Trade and urbanization: Evidence from Hungary*

Dávid Krisztián Nagy†

August 16, 2018

Abstract

I study how trade affects urbanization and welfare. To guide my investigation, I first develop a quantitative model of economic geography in which benefits from trading drive agglomeration around locations where trading activity takes place. As a result, increasing trade leads to urbanization and welfare gains. The model provides a simple formula according to which the degree of urbanization around trading locations is a sufficient statistic for the real income gains from trade. Next, I estimate the model using exogenous variation in trade due to the redrawing of Hungary's borders after the First World War. Besides explaining the decrease in urbanization near the country's new borders, the model also provides a tool to measure real income losses at any location, which are unobserved in the data. I find that the effects of the new borders on urbanization and real income are substantially heterogeneous across locations, due to the rich geography of frictions to trade and labor mobility.

1 Introduction

Trade is a key driver of the spatial distribution of population and economic activity. Locations with good access to trade, such as harbors, rivers and valleys, tend to have higher productivity, more firms, more people and higher income per capita. Good trading opportunities led to the rise of many large cities in history, such as Cairo, New York or Mumbai. As economies developed and self-sufficiency was gradually replaced by large-scale trade, these trading cities attracted more and more people, thus contributing to urbanization and allowing people to reap the benefits from both agglomeration and trade.

The extent to which trade can induce urbanization and welfare benefits, however, depends on spatial frictions, such as frictions to the movement of labor and the costs of

*I am grateful to Esteban Rossi-Hansberg for guidance, as well as to Treb Allen, Costas Arkolakis, Oleg Itskhoki, Réka Juhász, Miklós Koren, Jan de Loecker, Ildikó Magyari, Giacomo Ponzetto, Stephen Redding and seminar participants at Princeton University and the Barcelona Summer Forum for their helpful comments and suggestions. All errors are my own.

†CREI, Universitat Pompeu Fabra and Barcelona GSE. dnagy@crei.cat.

trading goods. If trade in goods is very costly, then even cities at the most advantageous locations can gain little from their position. If labor is largely immobile, then even dramatic increases in trade have little impacts on urbanization. Moreover, trade might induce urbanization but urbanization might also lead to trade as large cities can exploit economies of scale and specialize in a subset of goods. As a consequence, isolating the effect of trade on urbanization from its reverse requires looking for exogenous variation in trade.

To study the impact of trade on urbanization and welfare, I propose a quantitative model that incorporates a rich geography of frictions to the movement of goods and labor. In the model, a country is inhabited by a large number of workers. Each worker produces a specific good. Workers demand all goods produced in the economy, which gives rise to trade among them. Trade can take place at a subset of locations which I call *trading places*. Workers simultaneously choose a residential location where they live and a trading place where they trade. When making these choices, they keep in mind their real income from trading, the cost of shipping to and from the trading place, and the mobility frictions arising from their idiosyncratic tastes for residential locations. A worker trading at a place with good trading opportunities has an incentive to save on shipping costs, hence she is more likely to choose a residential location near the trading place than a location she likes for idiosyncratic reasons. This suggests that border regions of a country, by offering worse trading opportunities, must exhibit less urbanization around trading places, a key prediction that I verify in the data.

Next, I combine the theoretical model with Hungary's border changes after the First World War as a source of exogenous variation in trade. Drawn by the Allied Powers in 1920, Hungary's post-war borders offer a laboratory to study the effect of trade on urbanization since they had a dramatic effect on trade between the locations they isolated. On the one hand, historical evidence suggests that the post-war borders did not correspond to prior political, economic or ethnic boundaries (Kontler 2002, Teleki 1923). Hence, it is reasonable to assume that trade between the two sides of post-war borders was not subject to any frictions before the war. On the other hand, Hungary had hardly any interactions with its new neighbors in the years following the war (Teleki, 1923), making trade across the new borders essentially non-existent. Combining these large exogenous changes in trade with the model, I estimate the effect of trade on urbanization and real income.

I find that the model fits the data well, featuring a correlation with the predicted population distribution above 0.6 before and above 0.7 after the change in borders. Conforming to the model's main prediction on the positive effect of trade on urbanization, the data show that urbanization in regions near the new borders decreased relative to regions farther away.[1] In fact, the model can quantitatively replicate the difference in urbanization between border and interior regions. Finally, I use the model as a tool to measure the effects on real income, which are unobserved in the data. I find that the post-war borders

[1] I show that the trends of urbanization were identical between these two regions prior to the war.

had a large and heterogeneous impact on Hungarian locations, with the 2nd percentile of real income losses being at 13.9% and the 98th percentile at 17.6%.

This paper is related to three strands of the literature. First, it is related to recent quantitative models of trade and geography, such as Allen and Arkolakis (2014), Desmet et al. (2018), Monte et al. (2018), Redding (2016) and Redding and Sturm (2008).[2] These models can tractably accommodate a large number of locations that are asymmetric in their characteristics, such as in their geographic location and amenities. They can also tractably incorporate frictions to trade and the mobility of labor. Despite introducing such rich heterogeneity, these models do not depart from the standard assumption that every location hosts some trading activity. In reality, however, trade is often concentrated even on top of the large spatial concentration of population and economic activity. For instance, the Hungarian census reports that cities above 20,000 inhabitants hosted as much as 71.7% of employment in "trade and finance" even though they only hosted 33.8% of the country's population in 1930. My main contribution to this literature is that I acknowledge this empirical fact by assuming that only a subset of locations – trading places – host trading activities. In equilibrium, this leads to the concentration of population around trading places that offer good trading opportunities, creating a novel and empirically testable link between trade and urbanization. I find strong support for this novel link in the data.

The mechanism that agglomeration arises around trading locations relates the paper to a set of papers that show the presence of this force around locations that are suitable for certain specific trading activities. Bleakley and Lin (2012) point to the key role that portage sites played in selecting the location of many U.S. cities. Armenter et al. (2014) suggest that bridges played a role in U.S. city formation, while Coşar and Fajgelbaum (2016) and Fajgelbaum and Redding (2014) argue that ports acted as focal points of the concentration of economic activity in China and Argentina, respectively. Within this literature, my paper is closest to Fajgelbaum and Redding (2014) as I also develop a quantitative model to study concentration around trading locations. However, unlike Fajgelbaum and Redding (2014), my framework not only allows for trade between a trading place and the rest of the world, but also for trade across trading places. This gives a disadvantage to trading places near borders and thus implies that urbanization is lower in border regions. I use this key prediction, which would be absent without trade across trading places, together with the redrawing of Hungary's borders to estimate the model and learn about the quantitative effect of trade on urbanization.

Finally, the paper is also related to the literature studying the effect of countries' trade openness on agglomeration more generally. An early paper in this literature is Krugman and Livas Elizondo (1996), who suggest in a stylized three-location model that closed economies are more likely to give rise to agglomeration through linkages between consumers and firms.

[2] See Redding and Rossi-Hansberg (2017) for a survey of this rapidly growing literature.

Brülhart (2011), however, argues in a survey of the literature that this prediction is specific to the particular three-location setting and other stylized frameworks in fact deliver the opposite conclusion. Unlike these papers, I develop a multi-location framework in which I can quantitatively measure the effect of trade on agglomeration without the need to assume a stylized geography.

The structure of the paper is as follows. Section 2 presents the model, while Section 3 describes the data and provides suggestive empirical evidence for the model's main mechanism: the urbanization-fostering effect of trade. Section 4 provides details on the structural estimation I conduct to take the model to the data, a method similar to Ahlfeldt et al. (2015). Section 5 presents the results of the estimation. Section 6 concludes.

2 A model of trade and urbanization in space

2.1 Setup

A country S consists of a finite number of *locations* $r \in S$. The country is populated by $\overline{L}$ *workers*, each of whom produces a specific good that everyone views as different from the goods produced by other workers.[3] Production of goods requires labor only, and each worker is endowed with a fixed amount of labor that I normalize to one. Goods are tradable within the country subject to shipping costs. Goods are not tradable with the rest of the world. Hence, the borders of the country constitute an impassable barrier to trade.[4]

Goods can be traded at a subset of locations $\mu_1, \ldots, \mu_M \in S$, which I call *trading places*. Workers simultaneously choose a residential location r where they live, consume and produce, and a trading place m where they sell their product and buy the products of others.

2.1.1 Consumption

Workers are heterogeneous in their location tastes. Worker i, if chooses to live at location r and trade at trading place m, obtains utility

$$u_m(r,i) = a(r,i) + \left[\sum_{j=1}^{\overline{L}} c_m^j(r,i)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{1}$$

[3] Although the assumption that workers produce at home may be relevant in historical contexts, it is at odds with reality today. However, Appendix B presents a model with firms that employ workers to produce goods, and shows a formal isomorphism between the two models.

[4] In Section 4, I argue that the assumption of impassable borders is a good approximation to reality in Hungary after the First World War.

where $a(r,i)$ is the level of amenities that the worker consumes at her residential location, $c_m^j(r,i)$ is the worker's consumption of the product of worker j, and σ is the elasticity of substitution across goods. In what follows, I assume $\sigma > 1$, that is, goods are substitutes.

Amenities reflect location-specific features that increase any resident's wellbeing (such as the location having a nice view), but also idiosyncratic factors that might only be beneficial to some (such as family ties). In particular, they take the form

$$a(r,i) = a(r) + \varepsilon(r,i)$$

where $a(r)$ is the part of amenities that is common across workers, and $\varepsilon(r,i)$ is an idiosyncratic amenity shifter that represents heterogeneity across workers in their tastes for different locations. I assume that $\varepsilon(r,i)$ is iid across both workers and locations, and is distributed Gumbel:

$$\Pr(\varepsilon(r,i) \leq z) = e^{-e^{-z/\theta}}$$

θ is a positive constant that drives the degree of heterogeneity in idiosyncratic location tastes, and thus the dispersion of population in equilibrium.[5] As $\theta \to \infty$, heterogeneity in tastes becomes large enough such that each location hosts the same number of workers, irrespectively of the distribution of $a(r)$ and prices. On the other hand, as $\theta \to 0$, all workers draw the same $\varepsilon(r,i)$, hence heterogeneity in tastes disappears, and all workers choose the residential location that offers them the best combination of $a(r)$ and access to tradables. As a result, θ can also be viewed as a parameter showing the severity of frictions to labor mobility.

2.1.2 Production and shipping

Producing a unit of a good requires one unit of labor. Shipping the good from the residential location, that is, its place of production, to a trading place is subject to an "iceberg" type cost. In particular, if $\varsigma_1(\mu_m, r) \geq 1$ units are sent from residential location r to trading place m, only one unit arrives, while the remaining $\varsigma_1(\mu_m, r) - 1$ units melt away in transit.

Having shipped their goods to the trading place, workers engage in monopolistic competition. That is, worker j chooses the price of her product p_m^j, but takes the CES price index of all goods, P_m, as given. She also takes into account that shipping goods across trading places is possible, but also subject to an iceberg cost $\tau(\mu_m, \mu_o) \geq 1$. To keep the model tractable, I assume that these shipping costs are symmetric: $\tau(\mu_m, \mu_o) = \tau(\mu_o, \mu_m)$ for all m and o. The worker is also subject to a shipping cost when bringing a good she purchased home from the trading place. If she buys $\varsigma_2(\mu_m, r) \geq 1$ units at m, she is left with one unit at her residential location r.

[5] For similar formulations of the dispersion force in economic geography models, see Redding (2016), Desmet et al. (2018) or Appendix A.2 in Allen and Arkolakis (2014).

Shipping costs, as well as the fact that workers demand all goods, are responsible for the force of agglomeration in the model. Unless shipping costs are zero or infinitely high, workers have an incentive to move close to each other so they can save on shipping costs. This implies, first, that workers tend to choose more centrally located trading places. Second, it also implies that they tend to live close to the trading place they choose. In equilibrium, this agglomeration force is counterbalanced by the dispersion force coming from workers' tastes for certain idiosyncratic locations.

2.1.3 Equilibrium

Due to the additive separability of utility in amenities and tradables as well as the fact that amenities do not depend on the worker's trading place, workers who live at the same residential location r all choose the same place to trade at. Let us denote this trading place by $\mu(r)$. Given this, I define an equilibrium of the economy below.

Definition 1 *Given parameters* $\{\sigma, \theta, \overline{L}\}$*, geography* S*,* $\{\mu_1, \ldots, \mu_M\}$ *and functions* $a : S \to \mathbb{R}_+$*,* $\{\tau, \varsigma_1, \varsigma_2\} : S^2 \to \mathbb{R}_+$*, an equilibrium of the economy consists of a population distribution* $L : S \to \mathbb{R}_+$*; consumption levels* $c : [0, \overline{L}]^2 \times S \times \{1, \ldots, M\} \to \mathbb{R}_+$*; goods' prices and production levels* $\{p, x\} : [0, \overline{L}] \times \{1, \ldots, M\} \to \mathbb{R}_+$*; and a function that assigns a trading place to each residential location,* $\mu : S \to \{1, \ldots, M\}$*, such that the following hold:*

1. *Workers choose their consumption, production, price, residential location and trading place to maximize their utility (1) subject to the production technology and their budget constraint.*

2. *The market for each good clears at every trading place, implying*

$$x_m^j = \sum_o \tau(\mu_m, \mu_o)^{1-\sigma} \left(p_m^j\right)^{-\sigma} P_o^{\sigma-1} p_o L_o \tag{2}$$

   for any worker j*, where* x_m^j *denotes the worker's production level,* m *denotes the trading place where she sells her product, and* L_o *is the total number of workers trading at trading place* o*.*[6]

Appendix A shows that the spatial distribution of population is governed by the following system of equations in equilibrium:

$$\log L(r) = \nu + \theta^{-1}\left[a(r) + \varsigma(\mu(r), r)^{-1} \omega_{\mu(r)}\right] \tag{3}$$

$$\varsigma(\mu(r), r)^{-1} \omega_{\mu(r)} \geq \varsigma(\mu_m, r)^{-1} \omega_m \qquad \forall m \tag{4}$$

[6]The right-hand side of equation (2) follows from CES demand for worker j's product at any trading place o.

$$\omega_m^{\frac{\sigma(\sigma-1)}{2\sigma-1}} = \sum_o \omega_o^{-\frac{(\sigma-1)^2}{2\sigma-1}} L_o \tau\left(\mu_m, \mu_o\right)^{1-\sigma} \quad (5)$$

$$L_m = \sum_{r:\ m=\mu(r)} L\left(r\right) \quad (6)$$

where $\varsigma\left(\mu_m, r\right) = \varsigma_1\left(\mu_m, r\right)\varsigma_2\left(\mu_m, r\right)$, $L\left(r\right)$ is the population of residential location r, ν is a combination of parameters, and $\omega_m = \frac{p_m}{P_m}$ is the real income of a worker trading at trading place m.

To gather intuition for equations (3) to (6), note that equation (3) determines population at r as an increasing function of local amenities $a\left(r\right)$, a decreasing function of shipping costs to and from the trading place $\varsigma\left(\mu\left(r\right), r\right)$, and an increasing function of real income at the trading place, $\omega_{\mu(r)}$. Equation (4) shows how the choice of trading places takes place in equilibrium: workers at r choose the trading place that offers the best combination of proximity $\varsigma\left(\mu_m, r\right)^{-1}$ and real income ω_m. Equation (5) relates real income at m to the *market access* of m, which measures trading opportunities at m as it depends on the real income and population levels of neighboring trading places:

$$MA_m = \sum_o \omega_o^{-\frac{(\sigma-1)^2}{2\sigma-1}} L_o \tau\left(\mu_m, \mu_o\right)^{1-\sigma}. \quad (7)$$

Finally, equation (6) simply states that the size of each trading place m is equal to the total number of people who choose to trade at m.

Equation (3) sheds light on an important feature of the equilibrium. Take two residential locations r and s that have the same level of amenities, and from which consumers ship to the same trading place. Assume that consumers' shipping costs are an increasing function of distance. Then, if r is closer to the trading place than s, we have $\varsigma\left(\mu\left(r\right), r\right) < \varsigma\left(\mu\left(r\right), s\right)$, hence equation (3) implies $L\left(r\right) > L\left(s\right)$. That is, population decreases with distance from the trading place: *cities with a negative population gradient* form around trading places.

The next section uses equations (3), (5) and (7) to explore the relationship among trade, urbanization and real income in the model.

2.2 Urbanization and the gains from trade

In this section, I use the model outlined in Section 2.1 to define an intuitive measure of urbanization around trading places. Next, I show in Propositions 1 and 2 that the measure is not only intuitive but is also related to two famous objects in trade: market access and the gains from trade. These propositions allow me to qualitatively explore the relationship between urbanization and trade.

Definition 2 *The* ***urbanization index*** *at residential location r is the gradient of log* *population with respect to proximity (inverse shipping costs) to the trading place,*

$$UI(r) = \frac{\partial \log L(r)}{\partial \varsigma(\mu(r), r)^{-1}}.$$

If $UI(r)$ is large, then the gradient of the population distribution is steep, indicating that the close neighborhood of trading place $\mu(r)$ is highly urbanized. On the other hand, a low value of $UI(r)$ suggests that the population distribution is very dispersed in the surroundings of $\mu(r)$, hence the level of urbanization is low around $\mu(r)$.

The following proposition relates the urbanization index of a location to the market access of its trading place.

Proposition 1 *The urbanization index of* r *is related to the market access of* $\mu(r)$ *according to the equation*

$$UI(r) = \theta^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}}.$$

Proof. *Partially differentiating equation (3) with respect to* $\varsigma(\mu(r), r)^{-1}$ *and using equation (7) to substitute for* $\omega_{\mu(r)}$ *gives the result.* ■

Proposition 1 thus states that trading places with good trading opportunities (good market access) have more urbanized surroundings. This result follows from the trade-off between agglomeration and dispersion forces that shape the population distribution in the model. The force of agglomeration gives incentives for people to live close to their trading place, as this allows them to save on shipping costs. The force of dispersion, coming from people's idiosyncratic tastes for locations, counterbalances this agglomeration force in equilibrium. However, the force of agglomeration is naturally stronger around trading places that offer good trading opportunities. As a result, the population distribution is more concentrated and the degree of urbanization is higher around these trading places.

Note that, by equations (5) and (7), the urbanization index is a linear function of real income at $\mu(r)$:

$$UI(r) = \theta^{-1} \omega_{\mu(r)}$$

This allows me to relate the gains from trade to changes in the urbanization index in the following proposition.

Proposition 2 *Assume a change in trade due to an exogenous change in trade costs or country borders. Define the gains from trade at residential location* r *as the percentage change in the real income of location* r*'s residents,*

$$GFT(r) = \frac{\omega'_{\mu'(r)}}{\omega_{\mu(r)}} - 1$$

where variables with a prime indicate variables after the change in trade costs or borders. Then we have

$$GFT(r) = \frac{UI'(r)}{UI(r)} - 1.$$

That is, the change in the urbanization index is a sufficient statistic for the gains from trade.

The relationship between the gains from trade and urbanization is also intuitive. As trading opportunities increase, people move closer to their trading place to reap the benefits from increased trade. The extent to which they move closer, which is captured by the change in the urbanization index, conveys information about the gains from trade.

Proposition 2 immediately implies that borders, by imposing barriers on trade, are likely to negatively affect both urbanization and welfare. This effect must be especially pronounced in regions near the border, as these regions are likely to suffer the largest loss in their trading opportunities. Note that the opposite of this prediction arises in standard trade and geography models that assume that every location is a trading place, as Allen and Arkolakis (2014) or Redding and Sturm (2008). Such models predict that small border locations, such as villages, lose a larger fraction of their market access due to the border than large border locations such as cities. As a result, in these other models, cities lose less of their population and hence urbanization becomes higher near borders. This makes the urbanization-hindering effect of borders a novel prediction of my framework. In the remaining sections of the paper, I verify this prediction by studying the effect of a large-scale exogenous change in borders: the redrawing of Hungary's borders by the Allied Powers after the First World War.

3 Data and suggestive empirical evidence

3.1 Data

Hungary underwent dramatic changes in its borders due to the Treaty of Trianon in 1920: its area shrinked from 325,000 km^2 to 93,000 km^2, and its population fell from 20.9 million to 7.6 million (see Figure 1). These border changes were unexpected until 1918, the end of the First World War. Although Hungary was a multi-ethnic state before 1920 and ethnic minorities claimed for autonomy or occasionally even independence, the new borders were largely unrelated to political, economic and ethnic boundaries; see Kontler (2002) or Teleki (1923).

I use data from the 1910 and 1930 population censuses to examine the effects that changes in trade due to the change in borders had on the distribution of population. These censuses provide the number of inhabitants in each settlement of the country, as well as each settlement's area in square kilometers. Every location in space belongs to one and only one settlement; in other words, settlements' areas constitute a partition of the country's territory. Unfortunately, there are no data on settlements' actual geographic boundaries. Although the censuses themselves classify settlements into cities (*város*), towns (*nagyközség*) and villages (*kisközség*), I do not rely on this classification because it is largely based on history, with settlements with one or two thousand inhabitants that had once been important places being called cities but some with much larger population classified as towns or even as villages. Moreover, the classification changed substantially before the 1930 census for political reasons. Instead, I regard any settlement with more than 20,000 inhabitants as a city.[7]

Motivated by the model's prediction that cities form around trading places in equilibrium, I assume that each city had a trading place in its geographic center. Although this assumption is clearly an abstraction from reality in which trade can potentially happen at various locations, the censuses in fact support it. The 1930 census reports that cities hosted as much as 71.7% of workers employed in "trade and finance," even though they only hosted 33.8% of Hungarian population. This suggests that trading activity was indeed highly concentrated in cities. Moreover, an alternative threshold in which I classify settlements with more than 500 trade and finance workers as trading places would almost exactly coincide with my baseline classification of cities as trading places.

Four cities in Hungary's post-1920 territory went above the 20,000-inhabitant threshold between 1910 and 1930, hence I treat their geographic centers as trading places in 1930 but not in 1910. I do not need to assume that these trading places formed exogenously. A mapping outside the model can exist between (exogenous or endogenous) variables and the locations of trading places in any period. This mapping would need to be taken into account in counterfactual exercises but is irrelevant when comparing the actual equilibria of 1910 and 1930, in which I observe the locations of trading places in the data. Under the assumption that borders were redrawn exogenously, omitting the mapping cannot bias the estimation either, since the estimation solely relies on this exogeneity assumption.[8]

Some comments on the period of investigation are in order. In Hungary, starting from 1869, one population census was carried out in a decade. Therefore, the 1910 census is the last one providing a picture of the population distribution in Hungary before the border change. The 1920 census, which was carried out right after the change in borders, is not likely to reflect all the effects of border changes. This is because cities' relative population levels – abstracting from differences in birth and death rates – can only change through migration, and it is unlikely that migration fully took place over a period of months. Moreover, 1920 city populations are distorted by the fact that about 350,000 ethnic Hungarian

[7] I regard Budapest and its suburbs as one large city instead of a collection of cities in the analysis. Even though these cities were not united administratively until 1950, they were largely integrated economically already in 1910 (Hanák, 1988). Treating them as separate cities does not lead to a significant change in the results.

[8] See Section 4 for the specific exogeneity assumption I make.

refugees, who fled to the country in the previous years, were given temporary accommodation in school buildings and railway cars around railway stations of the largest cities. These "railway car towns" gradually disappeared by 1930 (Kontler, 2002). Population data in the 1941 census, on the other hand, is likely to be influenced by other factors including the Great Depression, the increase in trade between Hungary and Austria, Germany and Italy in the 1930s (Kosáry, 1941), the first two years of the Second World War, and significant new border changes between 1938 and 1940. Thus, the twenty-year window between 1910 and 1930 seems to be the best choice if one tries to measure the effects of the 1920 border changes on the spatial distribution of population.

3.2 Suggestive evidence for the model's mechanism

This section uses the data to provide suggestive evidence for the model's main prediction: the positive effect of trade on urbanization. Looking at the patterns without directly using the model of Section 2 has an obvious advantage: it does not impose the model's structure on the data. Its disadvantage is, however, that data on real income are not available, hence I can only look at the effect on simple urbanization measures that can be calculated from the data. To estimate how changes in trade due to the change in borders affected real income levels across Hungarian locations, I combine the full structure of the model with the data in Sections 4 and 5.

Hungary underwent a rapid phase of urbanization around the end of the 19th century and the beginning of the 20th century, as evident from Figure 2. The share of population living in cities over the post-1920 territory of Hungary increased by about 50% between 1890 and 1930, from 23.0% to 33.8%.[9] The increase in urbanization, however, was not uniform across regions. Figure 3 shows this by presenting the evolution of urbanization in two example counties: one in the center of the post-1920 territory, and one close to the new border. The county in the center exhibits an increase in the share of population living in cities over the entire period, although urbanization somewhat slows down after 1910. The county at the new border, on the other hand, suffers a decrease in urbanization after the border change, 1920. This is in line with the main prediction of the model.

To see if these examples reflect systematic patterns in the data, I plot the change in urbanization (measured by the share of people living in cities) between 1910 and 1930 against distance from the new border for all counties of Hungary in the left panel of Figure 4.[10] As expected, the relationship is increasing: counties near the new border exhibit

[9] The increase is even more dramatic, from 11.3% to 33.8%, if one looks at the change in urbanization over the entire territory of Hungary. However, this reflects the composition effect of the country losing its peripheral regions in 1920, which exhibited lower levels of urbanization. The increase is comparable if one focuses on the share of population living in settlements above $10,000$ inhabitants, which rose from 30.9% to 43.1% over the post-1920 territory.

[10] I measure the distance from the border as the shortest "as the crow flies" distance between the county seat and the border.

substantially slower (and often negative) growth in urbanization than counties farther away. The right panel of Figure 4 shows the result of a placebo exercise in which I plot distance from the new border against changes in urbanization between 1890 and 1910, a time period in which the post-1920 border could not shape urbanization patterns. No clear relationship emerges between these two variables, suggesting no significant difference in pre-trends of urbanization between counties that became peripheral after the change in borders and counties that remained central. Nevertheless, these counties exhibit different urbanization patterns once the new border is present, conforming to the model's predictions.[11]

One may wonder whether the slower growth in urbanization in border counties is due to a subset of industries only. Figure 5 seems to contradict this presumption for a rough industrial classification (agriculture vs manufacturing). Using data that the censuses provide on agricultural and manufacturing employment at the settlement level, I compute the share of cities in total county employment in both agriculture and manufacturing. Next, I plot the changes in these shares between 1910 and 1930 against distance from the new border. The increasing relationship that prevails in the aggregate is present in both industries, suggesting that the urbanization-fostering effect of trade operated in both manufacturing and agriculture.

4 Structural estimation

To obtain estimates on the welfare effects of Hungary's border changes, I employ the following strategy to combine the data with the structure of the model. Using the fact that the amenity function, $a(r)$, can take any form, I match the model exactly to cities' population levels in both 1910 and 1930. To this end, I assume the following amenity function: $a(r) = a_c$ if location r belongs to the area of city c, and $a(r) = 0$ otherwise.[12] Since I do not have data on cities' geographic boundaries but do have their area in square kilometers, I assume that each city had a circular shape around its geographic center, with the area of the circle being equal to the city's actual area as reported in the data. Then I use the model, separately for 1910 and 1930, to search for values of a_c that are consistent with each city having the same population as in the data.[13] To obtain predictions at a high level of spatial disaggregation, I use a very fine discretization of space when calculating the

[11] Due to the small number of observations, the estimated coefficients of distance from the border on county-level urbanization have large standard errors. Nevertheless, the coefficients are all positive and statistically significant at a 10% level, except in the placebo exercise in which the coefficient is not significantly different from zero at any standard significance level.

[12] Recall that consumers' utility is additively separable in amenities, hence $a(r)$ does not imply infinitely bad amenities.

[13] In principle, there could exist different sets of amenity levels that are consistent with city sizes observed in the data. However, running the searching procedure with many different initial values has always resulted in the same values of a_c. This suggests that the set of amenity levels leading to the observed city sizes is a singleton, at least for the specific geography and values of structural parameters used in this paper.

equilibrium on the computer: I split the territory of Hungary into 0.01° by 0.01° grid cells. This means about 400,000 grid cells in total. Despite the large number of locations, the relatively simple structure of the model leads to quick calculations: finding the values of a_c that rationalize the data takes a few minutes on a typical personal computer.

Besides $a(\cdot)$, I need to choose the specification of two functions: shipping costs across trading places $\tau(\cdot,\cdot)$, and shipping costs from and to the trading place $\varsigma(\cdot,\cdot)$, as well as the values of two structural parameters: σ and θ. I assume that both types of shipping costs are exponential functions of distance $d(\cdot,\cdot)$:

$$\begin{aligned}\tau(r,s) &= e^{\phi \cdot dist(r,s)}\\ \varsigma(r,s) &= e^{\psi \cdot dist(r,s)}\end{aligned}$$

The exponential formulation of trade costs is a frequently used assumption in the economic geography literature. Examples are Fujita et al. (1999), Rossi-Hansberg (2005) and Desmet and Rossi-Hansberg (2014). I measure distances "as the crow flies." It is unlikely that using road and rail distances instead would lead to a significant change in the results. This is because the Hungarian road and railroad network was very dense already in 1910, comparable in density to the networks of developed countries (Kontler, 2002).

The next step is to choose the values of shipping cost elasticities ϕ and ψ. I match ϕ to evidence on trade costs. The 1910 Yearbook of the Hungarian Statistical Office reports average prices of the three main products imported through the port of Fiume (now Rijeka, Croatia) both in Fiume and in Budapest. Wheat was 25% more expensive in Budapest than in Fiume, coffee was 15% more expensive, and rice was 8% more expensive. With $\phi = 3.51 \cdot 10^{-4}$, shipping costs between the CBDs of Fiume and Budapest are equal to 16.2%, which corresponds to the average of the above three numbers. Next, I choose the value of ψ such that the model matches the standard deviation of the sizes of settlements above $2,000$ inhabitants in 1910. Since the population density gradient around trading places is strictly increasing in shipping costs $\varsigma(\cdot,\cdot)$ – see equation (4) –, the standard deviation is strictly increasing in ψ in the model. This implies that the parameter is identified. The procedure pins down a value of $\psi = 1.02 \cdot 10^{-1}$.

I choose the value of structural parameter σ based on the fact that the elasticity of trade with respect to variable costs is $1-\sigma$ in the model. Following Simonovska and Waugh (2014), I set the value of this elasticity to negative four, which implies $\sigma = 5$. Finally, I structurally estimate the value of θ – the parameter directly influencing the effects of trade on urbanization and real income, as evident from Proposition 1 – using simulated method of moments. I impose the following moment condition, akin to the moment conditions applied by Ahlfeldt et al. (2015) to study the division of Berlin during the Cold War:

$$corr\left[\Delta a_c, dist\left(\mu_c, border\right)\right] = 0$$

where Δa_c is the change in city c's amenities between 1910 and 1930 as identified from the model, and $dist\,(\mu_c, border)$ is the distance of city c to Hungary's post-1920 border. This moment condition thus states that the placement of the new border, by being exogenous, was uncorrelated with changes in exogenous amenities in cities.

Finally, note that the model assumes that Hungary was a closed economy both before and after the change in its borders. Given that the territory of the country shrinked, the more restrictive of these two assumptions is the one that assumes autarky after 1920. Continuing trade with the locations that had previously been part of Hungary could have mattered for how much trade was lost as a result of the border change, and hence could bias my estimates of real income losses upwards. This is, however, not a concern as political conflicts led Hungary to essentially stop trading with its neighbors after the redrawing of borders (Teleki, 1923), and foreign trade started to grow substantially in the 1930s only (Kosáry, 1941).

5 Results

Table 1 presents the results of the estimation. The point estimate of θ is 11.49. By equation (3), this implies that increasing real income or amenities at a location by one unit leads to a 8.3% increase in the location's population. Although the standard error of θ is relatively large, the effect of border changes on average real income, calculated using the gains from trade formula of Proposition 2, is tightly identified. On average, the new border leads to a 15.55% decline in real income, with a standard deviation of 1.08 percentage points across locations.[14] Figure 6 presents the smoothed empirical density of the losses, calculated using the point estimate of $\theta = 11.49$. It can be seen from the figure that the redrawing of borders had a considerably heterogeneous impact on different locations; the 2nd and 98th percentiles of the distribution are 13.9% and 17.6%, almost four percentage points apart from one another. Table 1 also shows that this heterogeneity in losses led to an increase in country-level income inequality, measured as the standard deviation of individuals' (log) real income. The point estimate of the increase in inequality is 9.78%. However, the standard error of this estimate, unlike the one of average income losses, is relatively large. This is primarily due to the fact that real income inequality was relatively low in 1910, thus even relatively small noises in measured inequality in 1910 and 1930 lead to a large noise in the measured percentage change in inequality. Finally, the geography of the losses for the point estimate of $\theta = 11.49$ can be seen in Figure 7. This figure shows that the smallest losses were mostly incurred at central locations, while places around the new Eastern border suffered the largest decline in real income. This is not surprising as these were places that had been centrally located before 1920 but got very close to the border afterwards (see

[14] I weight by locations' population when calculating the average and the standard deviation of real income levels and losses.

Figure 1).

Recall from Proposition 2 that the losses in real income coincide with decreases in the model-based urbanization index. Hence, Figure 7 also suggests that urbanization decreased most dramatically near the new border. This prediction cannot be directly tested since the urbanization index has no observable counterpart in the data. However, Figure 8 presents the changes in an observable measure of urbanization: the ratio of population living in cities to those living in settlements above $2,000$ inhabitants, for each county. The left panel plots this measure against distance from the new border in the model, while the right panel plots the same relationship in the data. Both figures show an increasing relationship, with a slope of 0.08 per hundred kilometers in the model and 0.11 per hundred kilometers in the data. This indicates that the model is able to account for the vast majority of differences in urbanization between interior and border counties.

The bottom panel of Table 1 evaluates the model's ability to fit the population distribution in the data, both before and after the border change. In general, the model seems successful at predicting settlement sizes in both periods, especially after 1920. Of course, part of this success might be due to matching the population of settlements above $20,000$ inhabitants. However, the last two rows of the table show that the correlations are between 0.4 and 0.5 even when these settlements are excluded from the calculations.

This section showed that the theoretical model of Section 2 is successful in two respects. First, it is able to predict the effect of border changes on the population distribution with relatively large confidence. Second, it is able to explain the fact that border regions suffered larger decreases in urbanization than interior regions. In the model, this result is a direct consequence of the role that trade plays in the process of urbanization.

6 Conclusion

Urbanization is an ongoing process: the share of world population living in urban areas has increased from 30% in 1950 to 54% in 2014, and is expected to reach 66% by 2050, which will likely have profound effects on individuals' wellbeing, on development and on the environment (UN, 2014). Naturally, we aim to understand the sources of this large-scale process. In this paper, I explore trade as a possible source of urbanization. To this end, I develop a quantitative model of economic geography that, despite its flexible geographic structure, provides simple predictions on the effect of trade on urbanization and welfare. Next, I address the key challenge of identification, the endogeneity of trade, by using the redrawing of Hungary's borders after the First World War as a source of exogenous variation in trade. In line with the key prediction of the model, I find that urbanization decreased in regions near the new borders relative to regions farther away. The estimated model can quantitatively replicate the difference in urbanization between border and interior regions. Moreover, it can be used as a tool to measure the effects on real income, which are unobserved in the data.

A natural extension of the framework is one that models the endogenous formation of trading places. In Nagy (2017), the spatial concentration of a sector in which production is subject to increasing returns implies that trade only happens at a subset of locations in equilibrium. An alternative strategy would rely on explicitly modeling location choice in the trading sector. This would allow one to study how the spatial distribution of consumption, production and trading activity together respond to changes in borders, the natural environment or international trade.

References

- Ahlfeldt, G., Redding, S., Sturm, D., and Wolf, N. (2015): The economics of density: Evidence from the Berlin Wall. *Econometrica*, vol. 83(6), 2127–2189.

- Allen, T. and Arkolakis, C. (2014): Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, vol. 129(3), 1085–1140.

- Armenter, R., Koren, M., and Nagy, D. (2014): Bridges. Mimeo.

- Bleakley, H. and Lin, J. (2012): Portage and path dependence. *Quarterly Journal of Economics*, vol. 127(2), 587–644.

- Brülhart, M. (2011): The spatial effects of trade openness: a survey. *Review of World Economics*, vol. 147(1), 59–83.

- Coşar, K. and Fajgelbaum, P. (2016): Internal geography, international trade, and regional specialization. *American Economic Journal: Microeconomics*, vol. 8(1): 24–56.

- Desmet, K. and Rossi-Hansberg (2014): Spatial development. *American Economic Review*, vol. 104(4), 1211–1243.

- Desmet, K., Nagy, D., and Rossi-Hansberg, E. (2018): The geography of development. *Journal of Political Economy*, forthcoming.

- Fajgelbaum, P. and Redding, S. (2014): External integration, structural transformation and economic development: Evidence from Argentina 1870-1914. *NBER Working Paper* 20217.

- Fujita, M., Krugman, P., and Venables, A. (1999): The spatial economy. Cities, regions, and international trade. The MIT Press.

- Hanák, P. (ed.) (1988): One thousand years. A concise history of Hungary. Corvina.

- Kontler, L. (2002): A history of Hungary. Millenium in Central Europe. Palgrave Macmillan.
- Kosáry, D. (1941): A history of Hungary. Country Life Press.
- Krugman, P. (1991): Increasing returns and economic geography. *Journal of Political Economy*, vol. 99(3), 483–499.
- Krugman, P. and Livas Elizondo, R. (1996): Trade policy and the Third World metropolis. *Journal of Development Economics*, vol. 49, 137–150.
- Monte, F., Redding, S., and Rossi-Hansberg, E. (2018): Commuting, migration and local employment elasticities. *American Economic Review*, forthcoming.
- Nagy, D. (2017): City location and economic development. Mimeo.
- Redding, S. (2016): Goods trade, factor mobility and welfare. *Journal of International Economics*, vol. 101, 148–167.
- Redding, S. and Sturm, D. (2008): The costs of remoteness: Evidence from German division and reunification. *American Economic Review*, vol. 98(5), 1766–1797.
- Redding, S. and Rossi-Hansberg, E. (2017): Quantitative spatial economics. *Annual Review of Economics*, vol. 9, 21–58.
- Rossi-Hansberg, E. (2005): A spatial theory of trade. *American Economic Review*, vol. 95(5), 1464–1491.
- Simonovska, I. and Waugh, M. (2014): The elasticity of trade: Estimates and evidence. *Journal of International Economics*, vol. 92(1), 34–50.
- Teleki, P. (1923): The evolution of Hungary and its place in European history. The Macmillan Company.
- United Nations, Department of Economics and Social Affairs, Population Division (2014): World urbanization prospects: The 2014 revision. Highlights.

Table 1: Estimation results

θ	11.49
	(3.71)
Average GFT	-15.55%
	(0.28%)
Standard deviation of GFT	1.08%
	(0.24%)
Change in standard deviation of log real income, 1910 to 1930	9.78%
	(27.05%)
Correlation between model and data, settlements above 2000, 1910	0.6030
	(0.0829)
Correlation between model and data, settlements above 2000, 1930	0.7070
	(0.1003)
Correlation between model and data, settlements 2000 to $20,000$, 1910	0.4092
	(0.0953)
Correlation between model and data, settlements 2000 to $20,000$, 1930	0.4759
	(0.1194)

Bootstrap standard errors in parentheses.

Figure 1: Hungary before (green) and after (brown) the Treaty of Trianon, 1920

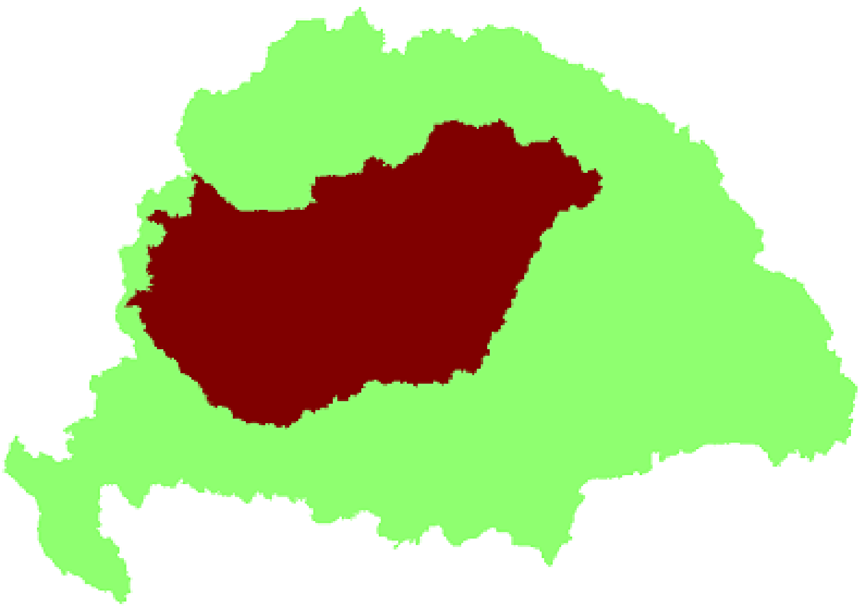


Figure 2: Urbanization in Hungary

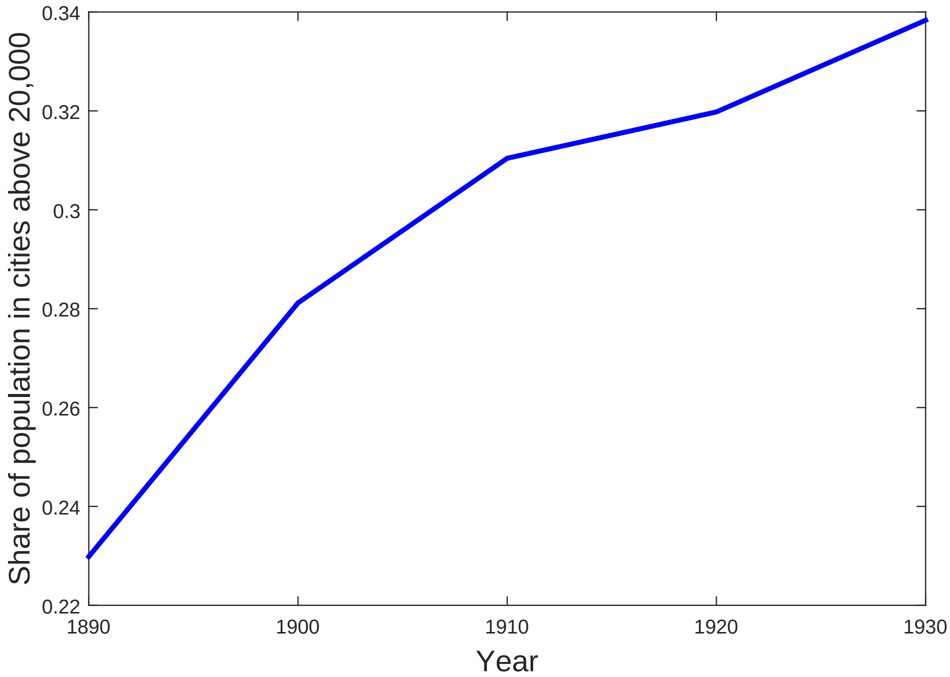


Figure 3: Urbanization in Pest-Pilis-Solt-Kiskun county (left) vs Győr county (right)

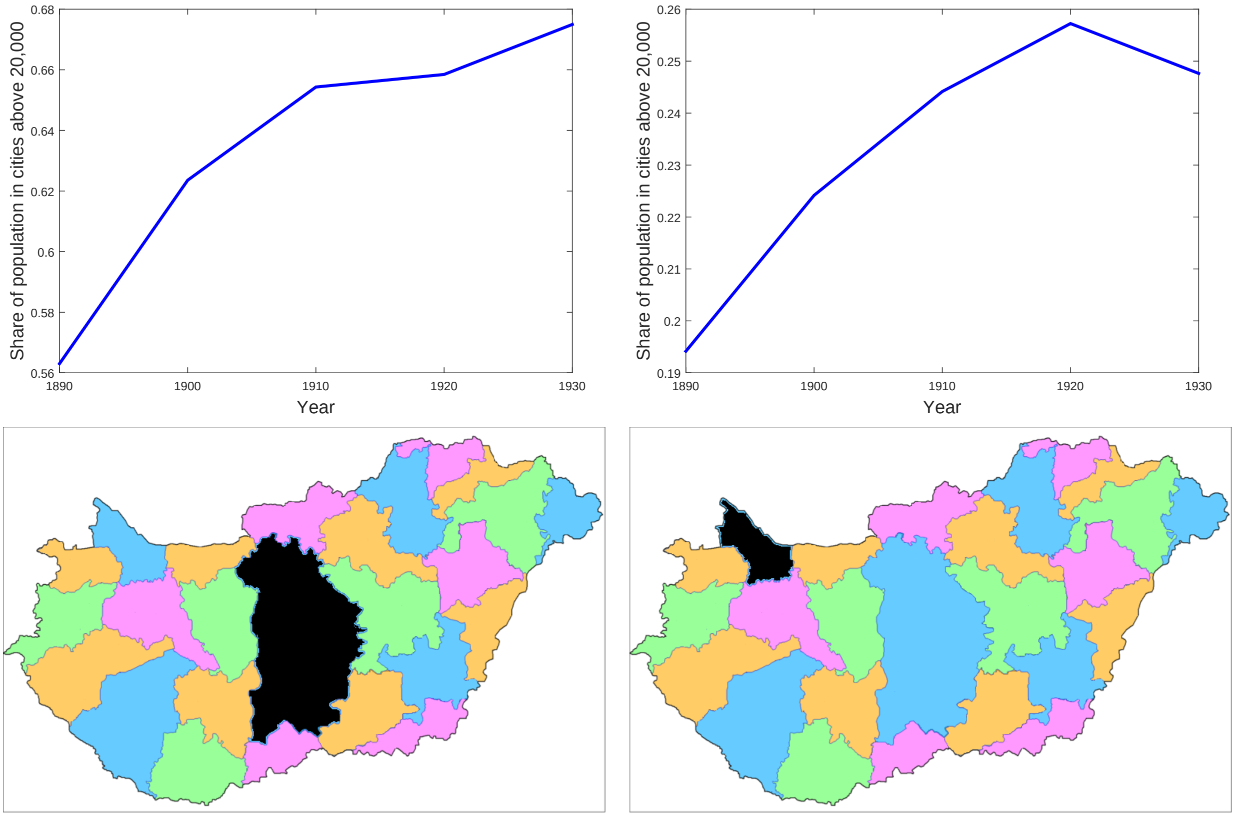


Figure 4: Change in urbanization as a function of distance from the new border: after the border change (left) vs before the border change (right)

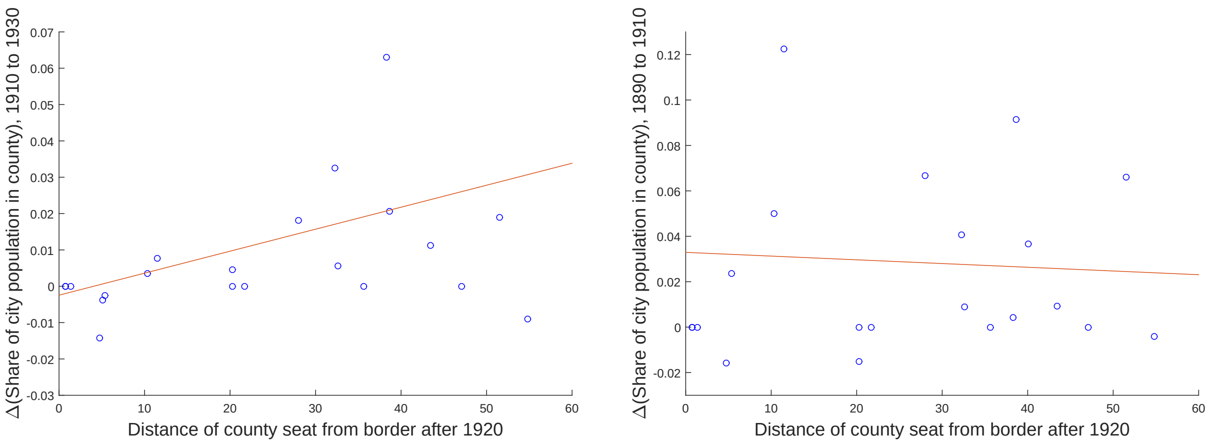


Figure 5: Change in urbanization in agriculture (left) vs manufacturing (right)

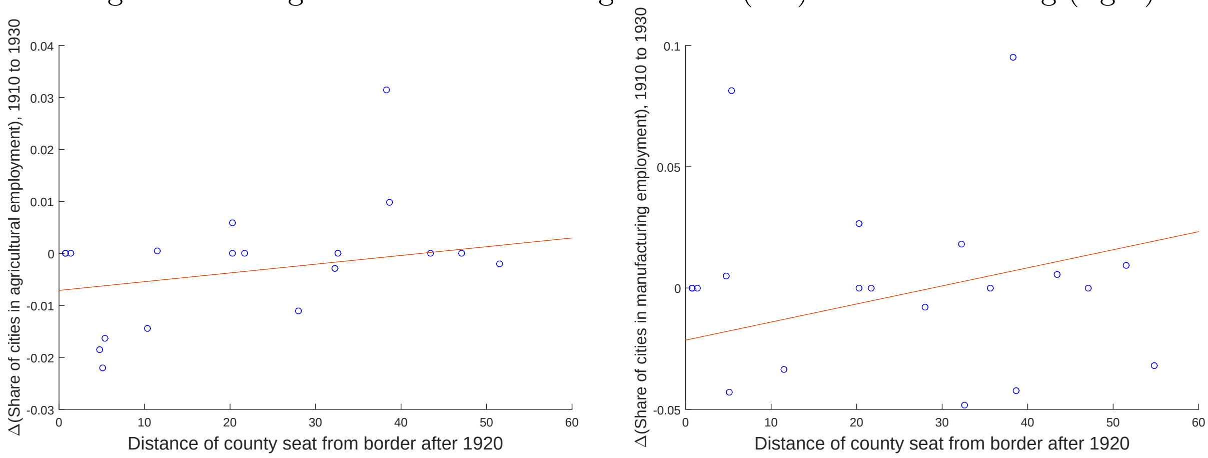


Figure 6: Empirical density of real income changes between 1910 and 1930

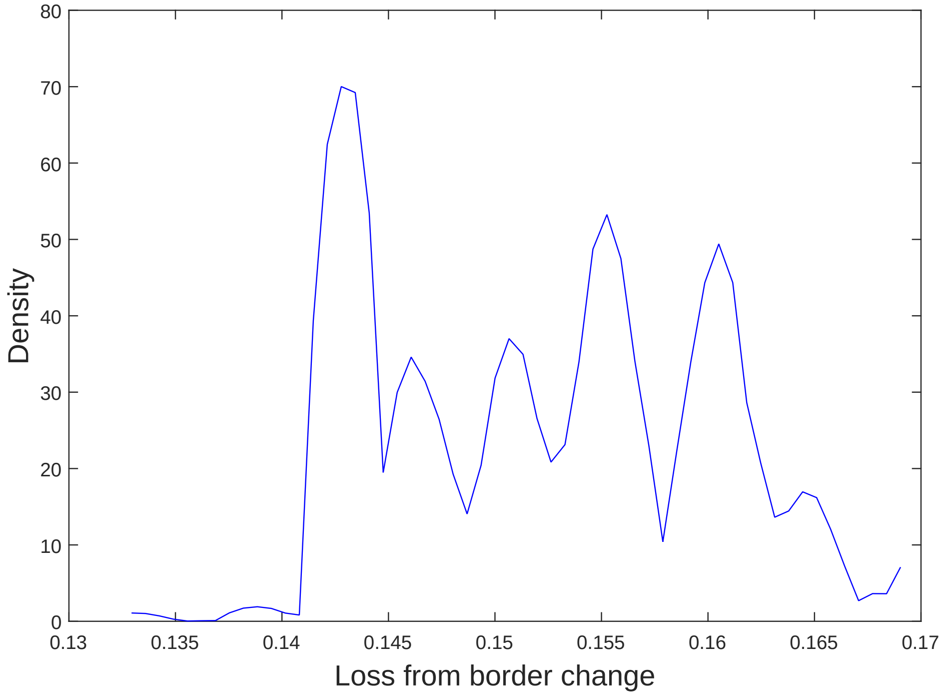


Figure 7: Map of real income changes between 1910 and 1930

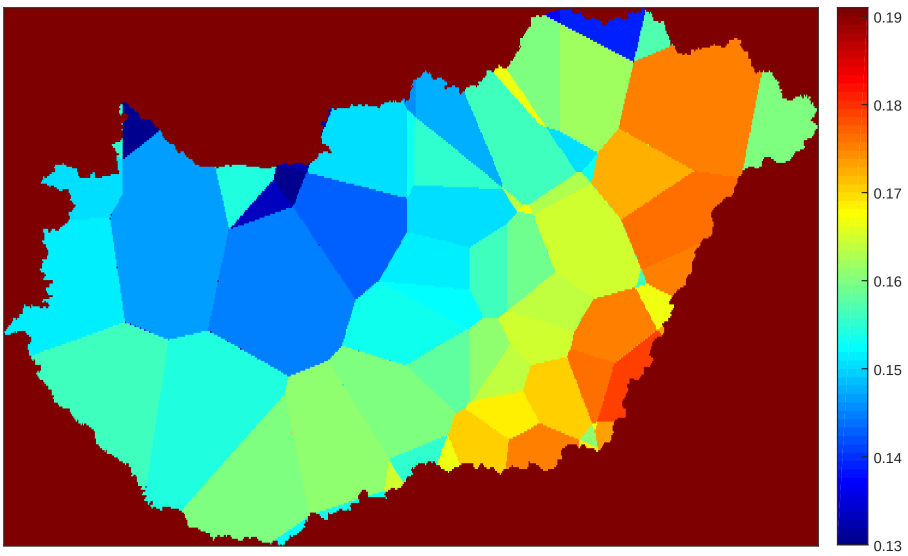
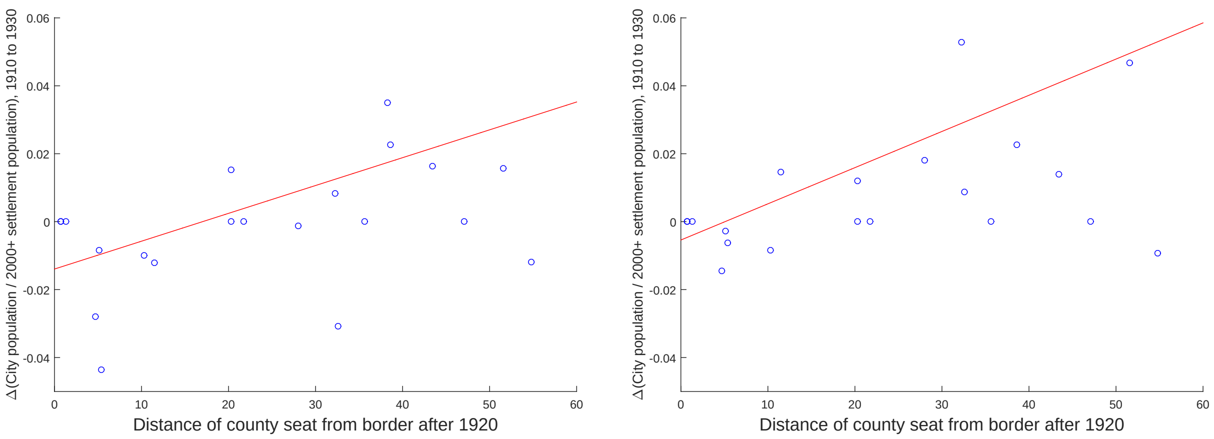


Figure 8: Change in population of cities relative to population of settlements above 2,000 inhabitants: model (left) vs data (right)



Appendix A: Derivation of equations (3) to (6)

First, use consumers' CES demand function for goods to write the price index at m as

$$P_m = \left[\sum_o p_o^{1-\sigma} L_o \tau \left(\mu_o, \mu_m \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \tag{A1}$$

Second, since workers do not value leisure and shipping costs take the iceberg form, their utility is strictly increasing in the output of their product. Although workers, being monopolists, could have an incentive to reduce output in order to increase their price and hence their revenue, they do not want to do that as long as $\sigma > 1$. The intuition for this result is that, whenever $\sigma > 1$, decreasing the price by 50% more than doubles demand and thus increases total revenue. As a consequence, workers all produce the maximum quantity, $x^j = 1$, and set a price p_m^j at which demand meets supply. As neither demand nor supply depends on the worker's index, the equilibrium price is common across all workers at the same trading place: $p_m^j = p_m$. These results allow me to write the goods market clearing condition (2) as

$$p_m^{\sigma} = \sum_o P_o^{\sigma-1} p_o L_o \tau \left(\mu_m, \mu_o \right)^{1-\sigma}. \tag{A2}$$

Third, note that the extreme value distribution of idiosyncratic amenities implies that the share of population living at r is given by

$$\frac{L(r)}{\overline{L}} = \frac{\left[e^{a(r) + \max_m \frac{p_m}{\varsigma(\mu_m, r) P_m}} \right]^{\theta^{-1}}}{\sum_s \left[e^{a(s) + \max_o \frac{p_o}{\varsigma(\mu_o, s) P_o}} \right]^{\theta^{-1}}}.$$

The maximizations on the right-hand side do not depend on the worker's index. As a consequence, workers living at a given residential location r all choose the same trading place m; in what follows, I denote this trading place by $\mu(r)$.[15] Using this and denoting real income by $\omega_m = \frac{p_m}{P_m}$, the previous formula reduces to

$$\log L(r) = \nu + \theta^{-1} \left[a(r) + \varsigma \left(\mu(r), r \right)^{-1} \omega_{\mu(r)} \right] \tag{3}$$

where $\nu = \log \left(\overline{L} \right) - \log \left[\sum_s e^{a(s) + \varsigma(\mu(s), r)^{-1} \omega_{\mu(s)}} \right]$. Also, by the definition of $\mu(r)$,

$$\varsigma \left(\mu(r), r \right)^{-1} \omega_{\mu(r)} \geq \varsigma \left(\mu_m, r \right)^{-1} \omega_m \qquad \forall m. \tag{4}$$

[15] If there is a tie between trading places m and o for a residential location r, I only consider the equilibrium in which workers choose the trading place with the smaller index, that is, $\mu(r) = \min\{m, o\}$. This issue never arises in practice when I take the model to the data.

Now use $P_m = \frac{p_m}{\omega_m}$ to rewrite equations (A1) and (A2):

$$p_m^{1-\sigma} \omega_m^{\sigma-1} = \sum_o p_o^{1-\sigma} L_o \tau \left(\mu_o, \mu_m\right)^{1-\sigma}$$

$$p_m^{\sigma} = \sum_o p_o^{\sigma} \omega_o^{1-\sigma} L_o \tau \left(\mu_m, \mu_o\right)^{1-\sigma}$$

Recall that shipping costs are assumed to be symmetric: $\tau\left(\mu_m, \mu_o\right) = \tau\left(\mu_o, \mu_m\right)$. In this case, the previous two equations can be reduced to one equation. This is done using the trick by Allen and Arkolakis (2014): guess that the price at trading place m takes the form

$$p_m = \omega_m^{\iota}.^{16}$$

Then note that for $\iota = \frac{\sigma-1}{2\sigma-1}$, both (A1) and (A2) imply

$$\omega_m^{\frac{\sigma(\sigma-1)}{2\sigma-1}} = \sum_o \omega_o^{-\frac{(\sigma-1)^2}{2\sigma-1}} L_o \tau \left(\mu_m, \mu_o\right)^{1-\sigma}. \tag{5}$$

Finally, equation (6) simply follows from the fact that L_m equals the number of people trading at m:

$$L_m = \sum_{r:\ m=\mu(r)} L\left(r\right). \tag{6}$$

I have thus derived equations (3) to (6).

Appendix B: An isomorphic model with firms

This section presents an alternative model in which goods are produced by monopolistically competitive firms that hire workers in competitive labor markets. I first outline the assumptions of the model. Next, I show the formal isomorphism between this model and the one presented in Section 2.

Assume that workers consume a CES aggregate of goods available in the economy, implying that the utility of worker i who chooses to live at location r and trade at trading place m is

$$u_m\left(r,i\right) = a\left(r,i\right) + \left[\sum_{k=1}^{K} c_m^k \left(r,i\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{A3}$$

where $c_m^k\left(r,i\right)$ is the worker's consumption of good k, and K is the endogenous number of available goods. Workers cannot produce any good themselves. However, it is possible to set up firms at trading places, which then hire workers to produce goods according

[16] Although there could, in principle, appear an intercept term on the right-hand side, I can set it to one since I have not normalized any price yet.

to the linear production technology of Section 2, in which one unit of labor is needed to produce one unit of output. Starting a firm requires additional $f > 0$ units of labor. Each firm has an incentive to differentiate its product from those of other firms, as it allows the firm to be a monopolist and charge a markup over its marginal cost. Firms hence engage in monopolistic competition with endogenous entry. The fixed startup cost leads to increasing returns internal to the firm. This, together with shipping costs across trading places, constitutes an agglomeration force, as in Krugman (1991).

Consider the problem of a firm producing good k at trading place m. Since the firm takes the wage as given, the price elasticity of demand is constant at σ and trade costs are of the iceberg type, the firm sets a mill price that is a constant markup over the wage:

$$p_m^k = p_m = \frac{\sigma}{\sigma - 1} w_m \tag{A4}$$

and, by free entry, each firm produces $f(\sigma - 1)$ units.

Finally, I assume that a worker residing at location r needs to pay an iceberg cost $\varsigma(\mu_m, r) \geq 1$ to commute to trading place m. The rest of the model's assumptions are unchanged. This allows me to define the equilibrium as follows.

Definition 3 *Given parameters $\{\sigma, \theta, f, \overline{L}\}$, geography S, $\{\mu_1, \ldots, \mu_M\}$ and functions $a : S \to \mathbb{R}_+$, $\{\tau, \varsigma\} : S^2 \to \mathbb{R}_+$, an equilibrium of the economy with firms consists of a population distribution $L : S \to \mathbb{R}_+$; consumption levels $c^k : S \to \mathbb{R}_+$; wages, goods' prices, production levels and the number of goods produced $\{p, w, x^k, K\} : \{1, \ldots, M\} \to \mathbb{R}_+$; and a trading place assignment function $\mu : S \to \{1, \ldots, M\}$ such that the following hold:*

1. *Workers choose their consumption, location and trading place to maximize their utility (A3).*

2. *Firms choose their prices and quantities to maximize profits, and profits are driven down to zero by free entry. Therefore, equation (A4) holds, and each firm produces $f(\sigma - 1)$ units of its good.*

3. *The market for labor clears at every trading place, that is,*

$$L_m = K_m [f(\sigma - 1) + f] = K_m f \sigma \tag{A5}$$

for all m, where L_m denotes the mass of workers commuting to m, and K_m denotes the number of goods produced at m.[17]

[17] For simplicity, I assume that the number of goods produced can take non-integer values. Note, however, that it is possible to normalize the fixed cost f, and hence the size of a firm, such that the number of goods produced is an integer at every location.

4. *The market for each good clears at every trading place, implying*

$$f(\sigma-1)=\sum_{o}\tau(\mu_m,\mu_o)^{1-\sigma}p_m^{-\sigma}P_o^{\sigma-1}w_oL_o \tag{A6}$$

for all m*, where* P_o *is the CES price index at* o*:*

$$P_o=\left[\sum_m K_m\left[\tau(\mu_m,\mu_o)p_m\right]^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$$

5. *The national labor market clears, that is,*

$$\overline{L}=\sum_m L_m.$$

The next proposition states that the equilibrium distribution of population, real wages and trading place assignments is isomorphic between this model and the one in Section 2.

Proposition 3 *Normalizing the fixed cost to* $f=\sigma^{-\sigma}(\sigma-1)^{\sigma-1}$*, the equilibrium conditions of the model with firms can be reduced to the system of equations (3) to (6). Hence, the equilibrium distribution of population, real income and trading place choices is the same as in the model of Section* 2*.*

Proof. *First, use equations (A4) and (A5) to write the price index at* m *as*

$$P_m=f^{\frac{1}{\sigma-1}}\sigma^{\frac{\sigma}{\sigma-1}}(\sigma-1)^{-1}\left[\sum_o w_o^{1-\sigma}L_o\tau(\mu_o,\mu_m)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{A7}$$

and the goods market clearing condition (A6) as

$$f(\sigma-1)=\left[\frac{\sigma}{\sigma-1}\right]^{-\sigma}w_m^{-\sigma}\sum_o\tau(\mu_m,\mu_o)^{1-\sigma}P_o^{\sigma-1}w_oL_o,$$

from which

$$w_m^{\sigma}=f^{-1}\sigma^{-\sigma}(\sigma-1)^{\sigma-1}\sum_o P_o^{\sigma-1}w_oL_o\tau(\mu_m,\mu_o)^{1-\sigma}. \tag{A8}$$

Second, the share of population living at r *is given by*

$$\frac{L(r)}{\overline{L}}=\frac{\left[e^{a(r)+\max_m\frac{w_m}{\varsigma(\mu_m,r)P_m}}\right]^{\theta^{-1}}}{\sum_s\left[e^{a(s)+\max_o\frac{w_o}{\varsigma(\mu_o,s)P_o}}\right]^{\theta^{-1}}}$$

just like in the model of Section 2*. Using this and denoting real income by* $\omega_m=\frac{w_m}{P_m}$*, the* *previous formula reduces to*

$$\log L(r) = \nu + \theta^{-1}\left[a(r) + \varsigma(\mu(r), r)^{-1}\omega_{\mu(r)}\right] \tag{3}$$

where $\nu = \log\left(\overline{L}\right) - \log\left[\sum_s e^{a(s)+\varsigma(\mu(s),r)^{-1}\omega_{\mu(s)}}\right]$*. Also, by the definition of* $\mu(r)$*,*

$$\varsigma(\mu(r), r)^{-1}\omega_{\mu(r)} \geq \varsigma(\mu_m, r)^{-1}\omega_m \qquad \forall m. \tag{4}$$

Now use $P_m = \frac{w_m}{\omega_m}$ *to rewrite equations (A7) and (A8):*

$$w_m^{1-\sigma}\omega_m^{\sigma-1} = f^{-1}\sigma^{-\sigma}(\sigma-1)^{\sigma-1}\sum_o w_o^{1-\sigma}L_o\tau(\mu_o, \mu_m)^{1-\sigma}$$

$$w_m^{\sigma} = f^{-1}\sigma^{-\sigma}(\sigma-1)^{\sigma-1}\sum_o w_o^{\sigma}\omega_o^{1-\sigma}L_o\tau(\mu_m, \mu_o)^{1-\sigma}$$

from which, following the same procedure as in Appendix A, I obtain

$$\omega_m^{\frac{\sigma(\sigma-1)}{2\sigma-1}} = \sum_o \omega_o^{-\frac{(\sigma-1)^2}{2\sigma-1}}L_o\tau(\mu_m, \mu_o)^{1-\sigma} \tag{5}$$

where I also used the normalization $f = \sigma^{-\sigma}(\sigma-1)^{\sigma-1}$*. Using the same argument as in Appendix A, equation (6) obviously follows:*

$$L_m = \sum_{r:\ m=\mu(r)} L(r). \tag{6}$$

■