CESifo WORKING PAPERS

7609
2019

April 2019

Urbanization, productivity differences and spatial frictions

Calin Arcalean, Gerhard Glomm, Ioana Schiopu

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7609*
Category 2: Public Choice

Urbanization, productivity differences and spatial frictions

Abstract

We study decentralized and optimal urbanization in a simple multi-sector model of a rural-urban economy focusing on productivity differences and internal trade frictions. We show that even in the absence of the typical externalities studied in the literature, such as agglomeration, congestion or public goods, the decentralized city size can be either too large or too small relative to that chosen by a planner. In particular, optimal urbanization exceeds decentralized levels when productivity differences in location specific non-traded goods is small, a case typically arising in developed economies. In contrast, developing countries are likely to display overurbanization. A numerical exercise calibrated to Brazilian data suggests that the wedges can be quantitatively important. Urban biased policies - placing a higher weight on the welfare of city dwellers - are closer to optimal policies than decentralized allocations whenever productivity differences in non-traded sectors are either very small or very large. For intermediate productivity differences, the urban bias leads to larger cities even relative to decentralized policies.

JEL-Codes: R120, R130, O180, J610.

Keywords: city size, productivity differences, multi-sector models, trade costs, welfare.

Calin Arcalean
ESADE – Ramon Llull University
Spain - 08172 Sant Cugat
calin.arcalean@esade.edu

Gerhard Glomm
Indiana University
USA - 47405 Bloomington IN
gglomm@indiana.edu

Ioana Schiopu
ESADE – Ramon Llull University
Spain - 08172 Sant Cugat
ioana.schiopu@esade.edu

The authors acknowledge financial support from Universitat Ramon Llull, Banc Sabadell, the Spanish Ministry of Education and Science (ECO2016-76855-P) and Generalitat de Catalunya (2017 SGR 640). The usual caveat applies.

1 Introduction

Economic development the world over has been associated with ever-increasing rates of urbanization. According to the U.N.'s The World Urbanization Prospects (2018), the fraction of the population that lives in urban areas has increased from about 30% in 1950 to about 55% in 2018 and is expected to reach 68% by the year 2050. In Europe and North America urbanization has already reached 74% and 82%, comparable with the levels reached in Latin America and the Caribbean (81%) and Oceania (68%) while Asia (50%) and Africa (43%) lag behind.

The size of cities and the associated degree of urbanization have been of interest to academics at least since Beckmann (1972) and Mills and de Ferranti (1971). This issue is also of great concern for policy makers. A recent report for the World Bank by Lall and Shalizi (2006) poses the following questions regarding urbanization: "To what extent is internal migration a desirable phenomenon and under what circumstances? Should governments intervene and if so with what types of interventions? What should be their policy objectives?" While there is a widespread consensus that urbanization represents a great opportunity for economic progress, recent literature, such as Glaeser (2014) or Fuller and Romer (2014) have emphasized the vital role of proper governance along the way.

An abundant literature on optimal urbanization shed light on the effects of congestion and scale economies[1], public goods[2], knowledge spillovers[3], dynamic human capital externalities or agglomeration forces[4].

In this paper we take a step back and focus on how sector specific productivity differences and trade frictions between urban and rural areas shape the allocation of people

[1] Seminal contributions by Mills and de Ferranti (1971) and Sheshinski (1973) consider models in which congestion influences optimal city size. In Henderson (1974), optimal city size balances scale economies and increasing commuting costs for workers.

[2] Arnott (1979) embeds public goods considerations and transportation costs in an explicitly spatial model of city size. Poelhekke and Van der Ploeg (2008) focus on mega-cities and conclude that due to insufficient public goods provision, urbanization is too large. Ott and Soretz (2010) obtain similar results. Albouy et al. (2018) find in calibrated version of their model that US cities might well be under-populated.

[3] Lucas (2001) models spatial structure, production externalities and choice of work sites and residential locations. Following papers by Eaton and Eckstein (1997), Black and Henderson (1999), Glaeser (1999) and Lucas (2004), on "knowledge spillovers" in urban areas, Holmes (2005) estimates these effects for a particular activity, namely locating sales offices. Carlino and Hunt (2007) build on the idea of knowledge spillovers and find a positive correlation between urban density and the rate of innovation. Berliant and Wang (2006) and Gautier and Teulings (2009) provide search theoretic micro-foundations for the knowledge spillovers mentioned above.

[4] In Drazen and Eckstein (1988), the competitive equilibrium allocation does not satisfy the Golden Rule and urbanization is too small relative to the Golden Rule. Following Lucas (1988) there is a sizeable literature on city growth/urbanization that stresses the role of human capital externalities or agglomeration forces. This literature includes Glomm (1992), Lucas (2004), Fan and Stark (2008), Itoh (2009) and others. Depending on the precise specification of the agglomeration effects, urbanization may exceed, fall short of, or be exactly equal to the optimal rate. Fan and Stark (2008), for example obtain "over-migration" since the migration of low skill workers can dilute the agglomeration effects.

across space, abstracting from any explicit externality or agglomeration effect.

Most of the literature on urbanization employs two location-two sector models under the assumption that the urban area, hosting a so called modern sector (typically manufacturing), trades in a frictionless way with the rural area which is home to the traditional sector (typically agriculture).

We generalize this framework in two ways. First, we allow for transportation costs in the tradable sectors. Atkin and Donaldson (2015) provide evidence of large intranational frictions in developing countries. Focusing on Canada, Agnosteva and Yotov (2019) demonstrate large differences in tradability across sectors. Second, we introduce location specific, non-tradable sectors, possibly with different productivity levels.[5]

Local non-traded sectors are an important fraction of employment. Focusing on US data, Jensen and Kletzer (2005) estimate that up to 60% of jobs are in non-traded sectors, even after accounting for the tradability of some service sectors. In France, Frocrain and Giraud (2017) estimate that the share of employment in non-traded sectors can reach $50\%-70\%$ in metropolitan areas with larger shares in rural areas. For a sample of Latin American countries, Reardon and Escobar (2001) document that the share of rural non-farm income in total rural income is between 22% and 68%. Lanjouw and Lanjouw (2001) disaggregate the non-farm rural sector and find that services, commerce and transportation are often the lions' share of non-farm income. They also document that in many countries the size of the non-farm rural sector is even expanding.

Differences in the level of productivity in the local non-traded sectors play a crucial role. Consistent with the data, we assume that rural services are less productive than urban services. There is a fundamental difference in productivity of small-scale local firms between the city and the countryside. Much evidence points in the direction of lower productivity in the countryside. Liedholm and Mead (1987) find that in three countries net returns per hour worked are increasing as one moves from small rural localities with fewer than $2,000$ inhabitants, to rural town with between $2,000$ and $20,000$ thousand inhabitants, to cities with more than $20,000$ inhabitants. Productivities in the cities sometimes exceed productivities in small rural localities by a factor of four. Porter and Bryden (2004) find lower levels of wages in rural vs. urban areas and lower wage growth rates as well. According to Wiggins and Proctor (2001) rural firms operate with lower productivity than urban firms because of a lack of access to physical, human, financial and social capital.

In our model both transportation costs and productivity differences determine the allocation of people across space, which in turn has general equilibrium effects. An individual's move, for example, from the rural to the urban area, increases demand for

[5] This may arise due to agglommeration effects or sorting by skills. For example, Davis and Dingle (2017) show that cities have a comparative advantage over rural areas in skill intensive production. In keeping with the normative focus of the paper, we follow the literature (see for example Albouy et al. 2018) and leave the microfoundations behind these productivity differences unspecified.

both traded and non-traded goods in the city. At the same time, it reduces the supply of labor and thus the production of both traded and non-traded goods in the countryside. As a result, the price of the rural traded good increases, which increases price indices in both locations. How much these price indices go up in relative terms depends on both the transportation costs for the traded goods and the productivity differences in non-traded sectors across locations. In addition, the income in the rural area increases as it is directly related to the price of the traded good.

We compare the decentralized equilibrium with the social planner's solution. While under decentralization migration stops when the level of utility is equalized across the two regions, migrants do not take into consideration how this combination of price and income effects changes with their decision. Therefore, the resulting competitive equilibrium is, in general, not Pareto optimal. In other words, under decentralization, migration equalizes the average utility in the city and the countryside, while the social planner, which is concerned with aggregate welfare, aims to balance marginal welfare across locations. Whether the decentralized equilibrium produces over or under urbanization depends on the productivity differences between sectors and the size of transportation costs. Specifically, decentralized urbanization is too low when productivity differences in the two non-traded sectors are small or when trading costs are relatively higher for the rural good. As productivity differences between urban and rural areas tend to decline with development, our results suggest that, for given transportation costs, the optimal city size is smaller than the decentralized equilibrium at low development levels but larger in developed economies.

Removing *both* transportation frictions in traded sectors and productivity differences in non-traded sectors restores the optimality of decentralized urbanization. This is due to the fact that prices of both traded and non-traded goods, as well as wages, fully reflect the changes in demand and supply - for both goods and labor - arising from relocation of agents. Thus, in this symmetric and frictionless case, geographical space no longer matters and utility equalization across locations also implies income equalization.

Allowing only for transportation frictions introduces a wedge between the purchasing power across locations, which depends on the allocation of people across locations and is ignored under decentralization. Assuming rural goods are more costly to ship than urban goods, the planner chooses a larger city relative to decentralization.

Nonetheless, when allowing for differences in both transportation costs and labor productivity in non-traded sectors, the planner may prefer a smaller city.

In particular, assuming city-based non-traded sectors are more productive leads, everything else equal, to a larger city which in turn leads to further reallocation of labor across sectors within each region. Intuitively, "too many" people are employed in the non-traded sector in the city while the opposite holds in the countryside. This leads to lower employment in rural traded sectors and to the corresponding increase in the relative price of these goods. Furthermore, the two mechanisms interact with one another, as the effects of productivity differences in non traded sectors are magnified by the asymmetry in transportation costs.

We complement the theoretical analysis with a simple quantitative exercise that allows us to gauge the importance of productivity difference and trade frictions for optimal urbanization. We rely on data from the Brazilian 1980 census to compute sectoral employment allocations across urban and rural areas. We parametrize the model to match these data in conjunction with other macroeconomic indicators under decentralized policies. The benchmark parameter values are consistent with significant differences in both productivity and trade frictions. Counterfactual numerical exercises suggest an optimal share of urban population 22% larger than that observed in the data. On the other hand, removing completely trade frictions while maintaining productivity differences in non-traded sectors results in the optimal urbanization being almost 5% smaller than the corresponding decentralized urbanization rate.

Finally, we relate our model to the literature on "urban-biased" policies by constructing a social planner's problem where the welfare of the city dwellers has a higher weight. We compare this "urban-biased" policy against decentralized and optimal policies. Interestingly, we find that relative to decentralization, urban biased policies yield allocations that move in the direction of optimal policies in advanced or very poor economies, i.e. when productivity differences in non-traded sectors are either very small or very large. In contrast, at intermediate levels of development, the urban bias results in too large cities relative to the optimal solution, which in turn is lower than the decentralized city size.

In the following section, we introduce the model. In section 3, we define and solve for the decentralized equilibrium and show it is unique. In section 4, we define the centralized optimum. In section 5, we compare the decentralized and the centralized regime. Section 6 illustrates the theoretical findings with a numerical example calibrated on Brazilian data. In section 7, we discuss "urban-biased" policies. Section 8 contains concluding remarks with directions for future work. Proofs and derivations are contained in the appendix.

2 A model of labor allocation across space and sectors

There are two locations, urban (u) and rural (r). Each is inhabited by mobile households with total mass normalized to unity. Each location produces one traded and one non-traded good. For simplicity, we label the urban traded good b, the rural traded good a and the region specific non-traded goods n^u and n^r respectively. These goods represent a broad mix of sectors, reflecting for example, the fact that the urban areas may produce a range of manufactured goods and services while the rural area will export, in addition to agriculture, some services and even manufactured goods. See, for example, Kolko (1999) and Henderson (2010). In our numerical examples we use an empirical tradability index developed in Gervais and Jensen (2015) to group various sectors into tradable/non-tradable.

Households in each region derive utility from consumption of the two traded goods and from the region specific non-traded good. There are iceberg trade costs associated with shipping the tradable goods between regions, τ_b and τ_a, where $\tau_b > 1$ and $\tau_a > 1$, i.e. if urban dwellers want to consume one unit of the rural traded good, they need to buy τ_a units. We assume that in each region labor markets are competitive and labor moves freely between sectors until wages are equalized.

Individuals. The urban residents maximize:

$$U^u(b^u, a^u, n^u) = \left[(b^u)^\theta + (a^u)^\theta + (n^u)^\theta\right]^{1/\theta}, \tag{1}$$

where b^u, a^u and n^u are consumption of traded urban good b, rural traded good a, and the urban non-traded good n^u, respectively. In the following, we assume that goods are substitutes in consumption, i.e. $\theta \in (0, 1]$.[6] The elasticity of substitution is $e = 1/(1-\theta)$.

We assume the urban traded good is the numeraire and normalize $p_b = 1$. The budget constraint of the urban consumer is:

$$b^u + p_a \tau_a a^u + p_{nu} n^u = I^u, \tag{2a}$$

where p_a and p_{nu} are the relative prices of a and n^u respectively. The term $\tau_a a^u$ is the quantity of good a that city dwellers buy in order to consume a^u units. I^u is the income of an urban resident.

Preferences of individuals in the rural area are identical to those in the urban area and given by:

$$U^r(b^r, a^r, n^r) = \left[(b^r)^\theta + (a^r)^\theta + (n^r)^\theta\right]^{1/\theta}, \tag{3}$$

where b^r, a^r and n^r are consumption of the urban good, rural traded and the rural non-traded good, respectively. The corresponding budget constraint is:

$$\tau_b b^r + p_a a^r + p_{nr} n^r = I^r, \tag{4}$$

while p_a and p_{nr} are the corresponding relative prices. The term $\tau_b b^r$ is the quantity of traded urban good that consumers in the rural area buy in order to consume b^r units. I^r is the income of a resident in the rural area.

Production. There are four production technologies in this economy, all of them

[6] This is in line with the recent trade literature, e.g. Costinot and Rodriguez-Clare (2014).

are linear in labor: $Y_b = A_b L_b$, $Y_a = A_a L_a$, $Y_{nu} = A_{nu} L_{nu}$ and $Y_{nr} = A_{nr} L_{nr}$, where L_b, L_a, L_{nu} and L_{nr} are employment allocations in the b-sector, a-sector and the urban and rural non-traded sectors, respectively. A_b, A_a, A_{nu} and A_{nr} are the corresponding sectoral exogenous productivities. In each sector there are competitive firms that maximize profits taking prices as given.

Assumption 1. *Productivity in the urban non-traded sector is higher than that in the rural non-traded sector:* $A_{nu} > A_{nr}$.

This assumption is consistent with empirical evidence by Liedholm and Mead (1987), Ciccone and Hall (1996), Wiggins and Proctor (2001), and Porter and Bryden (2004).

In the following, we analyze and compare three regimes: 1) the decentralized equilibrium, where households freely choose their region of residence, 2) the centralized optimum, where a social planner chooses the allocation of people between city and rural area that maximizes the aggregate welfare in the economy and 3) an "urban biased" regime where the social planner attaches a higher weight on the welfare of the city inhabitants.

3 Decentralized equilibrium

Denote with ϕ the mass of people residing in the urban region (the size of the city). As the total population in the economy is normalized to 1, $\phi \in (0, 1)$.

Labor market clearing implies: $\phi = L_b + L_{nu}$, and $1 - \phi = L_a + L_{nr}$. Denoting by ϕ_{nu} the fraction of people in the city working in the non-tradable sector, we can write

$$L_{nu} = \phi_{nu} \phi \tag{5}$$

and

$$L_b = (1 - \phi_{nu})\phi. \tag{6}$$

Similarly,

$$L_{nr} = \phi_{nr}(1 - \phi) \tag{7}$$

and

$$L_a = (1 - \phi_{nr})(1 - \phi). \tag{8}$$

Next, we solve the consumers' problems in the two regions. Denote $\gamma = \theta/(\theta - 1) < 0$ when $\theta \in (0, 1]$. The demands for tradable and non-tradable goods in the city are:

$$b^u = I^u \frac{1}{1 + (\tau_a p_a)^\gamma + (p_{nu})^\gamma}, \tag{9}$$

$$a^u = I^u \frac{(\tau_a p_a)^{\gamma - 1}}{1 + (\tau_a p_a)^\gamma + (p_{nu})^\gamma}, \tag{10}$$

$$n^u = I^u \frac{(p_{nu})^{\gamma-1}}{1 + (\tau_a p_a)^\gamma + (p_{nu})^\gamma}. \tag{11}$$

In the rural area, the corresponding demand functions are:

$$b^r = I^r \frac{(\tau_b)^{\gamma-1}}{(\tau_b)^\gamma + (p_a)^\gamma + (p_{nr})^\gamma} \tag{12}$$

$$a^r = I^r \frac{(p_a)^{\gamma-1}}{(\tau_b)^\gamma + (p_a)^\gamma + (p_{nr})^\gamma} \tag{13}$$

$$n^r = I^r \frac{(p_{nr})^{\gamma-1}}{(\tau_b)^\gamma + (p_a)^\gamma + (p_{nr})^\gamma}. \tag{14}$$

In each sector, labor is paid the corresponding value of marginal product: $w_b = A_b$, $w_a = p_a A_a$, $w_{nr} = p_{nr} A_{nr}$, $w_{nu} = p_{nu} A_{nu}$, where w_b, w_a, w_{nr}, and w_{nu} are the wages in the b-sector, a-sector, rural and urban non-traded sectors, respectively.

Now we define the decentralized equilibrium in which household freely decide their location by comparing the corresponding utilities in each location. As household are identical, in equilibrium the utility of residing in the city is equal to that in the rural area.

Definition 1. *A decentralized equilibrium consists of urban and rural households' consumption allocations* $(b^u, a^u, n^u, b^r, a^r, n^r)$*, prices of goods* (p_a, p_{nr}, p_{nu})*, wages* (w_b, w_a, w_{nu} *and* w_{nr})*, the fraction of labor residing in the city* ϕ*, sectoral employment shares within regions* (ϕ^{sr}, ϕ^{su})*, such that, given the exogenous trade costs* τ_a *and* τ_b*, and the exogenous sectoral productivities* A_b, A_a, A_{nu} *and* A_{nr} *:*

1. households maximize utility subject to the budget constraint, given wages and the prices of goods;

2. the competitive firms solve their optimization problem given the prices of goods and wages;

3. wages are equalized across sectors;

4. all good markets clear;

5. the labor markets clear in each region;

6. households' utilities are equal across regions.

Next, we solve for the decentralized allocations. Wage equalization across sectors in each region implies: $w_b = w_{nu}$ and $w_a = w_{nr}$. Thus, the income of an urban resident is:

$$I^u = A_b = p_{nu} A_{nu} \tag{15}$$

and the typical income in the rural area is:

$$I^r = p_a A_a = p_{nr} A_{nr}. \tag{16}$$

Thus, we can express p_{nu} and p_{nr} as functions of p_a and sectoral productivities:

$$p_{nu} = \frac{A_b}{A_{nu}} \tag{17}$$

and

$$p_{nr} = p_a \frac{A_a}{A_{nr}}. \tag{18}$$

Market clearing in non-traded good in the urban area implies $\phi s^u = A_{nu} L_{nu}$, where the left-hand side is the quantity demanded and the right-hand side is the quantity supplied. Substituting L_{nu} from (5) and demand for non-traded good in the urban area n^u from (11),we obtain:

$$\phi_{nu} = \frac{I^u}{A_{nu}} \frac{(p_{nu})^{\gamma-1}}{1 + (\tau_a p_a)^\gamma + (p_{nu})^\gamma}.$$

Plugging in $I^u = p_{nu} A_{nu}$, we can write the employment share in non-traded sector in the urban area as a function of prices and transportation frictions:

$$\phi_{nu} = \frac{(p_{nu})^\gamma}{1 + (\tau_a p_a)^\gamma + (p_{nu})^\gamma}.$$

By the same token, we obtain $(1 - \phi)n^r = A_{nr} L_{nr}$. Substituting L_{nr} from (7), the demand for non-traded good in the rural area n^r from (14) and $I^r = p_{nr} A_{nr}$, we obtain:

$$\phi_{nr} = \frac{(p_{nr})^\gamma}{(\tau_b)^\gamma + (p_a)^\gamma + (p_{nr})^\gamma}.$$

We substitute (17) and (18) in the expressions of ϕ_{nu} and ϕ_{nr} and get:

$$\phi_{nu} = \frac{(\frac{A_b}{A_{nu}})^\gamma}{1 + (\tau_a p_a)^\gamma + (\frac{A_b}{A_{nu}})^\gamma} \tag{19}$$

$$\phi_{nr} = \frac{(p_a \frac{A_a}{A_{nr}})^\gamma}{(\tau_b)^\gamma + (p_a)^\gamma + (p_a \frac{A_a}{A_{nr}})^\gamma}. \tag{20}$$

Next, we use the market clearing of the urban traded good b to obtain a relationship between the relative price p_a and the share of urban population ϕ :

$$\phi b^u + (1 - \phi)\tau_b b^r = A_b L_b. \tag{21}$$

The demand for good b in the urban and rural areas are, respectively:

$$b^u = \frac{A_b}{1 + (\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma} \text{ and } b^r = \frac{p_a A_a (\tau_b)^{\gamma-1}}{(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a \frac{A_a}{A_{nr}}\right)^\gamma},$$

where we substituted the expressions for I^u, I^r, p_{nu} and p_{nr} (equations (15), (16), (17) and (18), respectively). Next, we use the equation (19) for ϕ_{nu} to express L_b as a function of the endogenous variables p_a and ϕ :

$$L_b = (1-\phi_{nu})\phi = \frac{1+(\tau_a p_a)^\gamma}{1+(\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}\phi.$$

Plugging b^u, b^r and L_b into the market clearing condition for good b, (21) and rearranging terms yields:

$$\frac{(1-\phi)}{\phi}\frac{A_a(\tau_b)^\gamma}{A_b(\tau_a)^\gamma(p_a)^{\gamma-1}} = \frac{(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}{1+(\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}. \tag{22}$$

We express ϕ as a function of p_a :

$$\phi = \frac{A_a p_a(\tau_b)^\gamma\left[1+(\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right]}{A_a p_a(\tau_b)^\gamma\left[1+(\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right] + A_b(\tau_a)^\gamma(p_a)^\gamma\left[(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a\frac{A_a}{A_{nr}}\right)^\gamma\right]}. \tag{23}$$

Finally, in the decentralized regime the utility is equalized across regions: $U^u(b^u, a^u, n^u) = U^r(b^r, a^r, n^r)$.

We use the demand expressions for each good in the urban and rural areas (9), (10), (11), (12), (13) and (14) to obtain the indirect utilities in the two regions:

$$V^u = I_u\left[1+(\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}} \tag{24}$$

and

$$V^r = I_r\left[(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a\frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}}. \tag{25}$$

Equalizing V^u and V^r and substituting (15) and (16) yields:

$$\frac{A_b}{A_a p_a} = \left[\frac{(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}{1+(\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}\right]^{-\frac{1}{\gamma}}. \tag{26}$$

The expression (26) pins down the equilibrium p_a as a function of exogenous variables, sectoral productivities and transportation costs. The proposition below establishes the existence and uniqueness of the solution, as well as properties of p_a.

Proposition 1. *There exists a unique relative price $p_a > 0$ that solves equation (26). The equilibrium relative price p_a is increasing in A_{nu} and decreasing in A_{nr}. Also, p_a is increasing in τ_b and decreasing in τ_a.*

Proof in the Appendix.

A unique p_a implies a unique decentralized allocation of people across regions, ϕ_d. The fact that $\phi_d \in (0,1)$ can immediately be seen from studying equation (23).

Proposition 2. *Denote by ϕ_d the share of urban population in the decentralized equilibrium. There exists a unique interior $\phi_d \in (0,1)$ that satisfies equation (23), where p_a is given by (26). The decentralized share of urban population ϕ_d is increasing in p_a.*

Proof in the Appendix.

An increase in p_a raises the demand for the urban traded good both in the city and the country side, as well as the relative employment in this sector. When goods are substitutes, trade balance implies a larger share of population residing in the city.[7] In equilibrium, the relative price p_a and the city size are determined jointly by utility equalization and market clearing in tradables. Once the solutions for ϕ_d and p_a are found, the sectoral allocations of labor can be determined. The share of labor in urban and rural non-traded goods are given by equations (19) and (20), respectively. Studying the expressions (19) and (20), we can see that the share of labor in urban non-tradables ϕ_{nu} is increasing in p_a, while the opposite holds for ϕ_{nr}.

4 Centralized optimum

Now we focus on a centralized optimum, in which the fraction of people residing in the city is chosen by a planner that maximizes the aggregate welfare in the economy subject to wage equalization across sectors and clearing of labor and good markets.

Definition 2. *A centralized optimum consists of urban and rural households' consumption allocations $(b^u, a^u, n^u, b^r, a^r, n^r)$, prices of goods $(p_b, p_a, p_{nr}, p_{nu})$, wages ($w_b$, w_a, w_{nu} and w_{nr}), the fraction of people residing in the city ϕ, sectoral employment shares within regions (ϕ_{nr}, ϕ_{nu}), such that, given the exogenous trade costs τ_a and τ_b, and the exogenous sectoral productivities A_b, A_a, A_{nu} and A_{nr} :*

1. households maximize utility subject to the budget constraint given the prices of goods,

2. the competitive firms solve their optimization problem given the prices of goods and wages,

3. wages are equalized across sectors,

4. all good markets clear,

5. the labor markets clear in each region, and

6. the aggregate welfare function

$$W(\phi) = \phi U^u(\phi) + (1-\phi)U^r(\phi),$$

[7]The property also holds for moderate degrees of complementarity $\theta < 0$, given transportation costs are not too large.

is maximized, where U^u and U^r are the individual utilities in the urban and rural areas.

Next, we substitute the urban and rural incomes $I^u = A_b$ and $I^r = p_a A_a$ into equations (24) and (25) to express the indirect utilities as functions of the relative price $p_a(\phi)$ and the exogenous parameters:

$$V^u(\phi) = A_b \left[1 + (\tau_a p_a(\phi))^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}} \tag{27}$$

$$V^r(\phi) = A_a p_a(\phi) \left[(\tau_b)^\gamma + (p_a(\phi))^\gamma + \left(p_a(\phi)\frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}}. \tag{28}$$

Consequently, the social planner's problem is:

$$\max_{\phi} W(\phi) = \phi V^u(p_a(\phi)) + (1-\phi)V^r(p_a(\phi)), \tag{29}$$

where $V^u(p_a(\phi))$ and $V^r(p_a(\phi))$ are given by (27) and (28), respectively, and $p_a(\phi)$ satisfies the market clearing condition in the urban traded sector, (22).

Proposition 3. *Denote by ϕ_c the optimal share of urban population. There exists an interior solution $\phi_c \in (0,1)$ to the central planner's problem* (29).

Proof in the Appendix.

Proving uniqueness of the solution is challenging, as the function $W(\phi)$ is not concave over the entire interval $(0,1)$. However, simulations over the relevant parameter ranges reveal that the solution of the central planner has a unique global maximum.

5 Comparison of regimes

In the following we compare the decentralized fraction of urban population with the centralized optimum. We proceed by evaluating the slope of the aggregate welfare function $\partial W(\phi)/\partial\phi$ at the decentralized solution ϕ_d. If $\partial W/\partial\phi|_{\phi=\phi_d} > 0$, then then social planner prefers a bigger city than the one arising in the decentralized equilibrium, i.e. $\phi_c > \phi_d$. The opposite holds when $\partial W/\partial\phi|_{\phi=\phi_d} < 0$.

Differentiating W with respect to ϕ yields:

$$\frac{\partial W}{\partial \phi} = (1-\phi)\frac{\partial V^r(\phi)}{\partial \phi} + \phi\frac{\partial V^u(\phi)}{\partial \phi} + V^u(\phi) - V^r(\phi),$$

where $V^u(\phi)$ and $V^r(\phi)$ are given by (24) and (25).

When evaluating $\partial W(\phi)/\partial\phi$ at the decentralized allocation ϕ_d the last two terms cancel as the decentralized allocation satisfies $V^u(\phi_d) = V^r(\phi_d)$. Thus, $\partial W/\partial\phi|_{\phi=\phi_d}$

becomes:

$$\begin{aligned}\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d} &= (1-\phi_d)\left.\frac{\partial V^r(\phi)}{\partial \phi}\right|_{\phi=\phi_d} + \phi_d \left.\frac{\partial V^u(\phi)}{\partial \phi}\right|_{\phi=\phi_d} \\ &= -V^r(\phi_d)\frac{(1-\phi_d)(p_a)^{\gamma-1}}{(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a \frac{A_a}{A_{nr}}\right)^\gamma}\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d} \\ &+ V^r(\phi_d)\frac{(1-\phi_d)}{p_a}\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d} \\ &- V^u(\phi_d)\frac{\phi_d (p_a)^{\gamma-1}(\tau_a)^\gamma}{1+(\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d},\end{aligned} \tag{30}$$

where we used the fact that $(1-\theta)\,\gamma/\theta = -1$.

The expression above shows that the under decentralization the marginal worker ignores the effects of her migration decision on the other agents, both in the rural and the urban region. The terms in the expression above illustrate these effects that work through the prices. Intuitively, there are direct effects on prices that consumers pay and indirect effects on workers' incomes.

Consider a marginal increase in the size of the city ϕ. As shown in Proposition 2, a higher city size ϕ has a positive effect on the relative price of rural traded good p_a $\left(\partial p_a(\phi)/\partial\phi|_{\phi=\phi_d} > 0\right)$. Wage equalization in the rural area implies a higher price of rural non-traded goods p_{nr}. Higher prices have a negative effect on consumers' welfare V^r, which is reflected in the first term in the expression above. At the same time, as p_a goes up, the incomes in the rural area increase $\left(\partial I^r/\partial\phi = A_a\, \partial p_a(\phi)/\partial\phi|_{\phi=\phi_d} > 0\right)$.

In the urban area, a higher ϕ has a negative effect on welfare, driven by a higher relative price of the traded good p_a. There is no income effect or an effect through the price of tradables as the urban traded good is the numeraire. As the production functions are linear in labor, the relative price of urban non-traded sectors p_{nu} is constant.[8]

Using the fact that $V^u(\phi_d) = V^r(\phi_d)$ and $\frac{1-\theta}{\theta}\gamma = -1$, we can rewrite $\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d}$ as:

$$\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d} = V^u(\phi_d)\left\{\begin{array}{c} -\phi_d \underbrace{\frac{(\tau_a)^\gamma (p_a)^{\gamma-1}}{1+(\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}}_{\text{individual price effect on V}^u} - (1-\phi_d)\times \\ \times \underbrace{\frac{\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right](p_a)^{\gamma-1}}{(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}}_{\text{individual price effect on V}^r} + (1-\phi_d) \underbrace{\frac{1}{p^a}}_{\text{individual income effect on V}^r} \end{array}\right\}\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d}. \tag{31}$$

[8] A production function with decreasing returns in factors would generate an income effect, as well a change in the price of services.

Combining the price and the income effect in the rural area yields the net effect on individual welfare: $\frac{(\tau_b)^\gamma}{p_a}\left((\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma\right)^{-1}>0$.

Consequently, the net effect of a marginal increase of ϕ on the utility in the rural area is positive, that is, the positive income effect outweighs the negative price effect.

Thus, the net aggregate welfare effect of a marginal change in the size of the city at $\phi=\phi_d$ depends on the magnitude of opposite effects on individual welfare in each region, as well as their weights (ϕ_d and $1-\phi_d$), which depend on the magnitudes of the spatial frictions present in the model (the trade costs and the productivity gap between the rural and urban areas). When the net change is positive, $\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d}>0$, the centralized solution ϕ_c is higher than the decentralized one, ϕ_d. The opposite holds when $\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d}<0$.

Next, we rewrite

$$\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d}=\underbrace{\frac{V^u(\phi_d)(p_a)^{\gamma-1}}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}}_{D>0}\left\{\begin{array}{c}-\phi_d(\tau_a)^\gamma\\ +(1-\phi_d)\frac{(\tau_b)^\gamma}{(p^a)^\gamma}\frac{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}\end{array}\right\}\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d}$$

and use the fact that ϕ_d satisfies the market clearing condition for the urban traded good:

$$\frac{(1-\phi_d)}{\phi_d}\frac{A_a(\tau_b)^\gamma}{A_b(\tau_a)^\gamma(p_a(\phi_d))^{\gamma-1}}=\frac{(\tau_b)^\gamma+(p_a(\phi_d))^\gamma+\left(p_a(\phi_d)\frac{A_a}{A_{nr}}\right)^\gamma}{1+(\tau_a p_a(\phi_d))^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}.$$

We obtain:

$$\begin{aligned}\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d} &= D\left\{\begin{array}{c}-\phi_d(\tau_a)^\gamma\\ +(1-\phi_d)\frac{(\tau_b)^\gamma}{(p^a)^\gamma}\frac{\phi_d}{(1-\phi_d)}\frac{A_b(\tau_a)^\gamma(p_a)^{\gamma-1}}{A_a(\tau_b)^\gamma}\end{array}\right\}\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d}\\ &= D\phi_d(\tau_a)^\gamma(p_a)^{-1}\left(\frac{A_b}{A_a}-p_a\right)\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d}\end{aligned}$$

In the proof of Proposition 2 we have shown that $\partial\phi/\partial p_a>0$. Thus, $\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d}>0$.

Therefore, $sign\left[\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d}\right]=sign\left(\frac{A_b}{A_a}-p_a(\phi_d)\right)$.

The following proposition states the conditions under which the planner prefers a larger (smaller) city than the decentralized solution.

Proposition 4. *The social planner prefers a larger city than the decentralized outcome ($\phi_c>\phi_d$) when the following condition holds:*

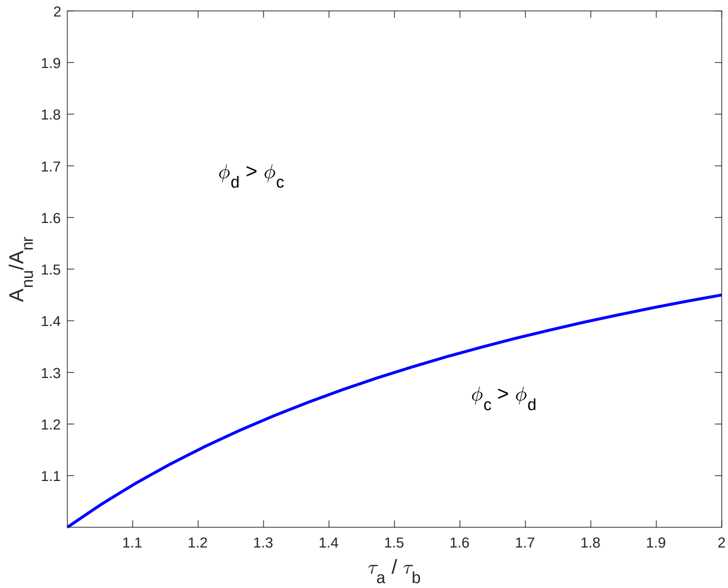
$$(A_{nu})^{-\gamma}-(A_{nr})^{-\gamma}<(A_a)^{-\gamma}\left[1-(\tau_a)^\gamma\right]-(A_b)^{-\gamma}\left[1-(\tau_b)^\gamma\right], \tag{32}$$

where $\theta\in(0,1)$ and $\gamma=\theta/(\theta-1)<0$.

Proof in the Appendix.

When $A_{nr} < A_{nu}$, $A_a < A_b$ and $\tau_a > \tau_b$, inequality (32) is more likely to hold when the productivity difference in non-tradables is small. Also, the condition is more likely to hold when the iceberg costs in non-tradables τ_a are large relative to τ_b, and the productivity differences in tradables is small across regions. Figure 1 shows the parameter space for which inequality (32) holds when A_{nu}/A_{nr} and τ_a/τ_b vary and A_a and A_b are fixed.

Figure 1: Comparison of decentralized and centralized policies



Parameter constellations below the curve imply $\phi_c > \phi_d$. $\tau_b = 1$, $\theta = 0.5$. $A_a = 0.9$, $A_{nr} = 1$, $A_b = 1$.

In order to get the intuition behind the result in Proposition 4, it is instructive to first consider the benchmark case in which (i) there are no transportation costs for tradable goods ($\tau_a = \tau_b = 1$) and (ii) the productivities in the non-tradable sectors are equal ($A_{nu} = A_{nr}$). In this case, the centralized and decentralized allocations are identical.

Proposition 5. *If $A_{nu} = A_{nr}$ and $\tau_a = \tau_b = 1$, then $\phi_c = \phi_d = 1/(1 + (A_b/A_a)^\gamma)$ and $p_a = A_b/A_a$.*

Proof. *Under the conditions above, the solution for p_a that solves equation (26) is $p_a = A_b/A_a$. Substituting this into (23) leads to $\phi = 1/(1 + (A_b/A_a)^\gamma)$. Moreover, $\partial W/\partial\phi|_{\phi=\phi_d} = A_b/A_a - p_a = 0$, hence $\phi_c = \phi_d$.*

Note that in this case, the average and the marginal utility levels are equal. Moving across locations does not lead to any change in the price of the non-tradables since $p_{nu} = A_b/A_{nu}$ and $p_{nr} = p_a A_a/A_{nu} = A_b/A_{nr}$ when $p_a = A_b/A_a$. Thus, $p_{nu} = p_{nr}$ for $A_{nu} = A_{nr}$. Furthermore, the production and the consumption prices for the tradable goods are equalized in both locations. We now analyze the role of each channel separately (differences in transportation costs and productivity gap in non-tradable sectors).

First, when $\tau_a = \tau_b = 1$ but $A_{nu} \neq A_{nr}$, and in particular $A_{nu} > A_{nr}$, the right-hand side of inequality (32) becomes zero while the left-hand side is positive, thus the planner prefers a smaller city (ϕ) and a lower price for the rural tradable good (p_a) than in the decentralized regime. As we have shown in Proposition 1, the price p_a is increasing in A_{nu}. Consider a change from a benchmark in which $A_{nu} = A_{nr}$ to a situation in which A_{nu} is higher and A_{nr} stays the same (such that now $A_{nu} > A_{nr}$). The resulting decentralized p_a increases relative to the benchmark, which yields a higher decentralized ϕ_d. The marginal migrant - which takes prices and incomes as given - ignores the general equilibrium effect of the increase in p_a on the welfare of both city and countryside residents. We can see from expression (31) that a higher A_{nu} magnifies the negative price effect in the individual welfare in the city. Also, a higher decentralized ϕ_d implies a larger aggregate price effect in the city, and a smaller positive net effect on rural welfare (lower $1-\phi_d$). Consequently, the negative welfare effect in the city offsets the positive effect in the rural area and the planner chooses a lower p_a and a smaller city size than the decentralized equilibrium. The larger the productivity difference in non-tradables between the rural and urban areas, the lower the city size preferred by the planner.

Second, consider the case of $A_{nu} = A_{nr}$ and $\tau_a > \tau_b = 1$ (there are positive transportation costs only for the tradable good produced in the rural area). The left-hand side of inequality (32) becomes zero while the right-hand side is positive, thus the planner prefers a larger city (ϕ) and a higher price for the rural tradable good (p_a) than in the decentralized regime. As equilibrium p_a is decreasing in τ_a (see Proposition 1), the resulting decentralized ϕ_d and p_a are lower than in the frictionless benchmark. At ϕ_d, a marginal increase in ϕ has a positive welfare effect in the rural area and negative in the city (see equation 31) that are ignored by the marginal migrant. However, the aggregate positive effect in the rural area dominates, as $1 - \phi_d$ is larger, and the planner can increase aggregate welfare by choosing a larger city. A larger productivity A_a magnifies the positive welfare effect in the rural area from having a larger city and makes social planner's solution (ϕ_c) even larger relative to the decentralized one (ϕ_d).

Positive transportation costs in the tradable good produced in the city ($\tau_b > 1$) change the effects discussed above in the opposite direction. The net effect (and the optimal city size) depends on the productivity difference in the tradable sectors (A_a relative to A_b) and the difference in trade costs. For example, if the differences in transportation costs are large and productivity differences in tradables are small (A_a is large relative to A_b and τ_a is large relative to τ_b), then the planner prefers a bigger city.

As the magnitude of productivity differences varies with the level of development, these results have implications on the optimal urban policies along the development path.

While direct evidence on rural-urban productivity differences at sector level is scant, more developed economies have in general a more uniform distribution of industries across space (see Imbs et al. (2012)). Furthermore, the productivity differences across sectors do vary with the level of development. For example, Acemoglu and Zilibotti (2001) document that the productivity differences between manufacturing (more concentrated in the city) and agriculture (more concentrated in the rural areas) are more pronounced in developing economies than in the more advanced ones. Finally, Duarte and Restuccia (2019) find that cross-country productivity differences are much larger in non-traded vs. traded services.

Therefore, we would expect A_{nu}/A_{nr} to be inversely related to the level of development. In this case, as shown in Figure 1, our results imply that, at a given level of transportation costs, the optimal city size is smaller than the decentralized equilibrium at low development levels (A_{nu}/A_{nr} large) and larger in developed economies (A_{nu}/A_{nr} small). The implication for advanced levels of development echoes, at an aggregate level, the result in Albouy et al. (2018), who find that most of the cities in the US are too small. However, while their model relies on a mix of location and fiscal externalities, our results stem from the interplay of transportation frictions and productivity differences between regions.

At intermediate levels of development, the ranking of decentralized and optimal urbanization depends on the relative transportation costs of the two tradable goods. Thus, for example, in a large economy, rural tradable goods, such as agricultural products or other natural resources, are likely to face larger transportation even at higher levels of development and thus $\phi_c > \phi_d$. The numerical example presented in the next section, calibrated using data from Brazil, approximates well such a case.

6 A numerical example

In this section we provide a numerical illustration of the model using data from Brazil in the 80s. Although the model is stylized, this exercise allows us to gauge the quantitative importance of the frictions at work.

We start by calibrating the decentralized version of the model to broadly match sectoral data on employment and productivity across urban and rural areas. Next, we use the calibrated parameters values to back out the optimal city size implied by our model.

The model's parameters are the sectoral labor productivity levels A_b, A_{nu}, A_a and A_{nr}, the transportation costs τ_a and τ_b and θ, driving the elasticity of substitution. In the following we normalize $A_b = 1$ and use $\theta = 0.8$ implying an elasticity of substitution of 5.[9] The rest of the model's parameters are calibrated as follows.

We use Brazilian census data from 1980, available from IPUMS-International to compute the distribution of workers across sectors and space.[10]

[9] We follow parametrizations used in the recent quantitative trade literature, e.g. Albrecht and Trevor Tombe (2016) and Costinot and Rodriguez-Clare (2014).

[10] See IPUMS (2018) for details on the IPUMS-International database.

Following the partition suggested by the model, we use the share of value added in tradable industries from Gervais and Jensen (2015) to construct a tradability index of broad sectors. Thus, sectors with a tradability index above median are considered tradable. These are: *agriculture, fishing, and forestry*, *mining*, *manufacturing*, *wholesale and retail trade*, *transportation, storage and communications* and *financial services and insurance*.[11]

We then compute the fraction of workers residing in urban areas, $\phi = 0.67$, the employment shares in non-tradable sectors in the city, $\phi_{nu} = 0.46$ and the countryside, $\phi_{nr} = 0.09$, as well as the ratio of rural to urban average income, $I_r/I_u = 0.29$. In addition, we use the 10-Sector database to retrieve the share of non-traded sectors in GDP, which is 0.307. In the model this corresponds to the aggregate production of non-tradables as a share of income $(\phi p_{nu} n^u + (1-\phi) p_{nr} n^r)/(\phi I_u + (1-\phi) I_r)$.

We then set the remaining parameters to jointly match the above data moments under the decentralized regime. Table 1 summarizes the calibration results:

Table 1: Calibrated numerical example

Parameter value	Target	Description	Data	Model
$A_a = 1.14$	ϕ	share of urban pop.	0.69	0.69
$A_{nu} = 0.91$	ϕ_{nu}	urban share of service empl.	0.46	0.41
$A_{nr} = 0.61$	ϕ_{nr}	rural share of service empl.	0.09	0.08
$\tau_a = 22.03$	$\frac{\phi p_{nu} n_u + (1-\phi) p_{nr} n_r}{\phi I_u + (1-\phi) I_r}$	share of non-tradables in GDP	0.307	0.367
$\tau_b = 1.07$	I_r/I_u	share of rural/urban income	0.29	0.29

Note that the calibration yields a very large transportation cost for rural traded goods, which more then offsets the productivity of the traded sector in this region. This is to be expected given that the model is silent about all externalities and frictions beyond those induced by spatial segmentation of markets. Finally, the rural non-traded sector is estimated to be roughly two-thirds as productive as its urban counterpart.

Using the model thus parametrized, we compute the centralized urbanization share ϕ_c that maximizes the aggregate welfare $\phi V^u + (1-\phi) V^r$. We find $\phi_c = 0.8416$, an increase of 22% relative to the decentralized urbanization.

Starting from this parametrization, we now analyze the role of the two frictions considered for the wedge between decentralized and optimal urbanization. Recall that assuming no productivity differences across non-traded sectors as well as zero transportation costs, decentralized urbanization is optimal, i.e. $\phi_c = \phi_d = 1/(1 + (A_b/A_a)^\gamma)$.

Everything else equal, assuming $\tau_a = \tau_b = 1$ yields a decentralized urbanization rate of $\phi_d = 0.52$, 5% higher than the corresponding centralized allocation, $\phi_c = 0.50$. Note that

[11] While Gervais and Jensen (2015) exclude some sectors due to data limitations, we consider agriculture, mining, fishing and forestry as tradable and construction as non-tradable. At the same time, as the IPUMS industry classification lumps together retail and wholesale trade, we consider this sector tradable.

in this case, urbanization is lower under both regimes as the demand for rural tradable good increase and so does the income in the rural area. Intuitively, a reduction in the transportation costs for goods produced in rural areas, be it exogenous or policy driven, could imply the need to adjust the urbanization policies as the decentralized population distribution may be suboptimally skewed towards cities.

While in-depth policy analysis is beyond the scope of this framework, the numerical exercise suggests nonetheless that spatial productivity differences and transportation costs could potentially generate significant departures from optimal urbanization rates.

7 The urban bias

The city dwellers and the inhabitants of the rural area have conflicting interests regarding the size of the city, as the individual welfare in the urban area $V^u(\phi)$ is decreasing in ϕ, while the opposite is true for $V^r(\phi)$. Starting with the seminal contribution of Lipton (1977), this tension has long been noted in the development literature, for example in Yang (1999)) or Bezemer and Headey (2008). In the spirit of this literature, we analyze the case when the central planner has an "urban bias", that is, it puts a higher weight on the total welfare in the city. Consequently, the social planner's problem with urban bias becomes:

$$\max_{\phi_p} W^p(\phi) = \delta\phi V^u(p_a(\phi)) + (1-\delta)(1-\phi)V^r(p_a(\phi)), \tag{33}$$

where $V^u(p_a(\phi))$ and $V^r(p_a(\phi))$ are given by (27) and (28), respectively, and $p_a(\phi)$ satisfies the market clearing condition in the urban traded sector, (22). Also, in the case of urban bias $\delta > 1/2$.

Denote by ϕ_p the solution chosen by the urban biased planner. It can be shown that an interior solution always exists in the interval $(0, 1)$. The proof is similar to that of Proposition 3.

Proposition 6. *$\phi_p > \phi_d$ when $(A_{nu})^{-\gamma} - (A_{nr})^{-\gamma} < (A_a)^{-\gamma}\left[1 - \kappa^{-\gamma}(\tau_a)^{\gamma}\right] + (A_b)^{-\gamma}\left[\kappa^{-\gamma}(\tau_b)^{\gamma} - 1\right]$, where $\kappa = (1-\delta)/\delta < 1$ and $\delta > 1/2$. The condition is sufficient but not necessary. When the condition holds, also $\phi_c > \phi_d$.*

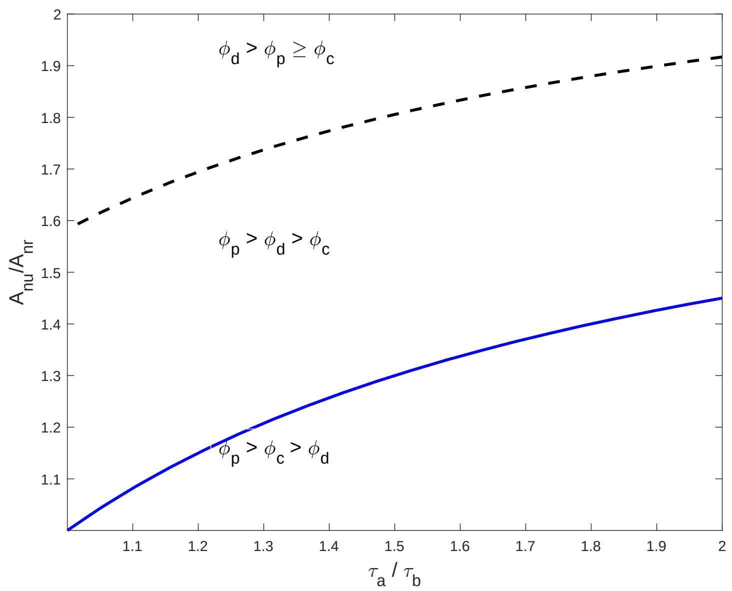
Proof in the Appendix.

The condition is more likely to hold when the difference in productivities between the non-tradable sectors is small, or when τ_a is large relative to τ_b. In this case, the urban bias leads to a city size that is larger than the decentralized equilibrium ($\phi_p > \phi_d$), while the central planner also prefers a large city size ($\phi_c > \phi_d$).

Ranking ϕ_p and ϕ_c is not possible analytically, therefore we resort to the numerical analysis of the model over different ranges of productivity differences and transportation costs. Figure 2 shows the boundaries of different rankings in the space of τ_a/τ_b and A_{nu}/A_{nr}.

For a given level of transportation frictions, τ_a/τ_b, we distinguish three regimes, depending on the productivity difference in the non-tradable sectors. When the productivity difference is either very low or very high, the urban biased allocation ϕ_p has the same direction as the centralized solution ϕ_c relative to the decentralized equilibrium ϕ_d. When A_{nu}/A_{nr} is small, the urban biased policy leads to a larger city than the decentralized one, overshooting the optimal solution ($\phi_p > \phi_c > \phi_d$). When A_{nu}/A_{nr} is very large, the urban biased policy yields a too small city relative to the decentralized equilibrium but larger than the unbiased central planner solution ($\phi_d > \phi_p \geq \phi_c$). This case would describe a situation of an urban biased policy that implements restrictions on the city growth, such as the well-known *hukou* household registration system in China. However, this urban biased policy is too restrictive relative to the unbiased planner's solution. In sharp contrast, at intermediate values of A_{nu}/A_{nr}, the urban biased policy goes against the optimal solution ($\phi_p > \phi_d > \phi_c$), resulting in an oversized city.

Figure 2: Comparison of decentralized, centralized and urban biased policies



Area below solid line: $\phi_c > \phi_d$. Area below dashed line $\phi_p > \phi_d$. $\tau_b = 1$, $\delta = 0.7$, $\theta = 0.5$. $A_a = 0.9$, $A_{nr} = 1$, $A_b = 1$.

Going back to our discussion on urbanization at different development stages, we note that urban biased policies yield allocations that move in the direction of optimal policies in advanced (A_{nu}/A_{nr} small) or very poor economies (A_{nu}/A_{nr} very large). However, while in a very developed country the urban bias could result in a too large city, in the least developed economies it could result in less severe restrictions on city size relative to the unbiased planner. At intermediate levels of development, the urban bias results in too large cities relative to the optimal solution, which in turn is lower than the decentralized city size.

8 Concluding remarks

We study decentralized and optimal urbanization in a simple multi-sector model focusing on productivity differences and transportation frictions in the context of trade between urban and rural areas. We show that even in the absence of the typical externalities studied in the literature, such as agglomeration, congestion or public goods, the decentralized city can be either too large or too small relative to that chosen by a benevolent planner. In particular, optimal urbanization exceeds decentralized levels when productivity differences in the production of location specific non-traded goods is small. A numerical exercise calibrated to Brazilian data suggests that the wedges can be quantitatively important.

Finally, we extend our model to incorporate urban-biased policies. We find that an urban biased planner will implement a city size that exceeds the optimal one when productivity differences in non-traded sectors are small, a situation likely to arise in more developed economies. In contrast, urban biased policies will result in restricted urbanization in developing economies where these productivity differences tend to be large.

Our theory uncovers important linkages between the industrial structure of an economy and the urbanization process and has stark policy implications that depend on the development stage of an economy. Nonetheless, further research is needed to better understand these policy implications, in particular by explicitly analyzing how the sector level productivity and transportation costs interact with the agglomeration and congestion externalities routinely studied in the urbanization literature.

References

Acemoglu, D. and F. Zilibotti (2001). Productivity differences. *The Quarterly Journal of Economics 116*(2), 563–606.

Agnosteva, Delina E., A. J. E. and Y. V. Yotov (2019). Intra-national trade costs: Assaying regional frictions. *European Economic Review 112*, 32 – 50.

Albouy, D., K. Behrens, F. Robert-Nicoud, and N. Seegert (2018). The optimal distribution of population across cities. *Journal of Urban Economics*.

Albrecht, L. and T. Trevor Tombe (2016). Internal trade, productivity, and interconnected industries: A quantitative analysis. *Canadian Journal of Economics 49*(1), 237 – 263.

Arnott, R. (1979). Optimal city size in a spatial economy. *Journal of Urban Economics 6*(1), 65 – 89.

Atkin, D. and D. Donaldson (2015). Who's getting globalized? the size and implications of intra- national trade costs. Technical report.

Beckmann, M. J. (1972). Von thunen revisited: A neoclassical land use model. *The Swedish Journal of Economics 74*(1), 1–7.

Berliant, Marcus, R. R. R. and P. Wang (2006). Knowledge exchange, matching, and agglomeration. *Journal of Urban Economics 60*(1), 69 – 95.

Bezemer, D. and D. Headey (2008). Agriculture, development, and urban bias. *World Development 36*(8), 1342 – 1364.

Black, D. and V. Henderson (1999). A theory of urban growth. *Journal of Political Economy 107*(2), 252–284.

Carlino, Gerald A., S. C. and R. M. Hunt (2007). Urban density and the rate of invention. *Journal of Urban Economics 61*(3), 389 – 419.

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *The American Economic Review 86*(1), 54–70.

Costinot, A. and A. Rodriguez-Clare (2014). *Chapter 4 - Trade Theory with Numbers: Quantifying the Consequences of Globalization*, Volume 4 of *Handbook of International Economics*, pp. 197 – 261. Elsevier.

Drazen, A. and Z. Eckstein (1988). On the organization of rural markets and the process of economic development. *The American Economic Review 78*(3), 431–443.

Duarte, M. and D. Restuccia (2019). Relative prices and sectoral productivity. Working papers, University of Toronto, Department of Economics.

Eaton, J. and Z. Eckstein (1997). Cities and growth: Theory and evidence from france and japan. *Regional Science and Urban Economics 27*(4), 443 – 474.

Fan, S. C. and O. Stark (2008). Rural-to-urban migration, human capital, and agglomeration. *Journal of Economic Behavior and Organization 68*(1), 234 – 247.

Frocrain, P. and P.-N. Giraud (2017, May). The evolution of tradable and non-tradable employment: evidence from France. Working papers, HAL.

Fuller, B. and P. Romer (2014). *Urbanization as Opportunity.* The World Bank.

Gautier, P. A. and C. N. Teulings (2009). Search and the city. *Regional Science and Urban Economics 39*(3), 251 – 265.

Gervais, A. and B. J. Jensen (2015). The tradability of services: Geographic concentration and trade costs, mimeo.

Glaeser, E. L. (1999). Learning in cities. *Journal of Urban Economics 46*(2), 254 – 277.

Glaeser, E. L. (2014). A world of cities: The causes and consequences of urbanization in poorer countries. *Journal of the European Economic Association 12*(5), 1154–1199.

Glomm, G. (1992). A model of growth and migration. *The Canadian Journal of Economics 25*(4), 901–922.

Henderson, J. (2010). Will the rural economy rebound in 2010? *Economic Review 95*(Q I), 95–119.

Henderson, J. V. (1974). The sizes and types of cities. *The American Economic Review 64*(4), 640–656.

Holmes, T. J. (2005). The location of sales offices and the attraction of cities. *Journal of Political Economy 113*(3), 551–581.

Imbs, J., C. Montenegro, and R. Wacziarg (2012). Economic integration and structural change. Technical report, Political Economy Seminar, Toulouse.

IPUMS (2018). Minnesota population center. integrated public use microdata series, international: Version 7.1 [dataset]. Technical report, Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D020.V7.1.

Itoh, R. (2009). Dynamic control of rural-urban migration. *Journal of Urban Economics 66*(3), 196 – 202.

Jensen, B. J. and L. G. Kletzer (2005). Tradable services: Understanding the scope and impact of services offshoring. *Brookings Trade Forum*, 75–133.

Kolko, J. (1999). Can i get some service here? information technology, service industries, and the future of cities.

Lall, Somik V., S. H. and Z. Shalizi (2006). *Rural-Urban Migration In Developing Countries : A Survey Of Theoretical Predictions And Empirical Findings*. The World Bank.

Lanjouw, J. O. and P. Lanjouw (2001). The rural non-farm sector: issues and evidence from developing countries. *Agricultural Economics 26*(1), 1–23.

Liedholm, C. and D. C. Mead (1987). Small scale industries in developing countries: Empirical evidence and policy implications. Technical report.

Lipton, M. (1977). *Why poor people stay poor : a study of urban bias in world development.* London : Canberra, ACT : Temple Smith ; Australian National University Press.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*(1), 3 – 42.

Lucas, R. E. (2001). Externalities and Cities. *Review of Economic Dynamics 4*(2), 245–274.

Lucas, R. E. (2004). Life earnings and rural?urban migration. *Journal of Political Economy 112*(S1), S29–S59.

Mills, E. S. and D. M. de Ferranti (1971). Market choices and optimum city size. *The American Economic Review 61*(2), 340–345.

Ott, I. and S. Soretz (2010). Productive public input, integration and agglomeration. *Regional Science and Urban Economics 40*(6), 538 – 549.

Poelhekke, S. and F. Van der Ploeg (2008). Globalization and the rise of mega-cities in the developing world. Technical report.

Porter, M. E., C. K. K. M. and R. Bryden (2004). Competitiveness in rural u.s. regions: Learning and research agenda. Technical report.

Reardon, Thomas, B. J. and G. Escobar (2001). Rural nonfarm employment and incomes in latin america: Overview and policy implications. *World Development 29*(3), 395 – 409.

Sheshinski, E. (1973). Congestion and the optimum city size. *The American Economic Review 63*(2), 61–66.

The World Urbanization Prospects (2018). *The World Urbanization Prospects 2018.* United Nations Department of Economic and Social Affairs, Population Division, United Nations.

Wiggins, S. and S. Proctor (2001). How special are rural areas? the economic implications of location for rural development. *Development Policy Review 19*(4), 427–436.

Yang, D. T. (1999). Urban-biased policies and rising income inequality in china. *American Economic Review 89*(2).

A Appendix A

Proof of Proposition 1.

1) Proof of existence and uniqueness of solution for p_a.

In equilibrium, p_a is the solution of the following equation:

$$\underbrace{\frac{A_b}{A_a p_a}}_{LHS(p_a)} = \underbrace{\left[\frac{(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a \frac{A_a}{A_{nr}}\right)^\gamma}{1 + (\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}\right]^{-\frac{1}{\gamma}}}_{RHS(p_a)}. \tag{A.1}$$

Denote $p_a = x$. For $\theta \in (0, 1]$, $\gamma = \theta/(\theta - 1) < 0$ and $\theta/(1 - \theta) = -\gamma > 0$. Thus, for $x \in (0, \infty)$, A.1 is equivalent with:

$$\left(\frac{A_b}{A_a}\right)^{-\gamma} x^\gamma = \frac{(\frac{\tau_b}{x})^\gamma + 1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma}{\left[1 + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right] (\frac{1}{x})^\gamma + (\tau_a)^\gamma}, \text{ or}$$

$$\underbrace{\left(\frac{A_b}{A_a}\right)^{-\gamma} (\tau_a x)^\gamma}_{LHS(x)} = \underbrace{(\frac{\tau_b}{x})^\gamma + 1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma - \left(\frac{A_b}{A_a}\right)^{-\gamma} \left[1 + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right]}_{RHS(x)}. \tag{A.2}$$

The $LHS(x)$ is monotonically decreasing in x with $lim_{x \to 0} LHS(x) = \infty$ and $lim_{x \to \infty} LHS(x) = 0$ and $RHS(x)$ is monotonically increasing in x with $lim_{x \to 0} RHS(x) = 1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma - \left(\frac{A_b}{A_a}\right)^{-\gamma} \left[1 + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right] < \infty$ and $lim_{x \to \infty} RHS(x) = \infty$. Thus, they intersect exactly once in the interval $(0, \infty)$.

2) Properties of p_a

We study equation (A.1). We note that $RHS(p_a)$ is decreasing in A_{nu}. From the properties of $RHS(p_a)$ and $LHS(p_a)$ studied above, it follows that when A_{nu} increases, $RHS(p_a)$ and $LHS(p_a)$ intersect at a higher point in the interval $(0, \infty)$.

By the same token, $RHS(p_a)$ is increasing in A_{nr}. From the properties of $RHS(p_a)$ and $LHS(p_a)$ studied above, it follows that when A_{nr} increases, $RHS(p_a)$ and $LHS(p_a)$ intersect at a lower point.

Similarly, we can show that equilibrium p_a is increasing in τ_b and decreasing in τ_a.

Proof of Proposition 2

We show that ϕ is increasing in p_a. Using the expression of ϕ, (23), rewrite

$$\phi = \frac{1}{1 + \frac{A_b (\tau_a)^\gamma}{A_a (\tau_b)^\gamma} (p_a)^{\gamma - 1} \frac{(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a \frac{A_a}{A_{nr}}\right)^\gamma}{1 + (\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}}.$$

Denote $p_a = x$ and $f(x) = \frac{A_b(\tau_a)^\gamma}{A_a(\tau_b)^\gamma} x^{\gamma-1} \frac{(\tau_b)^\gamma + x^\gamma + \left(x\frac{A_a}{A_{nr}}\right)^\gamma}{1+(\tau_a x)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}$. Thus,

$$\frac{\partial\phi}{\partial x} = -\frac{\partial f/\partial x}{(1+f(x))^2}, \tag{A.3}$$

where

$$\partial f/\partial x = (\gamma-1)\frac{f(x)}{x} + \frac{f(x)\gamma x^{\gamma-1}\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{(\tau_b)^\gamma + (x)^\gamma + \left(x\frac{A_a}{A_{nr}}\right)^\gamma} - \frac{\gamma f(x)(\tau_a)^\gamma x^{\gamma-1}}{1+(\tau_a x)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}, \text{ or}$$

$$\partial f/\partial x = (\gamma-1)\frac{f(x)}{x} + f(x)\gamma x^{\gamma-1}\frac{\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right] - (\tau_a)^\gamma(\tau_b)^\gamma}{\left[(\tau_b)^\gamma + (x)^\gamma + \left(x\frac{A_a}{A_{nr}}\right)^\gamma\right]\left[1+(\tau_a x)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right]}.$$

As $(\tau_a)^\gamma(\tau_b)^\gamma < 1$ and $\gamma < 0$, then $\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right] - (\tau_a)^\gamma(\tau_b)^\gamma > 0$. As $\gamma - 1 = 1/(\theta-1) < 0$ for $\theta < 1$, then $\partial f(x)/\partial x < 0$. Consequently, $\partial\phi/\partial x > 0$.

Proof of Proposition 3.

In order to prove existence of an interior solution in the interval $(0,1)$, we study the properties of welfare function when $\phi \to 0$ and $\phi \to 1$.

The aggregate welfare function is:

$$W(\phi) = \phi V^u(\phi) + (1-\phi)V^r(\phi),$$

where

$$\begin{aligned}
V^u(\phi) &= A_b\left[1+(\tau_a p_a(\phi))^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}} \\
V^r(\phi) &= A_a p_a(\phi)\left[(\tau_b)^\gamma + (p_a(\phi))^\gamma + \left(p_a(\phi)\frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}} \\
&= A_a\left[\left(\frac{\tau_b}{p_a(\phi)}\right)^\gamma + 1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}},
\end{aligned}$$

and $p_a(\phi)$ is given by (22).

Step 1. We study $\lim_{\phi\to 0} W(\phi)$ and $\lim_{\phi\to 1} W(\phi)$.

We study the first term of $W(\phi)$, $\phi V^u(\phi)$ and second term $(1-\phi)V^r(\phi)$ when $\phi \to 0$ and $\phi \to 1$.

Using (23), the first term $\phi V^u(\phi)$ can be written as:

$$\frac{A_aA_bp_a(\tau_b)^\gamma\left[1+(\tau_ap_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{\frac{\gamma-1}{\gamma}}}{A_ap_a(\tau_b)^\gamma\left[1+(\tau_ap_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]+A_b(\tau_a)^\gamma(p_a)^\gamma\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma\right]} \text{ or}$$

$$\frac{A_aA_b(\tau_b)^\gamma(p_a)^\gamma\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{\frac{\gamma-1}{\gamma}}}{A_a(p_a)^{1+\gamma}(\tau_b)^\gamma\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]+A_b(\tau_a)^\gamma(p_a)^{2\gamma}\left[(p_a)^{-\gamma}(\tau_b)^\gamma+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]} \text{ or}$$

$$\frac{A_aA_b(\tau_b)^\gamma\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{\frac{\gamma-1}{\gamma}}}{A_ap_a(\tau_b)^\gamma\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]+A_b(\tau_a)^\gamma(p_a)^\gamma\left[(p_a)^{-\gamma}(\tau_b)^\gamma+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}.$$

From equation (23) we can see that $\phi\to 0$ when $p_a\to 0$, and $\phi\to 1$ when $p_a\to\infty$. When $p_a\to 0$, $(p_a)^{-\gamma}\to 0$ and $(p_a)^\gamma\to\infty$ as $\gamma<0$. Thus $\lim_{\phi\to 0}\phi V^u(\phi)=0$. When $p_a\to\infty$, $(p_a)^\gamma\to 0$ as $\gamma<0$. Thus $\lim_{\phi\to 1}\phi V^u(\phi)=\frac{A_aA_b(\tau_b)^\gamma\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{\frac{\gamma-1}{\gamma}}}{A_a(\tau_b)^\gamma\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]+A_b(\tau_a)^\gamma(p_a)^{\gamma-1}(\tau_b)^\gamma}=$ $A_b\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}}>0$.

The second term of $W(\phi)$, $(1-\phi)V^r(\phi)$ is rewritten as:

$$\frac{A_aA_b(\tau_a)^\gamma(p_a)^{\gamma+1}\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma\right]^{1-\frac{1}{\gamma}}}{A_ap_a(\tau_b)^\gamma\left[1+(\tau_ap_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]+A_b(\tau_a)^\gamma(p_a)^\gamma\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma\right]} \text{ or}$$

$$\frac{A_aA_b(\tau_a)^\gamma(p_a)^{2\gamma}\left[(p_a)^{-\gamma}(\tau_b)^\gamma+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]^{1-\frac{1}{\gamma}}}{A_a(p_a)^{1+\gamma}(\tau_b)^\gamma\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]+A_b(\tau_a)^\gamma(p_a)^{2\gamma}\left[(p_a)^{-\gamma}(\tau_b)^\gamma+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]} \text{ or}$$

$$\frac{A_aA_b(\tau_a)^\gamma\left[(p_a)^{-\gamma}(\tau_b)^\gamma+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]^{1-\frac{1}{\gamma}}}{A_a(p_a)^{1-\gamma}(\tau_b)^\gamma\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]+A_b(\tau_a)^\gamma\left[(p_a)^{-\gamma}(\tau_b)^\gamma+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}.$$

When $p_a\to 0$, $(p_a)^{-\gamma}\to 0$ and $(p_a)^{1-\gamma}\to 0$ as $\gamma<0$. Thus $\lim_{\phi\to 0}(1-\phi)V^r(\phi)=$ $A_a\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}}>0$.
When $p_a\to\infty$, $(p_a)^\gamma\to 0$ as $\gamma<0$. Thus,
$\lim_{\phi\to 1}(1-\phi)V^r(\phi)=\frac{A_aA_b(\tau_a)^\gamma(p_a)^\gamma[(\tau_b)^\gamma]^{1-\frac{1}{\gamma}}}{A_a(\tau_b)^\gamma\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]+A_b(\tau_a)^\gamma(p_a)^{\gamma-1}(\tau_b)^\gamma}=0$.

Thus, $\lim_{\phi\to 0}W(\phi)=A_a\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}}>0$ and $\lim_{\phi\to 1}W(\phi)=A_b\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}}>0$.

Step 2. We study $\lim_{\phi\to 0}\partial W(\phi)/\partial\phi$ and $\lim_{\phi\to 1}\partial W(\phi)/\partial\phi$.

$$\frac{\partial W(\phi)}{\partial\phi}=V^u(\phi)+\phi\frac{\partial V^u(\phi)}{\partial\phi}-V^r(\phi)+(1-\phi)\frac{\partial V^r(\phi)}{\partial\phi}$$

First, we rewrite $\partial W(\phi)/\partial\phi$ as:

$$A_b\left[1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}}-\phi A_b\left[1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1}(\tau_a)^\gamma(p_a)^{\gamma-1}\frac{\partial p_a(\phi)}{\partial\phi}$$
$$-p_a A_a\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(\frac{p_a A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}}+(1-\phi)A_a\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(\frac{p_a A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}}\frac{\partial p_a(\phi)}{\partial\phi}$$
$$-(1-\phi)A_a(p_a)^\gamma\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(\frac{p_a A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1}\frac{\partial p_a(\phi)}{\partial\phi}\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right].$$

Further rearranging terms, we obtain $\partial W(\phi)/\partial\phi=$:

$$\underbrace{A_b\left[1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}}}_{g_1(\phi)}\underbrace{-\phi A_b\left[1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1}(\tau_a)^\gamma(p_a)^{\gamma-1}\frac{\partial p_a(\phi)}{\partial\phi}}_{g_2(\phi)}$$
$$\underbrace{-p_a A_a\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(\frac{p_a A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}}}_{g_3(\phi)}+\underbrace{(1-\phi)A_a(\tau_b)^\gamma\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(\frac{p_a A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1}\frac{\partial p_a(\phi)}{\partial\phi}}_{g_4(\phi)}.$$

Next, we calculate $\partial p_a(\phi)/\partial\phi$, which we will use to evaluate the limits of $g_2(\phi)$ and $g_4(\phi)$.

According to the implicit function theorem:

$$\frac{\partial p_a(\phi)}{\partial\phi}=-\frac{\partial Z/\partial\phi}{\partial Z/\partial p_a},$$

where $Z=\frac{(1-\phi)}{\phi}\frac{A_a(\tau_b)^\gamma}{A_b(\tau_a)^\gamma(p_a)^{\gamma-1}}-\frac{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}$. Furthermore:

$$\frac{\partial Z}{\partial\phi}=-\frac{1}{\phi^2}\frac{A_a(\tau_b)^\gamma}{A_b(\tau_a)^\gamma(p_a)^{\gamma-1}},$$

$$\frac{\partial Z}{\partial p_a}=\frac{(1-\phi)}{\phi}\frac{A_a(\tau_b)^\gamma(1-\gamma)}{A_b(\tau_a)^\gamma(p_a)^\gamma}+\gamma(\tau_a)^\gamma(p_a)^{\gamma-1}\frac{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}{\left[1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^2}-\frac{\gamma(p_a)^{\gamma-1}\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}.$$

We rewrite the second and third term in $\partial Z/\partial p_a$ using the fact that $Z=0$. Thus,
$\frac{\partial Z}{\partial p_a}=\frac{(1-\phi)}{\phi}\frac{A_a(\tau_b)^\gamma(1-\gamma)}{A_b(\tau_a)^\gamma(p_a)^\gamma}+\frac{(1-\phi)}{\phi}\frac{A_a(\tau_b)^\gamma}{A_b(\tau_a)^\gamma(p_a)^{\gamma-1}}\frac{\gamma(\tau_a)^\gamma(p_a)^{\gamma-1}}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}-\frac{(1-\phi)}{\phi}\frac{A_a(\tau_b)^\gamma}{A_b(\tau_a)^\gamma(p_a)^{\gamma-1}}\frac{\gamma(p_a)^{\gamma-1}\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}$

$$= \frac{(1-\phi)}{\phi}\frac{A_a(\tau_b)^\gamma}{A_b(\tau_a)^\gamma(p_a)^{\gamma-1}}\left[\frac{1-\gamma}{p_a}+\frac{\gamma(\tau_a)^\gamma(p_a)^{\gamma-1}}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}-\frac{\gamma(p_a)^{\gamma-1}\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}\right].$$

$$\frac{\partial p_a(\phi)}{\partial\phi}=\left\{\phi(1-\phi)(p_a)^{\gamma-1}\left[\frac{1-\gamma}{(p_a)^\gamma}+\frac{\gamma(\tau_a)^\gamma}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}-\frac{\gamma\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}\right]\right\}^{-1}. \tag{A.4}$$

We proceed in two steps.

A. We study $\lim_{\phi\to 0}\partial W(\phi)/\partial\phi$. When $\phi\to 0$, $(p_a)^\gamma\to\infty$, therefore $\lim_{\phi\to 0} g_1(\phi)=\infty$ and

$\lim_{\phi\to 0} g_3(\phi)=\lim_{\phi\to 0} A_a\left[(\tau_b)^\gamma(p_a)^{-\gamma}+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}}=A_a\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}}$.

Using the expression (A.4), we can write $g_2(\phi)$ as:

$$A_b(\tau_a)^\gamma\left[1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1}(1-\phi)^{-1}\times$$

$$\left[\frac{1-\gamma}{(p_a)^\gamma}+\frac{\gamma(\tau_a)^\gamma(p_a)^{\gamma-1}}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}-\frac{\gamma(p_a)^{\gamma-1}\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{(\tau_b)^\gamma+(p_a)^\gamma+\left(\frac{p_aA_a}{A_{nr}}\right)^\gamma}\right]^{-1} \quad \text{or}$$

$$A_b(\tau_a)^\gamma(p_a)^{-1-\gamma}\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1}(1-\phi)^{-1}\times$$

$$\times\left[\frac{1-\gamma}{(p_a)^\gamma}+\gamma(p_a)^{-\gamma-1}\frac{(\tau_a)^\gamma(\tau_b)^\gamma-\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[(\tau_b)^\gamma(p_a)^{-\gamma}+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}\right]^{-1} \quad \text{or}$$

$$A_b(\tau_a)^\gamma\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1}(1-\phi)^{-1}\times$$

$$\times\left[(1-\gamma)p_a+\gamma\frac{(\tau_a)^\gamma(\tau_b)^\gamma-\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{\left[(p_a)^{-\gamma}+(\tau_a)^\gamma+(p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[(\tau_b)^\gamma(p_a)^{-\gamma}+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}\right]^{-1}.$$

Thus, $\lim_{\phi\to 0} g_2(\phi)=0$.

Next, we write $g_4(\phi)$ as:

$$A_a(\tau_b)^\gamma \left[(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a \frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1} \phi^{-1}(p_a)^{\gamma+1} \times$$
$$\times \left[(1-\gamma)p_a + \gamma \frac{(\tau_a)^\gamma(\tau_b)^\gamma - \left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{\left[(p_a)^{-\gamma} + (\tau_a)^\gamma + (p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[(\tau_b)^\gamma(p_a)^{-\gamma} + 1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}\right]^{-1} \text{ or}$$
$$A_a(\tau_b)^\gamma \left[(\tau_b)^\gamma(p_a)^{-\gamma} + 1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1} \phi^{-1} \times$$
$$\times \left[(1-\gamma)p_a + \gamma \frac{(\tau_a)^\gamma(\tau_b)^\gamma - \left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{\left[(p_a)^{-\gamma} + (\tau_a)^\gamma + (p_a)^{-\gamma}\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[(\tau_b)^\gamma(p_a)^{-\gamma} + 1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}\right]^{-1}$$

We can see that $\lim_{\phi\to 0} g_4(\phi) = \infty$. Thus $\lim_{\phi\to 0} \partial W(\phi)/\partial\phi = \infty$.

B. We study $\lim_{\phi\to 1} \partial W(\phi)/\partial\phi$. When $\phi \to 1$, $(p_a)^\gamma \to 0$, therefore $\lim_{\phi\to 1} g_1(\phi) = A_b\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}}$,
$\lim_{\phi\to 1} g_3(\phi) = \infty$ and $\lim_{\phi\to 1} g_4(\phi) = 0$. Finally, we write $g_2(\phi)$ as:

$$A_b(\tau_a)^\gamma \left[1 + (\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right]^{-\frac{1}{\gamma}-1} (1-\phi)^{-1}(p_a)^\gamma \times$$
$$\times \left[1-\gamma+\gamma(p_a)^{-1} \frac{(\tau_a)^\gamma(\tau_b)^\gamma - \left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right]}{\left[1+(\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right]\left[(\tau_b)^\gamma + (p_a)^\gamma + \left(\frac{p_a A_a}{A_{nr}}\right)^\gamma\right]}\right]^{-1}.$$

We calculate
$\lim_{\phi\to 1}(1-\phi)^{-1}(p_a)^\gamma = \lim_{\phi\to 1} \frac{A_a p_a(\tau_b)^\gamma\left[1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]+A_b(\tau_a)^\gamma(p_a)^\gamma\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma\right]}{A_b(\tau_a)^\gamma\left[(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma\right]} = \infty$.

Thus, $\lim_{\phi\to 1} g_2(\phi) = \infty$. We conclude that $\lim_{\phi\to 0} \partial W(\phi)/\partial\phi = -\infty$.

Since $W(\phi)$ is differentiable on the interval $(0,1)$, these properties ensure the existence of an interior welfare maximum.

Proof of Proposition 4.

We compare $p_a(\phi_d)$ to A_b/A_a. When $\phi = \phi_d$, the relative price p_a satisfies the utility equalization condition (A.2) studied in Proposition 1.

We evaluate the $LHS(p_a)$ and $RHS(p_a)$ of this expression at $p_a = A_b/A_a$ to check when $p_a(\phi_d) < A_b/A_a$ or vice-versa.

$LHS\left(\frac{A_b}{A_a}\right) = (\tau_a)^\gamma$ and

$RHS\left(\frac{A_b}{A_a}\right) = (\tau_b)^\gamma\left(\frac{A_b}{A_a}\right)^{-\gamma} + 1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma - \left(\frac{A_b}{A_a}\right)^{-\gamma}\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]$

When $p_a(\phi_d) < A_b/A_a$, $LHS\left(\frac{A_b}{A_a}\right) < RHS\left(\frac{A_b}{A_a}\right)$, as $LHS(p_a)$ is decreasing over $(0,\infty)$ and $RHS(p_a)$ is increasing.

The condition is $(\tau_a)^\gamma < (\tau_b)^\gamma \left(\frac{A_b}{A_a}\right)^{-\gamma} + 1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma - \left(\frac{A_b}{A_a}\right)^{-\gamma}\left[1 + \left(\frac{A_b}{A_{nu}}\right)^\gamma\right]$, or
$(\tau_a)^\gamma (A_a)^{-\gamma} < (\tau_b)^\gamma (A_b)^{-\gamma} + (A_a)^{-\gamma} + (A_{nr})^{-\gamma} - \left[(A_b)^{-\gamma} + (A_{nu})^{-\gamma}\right]$.
Rearranging terms, we obtain:

$$(A_{nu})^{-\gamma} - (A_{nr})^{-\gamma} < (A_a)^{-\gamma}\left[1 - (\tau_a)^\gamma\right] - (A_b)^{-\gamma}\left[1 - (\tau_b)^\gamma\right].$$

If the condition above holds, then $\frac{A_b}{A_a} > p_a$ and $\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d} > 0$, so planner wants a larger city, i.e. $\phi_c > \phi_d$.

The opposite holds when $RHS\left(\frac{A_b}{A_a}\right) < LHS\left(\frac{A_b}{A_a}\right)$. Then, $\left.\frac{\partial W}{\partial \phi}\right|_{\phi=\phi_d} < 0$, so the planner wants a smaller city, $\phi_c < \phi_d$.

Proof of Proposition 6.

We evaluate the sign of $\left.\frac{\partial W^p}{\partial \phi}\right|_{\phi=\phi_d}$. If $\left.\frac{\partial W^p}{\partial \phi}\right|_{\phi=\phi_d}$, then $\phi_p > \phi_d$, where ϕ_p is the solution of $\frac{\partial W^p}{\partial \phi} = 0$.

$$\left.\frac{\partial W^p(\phi)}{\partial \phi}\right|_{\phi=\phi_d} = \left.\left[\delta V^u(\phi) + \phi\delta\frac{\partial V^u(\phi)}{\partial \phi} - (1-\delta)V^r(\phi) + (1-\delta)(1-\phi)\frac{\partial V^r(\phi)}{\partial \phi}\right]\right|_{\phi=\phi_d}.$$

$$\begin{aligned}\left.\frac{\partial W^p}{\partial \phi}\right|_{\phi=\phi_d} &= \phi_d \left.\frac{\partial V^u(\phi)}{\partial \phi}\right|_{\phi=\phi_d} + (1-\phi_d)\left.\frac{\partial V^r(\phi)}{\partial \phi}\right|_{\phi=\phi_d} \qquad \text{(A.5)}\\ &= -V^r(\phi_d)\frac{(1-\phi_d)(p_a)^{\gamma-1}}{(\tau_b)^\gamma + (p_a)^\gamma + \left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}\left[1 + \left(\frac{A_a}{A_{nr}}\right)^\gamma\right]\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d}\\ &+ V^r(\phi_d)\frac{(1-\phi_d)}{p_a}\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d}\\ &- V^u(\phi_d)\frac{\phi_d(p_a)^{\gamma-1}(\tau_a)^\gamma}{1 + (\tau_a p_a)^\gamma + \left(\frac{A_b}{A_{nu}}\right)^\gamma}\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d},\end{aligned}$$

As $V^u(\phi_d) = V^r(\phi_d)$, we can write $\left.\frac{\partial W^p(\phi)}{\partial \phi}\right|_{\phi=\phi_d}$ as

$$\begin{aligned}&\left.\left[\phi\delta\frac{\partial V^u(\phi)}{\partial \phi} + (1-\delta)(1-\phi)\frac{\partial V^r(\phi)}{\partial \phi}\right]\right|_{\phi=\phi_d} + (2\delta-1)V^u(\phi_d)\\ = \;& V^u(\phi_d)\left\{\begin{array}{c} -\delta\phi_d\frac{(\tau_a)^\gamma(p_a)^{\gamma-1}}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma} - \\ -(1-\delta)(1-\phi_d)\frac{\left[1+\left(\frac{A_a}{A_{nr}}\right)^\gamma\right](p_a)^{\gamma-1}}{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma} + (1-\delta)(1-\phi_d)\frac{1}{p^a}\end{array}\right\}\left.\frac{\partial p_a(\phi)}{\partial \phi}\right|_{\phi=\phi_d}\\ &+(2\delta-1)V^u(\phi_d).\end{aligned}$$

Rearranging terms, we obtain $\left.\frac{\partial W^p}{\partial \phi}\right|_{\phi=\phi_d}$ as:

$$V^u(\phi_d)\left\{\begin{array}{c}-\delta\phi_d\frac{(\tau_a)^\gamma(p_a)^{\gamma-1}}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}\\ +(1-\delta)(1-\phi_d)\frac{1}{(p^a)}\frac{(\tau_b)^\gamma}{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}\end{array}\right\}\left.\frac{\partial p_a(\phi)}{\partial\phi}\right|_{\phi=\phi_d}+(2\delta-1)V^u(\phi),\ \text{or}$$

$$\underbrace{\frac{V^u(\phi_d)(p_a)^{\gamma-1}}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}}_{D>0}\left\{\begin{array}{c}-\delta\phi_d(\tau_a)^\gamma\\ +(1-\delta)(1-\phi_d)\frac{(\tau_b)^\gamma}{(p^a)^\gamma}\frac{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}\end{array}\right\}\left.\frac{\partial p_a(\phi)}{\partial\phi}\right|_{\phi=\phi_d}+(2\delta-1)V^u(\phi)$$

We use the fact that ϕ_d satisfies the utility equalization condition: $\frac{(1-\phi_d)}{\phi_d}\frac{A_a(\tau_b)^\gamma}{A_b(\tau_a)^\gamma(p_a)^{\gamma-1}}=\frac{(\tau_b)^\gamma+(p_a)^\gamma+\left(p_a\frac{A_a}{A_{nr}}\right)^\gamma}{1+(\tau_a p_a)^\gamma+\left(\frac{A_b}{A_{nu}}\right)^\gamma}$.

We obtain:

$$\left.\frac{\partial W^p}{\partial\phi}\right|_{\phi=\phi_d}=\frac{\delta D\phi_d(\tau_a)^\gamma}{p_a}\left\{(\frac{1-\delta}{\delta}\frac{A_b}{A_a}-p_a(\phi_d)\right\}\left.\frac{\partial p_a(\phi)}{\partial\phi}\right|_{\phi=\phi_d}+(2\delta-1)V^u(\phi).$$

As $2\delta-1>0$, a sufficient condition for $\left.\frac{\partial W^p}{\partial\phi}\right|_{\phi=\phi_d}>0$ is $\frac{1-\delta}{\delta}\frac{A_b}{A_a}-p_a(\phi_d)>0$.

We use the fact that $p_a(\phi_d)$ satisfies the utility equalization condition ($A.2$).

$$\underbrace{\left(\frac{A_b}{A_a}\right)^{-\gamma}(\tau_a x)^\gamma}_{LHS(x)}=\underbrace{(\frac{\tau_b}{x})^\gamma+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma-\left(\frac{A_b}{A_a}\right)^{-\gamma}\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]}_{RHS(x)}$$

We evaluate the $LHS(p_a)$ and $RHS(p_a)$ of this expression at $(1-\delta)A_b/\delta A_a$ to check when $p_a(\phi_d)<(1-\delta)A_b/\delta A_a$.

$LHS\left(\frac{1-\delta}{\delta}\frac{A_b}{A_a}\right)=\left(\frac{1-\delta}{\delta}\right)^{-\gamma}(\tau_a)^\gamma$ and

$RHS\left(\frac{1-\delta}{\delta}\frac{A_b}{A_a}\right)=(\tau_b)^\gamma\left(\frac{1-\delta}{\delta}\frac{A_b}{A_a}\right)^{-\gamma}+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma-\left(\frac{A_b}{A_a}\right)^{-\gamma}\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]$

$LHS(p_a)$ is decreasing over $(0,\infty)$ and $RHS(p_a)$ is increasing. Thus, when $p_a(\phi_d)<(1-\delta)A_b/\delta A_a$,

$LHS\left(\frac{1-\delta}{\delta}\frac{A_b}{A_a}\right)<RHS\left(\frac{1-\delta}{\delta}\frac{A_b}{A_a}\right)$.

The condition is $(\tau_a)^\gamma\left(\frac{1-\delta}{\delta}\right)^{-\gamma}<(\tau_b)^\gamma\left(\frac{1-\delta}{\delta}\frac{A_b}{A_a}\right)^{-\gamma}+1+\left(\frac{A_a}{A_{nr}}\right)^\gamma-\left(\frac{A_b}{A_a}\right)^{-\gamma}\left[1+\left(\frac{A_b}{A_{nu}}\right)^\gamma\right]$.

Denote $\kappa=(1-\delta)/\delta<1$, when $\delta>1/2$. The condition above becomes:

$(\tau_a)^\gamma(\kappa A_a)^{-\gamma}<(\tau_b)^\gamma(\kappa A_b)^{-\gamma}+(A_a)^{-\gamma}+(A_{nr})^{-\gamma}-\left[(A_b)^{-\gamma}+(A_{nu})^{-\gamma}\right]$.

Rearranging terms, we obtain:

$$(A_{nu})^{-\gamma}-(A_{nr})^{-\gamma}<(A_a)^{-\gamma}\left[1-(\kappa/\tau_a)^{-\gamma}\right]-(A_b)^{-\gamma}\left[1-(\kappa/\tau_b)^{-\gamma}\right].$$

$$(A_{nu})^{-\gamma}-(A_{nr})^{-\gamma}<(A_a)^{-\gamma}-(A_b)^{-\gamma}+\kappa^{-\gamma}\left[\left(\frac{A_b}{\tau_b}\right)^{-\gamma}-\left(\frac{A_a}{\tau_a}\right)^{-\gamma}\right].$$

When $p_a(\phi_d) < (1-\delta)A_b/\delta A_a$, then $\left.\frac{\partial W^p}{\partial \phi}\right|_{\phi=\phi_d} > 0$. Thus, $\phi_p > \phi_d$.

When $\delta \in (1/2, 1)$, $(1-\delta)A_b/\delta A_a < A_b/A_a$. When $p_a(\phi_d) < A_b/A_a$, $\left.\frac{\partial W^c}{\partial \phi}\right|_{\phi=\phi_d} > 0$ and thus $\phi_c > \phi_d$.