CHAPTER 20

Transportation Costs and the Spatial Organization of Economic Activity

Stephen J. Redding[*], **Matthew A. Turner**[†]
[*]Economics Department and WWS, Princeton University Fisher Hall, Princeton, NJ, USA
[†]Economics Department, Brown University, Providence, RI, USA

Contents

Handbook of Regional and Urban Economics, Volume 5B
ISSN 1574-0080, http://dx.doi.org/10.1016/B978-0-444-59531-7.00020-X

Abstract

This chapter surveys the theoretical and empirical literature on the relationship between the spatial distribution of economic activity and transportation costs. We develop a multiregion model of economic geography that we use to understand the general equilibrium implications of transportation infrastructure improvements within and between locations for wages, population, trade, and industry composition. Guided by the predictions of this model, we review the empirical literature on the effects of transportation infrastructure improvements on economic development, paying particular attention to the use of exogenous sources of variation in the construction of transportation infrastructure. We examine evidence from different spatial scales, between and within cities. We outline a variety of areas for further research, including distinguishing reallocation from growth and dynamics.

Keywords

Highways, Market access, Railroads, Transportation

JEL Classification Codes

F15, R12, R40

20.1. INTRODUCTION

The organization of economic activity in geographic space depends crucially on the transportation of goods and people. Most production involves the movement of inputs such as raw materials, labor, and fuel from different locations. Most consumption requires either the conveyance of finished goods or the transfer of people to the points at which goods and services are supplied. The transportation sector as a whole typically accounts for around 5% of gross domestic product (GDP), and transportation networks constitute some of the largest investments ever made. In the United States, the Interstate Construction Program extended to 42,795 miles of highways with an estimated cost of $128.9 billion (1991 US dollars).[1] Multiplying estimates of the cost per interstate lane kilometer found in Duranton and Turner (2012) by the extent of the system gives

[1] US Department of Transportation, Federal Highway Administration, interstate cost estimates reported to Congress.

much larger values. In China, the National Trunk Highway System involved the construction of around 21,747 miles (35,000 km) of highways over a period of 15 years at an estimated construction cost of around $120 billion (in current price US dollars).[2]

Transportation technologies themselves have undergone large-scale changes over time, which have in turn reshaped the spatial organization of economic activity. For most of human history, the movement of goods and people was limited by the physical capabilities of humans and their animals. The invention of the railroad reduced transportation costs and created a hub-and-spoke transportation network that was characterized by substantial fixed costs (e.g., in stations and goods yards) and favored point-to-point travel between the central cities. The development of the internal combustion engine (and hence the automobile and truck) in turn created greater flexibility in transportation, benefiting lower-density locations relative to central cities.[3] Even within existing transportation technologies, such as maritime shipping, there have been large-scale changes in the organization of economic activity in the form of containerization and the adoption of new information and communication technologies such as the computer. These innovations have played an important part in the development of integrated logistics networks, which control the movement of a package from its origin to its destination, and integrate packaging, storage, transport, inventories, administration, and management. The discovery of entirely new modes of transportation, such as air travel, has further transformed the relative attractiveness of locations for economic activity.

This chapter describes our current understanding of the way that transportation costs and transportation infrastructure affect the organization of economic activity within a country. We first provide some basic facts about transportation costs within and between cities. Then we develop a multiregion model of economic geography as a framework to organize our discussion of the empirical literature. The existing empirical literature on the effects of transportation costs and infrastructure can be usefully divided into two parts. The first of these parts considers the role of transportation costs between cities and is mainly interested in the movement of goods, while the second considers the role of transportation costs within cities and is mainly interested in the movement of people. Our model unifies the analysis of within-city and between-city transportation, thereby allowing us to simultaneously consider the two previously disparate strands of the empirical literature. Analysis of our model yields structural equations corresponding to the reduced-form estimating equations on which the two parts of the empirical literature are based. The divergence between theoretically founded

[2] Faber (2015).

[3] See, for example, the discussion in Glaeser and Ponzetto (2013).

structural equations and reduced-form estimating equations, in turn, provides insight into the inference problems that reduced-form estimation must overcome. Finally, with a handful of exceptions, the existing literature provides only an incomplete understanding of general equilibrium effects of transportation infrastructure and little basis for welfare analysis. The model that we develop illustrates a possible direction for research on this issue.

The available empirical literature provides credible, causal estimates of the effect of roads, railroads, and subways on outcomes such as population density, land rents, and output. In addition to providing particular elasticity estimates, this literature is large enough to suggest three preliminary conclusions. First, that the effects of different types of infrastructure are similar across economies at different stages of development and are not especially sensitive to the spatial scale of the unit of observation. Second, that different modes of transportation are not interchangeable. Railroads affect production more than the population and the effects of railroads on the location of production vary systematically with the weight-to-value ratio of output, while the spatial organization of the population is more sensitive to roads and subways than to railroads. Finally, and unsurprisingly, institutions matter. The existing empirical literature suggests that politics plays an important role in the allocation of infrastructure and that these politics vary systematically across countries.

Determining the extent to which the effects of transportation infrastructure reflect growth or reorganization is fundamental to understanding its role in the spatial organization of economic activity. Indeed, this question is at the heart of Fogel's classic study of railroads in the late nineteenth century United States. While the current empirical literature provides credible causal estimates of the effects of transportation infrastructure, it is impossible for the reduced-form regressions conducted in almost all of the empirical studies that we survey to separately identify the effect of transportation infrastructure on the growth and reorganization of economic activity. We suggest two approaches to this problem: one is a simple extension of the existing reduced-form literature, and the second is an implementation of our structural model. The handful of articles which shed light on this question suggest that reorganization is often about as important as growth. This is an important area for further research.

The remainder of this chapter is structured as follows. Section 20.2 reports some descriptive evidence regarding transportation costs across countries and over time. Section 20.3 introduces the theoretical framework that we use to organize our discussion of the empirical evidence. Section 20.4 uses the model to develop a reduced-form framework for examining the impacts of transportation infrastructure on the distribution of economic activity between and within cities. Section 20.5 uses this reduced-form framework to review existing empirical evidence of these impacts. Section 20.6 discusses the interpretation of this existing evidence. Section 20.7 summarizes our conclusions.

20.2. STYLIZED FACTS ABOUT TRANSPORTATION

In this section, we present stylized facts about transportation costs for goods and people, both over a long historical time period and across countries.[4] The key features of the data are as follows. First, there is a secular decline in transportation costs for goods. Second, there is a change in the relative importance of different transportation modes over time (e.g., rail versus road versus air) and for value versus weight. Third, transportation costs for people continue to be important. Commuting costs remain substantial, both in terms of the opportunity cost of time and in terms of overall household expenditure.

20.2.1 Transportation costs for goods

To provide a rough indication of the real resources involved in the transportation sector over time, Figure 20.1 displays the share of the transportation sector in US GDP from the late nineteenth century to the late twentieth century.[5] The striking feature of this figure is the long secular decline in the share of the transportation sector, which is even

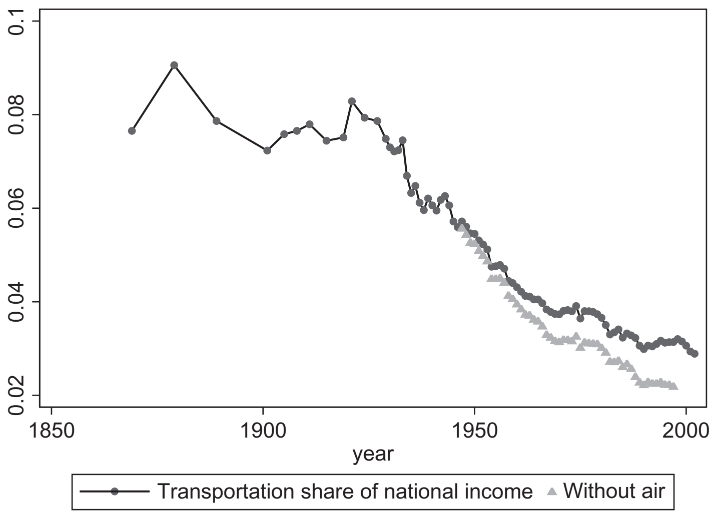


Figure 20.1 Share of transport sector in US GDP. *Sources: Department of Commerce (since 1929), and Historical Statistics of the United States (Martin Series) before then.*

[4] For a more detailed analysis of the evolution of transportation costs over time in the United States, see Glaeser and Kohlhase (2004).

[5] Figures 20.1–20.4 replicate similar figures that appear in Glaeser and Kohlhase (2004).

more rapid toward the end of the twentieth century if air transport is excluded. The share of US GDP attributed to transportation[6] fell from about 8% in 1929 to about 3% by 1990, of which about one-quarter is air transport. While these numbers are striking, they may reflect the increased importance of nontraded services rather than a decrease in the importance of transportation. In addition, while these GDP figures tell us about the resources devoted to moving goods, they do not tell us about the amount and value of the goods being moved.

To provide a more direct measure, Figure 20.2 displays the transportation costs for a given mode of transportation (railroads) in the United States over a similar time period (measured as costs per ton mile in 2001 dollars). The figure confirms a secular decline in transportation costs over time. The price per ton mile of rail freight fell from about 18.5 cents in 1890 to about 2 cents in 2000. Figure 20.3 compares the evolution of truck, rail, and pipeline transportation costs for the United States during the post-Second World War period (measured as revenue per ton mile in 2001 dollars). As is

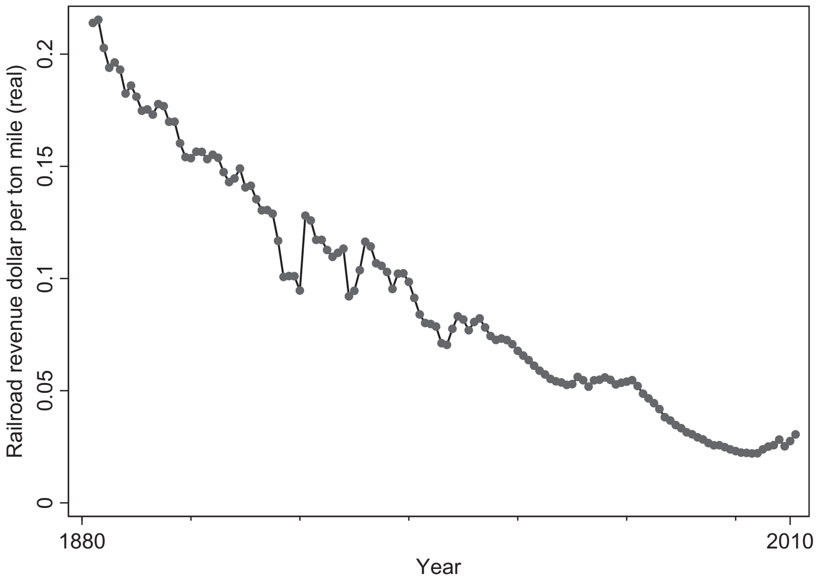


Figure 20.2 The costs of railroad transportation over time. *Sources: Historical Statistics of the United States (until 1970), and Bureau of Transportation Statistics annual reports 1994 and 2002.*

[6] Defined as rail, water, pipeline, trucking, warehousing, air transport, transportation services, and local and interurban rail transit.

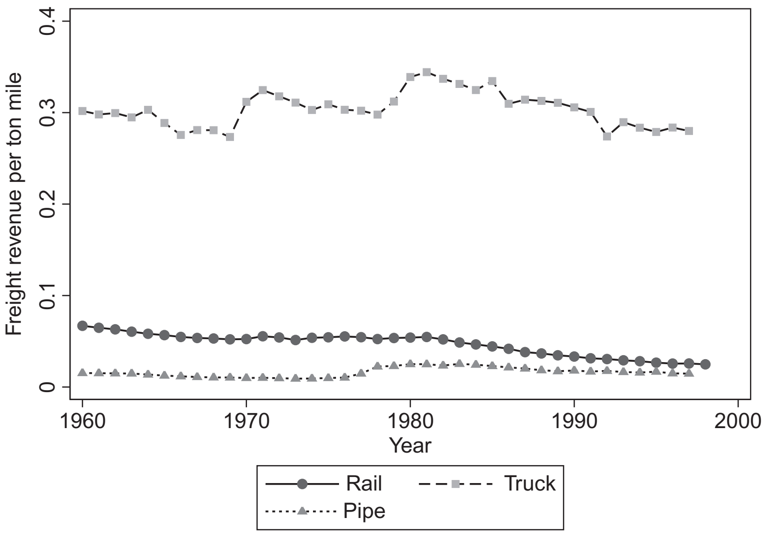


Figure 20.3 Revenue per ton mile (all modes). *Sources: Bureau of Transportation Statistics annual reports.*

apparent from the figure, truck transport is substantially more expensive than rail transport, and its real costs have fallen notably since the early 1980s.[7]

Figure 20.4 shows the evolution of ton miles of freight over time from the mid-1960s. Rail is relatively more important than trucks when we measure volume shipped rather than value because of a widely observed selection effect in which more expensive items are disproportionately shipped by the more expensive transportation mode.[8] As a share of the value of goods, Glaeser and Kohlhase (2004) find that for heavy low-value goods traveling by truck (e.g., lumber), the cost of shipping goods over an average shipment distance can be as high as 20% of the value of the good. For more typical sectors, this value is on the order of 5%. For goods traveling by rail, the corresponding values range from 0.1% to 2%. These findings highlight that the cost of moving freight has dropped dramatically to the point that freight transportation is about 3% of the US economy and that freight charges make up only a small share of the value of final output.

[7] These figures invite the question of why people use trucks at all, the nominally more costly mode. Although trucks have a higher cost per ton mile than rail, the real cost of quality-adjusted transportation services also depends on speed, flexibility, reliability, and a number of other attributes. The large-scale reallocation of transportation expenditure from rail to trucks following the invention of the internal combustion engine suggests that this invention was associated with a substantial reduction in the real cost of quality-adjusted transportation services, at least for many types of shipments and journeys.

[8] This is an example of the Alchian–Allen effect from the international trade literature or "shipping the good apples out."

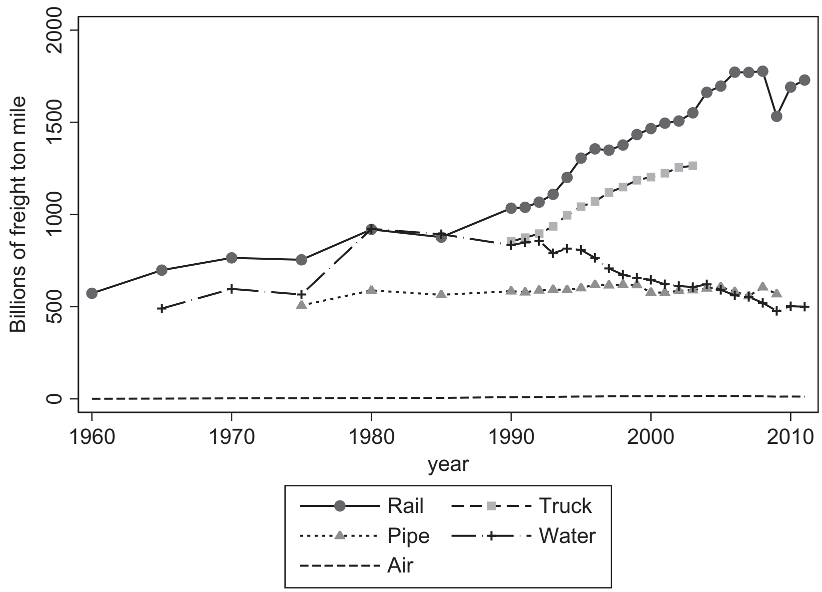
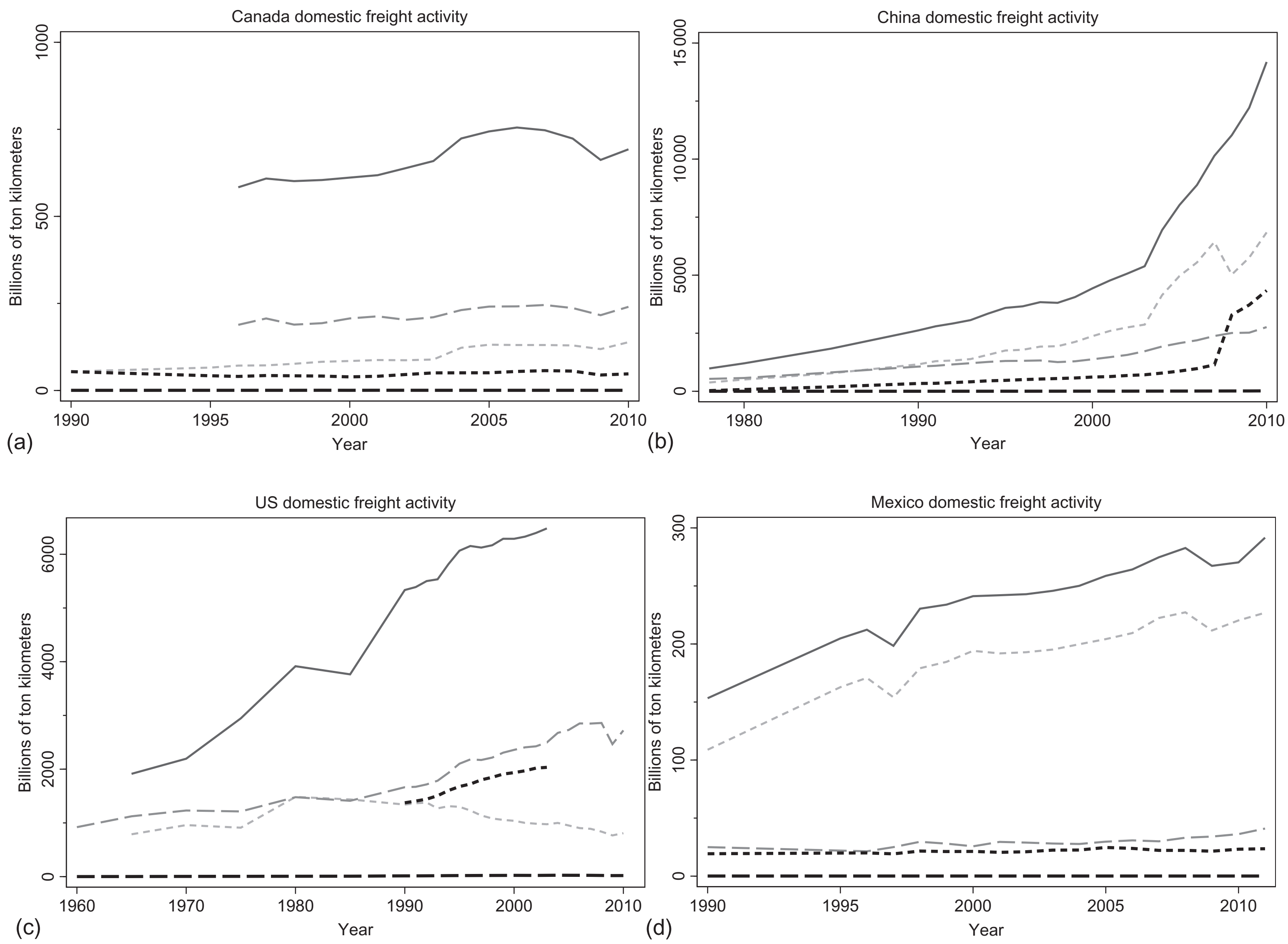


Figure 20.4 Ton miles of freight over time. *Sources: Bureau of Transportation Statistics annual reports.*

To show that these patterns are not specific to the United States, Figure 20.5 describes ton kilometers of domestic trade for seven countries by mode and year. While there are differences between countries, several patterns are clear. First, there is a general trend upward in the amount of domestic trade, as expected given a secular increase in the level economic activity over time. Second, trucking is not the dominant mode of travel in any of our countries by this metric. Third, the amount of material being moved is immense.[9]

Table 20.1 shows the value of international trade by mode for a sample of countries in 2007. In Figure 20.5 we see that the share of ton kilometers for goods that travel by air is negligible. In contrast, in Table 20.1, we see that the share of the value of goods traveling by air is often large. While the two sets of results are not directly comparable (one measures domestic trade and the other measures international trade), together they strongly suggest that high-value goods travel by air and low-value goods travel by ship or rail.

[9] To get a sense of this, a typical coal train in the United States is about 100 cars long, about 2 km, and each car carries about 100 tons of coal, which implies 10,000 tons per train. If such a train travels 100 km, it provides 1 million ton-km of freight service. To carry 5000 billion ton-km of freight per year, a bit less than the current US annual total, we require about 1200 such trains to operate 24 h per day, 365 days per year at 50 km per h.


Canada domestic freight activity
Billions of ton kilometers
0
500
1000
1990
1995
2000
2005
2010
Year
(a)
China domestic freight activity
Billions of ton kilometers
0
5000
10 000
15 000
1980
1990
2000
2010
Year
(b)
US domestic freight activity
Billions of ton kilometers
0
2000
4000
6000
1960
1970
1980
1990
2000
2010
Year
(c)
Mexico domestic freight activity
Billions of ton kilometers
0
100
200
300
1990
1995
2000
2005
2010
Year
(d)

(Continued)

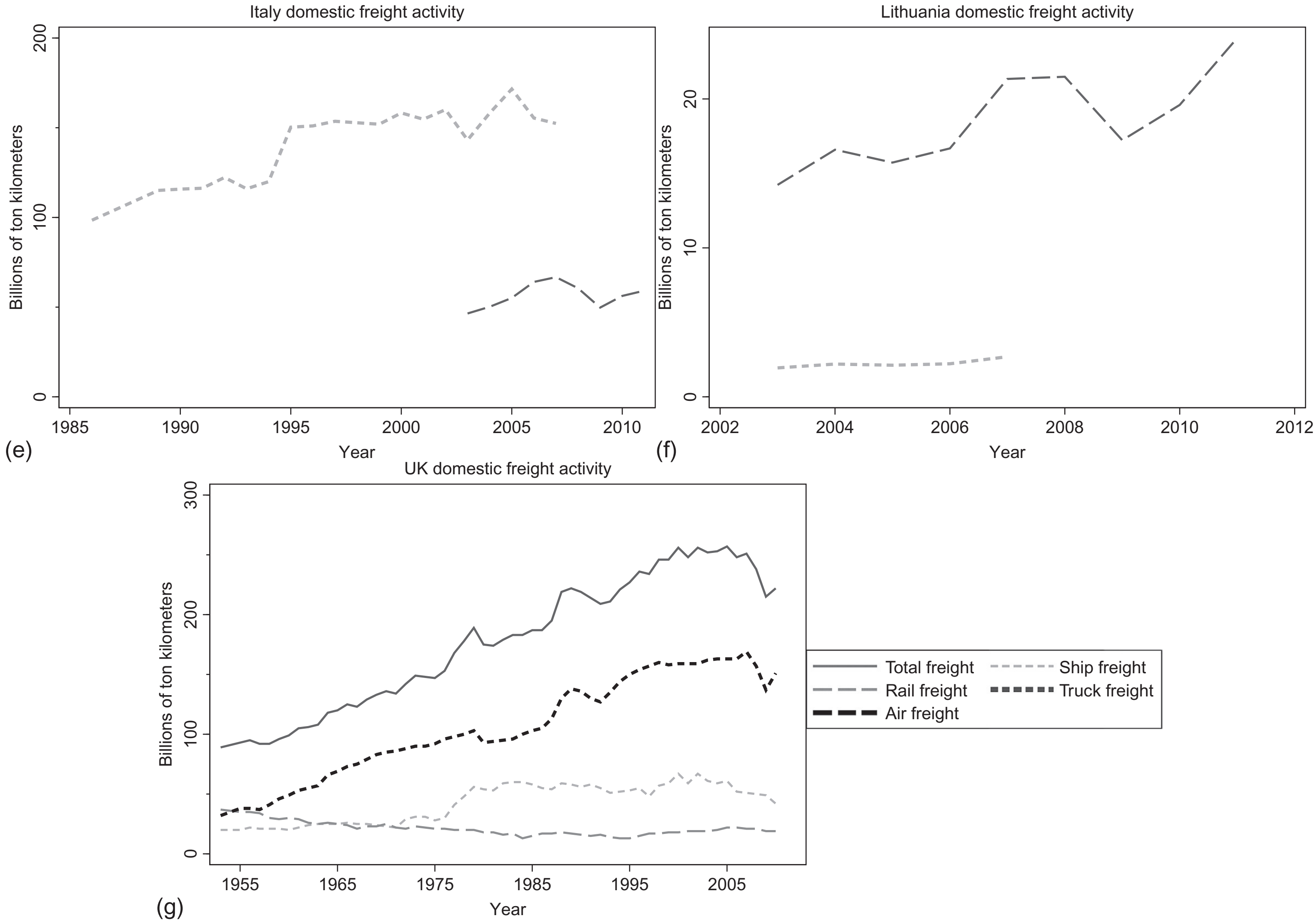


Figure 20.5 Ton kilometers of freight by year and mode for several countries. *Sources: (a) North American Transportation Statistics (2012c); (b) China Data Online (2010); (c) Bureau of Transportation Statistics (2012a); (d) North American Transportation Statistics (2012c); (e) Eurostat (2010); (f) Eurostat (2010); (g) Eurostat (2010).*

Table 20.1 Shares of total international trade by country and mode

	Total	Ship (%)	Air (%)	Truck (%)	Rail (%)
Austria	279	13	8	70	7
Belgium	750	26	14	44	2
Bulgaria	48	38	4	50	7
Canada	798	19	11	47	12
Cyprus	10	83	16	0	
Czech Republic	194	2	5	82	9
Denmark	186	24	8	11	0
Estonia	24	34	4	34	10
EU27	10,261	29	12	32	2
Finland	169	69	15	12	4
France	1105	25	11	9	0
Germany	2231	27	14	37	3
Greece	106	57	7	33	1
Hungary	164	1	10	75	9
Ireland	198	45	31	11	0
Italy	1003	40	9	36	3
Latvia	21	18	3	45	14
Lithuania	38	29	3	56	12
Mexico	554	24	6	56	8
Netherlands	995	27	9	39	1
Poland	259	17	3	55	9
Portugal	119	37	7	52	0
Romania	101	25	2	13	3
Slovakia	96	13	4	60	12
Slovenia	56	16	3	64	4
Spain	615	45	7	44	2
Sweden	313	25	10	15	1
United Kingdom					
United States	3116	45	25	18	4

Total trade for European countries is in billions of euros and for the other countries is in billions of US dollars.
Source: North American Transportation Statistics (2012d) and Eurostat (2012).

Table 20.2 compares employment in for-hire transportation by mode for Canada, Mexico, and the United States in 2002 or 2003 (depending on data availability). Transportation is typically smaller as a share of employment than as a share of GDP. The share of employment in transportation is about 3% for the United States and Canada, and almost 6% for Mexico. In all three countries, the largest fraction of transportation employment is devoted to trucking. Note that the share of labor devoted to for-hire transportation is close to the same as the share devoted to commuting.

A striking feature of international trade in goods is the extent to which the volume of trade in goods declines with distance. Hillberry and Hummels (2008) examine the

Table 20.2 Employment in for-hire transportation as share of total employment

Country	Year	Total employment	All transport	Air	Rail	Truck	Ship
Canada	2002	13	3.56	0.44	0.28	1.29	0.09
Mexico	2003	35	5.75	0.06	0.04	2.75	0.03
United States	2003	130	3.22	0.41	0.17	1.02	0.04

Total employment is in millions of people, and all others are percentages of the total.
Sources: North American Transportation Statistics (2012a) and Bureau of Transportation Statistics (2012b).

pattern of shipments between US mining, manufacturing, and wholesaling firms and find that three-quarters of all shipments, weighted by the value of shipments, begin and end in the same zip code, a conclusion that does not appear to be driven by shipments from wholesalers to retailers. Hummels (1999) documents the cost of air freight between the 1950s and the first decade of the twenty-first century and finds that it decreases by a factor of about 12.5, while the cost of shipping was approximately constant. For comparison, the corresponding decrease for rail, from Glaeser and Kohlhase (2004), is about a factor of 8 for 110 years.[10] Over the same 1955–2004 period, Hummels (1999) documents 5–7% increases in the value and weight of international trade and an 11% average annual increase in the share of the value of trade that travels by air. Limao and Venables (2001) use data describing market price to ship a standard 40 ft container from Baltimore, Maryland, to 1 of about 50 countries around the world in the late 1990s.[11] In a regression of total freight charge on a land-locked country indicator, sea distance and land distance to destination, they find that the cost to ship a standard container 1000 km by sea is about \$190, while to ship it the same distance over land the cost is about \$1380. Recalling that a standard container can hold about 30 tons, we find that this gives sea rates of about half a cent per ton mile and land rates of about 5 cents per ton mile, so overland travel is about 10× as expensive as sea travel. These rates seem somewhat low compared with the price of US truck and rail rates reported in Glaeser and Kohlhase (2004) (28 cents per ton mile for trucks and 3 cents per ton mile for rail). Finally, Clark et al. (2004) find that the cost of shipping all maritime freight to and from the United States is equal to about 5.25% of the value of freight and that port efficiency is an important contributor to this cost.

These facts paint a subtle picture. While the real costs of moving goods have fallen to astonishingly low levels and the weight of trade is immense, the fact that not all

[10] Note that rail was at its peak in 1890 and airfreight was novel in 1950.

[11] To get a sense of the nature of the sample, when the countries are ranked by kilometers of paved road per person, the median country is Kenya. For evidence regarding the role of containerization in reducing international transportation costs, see Bernhofen et al. (2013).

trade travels by the cheapest mode and that most trade travels very short distances suggests that the decline in the price per ton of moving goods is not leading to the "death of distance."

While it is natural to think of time costs as being most important for the movement of people, the rise of air trade suggests that the time in transit is an increasingly important part of the cost of transit for goods. A back of the envelope calculation bolsters this idea. The capacity of a typical 40 ft container is about 30 tons. From Duranton et al. (2014), the value per ton of an average US domestic shipment of electrical appliances is about \$6000 per ton. Thus, a typical container of US electrical appliances can hold about \$200,000 worth of freight. From Glaeser and Kohlhase (2004), shipping this container 1000 miles by rail will cost about $1000 \times 30 \times 0.023 = \$700$. At a 5% annual rate, daily interest on 1 million dollar cargo is $200{,}000 \times 0.05/365 = \$28$, so on a 5-day journey, the opportunity cost of travel time is equal to about one-fifth of the freight charges. An average ton of goods is worth less than one-tenth of this, while a typical ton of computer equipment is $15\times$ as valuable. At least for relatively high value to weight products, time in transit is important.

Moreover, the predominance of short-haul trade suggests that not only are transportation costs important, but that the geography of production is influenced by transportation costs. For example, the development of nineteenth century Chicago was heavily influenced by its location relative to its surrounding agricultural hinterland, as discussed in Cronon (1991). This points to an important econometric problem in interpreting the transportation cost data presented so far: these data describe equilibrium transportation costs. Therefore, they do not isolate the supply-side production function (or cost function) for transportation, but are rather influenced by both demand and supply. Although these data on transportation costs are still suggestive, they capture both the cost of transportation (supply) and the endogenous organization of economic activity in space in response to the cost of transportation (demand). This presents important and difficult econometric problems, to which we return below.

Another striking feature of microdata on trade and production is the finding of Atalay et al. (2013) that most vertically integrated firms actually ship very little between plants. From the above, we have the puzzling collection of facts: the cost of moving goods is a small fraction of their value, most shipments occur over very small distances, most shipments do not travel by the cheapest mode, and the time cost of freight is probably important. One possible way of rationalizing this combination of findings is that there is something valuable about proximity other than the reduction in transportation costs—that is, agglomeration effects including knowledge spillovers and idea flows. In this case, trade could decline rapidly with distance even in a world in which transportation costs are small, because most economic activity is clustered together for these other reasons and hence most economic interactions are over short distances.

Alternatively, one could question whether transportation costs are really as small as a share of value added as some of the figures above suggest. Arguably, labor used in transportation should be compared with labor used in production and we should take into account the same kinds of costs that we think about for commuting: time costs and scheduling costs.

20.2.2 Household travel and commuting

While the trade literature has typically focused on the movement of goods, another important source of transportation costs in the urban literature is the movement of people. These costs of transporting people remain substantial, both in terms of the opportunity cost of time and in terms of the share of overall household expenditure. Table 20.3 lists average round-trip commuting times in minutes in a sample of countries and years for which data were readily available. While we should be concerned that differences in commuting times across countries reflect sampling error and differences in survey methods, with this caveat, these data suggest that the country mean round-trip commute is about 40 min in the 2000–2005 window, where we have the most observations. These times are fairly closely clustered, with a standard deviation of just less than 8 min. If the

Table 20.3 Average round-trip commuting times in minutes, by country and year

	1980–1984	1990–1994	2005–2009	2010–2012	
Austria			35		
Belgium			41		
Canada		54	59	63	52
Denmark			38		
Finland			41		
France			36		
Germany			42		
Greece			32		
Ireland			36		
Italy			30		
Luxembourg			35		
Netherlands			47		
Portugal			29		
Spain			38		
Sweden			38		
United Kingdom			37	53	
United States	43	45	51		50

Sources: Statistics Canada (2010), Turcotte (2005), Eurofound (2000), Office of National Statistics (2009), and Census Bureau (2009).

"work day" consists of 8 h at work and time in commuting, then commuting consumes about 7.5% of labor. Alternatively, if we value time in commuting at half the wage (as is common in the transportation economics literature; see Small and Verhoef, 2007) and suppose an 8-h work day, then the value of commuting time is equal to about 3.5% of the value of labor. While this is a large number, it understates the cost of household travel by restricting attention to commuters and commuting trips.

Alternatively, Schafer (2000) summarizes 26 national household travel surveys from countries all over the world. Averaging across these surveys, again with the caveat about the comparability of surveys, he finds that the daily household travel time is about 73 min, with a standard deviation of about 12 min. If we value this time at half the wage and again suppose an 8-h work day, then the value of time spent in household travel is about 8% of the value of labor.[12] If we take the labor share of GDP to be close to the current US level at 0.6, then the time cost of household travel is between 2.4% and 4.8% of GDP.

Table 20.4 describes the shares of household expenditure on transportation for 26 countries and several years. Again noting the possibility of different methods across countries, we find that the mean expenditure share is about 16.2% for the 2000–2004 window and about 14.6% for the 2005–2009 window, with standard deviations of 5.4% and 3.7%, respectively. Schafer (2000) investigates these shares using older and somewhat more extensive national accounts data and finds that across countries the average expenditure share for household travel is about 11%, with a standard deviation of about 3%. Weighting the household transportation share by 0.6, about the share of expenditure in current US GDP, and adding time costs, we have that the total costs of household travel are between 9% and 11.4% of GDP.

Two further points are made in Schafer (2000). First, for country-level aggregates, per capita travel time and expenditure share are negatively correlated. Second, for Zambia, only 5% of all trips are longer than 10 km, while for the United States, 5% of trips are longer than 50 km. To the extent that these findings are driven by differences in transportation technologies, they suggest that the transferral of developed-country transportation technologies to developing countries is likely to lead to substantial changes in the spatial organization of economic activity.

[12] We note that this estimate is problematic for at least two reasons. First, it assigns the time cost of an average worker to an average traveler, when many travelers are likely to have a lower value of time. Second, it assigns the time cost of an average worker to an average commuter, when wages probably vary systematically with commuting distance. With this said, on the basis of these surveys, a rough guess would be that the aggregate time cost of household travel is somewhere between 3.5% and 8% of the aggregate value of labor in an economy.

Table 20.4 Percentage of household expenditure on all modes of transportation by year and country

	1990–1994	1995–1999	2000–2004	2005–2009	2010–2012
Austria		15	14	16	
Belgium	10	10	13	12	
Bulgaria			23	12	
Canada				15	14
China (rural)	1	3	6	10	10
China (urban)	1	5	9		14
Croatia				13	
Cyprus			23	16	
Czech Republic			26	19	
Denmark		12	11	10	
Estonia			13	16	
Finland		12	14	13	
France	15	13	13	13	
Germany	11	13	12	13	
Greece	12	12	14	12	
Hungary			25	23	
Ireland		14	14	9	
Italy	14	15	15	12	
Latvia			15	19	
Lithuania			17	15	
Luxembourg	12	14	15	16	
Macedonia				16	
Malta			19	25	
Mexico					21
Netherlands	9	9	11	11	
Norway				13	
Poland			19	13	
Portugal	27	23	21	15	
Romania			17	12	
Slovakia			24	15	
Slovenia			26	21	
Spain	14	14	15	12	
Sweden		11	10	12	
Turkey				16	
EU27				13	
United Kingdom					14
United States				18	16

Sources: Statistics Canada (2012b), China Data Online (2011), Eurostat (2009), OECD Stat Extracts (2012), and Bureau of Labour Statistics (2012).

Table 20.5 Carbon dioxide emissions from transportation and shares by mode (2007)

Country	Total CO_2 equivalents (Mt)	Transport (%)	Rail (%)	Truck (%)	Ship (%)	Air (%)
Canada	561	32	1.32	24.34	1.69	3.14
Mexico	456	35	0.42	32.53	0.57	1.43
United Kingdom	529	29	0.87	25.81	2.38	7.24
United States	5829	33	0.89	27.72	0.95	2.96

Sources: Natural Resources Canada (2012a), North American Transportation Statistics (2012b), and Department for Transport (2012).

20.2.3 External costs

We have so far concerned ourselves with private costs of transportation, time, and private expense. We now turn our attention to two costs of transportation that are rarely priced, carbon emissions and congestion.

Table 20.5 presents total 2007 carbon dioxide equivalent (CO_2e) emissions for the transportation sector for Canada, Mexico, the United States, and the United Kingdom. Total emissions for the United States in 2007 were about 7000 Mt, so the transportation sector accounts for about 30% of US emissions. To the extent that these costs of transportation are not priced, the market allocation of resources to the transportation sector will be, in general, inefficient. With a social cost of about US $30 per ton CO_2e, the social cost of CO_2e emissions from transportation in the United States is about US $21 billion per year. This is only about one-tenth of 1% of US GDP. Thus, while greenhouse gas emissions from transportation are important in an absolute sense, they are small relative to the total cost of transportation.

Parry et al. (2007) provides a comprehensive survey of the externalities to automobile use, including local air pollution, global air pollution, traffic congestion, traffic accidents, and other externalities (such as noise and highway maintenance costs). Couture et al. (2012) estimate that a lower bound on the deadweight loss from traffic congestion in the United States is on the order of US $100 billion per year, although we note that these costs are already reflected in the transportation expenditure data described above.

20.3. THEORETICAL FRAMEWORK

In this section, we outline a multiregion extension of the Helpman (1998) model that follows Redding and Sturm (2008) and Redding (2012). The model incorporates many locations, goods transportation costs within and between locations, and commuting costs within locations. We use the model to show the effects of improvements in transportation infrastructure on the spatial distribution of wages, land rents, the population, and trade within and between locations. Although the model does not capture all of the theoretical foundations considered in the regional literature and urban literature, it captures many of the standard ingredients, and we use its predictions to structure our review of the empirical evidence below.[13]

20.3.1 Preferences and endowments

The economy consists of a set of locations indexed by n or $i \in N$, where n will typically refer to a consuming region and i to a producing region. To refer to a pairwise quantity, such as a distance or a quantity of trade, we use two subscripts, with the first indicating the location of consumption and the second the location of production. The economy is populated by a mass of representative consumers, $\overline{L}$, who are mobile across locations and are endowed with a single unit of labor that is supplied inelastically with zero disutility. The effective supply of labor for each location i depends on its population (L_i) and commuting technology (b_i), where commuting costs are assumed to take the iceberg form. For each unit of labor residing in location i, only a fraction b_i is available for production, where $0 < b_i < 1$, and the remaining fraction $1 - b_i$ is lost in commuting. While we treat b_i as a primitive of the model here, it could in principle depend on equilibrium population density (e.g., if higher population density increases congestion costs).

Preferences are defined over a consumption index of tradeable varieties, C_n, and consumption of a nontradeable amenity, H_n, which can be interpreted as housing. For simplicity, we treat the stock of housing as a primitive of the model, although it could also in principle depend on equilibrium population density (e.g., if a higher population density increases the supply of housing). The upper level utility function is assumed to be of Cobb–Douglas form[14]:

$$U_n = C_n^{\mu} H_n^{1-\mu}, \quad 0 < \mu < 1. \tag{20.1}$$

The tradeable goods consumption index takes the standard constant elasticity of substitution form:

$$C_n = \left(\sum_{i \in N} M_i c_{ni}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}},$$

[13] The model builds on the new economic geography literature synthesized in Fujita et al. (1999). While this literature assumes firm product differentiation and monopolistic competition, the model shares many properties with perfectly competitive models such as the model of Eaton and Kortum (2002) (see Redding, 2012) or the Armington model of product differentiation by location (see Allen and Arkolakis, 2013). The organization of economic activity within countries has recently received renewed attention, as in Cosar and Fajgelbaum (2013) and Ramondo et al. (2012).

[14] For empirical evidence using US data in support of the constant housing expenditure share implied by the Cobb–Douglas functional form, see Davis and Ortalo-Magne (2011).

where σ is the elasticity of substitution between varieties, and we assume that varieties are substitutes ($\sigma > 1$); c_{ni} denotes consumption in country n of a variety produced in country i; we have used the fact that all varieties produced in location i are consumed in location n in the same amount c_{ni}. Varieties are assumed to be subject to iceberg trade costs. In order for one unit of a variety produced in location i to arrive at location n, a quantity $d_{ni} > 1$ must be shipped, so $d_{ni} - 1$ measures proportional trade costs. The price index dual to the tradeables consumption index C_n is given by

$$P_n = \left(\sum_{i \in N} M_i p_{ni}^{1-\sigma} \right)^{1/(1-\sigma)}, \tag{20.2}$$

where we have used the fact that the measure M_i of varieties produced in location i faces the same elasticity of demand and charges the same equilibrium price $p_{ni} = d_{ni} p_i$ to consumers in location n.

Applying Shephard's lemma to the tradeables price index, we find that the equilibrium demand in location n for a tradeable variety produced in location i is

$$x_{ni} = p_i^{-\sigma} (d_{ni})^{1-\sigma} (\mu v_n L_n) (P_n)^{\sigma-1}, \tag{20.3}$$

where $v_n L_n$ denotes total income, which equals total expenditure, and, with Cobb–Douglas utility, consumers spend a constant share of their income, μ, on tradeables.

With constant expenditure shares and an inelastic supply of the nontradeable amenity, the equilibrium price of this amenity depends solely on the expenditure share, $(1-\mu)$, total income, $v_n L_n$, and the supply of the nontradeable amenity, $\bar{H}_n$:

$$r_n = \frac{(1-\mu) v_n L_n}{\bar{H}_n}. \tag{20.4}$$

Total income is the sum of labor income and expenditure on the nontradeable amenity, which is assumed to be redistributed lump-sum to the location's residents:

$$v_n L_n = w_n b_n L_n + (1-\mu) v_n L_n = \frac{w_n b_n L_n}{\mu}, \tag{20.5}$$

where we have used the fact that only a fraction b_n of the labor in location i is used in production because of commuting costs. Therefore, total labor income equals the wage per effective unit of labor (w_n) times the measure of effective units of labor ($b_n L_n$).

20.3.2 Production technology

There is a fixed cost in terms of labor of producing tradeable varieties ($F > 0$) and a constant variable cost that depends on a location's productivity (A_i). Both the fixed cost and the variable cost are the same across all varieties produced within a location. The total amount of labor (l_i) required to produce x_i units of a variety in location i is

$$l_i = F + \frac{x_i}{A_i}, \tag{20.6}$$

where we allow productivity (A_i) to vary across locations to capture variation in production fundamentals.

Profit maximization implies that equilibrium prices are a constant markup over marginal cost:

$$p_{ni} = \left(\frac{\sigma}{\sigma - 1}\right) \frac{d_{ni} w_i}{A_i}. \tag{20.7}$$

Combining profit maximization and zero profits, we find the equilibrium output of each tradeable variety equals the following constant:

$$\overline{x} = x_i = \sum_n x_{ni} = A_i F(\sigma - 1). \tag{20.8}$$

Labor market clearing for each location implies that labor demand equals the effective labor supply in that location, which is in turn determined by population mobility. Using the constant equilibrium output of each variety (20.8) and the tradeables production technology (20.6), we can write the labor market clearing condition as follows:

$$b_i L_i = M_i \bar{l}_i = M_i F \sigma, \tag{20.9}$$

where $\bar{l}_i$ denotes the constant equilibrium labor demand for each variety. This relationship pins down the measure of tradeable varieties produced in each location as a function of the location's population, the commuting technology, and the parameters of the model.

20.3.3 Market access and wages

Given demand in all markets and trade costs, the free-on-board price (p_i) charged for a tradeable variety by a firm in each location must be low enough in order to sell the quantity $\overline{x}$ and cover the firm's fixed production costs. We saw above that prices are a constant markup over marginal cost. Therefore, given demand in all markets, the equilibrium wage in location i, w_i, must be sufficiently low in order for a firm to sell $\overline{x}$ and cover its fixed production costs. Using demand (20.3), profit maximization (20.7), and equilibrium output (20.8), we obtain the *tradeables wage equation*:

$$\left(\frac{\sigma}{\sigma - 1} \frac{w_i}{A_i}\right)^{\sigma} = \frac{1}{\overline{x}} \sum_{n \in N} (w_n b_n L_n)(P_n)^{\sigma - 1}(d_{ni})^{1-\sigma}. \tag{20.10}$$

This relationship pins down the maximum wage that a firm in location i can afford to pay given demand in all markets, trade costs, and the production technology. On the right-hand side of the equation, market n demand for tradeables produced in location i depends on the total expenditure on tradeable varieties, $\mu v_n L_n = w_n b_n L_n$, the tradeables price index, P_n, which summarizes the price of competing varieties, and bilateral trade costs, d_{ni}. Total demand for tradeables produced in location i is the weighted sum of the demand in all markets, where the weights are these bilateral trade costs, d_{ni}.

Following Redding and Venables (2004), we define the weighted sum of market demands faced by firms as *firm market access*, fma_i, such that the tradeables wage equation can be written more compactly as

$$w_i = \xi A_i^{\frac{\sigma-1}{\sigma}} [\text{fma}_i]^{1/\sigma}, \qquad \text{fma}_i \equiv \sum_{n \in N} (w_n b_n L_n)(P_n)^{\sigma-1}(d_{ni})^{1-\sigma}, \tag{20.11}$$

where $\xi \equiv (F(\sigma-1))^{-1/\sigma}(\sigma-1)/\sigma$ collects together earlier constants. Therefore, wages increase with both productivity A_i and firm market access (fma_i). Investments in transportation infrastructure that reduce the costs of transporting goods (d_{ni}) to markets ($(w_n b_n L_n)(P_n)^{\sigma-1}$) raise market access and wages. Improvements in the commuting technology (b_n) increase the effective supply of labor ($b_n L_n$) and hence total income, which also raises market access and wages.

20.3.4 Labor market equilibrium

With perfect population mobility, workers move across locations to arbitrage away real income differences. Real income in each location depends on per capita income (v_n), the price index for tradeables (P_n), and the price of the nontradeable amenity (r_n). Therefore, population mobility implies

$$V_n = \frac{v_n}{(P_n)^{\mu}(r_n)^{1-\mu}} = \overline{V}, \tag{20.12}$$

for all locations that are populated in equilibrium, where we have collected the constants $\mu^{-\mu}$ and $(1-\mu)^{-(1-\mu)}$ into the definition of V_n and $\overline{V}$.

The price index (20.2) that enters the above expression for real income depends on consumers' access to tradeable varieties, as captured by the measure of varieties and their free-on-board prices in each location i, together with the trade costs of shipping the varieties from location i to location n. We summarize consumers' access to tradeables using the concept of *consumer market access*, cma_n:

$$P_n = (\text{cma}_n)^{1/(1-\sigma)}, \qquad \text{cma}_n \equiv \sum_{i \in N} M_i (p_i d_{ni})^{1-\sigma}. \tag{20.13}$$

Substituting for v_n, P_n, and r_n, we can rewrite the labor mobility condition (20.12) to yield an expression linking the equilibrium population of a location (L_n) to its productivity

(A_n), its commuting technology (b_n), the supply of the nontraded amenity (H_n), and the two endogenous measures of market access introduced above (one for firms (fma_n) and one for consumers (cma_n)):

$$L_n = \chi b_n^{\frac{\mu}{1-\mu}} A_n^{\frac{\mu(\sigma-1)}{\sigma(1-\mu)}} \bar{H}_n (\text{fma}_n)^{\frac{\mu}{\sigma(1-\mu)}} (\text{cma}_n)^{\frac{\mu}{(1-\mu)(\sigma-1)}}, \tag{20.14}$$

where $\chi = \overline{V}^{-1/(1-\mu)} \xi^{\mu/(1-\mu)} \mu^{-\mu/(1-\mu)} (1-\mu)^{-1}$ is a function of the common real income $\overline{V}$.

Therefore, the equilibrium population (L_n) increases with the quality of the commuting technology (b_n), the productivity of the final goods production technology (A_n), and the supply of the nontraded amenity $(\bar{H}_n)$. Investments in transportation infrastructure that reduce the costs of transporting goods (d_{ni}) raise both firm market access and consumer market access $(\text{fma}_n$ and $\text{cma}_n)$ and hence increase the equilibrium population. Improvements in the commuting technology (b_n) also have positive indirect effects on the equilibrium population through higher firm and consumer market access.

From land market clearing (20.4) and total labor income (20.5), land prices can be written in terms of wages and the total population:

$$r_n = \frac{(1-\mu)}{\mu} \frac{w_n b_n L_n}{\bar{H}_n}. \tag{20.15}$$

Therefore, higher firm market access (fma_n) raises land prices through both higher wages (from (20.10)) and higher population (from (20.14)), while higher consumer market access (cma_n) raises land prices through a higher population alone (from (20.14)). Reductions in the cost of transporting goods (d_{ni}) raise land prices through both firm market access and consumer market access. Improvements in commuting technology (b_n) raise land prices directly and also indirectly through higher wages and a higher population.

20.3.5 Trade flows

Using constant elasticity of substitution demand, we can express the share of location n's expenditure on varieties produced in location i as

$$\pi_{ni} = \frac{M_i p_{ni}^{1-\sigma}}{\sum_{k\in N} M_k p_{nk}^{1-\sigma}}, \tag{20.16}$$

which, with use of the equilibrium pricing rule (20.7) and the labor market clearing condition for each location (20.9), can be written as

$$\pi_{ni} = \frac{b_i L_i (d_{ni} w_i)^{1-\sigma} (A_i)^{\sigma-1}}{\sum_{k\in N} b_k L_k (d_{nk} w_k)^{1-\sigma} (A_k)^{\sigma-1}}. \tag{20.17}$$

This expression for bilateral trade shares (π_{ni}) corresponds to a "gravity equation," in which bilateral trade between exporter i and importer n depends on both "bilateral resistance" (i.e., the bilateral goods of trading goods between exporter i and importer n (d_{ni}) in the numerator) and "multilateral resistance" (i.e., the bilateral costs for importer n of sourcing goods from all exporters k (d_{nk}) in the denominator). In this gravity equation specification, bilateral trade depends on characteristics of the exporter i (e.g., the exporter's wage w_i in the numerator), bilateral trade costs (d_{ni}), and characteristics of the importer n (i.e., the importer's access to all sources of supply in the denominator).[15]

Taking the ratio of these expenditure shares, we find the value of trade between locations (X_{ni}) relative to trade within locations (X_{nn}) is

$$\frac{X_{ni}}{X_{nn}}=\frac{\pi_{ni}}{\pi_{nn}}=\frac{b_i L_i (d_{ni} w_i)^{1-\sigma}(A_i)^{\sigma-1}}{b_n L_n (d_{nn} w_n)^{1-\sigma}(A_n)^{\sigma-1}}. \tag{20.18}$$

Therefore, transportation infrastructure improvements that reduce the cost of transporting goods within locations (d_{nn}) by the same proportion as they reduce the cost of transporting goods between locations (d_{ni}) leave the ratio of trade between locations to trade within locations unchanged. One potential example is building roads within cities that make it easier for goods to circulate within the city and to leave the city to connect with long-distance highways. Transportation cost improvements that reduce commuting costs for all locations (increase b_n and b_i) also leave the ratio of trade between locations to trade within locations unchanged.

In this model with a single differentiated sector, all trade takes the form of intraindustry trade, and transportation infrastructure improvements affect the volume of this intraindustry trade. More generally, in a setting with multiple differentiated sectors that differ in terms of the magnitude of trade costs (e.g., high value to weight versus low value to weight sectors), transportation infrastructure improvements also affect the pattern of interindustry trade and the composition of employment and production across sectors within locations.

20.3.6 Welfare

We now show how the structure of the model can be used to derive an expression for the welfare effects of transportation infrastructure improvements in terms of observables. Using the trade share (20.16), we can rewrite the price index (20.2) in terms of each location's trade share with itself and other parameters:

[15] For an insightful review of the gravity equation in the international trade literature, see Head and Mayer (2013).

$$P_n = \frac{\sigma}{\sigma - 1}\left(\frac{b_n L_n}{\sigma F \pi_{nn}}\right)^{\frac{1}{1-\sigma}} \frac{d_{nn} w_n}{A_n}. \tag{20.19}$$

From this expression for the price index and land market clearing (20.15), the population mobility condition (20.12) implies that the equilibrium population for each location can be written as

$$L_n = \left(\frac{\left(\frac{1}{\sigma F \pi_{nn}}\right)^{\frac{\mu}{\sigma-1}} \bar{H}_n^{1-\mu} b_n^{\frac{\mu\sigma}{\sigma-1}} A_n^{\mu}}{\mu\left(\frac{1-\mu}{\mu}\right)^{1-\mu}\left(\frac{\sigma}{\sigma-1}\right)^{\mu} \overline{V} d_{nn}^{\mu}}\right)^{\frac{\sigma-1}{\sigma(1-\mu)-1}}, \tag{20.20}$$

where terms in wages (w_n) have canceled and labor market clearing for the economy as a whole implies

$$\sum_{n \in N} L_n = \overline{L}. \tag{20.21}$$

This expression for the equilibrium population (20.20) has an intuitive interpretation. The population of each location n decreases with its domestic trade share (π_{nn}), since locations with low domestic trade shares have low consumption goods price indices, which increases their attractiveness to residents. The population of each location increases with the efficiency of its commuting technology (b_n), its productivity in production (A_n), its supply of housing ($\bar{H}_n$), and its transportation technology (inversely related to d_{nn}). The common level of utility across all locations ($\overline{V}$) is endogenous and is determined by the requirement that the labor market clears for the economy as a whole.

Rearranging the population mobility condition (20.20), we can write the real income in each location in terms of its population, trade share with itself, and other parameters.

$$V_n = \frac{\left(\frac{1}{\sigma F \pi_{nn}}\right)^{\frac{\mu}{\sigma-1}} L_n^{-\left(\frac{\sigma(1-\mu)-1}{\sigma-1}\right)} \bar{H}_n^{1-\mu} b_n^{\frac{\mu\sigma}{\sigma-1}} A_n^{\mu}}{\mu\left(\frac{1-\mu}{\mu}\right)^{1-\mu}\left(\frac{\sigma}{\sigma-1}\right)^{\mu} d_{nn}^{\mu}} = \overline{V}. \tag{20.22}$$

A key implication of this expression for real income is that the change in each location's trade share with itself and the change in its population are sufficient statistics for the welfare effects of improvements in transportation technology that reduce the costs of trading goods (see Redding, 2012):

$$\frac{V_n^1}{V_n^0} = \left(\frac{\pi_{nn}^0}{\pi_{nn}^1}\right)^{\frac{\mu}{\sigma-1}}\left(\frac{L_n^0}{L_n^1}\right)^{\left(\frac{\sigma(1-\mu)-1}{\sigma-1}\right)} = \frac{\overline{V}^1}{\overline{V}^0}, \tag{20.23}$$

where the superscripts 0 and 1 denote the value of variables before and after the improvement in transportation technology, respectively.

Similar sufficient statistics apply for the welfare effects of improvements in transportation technology that reduce commuting costs, although these welfare effects also depend directly on the change in commuting costs (through the resulting increase in the effective supply of labor):

$$\frac{V_n^1}{V_n^0} = \left(\frac{b_n^1}{b_n^0}\right)^{\frac{\mu\sigma}{\sigma-1}} \left(\frac{\pi_{nn}^0}{\pi_{nn}^1}\right)^{\frac{\mu}{\sigma-1}} \left(\frac{L_n^0}{L_n^1}\right)^{\left(\frac{\sigma(1-\mu)-1}{\sigma-1}\right)} = \frac{\overline{V}^1}{\overline{V}^0}. \tag{20.24}$$

While these improvements in transportation infrastructure have uneven effects on wages, land prices, and the population, the mobility of workers across locations ensures that they have the same effect on welfare across all populated locations.

To understand the relationship between changes in domestic trade shares and the welfare change from improvements in transportation technology that reduce goods trade costs, consider the extreme case where the transportation improvement allows goods trade between two previously autarkic locations. For locations closed to goods trade, domestic trade shares must equal 1. Once locations open to trade, they can specialize to exploit gains from trade with other locations, and domestic trade shares fall below 1. This fall in the domestic trade shares reflects the increase in specialization and is directly related to increases in real income, our measure of welfare.

To understand the relationship between changes in population and the changes in welfare following improvements in transportation technology that reduce goods trade costs, first note that labor mobility requires real wage equalization across populated locations. Therefore, if goods trade is opened between locations, and some locations (e.g., coastal regions) benefit more than other locations (e.g., interior regions) at the initial labor allocation, workers must relocate to arbitrage away real wage differences. Those locations that experience larger welfare gains from trade at the initial labor allocation will experience population inflows, which increases the demand for the immobile factor land and causes land prices to rise. In contrast, those locations that experience smaller welfare gains from trade at the initial labor allocation will experience population outflows, which decreases the demand for land and reduces land prices. This population reallocation continues until real wages are again equalized across all populated locations. Hence, these population changes also need to be taken into account in computing the welfare effects of the improvement in transportation technology.

Therefore, together, the change in a location's domestic trade share and the change in its population are sufficient statistics for the effects of a transportation improvement that reduces the costs of trading goods (d_{ni}). A transportation improvement that reduces the commuting costs for a region (b_n) also directly increases the supply of labor for that region, which is taken into account in the welfare formula.

20.3.7 General equilibrium

The general equilibrium of the model can be represented by the share of workers in each location ($\lambda_n = L_n/\bar{L}$), the share of each location's expenditure on goods produced by other locations (π_{ni}), and the wage in each location (w_n). Using labor income (20.5), the trade share (20.16), population mobility (20.20), and labor market clearing (20.21), the equilibrium triple $\{\lambda_n, \pi_{ni}, w_n\}$ solves the following system of equations for all $i, n \in N$ (see Redding, 2012):

$$w_i b_i \lambda_i = \sum_{n \in N} \pi_{ni} w_n b_n \lambda_n, \tag{20.25}$$

$$\pi_{ni} = \frac{b_i \lambda_i (d_{ni} w_i / A_i)^{1-\sigma}}{\sum_{k \in N} b_k \lambda_k (d_{nk} w_k / A_k)^{1-\sigma}}, \tag{20.26}$$

$$\lambda_n = \frac{\left[\bar{H}_n^{1-\mu}\left(\frac{1}{\pi_{nn}}\right)^{\frac{\mu}{\sigma-1}} b_n^{\frac{\mu\sigma}{\sigma-1}} A_n^{\mu} d_{nn}^{-\mu}\right]^{\frac{\sigma-1}{\sigma(1-\mu)-1}}}{\sum_{k \in N}\left[\bar{H}_k^{1-\mu}\left(\frac{1}{\pi_{kk}}\right)^{\frac{\mu}{\sigma-1}} b_k^{\frac{\mu\sigma}{\sigma-1}} A_k^{\mu} d_{kk}^{-\mu}\right]^{\frac{\sigma-1}{\sigma(1-\mu)-1}}}. \tag{20.27}$$

The assumption that $\sigma(1-\mu) > 1$ corresponds to the "no black hole" condition in Krugman (1991) and Helpman (1998). For parameter values satisfying this inequality, the model's agglomeration forces from love of variety, increasing returns to scale, and transportation costs (which are inversely related to σ) are not too strong relative to its congestion forces from an inelastic supply of land (captured by $1-\mu$). As a result, each location's real income monotonically decreases with its population, which ensures the existence of a unique stable nondegenerate distribution of the population across locations.

While the existence of a unique equilibrium ensures that the model remains tractable and amenable to counterfactual analysis, often the rationale for transportation investments is cast in terms of shifting the distribution of economic activity between multiple equilibria. To the extent that such multiple equilibria exist, their analysis requires either consideration of the range of the parameter space for which the model has multiple equilibria or the use of a richer theoretical framework.[16]

[16] A body of empirical literature has examined whether large and temporary shocks have permanent effects on the location of economic activity and has interpreted these permanent effects as either evidence of multiple equilibria or path dependence more broadly. See, for example, Bleakley and Lin (2012), Davis and Weinstein (2002), Maystadt and Duranton (2014), Redding et al. (2011), and Sarvimäki et al. (2010).

20.3.8 Counterfactuals

The system of equations for general equilibrium (20.25)–(20.27) can be used to undertake model-based counterfactuals in an extension of the trade-based approach of Dekle et al. (2007) to incorporate factor mobility across locations. The system of equations for general equilibrium must hold both before and after any counterfactual change in, for example, transportation infrastructure. Denote the value of variables in the counterfactual equilibrium with a prime (x') and the relative value of variables in the counterfactual and initial equilibria by a hat ($\hat{x} = x'/x$). Using this notation, we can rewrite the system of equations for the counterfactual equilibrium (20.25)–(20.27) as follows:

$$\hat{w}_i \hat{b}_i \hat{\lambda}_i Y_i = \sum_{n \in N} \hat{\pi}_{ni} \pi_{ni} \hat{w}_n \hat{b}_n \hat{\lambda}_n Y_n, \tag{20.28}$$

$$\hat{\pi}_{ni} \pi_{ni} = \frac{\pi_{ni} \hat{\lambda}_i \hat{b}_i \left(\hat{d}_{ni} \hat{w}_i / \hat{A}_i \right)^{1-\sigma}}{\sum_{k \in N} \pi_{nk} \hat{\lambda}_k \hat{b}_k \left(\hat{d}_{nk} \hat{w}_k / \hat{A}_k \right)^{1-\sigma}}, \tag{20.29}$$

$$\hat{\lambda}_n \lambda_n = \frac{\lambda_n \left(\hat{\bar{H}}^{1-\mu} \hat{\pi}_{nn}^{-\frac{\mu}{\sigma-1}} \hat{b}_n^{\frac{\mu\sigma}{\sigma-1}} \hat{A}_n^{\mu} \hat{d}_{nn}^{-\mu} \right)^{\frac{\sigma-1}{\sigma(1-\mu)-1}}}{\sum_{k \in N} \lambda_k \left(\hat{\bar{H}}_k^{1-\mu} \hat{\pi}_{kk}^{-\frac{\mu}{\sigma-1}} \hat{b}_k^{\frac{\mu\sigma}{\sigma-1}} \hat{A}_k^{\mu} \hat{d}_{kk}^{-\mu} \right)^{\frac{\sigma-1}{\sigma(1-\mu)-1}}}, \tag{20.30}$$

where $Y_i = w_i b_i L_i$ denotes labor income in the initial equilibrium.

Given an exogenous change in transportation infrastructure that affects the costs of trading goods ($\hat{d}_{ni}$) or the costs of commuting ($\hat{b}_n$), this system of equations (20.28)–(20.30) can be solved for the counterfactual changes in wages ($\hat{w}_n$), population shares ($\hat{\lambda}_n$), and trade shares ($\hat{\pi}_{ni}$). Implementing these counterfactuals requires only observed values of GDP, trade shares, and population shares $\{Y_n, \pi_{ni}, \lambda_n\}$ for all locations $i,n \in N$ in the initial equilibrium. For parameter values for which the model has a unique stable equilibrium ($\sigma(1-\mu) > 1$), these counterfactuals yield determinate predictions for the impact of the change in transportation costs. From the welfare analysis above, the changes in each location's population and its domestic trade share provide sufficient statistics for the welfare effect of transportation improvements that affect the costs of trading goods ($\hat{d}_{ni}$). In contrast, transportation improvements that affect the costs of commuting ($\hat{b}_n$) also have direct effects on welfare in addition to their effects through the population and domestic trade shares. With perfect population mobility, these welfare effects must be the same across all populated locations.

20.4. REDUCED-FORM ECONOMETRIC FRAMEWORK

20.4.1 A simple taxonomy

We survey the recent empirical literature investigating the effects of infrastructure on the geographic distribution of economic activity. The preponderance of this literature can be described with a remarkably simple taxonomy.

Let t index time periods, and, preserving the notation from above, let n and $i \in N$ index a set of geographic locations, typically cities or counties. Let L_{it} denote an outcome of interest for location i at time t: employment, population, rent, or centralization. Let x_{it} be a vector of location- and time-specific covariates, and finally, let b_{it} and d_{it} denote the transportation variables of interest. In particular, consistent with the notation in our theoretical model, let b_{it} denote a measure of transportation infrastructure that is internal to unit i, and let d_{it} denote a measure of transportation infrastructure external to unit i. For example, b_{it} could count radial highways within a metropolitan area, while d_{it} could indicate whether a rural county is connected to a highway network.

With this notation in place, define the "intracity regression" as

$$L_{it} = C_0 + C_1 b_{it} + C_2 x_{it} + \delta_i + \theta_t + \epsilon_{it}, \tag{20.31}$$

where δ_i denotes location-specific time-invariant unobservables, θ_t denotes a common time effect for all locations, and ϵ_{it} denotes the time-varying location-specific residual. The coefficient of interest is C_1, which measures the effect of within-city infrastructure on the city-level outcome.[17]

Similarly, define the "intercity regression" as

$$L_{it} = C_0 + C_1 d_{it} + C_2 x_{it} + \delta_i + \theta_t + \epsilon_{it}, \tag{20.32}$$

which differs from the intracity regression only in that the explanatory variable of interest describes transportation costs between unit i and other units, rather than within-city infrastructure.

These equations require some discussion before we turn to a description of the results. First, both estimating equations are natural reduced-form versions of Equation (20.14) or, if the outcome of interest is land rent, Equation (20.15). Thus, they are broadly consistent with the theoretical framework described earlier. Second, comparing the regression equations with their theoretical counterparts immediately suggests four inference problems that estimations of the intracity and intercity regressions should confront.

First, equilibrium employment or land rent depends on the location-specific productivity, A_n. This will generally be unobserved and thus will be reflected in the error terms of our regression equations. It is natural to expect that intracity and intercity infrastructure will depend on location-specific productivity and, hence, be endogenous in the two

[17] Moses (1958) and Moses and Williamson (1963) are pioneering studies on the role of automobiles and highways in reorganizing the distributions of population and economic activity within metropolitan areas.

regression equations. Second, equilibrium employment or land rent depends on the level of a location-specific amenity, H_n. In our model, this reflects a supply of housing, but in reality, it may also reflect unobserved location characteristics that augment or reduce the welfare of residents at a location. We might also be concerned that such amenities, to the extent that they are unobserved, affect infrastructure allocation and give rise to an endogeneity problem. More generally, the intercity and intracity regressions do not by themselves distinguish between the demand for and supply of transportation.

Third, Equations (20.14) and (20.15) involve expressions for market access not present explicitly in the estimating equations. To the extent that market access depends on transportation costs between cities, the treatment of market access in these estimations deserves careful attention. Fourth, to the extent that there are general equilibrium effects of transportation infrastructure on all locations, these are not captured by C_1. Instead, they are captured in the time effects θ_t and cannot be separated from other time-varying factors that are common to all locations without further assumptions. More generally, in general equilibrium, transportation investments between a pair of regions i and j can have effects on third regions k, which are not captured by the transportation variables for regions i and j.

20.4.2 Identification of causal effects

As discussed above, perhaps the biggest empirical challenge in estimating the intercity and intracity regressions is constructing the appropriate counterfactual for the absence of the transportation improvement. In particular, ordinary least squares (OLS) regressions comparing treated and untreated locations are unlikely to consistently estimate the causal effect of the transportation improvement, because the selection of locations into the treatment group is nonrandom. The main empirical approach to addressing this challenge has been to develop instruments for the assignment of transportation improvements that plausibly satisfy the exclusion restriction of affecting the economic outcome of interest only through the transportation improvement.[18] More formally, this approach to identifying the causal effects posits an additional first-stage regression that determines the assignment of transportation infrastructure:

$$\mathbb{I}_{it} = D_0 + D_1 x_{it} + D_2 z_{it} + \eta_i + \gamma_t + u_{it}, \tag{20.33}$$

where $\mathbb{I}_{it} \in \{b_{it}, d_{it}\}$ is the transportation variable of interest (depending on whether the specification is intracity or intercity), x_{it} are the location- and time-varying controls from the second-stage regression ((20.31) or (20.32)), η_i are location-specific time-invariant

[18] While the program evaluation literature suggests other complementary approaches, such as conducting randomized experiments with transportation improvements or the use of matching estimators, these have been less widely applied in this empirical literature.

unobservables, γ_t are time indicators, u_{it} is a time-varying location-specific residual, and z_{it} are the instruments or excluded exogenous variables.

Combining the second-stage equation ((20.31) or (20.32)) with the first-stage equation (20.33), we can estimate the impact of transportation infrastructure on the economic outcomes of interest (C_1) using two-stage least squares. Credible identification of the causal impact of transportation infrastructure requires that two conditions are satisfied: (1) the instruments have power in the first-stage regression ($D_2 \neq 0$) and (2) the instruments satisfy the exclusion restriction of affecting the economic outcomes of interest only through transportation infrastructure conditional on the controls x_{it}—that is, $\text{cov}(\epsilon_{it}, u_{it}) = 0$.

The existing literature has followed three main instrumental variables strategies. The first, the *planned route instrumental variable approach*, is an instrumental variables strategy which relies on planning maps and documents as a source of quasi-random variation in the observed infrastructure. The second, the *historical route instrumental variable approach*, relies on very old transportation routes as a source of quasi-random variation in observed infrastructure. The third, the *inconsequential place approach*, relies on choosing a sample that is inconsequential in the sense that unobservable attributes do not affect the placement of infrastructure. The plausibility of these identification strategies depends sensitively on the details of their implementation and is sometimes contentious. With this said, we briefly describe these identification strategies and the rationale for their use. We avoid discussion of the validity of these strategies in particular contexts. Broadly, the strategies we describe are the best approaches currently available for estimating the causal effects of transportation infrastructure on the organization of economic activity.

20.4.2.1 Planned route instrumental variable approach

Baum-Snow (2007) pioneers the planned route instrumental variable approach by using a circa 1947 plan for the interstate highway network as a source of quasi-random variation in the way the actual network was developed. In the specific context of Baum-Snow (2007), this means counting the number of planned radial highways entering a metropolitan area and using this variable to predict the actual number of interstate highway rays. Since the network plan was developed under a mandate to serve military purposes, the validity of this instrument hinges on the extent to which military purposes are orthogonal to the needs of postwar commuters. Several other empirical investigations into the effects of the US road and highway network exploit instruments based on the 1947 highway plan, while Hsu and Zhang (2012) develop a similar instrument for Japan. Michaels et al. (2012) use an even earlier plan of the US highway network, the "Pershing plan," as a source of quasi-random variation in the US highway network. Although Donaldson (2015) stops short of using hypothetical planned networks as instruments for realized networks, he does compare the development of districts without railroads and without planned railroads with those without railroads but with planned railroads.

That these sets of districts develop in the same way suggests that the planning process did not pick out districts on the basis of different unobservable characteristics.

20.4.2.2 Historical route instrumental variable approach

Duranton and Turner (2012) develop the historical route instrumental variable approach. In regressions predicting metropolitan statistical area (MSA)-level economic outcomes they rely on maps of historical transportation networks, the US railroad network circa 1898, and the routes of major expeditions of exploration of the United States between 1535 and 1850 as sources of quasi-random variation in the US interstate highway network at the end of the twentieth century. The validity of these instruments requires that, conditional on the controls, factors that do not directly affect economic activity in US metropolitan areas at the end of the twentieth century determine the configuration of these historical networks. A series of articles (Duranton and Turner, 2011, 2012; Duranton et al., 2014) use the two historical route instruments and the 1947 highway plan as sources of quasi-random variation in regressions predicting metropolitan total vehicle kilometers traveled, changes in metropolitan employment, and trade flows between cities as functions of the interstate highway network.

One distinctive feature of Duranton and Turner (2011, 2012) and Duranton et al. (2014) is the use of multiple instruments based on different sources of variation. With more instruments than endogenous variables, the specification can be estimated with either all or subsets of the instruments, and overidentification tests can be used as a check on the identifying assumptions. Conditional on one of the instruments being valid, these overidentification tests check the validity of the other instruments. Given that the instruments exploit quite different sources of variation in the data, if a specification passes the overidentification test, this implies that either all of the instruments are valid or an improbable relationship exists between the instruments and the errors of the first-stage and second-stage regressions.

Several other authors develop historical transportation networks as a source of quasi-random variation in modern transportation networks in other regions. Baum-Snow et al. (2012) rely on Chinese road and rail networks from 1962 as a source of quasi-random variation in road and rail networks after 2000. Garcia-Lopez et al. (2013) use eighteenth century postal routes and Roman roads for Spain. Hsu and Zhang (2012) rely on historical Japanese railroad networks. Martincus et al. (2012) use the Inca roads for Peru. Duranton and Turner (2012) provide a more detailed discussion of the validity of these instruments.

20.4.2.3 Inconsequential units approach

To estimate the intercity regression, researchers often rely on the inconsequential units approach to identification, sometimes in conjunction with one or both of the instrumental variables strategies described above. If we consider economically small units lying between large cities, then we expect that intercity links will traverse these units only when they lie along a convenient route between the two large cities. That is, we expect that the unobserved characteristics of units between large cities are inconsequential to the choice of route, and therefore that the connection status of these units will not depend on the extent to which these units are affected by the road. Chandra and Thompson (2000) pioneer this strategy in their analysis of the effect of access to the interstate highway system on rural counties in the United States. By restricting attention to rural highways, they hope to restrict attention to counties that received interstate highways "accidentally," by virtue of lying between larger cities. While it is difficult to assess the validity of this approach, some of the regressions reported in Michaels (2008) are quite similar to those in Chandra and Thompson (2000) but rely on the 1947 planned highway network for identification. That the two methods arrive at similar estimates is reassuring. Banerjee et al. (2012) also use the inconsequential units strategy in their analysis of the effects of Chinese transportation networks. In particular, they construct a hypothetical transportation network connecting historical treaty ports to major interior trading centers. Counties near these predicted networks are there accidentally in the same sense that rural counties may be accidentally near interstate highways in the United States. Similarly, and also for China, Faber (2015) constructs a hypothetical least-cost network connecting major Chinese cities and examines the impact of proximity to this network on outcomes in nearby rural counties.

These three econometric responses to the probable endogeneity of transportation infrastructure are widely used. Other approaches to this problem typically exploit natural experiments that, while they may provide credible quasi-random variation in infrastructure, are not easily extended to other applications.

20.4.3 Distinguishing growth from reorganization

As Fogel (1964) observes in his classic analysis of the role of railroad construction in the economic development of the nineteenth century United States, an assessment of the economic impacts of transportation infrastructure depends fundamentally on whether changes in transportation costs affect the amount of economic activity or reorganize existing economic activity. For example, the welfare implications of a road or light rail line that attracts preexisting firms are quite different from those of one that leads to the creation of new firms. Importantly, this issue is distinct from the endogeneity problem discussed above. The problem of endogeneity follows from nonrandom assignment of transportation infrastructure to "treated" observations. The problem of distinguishing between growth and reorganization persists even when transportation is assigned to observations at random. Even in the case in which a region experiences an exogenous change in transportation infrastructure, the observed effects on economic activity in the region can reflect either reorganization or growth. This same issue of distinguishing growth and reorganization appears in the literature evaluating place-based policies, as discussed in Neumark and Simpson (2014) in this volume.[19]

Figure 20.6 illustrates a simple hypothetical dataset with the same structure as that typically used to estimate the intercity and intracity estimating equations. Figure 20.6 describes a sample consisting of three regions: a region that is "treated" in some way that affects transportation costs in this region—for example, a new road; an untreated region which is typically near the treated region but is not subject to a change in transportation infrastructure; and everyplace else. The outcome variable of interest is y and the new road creates a units of this outcome in the treated region and displaces d units from the untreated region to the treated region.

Fundamentally, the intercity and intracity regressions estimate the effect of treatment on the difference between treated regions and untreated comparison regions. As the figure makes clear, the difference in the outcome between treated and untreated regions is $2d + a$, the compound effect of reorganization and growth. At its core, the problem of distinguishing between reorganization and growth requires us to identify two quantities. Without further assumptions, these two quantities cannot be separately identified if we estimate only a single equation, regardless of whether it is the intercity or intracity estimating equation. To identify both the growth and the reorganization effect, we must estimate two linearly independent equations.

In the context of the sample described in Figure 20.6, these two equations could involve a comparison of any two of the three possible pairs of regions—that is, treated

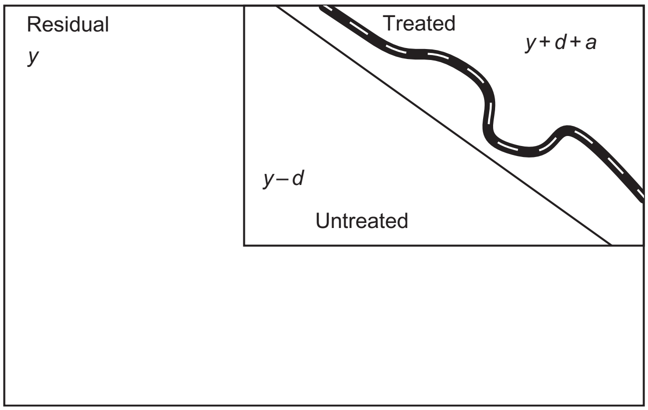


Figure 20.6 A simple hypothetical sample.

[19] For approaches to distinguishing growth and reorganization in this literature on place-based policies, see Criscuolo et al. (2012) and Mayer et al. (2013).

and untreated, untreated and residual, and treated and residual. Alternatively, with panel data, one could estimate the change in the treated region following the change in transportation costs and also the change in the untreated region following the change in the treated region. While the literature has carefully addressed the possibility that transportation costs and infrastructure are not assigned to regions at random, few authors conduct estimations allowing the separate identification of growth and reorganization.

While Figure 20.6 suggests simple methods for distinguishing between growth and reorganization, this reflects implicit simplifying assumptions. In particular, the new road in the treated district does not lead to migration of economic activity from the residual region to the untreated or the treated region and does not cause growth in the untreated or residual region. If we allow these effects, then the effect of a new road in the treated region is characterized by six parameters rather than two. Identifying all of these parameters will generally require estimating six linearly independent equations and will not generally be possible with cross-sectional data. In the context of "real data," with a more complex geography and many regions subject to treatment, distinguishing between growth and reorganization requires *a priori* restrictions on the nature of these effects.

The literature has, as yet, devoted little attention to what these identifying assumptions should be. As suggested by Figure 20.6, this problem can be resolved with transparent but ad hoc assumptions. Alternatively, the theoretical model described in Section 20.3 provides a theoretically founded basis for distinguishing between growth and reorganization which derives from the iceberg structure of transportation costs and assumptions about demand and production. Importantly, if the new road in the treated region affects the level of economic activity in all three regions, then no cross-sectional estimate can recover this effect. This requires time series data or cross-sectional data describing "replications" of Figure 20.6. More generally, for a penetration road or single transport project, it may be possible to construct plausible definitions of treated, untreated, and residual regions, as in Figure 20.6. However, for an evaluation of a national highway system, there may be no plausible residual regions, in which case we are necessarily in a general equilibrium world.

20.5. REDUCED-FORM EMPIRICAL RESULTS

20.5.1 Intracity infrastructure and the geographic organization of economic activity

20.5.1.1 Infrastructure and decentralization

Baum-Snow (2007) partitions a sample of US metropolitan areas into an "old central business district," the central business district circa 1950, and the residual suburbs. He then estimates a version of the intracity regression, Equation (20.31), in first differences, where the unit of observation is a US MSA, the measure of infrastructure is the count of radial interstate highways, and the instrument is a measure of rays based on the 1947 highway plan discussed above. He finds that each radial segment of the interstate highway network causes about a 9% decrease in the central city population. Since one standard deviation in the number of rays in an MSA is 1.5, this means that a one standard deviation increase in the number of rays causes an about 14% decrease in the central city population. To get a sense of the magnitude of this effect, the US population grew by 64% during his study period, the MSA population grew by 72%, and the constant-boundary central city population declined by 17%. Thus, the interstate highway system can account for almost the entire decline in old central city population densities. Note that, since Baum-Snow (2007) estimates the share of the population in the treated area, he avoids the problem of distinguishing between growth and reorganization. The share of the population in the central city reflects changes in the level of the central city and the suburb and migration between the two.

This result has been extended to two other contexts. Baum-Snow et al. (2012) conduct essentially the same regression using data describing Chinese prefectures between 1990 and 2010. They first partition each prefecture into the constant-boundary administrative central city and the residual prefecture and then examine the effect of several measures of infrastructure on the decentralization of the population and employment. They rely on historical routes (from 1962) as a source of quasi-random variation in city-level infrastructure. They find that each major highway ray causes an about 5% decrease in the central city population. No other measure of infrastructure—kilometers of highways, ring road capacity, kilometers of railroads, ring rail capacity, or radial rail capacity—has a measurable effect on the organization of the population in Chinese prefectures. Baum-Snow et al. (2012) also examine the effect of infrastructure on the organization of production. They find that radial railroads and highway ring capacity both have dramatic effects on the organization of production. In particular, each radial railroad causes about 26% of central city manufacturing to migrate to the periphery, while ring roads also have a dramatic effect. This effect varies by industry. Industries with relatively low weight-to-value ratios are more affected. None of the other infrastructure measures they investigate affect the organization of production.

Finally, Garcia-Lopez et al. (2013) consider the effect of limited-access highways on the organization of the population in Spanish cities between 1991 and 2011. Their unit of observation is 1 of 123 Spanish metropolitan regions. They conduct a version of the intracity regression in first differences to explain the change in the central city population between 1991 and 2011 as a function of changes in the highway network over the same period. They rely on three historical road networks as an instrument for changes in the modern network: the Roman road network; a network of postal roads, circa 1760; and a network of nineteenth century main roads. They find that each radial highway causes an about 5% decrease in the central city population, and that kilometers of central city or suburban highways have no measurable effect. Using a similar instrumentation strategy, Ángel Garcia-López (2012) examines the impact of transportation improvements on the location of the population within the city of Barcelona. Consistent with some of the findings discussed above, improvements to the highway and railroad systems are found to foster population growth in suburban areas, whereas the expansion of the transit system is found to affect the location of the population inside the central business district.[20]

While the decentralization articles mentioned above investigate the effect on central cities of infrastructure improvements which reduce the cost of accessing peripheral land, Ahlfeldt et al. (2014) consider the effect of changes in transportation cost between two adjacent parts of the same central city. Specifically, Ahlfeldt et al. (2014) consider the effect of the construction and destruction of the Berlin Wall, which separated West Berlin from the historical central business district. They examine population, employment, and land values in 1936, before the partition of the city, in 1986, shortly before reunification, and in 2006, 15 years after reunification. That is, when the cost of commuting from the West to the East was low, prohibitively high, and low again.

Methodologically, the analysis by Ahlfeldt et al. (2014) differs dramatically from the analyses in the centralization articles mentioned above. Their sample consists of approximately 16,000 "statistical blocks" comprising metropolitan Berlin, each with a population of about 250 people in 2005. Loosely, for each block, Ahlfeldt et al. (2014) record the location, population, land rent, and employment in the 3 years of their study. They use these data to estimate a first-differences variant of the intercity regression (Equation 20.32). The reduced-form results in Ahlfeldt et al. (2014) show that the construction of the Berlin Wall caused the central business district to migrate so that it was more nearly central in the territory of West Berlin, and that the removal of the Berlin Wall approximately reversed this process. The identifying assumption underlying this natural experiment is that change in access to economic activity following from division and reunification is uncorrelated with other changes in the way the city was organized, except through its effect on access to economic activity. In addition to these reduced-form results, Ahlfeldt et al. (2014) also conduct structural estimations, which we discuss later.

20.5.1.2 Infrastructure and miscellaneous city-level outcomes

Beyond the literature investigating infrastructure and decentralization, a series of articles by Duranton and Turner investigate the relationship between roads and employment growth, intercity trade, and driving.

Duranton and Turner (2012) examine employment growth in US MSAs between 1984 and 2004. Their principal regression is a variant of the intracity regression for which the outcome is employment growth between 1984 and 2004, and their measure of transportation is kilometers of interstate highways within city boundaries. They rely on the

[20] One issue that has received relatively little attention in the intracity literature is the role of transportation infrastructure in segregating cities and leaving some neighborhoods "on the wrong side of the tracks."

1947 highway plan, a map of the 1898 railroad network, and maps of historical routes of exploration as sources of exogenous variation in the interstate highway network. Their main finding is that a 10% increase in kilometers of interstate highways causes an about 1.5% increase in employment over 20 years. Alternatively, a one standard deviation in initial roads causes a change in employment growth of about 15% over 20 years. This is a bit under two-thirds of the sample average growth rate.

Duranton and Turner (2012) also estimate a second equation in which they examine the effect on employment growth of changes in the stock of roads in the nearest large city. In the context of Figure 20.6, this corresponds to looking for an effect in the treated region from changes in the residual region. They find no effect. This regression, together with their main intracity regression, provides a tentative basis for concluding that roads cause employment growth in cities rather than simply rearranging employment across cities.

In a second exercise, Duranton et al. (2014) investigate the relationship between intercity trade flows in 2007 and the interstate highway network. Their unit of analysis is a US "commodity flow survey area": a reporting unit somewhat larger than an MSA. They record the weight and value of pairwise trade flows between 69 such units and also aggregate flows in and out of each area by sector. On the basis of a method pioneered in Redding and Venables (2004) and Anderson and van Wincoop (2003), they develop two estimating equations. The first is a variant of the intercity regression and explains pairwise trade flows of weight and value as a function of pairwise interstate distance. The second is a variant of the intracity regression and predicts aggregate flows in and out of each city, by weight and value (irrespective of destination). In each case, they use the 1947 highway plan and the 1898 railroad network to derive instrumental variables. For the intracity regression, they also use instruments derived from routes of major explorations between 1530 and 1850. They arrive at three main findings. First, a 1% decrease in pairwise travel distance causes an about 1.4% increase in the value of pairwise trade and a 1.7% increase in its weight. Second, within-city highways affect the weight of exports, but not their value. Specifically, a 1% increase in the lane kilometers of within commodity flow survey area interstate highways causes an about 0.5% increase in the weight of exports but has no measurable effect on the value of exports. A 50-year panel of employment data confirms this result. Cities with more highways employ more people to make heavy manufactured goods, and conversely.

Finally, Duranton and Turner (2011) investigate the effect of the supply of roads and highways on the amount of driving in a city. More specifically, they conduct a version of the intracity regression. The outcome variable of interest is a measure of the total vehicle kilometers driven in a US MSA on particular road networks in a year and the explanatory variables of interest measure the extent of the road networks. They conduct this regression in levels, first differences and second differences. They also rely on maps of the 1947 highway plan, the 1898 railroad network, and routes of major expeditions of exploration between 1530 and 1850 as sources of exogenous variation in MSA roads. They establish a "fundamental law of road congestion," according to which driving increases by about 1% for each 1% increase in the stock of roadways, a finding that is robust across all of their specifications. They provide a rough decomposition of the sources of the marginal induced driving. About half comes from changes in individual behavior. Increases in commercial driving are less important. Migration in response to new roads and diversion of traffic from other networks appear to be least important. Hsu and Zhang (2012) replicate the analysis of Duranton and Turner (2011) using Japanese data. They arrive at the same conclusion: driving in Japanese cities increases by about 1% for each 1% increase in the extent of the road network.

While the above-mentioned articles are concerned with the relationship between overall traffic volumes and lane kilometers of roads, Couture et al. (2012) examine the determinants of driving speed in large US cities. Remarkably, their article is the first to estimate an econometric framework in which the supply and demand for travel are both explicitly modeled. The estimation results are used to construct a city-level index of driving speed and to undertake a welfare analysis of counterfactual changes in driving speed. Cities differ substantially in terms of driving speed, and the welfare gains from improvements in driving speed in the cities with the lowest driving speeds are found to be large. Taken together, these results are consistent with substantial deadweight losses from congestion.

Although most of the intracity literature is based on one of the three instrumental variables estimation strategies discussed above, the article by Gonzalez-Navarro and Quintana-Domeque (2013) is noteworthy for its use of a randomized experiment research design to examine the effects of road paving in Mexico. Homes in treatment streets that were paved experienced an increase in value of between 15% and 17% relative to those in control streets. The estimated rate of return to road pavement is 2% without taking into account externalities, but rises to 55% after incorporating externalities.

20.5.1.3 Subways and the internal organization of cities, and related other results

A large body of literature examines the effect of subways on the internal organization of cities. These articles typically consider a unit of analysis that is small relative to the city—for example, a census tract or zip code. The explanatory variable of interest is typically the distance to the subway. The outcome of interest is typically population or employment density, land prices, or ridership rates. That is, these articles perform a version of the intercity regression (here inaptly named), Equation (20.32), at a subcity scale of analysis. As we discussed in Sections 20.4.2 and 20.4.3, such regressions must overcome two problems: endogeneity and distinguishing between growth and reorganization.

The literature on subways is too large to survey exhaustively. We focus on three articles which provide, in our opinion, the best resolution to the endogeneity problem—Gibbons and Machin (2005), Billings (2011), and Ahlfeldt et al. (2014)—on two articles showing that within-city roads are associated with qualitatively similar density gradients as subways—Baum-Snow (2007) and Garcia-Lopez et al. (2013)—and finally, on two articles which provide cross-city evidence of the effects of subways—Baum-Snow and Kahn (2005) and Gordon and Willson (1984). Gibbons and Machin (2005) and Billings (2011), in particular, provide more extensive surveys.

Gibbons and Machin (2005) conduct a difference-in-differences estimate of the intercity estimation equation in order to evaluate the effect on London residential real estate prices of subway extension in the late 1990s. Their unit of observation is a "postcode unit," an administrative unit containing 10–15 households. They observe real estate transactions by postcode unit before and after the Docklands light rail extension in south London. As a consequence of this extension, parts of their sample experience a decrease in the distance to a subway station. This makes a difference-in-differences estimate possible: they compare the change in real estate prices in postcodes that experienced changes in subway access with the change in real estate prices in postcodes that did not.

They find that for properties within 2 km of a station, a 1 km reduction in station distance causes an about 2% increase in real estate prices. Usefully, Gibbons and Machin (2005) compare their difference-in-differences estimate with a more conventional cross-sectional estimate. They find that estimates based on cross-sectional variation alone are 3× as large as difference-in-difference estimates. This suggests that, as we might hope, subway station locations are not selected at random and more valuable land is more likely to receive subway service.

Billings (2011) and Ahlfeldt et al. (2014) also conduct difference-in-differences estimates of the effects of subways. For a newly opened light rail line in Charlotte, North Carolina, Billings (2011) finds that residential real estate prices within 1 mile of a station increase by about 4% for single-family homes and by about 11% for condominiums, and that light rail access has no effect on commercial property prices. Ahlfeldt et al. (2014) find that city blocks further than 250 m from a 1936 subway station experienced an about 13% smaller decrease in the price of floor space as a consequence of the division of Berlin than did those within 250 m. Glaeser et al. (2008) look at the effects of the New York city subway and find evidence that poor people move to be closer to subway stations.

Each of these three articles investigates the rate at which land rent declines with distance from a subway or light rail line. Baum-Snow (2007) and Garcia-Lopez et al. (2013) investigate how population density varies with distance to a highway. The unit of observation in Baum-Snow (2007) is a census tract. For each US census tract in a 1990 MSA, he observes the population density in 1970 and 1990 and the distance to an interstate highway. This allows him to estimate a variant of the intercity estimating equation for two cross sections and in first differences. He finds that a 10% decrease in the distance to a highway is associated with an about 0.13% increase in population density in 1970 and a slightly smaller increase in 1990. First-difference estimates are similar. Garcia-Lopez et al. (2013) arrive at similar estimates using Spanish data.

While each of these articles attempts to resolve the problem of endogenous placement of infrastructure, they do not provide a basis for determining whether subways cause growth or reorganization of nearby economic outcomes. In particular, they are unable to measure whether a change in a city's subway network affects city-level variables. In the context of Figure 20.6, this would correspond to asking whether a change in treated unit infrastructure affects the level of an outcome in all three regions. This question, which is of obvious public policy interest, requires cross-city data describing subways and city-level outcomes—that is, data which allow the estimation of the intracity regression (Equation 20.31). Since subways are relatively rare, data of this sort are difficult to assemble, and we know of only two such efforts to date. The first, by Gordon and Willson (1984), constructs a single cross section of 52 cities that describes population density, subway passenger kilometers per year, and a handful of city-level control variables. In a simple cross-sectional estimate of ridership on density, they find a strong positive relationship. Baum-Snow and Kahn (2005) construct disaggregated panel data describing a panel of 16 US metropolitan areas with subways. In addition to describing the extent of each city's subway network, their data describe ridership commuting times. Overall, they find little evidence that US subway expansions elicit large increases in ridership.

20.5.2 Intercity infrastructure and the geographic organization of economic activity

We now turn our attention to the effect of infrastructure that connects a unit of observation, typically a county, to the rest of the world. This most often involves estimating a version of the intercity regression. We first describe results for high-income countries and then turn to results for low-income countries and historical data.

20.5.2.1 High-income countries

Chandra and Thompson (2000) consider the effect of the interstate highway system on a sample of 185 nonmetropolitan US counties that received a highway after 1969, and 391 neighboring nonmetropolitan counties that did not. By restricting their attention to nonmetropolitan counties, Chandra and Thompson (2000) hope to restrict their attention to counties that were treated with highways "accidentally," and in particular, without regard for the effect of highways on the treated counties. This is the pioneering use of the inconsequential place approach to identification. Their outcome measures are aggregate annual earnings by county, year, and one-digit Standard Industrial Classification code, for all years from 1969 to 1993.

Chandra and Thompson (2000) estimate a distributed lag version of the intercity regression with county fixed effects. In particular, they include 24 dummies for the age of the highway connection in each year as explanatory variables. Their results are striking. They find a marginally positive 24-year effect of a highway connection on earnings in finance, insurance, real estate, transportation and public utilities, and retail and services. They find that the effect on earnings in manufacturing and farming is marginally negative. Overall, the 24-year effect on earnings of a highway connection of a nonmetropolitan county is a 6–8% increase. The effect on untreated neighboring counties is approximately the opposite. Overall, untreated neighboring counties see a decrease in total earnings of between 1% and 3%. Note that Chandra and Thompson (2000) estimate two distinct equations. In the context of Figure 20.6, the first predicts the effect of changes in infrastructure on the treated area, and the second predicts the effect of changes in infrastructure on neighboring untreated regions. Together, these two regressions are exactly what is required to distinguish between growth and reorganization. Importantly, Chandra and Thompson (2000) cannot reject the hypothesis that aggregate changes in earnings caused by a highway connection sum to zero across the whole sample of treated and neighboring counties.

Michaels (2008) considers a sample of 2000 counties in the United States that are more than 50% rural and had no highways in 1950—that is, the inconsequential place approach. He then identifies a subset of the interstate network constructed between 1959 and 1975 to serve intercity travel. His explanatory variable of interest is an indicator of whether a county is connected to this network at the end of the study period. He also relies on a planned route instrumental variable approach based on the 1947 highway plan. He considers a number of outcome variables, in particular, per capita earnings in trucking and retail sales, and the relative wages of skilled and unskilled workers. He finds that rural counties receiving highway connections experience about the same increase in trucking and retail earnings as Chandra and Thompson (2000) observe, the only two outcome variables common to the two articles. This is reassuring given the quite different identification strategies. He also finds that highways cause a small increase in the wage of skilled workers relative to that of unskilled workers.

In the first of two related, but methodologically quite different articles, Redding and Sturm (2008) consider the effect of the postwar partition of Germany on the organization of economic activity. They find that the population of German cities near the East–West border grew more slowly than that of those far from the border. That is, in response to an increase in the cost of travel between East Germany and West Germany, economic activity migrates away from the border region. Duranton et al. (2014) examine the effect of pairwise distance on pairwise trade of manufactured goods between US cities in 2007. They find that trade responds to highway distance rather than straight-line distance, that the effect of distance on trade is large, and that it is larger on the weight of goods than on their value. Unsurprisingly, Duranton et al. (2014) also find that trade by rail is less sensitive to distance than is trade by road. Curiously, Duranton (2014) replicates the analysis of Duranton et al. (2014) using data describing trade in Columbia rather than the United States. He reaches somewhat different conclusions: trade is less sensitive to distance, the value and the weight of trade are about equally sensitive to infrastructure, and the value of trade responds to infrastructure.

While most of the intercity literature has focused on roads, Sheard (2014) estimates the effects of airport infrastructure on relative sectoral employment at the metropolitan area level, using data from the United States. To address the potential endogeneity in the determination of airport sizes, the 1944 National Airport Plan is used as an instrument for the current distribution of airports. Airport size is found to have a positive effect on the employment share of tradeable services, controlling for overall local employment, but has no measurable effect on manufacturing or most nontradeable sectors. The effect of airport size on overall local employment is practically zero, suggesting that airports lead to specialization but not growth at the metropolitan area level. The implied elasticity of tradeable service employment with respect to airport size is approximately 0.22.

20.5.2.2 Low-income countries

Donaldson (2015) considers the effect of railroads on a sample of 235 "districts" covering the preponderance of India during the period from 1870 to 1930. He uses these data to estimate the intercity regression with district and year fixed effects. His outcome variable is the aggregate annual value of 17 agricultural crops per unit of district area. During this study period, agriculture accounted for about two-thirds of Indian GDP, and the 17 crops Donaldson considers accounted for 93% of the value of agricultural output. To investigate the probable endogeneity of railroads, Donaldson gathers data describing hypothetical planned railroad networks that were competitors to the realized network. He finds no difference in output between districts treated with planned networks and those not treated. This suggests that the realized network did not target the most productive districts.

Donaldson finds that districts with access to the railroad report about 17% higher real agricultural income per unit of district area than districts without railroads. Because Donaldson's regression equation contains year and district effects, this means that a district treated with a railroad connection sees its income increase by 17% relative to untreated districts. This is a large effect. Over the course of the 1870–1930 study period, India's real agricultural income increased by only about 22%, so a rail connection was equivalent to more than 40 years of economic growth.

In a related article, Donaldson and Hornbeck (2013) consider a sample of about 2200 counties in the continental United States between 1870 and 1890, a period of rapid rail expansion. They also perform a variant of the intercity regression, this time with county fixed effects, state-year fixed effects, and a cubic polynomial in latitude and longitude. The outcome variable of interest is the total value of a county's agricultural land.

Donaldson and Hornbeck (2013) find that counties treated with rail access in a year experience a 34% increase in aggregate agricultural land rent relative to others in the same state and year. If the share of agricultural land in production stays approximately constant during their study period, then this implies the same effect on output, nominally larger than the corresponding estimate for India. With this said, the rate of growth in the United

States was much higher during this period, so a rail link was equivalent to only about 7.5 years of economic growth, as opposed to more than 40 years for Indian districts.

Beyond the inclusion of county fixed effects and other controls, Donaldson and Hornbeck (2013) do not have a strategy to deal with the endogeneity of rail access in the specification discussed above. Instead, they conduct an alternative regression where the explanatory variable of interest is a measure of market access. Their measure of market access results from a model similar to the one we described in Section 20.3 and is well approximated by a "gravity" measure of population—that is, an inverse travel time weighted sum of county populations. They find that the effects of this measure are similar to those of the connection indicator. They also find that the effects of a restricted gravity measure, which excludes nearby counties, have a similar effect. That the two gravity measures have similar effects suggests that the effect of rail access on a county depends equally on rail access to places near and far away.

Haines and Margo (2008) conduct an analysis similar to that of Donaldson and Hornbeck (2013). They consider a sample of 655 counties in 12 US states and estimate the intercity regression in first differences. Their study period runs from 1850 to 1860, just before the 1870–1890 period that Donaldson and Hornbeck consider. They primarily consider the following outcome measures: share of urban population, agricultural wage, agricultural output per acre, and improved acreage share. Their measure of rail access is an indicator variable describing whether or not a rail line passes through a county in a year. They find that rail access is associated with a 10% increase in the share of a county's improved acreage, a 3% increase in farm wages, no effect on output per improved acre, a small increase in service sector employment, and a 4% decrease in agricultural employment. In spite of the fact that Haines and Margo (2008) consider many of the same counties as Donaldson and Hornbeck (2013), and that the two study periods are adjacent, these results are much smaller than those obtained by Donaldson and Hornbeck (2013).

Bogart (2009) uses a sample of about 3000 English parishes and townships between 1692 and 1798 to estimate the intercity regression in first differences. His dependent variable is land rent per acre. His measure of transportation is an indicator of whether a parish or township is close to a turnpike, an improved road maintained by tolls. He also conducts an instrumental variables variant of the first-differences intercity regression, where he uses proximity to a major trade route as an instrument for the presence of a turnpike. This is a variant of the inconsequential places approach developed in Chandra and Thompson (2000). Bogart (2009) finds that a turnpike increases parish or township land rent by about 11% in first-difference estimates and by about 30% in instrumental variable estimates.

Banerjee et al. (2012) use county-level Chinese data to estimate the intercity regression with provincial and year fixed effects, and county-level controls. They consider a sample of 310 Chinese counties, for which they observe per capita GDP annually from 1986 until 2006, a period when Chinese road and rail infrastructure expanded dramatically. They also consider a census of firms for a larger set of counties in a smaller number of years. To measure infrastructure, Banerjee et al. (2012) construct a hypothetical network connecting "treaty ports" to interior trading centers and use this network as an instrument. Again, this is a variant of the inconsequential places approach. Their measure of infrastructure is the distance from a county to a line in this hypothetical network, which predicts the proximity to both railroads and major highways.

Since Banerjee et al. (2012) have one instrument and two endogenous dependent variables, proximity to railroads and highways, they cannot separately identify the effects of roads and railroads. Instead, they present the results of an intercity regression in which the measure of transportation access is the distance to the hypothetical line. Therefore, as they acknowledge, these results are somewhat difficult to interpret. With this said, Banerjee et al. (2012) arrive at robust and interesting results. In particular, a 10% increase in the distance to a "line" causes an about 6% decrease in county GDP and has no effect on the growth of income. They find that the gradient for the density of firms is slightly steeper and that proximity to a line has no effect, or possibly a small negative effect, on the growth rate of firm density.

Storeygard (2012) uses a sample of 287 small cities in sub-Saharan Africa between 1992 and 2008 to estimate a first-differences variant of the intercity regression. This article is innovative in two regards. First, it uses "lights at night data" as a proxy measure for city GDP in small developing countries where data availability is limited.[21] Second, to generate time series variance in transportation costs, he causes constant network distances to interact with a measure of the price of oil on international markets. As he observes, the validity of this approach hinges on the claim that, conditional on controls, oil prices do not affect city lights except through transportation costs. Thus, more specifically, for a sample of 287 small cities, Storeygard (2012) estimates a variant of the intercity regression where the outcome of interest is a measure of average annual light intensity for constant-boundary cities, and the measure of transportation costs is the interaction of network distance with annual average oil prices, city fixed effects, and variables to control for other possible channels through which oil prices might affect light intensity. Storeygard (2012) estimates that doubling the distance between a sample city and the primate port city causes an about 6% reduction in GDP, and that this is close to the effect of a quadrupling of fuel costs.

Jedwab and Moradi (2013) provide evidence regarding the intercity regression using rail construction in colonial sub-Saharan Africa, where over 90% of African railroad lines were built before independence. Colonial railroads are found to have strong effects on commercial agriculture and urban growth before independence. A number of

[21] Henderson et al. (2012) pioneer the use of these data and show that they are highly correlated with national-level GDP, a result that Storeygard (2012) confirms at the subnational level.

identification strategies are used to provide evidence that these effects are causal, including placebo lines that were planned but not built and a version of the inconsequential units approach. Furthermore, using the fact that African railroads fell largely out of use after independence, owing to mismanagement and lack of maintenance, the article shows that colonial railroads had a persistent impact on cities. While colonial sunk investments (e.g., schools, hospitals, and roads) partly contributed to urban path dependence, the evidence suggests that railroad cities persisted because their early emergence served as a mechanism to coordinate contemporary investments for each subsequent period.

Faber (2015) also estimates a version of our intercity regression using a sample of about 1300 rural Chinese counties that are more than 50 km from a major city and that he observes in 1990, 1997, and 2006. For each county and year, he observes county-level GDP in three sectors—agriculture, industry, and services—as well as government expenditure. He also observes a rich set of county-level controls. His measure of infrastructure is the distance from the county centroid to the nearest segment of the trunk highway network, the limited access highway network that was substantially constructed during Faber's study period. To resolve the probable endogeneity of the network placement, he relies on two hypothetical networks. The first resembles the hypothetical network developed by Banerjee et al. (2012). The second describes the cost-minimizing network to connect a set of major cities targeted by plans for the realized network. Faber (2015) finds that industrial GDP, total GDP, and government revenue all *decrease* with proximity to the network. This result, which appears robust, is without precedent in the literature. Every other implementation of this research design we survey arrives at the opposite conclusion–that is, that transportation infrastructure attracts (or creates) economic activity.

Ghani et al. (2013) use the inconsequential units approach to estimate the intercity regression for "The Golden Quadrilateral Project," which upgraded the quality and width of 5846 km of roads in India. A difference-in-differences specification is used to compare non-nodal districts on the basis of their distance from the highway system. Positive treatment effects are found for non-nodal districts located 0–10 km from the Golden Quadrilateral that are not present in districts 10–50 km away, most notably for higher entry rates and increases in plant productivity.

20.6. DISCUSSION

20.6.1 Growth versus reorganization

Determining the extent to which the observed effects of infrastructure reflect changes in the level of economic activity versus a reorganization of existing activity is fundamental to understanding the effects of infrastructure and to policy analysis. The existing reduced-form literature generally does not provide a basis for separately identifying the two effects. In spite of this, we can suggest some tentative conclusions about the contributions of growth and reorganization to the observed effects of infrastructure. These conclusions are based on comparisons between four sets of estimation results.

First, Duranton et al. (2014) examine the effect of within-city highways on the composition and value of intercity trade for US cities. They find that an increase in within-city highways causes cities to become more specialized in the production of heavy goods, but has at most small effects on the total value of trade. Here, the primary effect of within-city highways is to reorganize economic activity, not to create it.

Second, from the results in Baum-Snow (2007), Garcia-Lopez et al. (2013), and Baum-Snow et al. (2012), respectively, the effects of a one standard deviation increase in the number of radial highways cause the central city population to decrease by 14%, 5%, and 17% where secular rates of city population growth were 72%, 30%, and 55%. Thus, the transportation network causes reorganizations of cities that are large compared with the forces affecting them. On the other hand, Duranton and Turner (2012) find that a one standard deviation increase in within-city lane kilometers of interstate highways causes an about 15% increase in the population over 20 years. Happily, the samples of cities and years considered by Baum-Snow (2007) and Duranton and Turner (2012) substantially overlap. While the comparison is somewhat strained, it suggests that growth and reorganization are about equally important.

Third, Banerjee et al. (2012) conduct intercity regressions where the outcome variable is the level of GDP, and where it is the growth of GDP. They find that transportation infrastructure (really, their hypothetical network connecting treaty ports and interior trading centers) has important effects on the level of output, but not on its growth.

Fourth, and finally, Chandra and Thompson (2000) find that interstate highways increase firm earnings in US counties treated with interstate highways at the expense of their untreated neighbors. Summing over the treated and untreated counties, they cannot reject the hypothesis of no change.

While our evidence here is fragmentary, it suggests two conclusions. First, within large cities, relocation of economic activity in response to transportation infrastructure is at least as important as the creation of economic activity. This conclusion is broadly consistent with current estimates of agglomeration effects: if output increases by 2% with each doubling of city size, then even if infrastructure can double population size, we will see only small increases in productivity. Second, for nonurban counties, the primary effect of treatment with highways or railroads is to attract economic activity at the expense of more remote areas, with some variation by industry.

20.6.2 The effects of transportation infrastructure on economic activity

20.6.2.1 Invariance across economies

Quite different data underlie the three decentralization articles. Baum-Snow (2007) considers a 40-year study period and a US unit of observation with a mean population around

160,000. Garcia-Lopez et al. (2013) consider a 20-year study period and a Spanish unit of observation with a mean population around 120,000. Baum-Snow et al. (2012) consider a 20-year study period and a Chinese unit of observation with a population near 4 million. In spite of this, the three studies find remarkably similar effects of highways on the decentralization of the population from central cities to suburbs; 5% per ray for Spanish cities, 9% per ray for US cities, and 5% per ray for Chinese cities. That the effect of radial highways on population decentralization is so nearly the same in such different contexts suggests that the effects of infrastructure are not sensitive to the scale of the analysis or the details of the economies where the cities are located.[22]

Other comparisons bolster this proposition. First, Duranton and Turner (2011) and Hsu and Zhang (2012) find, respectively, that a 1% increase in limited-access highways in a metropolitan area increases driving by 1% in US and Japanese metropolitan regions. Second, the effect of subways on land rent gradients appears to be about the same in suburban London as in Charlotte, North Carolina, while the effect of highway access on population density gradients appears similar in the United States and Spain.

Finally, with a few exceptions, there is broad agreement among the many articles that estimate the intercity regression: Chandra and Thompson (2000) find a 6–8% increase in firm earnings in counties adjacent to the interstate highway network; Michaels (2008) confirms the finding of Chandra and Thompson (2000) in the two industries where they overlap; Donaldson (2015) finds 17% higher real agricultural income for Indian districts with rail access; Haines and Margo (2008) find a 3% increase in farm wages for counties served by a railroad; Bogart (2009) finds an 11–30% increase in land rent for parishes served by a turnpike; Banerjee et al. (2012) find a 6% decrease in per capita income from doubling the distance to a hypothetical trade route; and Storeygard (2012) finds a 6% decrease in city light intensity from doubling the cost of travel to the primate city. Donaldson and Hornbeck (2013) and Faber (2015) are outliers, predicting a 34% increase in agricultural land rent for counties served by a railroad and a decrease in output for counties closer to a highway.

If we exclude the work of Faber (2015), and ignore the problem of comparing the gradient estimates of Banerjee et al. (2012) and Storeygard (2012) with discrete treatment effects in the others, these estimates are all within one order of magnitude.[23]

[22] It also suggests that the changes caused by radial highways may occur more rapidly than these 20- or 40-year study periods considered by extant research.

[23] Banerjee et al. (2012) consider the effect of the distance to a line, rather than an indicator for whether a line crosses a county. Therefore, their results cannot be compared directly with results based on treatment indicators. However, an average county in their sample is approximately 2000 km^2, the area of a square about 45 km on a side. Given this, doubling or quadrupling the distance from a county center to a line should usually be enough to remove an intersecting segment. This suggests that the effect of an indicator variable for line presence should be in the neighborhood of 6–12%.

Given the differences in the underlying economies that are the subject of these studies, this seems remarkable.

In sum, the literature suggests that transportation infrastructure has similar effects on the organization of economic activity across a range of countries and levels of development. More specifically, highways cause the decentralization of economic activity and an increase in its level in cities, highways cause a dramatic increase in driving, and highways and railroads cause an increase in economic activity in rural areas near highways. This conclusion is subject to four caveats. First, there is some disagreement among articles estimating the intercity regression. Second, although the methods and data used in these articles are similar, they are not identical, so comparisons between them need to be regarded with caution. Third, as we noted above, we do not have much basis for distinguishing growth from reorganization. Fourth, and finally, Duranton et al. (2014) and Duranton (2014) examine the effects of roads on trade in the United States and Columbia and find different effects.

20.6.2.2 Variability across activities and modes

While the literature surveyed above suggests a number of general results, it also provides suggestive evidence that different activities respond differently to changes in infrastructure. The three decentralization studies—Baum-Snow (2007), Baum-Snow et al. (2012), and Garcia-Lopez et al. (2013)—find that decreasing transportation costs leads the population to migrate to the lower-density periphery. Here, reductions in transportation costs reduce central city population density. Baum-Snow et al. (2012) find that manufacturing decentralizes along with the population.

Empirical results from the literature conducting intercity regressions also suggest heterogeneous responses by industries. Chandra and Thompson (2000) find different responses to the interstate highway access in rural counties by different sectors, a result confirmed in Michaels (2008). Haines and Margo (2008) find a shift of land into agriculture and of employment into services with rail access in nineteenth century United States counties. Duranton et al. (2014) find that US cities with more highways specialize in the production of heavier goods.

Finally, the gradient estimates in Banerjee et al. (2012) can be directly compared to within-city regressions estimating the effects of population density or land rent on proximity to a road—for example, Baum-Snow (2007) and Garcia-Lopez et al. (2013). This comparison suggests a much steeper gradient for economic activity near rural highways than near urban highways.

Broadly, these studies support the claim that the weight per unit value of output, land share of production, and sensitivity to agglomeration are all economically important determinants of how a firm or industry responds to changes in transportation infrastructure. The literature is as yet too incomplete to provide much insight into the relative importance of these different factors. More speculatively still, highways may have larger effects on the organization of economic activity in rural areas than in cities.

20.6.2.3 Political economy of infrastructure allocation

As discussed above, a central issue in evaluating the effects of transportation improvements is that these improvements are not randomly assigned. Implicit evidence for the process through which transportation investments are assigned can be obtained by comparing the OLS coefficients for the intercity and intracity regressions (which capture the impact of transportation investments assigned through the existing political process) with the instrumental variable coefficients (which capture the impact of transportation investments assigned through quasi-experimental variation). In Baum-Snow (2007) and Duranton and Turner (2012), instrumental variable estimates are larger in magnitude than OLS estimates. This suggests that the equilibrium allocation process assigns roads to places growing more slowly than a randomly selected city. Baum-Snow et al. (2012) and Garcia-Lopez et al. (2013) find contrary results for China and Spain. Thus, conditional on the validity of their respective identification strategies, these articles point to implicit differences in the political economy of infrastructure funding across countries.

Further research is needed explicitly examining the political economy of transportation infrastructure investments. Knight (2002) examines the US Federal Aid Highway Program, over which the House Committee on Transportation and Infrastructure and the Senate Environment and Public Works Committee have jurisdiction. The article finds evidence that measures of the political power of state delegations affect the allocation of funds, including a state's proportion of members serving on the transportation authorization committee, the proportion of a state's representatives in the majority party, and the average tenure of a state's representatives. Federal highway grants are found to crowd out state highway spending, leading to little or no increase in net spending.

20.6.3 General equilibrium effects

Generally, studies of the effect of infrastructure on the internal organization of cities do not consider the role of market access. This occurs despite the fact that market access is a component of the theoretical precursor of both the intercity and intracity regression equations. This appears to rest on the assumption, usually implicit, that cities are small open units and that we can examine changes in their internal structure and level of economic activity without reference to other cities. In fact, Duranton and Turner (2012) make this small open city assumption explicitly and attempt to test it by examining the effect on a target city of a change in the stock of roads in the nearest large city. While this is not a particularly satisfactory test, that they find no effect suggests that disregarding interactions between cities while studying the effect of transportation infrastructure on their internal workings is reasonable.

The problem of market access merits two further comments. First, for the purpose of examining pairwise trade flows, Redding and Venables (2004) develop a framework which allows the explicit estimation of market access and variants of estimating Equations (20.32) and (20.31) based on a two-step estimation procedure. It is this framework that Duranton et al. (2014) apply to their investigation of the effect of the interstate highway system on pairwise trade flows between US cities. Second, the extant empirical literature can be usefully divided into two classes. The first follows a long tradition of conducting city-level regressions that assume implicitly (or explicitly in the case of Duranton and Turner, 2012) that cities can be regarded as independent units. In this framework, what happens in each city is pinned down by the utility level in a residual rural sector. This implies that what happens in one city does not affect what happens in other cities. The second follows the trade or new economic geography literature—for example, Redding and Sturm (2008)—and supposes that the interactions between cities are important. An interesting area for further research is reconciling these two different approaches.

20.6.4 Structural estimation, general equilibrium, and welfare

The recent reduced-form literature has made important strides in identifying causal effects of infrastructure on economic activity in rural regions. Specifically, this literature estimates changes in economic activity by industry and changes in population for cities and rural regions. We are just beginning to investigate whether different modes of transportation have different effects. With this said, the existing literature provides at most suggestive evidence for the extent to which the observed effects of infrastructure reflect the reorganization or creation of economic activity. Progress on this issue appears to fundamentally require an econometric framework which is capable of dealing with general equilibrium effects such as the possibility that infrastructure moves activity from one unit to another.

In the remainder of this section, we discuss a number of studies that have used structural approaches to estimate intercity or intracity effects of transportation infrastructure. These studies highlight four main advantages of a structural approach. First, as discussed above, this approach enables general equilibrium effects to be captured. Second, a structural approach allows for the estimation or testing of specific economic mechanisms. Third, the estimated model can be used to quantify aggregate welfare effects (as, e.g., in Section 20.3.6). Fourth, the estimated model can be used to undertake counterfactuals and generate ex ante predictions for the effects of policies that have not yet been implemented (see, e.g., Section 20.3.8).

We begin with intercity studies. Redding and Sturm (2008) use the division of Germany after the Second World War and the reunification of Germany and in 1990 as a natural experiment to provide evidence in support of a quantitative model of economic geography. As discussed above, in the aftermath of division, cities in West Germany close to the East German–West German border experienced a substantial decline in population growth relative to other West German cities, and the estimated treatment effect is larger for small cities than for large cities. In a multiregion extension of the Helpman (1998) model, the treatment effect of division on border cities depends on two parameter combinations that capture (a) the strength of agglomeration and dispersion forces and (b) the elasticity of trade with respect to distance. For plausible values of these parameter combinations, the model can account quantitatively for both the average treatment effect of division and the larger treatment effect for small cities than for large cities. Smaller cities are more adversely affected by division, because they are disproportionately dependent on markets in other cities.

Donaldson (2015) combines a general equilibrium trade model with archival data from colonial India to investigate the impact of India's vast railroad network. The empirical analysis is structured around an extension of the analysis of Eaton and Kortum (2002) to incorporate multiple agricultural commodities that shares some features with the theoretical framework developed in Section 20.3. This model delivers four key theoretical predictions that are applied to the data. First, for goods that are traded between regions, price differences between those regions can be used to measure bilateral trade costs. Second, the model yields a gravity equation for bilateral trade flows that can be used to estimate the response of trade flows to trade costs. Third, railroads increase real income levels, as measured by the real value of land income per unit area. Fourth, as in the theoretical framework developed above, each location's trade share with itself is a sufficient statistic for welfare. Consistent with these predictions of the model, there is a strong and statistically significant estimated effect of railroads on real income levels, but this effect becomes statistically insignificant after controlling for the model's sufficient statistic of a region's own trade share. These results provide evidence that the estimated effects of railroads are capturing the goods trade mechanism emphasized in the model.[24]

To quantify the intercity effects of road construction, Duranton and Turner (2012) develop a system of cities model that they use to derive a system of equations for employment and roadway growth that can be estimated empirically. Utility in each city depends on the quality of amenities, consumption of a numeraire composite good, distance traveled, and consumption of land. Productivity in producing the composite good increases with city employment through a standard agglomeration economy. The cost of travel per

[24] Transportation infrastructure may not only promote internal trade within countries (as considered here) but may also enable the interior regions of countries to participate in external (international) trade, as examined in Fajgelbaum and Redding (2013) using the natural experiment of Argentina's integration into the world economy in the late nineteenth century.

unit of distance decreases with the length of the roadway and increases with aggregate vehicle traffic through a standard congestion effect. Population mobility implies that utility in each city is equalized with utility in the outside alternative of a rural area. Equilibrium city size is determined by the willingness of residents to drive to the city center. Using equalization of utility between cities and rural areas, together with equilibrium in land and travel markets, one can express equilibrium city employment as a power function of the length of roadways. If a partial adjustment process is specified, according to which city employment growth is a function of the distance between a city's actual population and its equilibrium population, the model delivers the following equation for city employment growth:

$$n_{it+1} - n_{it} = A_1 + ar_{it} + \lambda n_{it} + c_1 x_i + \epsilon_{1it}, \tag{20.34}$$

where n_{it} is the logarithm of employment in city i at time t, r_{it} is the logarithm of the roadway, x_i are controls for city characteristics, and ϵ_{1it} is a stochastic error. If a similar partial adjustment process for road construction is specified, we obtain an analogous equation for the city roadway growth:

$$r_{it+1} - r_{it} = A_2 + \theta r_{it} + \eta n_{it} + c_2 x_i + \epsilon_{2it}, \tag{20.35}$$

where ϵ_{2it} is a stochastic error. The equilibrium logarithm of roadway length is assumed to depend on the logarithm of the city population, the city characteristics controls, x_i, and instruments, z_i, that satisfy the exclusion restriction of affecting the city population only through roadways:

$$r_{it} = A_3 + c_3 n_{it} + c_4 x_i + c_5 z_i + \epsilon_{3it}, \tag{20.36}$$

where ϵ_{3it} is a stochastic error. The identification assumptions for instrument validity are

$$c_5 \neq 0, \tag{20.37}$$

$$\mathrm{Cov}(z, \epsilon_1) = 0, \tag{20.38}$$

$$\mathrm{Cov}(z, \epsilon_2) = 0. \tag{20.39}$$

As discussed above, the instrumental variables estimates imply that a 10% increase in a city's stock of interstate highways causes an about 1.5% increase in its employment growth over 20 years. These instrumental variables estimates are somewhat larger than the OLS estimates. Therefore, an additional kilometer of highway allocated to a city at random is associated with a larger increase in employment or the population than for a road assigned to a city by the prevailing political process. These results are consistent with the view that the existing political process tends to assign highways to more slowly growing cities.

The intercity study of Desmet and Rossi-Hansberg (2013) highlights the way in which a general equilibrium model can be used to quantify the relative importance of different mechanisms and evaluate welfare effects. This paper develops a system of cities model that incorporates heterogeneity in productivity, amenities, and congestion costs as determinants of city sizes. Congestion costs are modeled as depending on city-specific transportation infrastructure. Data on US MSAs are used to estimate these city characteristics and decompose the variation in city sizes into their contributions. All three characteristics are important for explaining the observed city size distribution. Eliminating differences across cities in any one characteristic leads to large population reallocations but has small welfare effects (population reallocations of as much as 40% can have welfare gains of as little as 2%). This pattern of results is consistent with the idea that welfare is approximately equalized across cities in the initial equilibrium, in which case the envelope theorem implies small welfare effects from population reallocations. In contrast, when the same method is applied to Chinese cities, eliminating differences across cities in any one characteristic leads to both large population reallocations and large changes in welfare. These contrasting results between the two countries are consistent with urban policies in China playing an important role in determining relative city sizes and aggregate welfare.

The intercity study of Allen and Arkolakis (2013) also uses a structural approach to quantify alternative economic mechanisms and evaluate welfare effects. The article develops an Armington model of trade and factor mobility that incorporates both an economic component and a geographic component. The economic component combines the gravity structure of trade in goods with labor mobility to determine the equilibrium distribution of economic activity on a space with any continuous topography of exogenous productivity and amenity differences and any continuous bilateral trade costs. To incorporate the possibility of agglomeration and dispersion forces, the overall productivity and amenity in a location can endogenously depend on its population. The article provides general conditions for the existence, uniqueness, and stability of the spatial economic equilibrium. The geographic component of the model provides a microfoundation for bilateral trade costs as the accumulation of instantaneous trade costs along the least-cost route between locations. The model combining these economic and geographic components is used to estimate the topography of trade costs, productivities, and amenities in the United States. Geographic location is found to account for at least 20% of the spatial variation in US income. The construction of the US interstate highway system is estimated to increase welfare by 1.1–1.4%, which is substantially larger than its cost.

We now turn to intracity studies. Until recently, theoretical models of internal city structure were highly stylized, which limited their usefulness for empirical research. Much of the theoretical literature has focused on the monocentric city model, in which firms are assumed to locate in a central business district and workers decide how close to live to this central business district.[25] Lucas and Rossi-Hansberg (2002) were the first to develop a model of a two-dimensional city, in which equilibrium patterns of economic activity can be nonmonocentric.[26] In their model, space is continuous and the city is assumed to be symmetric, so the distance from the center is a summary statistic for the organization of economic activity within the city. Empirically, however, cities are not perfectly symmetric because of variation in locational fundamentals, and most data on cities are reported for discrete spatial units such as blocks.[27]

To address these challenges, Ahlfeldt et al. (2014) develop a quantitative theoretical model of internal city structure that allows for a large number of discrete locations within the city that can differ in their natural advantages for production, residential amenities, land supply, and transportation infrastructure. The model remains tractable and amenable to empirical analysis because of the stochastic formulation of workers' commuting decisions that follows Eaton and Kortum (2002) and McFadden (1974). The city is populated by an endogenous measure of $\bar{H}$ workers, who are perfectly mobile within the city and the larger economy. Workers experience idiosyncratic shocks to the utility they derive from each possible pair of residence and employment locations within the city. Workers choose their residence and employment locations and consumption of residential land and a tradeable final good to maximize their utility. This idiosyncratic formulation of utility yields a gravity equation for the probability of commuting from i to j (π_{ij}):

$$\pi_{ij} = \frac{T_{ij}\left(d_{ij}Q_i^{1-\beta}\right)^{-\epsilon}\left(B_i w_j\right)^{\epsilon}}{\sum_{r=1}^{S}\sum_{s=1}^{S} T_{rs}\left(d_{rs}Q_r^{1-\beta}\right)^{-\epsilon}\left(B_r w_s\right)^{\epsilon}}, \tag{20.40}$$

where T_{ij} is a Fréchet scale parameter that determines the average attractiveness of the bilateral commute from residence location i to employment location j, d_{ij} is the iceberg cost in terms of utility of commuting between i and j, Q_i is land prices. B_i denotes amenities at residential location i and w_j denotes wages at employment location j.

In this setting, transportation technology influences the organization of economic activity within the city through the matrix of bilateral commuting costs d_{ij}. Both residential amenities (B_i) and final goods productivity (A_j, which determines w_j) are characterized by agglomeration economies and hence depend on the transportation technology through the endogenous employment distribution. Ahlfeldt et al. (2014) use the division and reunification of Berlin as an exogenous shock to structurally estimate the strength of

[25] The classic urban agglomeration models of Alonso (1964), Mills (1967), and Muth (1969) impose a monocentric city structure. While Fujita and Ogawa (1982) and Fujita and Krugman (1995) allow for nonmonocentricity, they model one-dimensional cities on the real line.

[26] For an analysis of optimal urban land use policies in such a setting, see Rossi-Hansberg (2004).

[27] For empirical evidence regarding the extent to which the organization of economic activity within cities is indeed symmetric, see Brinkman (2013).

the model's agglomeration and dispersion forces and to show that the model can account quantitatively for the observed changes in city structure. The model also provides a framework that can be used to analyze the effects of other public policy interventions, such as transportation infrastructure investments that reduce commuting costs d_{ij} between pairs of locations.

Another structural intracity approach is that of Combes et al. (2012), which develops a method for estimating congestion costs (which depend on transportation technology) using land transactions data. The key insight behind this method is that residential mobility implies that urban (dis)amenities and commuting costs are ultimately reflected in land prices. A system of cities model is developed, in which each city is monocentric and workers face costs of commuting to the central business district. The model highlights that the elasticity of urban costs with respect to the city population is the product of three quantities: the elasticity of unit land prices at the city center with respect to the population, the share of land in housing, and the share of housing in consumption expenditure. With implementation of this method, the article's preferred estimates for these three elasticities are 0.72, 0.25, and 0.23, respectively. From the product of these three parameters, the preferred elasticity of urban costs with respect to the city population is 0.041, which is close to existing estimates of agglomeration economies in the form of the elasticity of city productivity with respect to the city population. This finding that cities operate near aggregate constant returns to scale suggests that the fundamental trade-off of spatial economics—between agglomeration economies and congestion costs—may play only a limited role in explaining the observed distribution of city sizes. This prediction is in turn consistent with the observation that cities of vastly different sizes exist and prosper.

20.7. CONCLUSION

To determine the causal effect of infrastructure on the spatial organization of economic activity, the central inference problem that researchers must overcome is that infrastructure is not assigned to locations at random, but is assigned rather on the basis of many of the same unobserved location characteristics that affect economic activity. The recent empirical literature is organized around three main approaches to this problem: planned route instrumental variable approach, historical route instrumental variable approach, and the inconsequential places approach. While these approaches remain open to criticism and refinement, they are about as good as can be hoped for in an environment where experiments seem implausible.

This literature suggests a number of tentative conclusions about the effects of infrastructure. Most studies estimate that population or employment density falls between 6% and 15% with a doubling of the distance to a highway or railroad (where railroads are the primary mode of transportation). Highways decentralize urban populations, and with less certainty, manufacturing activity. They may also lead to a complementary concentration of services. Different sectors appear to respond differently to different modes of transportation and people respond differently from firms. The effects of infrastructure seem similar across countries at different stages of development.

While much effort has been directed to unraveling the problem of nonrandom assignment of infrastructure to places, much less has been directed to distinguishing between growth and reorganization. This distinction is clearly central to any understanding of the role of infrastructure and transportation costs in an economy. We suggest two approaches to resolving this problem. The first is a two-equation generalization of the current single-equation reduced-form models. The second relies on our structural model to resolve this problem. With this said, the literature does suggest that much of the estimated effect of transportation costs and infrastructure on the spatial organization of economic activity is probably due to reorganization rather than growth. Refining our understanding of this issue seems an obvious place for further research.

In addition to the largely reduced-form literature currently available, structural models of transportation costs and the spatial organization of economic activity are beginning to appear. Structural models have the important advantage of allowing for estimates of general equilibrium effects, such as the migration of economic activity in response to changes in transportation costs, on the basis of theoretically founded estimating equations. They also have obvious advantages for welfare and counterfactual analysis: available results suggest the importance of the "share of trade with self" as an indicator of welfare. With this said, there is disagreement in the literature on the fundamental assumptions underlying these models: in particular, whether we should think of cities as drawing people from the countryside or as competing with other cities for residents. Resolving this issue appears to be an important prerequisite for further progress.

Finally, the existing literature has devoted little attention, empirical or theoretical, to the dynamics of how transportation infrastructure affects economic development. In particular, there are few panel data studies conducting impulse response estimates. This seems to be an important, though difficult area for further research.

ACKNOWLEDGMENTS

We are grateful to Chang Sun and Tanner Regan for excellent research assistance. We also thank Nate Baum-Snow, Gilles Duranton, Will Strange, Vernon Henderson, and participants at the conference for the *Handbook of Regional and Urban Economics* for excellent comments and suggestions. Responsibility for any opinions, errors and omissions lies with the authors alone.

REFERENCES

Ahlfeldt, G.M., Redding, S.J., Sturm, D.M., Wolf, N., 2014. The economics of density: evidence from the Berlin Wall. Technical report. NBER Working paper 20354.

Allen, T., Arkolakis, C., 2013. Trade and the Topography of the Spatial Economy. Yale University. Processed.

Alonso, W., 1964. Location and Land Use: Toward a General Theory of Land Rent. Harvard University Press, Cambridge, MA.
Anderson, J.E., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. Am. Econ. Rev. 93, 170–192.
Ángel Garcia-López, M., 2012. Urban spatial structure, suburbanization and transportation in Barcelona. J. Urban Econ. 72, 176–190.
Atalay, E., Hortacsu, A., Syverson, C., 2013. Vertical Integration and Input Flows. University of Chicago. Processed.
Banerjee, A., Duflo, E., Qian, N., 2012. On the road: transportation infrastructure and economic growth in China. Technical report. NBER Working paper No. 17897.
Baum-Snow, N., 2007. Did highways cause suburbanization? Q. J. Econ. 122, 775–805.
Baum-Snow, N., Kahn, M.E., 2005. Effects of urban rail transit expansions: evidence from sixteen cities, 1970–2000. Brook. Wharton Pap. Urban Affa. 2005 1 (4), 147–197.
Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A., Zhang, Q., 2012. Roads, Railroads and Decentralization of Chinese Cities. University of Toronto. Processed.
Bernhofen, D.M., El-Sahli, Z., Kneller, R., 2013. Estimating the Effects of Containerization on World Trade. University of Nottingham. Processed.
Billings, S.B., 2011. Estimating the value of a new transit option. Reg. Sci. Urban Econ. 41, 525–536.
Bleakley, H., Lin, J., 2012. Portage: path dependence and increasing returns in U.S. history. Q. J. Econ. 127, 587–644.
Bogart, D., 2009. Turnpike trusts and property income: new evidence on the effects of transport improvements and legislation in eighteenth-century England. Econ. Hist. Rev. 62, 128–152.
Brinkman, J.C., 2013. Transportation Technologies, Agglomeration, and the Structure of Cities. Philadelphia Federal Reserve Bank. Processed.
Bureau of Labour Statistics, 2012. Consumer expenditure survey: average annual expenditures and characteristics of all consumer units. www.bls.gov (accessed 2013-06-01).
Bureau of Transportation Statistics, 2012a. National transportation statistics: table 1–49, US ton-miles of freight. www.rita.dot.gov (accessed 2013-05-21).
Bureau of Transportation Statistics, 2012b. National transportation statistics: table 3–23, employment in for-hire transportation and selected related industries. www.rita.dot.gov (accessed: 2013-05-02).
Census Bureau, 2009. Commuting (journey to work): travel time tables. www.census.gov (accessed 2013-06-01).
Chandra, A., Thompson, E., 2000. Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. Reg. Sci. Urban Econ. 30, 457–490.
China Data Online, 2010. China statistical yearbook 2010: table 16–9, freight ton-kilometers. www.chinadataonline.org (accessed 2013-05-23).
China Data Online, 2011. China statistical yearbook 2011: tables 10–5 (urban) and 10–25 (rural). www.chinadataonline.org (accessed 2013-06-18).
Clark, X., Dollar, D., Micco, A., 2004. Port efficiency, maritime transport costs, and bilateral trade. J. Dev. Econ. 75, 417–450.
Combes, P.P., Duranton, G., Gobillon, L., 2012. The Costs of Agglomeration: Land Prices in French Cities. University of Pennsylvania. Processed.
Cosar, K., Fajgelbaum, P., 2013. Internal Geography, International Trade, and Regional Specialization. University of California, Los Angeles. Processed.
Couture, V., Duranton, G., Turner, M.A., 2012. Speed. University of Toronto. Processed.
Criscuolo, C., Martin, R., Overman, H.G., Reenen, J.V., 2012. The causal effects of an industrial policy. Technical report. CEP Discussion Paper No 1113.
Cronon, W., 1991. Nature's Metropolis: Chicago and the Great West. W. W Norton & Co, New York, NY.
Davis, M.A., Ortalo-Magne, F., 2011. Household expenditures, wages, rents. Rev. Econ. Dyn. 14, 248–261.
Davis, D.R., Weinstein, D.E., 2002. Bones, bombs, and break points: the geography of economic activity. Am. Econ. Rev. 92, 1269–1289.

Dekle, R., Eaton, J., Kortum, S., 2007. Unbalanced trade. Am. Econ. Rev. 97, 351–355.
Department for Transport, 2012. Table tsgb0306: greenhouse gas emissions by transport mode. www.gov.uk (accessed 2013-06-08).
Desmet, K., Rossi-Hansberg, E., 2013. Urban accounting and welfare. Am. Econ. Rev. 103 (6), 2296–2327.
Donaldson, D., 2015. Railroads of the Raj: estimating the impact of transportation infrastructure. Am. Econo. Rev. (forthcoming).
Donaldson, D., Hornbeck, R., 2013. Railroads and american economic growth: a "market access" approach. Technical report. NBER Working paper No. 19213.
Duranton, G., 2014. Roads and Trade in Columbia. University of Pennsylvania. Processed.
Duranton, G., Morrow, P., Turner, M.A., 2014. Roads and trade: evidence from the US. Rev. Econ. Stud. 81, 681–724.
Duranton, G., Turner, M.A., 2011. The fundamental law of road congestion: evidence from US cities. Am. Econ. Rev. 101, 2616–2652.
Duranton, G., Turner, M.A., 2012. Urban growth and transportation. Rev. Econ. Stud. 79, 1407–1440.
Eaton, J., Kortum, S., 2002. Technology, geography, and trade. Econometrica 70, 1741–1780.
Eurofound, 2000. Third European survey on working conditions 2000: figure 37. www.eurofound.europa.eu (accessed 2013-06-17).
Eurostat, 2009. Mean consumption expenditure by detailed COICOP level. www.epp.eurostat.ec.europa.eu (accessed 2014-05-18).
Eurostat, 2010. Transport database. www.epp.eurostat.ec.europa.eu (accessed 2013-05-01).
Eurostat, 2012. Extra EU27 trade since 1999 by mode of transport (NSTR). www.epp.eurostat.ec.europa.eu (accessed 2013-06-17).
Faber, B., 2015. Trade integration, market size, and industrialization: evidence from China's National Trunk Highway system. Rev. Econ. Stud. (forthcoming).
Fajgelbaum, P., Redding, S., 2013. External Integration, Structural Transformation and Economic Development: Evidence from Argentina. Princeton University. Processed.
Fogel, R., 1964. Railroads and American Economic Growth: Essays in Econometric History. Johns Hopkins University Press, Baltimore, MD.
Fujita, M., Krugman, P., 1995. When is the economy monocentric? Von Thünen and Chamberlain unified. Reg. Sci. Urban Econ. 25, 505–528.
Fujita, M., Ogawa, H., 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. Reg. Sci. Urban Econ. 12, 161–196.
Fujita, M., Krugman, P., Venables, A., 1999. The Spatial Economy: Cities, Regions, and International Trade. MIT Press, Cambridge, MA.
Garcia-Lopez, M.A., Holl, A., Viladecans-Marsal, E., 2013. Suburbanization and Highways: When the Romans, the Bourbons and the First Cars Still Shape Spanish Cities. Universitat Autonoma de Barcelona and IEB.
Ghani, E., Goswami, A.G., Kerr, W.R., 2013. Highway to success in India: The impact of the Golden Quadrilateral project for the location and performance of manufacturing. Technical report. World Bank. Policy Research Working paper 6320.
Gibbons, S., Machin, S., 2005. Valuing rail access using transport innovations. J. Urban Econ. 57, 148–169.
Glaeser, E.L., Kohlhase, J.E., 2004. Cities, regions and the decline of transport costs. Pap. Reg. Sci. 83, 197–228.
Glaeser, E.L., Ponzetto, G.A.M., 2013. Did the death of distance hurt Detroit and help New York? Technical report. NBER Working paper 13710.
Glaeser, E.L., Kahn, M.E., Rappaport, J., 2008. Why do the poor live in cities? The role of public transportation. J. Urban Econ. 63, 1–24.
Gonzalez-Navarro, M., Quintana-Domeque, C., 2013. Roads to Development: Experimental Evidence from Urban Road Pavement. University of Toronto. Processed.
Gordon, P., Willson, R., 1984. The determinants of light-rail transit demand—an international cross-sectional comparison. Transp. Res. A Gen. 18, 135–140.
Haines, M.R., Margo, R.A., 2008. Railroads and local economic development: the United States in the 1850s. In: Rosenbloom, J.L. (Ed.), Quantitative Economic History: The Good of Counting. Routledge, London, UK, pp. 78–99.

Head, K., Mayer, T., 2013. What separates us? Sources of resistance to globalization. Can. J. Econ. 46, 1196–1231.

Helpman, E., 1998. The size of regions. In: Pines, D., Sadka, E., Zilcha, I. (Eds.), Topics in Public Economics: Theoretical and Applied Analysis. Cambridge University Press, Cambridge, pp. 33–54.

Henderson, J.V., Storeygard, A., Weil, D.N., 2012. Measuring economic growth from outer space. Am. Econ. Rev. 102, 994–1028.

Hillberry, R., Hummels, D., 2008. Trade responses to geographic frictions: a decomposition using micro-data. Eur. Econ. Rev. 52, 527–550.

Hsu, W.T., Zhang, H., 2012. The fundamental law of highway congestion: evidence from national expressways in Japan. Technical report. Working paper, Department of Economics, The National University of Singapore.

Hummels, D., 1999. Towards a geography of trade costs. Technical report. GTAP Working paper No. 17.

Jedwab, R., Moradi, A., 2013. Transportation Technology and Economic Change: Evidence from Colonial Railroads and City Growth in Africa. George Washington University. Processed.

Knight, B., 2002. Endogenous federal grants and crowd-out of state government spending: theory and evidence from the federal highway aid program. Am. Econ. Rev. 92, 71–92.

Krugman, P.R., 1991. Increasing returns and economic geography. J. Pol. Econ 99 (3), 483–499.

Limao, N., Venables, A.J., 2001. Infrastructure, geographical disadvantage, transport costs and trade. World Bank Econ. Rev. 15, 451–479.

Lucas Jr., R.E., Rossi-Hansberg, E., 2002. On the internal structure of cities. Econometrica 70, 1445–1476.

Martincus, C.V., Carballo, J., Cusolito, A., 2012. Routes, Exports, and Employment in Developing Countries: Following the Trace of the Inca Roads. World Bank. Processed.

Mayer, T., Mayneris, F., Py, L., 2013. The impact of urban enterprise zones on establishments' location decisions: evidence from French ZFUs. Technical report. CEPR Discussion Paper 9074.

Maystadt, J.F., Duranton, G., 2014. The Development Push of Refugees: Evidence from Tanzania, Wharton. Processed.

McFadden, D., 1974. The measurement of urban travel demand. J. Publ. Econ. 3, 303–328.

Michaels, G., 2008. The effect of trade on the demand for skill—evidence from the Interstate Highway System. Rev. Econ. Stat. 90, 683–701.

Michaels, G., Rauch, F., Redding, S.J., 2012. Urbanization and structural transformation. Q. J. Econ. 127, 535–586.

Mills, E.S., 1967. An aggregative model of resource allocation in a metropolitan area. Am. Econ. Rev. (Pap. Proc.) 57, 197–210.

Moses, L.N., 1958. Location and the theory of production. Q. J. Econ. 72, 259–272.

Moses, L.N., Williamson, H.F., 1963. Value of time, choice of mode, and the subsidy issue in urban transportation. J. Pol. Econ. 71, 247–264.

Muth, R.F., 1969. Cities and Housing. University of Chicago Press, Chicago, IL.

Natural Resources Canada, 2012. Comprehensive energy use database table 8: GHG emissions by transportation mode. www.oee.nrcan.gc.ca (accessed 2013-06-03).

Neumark, D., Simpson, H., 2014. Place-based policies. In: Duranton, G., Strange, W., Henderson, V. (Eds.), Handbook of Regional and Urban Economics, vol. 5. Elsevier, Amsterdam, pp. 1197–1288.

North American Transportation Statistics, 2012a. Table 2–3: employment in transportation and related industries. www.nats.sct.gob.mx (accessed 2013-05-23).

North American Transportation Statistics, 2012b. Table 4–2: greenhouse gas emissions by transportation sector. www.nats.sct.gob.mx (accessed 2013-06-03).

North American Transportation Statistics, 2012c. Table 5–2: domestic freight activity. www.nats.sct.gob.mx (accessed 2013-05-23).

North American Transportation Statistics, 2012d. Table 7–1: international merchandise trade by mode. www.nats.sct.gob.mx (accessed 2013-06-03).

OECD Stat Extracts, 2012. National accounts: table 5, final consumption expenditure of households. www.stats.oecd.org (accessed 2013-06-08).

Office of National Statistics, 2009. Labour force survey: 11.6 time taken to travel to work by workplace. www.ons.gov.uk (accessed 2013-06-08).

Parry, I.W.H., Walls, M., Harrington, W., 2007. Automobile externalities and policies. J. Econ. Liter. 45, 373–399.
Ramondo, N., Rodriguez-Clare, A., Saborio, M., 2012. Scale Effects and Productivity: Does Country Size Matter? University of California at Berkeley. Processed.
Redding, S., 2012. Goods trade, factor mobility and welfare. NBER Working paper.
Redding, S., Sturm, D., 2008. The costs of remoteness: evidence from German division and reunification. Am. Econ. Rev. 98, 1766–1797.
Redding, S., Venables, A.J., 2004. Economic geography and international inequality. J. Int. Econ. 62, 63–82.
Redding, S.J., Sturm, D.M., Wolf, N., 2011. History and industry location: evidence from German airports. Rev. Econ. Stat. 93, 814–831.
Rossi-Hansberg, E., 2004. Optimal urban land use and zoning. Rev. Econ. Dyn. 7, 69–106.
Sarvimäki, M., Uusitalo, R., Jäntti, M., 2010. The long-term effects of forced migration. Aalto University. Processed.
Schafer, A., 2000. Regularities in travel demand: an international perspective. J. Transp. Stat. 3, 1–31.
Sheard, N., 2014. Airports and urban sectoral employment. J. Urban Econ. 80, 133–152.
Small, K.A., Verhoef, E.T., 2007. The economics of urban transportation. Routledge, New York, NY.
Statistics Canada, 2010. Commuting to work: results of the 2010 general social survey, table 1. www.statcan.gc.ca (accessed 2013-06-03).
Statistics Canada, 2012. Cansim: table 384-0041, detailed household final consumption expenditure. www.statcan.gc.ca (accessed 2013-06-03).
Storeygard, A., 2012. Farther on down the road: transport costs, trade and urban growth in Sub-Saharan Africa. Tufts University. Working paper.
Turcotte, M., 2005. The time it takes to get to work and back: table 1. www.publications.gc.ca (accessed 2013-06-03).