holds at equilibrium, which we henceforth refer to as the *ex ante rationality constraint*. In words, at equilibrium expected profits and the urban consumer surplus of the marginal city dweller balance urban and entry costs.

3 Spatial equilibrium and urban income inequality

We now endogenize city size by imposing a no-arbitrage condition among locations. This is conventionally referred to as a *spatial equilibrium*. At a spatial equilibrium: (i) all agents optimally choose whether to move to the city or to remain in the countryside; (ii) urbanites optimally choose whether they want to be entrepreneurs or not; (iii) all entrepreneurs set profit-maximising prices for their products; (iv) all consumers maximise utility; and (v) all markets clear. Formally, a spatial equilibrium is given by a city size, H, and a marginal cost cutoff, c_1, such that: (i) entrepreneurs with $c \geq c_1$ fail to produce profitably; (ii) the identity (7) holds; and (iii) the ex ante rationality constraint (9) is satisfied. Put differently, at a spatial equilibrium, agents are either indifferent between the two locations if a city emerges, or they strictly prefer the countryside so that no city can form. Formally, either $\mathbb{E}(\Delta V) = 0$ if $H > 0$, or $H = 0$ if $\mathbb{E}(\Delta V) < 0$.

A spatial equilibrium also satisfies the *ex post rationality constraint* $\Delta V(c) \geq 0$ if, conditional on the realisation of c, no agent wants to change location. This will be the case in the presence of urban-to-rural migration costs when those costs are high. In what follows, we disregard the issue of return migration by assuming that the associated costs are large enough.[7]

[7]Formally, let f^M denote urban-to-rural migration costs. There is no return migration from the city to the countryside if and only if $\Pi(c) + \text{CS} - \theta H \geq -f^M$. At an interior spatial equilibrium with $\mathbb{E}(\Delta V) = 0$, this condition is equivalent to $\mathbb{E}[\Pi(c)] - \Pi(c_1) \leq f^M + f^E$. Obviously, choosing a sufficiently large f^M ensures that ex post rationality holds as return migration is too costly.

3.1 Spatial equilibria with one region: 'Urbanisation'

Only two types of spatial equilibria may arise in the case of an isolated city: an equilibrium in which no city forms and an equilibrium in which a city forms. We refer to the former as a *rural equilibrium* ($H^* = 0$, and thus $c_1^* = \alpha$), and to the latter as an *urban equilibrium* ($H^* > 0$, and thus $0 < c_1^* < \alpha$).

The set of equilibria of the model is determined by the free entry condition and the identity which pins down the mass of entrepreneurs. Using expression (7) to substitute for H in (9) yields the free entry condition as a function of c_1 alone:

$$f(c_1) \equiv \frac{\alpha - c_1}{2\eta}\left[\alpha - \frac{k-1}{k+2}c_1 - \frac{2\theta}{A}c_l^{-(k+1)}\right] - f^E \leq 0. \tag{10}$$

Condition (10) is central to studying the number and nature of spatial equilibria. An urban equilibrium is such that $f(c_1^*) = 0$, whereas a rural equilibrium is such that $f(\alpha) \leq 0$. As is standard in the literature, we will focus on *stable* equilibria only. A spatial equilibrium is (locally) stable if and only if any small perturbation of the population distribution is self-correcting and brings the economy back to its initial situation. A rural equilibrium is always stable whenever it exists since $f^E > 0$, whereas an urban equilibrium is *locally stable* if and only if $\partial f(c_1^*)/\partial c_1 > 0$.

We first establish the existence of a spatial equilibrium in our model and characterise the number of possible equilibria. We then derive their comparative static properties and discuss under which conditions what type of equilibrium arises. Concerning existence, we can show the following results:

Proposition 1 (existence and number of spatial equilibria) *The function f has either one or three positive roots, of which at most two are in $[0, \alpha)$. Consequently, there exist at most two stable spatial equilibria: an urban equilibrium and the rural equilibrium. If no stable urban equilibrium exists, then the rural equilibrium is unique. The spatial equilibrium associated with the smallest value of c_1 is always stable.*

Proof. See Appendix B.1. ■

Proposition 1 establishes that whenever a rural equilibrium does not exist there exists at least one stable urban equilibrium; this holds by continuity. Also, whenever a smallest root of f exists – which corresponds to the largest equilibrium city size – it is a stable spatial equilibrium as in Henderson (1974). The next proposition establishes that all equilibria have the same comparative static properties:

Proposition 2 (equilibrium properties) *At any stable spatial equilibrium, $1/c_1^*$ and H^* are both non-increasing in θ and f^E and non-decreasing in α and A.*

Proof. See Appendix B.2. ■

As shown by Proposition 2, lower commuting costs θ, a stronger preference α for the differentiated good, and a better natural advantage or urban productivity A all weakly increase city size and city productivity at any stable spatial equilibrium.[8]

We next investigate under what conditions which *type* of spatial equilibrium arises. Obviously, for any $f^E > 0$, the rural equilibrium exists and is stable: no single agent wants to sink f^E if nobody else enters the city. Conversely, when f^E is small enough, both rural and urban equilibria may coexist and coordination failures may arise. If $\theta \geq \theta^R$ and $f^E \geq f^R$, where

$$\theta^R \equiv \frac{3A\alpha^{k+2}}{2(k+2)} > 0 \tag{11}$$

and

$$f^R \equiv \frac{\alpha^2}{2\eta}\frac{k-1}{k+2} > 0, \tag{12}$$

then the rural equilibrium is the unique spatial equilibrium. Note that the sufficient conditions $f^E \geq f^R$ and $\theta \geq \theta^R$ for the rural equilibrium to be the unique stable spatial equilibrium are less likely to hold if urban productivity A is high and if consumers value urban output a lot, i.e., α is large. Otherwise, there exists a threshold $\theta^U(f^E)$, with $\theta^U(f^E) < \theta^R$ and $\lim_{f^E \to 0} \theta^U(f^E) = \theta^R$, such that there is at most one stable urban equilibrium if $\theta \leq \theta^U(f^E)$. There also exists an urban productivity threshold $A_{\min}$ such that $H = 0$ for all $A < A_{\min}$. To see this, note that the utility differential $f(c_1)$ in (10) is negative for all $c_1 \in [0, \alpha]$ at the limit $A \to 0$: cities cannot arise if urban productivity is too low. Appendix B.3 summarizes the equilibrium structure of the model.

We depict the equilibrium structure of the model in Figure 1, which plots the equilibrium city size H against the urban cost θ. We use bold lines to denote stable spatial equilibria (the dashed schedule illustrates the unstable urban equilibrium). As one can see, H is non-increasing in θ and $H = 0$ is the unique equilibrium beyond some threshold $\theta^U < \theta^R$. Also, the rural equilibrium $H = 0$ is a stable equilibrium for *any* θ.

We can summarise the key properties of the model by focusing on the 'urbanisation threshold' θ^U. No city can emerge for $\theta \geq \theta^U$ and/or for $A \leq A_{\min}$. Any improvement in the benefits of living in cities, either as consumers or entrepreneurs, makes the emergence of cities more likely and maps into larger equilibrium city sizes and higher city productivity. These findings are consistent with

[8]Recall that A is decreasing in love-for-variety γ so that stronger love-for-variety reduces city sizes in our model. This is in contrast to the representative-firm models of Ottaviano *et al.* (2002) and Peng *et al.* (2006). To understand this striking difference, note that γ does not enter the equilibrium consumer surplus directly in (5). Instead, a more pronounced taste for variety relaxes competition among entrepreneurs by making demands less elastic, thereby relaxing selection pressures by decreasing the equilibrium productivity cutoff $1/c_1$. Having on average less productive firms reduces the consumer surplus, thus giving agents less incentives to agglomerate in big cities. We thank an anonymous referee for pointing this out to us. Note that Behrens, Mion, Murata, and Südekum (2012) obtain a related counterintuitive result in a very different model. There, smaller urban costs *ceteris paribus* reduce productivity in cities, for any given city size, by making the survival of low productivity firms easier. The reason is that consumers become richer when urban costs fall, so that firms face less elastic demands.

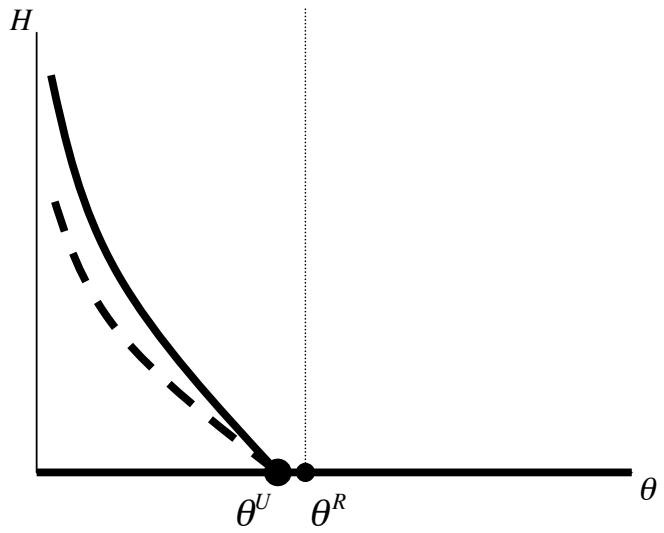

Figure 1: Equilibrium structure of the model with a single city

the three 'classical' conditions stressed in the literature for cities to emerge and to develop (e.g., Bairoch, 1988). First, there must be some demand for urban goods and services: the extent of this demand is captured by the parameter α and urban production is more valuable if it is produced at a larger scale in equilibrium. Second, the urban population must supply goods and services to sustain itself. It is able to produce more of these, the stronger its natural advantage, A. Last, any reduction in urban costs that stems from an improvement in urban transportation is conducive to urban development (Duranton and Turner, 2012). To sum up, a large α and low γ, θ, f^E or $c_{\max}$ are all conducive to the emergence of large cities.

3.2 Selection and urban income inequality

In our model, larger cities are more productive because of selection. What does this imply for the relationships between city size and city per capita income, and between city size and city income inequality? These are not trivial questions since only a share $G(c_1)$ of agents survive as entrepreneurs, whereas the remaining ones exit the market and consume from their endowments. The failure rate $1 - G(c_1)$ thus influences both moments of the income distribution of any city. We now show that selection increases both per capita income and urban inequality, two predictions that are robustly borne out by the U.S. data.

We first compute the average operating profit of all urbanites, including those who end up failing as entrepreneurs at equilibrium.[9] It is given by:

$$\overline{\Pi} = A\frac{Hc_1^{k+2}}{k+2} = \frac{c_1(\alpha - c_1)}{\eta(k+2)}, \tag{13}$$

[9]Recall that those who fail consume from their endowment and earn zero income. Hence, the average operating profit provides a measure of urban per capita income in our model.

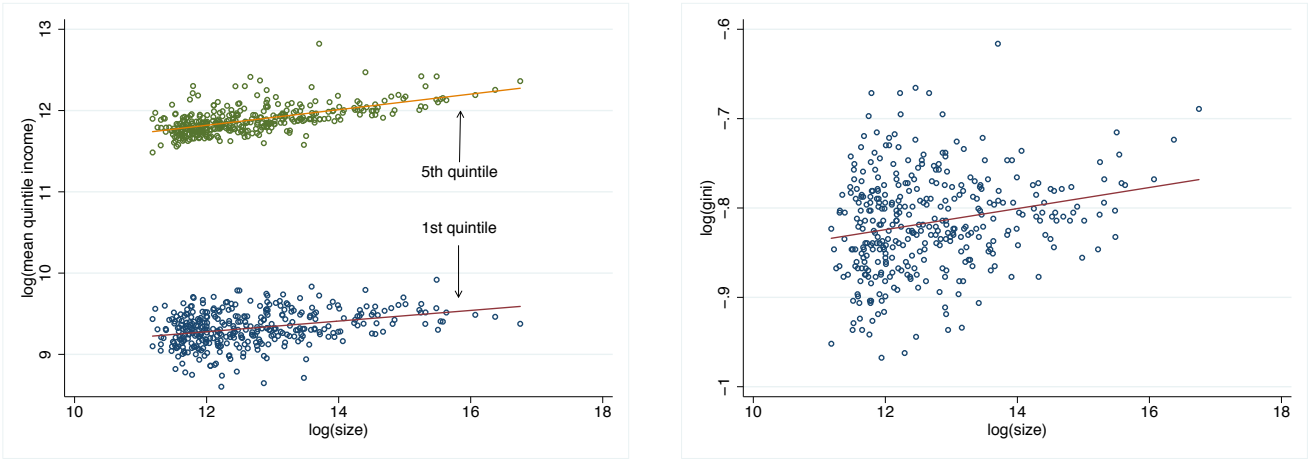

Figure 2: City size dilates the income distribution, larger cities have higher Gini coefficients.

where the second equality follows from the equilibrium size-productivity relationship (7). It is readily verified that $\partial\overline{\Pi}/\partial c_1 = (\alpha - 2c_1)/[\eta(2+k)]$. Hence, $\overline{\Pi}$ is $\cap$-shaped. The relationship is non-monotonic because operating in a large city involves both costs and benefits that are reflected in average profits: a large market size increases profits (the 'H' component in the expression for $\overline{\Pi}$) but it also induces tougher competition, thereby reducing markups and profits (the 'c_1' component in the expression for $\overline{\Pi}$).

The foregoing results suggest that selection makes larger cities more unequal. To measure inequality formally, define first $\overline{\Pi}_Q$ as the average profit earned by the top Q% of the distribution and let $\sigma_Q \equiv \overline{\Pi}_Q/\overline{\Pi}$ define the average income of the top Q% of the distribution relative to the overall average. We show in Appendix B.4 that σ_Q is equal to

$$\sigma_Q = \frac{k(k+1)(k+2)}{2}\left(\frac{c_1}{c_{\max}}\right)^{-k}\left[\frac{1}{k} - \frac{2}{k+1}\frac{q}{c_1} + \frac{1}{k+2}\left(\frac{q}{c_1}\right)^2\right], \tag{14}$$

where $q \equiv G^{-1}(Q)$ for $q \leq c_1$ and $\sigma_{Q_1} = (c_1/c_{\max})^{-k} > 1$ holds by definition (i.e. successful entrepreneurs represent only a fraction $(c_1/c_{\max})^k$ of the urban population but they earn all its income). We also compute the Gini coefficient of the income distribution in the city (see Appendix B.4 for details):

$$\text{Gini}(k, c_{\max}, c_1) = 1 - \frac{k+2}{4k+2}\left(\frac{c_1}{c_{\max}}\right)^k. \tag{15}$$

We are interested in the equilibrium relationship between city size and income inequality. It turns out that urban income inequality and city size are positively related for any of the foregoing measures of inequality:

Proposition 3 (city size and urban income inequality) *City size disproportionately benefits agents of the highest quintiles: for all $q < c_1$, $\partial\sigma_Q/\partial H > 0$. The Gini coefficient is: (i) increasing at an increasing rate in the productivity cutoff $1/c_1$; and (ii) increasing at a decreasing rate in city size H. Furthermore, conditional on c_1, the Gini coefficient is (iii) decreasing in k; and (iv) increasing in $c_{\max}$.*

Proof. See Appendix B.5. ■

Proposition 3 establishes three results. First, there is a 'superstar' effect whereby the elasticity of quantile income with respect to city size increases as we move up the income distribution. This relationship is illustrated by Figure 2 for a sample of US MSA's (see Supplemental Appendix C for more details on the data and the empirical implementation). As can be seen from the left panel of Figure 2, the elasticity of quintile mean income with respect to city size is positive for both the 1st and the 5th quintiles of the income distribution, and the elasticity pertaining to the 5th quintile is larger than the elasticity pertaining to the 1st quintile: *the skill premium is increasing in city size*. Second, larger cities are more unequal as measured by the income Gini coefficient. This relationship is depicted in the right panel of Figure 2. Last, Proposition 3 suggests that the expansion of the share of income accruing to the wealthiest comes at the expense of both the bottom half of the population of successful entrepreneurs and of those who simply fail. Thus, the positive relationship between city size and inequality is driven by both the top and the bottom of the income distribution.

To summarise our key findings, *larger urban areas generate more wealth and are at the same time more unequal than smaller cities.*[10] Our model suggests that the observed 40-year rise of the incomes of the 'working rich' relative to the population as a whole (Piketti and Saez, 2003) may at least be partly correlated with the intensive margin of urbanisation documented in the introduction, i.e., city growth. To the best of our knowledge, little attention has thus far been devoted to this aspect (see Moretti, 2013, for a complementary explanation based on the spatial sorting of college graduates).

4 Urban systems, trade, and income inequality

We have shown in the foregoing section that everything that makes cities more attractive favours their emergence, increases their equilibrium sizes, and makes them more unequal. Although we derived these results in a setting with a single city, we now show that the same logic goes through in an environment with numerous cities that trade with each other. More specifically, we extend our analysis to a symmetric environment with multiple cities and transportation costs in subsection 4.1, and we establish that lower transportation costs favour urbanization and the size of cities. It also makes them more unequal. We then illustrate the behavior of the model in an asymmetric setting in subsection 4.2. Since few clear analytical results are available here, we provide some numerical simulations using US data to illustrate the novel effects that arise in this environment, such as cities that are more centrally located may be bigger than more peripheral ones, how trade integration affects urban income inequality, and how lower transportation costs

[10] Glaeser *et al.* (2009) show that access to public transportation explains why central cities attract more poor people than the suburbs. Their analysis focuses on the determinants of the intra-city distribution of *poverty*, not *inequality*.

favour both margins of urbanization: the *number* and the *size* of cities.

Formally, consider an economy with Λ locations that can potentially host cities. We use subscripts l to denote variables that pertain to specific cities. Cities may differ in their underlying urban productivity, A_l, and in their bilateral transportation costs, modelled as iceberg trade costs and denoted by τ_{hl} for all city pairs: only one unit for every $\tau_{hl} > 1$ units of the urban good shipped from city h are actually delivered at city l. The equilibrium expressions for prices, quantities, profits, and therefore the expression of the consumer surplus (5) that pertains to the one-city case, namely, $\text{CS} = (2\eta)^{-1}(\alpha - c_1)\left[\alpha - c_l(k+1)/(k+2)\right]$, remain valid in this new environment. Appendices A.2–A.4 provide the computational details.

4.1 Symmetric urban systems

To set the stage, assume that locations are symmetric in two ways. First, they are equally amenable to urban development in the sense that $A_l = A > 0$ for all $l = 1, 2, \ldots, \Lambda$. Put differently, cities have the same underlying productivity distribution and there are no differences in natural advantage. Second, cities may trade their urban goods with each other at a common iceberg cost $\tau > 1$. We define the trade 'freeness' between any two cities as $\phi \equiv \tau^{-k} \in (0, 1)$, with $\phi = 0$ when trade costs are prohibitive and $\phi = 1$ in the absence of trade frictions.

We start by rewriting the expression governing the equilibrium relationship between the cutoffs, c_l, and city sizes, H_l, which replaces expression (7) in the previous section, as follows:

$$\frac{\alpha - c_l}{A\eta c_l^{k+1}} = H_l + \phi \sum_{h \neq l} H_h. \tag{16}$$

We can apply (16) to any two cities, e.g. $l = 1, 2$, and take the difference between the two resulting expressions to obtain:

$$\frac{\alpha - c_1}{A\eta c_1^{k+1}} - \frac{\alpha - c_2}{A\eta c_2^{k+1}} = (H_1 - H_2)(1 + \phi).$$

It then follows by inspection that $H_1 > H_2$ if and only if $c_1 < c_2$, that is, the larger city is also the most selective and thus the most productive. Note that city size differences translate into larger equilibrium productivity differences if trade is relatively free. We have thus shown:

Proposition 4 (size, selection, and productivity in an urban system) *Consider a symmetric environment with $A_l = A > 0$ for all cities l and bilateral iceberg costs equal to $\tau \in (1, +\infty)$ for all city pairs. Then selection is tougher and productivity is higher in larger cities at any spatial equilibrium:*

$$c_l \leq c_h \quad \Longleftrightarrow \quad H_l \geq H_h$$

and $\partial H_l / \partial c_l < 0$ and $\partial H_l / \partial c_h > 0$.

Proof. See Appendix C.1. ■

Proposition 4 establishes three important results. First, when accessibility differences are small enough, selection is tougher and, as a result, average productivity is higher in larger cities. This holds true irrespective of the number of cities in the urban system.[11] Second, the positive relationship between the number of available varieties and the toughness of selection in (16) implies the existence of a hierarchy of cities: larger cities offer a larger range of goods and services. Finally, while tougher selection in any city is beneficial to its own size, it negatively impacts the size of other nearby cities. We refer to this negative dependence as the 'cannibalisation effect' of cities (Dobkins and Ioannides, 2000). This is in line with the empirical results of Partridge, Rickman, Ali, and Olfert (2009) who find that quite large nearby cities cast 'agglomeration shadows', i.e., they inhibit the growth of other nearby cities.

A symmetric equilibrium exists in this symmetric environment. We denote the inverse productivity cutoff, common to all cities, and the common city size by c_1 and H, respectively. Let $\Phi \equiv (\Lambda - 1)\phi$ be a measure of overall market integration, with $\Phi = 0$ in the one-city case of Section 3, and $\Phi = \Lambda - 1$ in the limiting case of no trade frictions. It turns out that Φ and urban productivity A enter all equilibrium expression jointly as $(1 + \Phi)A$. Indeed, expected profits are equal to $(1 + \Phi)AHc_1^{k+2}/(k + 2)$ and the market clearing condition (16) simplifies to $(\alpha - c_1)/[(1 + \Phi)A\eta c_1^{k+1}] = H$. Plugging these into the free-entry condition (9) yields

$$\frac{\alpha - c_1}{2\eta}\left[\alpha - \frac{k-1}{k+2}c_1 - \frac{2\theta}{(1+\Phi)Ac_1^{k+1}}\right] - f^E \equiv f(c_1) \leq 0. \tag{17}$$

Rural and urban equilibria are defined as in the single-city case of Section 3. The whole analysis pertaining to the role of A in relation to the types and stability of equilibria and the comparative statics of the previous section readily extend to the role of Φ. The positive implications of a decrease in transport costs are isomorphic to those of an increase in the underlying productivity of the whole economy, thus implying that lower transport costs increase city productivity. Note that this result may not be as obvious as it sounds. Indeed, from the perspective of entrepreneurs in each city, lower inter-city trade costs and a larger number of trading partners mean both a better market access and tougher competition from entrepreneurs established in other cities. As it turns out, Proposition 5 below establishes that the agglomeration effect dominates in equilibrium:[12] Formally:

[11] We show in Appendix C.1 that the result of Proposition 4 extends to a situation of asymmetric trade costs when k is large enough. Note that Proposition 4 pertains to the equilibrium relationship between size and selection: at any multi-city equilibrium, size and productivity move in the same direction. Proposition 4, however, makes no statement as to which changes in exogenous parameters induce the change in city size and productivity in the first place.

[12] Using the German division and reunification as a natural experiment, Redding and Sturm (2008) establish a causal relationship between market access, which depends on trade frictions, and the size and growth of cities. Brülhart, Carrère, and Trionfetti (2012) establish a similar result for Austrian border regions using the fall of the Iron Curtain as an exogenous source of variation in market access.

Proposition 5 (city size, productivity, and inequality at the symmetric equilibrium) *A larger Φ (lower trade costs τ and/or more trading partners Λ), a larger A, or a lower θ: (i) are conducive to the emergence of cities ('extensive margin' of urban development); and (ii) they weakly increase the equilibrium sizes of cities ('intensive margin'). In addition, (iii) lower trade costs are associated with a more unequal distribution of earnings in all cities: $\partial \text{Gini}/\partial\tau < 0$.*

Proof. See Appendix C.2. ■

Part (iii) of Proposition 5 is a novel result relative to the one-city model. It establishes that lower trade frictions among symmetric cities make these cities more unequal. To establish this formally, we compute the Gini coefficient evaluated at the symmetric equilibrium and we write it as follows:

$$\text{Gini}(\Lambda, \tau, k; c_1) = 1 - \lambda(\Lambda, \tau, k) \left(\frac{c_1}{c_{\max}} \right)^k, \tag{18}$$

where $\lambda(\cdot)$ is a bundle of parameters too unwieldy to be revealing (its expression is relegated to Appendix C.3). The effect of trade integration on inequality in cities is threefold. First, cheaper transportation has a pro-competitive effect and this hurts the profits of *every* producer. Second, tougher selection raises the failure rate, i.e., $(c_1/c_{\max})^k$ falls. Last, lower transport costs also open distant markets to *some* entrepreneurs – the exporters. As a result, a lower τ unambiguously raises the exporters' share of profits in each city. Since only the most productive entrepreneurs export, it follows logically that lower transport costs increase income inequality by shifting profits from non-exporters to exporters.

4.2 Asymmetric urban systems

Though insightful, the ability of the symmetric equilibrium pattern to illustrate the fate of heterogeneous cities is limited. The analysis of asymmetric urban systems when cities interact through trade is quite involved (Fujita, Krugman, and Venables, 1999; Tabuchi, Thisse, and Zeng, 2005). By way of making progress, this subsection presents three numerical examples involving data on US cities (Supplemental Appendix B describes the data and the numerical procedure). We use information for 356 metropolitan and micropolitan statistical areas in the year 2007 that includes population size, total employment, city GDP, latitude, longitude, city surface, aggregate rent, and the Gini coefficient of income inequality. We approximate trade costs by $\tau_{hl} = d_{hl}^{\delta}$, where the distance d_{hl} is computed as the great circle distance between MSA centroids (by definition, $d_{hl} = d_{lh}$). In the numerical application, we relax the assumption that there are no trade costs internal to cities. More precisely, we approximate the internal distances by using the formula suggested by Redding and Venables (2004). We pick the distance elasticity of trade costs from Duranton *et al.* (2012), who estimate it for a sample of large US cities using Commodity Flow Survey data, and set it to $\delta = 0.81$. Calibrating the preference parameters (α, η, γ), as well as the shape parameter k and

the sunk entry cost f^E is difficult in this model and beyond the scope of our illustrative exercise. Hence, we somewhat arbitrarily set $(\alpha, \eta, \gamma, k, f^E) = (30, 1, 5.2, 2.3, 2.2)$. The qualitative results are not sensitive to that choice of parameter values.[13]

We first 'fit' the model to the initial observed distribution of city size and urban productivity, using the observed spatial structure between cities. We then selectively change either the distance elasticity δ or pairwise distances between cities and compute the implied equilibrium responses of city populations and productivities. We also report changes in Gini coefficients.

Illustration 1: Falling distance elasticity of trade costs. We first decrease the distance elasticity of trade costs by 10%, which can be viewed as an economy-wide improvement in transportation technology. The total increase in urban population is of 33,920,656 people, about 13.56% of the initial urban population. This corresponds to the intensive margin of urbanisation since the total number of cities is held fixed at 356. The average (unweighted) change in the cutoffs across MSAs is -11.16% which, given our value of k=2.3, corresponds to a productivity gain of 7.78%. The unweighted average change in population across MSAs is 19.36%. It is worth pointing out that small cities grow more than large cities in relative terms, so the size distribution tilts slightly. However, it generally remains fairly stable. Table 1 shows the five cities that move up the most ranks in the hierarchy, as well as the five cities that move down the most ranks. Cities that move up the size distribution tend to be relatively central, while cities that move down the size distribution tend to be located closer to the coasts: market access matters. It is further worth pointing out that changes in rankings take place in the bottom of the distribution, whereas – as expected - the top of the distribution remains very stable. Figure 3 summarises the changes in cutoffs and populations. Result (iii) in Proposition 5 suggests that this increased market integration should also increase income inequality in cities. Computing the implied changes in the Gini coefficients, the average (unweighted) change across the 356 MSAs is of 0.87%. In words, the result whereby deeper trade integration exacerbates income inequality in cities, continues to in the current asymmetric setup. The changes in the Gini coefficients are positive for all but one of the 356 cities. Small cities are especially prone to increasing inequality, being those whose population grows the most and that benefit the most from better access to markets. The distribution of changes in the Gini coefficients are depicted in the left panel of Figure 5.

Illustration 2: Rising distance elasticity of trade costs. We next increase the distance elasticity of trade costs by 5%. The aim of this exercise is to reveal the importance of cheap transportation for the sizes of cities and for the extent of urbanisation in general. The total decrease in urban population is of 35,223,500 people, about 14.09% of the initial urban population. The size of the

[13] We focus on the intensive margin of urbanisation, holding the number of cities fixed. Since the number of cities is fixed and because the shocks to trade costs are relatively small, the issue of multiple equilibria is crucial to none of our three numerical illustrations.

Table 1: Top five cities moving up or down the urban hierarchy (Numerical illustration 1).

MSA name	Initial population	Final population	Initial rank	Final rank	Change in rank
Flagstaff, AZ	127.450	176.734	291	262	-29
Cumberland, ML-WV	99.316	138.623	339	312	-27
Dothan, AL	139.499	188.421	270	244	-26
Brownsville-Harlingen, TX	387.210	553.005	127	102	-25
St. George, UT	133.791	184.264	278	253	-25
⋮	⋮	⋮	⋮	⋮	⋮
Santa Fe, NM	142.955	161.381	267	286	19
Jacksonville, NC	162.745	182.865	235	255	20
Hanford-Corcoran, CA	148.875	171.846	255	275	20
Yuba City, CA	164.138	184.641	230	252	22
Napa, CA	132.565	147.448	279	303	24

Notes: Population values are reported in 1000s. There are 356 cities in our numerical illustrations. Negative rank changes indicate cities that move up in the urban hierarchy.

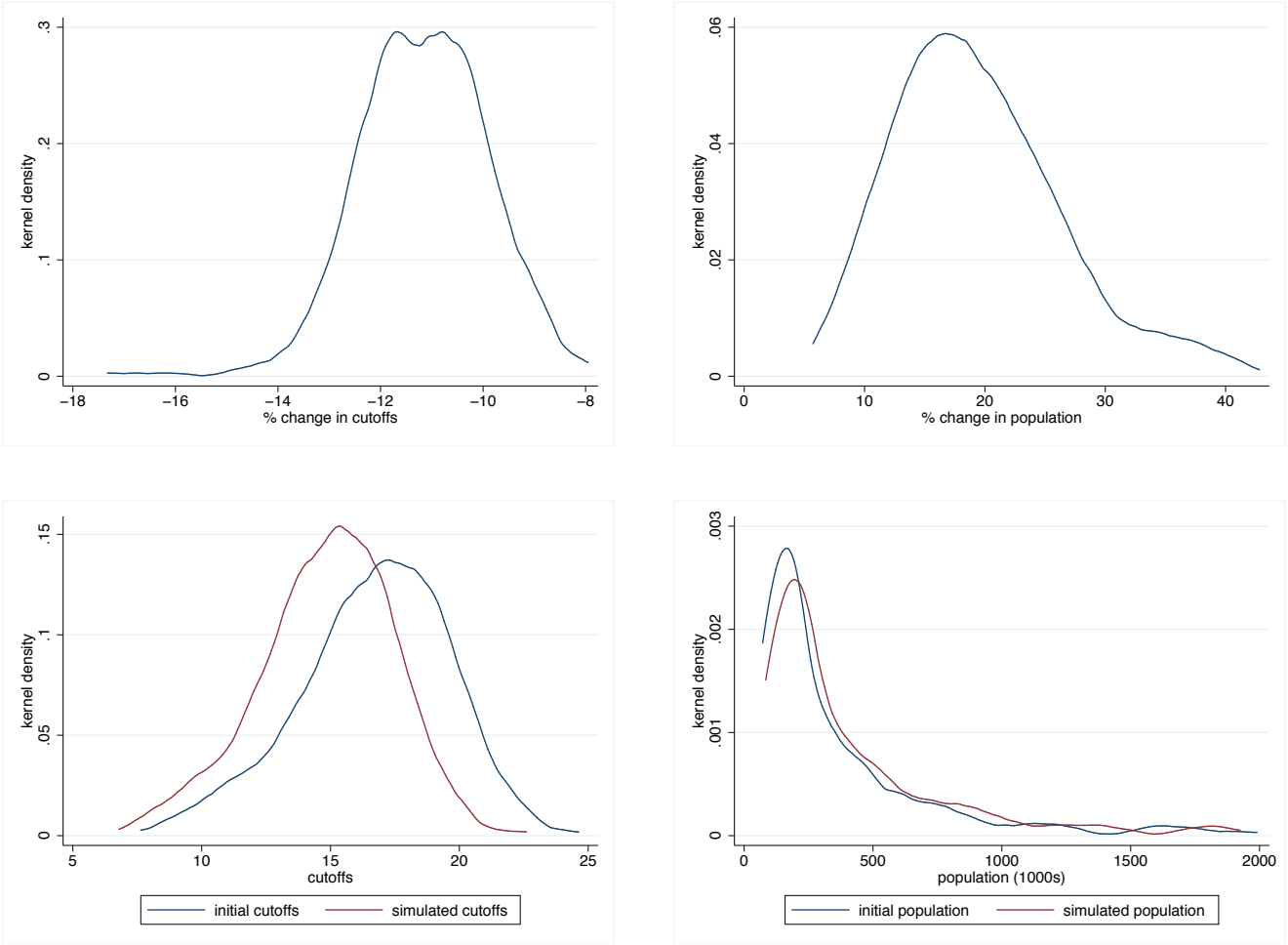

Figure 3: Summary results on productivity and size for Illustration 1.

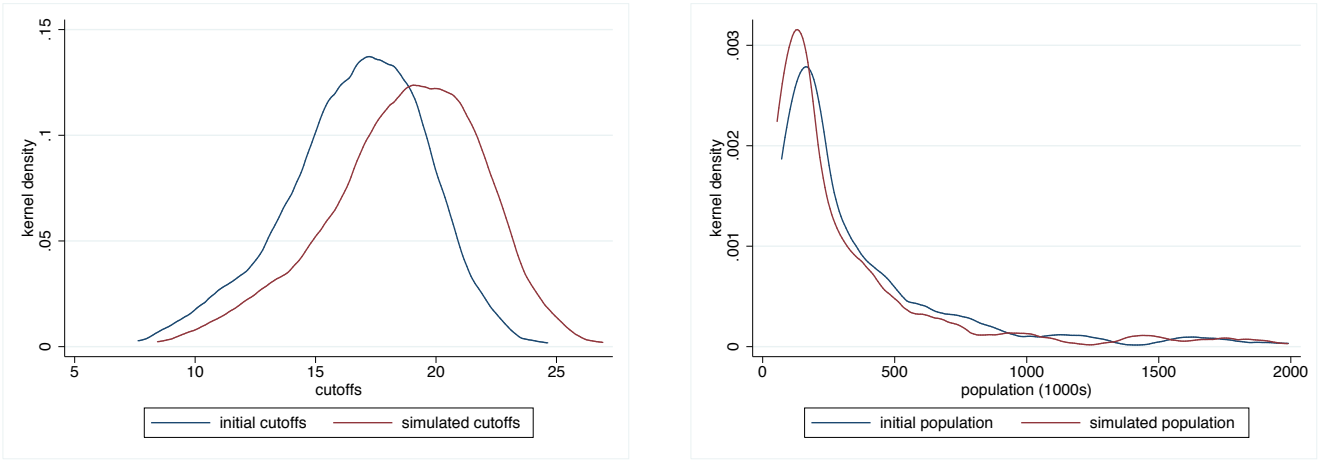

Figure 4: Summary results on productivity and size for Illustration 2.

total effect clearly reveals how important cheap transportation is for urbanisation and for the sizes of individual cities. The average (unweighted) change in the cutoffs across MSAs is 11.99% which, given our value of k=2.3, corresponds to a productivity loss of 8.36%. The unweighted average change in population across MSAs is -19.42%. It is worth pointing out that small cities lose more than large cities in relative terms, which is in line with the empirical findings of Brülhart, Carrère and Robert-Nicoud (2013). The size distribution remains, however, fairly stable again. Changes in rankings take place at the bottom of the distribution, whereas – as expected - the top of the distribution remains very stable. Figure 4 summarises the changes in cutoffs and populations. Again, we compute the implied changes in the Gini coefficients. The average (unweighted) change across the 356 MSAs is of -0.41%. In words, less trade reduces income inequality in cities in this asymmetric setup, which is in line with the formal result in Proposition 5, albeit derived in a symmetric environment. This reduction in earnings inequality is especially strong in large cities, as those lose more population and, therefore, see their market size shrink more (in absolute terms) than small cities. The changes in the Gini coefficients are negative for 352 cities and positive for the remaining 4. The distribution of changes in the Gini coefficients are depicted in the right panel of Figure 5.

Illustration 3: Transport improvements between New York and Chicago. We then reduce the distance between New York and Chicago by 50%, keeping all other distances and the distance elasticity of trade constant. This exercise can be viewed as simulating the impacts of a specific transportation infrastructure project that would make trade between selected city pairs more efficient. The average (unweighted) change in the cutoffs across MSAs is barely -0.001% which, given our value of $k = 2.3$, corresponds to a very small productivity gain. The changes for New York and Chicago – which are directly affected by the change in distance – are of course 'much' larger, namely -0.057% and -0.162% respectively. Changes in cutoffs in third cities can go in either direction: as Chicago and NYC grow, they provide both larger markets and tougher competition to

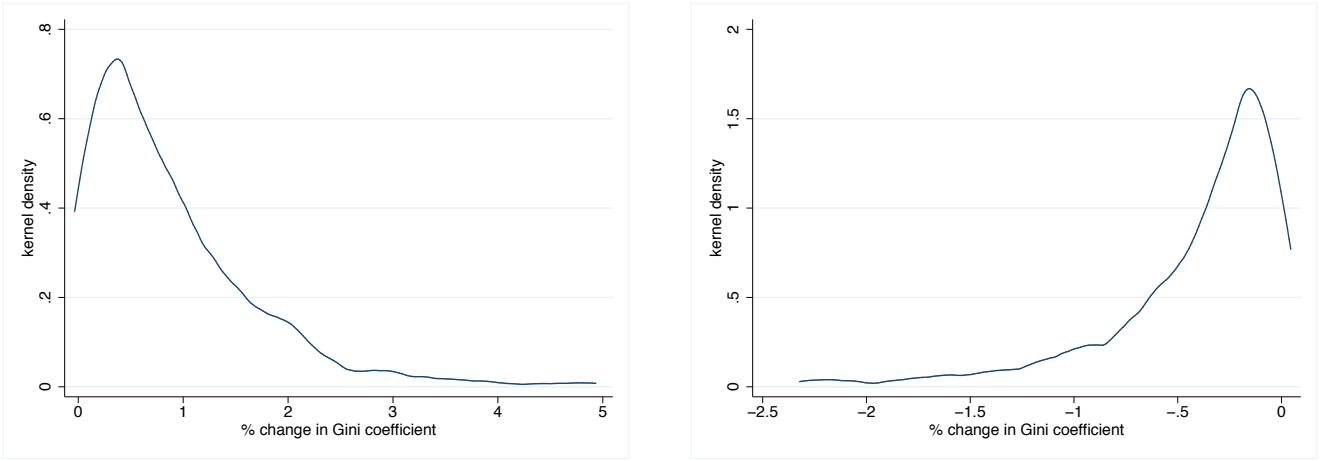

Figure 5: Summary of changes in the Gini coefficients (Illustration 1, left; Illustration 2, right).

nearby cities. In our numerical exercise, there are 75 cities for which productivity decreases. The unweighted average change in population across MSAs is -.0012%. Yet, New York and Chicago grow by 0.076% and 0.066%, respectively; this population gain that results from improved market access is in line with the empirical findings of Redding and Sturm (2008) and Brülhart *et al*. (2012, 2013). Note that 25 other MSAs also grow, whereas the 329 remaining ones actually lose population. Last, total urban population still increases by about 0.007%, which corresponds to 17,953 people. It is worth pointing out that our illustration reveals the presence of 'agglomeration shadows', as highlighted in subsection 4.1: although the total urban population increases and although New York, Chicago, and some other places grow, the better connection between New York and Chicago hurts the majority of other cities.

In all of the foregoing, the total number of cities was held fixed. In our final numerical example we relax that assumption and look in more depth at the extensive margin of urbanisation. That margin seems especially relevant in the context of the developing world, where new cities emerge and where small villages can quickly transform into major urban centres.

Illustration 4: Extensive margin of urban development. Small reductions in trade frictions may lead to 'massive urbanisation' in a more general asymmetric environment. To see this, we consider a comparative statics exercise that consists in comparing different equilibria as parameter values change. We decompose the passage from one equilibrium to another into several steps for illustrative purposes. Each of those steps highlights the interactions between the extensive and the intensive margins of urbanisation. The reader should keep in mind that there is no structural dynamic interpretation within our modelling framework.

Consider four asymmetrically located regions. Population is measured in thousands. The top panel of Figure 6 illustrates the initial configuration of the space-economy when trade costs are

relatively high.[14] The initial equilibrium configuration is of the 'core-periphery' type, with a single, relatively small city of about 60,000 inhabitants. The remaining three locations are empty as no city can form there at this stage.[15] Consider now a uniform decrease in all τ_{hl} of 0.1. The initial 'core-periphery' configuration is no longer stable: the indirect utility differential in regions 3 and 4 remains negative, but the indirect utility differential in location 2 turns positive. Hence, a second city forms in region 2. Imposing this configuration, we obtain $H_1^* = H_2^* = 85.59$ and $c_1^* = c_2^* = 2.33$. In words, a new city has formed (extensive margin), and the existing city has grown (intensive margin). In addition, these larger city sizes put downward pressure on the equilibrium cutoffs, i.e., there are productivity gains in the economy.

This is not the end of the story, however. This new two-city configuration is not an equilibrium as the indirect utility differential in regions 3 and 4 has become positive, too: the rise of city 1 and the emergence of city 2 offer additional and expanding markets for urban goods that locations 3 and 4 may produce. In turn, this yields rural-urban migration to regions 3 and 4 and two additional cities appear. The final stable equilibrium configuration is depicted in the bottom panel of Figure 6, and it has four large cities: 103,590 inhabitants, 102,780 inhabitants, 101,050 inhabitants, and 101,870 inhabitants, respectively. As one can also see from the figure, there have been large productivity gains between the initial equilibrium and the new equilibrium. Observe also that the size-productivity relationship, highlighted in Proposition 4, holds in our example.

In this example, the quantitative effect of a small change in trade costs on both population and productivity is substantial. Indeed, decreasing trade costs make city 1 grow by almost 73%, whereas the overall increase in the urban population amounts to 583%. Put differently, there is 'massive urbanisation' and the intensive margin contributes about 73% to urban growth, whereas the extensive margin contributes about 510% to that growth. Last, productivity in city 1 also increased substantially between the initial and the final configuration, namely by about 74%.

5 Conclusions

All empirical studies reveal that the elasticity of worker and firm productivity with respect to city size is positive and typically falls in the 3%–8% range. Less well known is the fact that larger cities are also more unequal: the average incomes of the highest income quantiles are magnified by city size, so that income inequality is increasing with urban size. The model we develop establishes clear links between city size, productivity, and inequality. It can shed light on phenomena such as urbanisation, and allows us to investigate the impacts of trade integration on city size and inequality. It is further able to cope with the various two-way interactions between size and productivity: a larger city size increases productivity via a selection process, whereas higher

[14] We use $\alpha = 12.2$, $\gamma = 2$, $\eta = 2$, $\theta = .1$, $c^{\max} = 10$, $f^E = 21$ and $k = 1.3$ throughout this numerical example.

[15] See Supplemental Appendix C for the procedure used to check the stability of equilibria.

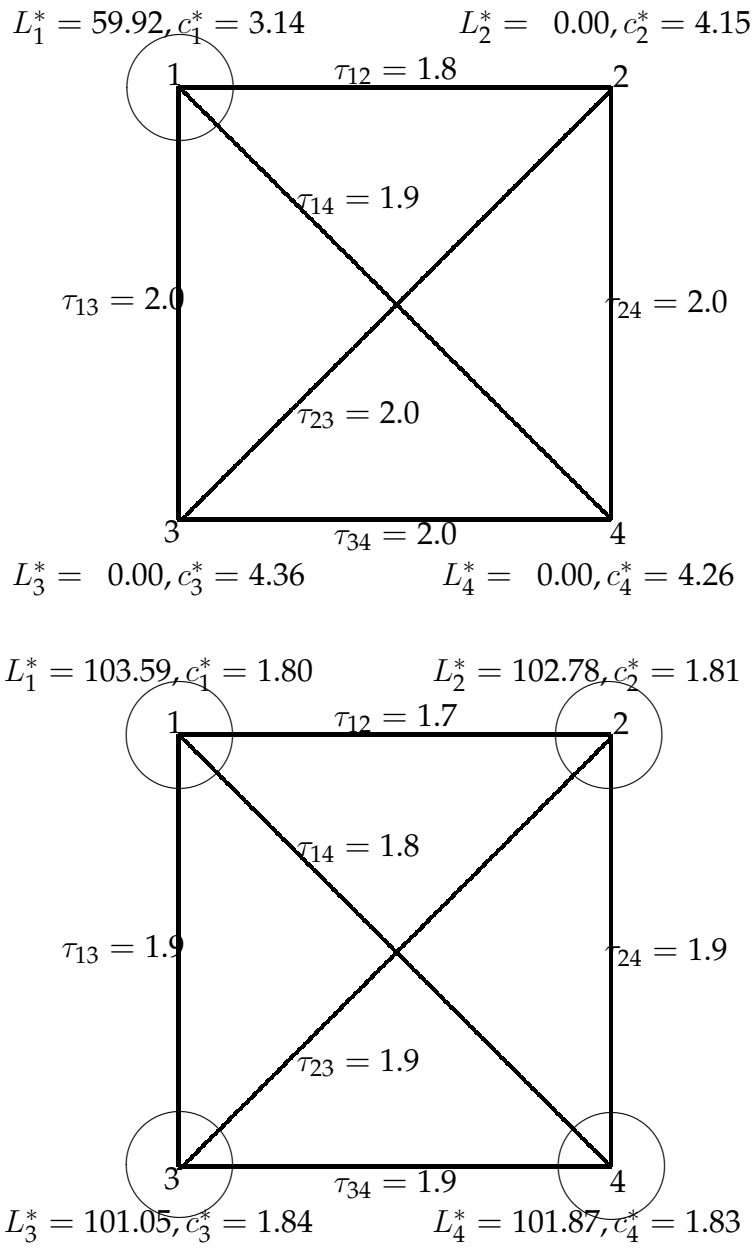

Figure 6: A case of massive urbanization with $\Lambda = 4$ regions (top = initial, bottom = final).

productivity increases city size by providing incentives for rural-urban migration. This circularity plays an important role in explaining episodes of rapid urbanisation and productivity gains, where both the number of cities and the size of individual cities rapidly change.

Avenues for future research include combining agglomeration, selection, and sorting in a unifying framework. To our knowledge, such a model is missing to date (in Behrens *et al.*, 2012, selection is trivial once sorting is controlled for). The theoretical analysis presented in this paper has also largely left untouched issues that can only be addressed in a rigorous manner by studying in a general way the asymmetric equilibria of the model. Studying the resulting urban hierarchies is a notoriously hard task but which is certainly worthwhile to undertake in order to garner additional

insights into the emergence and the evolution of urban systems. To sum up, there are plenty of theoretical and empirical avenues to be explored further, and we leave them open for future work in these directions.

Acknowledgements. This paper extends some parts of the working papers CEPR #7018, CEP #894, and CIRPÉE #09-19, and omits some others. We thank Andrea Galeotti and three anonymous referees for very helpful comments and suggestions. Donald Davis, Klaus Desmet, Gilles Duranton, Wen-Tai Hsu, Wilfried Koch, Muriel Meunier, Yasusada Murata, Volker Nocke, Issi Romem, Esteban Rossi-Hansberg, Takaaki Takahashi, Jacques Thisse as well as participants at seminars and conferences at Erasmus (Rotterdam), LSE, INRS Montréal, LMU Munich, Nagoya, Passau, and Warwick provided valuable comments and suggestions. Behrens is the holder of the *Canada Research Chair in Regional Impacts of Globalization*. Financial support from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada is gratefully acknowledged. Behrens and Robert-Nicoud gratefully acknowledge financial support from FQRSC Québec (Grant NP-127178). Behrens further gratefully acknowledges financial support from SSHRC's Standard Research Grants Program. Part of this paper was written while Robert-Nicoud was visiting the IES at Princeton, which he thanks for its hospitality. We also thank the organisers of the 2nd HSE Summer School in Tsarskoe Seolo (Pushkin), Russia, where we completed the current version of the paper, for their hospitality. Finally, like Dixit and Grossman (1982), "we blame each other for all remaining errors" ("Trade and protection with multistage production", *Review of Economic Studies* 42, 583-594).

References

[1] Asplund, M., and V. Nocke (2006) Firm turnover in imperfectly competitive markets, *Review of Economic Studies* 73, 295–327.

[2] Bairoch, P. (1988) *Cities and Economic Development: From the Dawn of History to the Present*. Chicago, IL: Univ. of Chicago Press.

[3] Baldwin, R.E., and T. Okubo (2006) Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting, *Journal of Economic Geography* 6, 323–346.

[4] Baum-Snow, N., and R. Pavan (2012) Inequality and city size. *Review of Economics and Statistics*, forthcoming.

[5] Behrens, K., G. Duranton, and F.L. Robert-Nicoud (2012) Productive cities: Sorting, selection and agglomeration. Processed. Available at `http://individual.utoronto.ca/gilles/Papers/Sorting.pdf`.

[6] Behrens, K., G. Mion, Y. Murata, and J. Südekum (2012) Spatial frictions. CEPR Discussion Paper #8572.

[7] Bleakley, H., and J. Lin (2012) Portage and path dependence, *Quarterly Journal of Economics* 127, 587–644.

[8] Brülhart, M., C. Carrère, and F.L. Robert-Nicoud (2013) Trade and towns: On the uneven effects of trade liberalization. In progress.

[9] Brülhart, M., C. Carrère, and F. Trionfetti (2012) How wages and employment adjust to trade liberalization: Quasi-experimental evidence from Austria, *Journal of International Economics* 86, 68–81.

[10] Combes, P.-Ph., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012) The productivity advantages of large cities: Distinguishing agglomeration from firm selection, *Econometrica* 80(6), 2543–2594.

[11] Davis, D.R., and J.I. Dingle (2012) A spatial knowledge economy, NBER Working Paper #18188.

[12] Dobkins, L.H., and Y. Ioannides (2000) Dynamic evolution of the size distribution of US cities. In: Huriot, J.-M. and J.-F. Thisse (eds.) *Economics of Cities: Theoretical Perspectives*. Cambridge, MA: Cambridge University Press, pp. 217–260.

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

[14] Duranton, G., and D. Puga (2001) Nursery Cities: Urban diversity, process innovation, and the life-cycle of product, *American Economic Review* 91, 1454–1477.

[15] Duranton, G., P.M. Morrow, and M.A. Turner (2012) Roads and trade. *Processed*, Univ. of Toronto and Univ. of Pennsylvania.

[16] Duranton, G., and M.A. Turner (2012) Urban growth and transportation, *Review of Economic Studies* 79(4), 1407–1440.

[17] Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2010) Spatial sorting: Why New York, Los Angeles and Detroit attract the greatest minds as well as the unskilled. CEPR Discussion Paper #8151.

[18] Egger, H., and U. Kreickemeier (2009) Firm heterogeneity and the labor market effects of trade liberalization, *International Economic Review* 50, 187–216.

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

[20] Fujita, M., P.R. Krugman, and A.J. Venables (1999) *The Spatial Economy. Cities, Regions and International Trade.* Cambridge, MA: MIT Press.

[21] Glaeser, E.L., M.E. Kahn, and J. Rappaport (2008) Why do the poor live in cities? The role of public transportation, *Journal of Urban Economics* 63, 1–24.

[22] Glaeser, E.L., J. Kolko and A. Saiz (2001) Consumer city, *Journal of Economic Geography* 1, 27–50.

[23] Glaeser, E.L., and D.C. Maré (2001) Cities and skills, *Journal of Labor Economics* 19, 316–342.

[24] Glaeser, E.L., M. Resseger, and K. Tobio (2009) Inequality in cities, *Journal of Regional Science* 49, 617–646.

[25] Handbury, J., and D.E. Weinstein (2012) Is New Economic Geography right? Evidence from price data. *Processed*, Univ. of Pennsylvania and Columbia Univ.

[26] Helpman, E., M.J. Melitz, and S.R. Yeaple (2004) Export versus FDI with heterogeneous firms, *American Economic Review* 94, 300–316.

[27] Henderson, J.V. (1974) The size and types of cities, *American Economic Review* 64, 640–656.

[28] Lee, S. (2010) Ability sorting and consumer city, *Journal of Urban Economics* 68, 20–33.

[29] Lucas, R.E. (1978) On the size distribution of business firms, *Bell Journal of Economics* 9, 508–523.

[30] Melitz, M.J., and G.I.P. Ottaviano (2008) Market size, trade, and productivity, *Review of Economic Studies* 75, 295–316.

[31] Michelacci, C., and O. Silva (2007) Why so many local entrepreneurs? *Review of Economics and Statistics* 89, 615–633.

[32] Moretti, E. (2013) Real wage inequality, *American Economic Journal: Applied Economics* 5(1), 65–103.

[33] Ottaviano, G.I.P., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and trade revisited, *International Economic Review* 43, 409–436.

[34] Ottaviano, G.I.P (2012) Agglomeration, trade, and selection, *Regional Science and Urban Economics* 42, 987–997.

[35] Partridge, M.D., D.S. Rickman, K. Ali, and M.R. Olfert (2009) Do New Economic Geography agglomeration shadows underlie current population dynamics across the urban hierarchy?, *Papers in Regional Science* 88, 445–466.

[36] Peng, S., J.-F. Thisse, and P. Wang (2006) Economic integration and agglomeration in a middle product economy, *Journal of Economic Theory* 131, 1–25.

[37] Piketty, T., and E. Saez (2003) Income inequality in the United States, 1913-1998, *Quarterly Journal of Economics* 118, 1–39.

[38] Redding, S.J., and D. Sturm (2008) The costs of remoteness: Evidence from German division and reunification, *American Economic Review* 98, 1766–1797.

[39] Redding, S.J., and A.J. Venables (2004) Economic geography and international inequality, *Journal of International Economics* 62, 53–82.

[40] Rosenthal, S. and W. 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*, North-Holland: Elsevier, pp. 2713–2739.

[41] Syverson, C. (2004) Market structure and productivity: A concrete example, *Journal of Political Economy* 112, 1181–1222.

[42] Tabuchi, T., J.-F. Thisse, and D.-Z. Zeng (2005) On the number and size of cities, *Journal of Economic Geography* 5, 423–448.

[43] Wheeler, C.H. (2001) Search, sorting, and urban agglomeration, *Journal of Labor Economics* 19, 879–899.

Appendix material

This extensive Appendix is structured as follows. Appendix **A** contains the guide to various calculations, Appendices **B** and **C** contain various proofs. In Appendix **A.1**, we spell out the urban structure and derive the urban costs. In Appendix **A.2**, we derive the expression of the consumer surplus. In Appendix **A.3** we derive the price equilibrium of the model. In Appendix **A.4**, we establish the expression for expected profits of an urban entrepreneur. In all these appendices, we provide the proofs for the general case with mulitple cities, but their adaption to the single-city case is straightforward. In Appendices **B.1** and **B.2** we prove Propositions 1 and 2, respectively. Appendix **B.3** summarises the equilibrium structure with a single city and characterises all equilibria. Appendix **B.4** provides details on the computation of the Gini coefficient, and Appendix **B.5** contains the proof of Proposition 3. In Appendices **C.1** and **C.2**, we prove Propositions 4 and 5, respectively. Last, Appendix **C.3** contains various compuations for the Gini coefficient in the case of symmetric trading cities and derives the comparative static results.

A. Guide to calculations

A.1. Urban costs. Assume that all city dwellers consume one unit of land, as in standard fixed lot-size models (see Fujita, 1989). Assume further that working in cities and consuming differentiated goods requires urban dwellers to commute to the 'Central Business District' (CBD). This CBD is located at $x = 0$, so that a city of size H stretches out from $-H/2$ to $H/2$. Without loss of generality, we normalise the opportunity cost of land at the urban fringe to zero: $R(H/2) = R(-H/2) = 0$. Each city dweller commutes to the CBD at constant unit-distance cost $\xi > 0$. Hence, an agent located at x incurs a commuting cost of $\xi|x|$. Because expected profits and consumer surplus do not depend on city location (see Section 2), the sum of commuting costs and land rent must be identical across locations at a residential equilibrium. This implies that

$$\underbrace{R\left(\frac{H}{2}\right)}_{=0} + \xi\frac{H}{2} = R(x) + \xi|x|,$$

for all x, which yields the equilibrium land rent schedule $R(x) = \xi\left(H/2 - |x|\right)$. The aggregate land rent is thus given by

$$\text{ALR} = \int_{-\frac{H}{2}}^{\frac{H}{2}} R(x)\mathrm{d}x = \frac{\xi}{4}H^2.$$

When ALR is equally redistributed to all agents, equilibrium total urban costs are given by

$$-\frac{\text{ALR}}{H} + R(x) + \xi|x| = \frac{\xi}{4}H.$$

Letting $\theta \equiv \xi/4 > 0$ then yields the expression θH for urban costs.

A.2. Consumer surplus. Denote by $D \equiv \int_{\mathcal{V}} d(\nu)\mathrm{d}\nu$ the demand for all varieties of the differentiated good supplied in the city. As can be seen from equation (1), marginal utility at zero consumption is bounded for each variety. Hence, consumers need not have positive demand for all of them. The inverse demand for each variety ν of that good is obtained by maximising (1) subject to (2) and can be expressed as follows:

$$p(\nu) = \alpha - \gamma d(\nu) - \eta D \tag{A.1}$$

whenever $d(\nu) \geq 0$. Expression (A.1) can be inverted to yield a linear demand system as follows:

$$q(\nu) \equiv Hd(\nu) = H\left[\frac{\alpha}{\eta N + \gamma} - \frac{p(\nu)}{\gamma} + \frac{\eta N}{\eta N + \gamma}\frac{\overline{p}}{\gamma}\right], \quad \forall \nu \in \mathcal{V}, \tag{A.2}$$

where $\overline{p} \equiv (1/N)\int_{\mathcal{V}} p(\nu)\mathrm{d}\nu$ stands for the average price. By definition, $\mathcal{V}$ is the set of varieties satisfying

$$p(\nu) \leq \frac{\gamma\alpha + \eta N\overline{p}}{\eta N + \gamma} \equiv p^d. \tag{A.3}$$

For any given value of love for variety γ, lower average prices $\overline{p}$ or a larger number of competing varieties N increase the price elasticity of demand and decrease the price bound p^d defined in (A.3). Stated differently, a lower $\overline{p}$ or a larger N generate a 'tougher' competitive environment, thereby reducing the maximum price at which entrepreneurs still face positive demand. Letting $p_{hl}(\nu)$ stand for the price of variety ν produced in h and sold in l, and $\mathcal{V}_{hl}$ be the set of varieties produced in h and consumed in l, the consumer surplus is given by:

$$\begin{aligned} \text{CS}_l &= \frac{\alpha^2 N_l}{2(\eta N_l+\gamma)} - \frac{\alpha}{\eta N_l+\gamma}\sum_h \int_{\mathcal{V}_{hl}} p_{hl}(\nu)\mathrm{d}\nu \\ &\quad + \frac{1}{2\gamma}\sum_h \int_{\mathcal{V}_{hl}} p_{hl}^2(\nu)\mathrm{d}\nu - \frac{\eta}{2\gamma(\eta N_l+\gamma)}\left[\sum_h \int_{\mathcal{V}_{hl}} p_{hl}(\nu)\mathrm{d}\nu\right]^2. \end{aligned} \tag{A.4}$$

The expression with a single city is obtained by letting $\mathcal{V}_{hl}=\mathcal{V}$, $p_{hl}=p$, $N_l=N$, and by eliminating the sum across h.

A.3. Price equilibrium. Let $\pi_{hl}(c)=\left[p_{hl}(c)-\tau_{hl}c\right]q_{hl}(c)$ denote operating profits, expressed as a function of the entrepreneur's inverse productivity c. The firms sets prices in order to maximise these profits for each market separately. Then, the profit maximising prices and output levels must satisfy (for $h\neq l$, with $\tau_{hl}=1$ substituted for when $h=l$):

$$p_{hl}(c)=\frac{\gamma\alpha+\eta N_l\overline{p}_l}{2(\eta N_l+\gamma)}+\frac{\tau_{hl}c}{2} \quad \text{and} \quad q_{hl}(c)=\frac{H_l}{\gamma}\left[p_{hl}(c)-\tau_{hl}c\right]. \tag{A.5}$$

Integrating the prices in (A.5) over all available varieties, summing across regions and rearranging yields the average delivered price in market l as follows:

$$\overline{p}_l=\frac{\gamma\alpha+\eta N_l\overline{p}_l}{2(\eta N_l+\gamma)}+\frac{\overline{c}_l}{2} \quad \Rightarrow \quad \overline{p}_l=\frac{\gamma\alpha+(\gamma+\eta N_l)\overline{c}_l}{2\gamma+\eta N_l}, \tag{A.6}$$

where

$$\overline{c}_l\equiv\frac{1}{N_l}\sum_h \tau_{hl}\int_{\mathcal{V}_{lh}} c\,\mathrm{d}G(c)$$

stands for the average delivered cost of surviving firms selling to l. Plugging (A.6) into (A.5), some straightforward rearrangements show that the equilibrium prices can then be expressed as follows:

$$p_{hl}(c)=\frac{c_l+\tau_{hl}c}{2}, \quad \text{where} \quad c_l\equiv\frac{2\alpha\gamma+\eta N_l\overline{c}_l}{2\gamma+\eta N_l}$$

denotes the *domestic cost cutoff in region l*. Only entrepreneurs with c 'sufficiently smaller' than c_l are productive enough to sell in city l. This can be seen by expressing q_{hl} in (A.5) more compactly as follows:

$$q_{hl}(c)=H_l\frac{c_l-\tau_{hl}c}{2\gamma}. \tag{A.7}$$

Clearly, selling in a 'foreign' market l when producing in h requires that $c \leq c_l/\tau_{hl}$, whereas the analogous condition for selling in the 'domestic' market is given by $c \leq c_l$. In what follows, we denote by c_{hl} the *export cost cutoff for firms producing in region h and selling to region l*. This cutoff must satisfy the zero-profit cutoff condition $c_{hl} = \sup\{c \mid \pi_{hl}(c) > 0\}$. From expressions (3) and (A.7), this condition can be expressed as either $p_{hl}(c_{hl}) = \tau_{hl}c_{hl}$ or $q_{hl}(c_{hl}) = 0$, which then yields: $c_{hl} = c_l/\tau_{hl}$. Clearly, $c_{hl} \leq c_l$ since $\tau_{hl} \geq 1$. Put differently, trade barriers make it harder for exporters to break even relative to their local competitors because of higher market access costs. Since $p_l^d = p_{ll}(c_l) = c_l$, the zero-profit cutoff condition (A.3) can be expressed as follows:

$$\frac{\gamma\alpha + \eta N_l \overline{p}_l}{\eta N_l + \gamma} = c_l, \quad \text{with} \quad \overline{p}_l = \frac{\alpha\gamma + (\gamma + \eta N_l)\overline{c}_l}{2\gamma + \eta N_l}.$$

We can thus solve for the mass of entrepreneurs selling in region l as follows:

$$N_l = \frac{2\gamma}{\eta}\frac{\alpha - c_l}{c_l - \overline{c}_l}. \tag{A.8}$$

Using the Pareto parametrisation of Section 3.3, the average price and the average marginal cost in region l are computed as follows: $\overline{p}_l = \frac{2k+1}{2k+2}c_l$ and $\overline{c}_l = \frac{k}{k+1}c_l$, i.e., they are both proportional to the domestic cutoff. Using this expression, as well as (A.8), we can then express the mass of sellers in l as follows:

$$N_l \equiv \sum_h H_h G(c_{hl}) = \frac{2\gamma(k+1)(\alpha - c_l)}{\eta c_l},$$

where the first equality comes from the definition of N_l. The consumer surplus is finally derived by substituting the equilibrium prices into (A.4) given in Appendix A.2.

A.4. Expected profits and sales. The expected profit in region l in the symmetric case under the Pareto parametrisation is given as follows:

$$\mathbb{E}(\Pi_l) = \frac{1}{H_l}\left[\sum_h \frac{H_h}{4\gamma}\int_0^{\frac{c_h}{\tau_{lh}}} (c_h - \tau_{lh}c)^2 \; H_l \mathrm{d}G_l(c)\right] = \frac{c_{\max}^{-k}[\sum_h \tau_{lh}^{-k} H_h c_h^{k+2}]}{2\gamma(k+1)(k+2)} = A\frac{\sum_h \tau_{lh}^{-k} H_h c_h^{k+2}}{k+2}.$$

Using this expression, and noting that neither the consumer surplus nor the urban costs depend on the entrepreneur's ability, we readily obtain expression (9). By the same token, the per capita sales (R for 'revenue') are equal to

$$\mathbb{E}(R_l) = \frac{1}{H_l}\left[\sum_h \frac{H_h}{4\gamma}\int_0^{\frac{c_h}{\tau_{lh}}} (c_h^2 - (\tau_{lh}c)^2) \; H_l \mathrm{d}G_l(c)\right] = (k+1)\mathbb{E}(\Pi_l).$$

Appendix B. Proofs for Section 3

We prove all propositions of Section 3 for a single city, and all propositions in Section 4 for an arbitrary number Λ of *symmetric* cities, one per region. Since the model is perfectly symmetric by

assumption, an equilibrium where all regions have the same size $H_l \equiv H$ and the same cutoff c_l always exists. Let $\Phi \equiv (\Lambda - 1)\tau^{-k}$ denote the 'freeness' of trade. The single-city case corresponds to the situation where $\Phi = 0$ (since $\Lambda = 1$), which also applies when $\tau \to \infty$ (trade is prohibitive). More generally, Φ is increasing in Λ and decreasing in τ and takes value $\Phi = \Lambda - 1$ when $\tau = 1$ (trade is costless). The reader can readily verify that all the proofs in this appendix apply to the special case where $\Phi = 0$, as in Section 3; and to the more general case where $\Phi > 0$, as in Section 4. It turns out that Φ and A enter all expressions together as $(1 + \Phi)A$ so that all comparative static exercises pertaining to the effect of a change in A readily extend to the effects of changes in the freeness of trade.

In the symmetric case with trade and with Λ regions, the free-entry condition (9) in each region reduces to:

$$\frac{(1+\Phi)A}{k+2} H c_l^{k+2} + \frac{\alpha - c_l}{2\eta}\left[\alpha - \frac{k+1}{k+2} c_l\right] - f^E - \theta H \leq 0. \tag{B.1}$$

Likewise, the identity (7) becomes

$$H = \frac{1}{(1+\Phi)A\eta} \frac{\alpha - c_l}{c_l^{k+1}}. \tag{B.2}$$

Substituting (B.2) into (B.1), and rearranging, we then obtain (17).

B.1. Proof of Proposition 1. Rewriting $f(\cdot)$ in decreasing order of its powers in c_1, we obtain:

$$f(c_1) = K_1 c_1^2 - K_2 c_1 \pm K_3 + K_4 c_1^{-k} - K_5 c_1^{-k-1},$$

where all coefficients K_i are strictly positive. Note that the constant K_3 (which is associated with c_1 to the power 0) may a priori be positive or negative, hence the $\pm$ sign in front of it. As one can see, in all cases *there are at most three sign changes from positive to negative or vice versa* between the coefficients of the consecutive powers. Let the number of positive roots be n and the number of sign changes be s. By Laguerre's (1883) generalisation of *Descartes' rule of signs*, we know that $n \leq s$ (i.e., there are at most as many positive roots as sign changes) and $(s - n)$ is an even number if $n < s$. Hence, there are either 3 or 1 positive roots in our case. Applying Laguerre's generalisation of Descartes' rule to the first and second derivatives of $f(\cdot)$ reveals that f' changes sign at most twice and that f'' changes signs at most once. The final part of the proposition results from the fact that $f(\cdot)$ increases from $-\infty$ at $c_1 = 0$. Hence, $\partial f / \partial c_1$ must be strictly positive at the smallest root (whenever one exists). By continuity, and the changes in the signs of the derivatives when there are multiple roots, it follows that there at most two stable equilibria. To see that the third root of $f(\cdot)$ is outside the relevant range $[0, \alpha]$, since $c_1 > \alpha$ implies a negative city size which does not make any economic sense, it is sufficient to know that $f(\alpha) = -f^E$ and $\lim_{c_1 \to +\infty} f(c_1) = \lim_{c_1 \to +\infty} f'(c_1) = +\infty$. Thus, the largest root of $f(\cdot)$ is (strictly) larger than α if (and only if) the parameter f^E is (strictly) positive.

B.2. Proof of Proposition 2. Turning to the comparative static results, using (10), it is readily verified that, for any given value of c_1, $f(\cdot)$ is strictly increasing in α or A and decreasing in θ. Assume that $c_1^* \in (0, \alpha]$ is a stable equilibrium. Two cases may arise: either $f(c_1^*) \leq 0$ (with $c_1^* = \alpha$ and $H^* = 0$), which corresponds to the rural equilibrium; or $f(c_1^*) = 0$ with $0 < c_1^* < \alpha$ and $H^* > 0$ at an urban equilibrium. Consider first an increase in θ. Then $f(\cdot)$ shifts down everywhere, so it must be that $f(c_1^*) < 0$ in the first case: the rural equilibrium remains stable, and $H^* = 0$ is trivially non-increasing from its initial value. In the second case, $f(c_1^*) < 0$ after the shift. Since stability implies $\partial f(c_1^*)/\partial c_1 > 0$ and by continuity of $f(\cdot)$, the new equilibrium must lie to the right of the previous one, hence c_1^* increases and H^* falls by (7). Consider next an increase in α or A. A symmetric argument to the foregoing ensures that c_1^* falls. The overall effect of a rise of A or α on H^* now involves a direct effect, seen in (7), that reinforces the indirect effect in the case of α but that works in the opposite direction in the case of A. This is because a more productive differentiated goods sector requires fewer entrepreneurs to produce the same quantity of output, ceteris paribus. In turn, fewer urban dwellers make it less costly to live in cities, thus triggering urban entry. The net effect turns out to be unambiguous, which can be established by contradiction. Assume that $\mathrm{d}A > 0$ but that $\mathrm{d}H^* < 0$. From (10), $\mathrm{d}H^* < 0$ implies that $(\alpha - c_1)\left[\alpha - (k-1)c_1/(k+2)\right]$ must fall. It turns out that this term is decreasing in c_1 over $(0, \alpha]$, thus $\mathrm{d}H^* < 0$ implies $\mathrm{d}c_1^* > 0$ by (10). However, we have previously established that $\partial c_1^*/\partial A < 0$, a contradiction. Therefore, $\partial H^*/\partial A > 0$.

In addition, all stable symmetric equilibrium city sizes H^* are non-decreasing in trade freeness Φ by the same token (remember that in the symmetric trading cities model, $(1+\Phi)A$ replaces the term A in the one-city model). The rest of the proof is identical. □

B.3. Types of equilibria. This appendix characterises under what conditions which type of equilibrium emerges. Although we restrict ourselves to the case where $f^E > 0$ in the main text, we also discuss the limit case where $f^E = 0$ in this appendix.

Proposition 6 (existence and stability of the rural equilibrium) *(i) The rural equilibrium exists and is stable for any $f^E > 0$. (ii) If $\theta \geq \theta^R$ and $f^E \geq f^R$, where*

$$\theta^R \equiv \frac{3A\alpha^{k+2}}{2(k+2)} > 0 \quad \text{and} \quad f^R \equiv \frac{\alpha^2}{2\eta}\frac{k-1}{k+2} > 0, \tag{B.3}$$

then the rural equilibrium is the unique spatial equilibrium. (iii) If $f^E = 0$ then the rural equilibrium exists and is a stable spatial equilibrium if and only if $\theta \geq \theta^R$.

Proof. (i) Condition (7) implies that $H^* = 0$ if and only if $c_1^* = \alpha$. Plugging this result into (10) shows that this inequality holds for any $f^E > 0$. Local stability of the rural equilibrium then immediately follows from the strict inequality. It is useful to show (iii) next. If $f^E = 0$, local

stability of the rural equilibrium requires that $\partial f/\partial c_1 > 0$ when evaluated at $\{H^*, c_1^*\} = \{0, \alpha\}$. Using (10), some straightforward computations show that this is equivalent to $\theta > \theta^R$, where θ^R is defined in (B.3). This establishes the stability of the rural equilibrium. To show its existence and to derive a sufficient condition for it to be the only equilibrium, add and subtract (B.3) in (7) to obtain:

$$\begin{aligned} f(c_1) &= \frac{\alpha - c_1}{2\eta}\left[\alpha - \frac{k-1}{k+2}c_1 - \frac{3\alpha}{k+2}\left(\frac{\alpha}{c_1}\right)^{k+1} + 2\frac{\theta^R - \theta}{Ac_1^{k+1}}\right] - f^E \\ &< \frac{\alpha - c_1}{2\eta}\left[(\alpha - c_1)\frac{k-1}{k+2} + 2\frac{\theta^R - \theta}{Ac_1^{k+1}}\right] - f^E, \end{aligned} \tag{B.4}$$

where the inequality stems from $c_1 < \alpha$. Imposing $\theta \geq \theta^R$, we further have

$$\frac{\alpha - c_1}{2\eta}\left[(\alpha - c_1)\frac{k-1}{k+2} + 2\frac{\theta^R - \theta}{Ac_1^{k+1}}\right] - f^E < \frac{\alpha}{2\eta}\left[\alpha\frac{k-1}{k+2} - 2\frac{\theta - \theta^R}{A\alpha^{k+1}}\right] - f^E \leq f^R - f^E, \tag{B.5}$$

where the first inequality in (B.5) is due to $c_1 < \alpha$ and where the second inequality comes from $\theta \geq \theta^R$. Consequently, when the right-hand side of (B.5) is (weakly) negative, then $f(c_1; \cdot) < 0$ for all values of c_1. In that case, the rural equilibrium is the unique equilibrium. A sufficient condition for this to be so is $f^E \geq f^R$, where $f^R \equiv \frac{\alpha^2}{2\eta}\frac{k-1}{k+2}$. This establishes the result. ■

Proposition 7 (existence and stability of the urban equilibrium) *(i) If $f^E = 0$ and $\theta \in (0, \theta^R)$, then there exists a stable urban equilibrium. (ii) If $f^E > 0$, then there exists a θ, denoted as $\theta^U(f^E)$ with $\theta^U(f^E) < \theta^R$ and $\lim_{f^E \to 0} \theta^U(f^E) = \theta^R$, such that there is at most one stable urban equilibrium if $\theta \leq \theta^U(f^E)$. (iii) There exists $A_{\min}$ such that $H = 0$ for all $A < A_{\min}$.*

Proof. (i) Let $f^E \geq f^R$ and $\theta \geq \theta^R$; then the rural equilibrium $H^* = 0$ is the unique equilibrium. (ii) Let $f^E = 0$ and $\theta > \theta^R$; then $H^* = 0$ is the unique stable equilibrium. (iii) Let $f^E = 0$ and $\theta \in (0, \theta^R)$; then there exists a unique pair $\{H^*, c_1^*\}$ in $\mathbb{R}_{++} \times (0, \alpha)$ that constitutes a stable equilibrium (the urban equilibrium). (iv) Let $f^E > 0$; then there exists a θ, denoted as $\theta^U(f^E)$ with $\theta^U(f^E) < \theta^R$ and $\lim_{f^E \to 0} \theta^U(f^E) = \theta^R$, such that there is at most one pair $\{H^*, c_1^*\}$ in $\mathbb{R}_{++} \times (0, \alpha)$ that constitutes a stable equilibrium if $\theta \leq \theta^U(f^E)$.

Parts (i) and (ii) are a re-statement of Proposition 6. (iii) We are looking for a candidate equilibrium with $\alpha > c_1$. In this case, (10) is equivalent to

$$\frac{2\theta}{A} \geq c_1^{k+1}\left(\alpha - \frac{k-1}{k+2}c_1\right), \tag{B.6}$$

the right-hand side of which is strictly concave in c_1, increasing at the limit $c_1 \to 0$, and its maximum value on $(0, \alpha]$ is given by $3\alpha^{k+2}/(k+2)$. Therefore, the condition $\theta < \theta^R$ is also sufficient to ensure that there exists a pair $\{H^*, c_1^*\}$ with $c_1^* \in (0, \alpha)$ and $H^* = H(c_1^*)$ from (7) that

is compatible with an equilibrium. We finally invoke the continuity of $f(\cdot)$ to establish (iv): at the limit $f^E \to 0$, there exists a finite $\theta^U(f^E)$ by (ii) such that a stable urban equilibrium exists, with $\lim_{f^E \to 0} \theta^U(f^E) = \theta^R$. Since $f(\cdot)$ is continuously differentiable in both f^E and θ, it must be the case that $\theta^U(f^E)$ is positive in the neighbourhood of $f^E = 0$ and, by $\partial f/\partial f^E < 0$ and $\partial f/\partial \theta < 0$, that $\theta^U(f^E)$ is smaller than θ^R for any f^E. (iv) If $A > A_{\min} \equiv \theta \alpha^{k+2}/(k+2)$, then (5) holds with a strict inequality for any $c_1 \in [0, \alpha]$ and no urban equilibrium exists as a result. ■

To extend the foregoing proofs to the multi-city case of Section 5, it suffices to replace θ^R by θ^R_Φ and A by $(1+\Phi)A$ in the proof above.

B.4. Measures of earnings inequality. We first show how to get the equilibrium share of earnings of the Q^{th} quintile, in expression (14). Recall $\sigma_Q \equiv \overline{\Pi}_Q/\overline{\Pi}$, where the average profit of the top quantile Q is defined as (and equal to)

$$\overline{\Pi}_Q(H, c_1) \equiv \frac{1}{G(q)} \int_0^q \Pi(c) \mathrm{d}G(c) = k \frac{H}{4\gamma} \left(\frac{c_1^2}{k} - \frac{2c_1 q}{k+1} + \frac{q^2}{k+2} \right). \tag{B.7}$$

Dividing this expression by (13) yields (14) in the text.

We next derive the Gini coefficient of income inequality as given by (15). Since all agents with $c \geq c_1$ have zero income, aggregate income in city l across all draws c is given by

$$W_l(c_1) \equiv H_l G(c_1) \overline{\Pi}(H_l, c_1) = A \frac{H_l^2 c_1^{k+2}}{k+2},$$

where $\overline{\Pi}(H_l, c_1)$ is from (13). The total income accruing to agents with draw $q \leq c_1$ is thus given by

$$W_l(q) \equiv H_l G(c_1) \overline{\Pi}_q(H_l, c_1) = \frac{k H_l^2}{4\gamma} \left(\frac{q}{c_{\max}} \right)^k \left(\frac{c_1^2}{k} - \frac{2q}{k+1} + \frac{q^2}{k+2} \right),$$

where $\overline{\Pi}_q(H_l, c_1)$ is from (B.7), and their income share is $W_l(q)/W_l(c_1)$. To compute the Gini coefficient, we have to link the income share with the population share. To do so, we need to switch to the distribution in terms of population shares (and not in terms of cost levels c). Let $y \equiv (q/c_{\max})^k$, i.e., $q = y^{1/k} c_{\max}$. Using this change in variables, the new upper bound for integration is given by $y = (c_1/c_{\max})^k$, and we obtain the integral of the Lorenz curve for the surviving agents as follows:

$$\int_0^{\left(\frac{c_1}{c_{\max}}\right)^k} \frac{W_l(y)}{W_l(c_1)} \mathrm{d}y - \int_0^{\left(\frac{c_1}{c_{\max}}\right)^k} x \mathrm{d}x = \frac{2+7k}{4+8k} \left(\frac{c_1}{c_{\max}} \right)^k - \frac{1}{2} \left(\frac{c_1}{c_{\max}} \right)^{2k} \tag{B.8}$$

To finally obtain the Gini coefficient, we need to add the integral of the Lorenz curve for the agents who do not produce. This is given by

$$\int_{\left(\frac{c_1}{c_{\max}}\right)^k}^1 (1-x) \mathrm{d}x = \frac{1}{2} \left[\left(\frac{c_1}{c_{\max}} \right)^k - 1 \right]^2 \tag{B.9}$$

Summing (B.9) and (B.8) then yields the Gini coefficient of the income distribution as follows:

$$\text{Gini}(k, c_1) = 1 - \frac{k+2}{4k+2}\left(\frac{c_1}{c_{\max}}\right)^k. \tag{B.10}$$

B.5. Proof of Proposition 3. To establish the first part of the proposition, let us differentiate (B.7) with respect to c_1; we obtain

$$\frac{\partial \sigma_Q}{\partial c_1} = -\frac{k(k+1)(k+2)}{2c_1}\left(\frac{c_1}{c_{\max}}\right)^{-k}\left(1 - \frac{q}{c_1}\right)^2 < 0$$

for all $q < c_1$, and thus $\partial \sigma_q / \partial H > 0$ by $\partial H / \partial c_1 < 0$. □

Let us next turn to the Gini coefficient: (i) It can be readily verified that $\partial(\text{Gini})/\partial c_1 < 0$ and $\partial^2(\text{Gini})/\partial c_1^2 < 0$. (ii) $\partial(\text{Gini})/\partial H > 0$ readily follows from the monotonicity of (7) and (15). To obtain the concavity of Gini with respect to H, invert (15) to get an expression for c_1 as a function of Gini, and substitute this for c_1 into (7). Then, standard algebra reveals that $\partial^2 H / \partial(\text{Gini})^2 > 0$ and thus $\partial^2(\text{Gini})/\partial H^2 < 0$. (iii) Using (15) again, we obtain:

$$\frac{\partial(\text{Gini})}{\partial k}(k, c_1) = [1 - \text{Gini}(k, c_1)]\left[-\frac{3}{(k+2)(2k+1)} + \ln\left(\frac{c_1}{c_{\max}}\right)^k\right],$$

which is negative by inspection (recall that $c_1 < c_{\max}$). (iv) The last part of the proposition immediately follows by inspection of (15). □

Appendix C. Proofs for Section 4

C.1. Proof of Proposition 4. The conditions in equation (16) hold at any spatial equilibrium. Rewriting these conditions for the multi-city case with heterogeneous bilateral trade costs in matrix form yields

$$\underbrace{\begin{bmatrix} 1 & \phi_{12} & \dots & \phi_{1\Lambda} \\ \phi_{21} & 1 & \dots & \phi_{2\Lambda} \\ \vdots & & \ddots & \vdots \\ \phi_{\Lambda 1} & \phi_{\Lambda 2} & \dots & 1 \end{bmatrix}}_{\mathbf{F}^\phi} \underbrace{\begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_\Lambda \end{bmatrix}}_{\mathbf{h}} = \underbrace{\frac{1}{A\eta}\begin{bmatrix} (\alpha - c_1)c_1^{-(1+k)} \\ (\alpha - c_2)c_2^{-(1+k)} \\ \vdots \\ (\alpha - c_\Lambda)c_\Lambda^{-(1+k)} \end{bmatrix}}_{\mathbf{x}}, \tag{C.1}$$

where $\mathbf{F}^\phi$ is a Λ-dimensional square matrix and $\mathbf{h}$ and $\mathbf{x}$ are both Λ-dimensional vectors. Straightforward rearrangement of (16) for regions h and l in the symmetric case yields

$$\frac{\alpha - c_l}{\alpha - c_h}\left(\frac{c_h}{c_l}\right)^{1+k} = \frac{(1-\phi)H_l + \phi\sum_i H_i}{(1-\phi)H_h + \phi\sum_i H_i}$$

which implies that

$$c_l < c_h \iff H_l > H_h.$$

To get the second set of results, recall that the solution to the linear system $\mathbf{Fh} = \mathbf{x}$ is given by $\mathbf{h} = \det(\mathbf{F})^{-1}\text{cof}(\mathbf{F})^T\mathbf{x}$, where $\text{cof}(\mathbf{F})$ denotes the matrix of cofactors associated with $\mathbf{F}$. Hence

$$\frac{\partial H_l}{\partial c_l} = \frac{(-1)^{2l}\det(\mathbf{F}_{l,l})}{\det(\mathbf{F})}\frac{\partial \mathbf{x}_l}{\partial c_l},$$

where $\det(\mathbf{F}_{l,l})$ is the minor of the $(\Lambda-1)\times(\Lambda-1)$ square matrix cut down from $\mathbf{F}$ by removing its l^{th} column and its l^{th} row. The matrix $\mathbf{F}_{l,l}$, like $\mathbf{F}$, has only 1's on its main diagonal and ϕ off its main diagonal. Thus, its determinant is also positive, i.e. $\det(\mathbf{F}_{l,l}) = (1-\phi)^{\Lambda-2}[1+(\Lambda-2)\phi] > 0$. Since $\partial \mathbf{x}_l/\partial c_l < 0$, the result follows. By the same token,

$$\frac{\partial H_l}{\partial c_h} = \frac{(-1)^{h+l}\det(\mathbf{F}_{h,l})}{\det(\mathbf{F})}\frac{\partial \mathbf{x}_h}{\partial c_h}.$$

One can verify that for any row h and any $m > 1$, $\mathbf{F}_{h,l}$ and $\mathbf{F}_{h,l+m}$ (with $l \neq h$ and $l+m \neq h$) differ by exactly m row and/or column permutations. Hence, we know from the permutation properties of the determinants that $\det(\mathbf{F}_{h,l}) = (-1)^m\det(\mathbf{F}_{h,l+m})$, which implies that $(-1)^{h+l}\det(\mathbf{F}_{h,l}) = (-1)^{h+l+m}\det(\mathbf{F}_{h,l+m})$. In words, $\partial H_l/\partial c_h$ has a constant sign for all $h \neq l$. To obtain that sign, we can proceed as follows. Take an arbitrary row h, and choose an adjacent column to the diagonal. Form the minor of $\mathbf{F}_{h,h-1}$ or $\mathbf{F}_{h,h+1}$. This has the same structure than $\mathbf{F}_{h,h}$, except for one term on the diagonal that has been replaced by ϕ. Now perform a column expansion on that column to see that $\det(\mathbf{F}_{h,h}) = \det(\mathbf{F}_{h,l}) + (1-\phi)\det(\widetilde{\mathbf{F}})$, where $\widetilde{\mathbf{F}}$ has the same structure than F but is a $\Lambda-2$ square matrix. Using the determinants, we then obtain that $\det(\mathbf{F}_{h,h-1}) = \det(\mathbf{F}_{h,h+1}) = \phi(1-\phi)^{\Lambda-2}$. This then shows that

$$\frac{\partial H_l}{\partial c_{l+1}} = \underbrace{\frac{(-1)^{2l+1}\phi(1-\phi)^{\Lambda-2}}{\det(\mathbf{F})}}_{<0}\frac{\partial \mathbf{x}_h}{\partial c_h}.$$

which, together with $\partial \mathbf{x}_h/\partial c_h < 0$ and the invariance of the sign of $\partial H_l/\partial c_h$ for $h \neq l$, completes the proof.

To obtain the limit result for large values of k, observe that the ith row-sum of non-diagonal elements of $\mathbf{F}^\phi$ is given by $\sum_{j\neq i} d_{ij}^{-\gamma k}$. Clearly, this sum limits zero as k gets large, so that $\mathbf{F}^\phi$ is diagonal dominant for k large enough. Since $\mathbf{F}^\phi$ is symmetric and positive, diagonal dominance then implies that the matrix is positive definite. Hence $\det(\mathbf{F}^\phi) > 0$. Furthermore, all minors $\mathbf{F}^\phi_{l,l}$ on the main diagonal are positive as they are also associated with a positive diagonal-dominant and symmetric (i.e., positive definite) sub-matrix. This establishes the result. □

C.2. Proof of Proposition 5. We first show that a smaller value of Φ makes the rural equilibrium more likely to occur. Note that when $f^E = 0$, local stability of the rural equilibrium requires that $\partial f/\partial c_l > 0$ when evaluated at $\{H^*, c_l^*\} = \{0, \alpha\}$. This is equivalent to $\theta > \theta_\Phi^R$, where θ_Φ^R is given by

$$\theta_\Phi^R \equiv (1+\Phi)3A\frac{\alpha^{k+2}}{2(k+2)} = (1+\Phi)\theta^R.$$

As in the single-city case, the rural equilibrium exists and is stable for all $\theta \geq \theta_\Phi^R$, whereas the urban equilibrium is the unique stable equilibrium when $\theta < \theta_\Phi^R$. Clearly, θ_Φ^R is increasing in Φ (i.e., with freer trade), which proves our claim. Turn next to the case where $f^E > 0$. We have already established in the proof of Proposition 7 that $f(\cdot)$ is continuously decreasing in both f^E and θ, which implies that the equilibrium city size is decreasing in f^E and increasing in Φ at the stable urban equilibrium. □

C.3. Gini coefficient and comparative statics. By inspection of (18), Gini_l is decreasing in both c_l and $\lambda(\cdot)$. Thus, to establish the result, it is sufficient to show that $\partial\lambda/\partial\tau < 0$. Let $z(\Lambda, \tau, k) \equiv -\lambda(\Lambda, \tau, k)/2$ so that (18) may be rewritten as

$$\text{Gini}_l(\Lambda, \tau, k; c_l) = 1 + 2z(\Lambda, \tau, k)\left(\frac{c_l}{c_{\max}}\right)^k,$$

with $z(\cdot) < 0$ for all Λ, τ and k. Fastidious calculations similar to those leading to (B.8) in appendix C.6 yield

$$\begin{aligned} z(\Lambda,\tau,k) &= -1 + \frac{\phi}{2(1+2k)}\frac{(\Lambda-1)\left[(\tau-1)^2(1+2k)(2+k)(1+k)+2(\tau-1)(2+k)(1+3k)+2+7k\right]}{2\tau^2+(\Lambda-1)\left[(\tau-1)^2(2+k)(1+k)+2(\tau-1)(2+k)+2\right]} \\ &\quad + \frac{1}{2(1+2k)}\frac{(2+7k)\tau^2}{2\tau^2+(\Lambda-1)\left[(\tau-1)^2(2+k)(1+k)+2(\tau-1)(2+k)+2\right]} \end{aligned}$$

from which it follows that $-2z(1,\tau,k) = (2+k)/(2+4k)$ and that $-2z(\Lambda,1,k) = (2+k)/(2+4k)$. We are now equipped to prove the result. Differentiating $z(\Lambda,\tau,k)$ with respect to τ yields:

$$\begin{aligned} \frac{\partial z(\Lambda,\tau,k)}{\partial\tau} &= -\frac{k(\Lambda-1)}{1+2k}\left\{\frac{\phi\left[\tau^2\kappa_2+\tau\kappa_1+\kappa_0\right]}{\left\{(\Lambda-1)\left[\tau^2(1+k)(2+k)-\tau(2+k)2k+(1+k)k\right]+2\tau^2\right\}^2}\right. \\ &\quad + \frac{\tau(2+7k)\left[(\tau-1)(2+k)+1\right]}{\left\{(\Lambda-1)\left[\tau^2(1+k)(2+k)-\tau(2+k)2k+(1+k)k\right]+2\tau^2\right\}^2} \\ &\quad \left. - \frac{\partial\phi}{\partial\tau}\frac{1}{k}\frac{(\tau-1)^2(1+k)(2+k)(1+2k)+2(\tau-1)(2+k)(3k+1)+(2+7k)}{(\Lambda-1)\left[\tau^2(1+k)(2+k)-\tau(2+k)2k+(1+k)k\right]+2\tau^2}\right\} \end{aligned}$$

where $\kappa_2 \equiv (\Lambda-1)(1+k)(2+k)^2 - 4k(2+k)$, $\kappa_1 \equiv 3(\Lambda-1)(1+k)(2+k) - 2k(7+2k)$ and $\kappa_0 \equiv (\Lambda-1)(2+k) - 6k$ all have ambiguous signs; therefore, the term in the first line of the right-hand side above cannot be signed a priori. By contrast, the terms on the second and third

lines are positive by inspection. However, if $\phi\left[\tau^2\kappa_2+\tau\kappa_1+\kappa_0\right]$ is negative, then it is larger than $\tau^2\kappa_2+\tau\kappa_1+\kappa_0$ and, adding the terms of the first and second lines, implies that

$$\begin{aligned}
&\phi\left[\tau^2\kappa_2+\tau\kappa_1+\kappa_0\right]+\tau(2+7k)\left[(\tau-1)(2+k)+1\right]\\
&\quad>\tau^2\kappa_2+\tau\kappa_1+\kappa_0+\tau(2+7k)\left[(\tau-1)(2+k)+1\right]\\
&\quad=(2+k)(\Lambda-1)\left[(1+k)(2+k)(\tau-1)^2+3(1+k)(\tau-1)+1\right]\\
&\qquad+(2+k)\left[(2+3k)(\tau-1)^2+3(1+k)(\tau-1)+1\right]>0
\end{aligned}$$

which in turn implies that $\partial z(\Lambda,\tau,k)/\partial\tau<0$ for all Λ, τ and k. We have already established in Proposition 5 that selection gets tougher as trade gets freer ($\partial c_l/\partial\tau>0$), therefore $\partial(\text{Gini}_l)/\partial\tau\equiv 2\left[c_l(\cdot)/c_{\max}\right]^k\left\{\partial z(\cdot)/\partial\tau+z(\cdot)c_l^{-1}\partial c_l(\cdot)/\partial\tau\right\}<0$. □

For the sake of completeness, note that

$$\begin{aligned}
\frac{\partial z(\Lambda,\tau,k)}{\partial\Lambda} &= -\frac{1}{1+2k}\left\{\frac{-\tau^2\phi\left[(1+k)(2+k)(1+2k)(\tau-1)^2+2(2+k)(1+3k)(\tau-1)+2+7k\right]}{\left\{(\Lambda-1)\left[\tau^2(1+k)(2+k)-\tau(2+k)2k+(1+k)k\right]+2\tau^2\right\}^2}\right.\\
&\quad\left.+\frac{(2+7k)\tau^2\left[(1+k)(2+k)(\tau-1)^2+2(2+k)(\tau-1)+2\right]}{2\left\{(\Lambda-1)\left[\tau^2(1+k)(2+k)-\tau(2+k)2k+(1+k)k\right]+2\tau^2\right\}^2}\right\}\\
&< -\frac{k}{1+2k}\frac{\tau^2(\tau-1)(2+k)\left[3(1+k)(\tau-1)+2\right]}{2\left\{(\Lambda-1)\left[\tau^2(1+k)(2+k)-\tau(2+k)2k+(1+k)k\right]+2\tau^2\right\}^2}<0.
\end{aligned}$$

Therefore, *given* c_l, granting access to more urban markets increases wages of the less productive exporters relative to the wages of the most productive ones; this positive effect is strong enough to overcome the negative one on income inequality that arises as a result of the wages of all successful entrepreneurs going up. However, since selection gets tougher as trade gets freer ($\partial c_l/\partial\tau>0$), the two effects work in opposite directions. Our numerical simulations suggest that the latter indirect effect always dominates the former, direct effect. More precisely, the fact that a larger Λ increases the Gini coefficient *is entirely due to the increase in selection*. By contrast, the fact that a lower τ increases the Gini is due to the increase in selection *and* to the increase of the profits of the most productive entrepreneurs relative to those of the least productive entrepreneurs.

References

[1] Fujita, M. (1989) *Urban Economic Theory: Land Use and City Size*. Cambridge, MA: Cambridge Univ. Press.

[2] Laguerre, E.N. (1883) Mémoire sur la théorie des équations numériques, *Journal de Mathématiques Pures et Appliquées* 3, 99-146. Translated into English by S.A. Levin (2002), On the theory of numeric equations, Stanford University.

Supplemental Appendix with Extra Material for "Survival of the Fittest in Cities"

A. Theoretical extensions

A.1. Return migration and inequality. Assume that, upon learning their inverse ability c, entrepreneurs who fail to be successful may return to the countryside at no cost. Assume further that all agents know this piece of information and include it in their entry decision. Starting from the equilibrium conditions of the benchmark model, the mass of people staying in the city is now $G(c_1)H$ with $G(c_1) = (c_1/c_{\max})^k$. Let $\widetilde{A} \equiv 1/[2(k+1)\gamma]$. The market clearing condition, which characterises the mass of varieties actually supplied to consumers, may be rewritten as:

$$\frac{\alpha - c_1}{\widetilde{A}\eta c_1} = G(c_1)H.$$

Next, the profit is given by $\Pi(c) = G(c_1)\frac{H}{4\gamma}(c_1 - c)^2$, so that the average profits of the stayers may be written as

$$\widetilde{\Pi} = \int_0^{c_1} \frac{\alpha - c_1}{4\gamma c_1 \widetilde{A}\eta}(c_1 - c)^2 k \frac{c^{k-1}}{{c_1}^k}\mathrm{d}c = \frac{1}{\eta(2+k)}(\alpha - c_1)c_1. \tag{A.1}$$

Note that (A.1) is positive and concave, as well as decreasing in c_1 over $(\alpha/2, \alpha)$. Furthermore, it is readily verified that

$$\begin{aligned}\widetilde{\Pi}(q) &\equiv \int_0^q \frac{\alpha - c_1}{4\gamma c_1 \widetilde{A}\eta}(c_1 - c)^2 k \frac{c^{k-1}}{c_1^k}\mathrm{d}c \\ &= \frac{(\alpha - c_1)k(k+1)}{2\eta}\left[\frac{{c_1}^{-k+1}q^k}{k} - \frac{2{c_1}^{-k}q^{k+1}}{k+1} + \frac{c_1^{-k-1}q^{k+2}}{k+2}\right],\end{aligned}$$

so that the share of profits accruing to entrepreneurs with a draw smaller than q is given by

$$\sigma(q) \equiv \frac{\widetilde{\Pi}(q)}{\widetilde{\Pi}} = \frac{k(k+1)(k+2)}{2}\left[\frac{{c_1}^{-k}q^k}{k} - \frac{2{c_1}^{-k-1}q^{k+1}}{k+1} + \frac{c_1^{-k-2}q^{k+2}}{k+2}\right],$$

which depends on the inverse average productivity c_1. For any given q, the income share is larger in larger cities (smaller c_1). The Gini coefficient can then finally be computed as follows:

$$\text{Gini} = 1 - 2\left[1 - \int_0^{c_l} \sigma(q)k\frac{q^{k-1}}{{c_1}^k}\mathrm{d}q\right] = \frac{3k}{4k+2}, \tag{A.2}$$

which is independent of city size and solely depends on the distributional parameter $k \geq 1$, despite the fact that $\sigma(q)$ is a function of c_1. Thus, the model with return migration delivers the counterfactual prediction that city size does not matter for income inequality.

A.2. Costly access to urban diversity in the countryside. Assume, contrary to what we did in the main body of the paper, that each region l has its *own countryside* (and not a 'common pool'). Consumers in the countryside associated with city l , henceforth denoted by l_0, can access all goods that are available in the city, but at a higher cost. More precisely, if residents in city l can access goods from city h at trade cost τ_{hl}, rural residents have to pay $\tau_{hl_0} = \xi\tau_{hl}$, with $\xi > 1$. We assume that ξ is common to all countrysides, but this assumption is immaterial for our analysis and we could easily relax it.

Let us subscript all expressions for the countryside l_0 by 0. It can readily be verified using appendices A.2 and A.3 that all expressions remain basically unchanged. In particular, the consumer surplus in the countryside is given by:

$$\mathrm{CS}_{l_0} \equiv \mathrm{CS}(c_{l_0}) = \frac{\alpha - c_{l_0}}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_0\right). \tag{A.3}$$

The number of sellers in countryside l_0 must satisfy

$$\frac{\alpha - c_{l_0}}{A_0\eta c_{l_0}^{k+1}} \equiv \sum_h \tau_{hl}^{-k} H_h, \tag{A.4}$$

where $A_0 = \xi^{-k}A < A$. The *indirect utility differential* between remaining in the countryside or moving to city l, given by

$$\Delta V_l(c) \equiv \Pi_l(c) + (\mathrm{CS}_l - \mathrm{CS}_{l_0}) - \theta H_l - f^E, \tag{A.5}$$

is obviously always smaller than when access to urban goods is prohibitive in the countryside. When urban goods are available in the countryside, cities will grow less strongly since the urban consumption premium decreases. This is one explanation for urban giants in the Third World – access to all sorts of goods and services that are just inexistent outside of cities.

Let $c_{l_0} \equiv c_0$ for simplicity. In the case of a single city, the free entry condition reduces to

$$\begin{aligned} f\left(H, c_1, c_0\right) \equiv\ & \frac{1}{k+2}\left[AHc_1^{k+2} + A_0(\overline{L} - H)c_0^{k+2}\right] + \frac{\alpha - c_1}{2\eta}\left[\alpha - \frac{k+1}{k+2}c_1\right] \\ & - \frac{\alpha - c_0}{2\eta}\left[\alpha - \frac{k+1}{k+2}c_0\right] - \theta H - f^E \leq 0 \end{aligned} \tag{A.6}$$

with the complementary slackness condition $Hf\left(H, c_1, c_0\right) = 0$. The first term (expected profits) is strictly increasing with H since access to urban consumers increases entrepreneurs' profits. The second term is always positive, but less so the smaller is ξ (in the limit, when $c_0 \to \alpha$, it boils down to the corresponding expression in the main body of the paper; or it vanishes if $\xi = 1$, in which case only the profits/urban costs tradeoff matters).

Turning to condition (A.4), it can be solved for H as follows in the countryside:

$$H = \frac{\alpha - c_0}{A_0\eta c_0^{k+1}}, \tag{A.7}$$

and for H as follows in the city:

$$H = \frac{\alpha - c_1}{A\eta c_1^{k+1}}. \tag{A.8}$$

The foregoing conditions (A.7) and (A.8) reveal that $c_0 > c_1$ for all $\xi > 1$, i.e., the consumer surplus is lower in the countryside. Clearly, c_0 can be expressed as a function of the city cutoff, $c_0 = f(c_1)$. We thus have

$$\begin{aligned} f(H, c_1, c_0) \equiv & \frac{1}{k+2}\left[AHc_1^{k+2} + A_0(\overline{L} - H)f(c_1)^{k+2}\right] + \frac{\alpha - c_1}{2\eta}\left[\alpha - \frac{k+1}{k+2}c_1\right] \\ & - \frac{\alpha - f(c_1)}{2\eta}\left[\alpha - \frac{k+1}{k+2}f(c_1)\right] - \theta H - f^E \leq 0 \end{aligned} \tag{A.9}$$

which we need to examine – combined with (A.8) – to determine the cutoffs c_1 and the city size H. The remaining variables (rural population and rural cutoff) can then be readily retrieved.

Let us look at the equilibrium structure numerically. A preliminary investigation confirms a few results. First, as shown above, $c_0 > c_1$ (which is obvious). Second, the structure of equilibria seems to be the same as the one in the main body of the paper. There is *always* a rural equilibrium (obvious), and at most one stable urban equilibrium. We can depict an example as follows in the two panels of Figure 7 below. The dashed curve is the loci of (c_1, c_0) such that $f(H, c_1, c_0) = 0$, where we have replaced H by its expression in (A.8). All points below that curve (to the south-west) are such that $f < 0$, whereas all points above that curve (to the north-east) are such that $f > 0$. The solid curve depicts the loci of (c_1, c_0) that satisfy (A.7) and (A.8). Clearly, that loci lies above the 45 degree line for all admissible couples (i.e., such that $c_1 \leq \alpha$ and $c_0 \leq \alpha$). Actually, in the example below, $\alpha = 10$ and the solid bold loci must cut the 45 degree line at $c_1 = \alpha$ and $c_0 = \alpha$. Below those values, it is always above the 45 degree line, i.e., $c_0 > c_1$ as it must be.

In the left panel of Figure 7, there is only a rural equilibrium. This is easy to see since the bold locus lies always in the zone where $f < 0$. In words, for all values of $c_1 < \alpha$ and $c_0 < \alpha$, people want to stay in the countryside. Hence, the equilibrium is such that $f(\alpha, \alpha) < 0$ and $H = 0$ (with $\overline{L}$ people remaining in the countryside). The left panel of Figure 7 is drawn for a high value of the fixed entry costs $f^E = 100$.

Now keep all parameters unchanged and decrease the fixed entry costs f^E to 10. In that case, as can be seen from the right panel of Figure 7, the dashed locus shifts down, whereas the solid locus remains the same. We now have an intersection between the two loci, i.e., a point where $f = 0$ and (A.7) and (A.8) hold. At that point, c_1 and c_0 are such that agents are indifferent at the margin between staying in the countryside or being in the city. Moving up the solid locus raises c_1 (thus shrinking the city) and yields $f > 0$: hence, if people would leave the city, they would be willing to move back. Moving down the locus decreases c_1 (thus making the city bigger) and yields $f < 0$: hence, if people would enter the city, they would be worse off and willing to move back. The intersection between the solid and the dashed loci thus yields the unique stable urban equilibrium. The rural equilibrium at (α, α) is obviously unstable. Observe that decreasing f^E

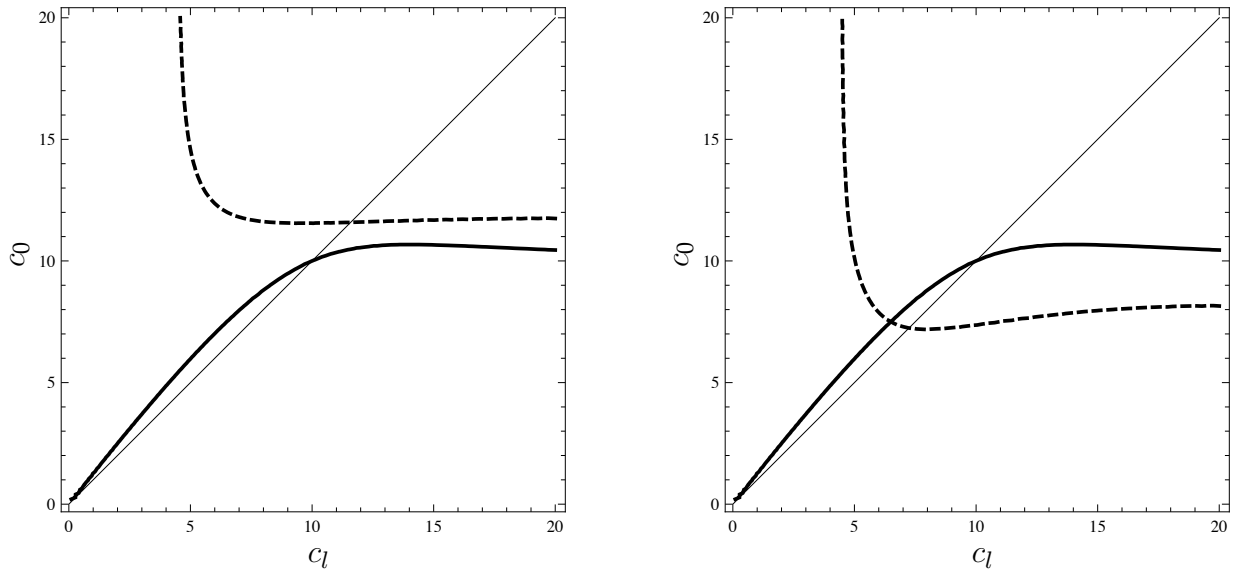

Figure 7: Only a rural equilibrium exists (left panel); urban equilibrium exists (right panel)

shifts down the dashed locus, which corresponds to smaller values of c_1 and c_0 and, therefore, larger equilibrium city sizes H_l.

B. Data for empirics and numerical illustrations

Our data on MSA sizes, average hourly wages, mean wages by income quintiles, aggregate rent, and income inequality comes from the US Census Bureau's American Community Survey 2007. The data on employment comes from the BLS, whereas metropolitan GDP comes from the BEA. The geographical data comes from the 2000 US Census Gazetteer county geography file. We aggregate up to the MSA-level using the county-to-MSA concordance tables for 2007. The geographical coordinates of an MSA are county-population weighted average centroids of the counties in the MSA. The MSA surface area – land surface only – is obtained from the same data source as the sum across constituent counties.

Following Corcos, Del Gatto, Mion, and Ottaviano (2012) and Behrens, Mion, Murata, and Südekum (2012), and using the properties of the Pareto distribution, the cutoffs are computed as follows:

$$c_l = \frac{k+1}{k}\frac{1}{\text{gdpc}_l},$$

where gdpc_l is GDP per employee in MSA l. Following Redding and Venables (2004), internal trade costs in a city are approximated by: : $d_{ii} = (2/3)\sqrt{\text{surface}_i/\pi}$. The numerical procedure we use to simulate the model is then as follows:

1. We back out the unobservable A_l terms from the identity (7) as follows:

$$\widehat{A}_l = \frac{\alpha - c_l}{\eta c_l^{k+2}} \frac{1}{\sum_h d_{hl}^{-k\delta} H_h}. \tag{B.1}$$

2. Using the values of $\widehat{A}_l$, we compute the corresponding upper bounds:

$$\widehat{c}_{l,\max} = \left[\frac{1}{2\widehat{A}_l (k+1)\gamma} \right]^{1/k}. \tag{B.2}$$

   Observe that since there are trade costs within each city, the 'domestic cutoff' is not given by c_l but by c_l/τ_{ll}. We makes sure in the application that $c_l/\tau_{ll} < \widehat{c}_{l,\max}$ for all l.

3. We then use the free entry conditions to get the unobservable $\widehat{\theta}_l$ terms:

$$\frac{\widehat{A}_l}{k+2} \sum_h \delta_{lh}^{-\delta k} H_h c_h^{k+2} + \frac{\alpha - c_l}{2\eta} \left[\alpha - \frac{k+1}{k+2} c_l \right] - \widehat{\theta}_l H_l - f_E = 0. \tag{B.3}$$

   We make sure in the application that $\widehat{\theta}_l$ is positive for all l.

4. Finally, we can run numerical illustrations. To this end, we proceed as follows. First, we either reduce δ – the distance elasticity of trade costs – by 10%, or increase it by 5%, or we reduce the distance d_{hl} between New York and Chicago by 50%. Then, for each case we solve the system of $2 \times K$ equations given by (B.1) and (B.3) for the $2 \times K$ unknowns H_h and c_h that would be observed in the new equilibrium. The upper bounds, $\widehat{c}_{l,\max}$, the commuting costs, $\widehat{\theta}_l$, and the parameters $(\alpha, \eta, \gamma, k, f_E)$ are all held constant in those exercises.

We compute the Gini index for all cities as follows. Take an arbitrary city ℓ. Define the accessibility of destination cities from ℓ as $c_j/\tau_{\ell j}$, rank destination cities from the most accessible to the least accessible (assuming ties away without loss of generality so as to simplify notation) and drop the origin subscript ℓ from $\tau_{\ell h}$ (where h is the destination city) for simplicity so that

$$c_1/\tau_1 > c_2/\tau_2 > ... > c_N/\tau_N,$$

with $N = 356$ in our case. Obviously, if firm c is more productive than firm z, then c serves at least as many markets as z. We then consider $N + 1 = 357$ firm categories, with firms in category n serving n markets, $n \in \{0, .., N\}$.

The aggregate earnings of all categories taken together are equal to

$$\Pi_\ell \equiv \frac{c_{\ell,\max}^{-k}}{2\gamma(1+k)(2+k)} \sum_{h=1}^{N} \tau_h^{-k} H_h c_h^{2+k}. \tag{B.4}$$

Now, consider the earnings of firms serving exactly one market (the most accessible of all), i.e. market 1. These are equal to

$$\Pi_\ell^1 \equiv \frac{1}{4\gamma}\int_{c_2/\tau_2}^{c_1/\tau_1} H_1\left(c_1 - c\tau_1\right)^2 \mathrm{d}G(c)$$

By this logic, it follows that the aggregate earnings of firms serving markets 1 to n (that is, the earnings of firms serving the n most accessible markets) are equal to

$$\begin{aligned}
\Pi_\ell^{1,...,n} &\equiv \frac{1}{4\gamma}\sum_{h=1}^{n} H_h \int_{c_{n+1}/\tau_{n+1}}^{c_h/\tau_h} \left(c_h - c\tau_h\right)^2 \mathrm{d}G(c) \\
&= \frac{k}{4\gamma c_{\ell,max}^k}\sum_{h=1}^{n} H_h \frac{c_h^2}{k}\left[\left(\frac{c_h}{\tau_h}\right)^k - \left(\frac{c_{n+1}}{\tau_{n+1}}\right)^k\right] \\
&- \frac{k}{4\gamma c_{\ell,max}^k}\sum_{h=1}^{n} H_h \frac{2c_h\tau_h}{1+k}\left[\left(\frac{c_h}{\tau_h}\right)^{1+k} - \left(\frac{c_{n+1}}{\tau_{n+1}}\right)^{1+k}\right] \\
&+ \frac{k}{4\gamma c_{\ell,max}^k}\sum_{h=1}^{n} H_h \frac{\tau_h^2}{2+k}\left[\left(\frac{c_h}{\tau_h}\right)^{2+k} - \left(\frac{c_{n+1}}{\tau_{n+1}}\right)^{2+k}\right], \qquad \text{(B.5)}
\end{aligned}$$

for all $n < N$, with $c_{n+1}/\tau_{n+1} = 0$ for $n = N$. Using (B.4) and (B.5) yields the following expression for the earnings share of firms of categories 1 to n:

$$\begin{aligned}
L_\ell(n) &\equiv \frac{\Pi_\ell^{1,...,n}}{\Pi_\ell} \\
&= \frac{(1+k)(2+k)}{2}\frac{1}{\sum_{h=1}^N \tau_h^{-k} H_h c_h^{2+k}}\sum_{h=1}^{n} H_h c_h^2\left[\left(\frac{c_h}{\tau_h}\right)^k - \left(\frac{c_{n+1}}{\tau_{n+1}}\right)^k\right] \\
&- k(2+k)\frac{1}{\sum_{h=1}^N \tau_h^{-k} H_h c_h^{2+k}}\sum_{h=1}^{n} H_h c_h\tau_h\left[\left(\frac{c_h}{\tau_h}\right)^{1+k} - \left(\frac{c_{n+1}}{\tau_{n+1}}\right)^{1+k}\right] \\
&+ \frac{k(1+k)}{2}\frac{1}{\sum_{h=1}^N \tau_h^{-k} H_h c_h^{2+k}}\sum_{h=1}^{n} H_h \tau_h^2\left[\left(\frac{c_h}{\tau_h}\right)^{2+k} - \left(\frac{c_{n+1}}{\tau_{n+1}}\right)^{2+k}\right],
\end{aligned}$$

with $c_{N+1}/\tau_{N+1} = 0$. A little bit of algebra confirms that $L_\ell(N) = 1$ holds, as should be the case.

Let $q_0 \equiv \Pr(c > c_1/\tau_1) = 1 - [c_1/(c_{\ell,\max}\tau_1)]^k$ denote the fraction of entrepreneurs who fail to serve any market and let $q_n \equiv \Pr(c > c_{n+1}/\tau_{n+1}) = 1 - [c_{n+1}/(c_{\ell,\max}\tau_{n+1})]^k$ denote the fraction of entrepreneurs who serve at most n markets, with $q_N = 1$. Then, for all $n \in \{1, ..., N-1\}$,

$$\begin{aligned}
L_\ell\left(q_n\right) &= \frac{(1+k)(2+k)}{2}\frac{1}{\sum_{h=1}^N \tau_h^{-k} H_h c_h^{2+k}}\sum_{h=1}^{n} H_h c_h^2\left[\left(\frac{c_h}{\tau_h}\right)^k - 1 + q_n c_{\ell,\max}^k\right] \\
&- k(2+k)\frac{1}{\sum_{h=1}^N \tau_h^{-k} H_h c_h^{2+k}}\sum_{h=1}^{n} H_h c_h\tau_h\left[\left(\frac{c_h}{\tau_h}\right)^{1+k} - \left(1 - q_n c_{\ell,\max}^k\right)^{1+1/k}\right] \\
&+ \frac{k(1+k)}{2}\frac{1}{\sum_{h=1}^N \tau_h^{-k} H_h c_h^{2+k}}\sum_{h=1}^{n} H_h \tau_h^2\left[\left(\frac{c_h}{\tau_h}\right)^{2+k} - \left(1 - q_n c_{\ell,\max}^k\right)^{1+2/k}\right], \qquad \text{(B.6)}
\end{aligned}$$

and $L_\ell(q_N) = L_\ell(1) = 1$. This is a Lorenz curve.

We compute the Lorenz curve numerically as follows: the vector of city sizes $\{H_h\}$ is the empirical one, the vector of inverse productivity cutoffs $\{c_h\}$ comes from (B.2), and the vector $\{c_{\ell,\max}\}$ comes from Step 2 above. We pick $k = 2.3$ and we compute the vector of bilateral trade costs as $\tau_h = d_{\ell h}^{0.81}$ for $h \neq \ell$ (otherwise we use $d_{\ell\ell} = (2/3)\sqrt{\text{surface}_\ell/\pi}$), as we explain in the text. Finally, we compute the vector q_n plugging these values into the definition $q_n \equiv 1 - [c_{n+1}/(c_{\ell,max}\tau_{n+1})]^k$.

Finally, we can use expression (B.6) to construct a piecewise linear approximation of the Gini in city ℓ as

$$\widehat{Gini}_\ell(\{c_h\},\{H_h\}) \;=\; 1 - \sum_{n=1}^{N}[L(q_n) - L(q_{n-1}](q_n - q_{n-1}). \tag{B.7}$$

We compute counterfactual Ginis following the same procedure.

C. Numerical procedure to check stability of equilibria

In the numerical application of Section 4.2, we can check the stability of any potential equilibrium candidate (including arbitrary corner solutions) as follows. Let Ω^+ and Ω^0 denote the sets of regions with and without a city at equilibrium, respectively. Assume that there are z regions without a city. The numerical procedure for constructing equilibria and for checking their stability is then as follows.

Let $\mathbf{c} = (c_1\ c_2\ \ldots\ c_\Lambda)$ and let $\mathbf{H} = (H_1\ H_2\ \ldots\ H_\Lambda)$. First, the non-positive expected profit is given by

$$\mathbb{E}(\Delta V_l) = \frac{A}{k+2}\sum_{h\in\Omega^+}\phi_{lh}H_h^* c_h^{*2+k} + \frac{(\alpha - c_l^*)}{2\eta}\left[\alpha - \frac{1+k}{2+k}c_l^*\right] - f^E - \theta H_l^* \equiv f_l(\mathbf{c},\mathbf{H}) \leq 0 \tag{C.1}$$

for any region. Condition (C.1) must hold with equality for all $l \in \Omega^+$ and with strict inequality for all $l \in \Omega^0$. Second, for any l, the identity for the masses of sellers can be rewritten as:

$$g_l(\mathbf{c},\mathbf{H}) \equiv A\frac{\alpha - c_l^*}{\eta c_l^{*1+k}} - \sum_{h\in\Omega^+}\phi_{hl}H_h^* \equiv 0 \tag{C.2}$$

Conditions (C.1) for $l \in \Omega^+$ and conditions (C.2) for all $l = 1, 2, \ldots, \Lambda$ constitute a system of $2\Lambda - z$ equations in the $2\Lambda - z$ unknowns $\{H_l\}_{\Omega^+}$ and $\{c_l\}_\Lambda$. Denote by $(\mathbf{c}^*, \mathbf{H}^*)$ a solution to that system. If $f_l(\mathbf{c}^*, \mathbf{H}^*) < 0$ for all $l \in \Omega^0$, this solution is an equilibrium candidate.

To check whether this solution is a stable equilibrium we proceed as follows. We can uniquely solve the set of equations $g_l(\mathbf{c},\mathbf{H}) = 0$ for all $l \in \Omega^+$ for the $H_l = H_l(\mathbf{c}_{\Omega^+})$ for $l \in \Omega^+$. Note that $\mathbf{c}_{\Omega^+}$ denotes the $\Lambda - z$ dimensional vector of the $\{c_l\}_{\Omega^+}$. We can thus substitute out the $\{H_l\}_{\Omega^+}$ in (C.1). Since $H_l = 0$ for $l \in \Omega^0$, we obtain a system of $\Lambda - z$ equations in the Λ variables c_l.

To check whether no deviation from one city to another is profitable, we have to make sure that the Jacobian associated with $\{f_l\}_{\Omega^+}$ in the variables $\{c_l\}_{\Omega^+}$ is positive definite at $\mathbf{c}^*$ (recall that we already substituted out the H_l) on the subset generated by the constraints (C.2) for $l \in \Omega^0$. After substituting the positive H_l into (C.2), the constraints are given by

$$g_l(\mathbf{c}) \equiv A\frac{\alpha - c_l}{\eta c_l^{1+k}} - \sum_{h\in\Omega^+} \phi_{hl} H_h(\mathbf{c}_{\Omega^+}) \equiv 0 \tag{C.3}$$

for all $l \in \Omega^0$. These constraints define on a one-to-one basis the equilibrium relationships between any c_i with $i \in \Omega^0$ and the set of variables c_l with $l \in \Omega^+$. Applying the implicit function theorem, we then obtain $\mathrm{d}c_l/\mathrm{d}c_i$ for $l \in \Omega^0$ and $i \in \Omega^+$. This finally allows to compute the $\Lambda - z$ square matrix of the Jacobian of $\{f_l\}_{\Omega^+}$ in $\{c_l\}_{\Omega^+}$, *taking into account the general equilibrium constraints via the* $\mathrm{d}c_l/\mathrm{d}c_i$ *terms*. It can readily be evaluated at the equilibrium candidate $(\mathbf{c}^*, \mathbf{H}^*)$. The equilibrium candidate is (locally) stable if this Jacobian is positive definite (which is the higher-dimensional extension of the simple stability condition $\mathrm{d}f(\cdot)/\mathrm{d}c_l > 0$ used in the simple cases).

References

[1] Behrens, K., G. Mion, Y. Murata, and J. Südekum (2012) Spatial frictions. CEPR Discussion Paper #8572.

[2] Corcos, G., M. del Gatto, G. Mion, and G.I.P. Ottaviano (2012) Productivity and firm selection: quantifying the 'new' gains from trade, *Economic Journal* 122, 754–798.