# Trade and urbanization: Evidence from Hungary*

Dávid Krisztián Nagy†

December 28, 2020

**Abstract**

I study how trade affects urbanization and real income. To this end, I exploit large-scale exogenous changes in trade stemming from the redrawing of Hungary's borders after the First World War. I show that, after the border change, urbanization in counties near the new border decreased significantly relative to counties farther away. This is despite the fact that these counties exhibited similar economic characteristics and urbanization trends prior to the border change. I rationalize these findings in a spatial model of trade and urbanization. In the model, benefits from trading drive agglomeration around locations where trading activity takes place. As a result, increasing trade leads to urbanization and real income gains. To measure real income changes after the redrawing of Hungary's borders, which are unobserved in the data, I structurally estimate the model using the border change as a source of exogenous variation. I find a 15.55% decrease in average real income after the redrawing of borders, with the largest losses concentrated in border regions.

## 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.

*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, Joan Monràs, Giacomo Ponzetto, Stephen Redding, two anonymous referees, as well as seminar participants at the Barcelona Summer Forum, CREI, LSE, Princeton University and the University of Padova for their helpful comments and suggestions. All errors are my own.

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

Identifying the extent to which trade can induce urbanization and income gains is challenging, however. This is due to a classical simultaneity problem: while trade might induce urbanization, urbanization is also likely to induce 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.

In this paper, I estimate the effect of trade on urbanization and real income by exploiting a unique historical episode: Hungary's large-scale border changes after the First World War. Drawn by the Allied Powers in 1920, Hungary's post-war border offers a laboratory to study the effects of trade. On the one hand, historical evidence suggests that the border did not correspond to prior political, economic or ethnic boundaries (Kontler, 2002; Teleki, 1923). Hence, it is reasonable to assume that trade between the opposite sides of the post-war border was not subject to any frictions before 1920. On the other hand, political conflict between Hungary and its new neighbors led to little trade across the border in the years after 1920 (Csató, 2000; Teleki, 1923). As a result, Hungarian counties in the new border's close proximity experienced a dramatic and exogenous decrease in trade with nearby locations that ended up on the opposite side of the border.

My reduced-form empirical strategy exploits this exogenous variation by comparing urbanization in Hungarian counties near the new border to urbanization in counties farther away. If trade induces urbanization and borders cut trade, then border counties should experience a slowdown in urbanization relative to more centrally located counties. This is precisely what I find. In my baseline specification, doubling distance from the new border implies a significant, 0.751 percentage point increase in a county's urbanization between 1910 and 1930.[1] This is more than half of a county's average increase in urbanization during the same period. I also argue that this large estimated effect is not driven by other differences between counties closer and counties farther away from the new border. First, I show that these two groups of counties were extremely similar in their observable characteristics before the border change. Second, I show that the differential trends in urbanization between these two groups of counties were not present before 1910.

To rationalize these empirical findings, I develop a quantitative spatial model of trade and urbanization. There is a strand of quantitative spatial models in which external trade induces urbanization near ports through which trade happens with other countries (Coşar and Fajgelbaum, 2016; Fajgelbaum and Redding, 2018). In what follows, I refer to these models as "port trade" models. Such a mechanism is likely to be important if external trade through ports plays a dominant role, such as in today's China (Coşar and Fajgelbaum, 2016) or in 19th-century Argentina (Fajgelbaum and Redding, 2018). Early-20th century Hungary, however, did not exhibit this spatial structure as the vast majority of trade was

[1] I measure urbanization as the share of the county's population that lived in cities above 20,000 inhabitants. Further details on the construction of this measure are provided in Section 2.2.

internal (that is, within-country) trade.[2] As a result, my starting point from a modeling perspective is the class of quantitative spatial models with internal trade, such as Allen and Arkolakis (2014), Donaldson and Hornbeck (2016) and Redding (2016). In what follows, I refer to these models as "internal trade" models.

A common feature of the above "internal trade" models is that trade affects a location's population through the location's *market access*, a structural term that depends on trade costs and the populations of the location's trading partners. The model I develop in this paper also has this feature, but its structure needs to depart from the structure of these existing models. Intuitively, the reason is that the above "internal trade" models predict a constant elasticity of population to market access. This implies that a 1% increase in market access changes population by the same percentage at any location that experiences this increase in trading opportunities. Therefore, trade affects population but has no heterogeneous impact between urban and rural locations and therefore does not affect urbanization.

In my model, I move away from this constant elasticity to replicate the positive effect of trade on urbanization observed in the data. To accomplish this, I borrow a feature of "port trade" models: the notion that trade takes place at a subset of locations, which I call *trading places*.[3] Besides this, I choose the functional form of utility to generate a positive impact of market access on the population gradient around trading places, a model-based measure of urbanization that I call the *urbanization index*. I prove that the urbanization index is increasing in market access in the model. I also prove that the real income gains from trade are closely linked to changes in the urbanization index: as trading opportunities increase, people move closer to their trading places and thus increase urbanization, while also reaping the gains from increased trade.

I structurally estimate the model to measure the effects of Hungary's new borders on Hungarian residents' real income, which is unobserved in the data. The core of the estimation is a moment condition that relies on the exogeneity of the post-1920 border, akin to the moment conditions used in Ahlfeldt et al. (2015). I find that the new border had a large and heterogeneous impact on Hungarian locations. The average loss in real income per capita was 15.55%, while the 2nd and 98th percentiles were 13.9% and 17.6%, respectively. I show that border regions experienced the largest drop in trade, real income and urbanization. I also test the model's fit to the spatial distribution of population and find correlations between the model and the data that were around 0.6 and 0.7 before and after the border change, respectively. Finally, I show the robustness of the results to alternative values of structural parameters, to an alternative moment condition, and to incorporating multiple sectors in the model.

[2]Hungary's total exports equaled 13.1% of the country's GDP in 1928 (Csató, 2000; Eckstein, 1955). For comparison, they equal 83.3% of Hungary's GDP today.

[3]When taking the model to the data, the locations of cities and evidence on employment in trade help me identify the locations of trading places.

This paper is related to an extensive literature that studies the reduced-form effect of trade on various economic outcomes. These outcomes include city populations and employment (Bleakley and Lin, 2012; Brülhart, Carrère and Trionfetti, 2012; Brülhart, Carrère and Robert-Nicoud, 2018; Ducruet et al., 2020; Ellingsen, 2020; Redding and Sturm, 2008), regional populations and economic activity (Brooks, Gendron-Carrier and Rua, 2019; Campante and Yanagizawa-Drott, 2018) and country development (Feyrer, 2009; Feyrer, 2019; Pascali, 2017). Similar to my paper, these papers rely on large-scale shocks to trade for identification. Redding and Sturm (2008) is the closest paper in this regard, as it also exploits a change in country borders: the division and reunification of Germany after the Second World War. My contribution to this literature is that, to the best of my knowledge, I am the first to study the reduced-form effect of trade on urbanization, that is, on the share of population living in cities.

The paper is also related to the rapidly growing literature developing quantitative spatial models of trade. One strand of this literature (Coşar and Fajgelbaum, 2016; Fajgelbaum and Redding, 2018) proposes models of how external trade through ports shapes the spatial distribution of activity within countries ("port trade" models). Another strand (Allen and Arkolakis, 2014; Donaldson and Hornbeck, 2016; Monte, Redding and Rossi-Hansberg, 2018; Redding, 2016) focuses on the role of internal trade instead ("internal trade" models). As internal trade played a dominant role in early-20th century Hungary, I propose a model that belongs to this second strand. However, I contribute to the "internal trade" literature by developing a model in which trade has a positive impact on urbanization.[4] To this end, among other things, I borrow a model ingredient from the "port trade" literature: the fact that trade takes place at a subset of all locations. Thus, this paper can be viewed as bridging the gap between "port trade" and "internal trade" models. I combine ingredients of these models such that I can both speak to the empirical findings and measure the real income gains from trade in a setting in which internal trade gives rise to urbanization.[5]

Finally, the paper is related to the literature studying the effect of countries' trade openness on agglomeration more generally. An early paper in this literature is the one by Krugman and Livas Elizondo (1996). They suggest in a stylized three-location model that, in the presence of external trade, economies are less likely to give rise to agglomeration through linkages between consumers and firms. 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 study the relationship between internal trade and agglomeration. Also, I develop a

[4]In Section 3.3, I show that existing "internal trade" models do not feature a positive effect of trade on the urbanization index. As I argue in that section, this is related to the fact that the elasticity of population to market access is constant in these models.

[5]A related paper in this regard is Armenter, Koren and Nagy (2014), which models trade in a world with bridges that can be used to trade between two banks of a river. However, this model is not used to quantify the effects of trade on urbanization and real income.

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 describes Hungary's border changes after the First World War and estimates the reduced-form effect of the new border on urbanization. Section 3 presents the model and relates it to existing "internal trade" models. Section 4 provides details on the structural estimation I conduct to take the model to the data. Section 5 presents the results of the estimation, while Section 6 presents robustness checks. Section 7 concludes.

# 2 Reduced-form evidence: The effect of borders on urbanization

In this section, I exploit the unique historical experiment provided by the 1920 redrawing of Hungary's borders to empirically investigate the effect of trade on urbanization. In Section 2.1, I briefly describe the historical background behind these border changes, their unexpected and exogenous nature, and the dramatic impact they had on trade between locations separated by the new border. In Section 2.2, I discuss the data used to estimate the effect of the new border on urbanization. In Section 2.3, I present my baseline empirical specification and the results, which point to a significant negative effect of the new border on urbanization. In Section 2.4, I provide further evidence that this significant negative effect is driven by a differential impact of the border on city populations as opposed to rural populations. I show that the results are robust to a set of alternative specifications in Section 2.5.

## 2.1 Historical background

The First World War formally ended for Hungary when the country's delegation signed the *Treaty of Trianon* in Versailles, France on June 4, 1920 (Kontler, 2002). The treaty regulated the status of the Hungarian Kingdom and defined its new borders. The country's borders changed dramatically with the treaty: its land area shrinked from 325,000 km$^2$ to 93,000 km$^2$ (see Figure 1), and its population fell from 20.9 million to 7.6 million.

The territories separated by the Treaty of Trianon were integrated for several centuries. Between the 12th and the 16th centuries, both sides of the post-1920 border were under the rule of the King of Hungary. Starting at the end of the 17th century, this was followed by the rule of the Emperor of Austria on both sides of the border. With the exception of Croatia, which had its own institutions but was still highly integrated with the rest of the kingdom, there were no institutional differences between the opposite sides of the border prior to 1920. There were no restrictions on the flow of goods or people either (Kosáry, 1941). Even though, to the best of my knowledge, no historical internal trade data exist for Hungary, there is no reason to believe that trade between the opposite sides of the border was not widespread before 1920. Another fact that suggests a high level of trade integration is that the Hungarian road and railroad network was as dense as the networks of developed countries by 1910 (Kontler, 2002).

The border change involved in the treaty was completely 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 border was drawn such that it was unrelated to past political, economic and ethnic boundaries (Kontler, 2002; Teleki, 1923). Instead, the new border was a rather ad hoc compromise between neighbor countries' ambitious territorial claims, also supported by France, and a breakup of Hungary along ethnic lines, supported by Britain and the United States (Zeidler, 2014).

After the treaty, the Hungarian government made the revision of the new border its primary foreign policy goal (Kontler, 2002). Not surprisingly, this created a hostile atmosphere with neighboring countries, which also poisoned trading relationships. Teleki (1923) provides a lengthy discussion of how the hostile environment made it orders of magnitude harder to trade with the opposite side of the border. Among other things, he recalls petty disputes over tiny quantities of timber that were meant to be shipped from Serbia to Hungary in the early 1920s.

The trading relationships between Hungary and its neighbors improved slightly over the course of the 1920s. Although the country's foreign trade only started to grow substantially in the 1930s (Kosáry, 1941), Hungary's exports to its neighbors (Austria, Czechoslovakia, Romania and Yugoslavia) accounted for 8.3% of the country's GDP in 1928 (Csató, 2000; Eckstein, 1955). This number seems non-negligible, yet it is substantially below Hungary's exports to these countries today (18.6% of GDP).[6] Moreover, the 8.3% number includes trade with parts of these countries that never belonged to Hungary, such as what is today the Czech Republic or most of Austria. All in all, the available historical evidence suggests that the redrawing of Hungary's borders led to a dramatic decline in trade between the territories isolated by the new border.

## 2.2 Data

In this section, I briefly describe the data used in my empirical analysis. First, I provide information on the sources and the structure of the data. Next, I describe how I use the data to define two objects that will play a key role in the analysis: *cities* and the level of *urbanization*. Finally, I discuss how I choose the period of investigation.

[6]Hungary's total exports equaled 13.1% of GDP in 1928 (Csató, 2000; Eckstein, 1955), while they are 83.3% of GDP today.

**Population data.** The key piece of data I use is population data, which come from the 1910 and 1930 population censuses of the Hungarian Kingdom. These censuses provide the population of each *settlement* (település) in Hungary. They also group settlements into larger geographic units named *counties* (megye). As county definitions changed between 1910 and 1930, I use the 1930 county definitions throughout the analysis. Figure 2 shows a map of the 25 counties in 1930. For later use, I also geolocate the geographic center of each settlement whose population was above 2,000 inhabitants in 1910.

**Data on outmigration.** In Sections 2.4 and 4.5, I use data on the number of residents moving abroad. A 1918 special issue of the Hungarian Statistical Office reports this number by 1910 county for each year between 1900 and 1910. I sum across the years to obtain the total number of outmigrants from 1900 until 1910 by county. As the years before 1910 were characterized by massive outmigration, these numbers are large. In total, 1,050,000 people (5.0% of Hungary's 1910 population) moved abroad during these eleven years, most of them to the United States (Jagadits, 2020).

**Sectoral employment data.** Some robustness checks of Section 2.5 use sectoral employment data. The 1910 and 1930 censuses report these data at the settlement level for broad sectors such as "manufacturing" or "trade and finance." Based on these data, I define three sectoral categories: *agriculture* (consisting of "agriculture and gardening" and "other branches of farming"), *manufacturing* (consisting of "manufacturing") and an *other sector* (consisting of all the remaining sectors). Next, I aggregate employment in these sectors to the county level. This procedure thus yields employment levels in agriculture, manufacturing and the other sector for each county, both in 1910 and 1930.

**Defining cities.** Settlements were classified as cities (város), towns (nagyközség) or villages (kisközség) in the 1910 and 1930 censuses. However, 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 define any settlement with more than 20,000 inhabitants as a city.[7]

**Defining the level of urbanization.** The key object of investigation in my empirical analysis will be the level of urbanization and, in particular, how it changed with the 1920 change in borders. To this end, I construct a measure of urbanization at the county level by computing the percentage share of total county population that lived in cities (i.e., in

[7] I define Budapest and its suburbs as one 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.

settlements above 20,000 inhabitants):

$$\text{Urbanization}_c = 100 \cdot \frac{\text{City population of county } c}{\text{Total population of county } c}$$

for both 1910 and 1930. This is the measure of urbanization I use in the empirical analysis of Sections 2.3, 2.4 and 2.5.

**Period of investigation.** In Hungary, starting from 1869, one population census was carried out in each decade. Therefore, I use data from the 1910 census, as it is the last census providing a picture of the population distribution in Hungary before the border change. I combine it with data from the 1930 census, which reflects the population distribution ten years after the border change. It needs to be noted that there was a census in 1920 as well, carried out right after the change in borders. However, this census is unlikely to reflect all the effects of the border change. 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 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). 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. In Appendix A, I rerun the main empirical specifications of Sections 2.3 and 2.4 on data from the 1920 census. In line with the argument above, I show in this appendix that the effects of the border change cannot be identified from the 1920 data.

## 2.3 Urbanization decreased near the post-1920 border

Hungary underwent a rapid phase of urbanization around the end of the 19th century and the beginning of the 20th century. Figure 3 presents the evolution of urbanization (that is, the share of people living in cities) over the post-1920 territory of Hungary between 1890 and 1930. As the figure shows, urbanization increased by about 50% during these 40 years, from 23.0% to 33.8%.[8] This is almost identical to the increase in urbanization in India over the last four decades (Bhagat, 2011).

The increase in urbanization, however, was not uniform across regions. Figure 4 shows this by presenting the evolution of urbanization in two example counties: one in the center

[8]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.

of the post-1920 territory, and one close to the new border. The county in the center exhibited an increase in the share of population living in cities over the entire period, although urbanization somewhat slowed down after 1910. The county at the new border, on the other hand, suffered a decrease in urbanization after the border change, 1920. This is in line with the idea that the new border, by cutting trade, slowed down urbanization in its close proximity.

To see if these examples reflect a systematic pattern in the data, I regress county-level changes in urbanization between 1910 and 1930 (that is, the change in the share of county population that lived in cities) on the county's (log) distance from the new border,[9]

$$\text{Urbanization}_{c,1930} - \text{Urbanization}_{c,1910} = \beta_0 + \beta_1 \log\left(dist_c\right) + \epsilon_c \tag{1}$$

in my main specification. The coefficient of interest is $\beta_1$. Trade fostering urbanization is consistent with $\beta_1 > 0$. That is, counties that became more isolated from trade after the 1920 change in borders should see a decrease in urbanization relative to counties that remained central.

Column (1) of Table 1 presents the estimate of $\beta_1$. The estimate shows that a county twice as far from the new border exhibited an average 0.751 percentage points higher increase in urbanization. The effect is significant not only in a statistical, but also in a quantitative sense, as it is more than half of a county's average change in urbanization between 1910 and 1930 (1.25 percentage points). Figure 5 presents the relationship between distance from the border and the change in urbanization in a scatterplot. Looking at Figure 5, one can again see that distance from the new border was a non-negligible factor in the evolution of urbanization. While most counties close to the border saw an absolute decrease in urbanization, the majority of counties farther away saw increases between 0 and 4 percentage points.

One concern about the results of column (1) is that counties near the new border may have been special in their economic characteristics prior to the border change. In columns (2) to (5) of Table 1, I provide evidence that this was not the case. In particular, I find that 1910 population density and sectoral composition did not differ significantly between counties closer and counties farther away from the new border. The differences are not only statistically but also quantitatively insignificant. A county twice as far from the new border had, on average, 0.006 log points lower population density in 1910, while 1910 population density had a standard deviation of 0.208 log points across counties. A county twice as far from the new border had a 0.480 percentage points lower 1910 employment share in agriculture (the standard deviation of this variable is 10.0 percentage points), a

[9] I measure distance from the border as the shortest "as the crow flies" distance between the settlement that served as the *county seat* (the seat of the county's government) and the border. In the main specifications of Sections 2.3 and 2.4, I restrict the sample to counties that are at most 60 kilometers from the new border. Section 2.5 shows that results are robust to including all counties in the analysis.

0.278 percentage points higher share in manufacturing (standard deviation 4.9 percentage points) and a 0.201 percentage points higher share in the rest of the economy (standard deviation 5.5 percentage points).

Finally, the last column of Table 1 presents the results of a placebo exercise in which I regress changes in urbanization between 1890 and 1910 on distance from the post-1920 border. As the table shows, the relationship between these two variables is not statistically significant. Moreover, the point estimate is less than one third of the estimated $\beta_1$ coefficient in column (1), and the $R^2$ of the placebo regression is only 0.005, down from 0.125 in the main specification. These findings confirm that post-1920 borders were not significantly related to county-level trends of urbanization prior to the border change.

This section presented two sets of results. First, there were no significant differences in 1910 economic characteristics or pre-trends of urbanization between Hungarian counties that became peripheral after the 1920 change in borders and counties that remained central. Second, these counties exhibited significantly different urbanization patterns once the new border was present, in a direction that is consistent with the urbanization-fostering effect of trade.

Looking at the patterns in a reduced-form fashion, instead of using the model of Section 3, 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.

## 2.4 Decomposition: The effect of the new border on city and rural populations

In this section, I examine the forces behind the significant negative effect of the post-1920 border on urbanization. Urbanization, defined as the share of total county population that lived in cities, can decrease in border counties if city populations decrease more, or increase less, than rural populations in these counties. In what follows, I run the following specifications to decompose the change in urbanization into changes in city and rural populations:

$$\log\left(\text{City population}_{c,1930}\right) - \log\left(\text{City population}_{c,1910}\right) = \gamma_0^C + \gamma_1^C \log\left(dist_c\right) + \eta_c^C \quad (2)$$

$$\log\left(\text{Rural population}_{c,1930}\right) - \log\left(\text{Rural population}_{c,1910}\right) = \gamma_0^R + \gamma_1^R \log\left(dist_c\right) + \eta_c^R \quad (3)$$

where $c$ indexes counties. Thus, specification (2) regresses the change in a county's (log) total city population on the county's (log) distance from the new border, while specification

(3) regresses the change in (log) total rural population – that is, county population minus city population – on the same variable. The coefficients of interest are $\gamma_1^C$ and $\gamma_1^R$. Based on the results of Section 2.3, we expect $\gamma_1^C > \gamma_1^R$. That is, counties that became more isolated from trade after the 1920 border change should see a larger decrease, or a smaller increase, in the population living in their cities than in their rural areas.

Columns (1) and (2) of Table 2 present the estimates of $\gamma_1^C$ and $\gamma_1^R$. The results confirm that $\gamma_1^C > \gamma_1^R$, providing further evidence on the urbanization-hindering effect of borders. In particular, the results show that $\gamma_1^C$ is positive and significantly different from zero, while $\gamma_1^R$ is small and insignificant.

The effect of the new border on city populations, $\gamma_1^C$, is particularly striking for two reasons. First, some countries had zero city population in both 1910 and 1930; these observations are dropped when taking logs.[10] Nonetheless, the coefficient is statistically significant, even with this reduced sample size (15 counties). Second, the magnitude of the point estimate, 0.120, suggests a quantitatively large impact of the new border on city populations. This point estimate is more than half of a county's average increase in its city population between 1910 and 1930 (0.224 log points). It implies that doubling distance from the border led to an approximately 12% increase in the number of people living in cities.

The effect of the new border on rural populations, $\gamma_1^R$, is insignificant both in a statistical and in a quantitative sense. The point estimate, $-0.012$, suggests that a county twice as far from the new border experienced an approximately 1.2% drop in its rural population. This estimate is small, both compared to the effect on city populations (0.120) and to a county's average change in its rural population between 1910 and 1930 (+0.092 log points).

The lack of rural populations' response to the border change could, in principle, come from rural residents facing disproportionately high barriers to mobility. This could either stem from cultural differences, or from the fact that rural residents had lower income than city residents. In fact, there is a literature suggesting that mobility is positively related to income in current developing countries (Bryan, Chowdhury and Mobarak, 2014), although evidence on this is somewhat mixed (Bazzi, 2017). Unfortunately, the censuses do not provide data on mobility across settlements. Nevertheless, there are two pieces of evidence that point against this explanation. First, rural populations *did* change from 1910 to 1930, by +0.092 log points in the average county and a standard deviation of 0.050 log points across counties. These are fairly large changes (approximately 9.2% and 5.0%) over a 20-year period. It is just that these changes in rural populations were not related to distance from the new border, as column (2) of Table 2 shows.

Another piece of evidence against rural residents' lower mobility comes from the data on outmigration. As described in Section 2.2, I have data on the number of residents

[10] In principle, counties that had zero city population in one year but positive city population in the other would be dropped as well. However, there are no such counties in the data.

moving abroad from 1900 until 1910 by county. If rural residents face disproportionately high barriers to mobility, then one would expect that fewer of them could move abroad as well. This is not what I find. Figure 6 plots the number of outmigrants relative to 1910 county population against counties' level of urbanization in 1910. The relationship between these two variables is *negative*, suggesting that rural regions saw a higher fraction of their residents moving.[11] Some of the most rural counties saw 10 to 20% of their population moving abroad between 1900 and 1910, which is even higher in magnitude than a county's average change in its rural population between 1910 and 1930. Although cross-country and within-country migration patterns might differ, these findings provide evidence that a high fraction of rural residents could leave their places of residence in early-20th century Hungary.

Columns (3) and (4) of Table 2 present the results of two placebo exercises in which I regress the 1890 to 1910 change in city and rural populations on distance from the post-1920 border. These results confirm that neither city nor rural populations exhibited pre-trends that were systematically related to distance from the new border. In the case of city populations (column 3), the estimated coefficient (0.040) is statistically insignificant, and the point estimate is only one third of the estimated effect after 1910. Moreover, the $R^2$ of the regression drops from 0.299 in the main specification to 0.042 in the placebo specification. The estimated coefficient is also small and statistically insignificant for rural populations (column 4). In this case, the $R^2$ of the regression drops from 0.095 in the main specification to 0.006 in the placebo specification.

This section showed that the negative effect of post-1920 borders on urbanization is driven by a significant and large decrease in city populations in counties near the new border, as rural populations did not change significantly in these counties. Thus, it was the substantial effect on city populations that led to less urbanization in border counties.

## 2.5 Robustness

This section presents a series of robustness checks to the key empirical finding of Sections 2.3 and 2.4: the fact that changes in urbanization and city populations between 1910 and 1930 were positively related to distance from the post-1920 border. One important concern is that changes in sectoral composition, if related to distance from the new border, could drive the results. This may be true even if counties near the new border and counties farther away were not significantly different in their sectoral composition prior to the border change (columns 3 to 5, Table 1).

In columns (1) to (3) of Table 3, I investigate whether county-level sectoral employment shares changed differentially with distance from the new border. I find that a county twice

[11] Regressing the share of population abroad on the level of urbanization, the estimated coefficient is −0.088, significantly different from zero at a 1% level. Dropping counties with zero urbanization decreases the point estimate somewhat to −0.051, but the estimate remains statistically significant at a 1% level.

as far from the border increased its employment share in agriculture by 1.78 percentage points, while it decreased its employment share in the other sector by 1.23 percentage points. The effect of the new border on the manufacturing employment share was statistically insignificant. Thus, sectoral composition did change systematically with distance from the post-1920 border.

It is important to note that these empirical findings do not contradict any predictions of the model I develop in Section 3. As that model only has one sector, it does not provide any predictions on how sectoral composition should change with border changes. Nonetheless, it is an empirical concern that my headline finding, the positive relationship between urbanization and distance from the new border, could be fully driven by these sectoral changes.

To address this concern, I first run a version of equations (1) and (2) in which I control for changes in the shares of agriculture and the other sector (the two sectors on which I found significant effects in columns 1 and 3 of Table 3) on the right-hand side. Thus, these specifications ask whether there are any effects of the border on urbanization and city populations that do not operate through changing sectoral shares. Columns (4) and (5) of Table 3 present the results of these specifications. I find that, once sectoral changes are controlled for, the effect of distance from the border on urbanization, if anything, is larger in magnitude and still statistically significant. The effect on city populations is effectively unchanged, both in terms of magnitude and in terms of statistical significance.

Although the results of columns (4) and (5) are reassuring, it is important to note that sectoral changes are likely endogenous. As a consequence, controlling for them in the reduced form cannot alleviate all concerns about how they affected the relationship between the new border and urbanization. Therefore, I also take a structural approach to address these concerns. In particular, I develop a multi-sector extension of my baseline one-sector model and take it to the data in Section 6.2. I show that this extended model delivers quantitative results that are very similar to the results of my baseline model.

In columns (6) to (11), I conduct further robustness checks. Column (6) includes distance, rather than log distance, on the right-hand side of equation (1). Column (7) includes all Hungarian counties in the sample, while column (8) only includes those that are at most 40 kilometers from the new border. Columns (9) to (11) repeat these alternative specifications for equation (2). Reassuringly, I find that distance from the new border had a significant positive effect on urbanization and city populations in all of these specifications.

# 3 A model of trade and urbanization in space

Motivated by the empirical findings of Section 2, I develop a quantitative spatial model of trade in this section. Section 3.1 presents the setup. Section 3.2 offers a model-based measure of urbanization and shows the main prediction of the model, which rationalizes the reduced-form evidence of Section 2: the fact that trade fosters urbanization. Finally, Section 3.3 shows that this prediction of the model is *not* shared by existing "internal trade" models of economic geography, and provides intuition for this negative result.

## 3.1 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.[12] 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.[13] Goods are not tradable with the rest of the world. Hence, the borders of the country constitute an impassable barrier to trade.[14]

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.

### 3.1.1 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

$$u_m(r,i) = a(r,i) + \varsigma(\mu_m, r)^{-1} \left[ \sum_{j=1}^{\overline{L}} c_m^j(r,i)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{4}$$

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

As goods are substitutes, workers demand a positive quantity of each worker's product in equilibrium, which gives rise to trade across all workers. Thus, the model belongs to

[12]Although the assumption that workers produce at home may be relevant in historical contexts, it is at odds with reality today. However, Appendix C.1 presents a model with firms that employ workers to produce goods, and shows a formal isomorphism between the two models.

[13]The model does not explicitly feature nontradable goods such as housing. Even though some quantitative economic geography models do not feature nontradables either (Allen and Arkolakis, 2014; Desmet, Nagy and Rossi-Hansberg, 2018; Donaldson and Hornbeck, 2016; Nagy, 2020), others do (Monte et al., 2018; Redding, 2016). However, Appendix C.2 presents a model in which workers also consume a homogenous nontradable good at their residential location, and shows that this extended model is isomorphic to my baseline model.

[14]In Section 2.1, I argued that the assumption of impassable borders is a good approximation to reality in Hungary after the First World War.

the tradition of "internal trade" models of economic geography discussed in Section 1. In Section 3.3, I further elaborate on the connection between the model and existing "internal trade" models. In that section, I show why the model of this paper delivers an urbanization-fostering effect of trade (Section 3.2), while existing "internal trade" models of economic geography do not have this feature.

Although the model features a large number of goods, these goods are not grouped into sectors. This simplifies the structure of the model and allows for a theoretical characterization of the effect of trade on urbanization and real income in Section 3.2. However, it also implies that the model cannot speak to the sector-level empirical results found in Section 2.5. Moreover, one might wonder to what extent the quantitative findings of the paper are influenced by the lack of sectors in the model. To address these concerns, I develop an extension of the model with multiple sectors in Section 6.2, take it to sector-level data, and show that my headline quantitative findings carry over to this multi-sector framework.

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:

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

$\theta$ is a positive constant that drives the degree of heterogeneity in idiosyncratic location tastes, and thus the dispersion of population in equilibrium.[15] 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, $\theta$ can also be viewed as a parameter showing the severity of frictions to labor mobility.

### 3.1.2 Production

Producing a unit of a good requires one unit of labor. Having produced their goods and having shipped them 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

[15] For similar formulations of the dispersion force in economic geography models, see Redding (2016) or Desmet et al. (2018).

goods at $m$, $P_m$, as given. She also takes into account that shipping goods across trading places is possible, but is subject to an iceberg shipping 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 $o$.

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.

### 3.1.3 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 $a: S \to \mathbb{R}_+$, $\{\tau, \varsigma\}: S^2 \to \mathbb{R}_+$, 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 (4) subject to the production technology and their budget constraint.*

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

$$x_m^j = \sum_o \tau(\mu_m, \mu_o)^{1-\sigma} \left(p_m^j\right)^{-\sigma} P_o^{\sigma-1} p_o L_o \tag{5}$$

*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 $o$.*[16]

[16]The right-hand side of equation (5) follows from CES demand for worker $j$'s product at any trading place $o$.

Appendix B.1 shows that the spatial distribution of population is governed by the following system of equations in equilibrium:

$$\log L(r) = \nu + \theta^{-1}\left[a(r) + \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}}\right] \tag{6}$$

$$\varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \geq \varsigma(\mu_m, r)^{-1} MA_m^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \qquad \forall m \tag{7}$$

$$MA_m = \sum_o MA_o^{-\frac{\sigma-1}{\sigma}} L_o \tau(\mu_m, \mu_o)^{1-\sigma} \tag{8}$$

$$L_m = \sum_{r:\ m=\mu(r)} L(r) \tag{9}$$

where $L(r)$ is the population of location $r$, $\nu$ is a combination of parameters, and $MA_m$ is the *market access* of trading place $m$, implicitly defined by equation (8).

To gather intuition for equations (6) to (9), note that equation (6) determines population at $r$ as an increasing function of local amenities $a(r)$, a decreasing function of shipping costs to the trading place $\varsigma(\mu(r), r)$, and an increasing function of market access at the trading place, $MA_{\mu(r)}$. Equation (7) 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 $\varsigma(\mu_m, r)^{-1}$ and market access. Equation (8) implies that a trading place $m$ has better market access if it has large trading places (large $L_o$) in its surroundings (low $\tau(\mu_m, \mu_o)$). Appendix B.1 also shows that the level of real income at a trading place, $\omega_m = \frac{p_m}{P_m}$, is an increasing function of market access:

$$\omega_m = MA_m^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \tag{10}$$

Finally, equation (9) 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 (6) sheds light on an important feature of the equilibrium. Take two 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 $\varsigma(\mu(r), r) < \varsigma(\mu(r), s)$, hence equation (6) implies $L(r) > L(s)$. That is, population decreases with distance from the trading place: *cities with a negative population gradient* form around trading places.

In the next section, I use equations (6), (8) and (10) to further explore the relationship among trade, urbanization and real income in the model.

## 3.2 Urbanization and the real income gains from trade

In this section, I use the model outlined in Section 3.1 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 real income gains from trade. These propositions allow me to show that trade induces urbanization and real income gains in the model.

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

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

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

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 $\mu\left(r\right)$ through to the equation*

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

*Proof.* Partially differentiating equation (6) with respect to $\varsigma\left(\mu\left(r\right),r\right)^{-1}$ gives the result. □

As the exponent on market access, $\frac{2\sigma-1}{\sigma(\sigma-1)}$, is positive, Proposition 1 states that trading places with good trading opportunities (good market access) have more urbanized surroundings. This also implies that borders, by imposing barriers on trade, are likely to negatively affect urbanization. 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. Thus, Proposition 1 shows that the model has the ability to replicate the reduced-form findings of Section 2, which show declining urbanization near Hungary's new borders.

At an intuitive level, the result of Proposition 1 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 Proposition 1 and equation (10), 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 real income 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 shipping costs or country borders. Define the real income gains from trade at 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 shipping 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 real income gains from trade.*

The relationship between the real income 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 how much their real income levels increase as a result of trade. In Section 4, I structurally estimate the model to quantify the real income effects of the redrawing of Hungary's borders.

### 3.3 Urbanization and trade: A more general "internal trade" framework

Is the urbanization-fostering effect of trade a common prediction of "internal trade" models of economic geography? In this section, I show that the answer to this question is negative and provide intuition for this negative result. To this end, I generalize the model of Section 3.1 by changing the utility function of a worker $i$ living at $r$ and trading at $m$ to

$$u_m(r,i) = a(r,i) + F\left(\varsigma(\mu_m, r)^{-1}\left[\sum_{j=1}^{\overline{L}} c_m^j(r,i)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}\right) \tag{11}$$

such that $F' > 0$. Thus, $F(\cdot)$ drives the worker's utility as a function of her consumption of tradables. The model of Section 3.1 is a special case in which $F(x) = x$. The rest of the model, including the definition of the equilibrium, is unchanged relative to Section 3.1.

Appendix B.2 derives the equilibrium conditions of this more general model. It shows that equation (6) becomes

$$\log L(r) = \nu + \theta^{-1}\left[a(r) + F\left(\varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}}\right)\right] \tag{12}$$

while the remaining equilibrium conditions, (7) to (9), are unchanged as they are not influenced by the $F(\cdot)$ function. Thus, the equilibrium distribution of population and market access is driven by equations (12), (7), (8) and (9) in the more general model.

As already mentioned, the model of Section 3.1 is a special case of this more general framework, such that $F(x) = x$. Another interesting special case is $F(x) = \log(x)$. In this latter case, equation (12) can be rearranged to obtain

$$L(r) = e^{\nu+\theta^{-1}a(r)}\varsigma(\mu(r), r)^{-\theta^{-1}} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}\theta^{-1}} \tag{13}$$

from which, summing across locations that trade at $m$,

$$L_m = A_m MA_m^{\frac{2\sigma-1}{\sigma(\sigma-1)}\theta^{-1}} \tag{14}$$

where

$$A_m = \sum_{r:m=\mu(r)} e^{\nu+\theta^{-1}a(r)}\varsigma(\mu_m, r)^{-\theta^{-1}}.$$

Expressing $MA_m$ from equation (14) and plugging it into equation (8) yields

$$A_m^{-\frac{\sigma(\sigma-1)}{2\sigma-1}\theta} L_m^{\frac{\sigma(\sigma-1)}{2\sigma-1}\theta} = \sum_o A_o^{\frac{(\sigma-1)^2}{2\sigma-1}\theta} L_o^{1-\frac{(\sigma-1)^2}{2\sigma-1}\theta}\tau(\mu_m, \mu_o)^{1-\sigma}. \tag{15}$$

Equation (15) resembles the equation determining the spatial distribution of population in the "internal trade" models of Allen and Arkolakis (2014), Donaldson and Hornbeck (2016) and Redding (2016). The only difference is that $A_m$ is endogenously determined in this model, while it would be exogenous in Allen and Arkolakis (2014), Donaldson and Hornbeck (2016) and Redding (2016). However, $A_m$ is in fact exogenous if trading place assignments to locations are exogenous.[17] In this special case, $A_m$ becomes an exogenous function of location amenities and shipping costs, and equation (15) becomes isomorphic to the equilibrium conditions of Allen and Arkolakis (2014), Donaldson and Hornbeck (2016) and Redding (2016). Hence, the general framework of this section nests not only the model of Section 3.1 but also these models as special cases.[18]

[17]One sufficient condition for trading place assignments being exogenous is that for every location $r$, there exists a unique trading place $\mu(r)$ such that $\varsigma(\mu(r), r)$ is finite but $\varsigma(s, r) = \infty$ for any $s \neq \mu(r)$.

[18]It needs to be noted that these models feature additional dimensions of heterogeneity across locations: exogenous differences in productivity (Allen and Arkolakis, 2014), or in productivity and the amount of land (Donaldson and Hornbeck, 2016; Redding, 2016). The formal isomorphism holds between the model

Equation (12) also allows me to express the urbanization index, as defined by Definition 2 of Section 3.2, in the more general model as

$$UI(r) = \theta^{-1} F' \left( \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \right) MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}}. \tag{16}$$

As equation (16) shows, the urbanization index is not necessarily increasing in market access in the more general model, since the relationship between these two objects is driven by the shape of $F'(\cdot)$. The following proposition provides a sufficient condition on the $F(\cdot)$ function under which urbanization is increasing in market access.

**Proposition 3.** *The urbanization index is increasing in market access if* $\frac{F''(x)}{F'(x)} x > -1$ *for any* $x > 0$.

*Proof.* Differentiating (16) with respect to market access yields

$$\begin{aligned} \frac{\partial UI(r)}{\partial MA_{\mu(r)}} = \theta^{-1} \frac{2\sigma-1}{\sigma(\sigma-1)} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}-1} \Bigg[ & F' \left( \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \right) \\ & + F'' \left( \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \right) \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \Bigg]. \end{aligned}$$

If the inequality $\frac{F''(x)}{F'(x)} x > -1$ holds for any $x$, then it holds for $x = \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}}$. As a result, we have

$$F' \left( \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \right) + F'' \left( \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \right) \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}} > 0$$

and hence $\frac{\partial UI(r)}{\partial MA_{\mu(r)}} > 0$. □

In the model of Section 3.1, $F(x) = x$, implying $F''(x) = 0$ and therefore $\frac{F''(x)}{F'(x)} x > -1$. Thus, the inequality holds and urbanization is increasing in market access. On the other hand, the special case $F(x) = \log(x)$ (the version of the model for which I showed isomorphisms with existing "internal trade" models) does not satisfy the inequality. This is because $\frac{F''(x)}{F'(x)} x = -1$ in this case. In fact, we obtain

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

implying that urbanization is independent of market access in this class of models.

of Section 3.3 and special cases of these models *without* these additional dimensions of heterogeneity. Differences in productivity and land would show up as additional exogenous terms multiplying the left- and right-hand sides of equation (15) but would not influence the elasticity of a location's population to market access in equation (13). As I argue later, the main conclusion of this section – the fact that trade does not foster urbanization in these models – hinges on the fact that this elasticity is a constant. As a result, the conclusion would be unchanged even with these additional dimensions of heterogeneity.

Why does trade not foster urbanization in existing "internal trade" models of economic geography? This can be best understood from looking back to equation (13). As this equation makes clear, the elasticity of a location's population to its market access is constant in these models. In other words, a 1% increase in market access triggers the same population response at all locations that experience this increase in trading opportunities. As a result, trade affects population but has no heterogeneous impact between urban and rural locations.

The model of Section 3.1 moves away from this constant elasticity. It does so by not only borrowing the concept of trading places from the "port trade" literature but also choosing the functional form of $F(\cdot)$ such that trade has a positive effect on urbanization around trading places. This is what allows the model to replicate the reduced-form effects of Hungary's 1920 border change on urbanization estimated in Section 2. The next section uses the structure of the model to estimate the real income effects of the border change.

# 4 Structural estimation

To estimate the real income effects of Hungary's 1920 border change, I combine the model with the data in this section. As a first step, I define the two key spatial units of analysis: locations (Section 4.1) and trading places (Section 4.2). Next, I choose the functional form of shipping costs across trading places $\tau(\cdot,\cdot)$ and between trading places and workers' locations $\varsigma(\cdot,\cdot)$ (Section 4.3). Next, I choose the value of structural parameter $\sigma$ and the structural parameters entering shipping costs (Section 4.4). Finally, I simultaneously back out the distribution of city amenities that rationalize the data and structurally estimate the value of key parameter $\theta$ using the exogeneity of the 1920 border change (Section 4.5).

## 4.1 Defining locations

To obtain predictions at a high level of spatial disaggregation, I use a very fine discretization of space when defining locations in the data. In particular, I set up a spatial grid of the territory of Hungary, with grid cells of the size 0.01° by 0.01° (approximately 1 by 1 kilometer). In both 1910 and 1930, I define a location as a grid cell that belongs to Hungary in that year. Having defined locations in this way, the model can deliver predictions on population and economic activity at each of the approximately 380,000 locations in 1910 and approximately 110,000 locations in 1930.

Despite the large number of locations, the relatively simple structure of the model leads to quick calculations. In particular, backing out the distribution of city amenities that rationalize the data only takes a few minutes on a typical personal computer. Simulating the model with a given set of amenities, which I do for alternative borders in Section 5.4, requires a similar amount of time.

## 4.2 Defining trading places

Motivated by the model's prediction that cities form around trading places, I assume that each city in the data had a trading place in its geographic center, both in 1910 and in 1930.[19] 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 the sector "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.

This definition of trading places implies that Hungary had 58 trading places in 1910. Out of these 58 trading places, 35 remained in the country after the 1920 border change. Moreover, four settlements in Hungary's post-1920 territory exceeded the 20,000-inhabitant threshold between 1910 and 1930. Hence, I treat their geographic centers as trading places as well, making the total number of trading places 39 in 1930. Importantly, I do not need to assume that 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 model counterfactuals 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.[20]

## 4.3 Choosing the functional form of shipping costs

Taking the model to the data requires me to specify the functional form of shipping costs across trading places $\tau(\cdot,\cdot)$ and shipping costs between trading places and workers' locations $\varsigma(\cdot,\cdot)$. I assume that both types of shipping costs were exponential functions of distance $dist(\cdot,\cdot)$:

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

[19] Recall that I defined cities as settlements above 20,000 inhabitants in the data, and geolocated their centers in Section 2.2.

[20] See Section 4.5 for the specific exogeneity assumption I make in the estimation. Also note that while the issue of trading place endogeneity is not present in the estimation, it *is* present when I simulate model counterfactuals with alternative borders in Section 5.4. I discuss how I address the issue in that section.

The exponential formulation of shipping costs is a frequently used assumption in the economic geography literature (Fujita, Krugman and Venables, 1999; Rossi-Hansberg, 2005; 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).

Also 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 large concern, given the historical evidence on little trade over the 1920s between Hungary and its neighbors (Section 2.1).

## 4.4 Choosing the values of structural parameters $\phi$, $\psi$ and $\sigma$

To calculate shipping costs, I need to choose the values of shipping cost elasticities $\phi$ and $\psi$. I match $\phi$ to evidence on shipping 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 geographic centers of Fiume and Budapest are equal to 16.2%, which corresponds to the average of the above three numbers.

I choose the value of $\psi$ such that the model matches the standard deviation of the population 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 (6) –, the standard deviation is strictly increasing in $\psi$ in the model. This implies that the parameter is identified. The procedure pins down a value of $\psi = 1.02 \cdot 10^{-1}$.

Finally, I choose the value of structural parameter $\sigma$ 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 $\sigma = 5$.

## 4.5 Backing out city amenities and estimating parameter $\theta$

Using the fact that the amenity function, $a(r)$, can take any form, I match the model exactly to cities' population levels in the data, both in 1910 and in 1930.[21] To this end, I assume the following amenity function: $a(r) = a_c$ if location $r$ is within the geographic

[21]Section 2.2 provides additional information on the population data.

boundaries of city $c$, and $a(r) = 0$ otherwise.[22] Thus, $a_c$ can be interpreted as the level of amenities that city $c$ provides to its residents.

Unfortunately, the censuses do not provide data on cities' actual geographic boundaries. However, they do provide their land area in square kilometers. Thus, 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 land 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.[23] I also match the total population of the country in both 1910 and 1930 by setting total population $\bar{L}$ to the population of Hungary reported by the census in each of these two years.

As this procedure of backing out amenities uses the structure of the model, it naturally depends on the values of shipping costs and structural parameters. Note that I chose these already in Sections 4.3 and 4.4, with the exception of parameter $\theta$. In fact, $\theta$ is a crucial parameter for the estimation of real income effects, as it directly influences the effects of trade on urbanization (Proposition 1) and therefore on real income (Proposition 2).

Given the central role that $\theta$ plays in shaping real income and urbanization, I exploit the natural experiment provided by the redrawing of Hungary's borders to estimate this parameter. I conduct the estimation using simulated method of moments. More precisely, I impose the following moment condition, akin to the moment conditions applied by Ahlfeldt et al. (2015) to study the division of Berlin after the Second World War:

$$corr\left[\Delta a_c, dist\left(\mu_c, \text{border}\right)\right] = 0 \tag{17}$$

where $\Delta a_c$ is the change in city $c$'s amenities between 1910 and 1930 as implied by the model, and $dist\left(\mu_c, border\right)$ 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.[24] The values of $\Delta a_c$ come from the procedure of backing out amenities described above. Thus, in practice, I simultaneously search for the set of amenities that rationalize the data in 1910 and 1930 and the value of $\theta$ that satisfies moment condition (17).

How plausible is the assumption that changes in city amenities were uncorrelated with distance from the new border? In the model, amenities serve as structural residuals and thus capture any factors influencing population that are not linked to market access. A concern is

[22] Recall that consumers' utility is additively separable in amenities, hence $a(r) = 0$ does not imply infinitely bad amenities.

[23] In principle, there could exist different sets of amenity levels that are consistent with the city populations 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 populations is a singleton, at least for the specific geography and values of structural parameters used in this paper.

[24] When calculating (17), I use the set of 35 cities that already exceeded the 20,000-inhabitant threshold in 1910 and remained in the country after the border change.

that these factors could have changed in a way between 1910 and 1930 that is systematically related to distance from the new border. One such factor could be related to labor mobility. In particular, if frictions to mobility had been increasing in distance, the new border could have made its surroundings more isolated not only from trade but also from labor flows. As the censuses do not provide data on mobility across settlements, I cannot study the relationship between internal mobility and distance directly. However, my 1910 data on outmigration (Section 2.2) offer an indirect way of studying this relationship. Historical evidence suggests that most outmigration took place through Fiume (now Rijeka, Croatia), Hungary's major sea port (Jagadits, 2020). If frictions to mobility had been increasing in distance, then one would expect more outmigration from counties near Fiume than from counties farther away. Plotting outmigration against distance to Fiume, I find no systematic relationship between these two variables (Figure 7).[25] Needless to say, outmigration may exhibit different patterns than internal migration. Yet, it is reassuring that outmigration flows do not show any signs of being related to distance.

Other factors that could have changed amenities differentially in border cities are increased police or military presence (which is also discussed by Redding and Sturm (2008) in the context of Germany), or government subsidies aimed specifically at these cities. To the best of my knowledge, no data are available on these factors. Historical evidence suggests that government subsidies were present, primarily in the form of relocating universities to cities near the new border (Kosáry, 1941). If these subsidies had attracted more people to border cities, they would bias my estimated losses in urbanization and real income downwards. Although I could not find any direct evidence on this, it is not unreasonable to think that such policies were at least partly aimed at mitigating the losses from these cities' declining trading opportunities.

The structural estimation identifies a point estimate of $\theta = 11.49$, with a standard error of 3.71.[26] Recall that $\theta$ captures the heterogeneity of workers' idiosyncratic tastes for locations. Equations (6) and (10), together with the fact that $\varsigma(\mu_m, \mu_m) = 1$, also imply that a one unit change in real income at a trading place ($\omega_m$) leads to a $\theta^{-1}$ change in log population at the trading place ($\log L(\mu_m)$). As this relationship involves real income *levels* rather than logs, the value of $\theta^{-1}$ is not directly comparable to the elasticity of population to real income, the object typically estimated in the literature.[27] Nonetheless, I can use $\theta^{-1}$ to evaluate this elasticity at any real income level $\bar{\omega}$ since the elasticity of population

[25] Regressing outmigration on (log) distance to Fiume, the point estimate of the regression coefficient is positive (+0.012), implying, if anything, more outmigrants from counties far away from the port. However, the relationship is not statistically significant (p-value 0.440).

[26] I compute standard errors using a bootstrap procedure. In particular, I resample (with replacement) the set of cities used in the calculation of (17) and recompute the combination of $a_c$ (both in 1910 and 1930) and $\theta$ that simultaneously match city populations in both years and satisfy moment condition (17) for the new sample. I repeat this procedure 25 times.

[27] In light of the discussion of Section 3.3, it is not surprising that the model does not imply a constant elasticity of population to real income. As I argued there, the model's key prediction that trade fosters urbanization is closely linked to this non-constant elasticity.

to real income equals

$$\frac{\partial \log L(\mu_m)}{\partial \log \omega_m}(\bar{\omega}) = \frac{\partial \log L(\mu_m)}{\partial \omega_m}(\bar{\omega}) \cdot \frac{\partial \omega_m}{\partial \log \omega_m}(\bar{\omega}) = \frac{\partial \log L(\mu_m)}{\partial \omega_m}(\bar{\omega}) \cdot \bar{\omega} = \theta^{-1}\bar{\omega}.$$

Evaluating $\frac{\partial \log L(\mu_m)}{\partial \log \omega_m}(\bar{\omega})$ at the minimum and maximum real income level implied by the model for 1910 and 1930, I find that the elasticity of population to real income varied between 2.29 and 2.89. These values are in the ballpark of the elasticities estimated by Morten and Oliveira (2018) for late-20th century Brasil (1.91), by Tombe and Zhu (2019) for today's China (2.54), and by Monte et al. (2018) for today's United States (3.30).

# 5 Results

I present the results of the structural estimation in this section. In Section 5.1, I show that, as expected, the new border led to a dramatic decrease in trade in its surroundings. In Section 5.2, I present the estimated changes in real income and urbanization. In Section 5.3, I investigate the model's fit to settlement-level population data coming from the census. In Section 5.4, I study the importance of territories lost to different countries (Austria, Czechoslovakia, Romania and Yugoslavia) in shaping the overall losses from the new border.

## 5.1 The effect of the new border on trade

With the estimated model at hand, I can measure the total amount of trade that a trading place $m$ has with other trading places, given by

$$\text{Trade}_m = \sum_{o \neq m} \tau(\mu_m, \mu_o)^{1-\sigma} p_m^{1-\sigma} L_m P_o^{\sigma-1} p_o L_o \tag{18}$$

both in 1910 and in 1930.[28] This allows me to quantify the fraction of trade at $m$ lost in the model (in percentages) as

$$100 \cdot \frac{\text{Trade}_{m,1910} - \text{Trade}_{m,1930}}{\text{Trade}_{m,1910}}.$$

These losses are substantially larger near the post-1920 border. This can be seen in Figure 8, which plots the fraction of trade lost at a trading place against distance between the trading place and the new border. While some trading places in the border's close proximity suffered losses in the range of 20 to 60%, most trading places farther away lost

[28]Equation (18) follows from two facts: (i) the quantity of a good $j$ produced at $m$ and sold at $o$ equals $x_{mo}^j = \tau(\mu_m, \mu_o)^{1-\sigma} \left(p_m^j\right)^{-\sigma} P_o^{\sigma-1} p_o L_o$ (see the right-hand side of equation 5), and (ii) the prices of individual goods equalize, i.e., $p_m^j = p_m$ (as shown in Appendix B.1), implying that the total amount of trade between $m$ and $o$ equals $p_m x_{mo}^j L_m$.

substantially less, or even saw moderate gains in their amount of trade with others.[29] While, in principle, changes in trade may also stem from changes in city amenities between 1910 and 1930, this could only influence the relationship between trade and distance from the new border if changes in amenities were correlated with distance from the new border. One advantage of my estimation strategy is that, by imposing moment condition (17), it rules out this possibility.

Regressing the fraction of trade lost on (log) distance from the new border, I obtain a point estimate of $-14.30$, significantly different from zero at a 5% level (column 1 of Table 4). This implies that a trading place twice as far from the new border experienced a 14.30% smaller drop in trade. These findings underscore that the border isolated regions in its close proximity from their trading partners, to a degree that is both statistically and economically significant. In the next section, I study how this large shock to trade translated into losses in Hungarian residents' real income and the degree of urbanization.

## 5.2 The effect of the new border on real income and urbanization

The top panel of Table 5 presents the estimated effects on Hungarian residents' real income. The effect of border changes on average real income per capita, calculated using the real income gains from trade (GFT) formula of Proposition 2, is tightly identified. On average, the new border led to a 15.55% decline in real income (standard error 0.28%), with a standard deviation of 1.08% across locations (standard error 0.24%).[30] Figure 9 presents the smoothed empirical density of the losses, calculated at the point estimate of $\theta = 11.49$. It can be seen from the figure that the redrawing of borders had a considerably heterogeneous impact across Hungarian locations. The 2nd and 98th percentiles of the distribution are 13.9% and 17.6%, almost four percentage points apart from one another. The top panel of Table 5 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 9.68%. 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.

Figure 10 shows the geography of real income losses at the point estimate of $\theta = 11.49$. The figure confirms that the smallest losses were mostly incurred at central locations, while

[29]The gains in more central regions might be due to the reallocation of population toward these regions, as well as to overall population growth between 1910 and 1930, which, everything else fixed, increases trade (see equation 18). Section 5.2, however, shows that real income per capita fell even in these more central regions.

[30]I weight by locations' 1930 population when calculating the average and the standard deviation of real income losses. Weighting by 1910 population yields virtually identical results: an average of 15.58% and a standard deviation of 1.09%.

places near 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 Figure 1).

Changes in the urbanization index $UI(r)$, the model-implied measure of urbanization, coincide with changes in real income according to Proposition 2. Hence, the real income losses shown in Figure 10 can also be interpreted as the model-implied declines in urbanization due to the new border. As already discussed, the largest declines can be seen near the new Eastern border, while the smallest declines are mostly experienced by central locations. This prediction of the model is in line with the reduced-form evidence of Section 2, which shows a significant negative impact of the new border on neighboring counties' degree of urbanization and city populations.

The predictions that the model delivers on the urbanization index cannot be tested directly as the urbanization index has no observable counterpart in the data. However, Figure 11 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 in the data by county.[31] The left panel of Figure 11 plots this measure against distance from the new border in the model, while the right panel plots the same relationship in the data. The figures suggest that, both in the model and in the data, counties near the new border saw smaller increases, or even decreases, in their level of urbanization between 1910 and 1930. Columns (2) and (3) of Table 4 present the results of regressing these changes in observed urbanization on (log) distance from the new border. According to the point estimates, a county twice as far from the new border exhibited a 0.833 percentage points higher increase in the population of its cities relative to its smaller settlements in the model. The corresponding number is 1.259 in the data. While the effect is statistically significant in the data, it is slightly outside the 10% significance bound in the model.[32] Although the lack of statistical significance in the model warrants caution when interpreting these estimates, the point estimates seem to suggest that the model is able to account for the vast majority of the observed differences in urbanization in the data.

## 5.3 Model fit to settlement populations

In this section, I evaluate the model's ability to fit the population distribution in the data, both before and after the border change. To this end, I use census data on the land area of settlements whose population was above 2,000 inhabitants in 1910, together with the geolocation of these settlements. I compute the model-implied population of these settlements in the same way as I compute the model-implied population of cities (Section

[31]Section 5.3 provides details on how I compute the population of these settlements in the model.

[32]The p-value of the model estimate is 0.133. The estimated effect turns significant with a point estimate of 0.080 (p-value 0.098) if I replace log distance by distance on the right-hand side of the regression. In that case, the corresponding point estimate in the data is 0.107 (p-value 0.085).

4.5). That is, I assume that each settlement had a circular shape around its geographic center, such that the area of the circle equals the settlement's land area reported in the census. Next, I aggregate the total population of locations that are within the boundaries of the settlement, in both 1910 and 1930. Finally, I calculate the correlation between these model-implied populations and the population of the same settlements in the 1910 and 1930 census data (Section 2.2).

The results are reported in the bottom panel of Table 5. In general, the model seems successful at predicting settlement populations in both periods, but especially after the border change. Of course, part of this success might be due to the fact that I match the population of cities exactly in the estimation (Section 4.5). However, the last two rows of the table show that the correlations are between 0.4 and 0.5 even if cities (that is, settlements above 20,000 inhabitants) are excluded from the calculations. These findings suggest that the model is able to capture a quantitatively relevant part of the spatial distribution of Hungary's population, particularly after the change in borders.

## 5.4 How important were the different territories lost?

The previous sections used the model to show that the 1920 border change led to substantial losses in Hungarian residents' real income. This may not be surprising, given that the country lost more than two thirds of its land area and almost two thirds of its population in 1920. But among the territories that were lost in 1920, which were the most or least influential? Since the model can be simulated under alternative borders, it can be used to address this question as well. In this section, I use the model to disentangle the overall real income loss from the border change into the individual losses induced by territories joining different countries after 1920.

Figure 12 shows how the territory detached of Hungary was divided across neighboring countries in the Treaty of Trianon. The Southeast (shown in blue in the map) joined Romania; the North (orange) became part of Czechoslovakia; the West (red) joined Austria; and the Southwest and the South (black) became part of Yugoslavia.[33] To disentangle Hungary's overall real income loss into losses induced by these four territories, I simulate the model in four counterfactual scenarios. In each of the four counterfactuals, Hungary loses *only one* of the four territories. Then I calculate the model-impied loss in average real income in each of the four counterfactuals relative to the 1910 equilibrium.[34]

[33] A few villages in the North joined Poland, while the city of Fiume (now Rijeka, Croatia) was part of Italy until the end of the Second World War. For simplicity, I abstract from these small changes and assume that these two territories and their populations became part of Czechoslovakia and Yugoslavia, respectively.

[34] I keep city amenities fixed at their 1910 levels in the counterfactuals. To compute total population $\bar{L}$ in each of the four counterfactuals, I allocate Hungary's total population decline between 1910 and 1930 among the four territories according to their population shares in the year of the border change. This procedure thus guarantees that the sum of the four population losses equals the actual population decline between 1910 and 1930.

In Section 4.2, I discussed that my estimation strategy is robust to the potential endogeneity of the formation of trading places. This is because I can use the actual set of trading places in both 1910 and 1930, without the need to take a stand on what leads to their emergence. Obviously, this is no longer true in the counterfactual scenarios studied in this section. As an example, consider the scenario in which only the territory attached to Romania is lost. An alternative set of trading places could emerge in the rest of Hungary in response to this border change, and there is nothing in the data that informs us about this counterfactual set of trading places. At the same time, one *needs* to assume a set of trading places so that the model can be simulated in this scenario. In the rest of the section, I proceed with the most conservative assumption in the four counterfactual scenarios. Specifically, I assume that no new trading place forms and no trading place disappears over the territory that remains part of Hungary. In other words, I assume that the set of trading places remains the same as in 1910, with the exception of trading places that were part of the lost territory.

Table 6 presents the results of the four counterfactuals. Column (1) reports the fraction of Hungary's 1910 land area lost in each of the four scenarios, while column (2) reports the corresponding losses in average real income per capita in the remaining territory. In terms of land area, the territory lost to Romania was the largest. Therefore, it is no surprise that it also accounted for the largest real income loss according to the model. On the other end of the spectrum, the territory lost to Austria was small both in its land area and its effect on real income.

In terms of both their land area and their effects on real income, Czechoslovakia and Yugoslavia were in between Romania and Austria. In fact, these two territories were almost identical in their land area. Interestingly however, the territory lost to Yugoslavia was responsible for a larger loss in real income. A reason for this could be that the territory joining Yugoslavia was more urbanized. Whereas the part of Hungary lost to Czechoslovakia only had three cities in 1910, with a total city population of 144,771, the part lost to Yugoslavia had 11 cities with a total city population of 289,438. Moreover, many of these 11 cities were located close to the new border, including the largest one (Szabadka, with a population of 94,610). Apparently, losing access to these cities had a substantial impact on Hungarian residents' real income according to the model. This result underscores the rich interactions between urbanization and the real income gains from trade. While trade induces urbanization, highly urbanized areas foster trade and therefore induce substantial economic gains, especially in their surroundings.

Column (2) of Table 6 also shows that the sum of the real income losses from losing only one territory ($4.61\% + 4.01\% + 3.54\% + 0.28\% = 12.44\%$) is below the loss from losing all four at the same time (15.55% in Section 5.2).[35] Although losing trading opportunities

[35]Importantly, this difference is not due to city amenities changing or four new trading places forming between 1910 and 1930, both of which I took into account in Section 5.2 but not in the counterfactuals

with certain territories is hurtful, especially for nearby regions, new trade links may form with territories to which access is not lost. The resulting spatial reorganization of economic activity has the potential to mitigate the losses from declining trading opportunities.

# 6 Robustness

In this section, I provide robustness to my headline quantitative finding, that is, to the fact that the redrawing of Hungary's borders led to a 15.55% decline in average real income. I consider alternative values for the model's structural parameters and an alternative moment condition for the structural estimation in Section 6.1. In Section 6.2, I develop an extension of the model of Section 3.1 that features multiple sectors and take it to the data.

## 6.1 Alternative structural parameters and estimation strategies

This section studies how the values of calibrated model parameters ($\phi$, $\psi$ and $\sigma$) and the moment condition used to estimate parameter $\theta$ (equation 17) influence the estimated average real income loss from the new border. In the case of the calibrated parameters, I change the value of each by either +10% or −10%, keeping the remaining parameters unchanged, and redo the structural estimation of the model. In the case of moment condition (17), a concern is that this condition takes into account distance from the new border but not how much a location's distance from Hungary's border *changed* in 1920. For instance, certain locations in the West got close to the new border but they were near the country border already before 1920 (Figure 1). To address this concern, I estimate the model with an alternative moment condition that involves the change in distance from the border:

$$corr\left[\Delta a_c, dist\left(\mu_c, \text{old border}\right)/dist\left(\mu_c, \text{new border}\right)\right] = 0 \qquad (19)$$

where $dist\left(\mu_c, \text{old border}\right)$ and $dist\left(\mu_c, \text{new border}\right)$ denote city $c$'s distance from the old and the new border, respectively. As the country's territory shrinked in 1920, we have $dist\left(\mu_c, \text{old border}\right)/dist\left(\mu_c, \text{new border}\right) > 1$ for all $c$, and the closer the value of this distance ratio is to one, the less city $c$'s distance from the border changed.

Table 7 presents the point estimate of $\theta$ and the average real income change for each of the seven robustness exercises. As expected, higher values of shipping cost parameters $\phi$ and $\psi$ imply smaller losses from the new border. Under higher shipping costs, locations trade less with each other, implying that they lose less from a border that cuts their trading opportunities with other locations. Similarly, a higher value of the elasticity of substitution across tradables ($\sigma$) decreases the losses from the new border. If goods are less

of this section. If I simulate a fifth counterfactual with the actual 1930 borders of Hungary but keep city amenities and the set of trading places fixed at their 1910 levels, the average loss in real income remains virtually unchanged (15.58% instead of 15.55%).

differentiated (that is, if $\sigma$ is higher), then locations rely less on trade with other locations, implying again that the border hurts the economy less. Unsurprisingly, decreasing $\phi$, $\psi$ or $\sigma$ moves the losses from the new border in the opposite direction. Overall, changing these structural parameters has little effect on the estimated real income losses, which vary between $-13.82\%$ and $-17.73\%$ in the robustness exercises.

Using moment condition (19) in the estimation changes the estimated value of $\theta$ quite a bit but leaves the average real income loss almost unchanged (15.64% instead of 15.55%). The reason for this small difference might be that changes in distance from the border were highly correlated with distance from the new border: the correlation coefficient between these two variables for the 35 cities that enter moment conditions (17) and (19) was $-0.51$. Central locations that remained far from the new border were typically the farthest from the old border as well, implying that they experienced the smallest change in their distance from the border. In other words, the identifying variation behind moment condition (19) is similar to the one behind moment condition (17). This might explain the small effect that switching to (19) has on estimated real income losses.

## 6.2 Multiple sectors

In this section, I extend the model of Section 3.1 by introducing multiple sectors and take the extended model to the data. I present the setup in Section 6.2.1, describe how I take the model to sector-level data in Section 6.2.2, and present the results in Section 6.2.3.

### 6.2.1 Setup

The economy has a finite number of $N$ sectors, indexed by $n$ or $\ell$. Every worker chooses in which sector she wants to produce. Within each sector, every worker produces a specific differentiated good and exchanges it for other goods at a trading place, just like in the model of Section 3.1. Goods are all tradable, but their iceberg shipping costs across trading places can vary by sector. Workers consume all goods produced in the economy, aggregating goods' sector-level CES aggregates in a Cobb–Douglas utility function.

Workers have heterogeneous tastes for the location where they live and the sector in which they produce. Worker $i$, if chooses to live at location $r$, produce in sector $n$ and trade at trading place $m$, obtains utility

$$u_m(r,n,i) = a(r,n,i) + \varsigma(\mu_m, r)^{-1}\bar{\alpha}\prod_{\ell}\left[\sum_{j\in\ell} c_m^j(r,i)^{\frac{\sigma-1}{\sigma}}\right]^{\alpha_\ell\frac{\sigma}{\sigma-1}} \tag{20}$$

where $c_m^j(r,i)$ denotes the worker's consumption of the product of worker $j$, the summation across $j \in \ell$ means summing across all workers $j$ who produce in a given sector $\ell$, $\sigma > 1$ is the elasticity of substitution across products within a sector, and $\alpha_\ell$ denotes workers'

Cobb–Douglas spending share on sector $\ell$.[36]

As in the model of Section 3.1, amenities consist of a part common across workers as well as an idiosyncratic shifter:

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

where the only novel part is that both $a(r,n)$ and $\varepsilon(r,n,i)$ depend on the worker's sector $n$. The idiosyncratic term is iid across workers, locations and sectors, and is distributed Gumbel:

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

implying that parameter $\zeta$ reflects the heterogeneity in workers' tastes for location-sector pairs.

The rest of the model is unchanged relative to the model of Section 3.1. In Appendix D.1, I define the equilibrium of the multi-sector model and derive its equilibrium conditions.

### 6.2.2 Structural estimation

Based on the available data (see Section 2.2 for details), I define three sectors of the economy: agriculture (A), manufacturing (M) and an "other sector" comprising the rest of the economy (O). As explained in Section 2.2, the censuses provide me with employment in these three sectors in every settlement of the country, both in 1910 and in 1930.

I follow the structural estimation of the one-sector model as closely as possible when taking the multi-sector model to these data. I keep the definition of locations and trading places fixed relative to Section 4. As for shipping costs, recall that the multi-sector model allows them to vary by sector. Unfortunately however, there is no data available on sectoral shipping costs. Therefore, I choose a common shipping cost elasticity $\phi$ across sectors as my baseline, but also estimate the multi-sector model with different shipping cost elasticities for robustness (see Section 6.2.3). I calibrate the other shipping cost elasticity, $\psi$, analogously to Section 4. I choose the same elasticity of substitution across goods as in Section 4 ($\sigma = 5$).

I match sectoral spending shares $\alpha_A$, $\alpha_M$ and $\alpha_O$ to the shares of these sectors in Hungary's GDP in 1910, estimated by economic historians (Probáld, 2009). Next, I assume the following amenity function in sector $n$: $a(r,n) = a_c(n)$ if location $r$ is within the geographic boundaries of city $c$, and $a(r,n) = a(n)$ otherwise. I normalize agricultural amenities outside cities to zero: $a(A) = 0$, and back out manufacturing and other-sector amenities outside cities, $a(M)$ and $a(O)$, to match the shares of these sectors in aggregate employment, separately for 1910 and 1930. Finally, I back out amenities by city-sector $a_c(n)$ to match employment by city-sector, again separately for 1910 and 1930.

[36]Naturally, $\sum_\ell \alpha_\ell = 1$. $\bar{\alpha} = \prod_\ell \alpha_\ell^{-\alpha_\ell}$ is a constant that simplifies the subsequent formulas algebraically.

This leaves me with estimating $\zeta$, the parameter driving the heterogeneity of preferences for location-sector pairs. I estimate this parameter using a moment condition akin to (17):

$$corr\left[\frac{\sum_n L_{c,1910}(n)\,\Delta a_c(n)}{\sum_n L_{c,1910}(n)}, dist\left(\mu_c, border\right)\right] = 0 \qquad (21)$$

where $L_{c,1910}(n)$ denotes the 1910 employment of city $c$ in sector $n$, and $\Delta a_c(n)$ is the change in city $c$'s amenities in sector $n$ between 1910 and 1930, coming from the procedure described above. This moment condition thus states that the placement of the new border, by being exogenous, was uncorrelated with changes in average amenities in cities, weighted by sectoral shares prior to the border change. As in Section 4, I simultaneously search for the value of $\zeta$ satisfying (21) and the set of city-sector amenities that rationalize employment by city-sector in 1910 and 1930. Appendix D.2 provides further details on the estimation.

### 6.2.3 Results

Table 8 presents the results of the estimation of the multi-sector model. In row 1, I assume equal shipping cost parameters across sectors, i.e., $\phi_A = \phi_M = \phi_O$. The point estimate of $\zeta$ is 31.25, while the average real income loss from the border change equals 16.04%. This is reassuringly close to the corresponding estimate in the one-sector model (15.55%).

Rows 2 to 4 of Table 8 estimate the model with alternative shipping cost parameters. Recall that, according to the reduced-form estimates of Section 2.5 (columns 1 to 3 in Table 3), the employment share of agriculture decreased in border counties after 1920. The manufacturing share did not change significantly, while the share of the other sector increased in border counties. These results are consistent with a world in which shipping costs were the lowest in agriculture, followed by manufacturing and then by the other sector. Therefore, row 2 of Table 8 estimates the model under the assumptions that shipping cost parameter $\phi$ was the 50% higher in manufacturing and 100% higher in the other sector than in agriculture.[37] Row 3 estimates the model under the assumptions that the manufacturing parameter equaled the agricultural one but the parameter of the other sector was 100% higher. Finally, row 4 estimates the model under even more extreme cross-sector differences in shipping costs ($\phi_M$ 100% higher, $\phi_O$ 200% higher than $\phi_A$).

The losses from the border are somewhat lower under higher shipping costs, in line with the intuition of Section 6.1. However, the estimates of average real income losses vary in a tight range around the estimate coming from my baseline one-sector framework (between 14.97% and 16.04%). Taken together, the results of this section suggest that my headline quantitative finding is robust to incorporating multiple sectors in the model.

[37] I keep the shipping cost elasticity of agriculture, $\phi_A$, at the level of parameter $\phi$ calibrated in Section 4. This is consistent with the fact that I used agricultural price data to calibrate this parameter.

# 7 Conclusion

Urbanization is an ongoing process: the share of world population living in urban areas 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. 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. I find that urbanization decreased in counties near the new border relative to counties farther away. I rationalize these findings in a quantitative spatial model of trade that, despite its flexible geographic structure, provides simple predictions on the effects of trade on urbanization and real income. The estimated model fits the data well. 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 (2020), 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.

- Bazzi, S. (2017): Wealth heterogeneity and the income elasticity of migration. *American Economic Journal: Applied Economics*, vol. 9(2), 219–255.

- Bhagat, R. B. (2011): Emerging pattern of urbanisation in India. *Economic and Political Weekly*, vol. 46(34), 10–12.

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

- Brooks, L., Gendron-Carrier, N. and Rua, G. (2019): The local impact of containerization. Mimeo.

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

- Brülhart, M., Carrère, C. and Trionfetti, F. (2012): How wages and employment adjust to trade liberalization: Quasi-experimental evidence from Austria. *Journal of International Economics*, vol. 86(1), 68–81.

- Brülhart, M., Carrère, C. and Robert-Nicoud, F. (2018): Trade and towns: Heterogeneous adjustment to a border shock. *Journal of Urban Economics*, vol. 105, 162–175.

- Bryan, G., Chowdhury, S. and Mobarak, A. (2014): Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. *Econometrica*, vol. 82(5), 1671–1748.

- Campante, F. and Yanagizawa-Drott, D. (2018): Long-range growth: Economic development in the global network of air links. *Quarterly Journal of Economics*, vol. 133(3), 1395–1458.

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

- Csató, T. (2000): Külkereskedelem. In: *Magyarország a XX. században*, vol. 2, ed. Tarsoly, I. Babits.

- 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*, vol. 126(3), 903–983.

- Donaldson, D. and Hornbeck, R. (2016): Railroads and American economic growth: A "market access" approach. *Quarterly Journal of Economics*, vol. 131(2), 799–858.

- Ducruet, C., Juhász, R., Nagy, D. and Steinwender, C. (2020): All aboard: The effects of port development. Mimeo.

- Eckstein, A. (1955): National income and capital formation in Hungary, 1900–1950. *Income and wealth*, series V, ed. Kuznets, S. International Assocation for Research in Income and Wealth.

- Ellingsen, S. (2020): Free and protected: Trade and breaks in long-term persistence. Mimeo.
- Fajgelbaum, P. and Redding, S. (2018): Trade, structural transformation and development: Evidence from Argentina 1869-1914. *NBER Working Paper* 20217.
- Feyrer, J. (2009): Distance, trade, and income – The 1967 to 1975 closing of the Suez canal as a natural experiment. *NBER Working Paper* 15557.
- Feyrer, J. (2019): Trade and income – Exploiting time series in geography. *American Economic Journal: Applied Economics*, vol. 11(4), 1–35.
- 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.
- Jagadits, A. (2020): The effect of emigration on local structural change: Evidence from (Austria-)Hungary in the age of mass migration. Mimeo.
- 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*, vol. 108(12), 3855–3890.
- Morten, M. and Oliveira, J. (2018): The effects of roads on trade and migration: Evidence from a planned capital city. *NBER Working Paper* 22158.
- Nagy, D. (2020): Hinterlands, city formation and growth: Evidence from the U.S. westward expansion. Mimeo.
- Pascali, L. (2017): The wind of change: Maritime technology, trade and economic development. *American Economic Review*, vol. 107(9), 2821–2854.
- Probáld, F. (2009): General characteristics of economic development. In: *Hungary in maps*, ed. Kocsis, K. and Schweitzer, F. Geographical Institute, Hungarian Academy of Sciences.

- 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.
- 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.
- Tombe, T. and Zhu, X. (2019): Trade, migration, and productivity: A quantitative analysis of China. *American Economic Review*, vol. 109(5), 1843–1872.
- United Nations, Department of Economics and Social Affairs, Population Division (2014): World urbanization prospects: The 2014 revision. Highlights.
- Zeidler, M. (2014): Treaty of Trianon. In: *1914-1918 online. International Encyclopedia of the First World War*. Freie Universität Berlin.

Table 1: Urbanization decreased near the new border

| | (1) Change in urbanization 1910 to 1930 | (2) Log population density 1910 | (3) Share of agriculture 1910 | (4) Share of manufacturing 1910 | (5) Share of other sector 1910 | (6) Change in urbanization 1890 to 1910 |
|---|---|---|---|---|---|---|
| $\log(dist)$ | 0.751** | −0.006 | −0.480 | 0.278 | 0.201 | 0.228 |
| | (0.359) | (0.028) | (1.47) | (0.658) | (0.896) | (0.568) |
| $R^2$ | 0.125 | 0.002 | 0.004 | 0.005 | 0.002 | 0.005 |
| $N$ | 22 | 22 | 22 | 22 | 22 | 22 |

Column (1) presents the results of estimating equation (1), while columns (2) to (6) present the results of regressing other variables on $\log(dist)$. The unit of observation is a county in post-1920 Hungary, and $\log(dist)$ is the log of distance (in kilometers) between the county seat and the post-1920 border of Hungary. All dependent variables except *log population density 1910* are measured in percentage points. *Change in urbanization 1910 to 1930* is the change in the share of people living in cities above 20,000 inhabitants between 1910 and 1930. *Log population density 1910* is the log of population density (individuals per square kilometer) in 1910. *Share of agriculture*, *manufacturing* and *other sector, 1910* refer to the employment shares of agriculture, manufacturing and the rest of the economy, respectively. *Change in urbanization 1890 to 1910* denotes the change in the share of people living in cities above 20,000 inhabitants between 1890 and 1910. In all the columns, I restrict the sample to counties whose distance from the post-1920 border is less than 60 kilometers. Heteroskedasticity-robust standard errors in parentheses. *: significant at 10%; **: significant at 5%. Source: Censuses of the Hungarian Kingdom, 1890, 1910 and 1930.

Table 2: The effect of the new border on city and rural populations

| | (1) Change in log city population 1910 to 1930 | (2) Change in log rural population 1910 to 1930 | (3) Change in log city population 1890 to 1910 | (4) Change in log rural population 1890 to 1910 |
|---|---|---|---|---|
| $\log(dist)$ | 0.120** | −0.012 | 0.040 | −0.005 |
| | (0.042) | (0.007) | (0.062) | (0.014) |
| $R^2$ | 0.299 | 0.095 | 0.042 | 0.006 |
| $N$ | 15 | 22 | 13 | 22 |

Column (1) presents the results of estimating equation (2), while column (2) presents the results of estimating equation (3). The unit of observation is a county in post-1920 Hungary, and $\log(dist)$ is the log of distance (in kilometers) between the county seat and the post-1920 border of Hungary. *Change in log city population 1910 to 1930* is the log change in the number of people living in cities above 20,000 inhabitants between 1910 and 1930. *Change in log rural population 1910 to 1930* is the log change in the number of people living outside cities above 20,000 inhabitants between 1910 and 1930. *Change in log city population 1890 to 1910* and *Change in log rural population 1890 to 1910* denote changes in the same variables between 1890 and 1910. In all the columns, I restrict the sample to counties whose distance from the post-1920 border is less than 60 kilometers. Heteroskedasticity-robust standard errors in parentheses. *: significant at 10%; **: significant at 5%. Source: Censuses of the Hungarian Kingdom, 1890, 1910 and 1930.

Table 3: Urbanization and city populations decreased near the new border: Robustness

| | (1) Change in share of agriculture 1910 to 1930 | (2) Change in share of manufacturing 1910 to 1930 | (3) Change in share of other sector 1910 to 1930 | (4) Change in urbanization 1910 to 1930 | (5) Change in log city population 1910 to 1930 |
|---|---|---|---|---|---|
| log(*dist*) | 1.78** | −0.547 | −1.23** | 1.05** | 0.119** |
| | (0.798) | (0.339) | (0.502) | (0.468) | (0.052) |
| Change in share of agriculture '10-'30 | | | | 0.160 | 0.266 |
| | | | | (0.219) | (1.39) |
| Change in share of other sector '10-'30 | | | | 0.472 | 0.581 |
| | | | | (0.383) | (2.70) |
| Maximum distance | 60 km | 60 km | 60 km | 60 km | 60 km |
| $R^2$ | 0.138 | 0.091 | 0.142 | 0.253 | 0.301 |
| $N$ | 22 | 22 | 22 | 22 | 15 |

| | (6) Change in urbanization 1910 to 1930 | (7) Change in urbanization 1910 to 1930 | (8) Change in urbanization 1910 to 1930 | (9) Change in log city population 1910 to 1930 | (10) Change in log city population 1910 to 1930 | (11) Change in log city population 1910 to 1930 |
|---|---|---|---|---|---|---|
| *dist* | 0.061* | | | 0.006* | | |
| | (0.034) | | | (0.003) | | |
| log(*dist*) | | 0.555** | 0.645** | | 0.058* | 0.123** |
| | | (0.257) | (0.286) | | (0.032) | (0.048) |
| Maximum distance | 60 km | — | 40 km | 60 km | — | 40 km |
| $R^2$ | 0.138 | 0.080 | 0.245 | | 0.092 | 0.443 |
| $N$ | 22 | 25 | 17 | 15 | 18 | 11 |

Each column in this table corresponds to a regression. In all the regressions, the unit of observation is a county in post-1920 Hungary. *dist* is the distance (in kilometers) between the county seat and the post-1920 border of Hungary, while log(*dist*) is the log of the same variable. The dependent variables in columns (1) to (4) and columns (6) to (8) are measured in percentage points, while the dependent variables in column (5) and columns (9) to (11) are measured in logs. *Change in share of agriculture*, *manufacturing* and *other sector, 1910 to 1930* (or *'10-'30*) refer to the 1910 to 1930 change in the employment shares of agriculture, manufacturing and the rest of the economy, respectively. *Change in urbanization 1910 to 1930* is the change in the share of people living in cities above 20,000 inhabitants between 1910 and 1930. *Change in log city population 1910 to 1930* is the log change in the number of people living in cities above 20,000 inhabitants. Columns (1) to (6) and column (9) restrict the sample to counties whose distance from the post-1920 border is less than 60 kilometers. Columns (7) and (10) consider the entire set of counties, while columns (8) and (11) restrict the sample to counties less than 40 kilometers from the post-1920 border. Heteroskedasticity-robust standard errors in parentheses. *: significant at 10%; **: significant at 5%. Source: Censuses of the Hungarian Kingdom, 1910 and 1930.

Table 4: Model-implied regressions

| | (1) Fraction of trade lost | (2) Change in population of cities relative to settlements above 2,000 inhabitants | (3) |
|---|---|---|---|
| | Model | Model | Data |
| log(*dist*) | −14.30** | 0.833 | 1.259* |
| | (6.02) | (0.532) | (0.632) |
| $R^2$ | 0.172 | 0.077 | 0.112 |
| $N$ | 23 | 22 | 22 |

Column (1) presents the results of regressing the fraction of trade lost with the border change on log distance from the post-1920 border of Hungary. Column (2) presents the results of regressing the change in the population of cities (settlements above 20,000 inhabitants) relative to the population of settlements above 2,000 inhabitants between 1910 and 1930 on log distance from the border in the model, while column (3) presents the corresponding results in the data. In column (1), the unit of observation is a trading place that existed in 1910 and remained in Hungary after 1920. In columns (2) and (3), the unit of observation is a county in post-1920 Hungary. log(*dist*) is the log of distance (in kilometers) between the trading place (column 1) or the county seat (columns 2 and 3) and the post-1920 border. The dependent variable of column (1) is measured in percentages, while the dependent variable of columns (2) and (3) is measured in percentage points. *Fraction of trade lost* is measured at trading place $m$ as $100 \cdot \left(\text{Trade}_{m,1910} - \text{Trade}_{m,1930}\right)/\text{Trade}_{m,1910}$. I restrict the sample to trading places (column 1) or counties (columns 2 and 3) whose distance from the post-1920 border is less than 60 kilometers. Heteroskedasticity-robust standard errors in parentheses. *: significant at 10%; **: significant at 5%. Source: Censuses of the Hungarian Kingdom, 1910 and 1930.

Table 5: Estimated real income changes and model fit

Estimated real income changes between 1910 and 1930

| | |
|---|---|
| Average real income change | −15.55% |
| | (0.28%) |
| Standard deviation of real income changes | 1.08% |
| | (0.24%) |
| Change in standard deviation of log real income | 9.68% |
| | (27.05%) |

Model fit

| | |
|---|---|
| Correlation between model and data, settlements above 2,000, 1910 | 0.603 |
| | (0.083) |
| Correlation between model and data, settlements above 2,000, 1930 | 0.707 |
| | (0.100) |
| Correlation between model and data, settlements 2,000 to 20,000, 1910 | 0.409 |
| | (0.095) |
| Correlation between model and data, settlements 2,000 to 20,000, 1930 | 0.476 |
| | (0.119) |

The top panel presents the estimates of average real income change, standard deviation of real income changes and change in inequality (standard deviation of log real income) between 1910 and 1930. All three variables are calculated from location-level real income changes, with locations' 1930 population as weights. The bottom panel presents correlations between settlements' model-implied and actual populations in 1910 and 1930 to measure model fit. Bootstrap standard errors in parentheses.

Table 6: The effects of territories lost to other countries

| Territory lost to... | (1)<br>Change in land area (%) | (2)<br>Change in average real income (%) |
|---|---|---|
| Romania | −31.7 | −4.61 |
| Yugoslavia | −19.4 | −4.01 |
| Czechoslovakia | −19.1 | −3.54 |
| Austria | −1.2 | −0.28 |
| Total change | −71.4 | −15.55 |

This table presents the results of the counterfactuals of Section 5.4. Column (1) shows the change in Hungary's land area if one of the four territories (the territory joining Romania, Yugoslavia, Czechoslovakia or Austria) is lost. Column (2) shows the corresponding change in average real income. The last row presents the change in land area and real income if all four territories are lost at the same time, thus repeating the results of the baseline estimation. The color coding of the four territories coincides with the color coding of Figure 12.

Table 7: Average real income change between 1910 and 1930: Robustness

| | $\theta$ | Average real income change |
|---|---|---|
| Baseline estimation | 11.49 | −15.55% |
| 1. 10% higher shipping cost parameter $\phi$ | 11.76 | −15.41% |
| 2. 10% lower shipping cost parameter $\phi$ | 11.36 | −15.70% |
| 3. 10% higher shipping cost parameter $\psi$ | 11.90 | −15.53% |
| 4. 10% lower shipping cost parameter $\psi$ | 11.11 | −15.58% |
| 5. 10% higher elasticity of substitution $\sigma$ | 7.25 | −13.82% |
| 6. 10% lower elasticity of substitution $\sigma$ | 21.28 | −17.73% |
| 7. $corr\left[\Delta a_c, dist\left(\mu_c, \text{old border}\right)/dist\left(\mu_c, \text{new border}\right)\right] = 0$ | 7.41 | −15.64% |

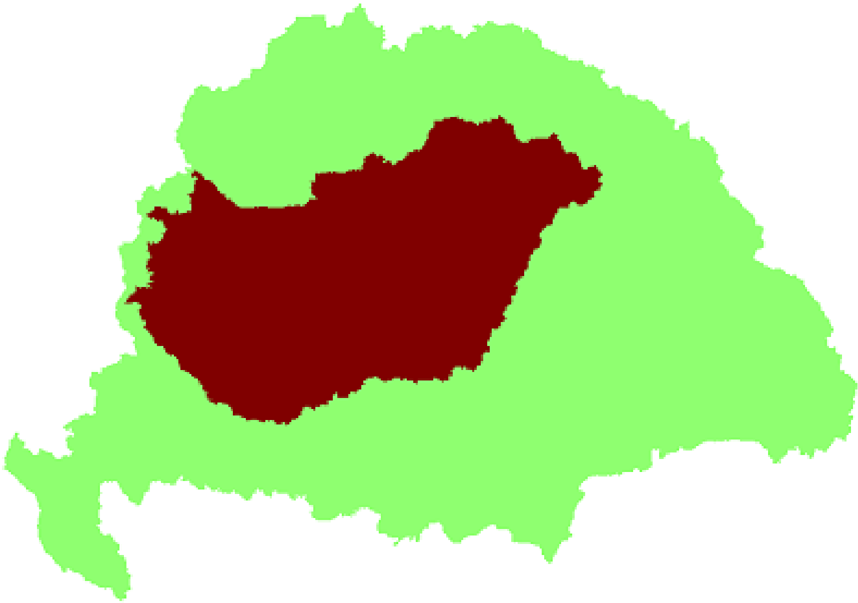
This table presents the results of the seven robustness exercises of Section 6.1. For each exercise, the table shows the point estimate of parameter $\theta$ and the average of locations' real income changes, weighted by locations' 1930 population. The first row repeats the corresponding results of the baseline estimation.

Table 8: Average real income change between 1910 and 1930: Multi-sector models

| | $\phi_M/\phi_A$ | $\phi_O/\phi_A$ | $\zeta$ | Average real income change |
|---|---|---|---|---|
| One-sector model | — | — | — | −15.55% |
| Multi-sector models | | | | |
| 1. Equal shipping costs across sectors | 1 | 1 | 31.25 | −16.04% |
| 2. More costly trade in M, even more in O | 1.5 | 2 | 30.30 | −15.49% |
| 3. More costly trade in O but not in M | 1 | 2 | 30.30 | −15.64% |
| 4. Even more costly trade in both M and O | 2 | 3 | 29.41 | −14.97% |

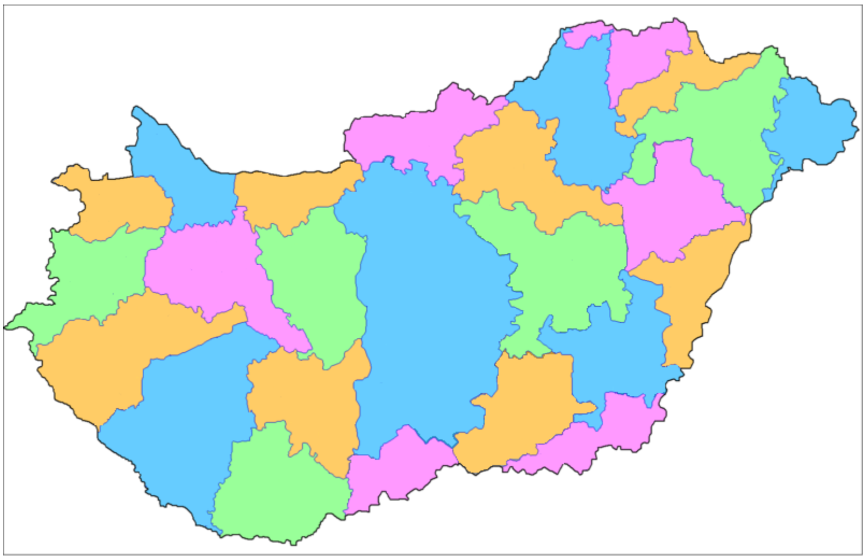
This table presents the results of the four robustness exercises conducted with the multi-sector model of Section 6.2. For each exercise, the table shows the ratio of sectoral shipping cost parameters fed into the estimation, the point estimate of parameter $\zeta$, and the average of location-sector pairs' real income changes, weighted by sectors' 1930 employment shares and locations' 1930 population. The first row repeats the corresponding average real income change in the one-sector model.

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



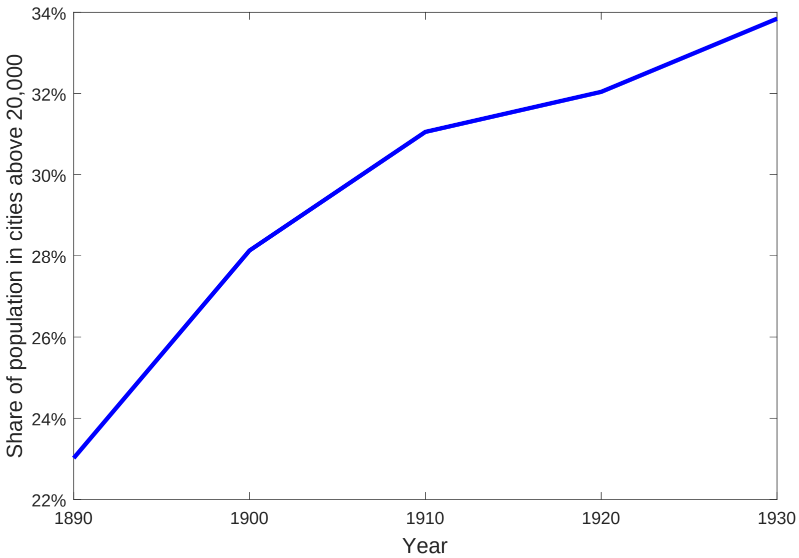
This map presents the change in Hungary's borders due to the Treaty of Trianon in 1920. The green and brown areas together constitute the country's territory before the treaty. The brown area constitutes the country's territory after the treaty.

Figure 2: Counties of Hungary, 1930



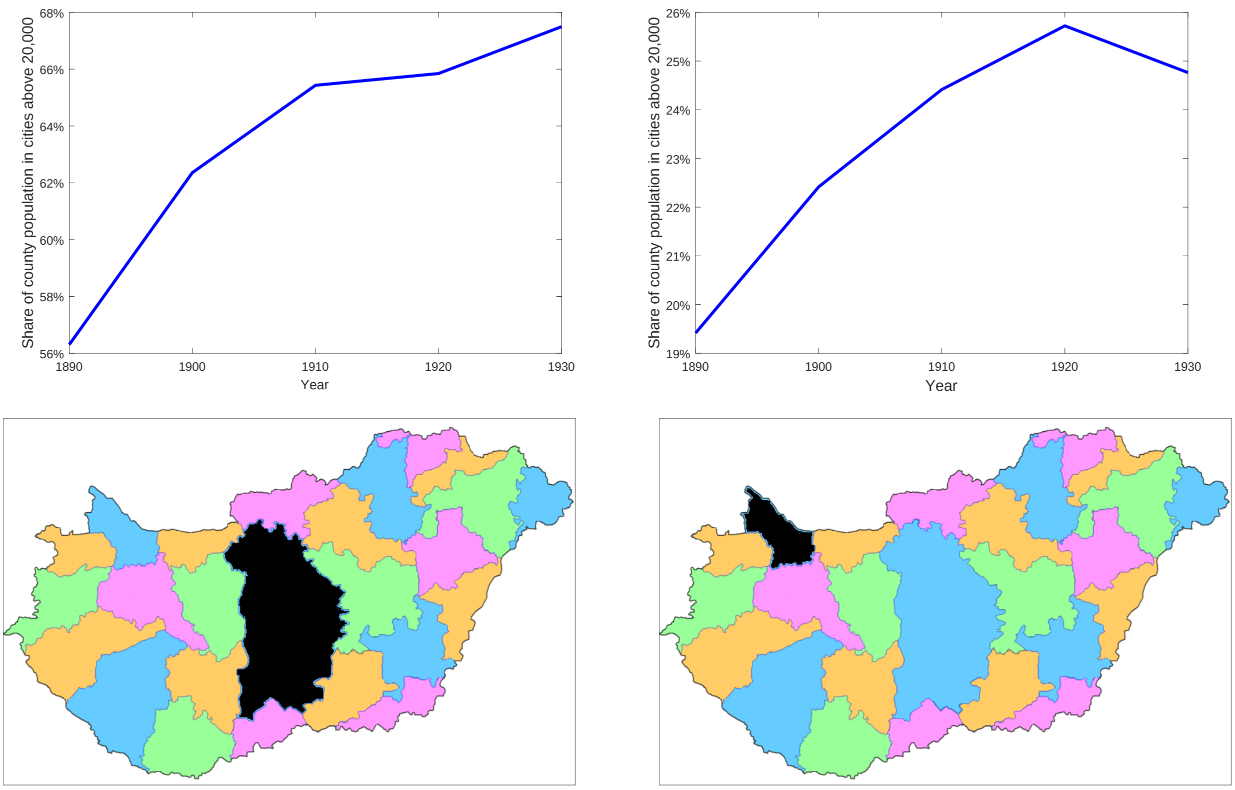
This map shows the 25 counties of Hungary in 1930.

## Figure 3: Urbanization in Hungary



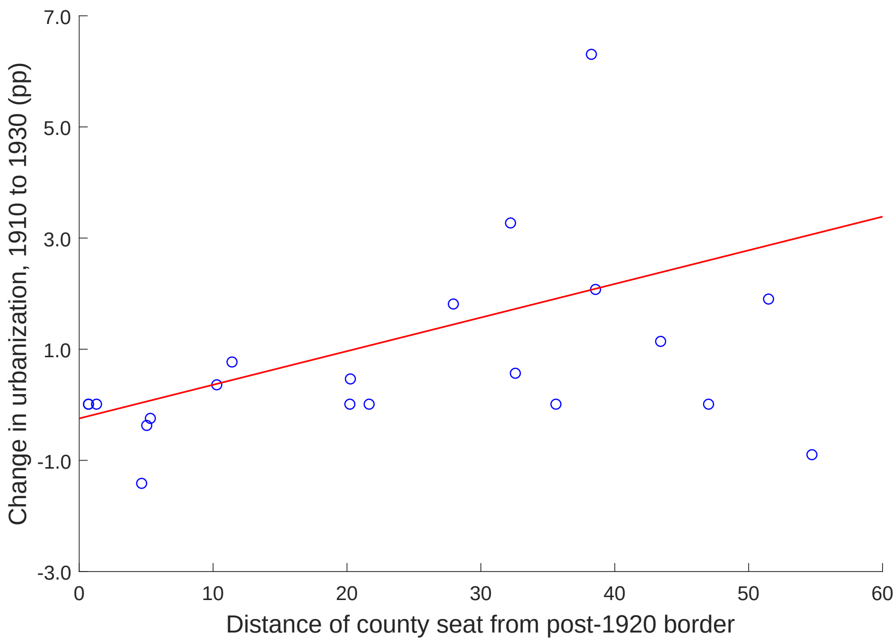

This graph shows the evolution of urbanization (the share of people living in cities above 20,000 inhabitants) over the post-1920 territory of Hungary between 1890 and 1930. Source: Censuses of the Hungarian Kingdom, 1890, 1900, 1910, 1920 and 1930.

## Figure 4: Urbanization in Pest-Pilis-Solt-Kiskun county (left) vs Győr county (right)



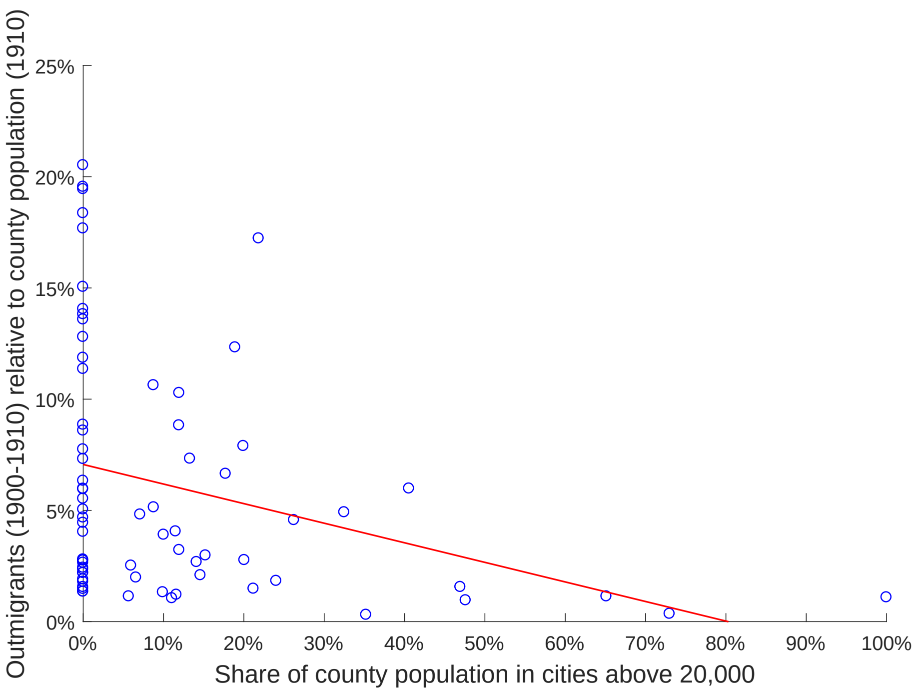

These graphs show the evolution of urbanization (the share of people living in cities above 20,000 inhabitants) in Pest-Pilis-Solt-Kiskun county (top left graph) and Győr county (top right graph) between 1890 and 1930. The bottom left and bottom right maps show the locations of Pest-Pilis-Solt-Kiskun county and Győr county, respectively. Source: Censuses of the Hungarian Kingdom, 1890, 1900, 1910, 1920 and 1930.

## Figure 5: Urbanization decreased near the new border



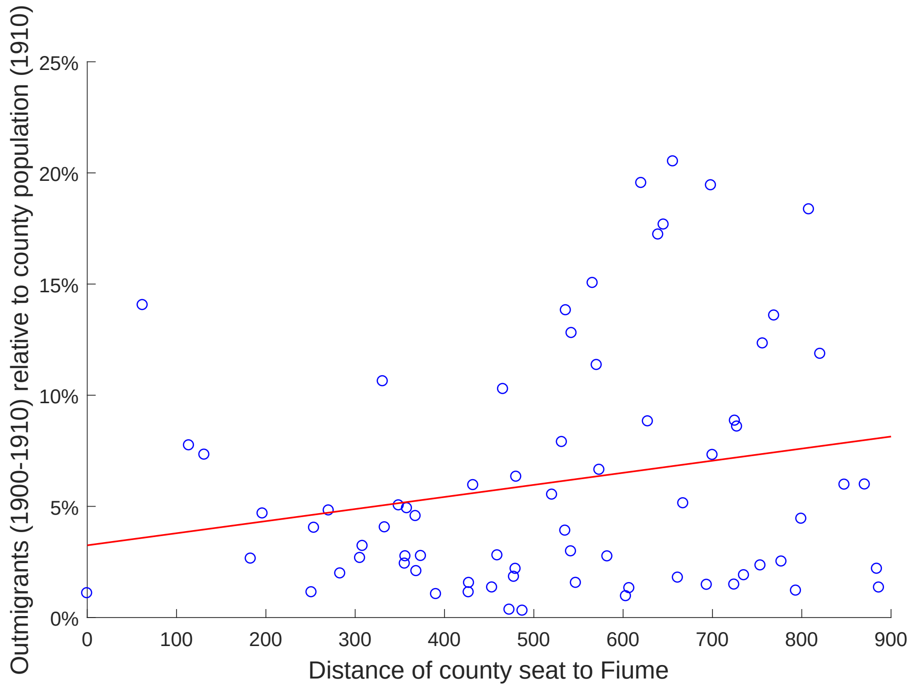

This figure presents a scatterplot of changes in urbanization against distance from the new border. Each blue circle corresponds to a county in post-1920 Hungary. The horizontal axis represents distance (in kilometers) between the county seat and the post-1920 border of Hungary. The vertical axis represents the 1910 to 1930 change (in percentage points) in the share of people who live in cities (settlements above 20,000 inhabitants). The regression line between these two variables is shown in red. See column (6) in Table 3 for the estimated coefficient of this regression.

## Figure 6: Outmigration decreased in the level of urbanization



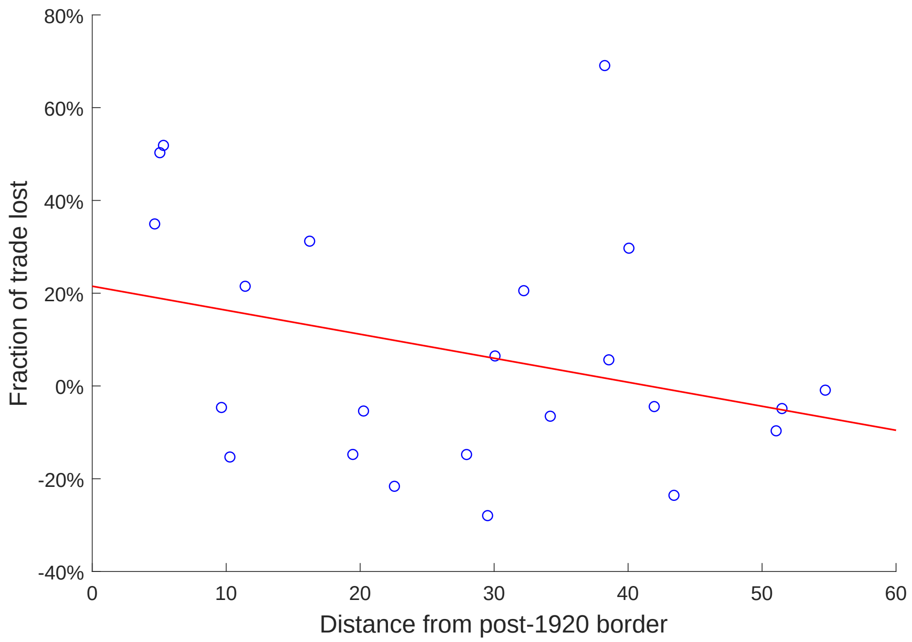

This figure presents a scatterplot of outmigration between 1900 and 1910 against urbanization in 1910. Each blue circle corresponds to a county in 1910 Hungary. The horizontal axis represents the share of people who live in cities (settlements above 20,000 inhabitants), in percentages. The vertical axis represents the number of people who moved abroad from the county between 1900 and 1910 relative to 1910 county population. The regression line between these two variables is shown in red.

Figure 7: Outmigration was not decreasing in distance to the main sea port



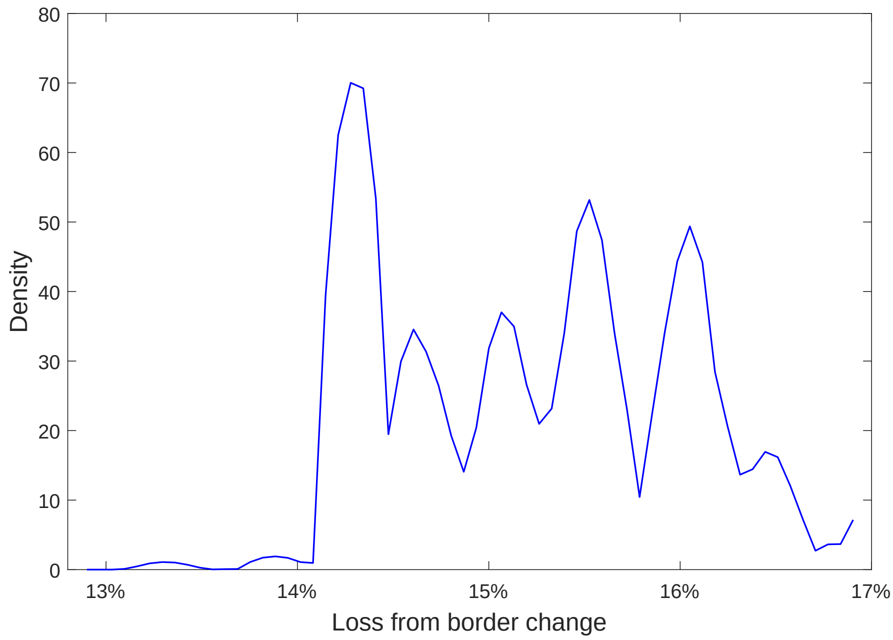

This figure presents a scatterplot of outmigration between 1900 and 1910 against distance to Fiume, the main sea port through which outmigration took place. Each blue circle corresponds to a county in 1910 Hungary. The horizontal axis represents distance (in kilometers) between the county seat and the port of Fiume (now Rijeka, Croatia). The vertical axis represents the number of people who moved abroad from the county between 1900 and 1910 relative to 1910 county population. The regression line between these two variables is shown in red.

Figure 8: Larger model-implied losses in trade near the new border



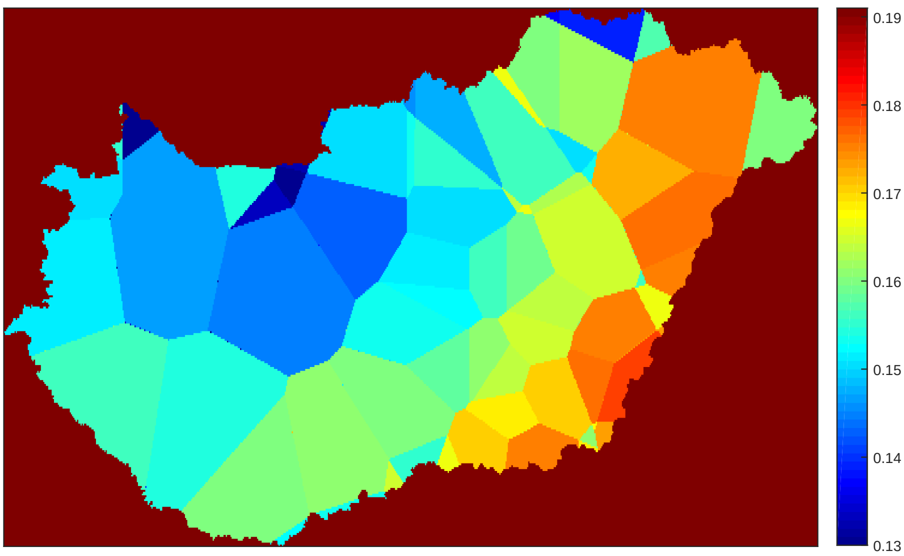

This figure presents a scatterplot of model-implied losses in trade against distance from the new border. Each blue circle corresponds to a trading place in 1910 Hungary that remained in the country after the border change. The horizontal axis represents distance (in kilometers) between the trading place and the post-1920 border of Hungary. The vertical axis represents the fraction of 1910 trade lost by 1930 in percentages, measured as $100 \cdot (\text{Trade}_{m,1910} - \text{Trade}_{m,1930}) / \text{Trade}_{m,1910}$ at trading place $m$. The regression line between these two variables is shown in red.

Figure 9: Empirical density of real income losses between 1910 and 1930



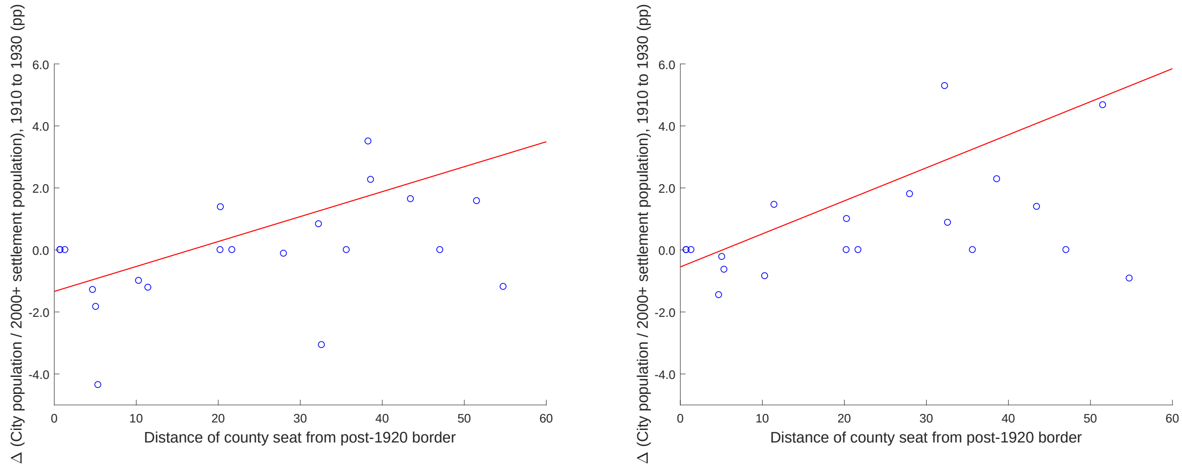

This graph presents the smoothed empirical density of locations' model-implied real income losses between 1910 and 1930. The densities are weighted by locations' 1930 populations. Smoothing is done with the Epanechnikov kernel. The values are trimmed at 12.9% and 16.9% for better visibility.

Figure 10: Map of real income losses between 1910 and 1930



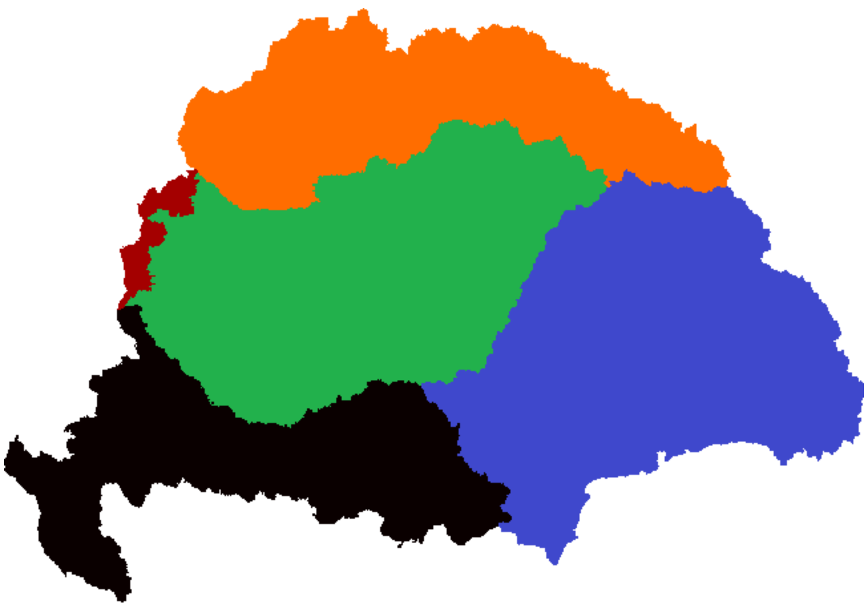

This map shows Hungarian locations' model-implied real income losses between 1910 and 1930.

Figure 11: Change in population of cities relative to population of settlements above 2,000 inhabitants: model (left) vs data (right)




These figures present scatterplots of changes in the population of cities relative to the population of settlements above 2,000 inhabitants against distance from the new border. The left figure shows this relationship in the model, while the right figure shows it in the data. Each blue circle corresponds to a county in post-1920 Hungary. The horizontal axis represents distance (in kilometers) between the county seat and the post-1920 border of Hungary. The vertical axis represents the 1910 to 1930 change in the population of cities (settlements above 20,000 inhabitants) relative to the population of settlements above 2,000 inhabitants. In each figure, the regression line between the two variables is shown in red. Source: Censuses of the Hungarian Kingdom, 1910 and 1930.

Figure 12: The division of Hungary's territory in the Treaty of Trianon, 1920



This map shows which regions of pre-1920 Hungary joined which country after the Treaty of Trianon. The entire colored area is Hungary before 1920. The territory lost to Romania is shown in **blue**; the territory lost to Czechoslovakia (as well as a few villages lost to Poland) is shown in **orange**; the territory lost to Austria is shown in **red**; and the territory lost to Yugoslavia (named *Kingdom of Serbs, Croats and Slovenes* until 1929, as well as Fiume lost to Italy) is shown in **black**. Post-1920 Hungary is shown in **green**.

# Appendix

## A The effects of borders on urbanization by 1920

In this appendix, I rerun the main empirical specifications of Section 2 on data from the 1920 census and show that the effects of the 1920 border change cannot be identified from these data. This finding suggests that the effects of the border change could not fully unfold over the few months between the Treaty of Trianon and the 1920 census.

Rerunning the empirical specifications of Section 2 amounts to re-estimating equations (1), (2) and (3) using 1920 instead of 1930 population data. Other than this change, I follow the same strategy to estimate these three equations as in Section 2.

Table 9 presents the results, with columns (1), (2) and (3) in the top panel corresponding to equations (1), (2) and (3), respectively. To ease the comparison, columns (5), (6) and (7) in the bottom panel repeat the corresponding results obtained from 1930 data (also available in Tables 1 and 2).

Table 9 shows that the positive effects of distance from the new border on urbanization and city populations, which are both significant and large by 1930, are not present in the 1920 data. In the case of city populations (column 2), the estimated coefficient is insignificant and the point estimate is only one tenth of the estimate in 1930 (column 6). The point estimate is similarly small for rural populations (column 3), although this effect is also insignificant and small by 1930 (column 7), as also discussed in Section 2.4. Finally, the estimated effect of distance on urbanization is insignificant but the point estimate is negative by 1920 (column 1), in sharp contrast to the significant positive effect by 1930 (column 5). Investigating this negative point estimate further, it turns out to be driven by one border county (Zemplén county). A settlement in this small county surpassed the 20,000-inhabitant threshold by 1920, becoming a city in the 1920 census, but fell below 20,000 inhabitants again by 1930. Plotting the change in urbanization by 1920 against distance from the new border in a scatterplot (Figure 13), this county shows up as an outlier in the top left corner. The remaining counties all exhibit nearly zero change in urbanization, with no clear relationship between their change in urbanization and their distance from the border. In line with this, I find that dropping the outlier county makes the effect on urbanization by 1920 slightly positive but, most importantly, small (less than one tenth of the effect by 1930) and statistically insignificant (column 4 of Table 9).[38]

Taking these results together, I conclude that the effect of the new border on urbanization and city populations is not yet detectable from the 1920 census. This is why I use the 1930 census both in my reduced-form empirical specifications and when taking the model to the data.

[38]Column (8) of Table 9 shows that the effect of the new border on urbanization remains statistically significant and largely unchanged in magnitude if I drop the same county in 1930.

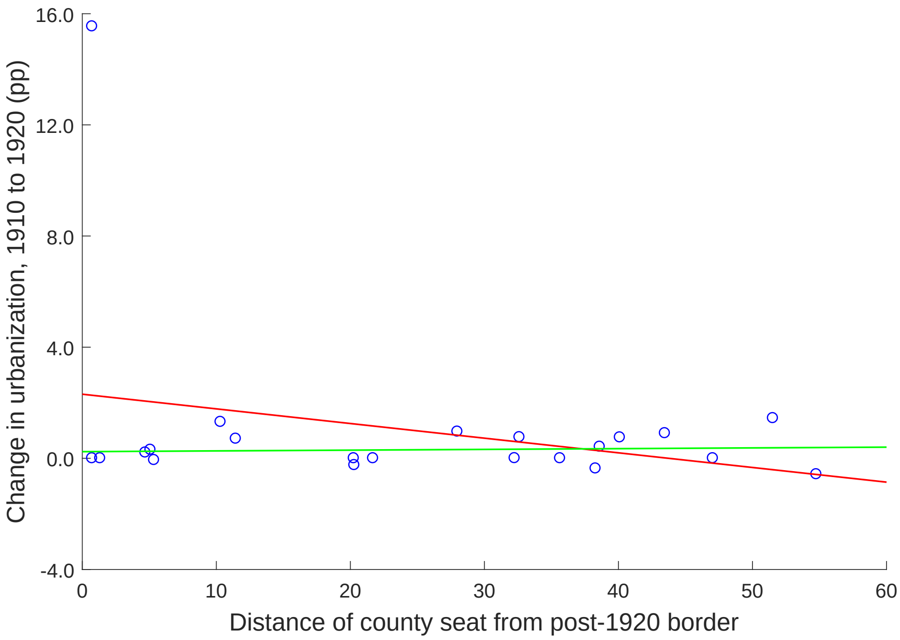
Table 9: The effect of the new border by 1920 (top panel) vs 1930 (bottom panel)

| | (1) Change in urbanization 1910 to 1920 | (2) Change in log city population 1910 to 1920 | (3) Change in log rural population 1910 to 1920 | (4) Change in urbanization 1910 to 1920 |
|---|---|---|---|---|
| log(*dist*) | −1.12 | 0.012 | 0.006 | 0.068 |
| | (1.01) | (0.016) | (0.014) | (0.076) |
| Drop Zemplén county | No | No | No | Yes |
| $R^2$ | 0.213 | 0.025 | 0.025 | 0.023 |
| $N$ | 22 | 15 | 22 | 21 |

| | (5) Change in urbanization 1910 to 1930 | (6) Change in log city population 1910 to 1930 | (7) Change in log rural population 1910 to 1930 | (8) Change in urbanization 1910 to 1930 |
|---|---|---|---|---|
| log(*dist*) | 0.751** | 0.120** | −0.012 | 0.855* |
| | (0.359) | (0.042) | (0.007) | (0.423) |
| Drop Zemplén county | No | No | No | Yes |
| $R^2$ | 0.125 | 0.299 | 0.095 | 0.124 |
| $N$ | 22 | 15 | 22 | 21 |

Column (1) presents the results of estimating equation (1), column (2) presents the results of estimating equation (2), and columns (3) presents the results of estimating equation (3) on 1910 and 1920 census data. Column (4) is identical to column (1), except that I drop Zemplén county, an outlier in its change in urbanization. Columns (5) to (8) repeat the same regressions as columns (1) to (4), but using 1930 census data. The unit of observation is a county in post-1920 Hungary, and log(*dist*) is the log of distance (in kilometers) between the county seat and the post-1920 border of Hungary. The dependent variables in columns (2), (3), (6) and (7) are measured in logs, while the dependent variables in columns (1), (4), (5) and (8) are measured in percentage points. *Change in urbanization* is the change in the share of people living in cities above 20,000 inhabitants. *Change in log city population* is the log change in the number of people living in cities above 20,000 inhabitants. *Change in log rural population* is the log change in the number of people living outside cities above 20,000 inhabitants. In all the columns, I restrict the sample to counties whose distance from the post-1920 border is less than 60 kilometers. Heteroskedasticity-robust standard errors in parentheses. *: significant at 10%; **: significant at 5%. Source: Censuses of the Hungarian Kingdom, 1910, 1920 and 1930.

Figure 13: The effect of the new border on urbanization by 1920




This figure presents a scatterplot of changes in urbanization between 1910 and 1920 against distance from the new border. Each blue circle corresponds to a county in post-1920 Hungary. The horizontal axis represents distance (in kilometers) between the county seat and the post-1920 border of Hungary. The vertical axis represents the 1910 to 1920 change (in percentage points) in the share of people who live in cities (settlements above 20,000 inhabitants). The regression line between these two variables is shown in red, while the regression line after dropping the outlier county in the top left corner (Zemplén county) is shown in green.

# B Derivation of equilibrium conditions

## B.1 Model of Section 3.1

In this section, I derive equations (6) to (9), which characterize the equilibrium of the model of Section 3.1. First, I use consumers' CES demand function for goods to write the price index at $m$ as

$$P_m = \left[ \sum_o p_o^{1-\sigma} L_o \tau \left( \mu_o, \mu_m \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \tag{B1}$$

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 (5) as

$$p_m^\sigma = \sum_o P_o^{\sigma-1} p_o L_o \tau \left( \mu_m, \mu_o \right)^{1-\sigma}. \tag{B2}$$

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

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

The maximizations on the right-hand side do not depend on the worker's index. As a consequence, workers living at a given location $r$ all choose the same trading place $m$; in what follows, I denote this trading place by $\mu(r)$.[39] Using this and denoting real income by $\omega_m = \frac{p_m}{P_m}$, the previous formula reduces to

$$\log L(r) = \nu + \theta^{-1}\left[a(r) + \varsigma(\mu(r), r)^{-1}\omega_{\mu(r)}\right] \tag{B3}$$

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

$$\varsigma(\mu(r), r)^{-1}\omega_{\mu(r)} \geq \varsigma(\mu_m, r)^{-1}\omega_m \qquad \forall m. \tag{B4}$$

Now use $P_m = \frac{p_m}{\omega_m}$ to rewrite equations (B1) and (B2):

$$p_m^{1-\sigma}\omega_m^{\sigma-1} = \sum_o p_o^{1-\sigma} L_o \tau(\mu_o, \mu_m)^{1-\sigma} \tag{B5}$$

$$p_m^{\sigma} = \sum_o p_o^{\sigma}\omega_o^{1-\sigma} L_o \tau(\mu_m, \mu_o)^{1-\sigma} \tag{B6}$$

Recall that shipping costs are assumed to be symmetric: $\tau(\mu_m, \mu_o) = \tau(\mu_o, \mu_m)$. 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

$$p_m = \omega_m^{\iota}.$$[40]

Then note that for $\iota = \frac{\sigma-1}{2\sigma-1}$, both (B5) and (B6) imply

$$\omega_m^{\frac{\sigma(\sigma-1)}{2\sigma-1}} = \sum_o \omega_o^{-\frac{(\sigma-1)^2}{2\sigma-1}} L_o \tau(\mu_m, \mu_o)^{1-\sigma}. \tag{B7}$$

Defining market access as the right-hand side of equation (B7), we obtain the following

[39] If there is a tie between trading places $m$ and $o$ for a 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.

[40] 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.

relationship between real income and market access at any trading place $m$:

$$\omega_m = MA_m^{\frac{2\sigma-1}{\sigma(\sigma-1)}}$$

Using this relationship to substitute for $\omega_m$ in equations (B3), (B4) and (B7) yields equations (6), (7) and (8), respectively. Finally, equation (9) simply follows from the fact that $L_m$ equals the number of people trading at $m$.

## B.2 Model of Section 3.3

In this section, I derive the equilibrium conditions of the model of Section 3.3. Workers' utility (11) and the extreme value distribution of idiosyncratic amenities imply that the share of population choosing to live at $r$ equals

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

in this model, from which

$$\log L(r) = \nu + \theta^{-1}\left[a(r) + F\left(\varsigma(\mu(r), r)^{-1}\omega_{\mu(r)}\right)\right]. \tag{B8}$$

The remaining equilibrium conditions, (7) to (9), are unchanged as they are not influenced by the $F(\cdot)$ function. As a result, equation (10) holds as well. Plugging equation (10) into (B8) yields equation (12).

# C Isomorphic models

## C.1 A model with firms

This section presents a model in which goods are not produced by workers at home but by monopolistically competitive firms that hire workers in competitive labor markets. I first outline the assumptions of the model. Next, I show a formal isomorphism between this model and the one presented in Section 3.1.

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

$$u_m(r,i) = a(r,i) + \varsigma(\mu_m, r)^{-1}\left[\sum_{k=1}^{K} c_m^k(r,i)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{C1}$$

where $c_m^k(r,i)$ 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 to the linear production technology of Section 3.1, 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 $\sigma$ and shipping costs are of the iceberg type, the firm sets a mill price that is a constant markup over the wage:

$$p_m^k = p_m = \frac{\sigma}{\sigma - 1} w_m \tag{C2}$$

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

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 (C1).*
2. *Firms choose their prices and quantities to maximize profits, and profits are driven down to zero by free entry. Therefore, equation (C2) holds, and each firm produces $f(\sigma - 1)$ units of its good.*
3. *The market for labor clears at every trading place, that is,*

$$L_m = K_m [f(\sigma - 1) + f] = K_m f \sigma \tag{C3}$$

*for all $m$, where $L_m$ denotes the mass of workers commuting to $m$, and $K_m$ denotes the number of goods produced at $m$.*[41]

[41] 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*

$$f(\sigma-1)=\sum_o \tau(\mu_m,\mu_o)^{1-\sigma} p_m^{-\sigma} P_o^{\sigma-1} w_o L_o \tag{C4}$$

*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, market access and trading place assignments is isomorphic between this model and the one in Section 3.1.

**Proposition 4.** *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 (6) to (9). Hence, the equilibrium distribution of population, market access and trading place choices is the same as in the model of Section 3.1.*

*Proof.* First, use equations (C2) and (C3) to write the price index at $m$ as

$$P_m=f^{\frac{1}{\sigma-1}}\sigma^{\frac{\sigma}{\sigma-1}}(\sigma-1)^{-1}\left[\sum_o w_o^{1-\sigma}L_o\tau(\mu_o,\mu_m)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{C5}$$

and the goods market clearing condition (C4) as

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

from which

$$w_m^{\sigma}=f^{-1}\sigma^{-\sigma}(\sigma-1)^{\sigma-1}\sum_o P_o^{\sigma-1}w_oL_o\tau(\mu_m,\mu_o)^{1-\sigma}. \tag{C6}$$

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

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

just like in the model of Section 3.1. Using this and denoting real income by $\omega_m=\frac{w_m}{P_m}$, the previous formula reduces to

$$\log L(r) = \nu + \theta^{-1}\left[a(r) + \varsigma(\mu(r), r)^{-1}\omega_{\mu(r)}\right] \tag{C7}$$

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

$$\varsigma(\mu(r), r)^{-1}\omega_{\mu(r)} \geq \varsigma(\mu_m, r)^{-1}\omega_m \qquad \forall m. \tag{C8}$$

Now use $P_m = \frac{w_m}{\omega_m}$ to rewrite equations (C5) and (C6):

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

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

from which, following the same procedure as in Appendix B.1, I obtain

$$\omega_m^{\frac{\sigma(\sigma-1)}{2\sigma-1}} = \sum_o \omega_o^{-\frac{(\sigma-1)^2}{2\sigma-1}}L_o\tau(\mu_m, \mu_o)^{1-\sigma} \tag{C9}$$

where I also used the normalization $f = \sigma^{-\sigma}(\sigma-1)^{\sigma-1}$. Defining market access as the right-hand side of equation (C9) and following the same steps as in the end of Appendix B.1, equations (6) to (9) follow. □

## C.2 A model with nontradable goods

This section presents a model in which workers also consume a nontradable good produced at their residential location. In particular, I assume that each worker demands one unit of a homogenous nontradable good, whose price she takes as given. After purchasing the unit of the nontradable good, she spends the rest of her income on tradables, just like in the model of Section 3.1. One interpretation of the nontradable good is the housing consumed by the worker.

For simplicity, I assume that the nontradable good is produced using the same CES aggregate of tradables as what enters the worker's utility function. One unit of the CES aggregate of tradables can be used to produce $\eta(r) > 0$ units of the nontradable good. Thus, $\eta(r)$ measures productivity with which nontradables can be produced at $r$. I assume that $\eta(r)$ is exogenously given at every location, but allow it to vary across locations.

There is perfect competition in the market for the nontradable good at each location. This and the constant returns production technology of nontradables imply that the price of nontradables, $R(r)$, becomes equal to their unit cost of production:

$$R(r) = \eta(r)^{-1}\varsigma(\mu_m, r)P_m \tag{C10}$$

if $m$ is the trading place of location $r$, where I have used the fact that a unit of the CES aggregate of tradable inputs costs $P_m$ and the cost of shipping them from $m$ to $r$ equals $\varsigma(\mu_m, r)$.[42]

In what follows, I define the model's equilibrium and derive the equilibrium conditions under these assumptions. Next, I show a formal isomorphism between this model and the one presented in Section 3.1.

**Definition 4.** *Given parameters $\{\sigma, \theta, \overline{L}\}$, geography $S$, $\{\mu_1, \ldots, \mu_M\}$ and functions $\{a, \eta\} : S \to \mathbb{R}_+$, $\{\tau, \varsigma\} : S^2 \to \mathbb{R}_+$, an **equilibrium** of the economy with nontradable goods consists of a population distribution $L : S \to \mathbb{R}_+$; tradable goods' consumption levels $c : [0, \overline{L}]^2 \times S \times \{1, \ldots, M\} \to \mathbb{R}_+$; tradable goods' prices and production levels $\{p, x\} : [0, \overline{L}] \times \{1, \ldots, M\} \to \mathbb{R}_+$; nontradable goods' prices and quantities $\{R, H\} : S \to \mathbb{R}_+$; 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 (4) subject to the production technology of tradables and their budget constraint.*

2. *Perfect competition drives down profits to zero in the production of nontradables. As a result, equation (C10) holds.*

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

$$x_m^j = \sum_o \tau(\mu_m, \mu_o)^{1-\sigma} \left(p_m^j\right)^{-\sigma} P_o^{\sigma-1} \sum_{r:o=\mu(r)} \left[(p_o - R(r)) L(r) + \eta(r)^{-1} \varsigma(\mu_o, r) P_o H(r)\right] \tag{C11}$$

*for any worker $j$, where $x_m^j$ denotes the worker's production level, $m$ denotes the trading place where she sells her product, $(p_o - R(r)) L(r)$ equals total spending on tradables by workers living at location $r$, and $\eta(r)^{-1} \varsigma(\mu_o, r) P_o H(r)$ equals total spending on tradable inputs in the production of nontradables at $r$.*

4. *The market for each nontradable good clears at every location, implying*

$$H(r) = L(r) \tag{C12}$$

*as each worker demands one unit of nontradables.*

In equilibrium, a worker $i$ living at $r$ and trading at $m$ obtains indirect utility

$$u_m(r, i) = a(r, i) + \varsigma(\mu_m, r)^{-1} \frac{p_m - R(r)}{P_m}.$$

[42] By perfect competition, profits made in the production of nontradables equal zero at every location. As a consequence, it does not matter who produces these nontradables. For completeness, I assume that workers are the ones producing them.

as her total spending on tradables equals her production income $p_m$ minus her spending on the nontradable good, $R(r)$. Combining this with equation (C10) and rearranging yields

$$u_m(r,i) = a(r,i) - \eta(r)^{-1} + \varsigma(\mu_m, r)^{-1} \frac{p_m}{P_m}.$$

The extreme value distribution of idiosyncratic amenities implies that the share of population living at $r$ is then given by

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

from which

$$\log L(r) = \nu + \theta^{-1}\left[a(r) - \eta(r)^{-1} + \varsigma(\mu(r), r)^{-1}\omega_{\mu(r)}\right] \tag{C13}$$

where $\omega_m = \frac{p_m}{P_m}$, and $\nu = \log\left(\overline{L}\right) - \log\left[\sum_s e^{a(s)-\eta(s)+\varsigma(\mu(s),s)^{-1}\omega_{\mu(s)}}\right]$.

To obtain the remaining equilibrium conditions, combine (C11) with (C10) and (C12) to get

$$x_m^j = \sum_o \tau(\mu_m,\mu_o)^{1-\sigma}\left(p_m^j\right)^{-\sigma} P_o^{\sigma-1} \sum_{r:o=\mu(r)} p_o L(r) = \sum_o \tau(\mu_m,\mu_o)^{1-\sigma}\left(p_m^j\right)^{-\sigma} P_o^{\sigma-1} p_o L_o$$

which is the same as the market clearing condition for tradables, (5), in the model of Section 3.1. Following the same steps as in Appendix B.1, one can then derive the following set of equilibrium conditions:

$$\varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \geq \varsigma(\mu_m, r)^{-1} MA_m^{\frac{2\sigma-1}{\sigma(\sigma-1)}} \qquad \forall m \tag{C14}$$

$$MA_m = \sum_o MA_o^{-\frac{\sigma-1}{\sigma}} L_o \tau(\mu_m,\mu_o)^{1-\sigma} \tag{C15}$$

$$L_m = \sum_{r:\ m=\mu(r)} L(r) \tag{C16}$$

which are identical to conditions (7), (8) and (9) in the model of Section 3.1. One also obtains the same relationship between real income and market access as in the model of Section 3.1:

$$\omega_m = MA_m^{\frac{2\sigma-1}{\sigma(\sigma-1)}}$$

Substituting this into equation (C13) yields

$$\log L(r) = \nu + \theta^{-1}\left[a(r) - \eta(r)^{-1} + \varsigma(\mu(r), r)^{-1} MA_{\mu(r)}^{\frac{2\sigma-1}{\sigma(\sigma-1)}}\right]. \tag{C17}$$

The next proposition introduces a mapping between the two models' parameters under which the equilibrium distribution of population, market access and trading place assignments is isomorphic between this model and the one in Section 3.1.

**Proposition 5.** *If* $a(r) - \eta^{-1}(r)$ *in the model with nontradables equals* $a(r)$ *in the model of Section 3.1, then the equilibrium conditions of the two models become identical. As a result, the equilibrium distribution of population, market access and trading place choices is the same between the two models.*

*Proof.* Equations (C14), (C15) and (C16) are identical to their counterparts, equations (7), (8) and (9) in the model of Section 3.1. Replacing $a(r) - \eta(r)^{-1}$ by $a(r)$ in (C17), this equation also becomes identical to its counterpart, equation (6) in Section 3.1. □

Proposition 5 shows that the model with nontradables only differs from the model of Section 3.1 in the interpretation of its location-specific structural residuals. The model of Section 3.1 interprets these residuals as amenities, whereas the model with nontradables interprets them as a combination of amenities and local productivity in nontradables. However, as these structural residuals are backed out of the model to match data on city populations (Section 4.5), one does not necessarily need to take a stand on their interpretation. Thus, the estimated effects of the new border on trade, real income and urbanization remain unchanged if one reinterprets the residuals as if they came from the model presented in this section.

# D Multi-sector model: Details

This appendix provides further details on the multi-sector model introduced in Section 6.2.

## D.1 Equilibrium conditions

Workers' utility (20) and the extreme value distribution of idiosyncratic amenities imply that the share of workers who choose to live at $r$ and produce in sector $n$ equals

$$\frac{L(r,n)}{\bar{L}} = \frac{\left[e^{a(r,n)+\bar{\alpha}\max_m \varsigma(\mu_m,r)^{-1}\frac{p_m(n)}{P_m}}\right]^{\zeta^{-1}}}{\sum_s \sum_\ell \left[e^{a(s,\ell)+\bar{\alpha}\max_o \varsigma(\mu_o,s)^{-1}\frac{p_o(\ell)}{P_o}}\right]^{\zeta^{-1}}} \tag{D1}$$

where $p_m(n)$ denotes the price of the worker's product at trading place $m$,[43] $\varsigma(\mu_m, r)$ denotes the worker's utility cost of shipping between home and $m$, and $P_m$ denotes the price index of all goods at $m$. Cobb–Douglas utility implies

$$P_m = \bar{\alpha} \prod_\ell P_m(\ell)^{\alpha_\ell}$$

[43] Appendix D.2 shows that this price does not depend on the worker's index.

such that $P_m(\ell)$ is the CES price index of sector-$\ell$ goods at $m$. Note that, by (D1), $L(r,n) > 0$. That is, every location has a positive number of workers in each sector.

Denoting the trading place that a worker living at $r$ and producing in sector $n$ chooses by $\mu(r,n)$ and denoting the worker's (normalized) real income at trading place $m$ by $\omega_m(n) = \bar{\alpha}\frac{p_m(n)}{P_m}$, equation (D1) reduces to

$$\log L(r,n) = \nu + \zeta^{-1}\left[a(r,n) + \varsigma(\mu(r,n), r)^{-1}\omega_{\mu(r,n)}(n)\right] \tag{D2}$$

where $\nu = \log\left(\bar{L}\right) - \log\left[\sum_s \sum_\ell \left[e^{a(s,\ell)+\bar{\alpha}\max_o \varsigma(\mu_o,s)^{-1}\frac{p_o(\ell)}{P_o}}\right]^{\zeta^{-1}}\right]$. Also, by the definition of $\mu(r,n)$,

$$\varsigma(\mu(r,n), r)^{-1}\omega_{\mu(r,n)}(n) \geq \varsigma(\mu_m, r)^{-1}\omega_m(n) \qquad \forall m. \tag{D3}$$

As in the model of Section 3.1, producing a unit of a good requires a unit of labor, and workers engage in monopolistic competition at the trading place where they ship their product. Shipping goods across trading places is subject to an iceberg cost $\tau(\mu_m, \mu_o; n)$ in sector $n$. These costs are symmetric across trading places, that is, $\tau(\mu_m, \mu_o; n) = \tau(\mu_o, \mu_n; n)$ for any $m$, $o$ and $n$.

As $\sigma > 1$, any worker $j$ has an incentive to produce the maximum possible quantity of her product, $x^j = 1$, by the same argument as in Appendix B.1. Also, the worker sets a price $p^j_m$ at which demand for her product meets supply. Just like in the model of Section 3.1, neither demand nor supply depends on the worker's index, implying that the equilibrium price is common across workers within sector and trading place: $p^j_m = p_m(n)$.

I define the equilibrium of the multi-sector economy as follows.

**Definition 5.** *Given parameters* $\{\sigma, \zeta, \overline{L}, \alpha_1, \ldots, \alpha_N\}$, *geography* $S$, $\{\mu_1, \ldots, \mu_M\}$ *and functions* $a : S \times N \to \mathbb{R}_+$, $\tau : S^2 \times N \to \mathbb{R}_+$, $\varsigma : S^2 \to \mathbb{R}_+$, *an* ***equilibrium*** *of the multi-sector economy consists of a distribution of workers across locations and sectors* $L : S \times N \to \mathbb{R}_+$; *consumption levels* $c : \left[0, \overline{L}\right]^2 \times S \times \{1, \ldots, M\} \to \mathbb{R}_+$; *goods' prices and production levels* $\{p, x\} : \left[0, \overline{L}\right] \times \{1, \ldots, M\} \to \mathbb{R}_+$; *and a function that assigns a trading place to each residential location-sector pair,* $\mu : S \times N \to \{1, \ldots, M\}$, *such that the following hold:*

1. *Workers choose their consumption, production, price, residential location, sector and trading place to maximize their utility (20) subject to the production technology and their budget constraint.*

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

$$x^j_m = \sum_o \tau(\mu_m, \mu_o; n)^{1-\sigma}\left(p^j_m\right)^{-\sigma} P_o(n)^{\sigma-1}\alpha_n \sum_\ell p_o(\ell) L_o(\ell) \tag{D4}$$

*for any worker* $j$ *producing in sector* $n$, *where* $x^j_m$ *denotes the worker's production* *level, $m$ denotes the trading place where she sells her product, and $L_o(\ell)$ is the total number of workers in sector $\ell$ trading at trading place $o$:*[44]

$$L_o(\ell) = \sum_{r:o=\mu(r,\ell)} L(r,\ell) \tag{D5}$$

Using the facts that $x^j_m = 1$ and $p^j_m = p_m(n)$, market clearing condition (D4) can be written as

$$p_m(n)^\sigma = \alpha_n \sum_o P_o(n)^{\sigma-1} \left[\sum_\ell p_o(\ell) L_o(\ell)\right] \tau(\mu_m,\mu_o;n)^{1-\sigma}$$

while the price index of sector $n$ at trading place $m$ equals

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

Using the definition of $\omega_m(n)$, one can rewrite the last two equations as

$$\omega_m(n)^\sigma \prod_\ell P_m(\ell)^{\alpha_\ell\sigma} = \alpha_n \sum_o P_o(n)^{\sigma-1} \prod_\ell P_o(\ell)^{\alpha_\ell} \left[\sum_\ell \omega_o(\ell) L_o(\ell)\right] \tau(\mu_m,\mu_o;n)^{1-\sigma} \tag{D6}$$

and

$$P_m(n)^{1-\sigma} = \sum_o \omega_o(n)^{1-\sigma} \prod_\ell P_o(\ell)^{\alpha_\ell(1-\sigma)} L_o(n)\, \tau(\mu_o,\mu_m;n)^{1-\sigma}. \tag{D7}$$

Equations (D2), (D3), (D5), (D6) and (D7) pin down the equilibrium distribution of real income in each trading place-sector pair $\omega_m(n)$, the number of people living in each location-sector pair $L(r,n)$, the number of people trading in each trading place-sector pair $L_m(n)$, trading place assignments to locations and sectors $\mu(r,n)$, and sectoral price indices at each trading place $P_m(n)$. Unlike in the model of Section 3.1 (see Appendix B.1), equations (D6) and (D7) cannot be reduced to a single equation in the multi-sector model. Nonetheless, the system defined by (D2), (D3), (D5), (D6) and (D7) turns out to be numerically as tractable as the model of Section 3.1.

## D.2 Structural estimation

This section provides further details on the structural estimation of the multi-sector model. As Section 6.2.2 described, I define three broad sectors of the economy: agriculture (A), manufacturing (M) and an "other sector" comprising the rest of the economy (O). Probáld (2009) estimates that agriculture was responsible for 44% of Hungary's GDP in 1910, while manufacturing accounted for approximately 25% of GDP. This allows me to set sectoral

[44]The right-hand side of equation (D4) follows from CES demand for worker $j$'s product at any trading place $o$.

spending shares, which equal GDP shares in the model, to $\alpha_A = 0.44$, $\alpha_M = 0.25$ and $\alpha_O = 1 - 0.44 - 0.25 = 0.31$.

Given that the value of shipping cost parameter $\phi$ was calibrated to agricultural price data in Section 4, I set the corresponding parameter in agriculture to the same value: $\phi_A = 3.51 \cdot 10^{-4}$. In my baseline estimation (row 1 of Table 8), I set the corresponding parameters in manufacturing and the other sector to the same value as well: $\phi_A = \phi_M = \phi_O$. For rows 2 to 4 of Table 8, I re-estimate the model with different values of $\phi_M$ and $\phi_O$.

I calibrate the value of the other shipping cost parameter $\psi$ in a procedure that is similar to the one used in Section 4. In particular, I choose $\psi$ such that the model matches the standard deviation of the employment of settlements above 5,000 inhabitants.[45] This procedure pins down a value of $\psi = 2.5 \cdot 10^{-1}$, which is slightly higher but is in the same ballpark as in the one-sector model.

Finally, I estimate parameter $\zeta$, the values of city-sector amenities $a_c(n)$ and the values of sector amenities outside cities $a(n)$ to match city-sector employment data and aggregate sectoral employment levels in both 1910 and 1930, as well as moment condition (21). For total population $\bar{L}$, I use total country-level employment both in 1910 and in 1930. This procedure relies on iterating on equations (D2), (D3), (D5), (D6), (D7) and (21) until all these targeted moments are matched by the model.

[45] In Section 4, I set this threshold to 2,000 inhabitants. However, several settlements below 5,000 inhabitants have zero employment in certain sectors. By using a somewhat higher threshold, I can avoid these zeros influencing the results.