Quantitative economic geography meets history: Questions, answers and challenges*

Dávid Krisztián Nagy†

November 1, 2020

Abstract

A rapidly growing literature uses quantitative general equilibrium models of economic geography to study the economic impact of historical events such as the railroad revolution, industrial take-off, structural transformation and wars. I identify three key challenges facing this literature: the availability of historical data, the tractability of model structure, and issues related to identification. I review the literature by discussing how it has been addressing each of these challenges. While doing so, I point out the rich set of questions that this literature can address, as well as the methodological innovations it has conducted to answer these questions.

1 Introduction

How much economic growth was induced by railroads? How much did the falling costs of trade contribute to the current world income distribution? What was the economic impact of city bombing, of war-induced border changes, or population expulsions? How did 19th-century pollution reshape developed-world cities? Each of these questions calls for studying the economic consequences of large-scale historical events. And, although these events took place in the past, they might be relevant for today's economists for at least two reasons. First, they might provide explanations for why the current economy looks how it looks. For instance, understanding the origins of the world income distribution has been a long-held goal among researchers (Acemoglu, 2009). Second, similar large-scale events may take place in the future, and economists and policymakers naturally want to understand what they will induce. Pollution, for instance, is at least as much of a problem for Chinese cities today as for 19th-century cities in the developed world (Ebenstein et al., 2015).

What is also common across these large-scale historical events is that they are *inherently geographic*. Railroads connect certain locations, but not others. Bombing and pollution

*I am grateful to Stephan Heblich and Giacomo Ponzetto for helpful comments and suggestions. All errors are my own.

†CREi, Universitat Pompeu Fabra and Barcelona GSE. Email: dnagy@crei.cat.

affect certain parts of a city but might leave other parts intact. Border changes or population expulsions happen across certain points in space. This is not to say that other locations in the economy are unaffected by these events. They might experience indirect effects through their *spatial linkages* – trade, commuting, or factor mobility – with directly affected locations. As a result, any methodology aimed at studying the effects of these geographic events needs to embrace the notion of distinct locations connected through spatial linkages. Traditional general equilibrium macromodels do not have this feature, as they model the economy as a dimensionless point in space. This is precisely why researchers interested in inherently geographic questions moved towards modeling the economy as a set of locations, linked to one another through trade and factor mobility (Fujita and Thisse, 2002).

Another prominent methodology used to study the effects of geographic events is reduced-form empirics. A clear advantage of this approach over structural modeling is that it is devoid of the issue of model misspecification, which lends additional credibility to its results.[1] At the same time, the core principle of this methodology is the comparison of individuals or locations affected by an event (the "treated group") to those unaffected by the event (the "control group"). In a world with spatial linkages, it is likely that a large-scale event affects every location. In this case, reduced-form empirics can identify the effect of the event on locations designated as "treated" *relative* to locations designated as "control." As described in Redding and Turner (2015), such a relative effect might reflect the *relocation* of existing economic activity between treated and control locations, rather than actual (positive or negative) *growth* as a result of the event. Reduced-form empirical techniques cannot distinguish between such growth and relocation effects. As a consequence, unlike structural general equilibrium models, they cannot be used to measure the effects on the economy as a whole.

Quantitative general equilibrium models of economic geography are able to combine the advantages of structural macromodels and reduced-form empirics. On the one hand, they embrace the realism of multiple locations connected through complex spatial linkages. This allows the researcher to combine the model with rich spatial data, similar to empirics. On the other hand, quantitative modeling, structural by nature, allows us to measure the effects of large-scale events on the whole economy and distinguish growth from the mere relocation of economic activity.

Needless to say, these advantages of quantitative modeling come at a cost. In particular, using quantitative models to study history imposes three key challenges on the researcher. One of these challenges stems from the sparse nature of historical data. Another is that the realism of model structure often comes at the expense of tractability. A final challenge, shared with reduced-form empirics, is that identifying the effects of events may be subject

[1] This argument is made, for instance, in Baum-Snow and Ferreira (2015), who also discuss the methodological challenges faced by the empirical economic geography literature.

to biases due to omitted variables and endogeneity.

In the remainder of this article, I review the rapidly growing literature that uses quantitative general equilibrium models of economic geography to study the economic impact of large-scale historical events. I do so by discussing the above three challenges and the ways the literature has been addressing each of them. Thus, Section 2 focuses on the scarcity of historical data, Section 3 discusses the challenge of model tractability, while Section 4 looks at issues related to identification. Section 5 concludes the article by suggesting avenues for further research in the field.

2 Availability of historical data

Available data become sparser as one goes back in time. This is by no means a problem that only economists face. In archaeology, conclusions normally need to be drawn from scattered, small remains. Identifying the sex and age of skeletal remains, for instance, relies primarily on examining the pelvis, the skull, and other large bones. Quite often, such remains are not available. In that case, the calcification of teeth and the maturation of non-cranial bones may provide information on age, while comparing the two may be informative about sex (Ubelaker, 2008). However, even these might be absent. What is left of the entire pre-human species *Australopithecus bahrelghazali*, for example, is three partial jawbones and a tooth (Brunet, 2010).

Similarly, economic geographers studying historical questions need to draw conclusions from a small amount of sparse data, which become increasingly sparser as one goes back to the more and more distant past. If anything, the problem is even more severe in the case of quantitative models, as taking such models to the data typically requires macroeconomic data such as GDP, population or land values at the location level. In what follows, I briefly review the availability of historical data sources on GDP, sectors and occupations, land values, population, transportation networks, commuting and geographic characteristics that have been used in the quantitative economic geography literature. Next, I discuss what Geographic Information Systems (GIS) and structural modeling can add to the information embedded in these sparse data.

GDP. National accounting is almost exclusively a 20th-century phenomenon. Official annual estimates of national income were first compiled by the Australian government in 1886, followed by Canada in 1925 (Bos, 1992). With Simon Kuznets joining the National Bureau of Economic Research in 1929, the United States soon became the leader in preparing national income estimates. In 1939, the U.S. Commerce Department produced the first estimates of income at the state level (Carson, 1975). However, it was not until 1947 that the first international guidelines on national accounting were prepared (Bos, 1992). Thus, official data on GDP are virtually unavailable for the study of historical questions that predate the mid-20th century.

That said, economic historians have produced estimates of past GDP based on available population, income and production data. One notable example is the Maddison database, which includes estimates of country GDP going back to the year 1 (Bolt et al., 2018). In the case of the U.S., Easterlin (1960), Gallman (1966), David (1967) and Weiss (1992) estimate per capita income of the country and its regions for various overlapping periods during the 19th century, relying primarily on sectoral data. The vast differences across their estimates already suggest that they are likely subject to large measurement errors. Hence, using them as direct inputs into quantitative models could cast doubt on the model's quantitative predictions. Nevertheless, such estimates of historical GDP might still be suitable for testing the model's qualitative predictions. For example, Nagy (2020) finds that, according to a spatial model of the 19th-century U.S. economy, the U.S. Northeast had higher per capita income than the South and the Midwest before the Civil War. This pattern of income distribution across U.S. regions is in line with the above historical estimates. Such tests, even though they do not force us to fully trust historical estimates of GDP, can lend additional credibility to models that can replicate similar broad patterns observable in the estimates.

Sector- and occupation-level data. The 19th century saw the first sectoral censuses conducted. These censuses tend to provide data on employment, the number of establishments, land use, and/or output in certain important sectors such as agriculture or manufacturing. Quite often, however, they do not cover the entire economy.

In the United States, James Madison suggested that the first, 1790 population census include occupational statistics. Though this did not happen in the end, the 1810 census asked enumerators to collect data on the country's manufacturing establishments. This became the first U.S. Census of Manufactures, and it was followed by similar censuses conducted in 1820 and 1840. However, the quantity and quality of the collected data in these first three censuses disappointed contemporaries.[2] Major improvements were made afterwards, which makes the 1850 and subsequent decennial Censuses of Manufactures substantially more reliable than the previous ones (Fishbein, 1973). In the meantime, the first U.S. Census of Agriculture was conducted in 1840,[3] and was followed by decennial censuses of this sector since then. The quantitative economic geography literature has used these historical Censuses of Agriculture and Manufactures. In particular, Donaldson and Hornbeck (2016) use Census of Agriculture data to measure the effect of late-19th century railroads on the value of agricultural land. Eckert and Peters (2018), on the other hand,

[2]One reason for this is that the instructions given to enumerators were vague, especially in 1810. Another reason is that several establishments refused to provide data, fearing that the information would be used for tax purposes.

[3]The 1840 census already includes detailed data on livestock, prices and output by crop, as well as estimates of the value of agricultural production by county (U.S. Census, 1841).

use Census of Manufactures data from 1880 on to study the spatial patterns of structural change towards manufacturing.

Another country with a long history of sectoral employment data is the United Kingdom. U.K. population censuses included a question about occupation from 1851 on. Moreover, Shaw-Taylor et al. (2010) compile data on the occupation of males in 1710 (for England) and 1817 (for England and Wales) based on thousands of baptism records. The occupational categories are such that the sector can be inferred from them; examples of occupations are "grower of minor crops," "coal miner," "textile products maker" and "rag dealer." Trew (2014) and Trew (2020) use these occupational data at the level of *registration districts*[4] to study spatial structural change in England during the industrial revolution.[5]

Value of land and structures. The value of land and buildings on farms were included in the U.S. Census of Agriculture from 1850 on. Naturally, these reflect not only land values but also the value of capital embedded in buildings. To separate the two, Fogel (1964) estimates the value of agricultural land alone at the state level. Donaldson and Hornbeck (2016) use these estimates to impute the value of agricultural land at the county level, assuming that the value of buildings relative to land was uniform within states.

Historical data on the value of non-agricultural land and structures are also sparse. In England and Wales, however, the value of properties were constantly assessed by the government since 1601 for tax purposes. These so-called *rateable values* again reflect both the value of land and the value of structures built on it. To study the economic impact of the steam railway on London, Heblich, Redding and Sturm (2020) use these rateable values by borough in the Greater London area. In the case of Berlin, Ahlfeldt et al. (2015) use high-resolution data on the value of undeveloped land (i.e., without structures) collected by a government-appointed building surveyor (Kalweit, 1937) to examine the effect of the Berlin Wall on the city's economy.

Population. Due to the long history of government-administered population censuses, population data tend to be the most widely available and the highest in quality.[6] As a result, these data have been widely used in the quantitative economic geography literature. Historical U.S. population data at the county level are used in Allen and Donaldson (2018), Desmet and Rappaport (2015), Donaldson and Hornbeck (2016), Eckert and Peters (2018) and Nagy (2020), among others. Census-based population data at similar levels of geographic disaggregation are used in quantitative historical studies on Argentina (Fajgelbaum and Redding, 2018), Germany (Peters, 2019; Santamaria, 2020) and the United

[4]The average registration district in England and Wales had 28,700 inhabitants in 1851.

[5]Further quantitative economic geography papers use historical sector-level data from Argentina (Fajgelbaum and Redding, 2018), Germany (Peters, 2019) and India (Donaldson, 2018).

[6]This does not mean that population data cannot be patchy sometimes. 1850 census data on San Franscisco were destroyed in a fire, while data on Contra Costa and Santa Clara counties were lost on the way to the San Francisco office (U.S. Census, 1852).

Kingdom (Trew, 2014; Trew, 2020). Normally, census data also allow for distinguishing urban from rural population. This has been an advantage for papers investigating the historical drivers of urbanization (Fajgelbaum and Redding, 2018; Nagy, 2018; Nagy, 2020; Redding and Sturm, 2008).

Historical census data can also be used to examine the drivers of the distribution of population within cities. Within-city analysis necessitates a distiction between *residential* population (i.e., number of people by location of residence) and *workplace* population (i.e., number of employed people by workplace location). Data on the former are more widely available, as censuses are normally conducted on residential locations. Heblich, Redding and Sturm (2020), for instance, use residential population by borough in Greater London from 1801 until 1921 to document that the steam railway allowed people to move away from downtown areas. To study the effect of the Berlin Wall, Ahlfeldt et al. (2015) use street-level residential population data on Berlin from the 1933 census. They, however, also observe workplace population. This is because company-level employment data and a registry of company locations allow them to construct employment data at the level of census blocks (Ahlfeldt et al., 2015).

Population censuses often include data on socioeconomic status such as education, occupation and – more recently – income. As a result, they can also be used to examine the drivers of residential segregation. Various studies have documented that residential segregation patterns exhibit persistence over time. That is, neighborhoods that were poor in the past also tend to be poor today. Multiple explanations have been offered for this phenomenon, based on persistent natural amenities (Lee and Lin, 2018), differences in 19th-century pollution across Western and Eastern parts of cities (Heblich, Trew and Zylberberg, 2020), and the war-time bombing of certain neighborhoods (Redding and Sturm, 2016).

Finally, certain research questions ask for population data that predate the official censuses. Similar to GDP, the Maddison project estimates country populations back to the year 1 (Bolt et al., 2018). At a higher level of spatial disaggregation, the *History Database of the Global Environment* (HYDE) combines an assortment of historical sources to estimate the population of each 5 by 5 arc minute grid cell of the Earth, going back 12,000 years (Klein Goldewijk, Beusen and Jansen, 2010). Delventhal (2018) uses these data for the year 1000, aggregated up to a 3° by 3° resolution, to calibrate a dynamic spatial model of the world economy. He uses the model to quantify the extent to which falling trade costs in the last 1000 years contributed to the current world income distribution.

Transportation links. A substantial portion of the quantitative economic geography literature has pointed out the importance of spatial linkages across locations. First, spatial linkages might be crucial in transmitting the effects of historical events from locations directly affected by the event to other locations in the economy. Second, the historical event that the researcher aims to evaluate might be the establishment of a spatial linkage itself. For instance, a large literature has studied the general equilibrium effects of U.S. railroads, starting with the seminal work of Fogel (1964).[7] Conducting such an exercise requires collecting a database of transportation links across locations. Such a database essentially consists of a series of maps, one by time period and mode of transportation (such as water, rail and road; or, in the case of within-city transportation, walking, car and public transport). Using data on the cost of transportation by mode, the next task is determining the overall transport cost between any pair of locations in each time period. This is normally done using least-cost algorithms (Dijsktra algorithm, Fast Marching algorithm).

Some studies use data on the location of planned, rather than actual, transportation links. For example, Santamaria (2020) uses planned highways in 1930s Germany to calibrate an economic geography model with endogenous infrastructure placement. Brinkman and Lin (2019) use planned freeways as instruments for the location of actual freeways, which are endogenously affected by opposition to construction due to freeway disamenities.

Commuting and other travel flows. Commuting data are useful to calibrate quantitative models of within-city geography, as such data create a mapping between residential and workplace locations. Heblich, Redding and Sturm (2020), for instance, observe 1921 commuting data across boroughs in the city of London. They combine these data with population by borough to recover the location of people's workplaces in London, which are unobserved. Unfortunately, such comprehensive commuting data are rarely available in other historical contexts. However, it is sometimes possible to rely on survey data. For example, Brinkman and Lin (2019) use data from 1950s travel surveys for Chicago and Detroit to show that freeways acted as barriers to travel flows aimed at the opposite side of the freeway.[8]

Geographic characteristics. It should come as no surprise that various studies in quantitative economic geography rely on data on natural geographic characteristics. First, natural features such as rivers and oceans had an influence on transportation routes, particularly during historical times. Thus, they serve as important inputs to the creation of

[7] Fogel's pioneering idea was that the aggregate impact of railroads on U.S. agriculture needs to be assessed by comparing the real-world economy to a hypothetical economy in which the transportation of agricultural goods can only happen through other modes – an exercise that today's economists call a *counterfactual*. Motivated by Fogel's analysis, Donaldson and Hornbeck (2016) conduct the same exercise in a quantitative economic geography model. Similar to Fogel, they find that the effect of railroads on U.S. agriculture was rather modest. However, Pérez-Cervantes (2014) finds, conducting a large number of counterfactuals involving alternative railroad networks in a similar model, that certain rail connections were much more influential than others. Finally, using a dynamic spatial model of city formation, Nagy (2020) argues that, even with a limited effect on the agricultural sector itself, railroads had a large aggregate impact on the U.S. economy by fostering the development of cities.

[8] Historical trade data at a higher spatial resolution than cross-country flows are even harder to find. To study the role of trade in Bronze Age city formation, Barjamovic et al. (2019) use the number of mentions on clay tablets as a proxy of trade flows across ancient Assyrian cities.

transportation links databases (Delventhal, 2018; Donaldson and Hornbeck, 2016). Second, nature matters for locations' agricultural potential. High-resolution spatial datasets on agricultural potential include the Food and Agriculture Organization's Global Agro-Ecological Zones database (FAO GAEZ), the Caloric Suitability Index (Galor and Özak, 2016), and the index of agricultural land suitability built by Ramankutty et al. (2002). These datasets are used as inputs to quantitative historical studies such as Delventhal (2018) and Nagy (2020).

Spatial data are often processed in Geographic Information Systems (GIS). A Geographic Information System is a versatile computer-based tool that helps with the "storage, analysis, output, and distribution of spatial data and information" (Bolstad, 2005). As the quote highlights, GIS is useful for at least three purposes. First, it allows for the efficient *storage* of high-resolution spatial data. Second, it can perform various sorts of *analysis* on spatial data, such as merging data up to higher units of disaggregation, measuring distances across spatial units, creating distance buffers around them, and so on. Finally, GIS is especially well-suited to creating nice visual *output* of spatial data, such as maps. Due to these advantages of GIS, various historical spatial datasets are primarily made available in GIS-compatible format. For instance, the *National Historical Geographic Information System* (NHGIS) is a free online source of GIS-compatible U.S. census and survey data that go back to 1790 (Manson et al., 2017).

Needless to say, no processing tool can compensate for the sparse nature of historical spatial data. One advantage of quantitative modeling, however, is that the model itself can do this job. One notable example is Barjamovic et al. (2019), who combine the extremely patchy data available from ancient Assyria (a proxy of trade flows across Assyrian cities and the locations of cities that have been found by archaeologists) with a quantitative trade model to predict the locations of cities that have not yet been found. A general issue with this approach is that the structure of the model, which identifies the unobserved data, is an assumption that is untestable in practice. Nonetheless, even though one cannot test the structure of the model in its entirety, one can test some of its implications on observed outcomes. For example, Barjamovic et al. (2019) contrast their model's predictions on merchants' itineraries to actual itineraries found. Although they do not use the actual itineraries in the model's quantification, they find a good fit between the model-implied and the actual itineraries. Moreover, they find that their location predictions for lost cities often coincide with archaeologists' conjectures, and that their model-based method of locating cities works well when predicting the locations of known cities. Such tests are crucial to show that the model is a credible enough tool to fill the gaps in sparse historical data.

3 Model tractability

Incorporating the complex nature of real-world spatial interactions often comes at the expense of tractability in quantitative modeling. To highlight the specific model ingredients that might prove challenging from a tractability perspective, I proceed in this section as follows. In Section 3.1, I present a class of quantitative spatial models with trade, labor mobility and commuting that, even though not solvable in closed form, prove tractable in various respects. Specifically, (1) the existence and uniqueness of the model's equilibrium can be characterized theoretically; (2) a simple algorithm can be used to solve for the equilibrium on the computer; (3) unobserved location fundamentals that rationalize the observed data as an equilibrium can be uniquely recovered for given structural parameters (that is, the model can be *inverted*). In Section 3.2, I consider additional model ingredients absent from this class of models: dynamics, multiple sectors, and endogenous infrastructure development. I discuss how these additional model ingredients can pose challenges to tractability. I also point out how existing quantitative historical studies that incorporate some of these additional ingredients have addressed these challenges.

3.1 A tractable class of quantitative spatial models

In this section, I develop a quantitative spatial model of a set of locations, where locations are linked through trade, labor mobility and commuting (Section 3.1.1).[9] Next, I show that this model features a tractable structure; more precisely, the existence and uniqueness of the model's equilibrium can be characterized theoretically, the equilibrium can be computed using a simple algorithm, and the model can be inverted to recover unobserved location fundamentals that rationalize the observed data as an equilibrium (Section 3.1.2). Finally, I show isomorphisms between the model and a set of other models, some of which have been used in the quantitative economic geography literature (Section 3.1.3). As a result of these isomorphisms, the tractability of model structure I show for the model of Section 3.1.1 applies to this entire class of quantitative spatial models.

3.1.1 Model setup

The economy consists of a discrete set of S locations, indexed by r, s, or u. $\bar{L} > 0$ workers inhabit the economy, where $\bar{L}$ is exogenously given. There is one sector producing tradable goods. Within this sector, each location produces one good that workers view as different from the goods produced at other locations.[10] Hence, I index each tradable good by the index of its production location. Besides tradables, workers also consume

[9] The model is closest in its structure to Monte, Redding and Rossi-Hansberg (2018), Heblich, Redding and Sturm (2020), and Allen and Arkolakis (2014).

[10] In trade and geography, this is called the Armington assumption (Anderson, 1979). Section 3.1.3 shows isomorphisms between the model and alternative models in which I relax this assumption.

housing, a homogenous nontradable good that is available in exogenous positive supply at each location. Housing payments go to immobile local landlords, who spend their entire income on tradables and have the same preferences over tradables as workers (Monte et al., 2018).

Consumption and location choice. Each worker chooses a residential location to live. They also choose a – potentially different – workplace location, where they inelastically supply the one unit of labor they own. Workers are atomistic, implying that they take wages, the prices of tradables and the price of housing as given at every location.

If worker i chooses to live at location r and work at location s, she obtains utility

$$U_i(r,s) = a_i(r,s)\,\kappa(r,s)^{-1}\left[\sum_{u=1}^{S} q_i(r,s;u)^{\frac{\sigma-1}{\sigma}}\right]^{\nu\frac{\sigma}{\sigma-1}} H_i(r,s)^{1-\nu}$$

where $a_i(r,s)$ denotes the amenities enjoyed by the worker at r and s, $\kappa(r,s) \geq 1$ is the cost of commuting between residence r to workplace s,[11], $H_i(r,s)$ is the quantity of housing consumed by the worker at her residence r, and $q_i(r,s;u)$ is her consumption of tradable good u, where σ is the elasticity of substitution across goods. I make the standard assumption that $\sigma > 1$. That is, tradables are imperfect substitutes.

Amenities are drawn from a Fréchet distribution that is independent across workers and residence-workplace pairs. This is a standard assumption in the literature (Ahlfeldt et al., 2015; Redding, 2016), and captures the idea that workers have different idiosyncratic tastes for different locations. More precisely, I assume that the cumulative distribution function of $a_i(r,s)$ takes the form

$$Pr\left[a_i(r,s) \leq a\right] = e^{-\bar{a}(r,s)N(r)^{-\lambda}a^{-\epsilon}}$$

where $\bar{a}(r,s) > 0$ is the exogenous fundamental amenity level of the pair (r,s), $N(r)^{-\lambda}$ is a congestion disamenity that depends on the population of the residential location r, $N(r)$, with elasticity $\lambda \geq 0$,[12] and $\epsilon > 0$ is a parameter driving the dispersion in idiosyncratic location tastes.

Production and trade. Each tradable good is produced by a large number of perfectly competitive firms that face a constant returns to scale production technology. As a result, there exists a representative firm. The representative firm of location s faces the production technology

$$q(s) = A(s)\,L(s)$$

[11] Commuting between two locations can be infinitely costly, in which case $\kappa(r,s) = \infty$. Thus, the model embeds no commuting as a special case ($\kappa(r,s) = \infty$ for any $r \neq s$).

[12] A generalization of the model would allow for the disamenity to depend on both residential and workplace population: $N(r)^{-\lambda_1} L(s)^{-\lambda_2}$, where $L(s)$ denotes the number of people working at r. All the theoretical results of Section 3.1.2 carry over to this more general case.

where $q(s)$ is the firm's output, $A(s)$ is the location's productivity that the firm takes as given, and $L(s)$ is employment. Employment at s potentially has an effect on productivity:

$$A(s) = \bar{A}(s) L(s)^{\alpha},$$

a relationship that the firm does not internalize. This is a formulation of agglomeration externalities that is standard in the literature (Ciccone and Hall, 1996; Allen and Arkolakis, 2014). $\bar{A}(s) > 0$ is the fundamental productivity level of location s, which is exogenous.

Trade in tradable goods is subject to *iceberg* costs. That is, if tradable goods start to be shipped from a location to another, only a fraction of these goods arrives. In particular, $\tau(s, r) \geq 1$ units of good s need to be shipped from the production location s so that one unit arrives at r. The remaining units melt away in transit. As firms take all prices as given, no arbitrage guarantees that the ratio of the good's price between r and s also equals $\tau(s, r)$, as long as the good is traded between these two locations.[13]

Equilibrium. An equilibrium of the model is a set of prices and quantities such that workers choose their consumption levels as well as residence and workplace locations to maximize their utility; landlords choose their consumption levels to maximize their utility; firms choose their employment and output to maximize their profits; and labor, housing and tradable goods markets clear. Section A.1 of the Appendix shows that the equilibrium conditions of the model can be reduced to a system of $5S + S^2 + 1$ equations. To compute the equilibrium, the researcher needs to solve these equations for $5S + S^2 + 1$ endogenous variables; namely, for the residential population, the employment level, the wage level, the spending on tradables and the ideal price index at each location, the commuting flows between each pair of locations, as well as the economy-wide level of workers' expected utility.

3.1.2 Model tractability: theoretical results

This section presents three theoretical results that highlight the tractability of the model. Theorem 1 shows that the existence and uniqueness of the model's equilibrium are guaranteed by a condition that only depends on the model's structural parameters. As I argue, this condition tends to hold if the model's agglomeration force is not too strong. Next, Theorem 2 shows that the theoretical condition for equilibrium uniqueness also guarantees that a simple algorithm can be used to solve for the model's equilibrium on the computer. Finally, Theorem 3 shows that, for a given set of structural parameters, the model inversion identifies a unique set of unobserved amenities and productivities (up to scale) that rationalize the observed data as an equilibrium.

[13] Similar to commuting costs, trade can be prohibitively costly across certain pairs of locations, in which case $\tau(s, r) = \infty$.

In the proofs of Theorems 1 to 3, I rely heavily on the set of theoretical results presented in Allen, Arkolakis and Li (2020). For brevity, the proofs are relegated to Section A.2 of the Appendix.

Theorem 1. *The model's equilibrium exists and is unique under a condition that only depends on the model's structural parameters.*

The Appendix presents the specific condition under which equilibrium existence and uniqueness are guaranteed. I have not included the condition in the main text as it is a complex function of matrices whose entries depend on the model's structural parameters. Intuitively, the condition is more likely to hold if the model's agglomeration force (the agglomeration externality in production) is weak relative to the congestion forces (housing, the congestion disamenity, and the dispersion of idiosyncratic location tastes). This result should not be surprising. Under strong agglomeration forces, the concentration of economic activity at a certain location can sustain itself, which may give rise to different equilibria in which concentration arises at different locations.

If the equilibrium is guaranteed to be unique, then the researcher can be sure that a numerical procedure that has found an equilibrium has found the *only* equilibrium of the model. In general, however, a uniqueness result does not suggest a numerical procedure that can be used to find the equilibrium. That said, the following theorem suggests such a procedure. Moreover, it shows that the procedure is guaranteed to find the equilibrium if the condition for uniqueness in Theorem 1 holds.

Theorem 2. *Assume that the sufficient condition for existence and uniqueness from Theorem 1 holds. Then a simple iterative algorithm is guaranteed to converge to the equilibrium spatial distribution of residential population, employment, wages, spending on tradables and price indices. Finally, workers' expected utility and commuting flows can be obtained in closed form as a function of the former distributions.*

The simple iterative algorithm, as shown in Section A.2 of the Appendix, consists of guessing any initial distribution of residential population, employment, wages, spending on tradables and price indices, and then updating these distributions using the model's equilibrium conditions until convergence.

Although the convergence of the simple iterative algorithm is guaranteed by Theorem 2, the algorithm would be of little practical use if it were slow. However, Allen et al. (2020) show that the rate of convergence for such algorithms depends on how far the values of structural parameters are from the boundary of the region in which the condition of Theorems 1 and 2 holds. In practice, such algorithms tend to be very quick unless the values of structural parameters are extremely close to this boundary.

Theorems 1 and 2 provide a powerful tool to the researcher. As long as the values of the model's structural parameters guarantee uniqueness (Theorem 1), the researcher can use the simple iterative algorithm of Theorem 2 to compute the equilibrium spatial distribution of economic activity under any trade costs, commuting costs, and so on. For instance, she can simulate a counterfactual world in which a new railroad is built to connect particular locations.

Of course, the results of such counterfactual exercises depend on the underlying distribution of fundamental amenities and productivity. In practice, these fundamentals are unobserved. Thus, it is important to recover their spatial distributions before conducting counterfactual exercises with the model. A standard method for this consists of finding the distributions of fundamentals that rationalize the real-world data as an equilibrium (*model inversion*).

To invert the model, the researcher should ideally know if there exists a unique set of fundamentals that rationalize the observed data. Under uniqueness, if she finds a distribution of fundamentals that rationalize the data, she can be assured that it is the *only* distribution that does so. Theorem 3 shows that this sort of uniqueness always holds in this model.

Theorem 3. *Assume that the researcher observes the values of structural parameters, the matrix of commuting flows $L(r,s)$, commuting costs $\kappa(r,s)$ and trade costs $\tau(r,s)$, as well as wages $w(s)$ at every location. Then the researcher can recover the set of fundamental amenities $\bar{a}(r,s)$ and productivities $\bar{A}(s)$ uniquely (up to scale).*[14]

Although Theorem 3 guarantees the uniqueness of model fundamentals (conditional on structural parameters and the observed data), it does not offer an algorithm to compute the distribution of these fundamentals, such as the algorithm provided by Theorem 2. In fact, one can show that the simple iterative algorithm of Theorem 2 is not guaranteed to work for the model inversion. That said, there exist relatively minor departures from it that can be used. One example is the algorithm used to invert the model in Desmet, Nagy and Rossi-Hansberg (2018). This algorithm relies on an approximation to the simple iterative algorithm that is guaranteed to converge to the equilibrium distribution of fundamentals.

3.1.3 Isomorphisms with other quantitative spatial models

In this section, I show that the tractability results of Section 3.1.2 carry over to an entire class of quantitative economic geography models that make different assumptions on consumption, production, and market structure. I show this by presenting a set of isomorphisms between the model of Section 3.1.1 and models featuring these alternative assumptions. Finally, I point out which models used in the quantitative economic geography

[14]Up to scale means that the researcher cannot recover the worldwide level of either fundamental. For instance, multipliying each location pair's amenity level by the same constant yields the same equilibrium (except that each worker's unobserved utility is multiplied by a constant).

literature belong to this class. The formal proofs of the isomorphisms are relegated to Section A.3 of the Appendix.

Land use in production. I first present an isomorphism between the model of Section 3.1.1 and a model in which the production of tradables requires two inputs: labor and land. The two inputs enter the production function with Cobb–Douglas shares μ and $1-\mu$:

$$q(s) = A(s) L(s)^{\mu} \Lambda(s)^{1-\mu}$$

where $\Lambda(s)$ is the quantity of land used by the firm. Land is available in fixed positive supply at each location. In equilibrium, land rents adjust at every location until the local land market clears. Rents are redistributed to local employees with equal shares. Thus, the income of an employee at location s equals her wage income plus a $1/L(s)$ fraction of rents at s.

Next, I show isomorphisms between the model of Section 3.1.1 and two models in which locations endogenously specialize in the set of tradable goods they produce. In other words, I relax the assumption that each location produces its own good.

Endogenous specialization, driven by comparative advantage. In this model, specialization is driven by differences in comparative advantage, as in Eaton and Kortum (2002). There exists a continuum of tradable goods, indexed by $\omega \in [0,1]$. Workers have constant elasticity of substitution (CES) preferences over tradables:

$$U_i(r,s) = a_i(r,s)\,\kappa(r,s)^{-1}\left[\int_0^1 q_i(r,s;\omega)^{\frac{\sigma-1}{\sigma}}\,d\omega\right]^{\nu\frac{\sigma}{\sigma-1}} H_i(r,s)^{1-\nu}$$

and purchase each tradable good ω from the location that offers the good at the lowest price at r (including trade costs). Immobile local landlords consume tradables only, and have the same preferences over them as workers.

Every location hosts a large number of perfectly competitive firms that can produce any tradable good ω. However, locations are not equally well-suited to producing a given good. In particular, the productivity of a good is heterogeneous across locations. As in Eaton and Kortum (2002), I assume that the productivity of good ω at location s, $A_\omega(s)$, is drawn from a Fréchet distribution, independent across goods and locations:

$$Pr\left[A_\omega(s) \leq z\right] = e^{-[z/A(s)]^{-\theta}}$$

where $\theta > \sigma - 1$ is a parameter driving the dispersion of productivity draws, and $A(s) > 0$ is the productivity of the location. As a result of these assumptions, locations specialize in the set of tradable goods they are most productive at relative to other locations.

Endogenous specialization, driven by increasing returns. Finally, I show an isomorphism between the model of Section 3.1.1 and a model in which locations' endogenous specializiation in tradables is driven by increasing returns, as in Krugman (1991). There exists a continuum of tradable goods, indexed by $\omega \in [0, N]$, where the set of tradables N is determined endogenously. As in the previous model, workers have constant elasticity of substitution (CES) preferences over tradables:

$$U_i(r,s) = a_i(r,s)\,\kappa(r,s)^{-1}\left[\int_0^N q_i(r,s;\omega)^{\frac{\sigma-1}{\sigma}}\,d\omega\right]^{\nu\frac{\sigma}{\sigma-1}} H_i(r,s)^{1-\nu}$$

while landlords have the same preferences over tradables.

Firms produce tradable goods with a technology that is subject to increasing returns, as in Krugman (1991). In particular, producing $q_\omega(s)$ units of good ω at location s requires $q_\omega(s)/A(s)$ units of labor, plus an additional fixed $f > 0$ units of labor required for start-up. This latter element of the technology is a fixed cost, which implies that production is subject to increasing returns.

Firms are aware that consumers differentiate across tradable goods. As a result, each good is only produced by one firm in equilibrium, and the firm uses its monopoly power to set a price of its good above the marginal cost. Firms are, however, also aware that they are atomistic relative to the whole set of tradables produced, and therefore cannot influence location-wide prices (monopolistic competition). In equilibrium, each firm operates at its efficient size, implying that its labor demand is a function of local productivity and structural parameters only. Local labor market clearing then determines the equilibrium mass of tradables that a location specializes in.

As a result of these isomorphisms, special cases of the model are isomorphic to a whole range of existing models in quantitative economic geography. In particular,

- the model of Section 2 in Redding (2016) is isomorphic to a special case of the model with specialization through comparative advantage in which there are no agglomeration externalities ($\alpha = 0$), no congestion disamenities ($\lambda = 0$) and no commuting;

- the model of Section 3 in Redding (2016) is isomorphic to a special case of the model with specialization through increasing returns in which there are no congestion disamenities and no commuting;

- the model of Redding and Sturm (2008) is isomorphic to a special case of the model with specialization through increasing returns in which there are no congestion disamenities, no commuting and no location taste heterogeneity (corresponding to the limiting case of $\epsilon \to \infty$);

- the model of Allen and Arkolakis (2014) is isomorphic to a special case of the model with no housing ($\nu = 1$) and no commuting;

- the model of Monte et al. (2018) is isomorphic to a special case of the model with specialization through increasing returns in which there are no congestion disamenities.

As a consequence of these isomorphisms, the tractability results of Section 3.1.2 carry over to the above frameworks as well, underscoring again that a broad class of quantitative economic geography models exhibit tractability. This is true despite the fact that these models incorporate various dimensions of real-world heterogeneity, such as differences in amenities, productivity, trade costs, and commuting costs across locations. The next section discusses how incorporating further realism might prove challenging from a tractability perspective.

3.2 Additional model ingredients: challenges to tractability

In this section, I consider a set of model ingredients that are absent from the class of quantitative models discussed in Section 3.1. My primary focus is on how these additional ingredients can pose challenges to tractability, and how existing quantitative studies have addressed these challenges.

Dynamics. Many historical questions are dynamic by nature: How did transport infrastructure foster economic growth? How persistent has the spatial distribution of economic activity been? Answering these questions with a quantitative spatial model necessarily requires incorporating dynamics in the model. It is not hard to see why incorporating dynamics can be challenging from the point of view of tractability. In a model in which the future enters economic agents' objective functions and influences the decisions they make today, agents need to predict the future evolution of prices. In a model with spatial linkages across locations, the problem is even more complex: agents need to predict the evolution of prices *at each location*. Without simplifying the problem, the curse of dimensionality makes it infeasible to compute the equilibrium, even if the number of locations is small.

Various simplifications have been offered to this general dynamic problem. One possible simplification is, of course, assuming that agents do not care about the future and thus make static decisions at every point in time. Under this assumption, the model still needs to be solved for every time period, but it can be solved as a sequence of static problems. A case in which such an assumption is justifiable is one in which a time period is about as long as a person's lifetime (Delventhal, 2018). A similar assumption can be made if agents live for two time periods, but their decision with dynamic consequences is made in only one of those two periods (Allen and Donaldson, 2018). In the case of firms, a possible microfoundation for the assumption that agents do not care about the future is that irrespectively of their decisions today, their future profits are driven down to zero by free entry (Desmet and Rossi-Hansberg, 2014; Nagy, 2020). Another simplifying assumption can be the absence of trade costs, in which case prices equalize across locations and agents only need to predict the future evolution of one worldwide price (Eckert and Peters, 2018).

Multiple sectors. The class of models discussed in Section 3.1 feature only one sector, but various historical questions call for incorporating multiple sectors in the model. This is particularly true for questions related to structural transformation (Delventhal, 2018; Eckert and Peters, 2018; Fajgelbaum and Redding, 2018; Nagy, 2020). Although the tractability properties of Section 3.1 carry over to multi-sector models under certain restrictive assumptions (constant expenditure shares, no input-output linkages across sectors, no agglomeration externalities – Allen and Arkolakis, 2014), such assumptions prove too restrictive in the context of structural transformation. One of the major empirical facts documented about structural transformation, for instance, is that sectoral expenditure shares do change over time (Herrendorf, Rogerson and Valentinyi, 2014). As a result, these studies cannot rely on the theoretical tractability results of Section 3.1.

Endogenous infrastructure development. Another assumption of the class of models discussed in Section 3.1 is that trade and commuting costs are exogenous. Thus, the development of transport infrastructure can be fed into these models as an exogenous reduction in trade (and possibly commuting) costs across certain locations. This methodology cannot account for the fact that infrastructure development is endogenous: whether to develop infrastructure across certain locations is a decision made by agents who take into account the benefits and costs of infrastructure development. Studying the forces behind these infrastructure development decisions is in fact the focus of various quantitative historical studies (Santamaria, 2020; Swisher, 2014; Trew, 2020). However, without further assumptions, the problem of endogenous infrastructure development is subject to a similar curse of dimensionality as the dynamic problem discussed earlier in this section. The reason for this is simple: the number of possible links across locations increases exponentially with the number of locations. Therefore, the problem quickly gets computationally out of hand as the number of locations becomes large. A further complication is that infrastructure developers might engage in strategic interaction with one another (Swisher, 2014).

Simplifications to the problem offered in this literature rely on reducing the number of links and restricting strategic behavior (Swisher, 2014), assuming one-dimensional space and free entry in infrastructure development (Trew, 2020), or making infrastructure development location-specific rather than link-specific (Santamaria, 2020; Fajgelbaum and Schaal, 2020; Ducruet et al., 2020). That said, even these simplifying assumptions do not necessarily guarantee certain aspects of tractability, such as equilibrium uniqueness. Facing this issue, Trew (2020) assumes that the infrastructure allocation maximizing economy-wide net rents is selected by the U.K. Parliament in the case of multiple equilibria with different allocations.

As discussed above, the tractability issues arising in quantitative models with such additional ingredients often prove challenging and require the researcher to make highly restrictive assumptions. This, of course, raises the question of whether quantitative modeling is the way to go to answer these important historical questions. However, one needs to weigh these costs of quantitative models against their benefits. There are at least four benefits that need to be taken into account. First, quantitative models can be used to infer missing data points, as discussed in Section 2. Second, unlike reduced-form empirical techniques, they are able to measure the aggregate general equilibrium impact of historical events and therefore distinguish growth from reallocation. Third, those quantitative models that feature dynamics can be used to study the long-run effects of events. Due to various other events taking place over long time periods, such long-run effects are often impossible to recover empirically. Finally, issues of tractability are ubiquitous in structural modeling and are by no means restricted to the field of economic geography. Therefore, new techniques developed to address tractability issues may spill over and benefit other areas of economics as well.

4 Identification

One of the largest challenges facing any empirical investigation in economic geography is the identification of causal effects (Redding and Turner, 2015). In the case of quantitative economic geography studies, this challenge can arise in two separate places. First, the challenge of identification is present when the quantitative model is estimated or calibrated to data. Second, besides taking the model to the data, a large number of quantitative studies also estimate the reduced-form effects of spatial events on local outcomes. This artice focuses on the first issue, as it is the one directly related to quantitative modeling. For a review of identification challenges faced by reduced-form empirical work in economic geography, see Baum-Snow and Ferreira (2015).

The key challenge involved in identifying the effects of historical spatial events stems from the fact that the event itself is often endogenous. Transport infrastructure construction, trade, or the spatial concentration of population are the outcomes of decisions made by economic agents. A model in which they occur exogenously misses a set of structural equations that determine these outcomes as a function of other economic variables. Taking the model to the data without these missing equations can bias the estimation or calibration of model parameters entering the equations that are present in the model. A related issue is that of omitted variables. If the set of locations directly affected by the historical event are special in other respects, then the estimation or calibration might falsely attribute the effects of these omitted variables to the event itself.

Quantitative historical studies have developed three broad strategies to overcome the challenge of identification. Some studies rely on natural experiments to estimate the parameters of the model. Others use instrumental variables that are arguably exogenous. Finally, some studies identify the model's parameters using data that only come from the period before the event occurred. In what follows, I briefly review the studies that have followed these three strategies.

Estimation of model parameters using natural experiments. A number of quantitative studies exploit natural experiments. In a spatial context, a certain event constitutes a natural experiment if locations' exposure to it is as good as random. That is, which locations are exposed to the event and to what extent are not related to these (or other) locations' pre-event characteristics. The redrawing of borders and expulsion of populations that follow wars often provide such natural experiments. Ahlfeldt et al. (2015), for example, use the division of Berlin after the Second World War as a natural experiment to study the strength of agglomeration economies within the city. The division of Berlin ultimately resulted in the construction of the Berlin Wall, which cut the Western part of the city from the Eastern part. Western locations close to the wall experienced a dramatic decline in the surrounding concentration of population and economic activity. The decline was naturally smaller for locations farther away from the wall. As the location of the wall was chosen based on non-economic (military) considerations, the extent to which different locations suffered a decline in surrounding concentration was *as if* these declines were randomly assigned to locations. Thus, Ahlfeldt et al. (2015) can use the division of Berlin as a natural experiment to estimate the model's parameters driving the extent to which locations benefit from surrounding concentration (agglomeration economies).

Ahlfeldt et al. (2015) estimate their quantitative spatial model of Berlin using Generalized Method of Moments (GMM). This method relies on setting up a series of *moment conditions*, i.e., conditions that need to be satisfied by the estimated parameters of the model. Ahlfeldt et al. (2015) base these conditions on the natural experiment provided by the division of the city. In particular, these moment conditions state that a location's proximity to the Berlin Wall was unrelated to the change in the levels of fundamental amenities and productivity at the location. This condition has to be satisfied if the placement of the wall was truly exogenous.

Nagy (2018) and Peters (2019) employ similar strategies to look at the effect of trade on urbanization and the effect of population on local economic growth, respectively. In Nagy (2018), the natural experiment stems from the redrawing of Hungary's borders after the First World War. This redrawing of borders disproportionately reduced the trading opportunities of regions close to the new border. Nagy (2018) develops a quantitative model in which trading opportunities affect urbanization, and estimates the key parameter of the model using the variation in changing trading opportunities due to different regions being differentially exposed to the new border. In the case of Peters (2019), the natural experiment comes from the expulsion of ethnic Germans from Eastern European countries after the Second World War. These refugees were settled in Germany, increasing local population substantially but differentially across locations. As the placement of the refugees was driven by non-economic considerations, the variation in population increases across locations can be used to estimate the parameters of a quantitative model in which local population drives local growth.

Estimation of model parameters using instrumental variables. Instrumental variables provide an alternative to the use of natural experiments, which are rare in history. In the context of measuring the effects of spatial events, an instrumental variable is a variable that is correlated with locations' (possibly endogenous) exposure to the event, but does not affect locations' economic outcomes in other ways. Heblich, Trew and Zylberberg (2020), for instance, study the persistence of neighborhood segregation in U.K. cities using a dynamic model of residential sorting. A key structural equation of the model is one that links current to past segregation. Estimating the equation with ordinary least squares would be subject to the issue that past segregation is endogenous. As a result, Heblich, Trew and Zylberberg (2020) use 19th-century pollution as an instrument for past segregation. This instrument has a significant effect on past segregation, as richer residents sorted to less polluted neighborhoods historically. This method of identification is valid as long as variation in 19th-century pollution across neighborhoods only affects current segregation through segregation in the past. Although this assumption is untestable by definition, Heblich, Trew and Zylberberg (2020) also show that a key determinant of 19th-century pollution was whether a neighborhood is on the Eastern or Western side of a city (as winds typically blow from West to East). Thus, a substantial part of the identifying variation comes from geography, which is exogenous.

Various other quantitative historical studies rely on geography-based instruments to estimate the parameters of quantitative models. Heblich, Redding and Sturm (2020) use straight-line distances across boroughs to instrument rail travel times in early-20th century London. Fajgelbaum and Redding (2018) use least-cost paths to major ports and Spanish colonial postal routes as instruments for the location of railroads in Argentina. In Brinkman and Lin (2019), planned freeways and historical exploration routes serve as instruments for the location of actual freeways in the United States.

Calibration of the model to pre-event data. In a few studies, the quantitative model is calibrated to data that precede the event whose effect the model is used to measure. As a result, the potential endogeneity of the event cannot bias the identification of the model's parameters. For example, Nagy (2020) evaluates the effect of early-19th century U.S. railroads on city development and aggregate growth. Though the placement of railroads is likely endogenous, the calibration of the model is fully based on data from the period prior to railroad construction. One concern could still be that people anticipated the placement of railroads at certain locations even before they started to be built. However, this is unlikely in the case of the steam railroad, which came into existence and spread out from Britain over the course of a few years in the 1820s and 1830s. Other quantitative historical studies calibrating model parameters to pre-event data include Redding and Sturm (2008) and Santamaria (2020), who take their models to pre-WWII German data to study the impact of the 1947 division of Germany on city populations and highway construction, respectively.

As in the case of new techniques meant to increase model tractability, novel ways of identification developed in the quantitative historical geography literature may benefit other researchers. Historical instruments, for instance, are often useful to obtain exogenous variation in present-day outcomes. Among others, Duranton, Morrow and Turner (2014) and Duranton and Turner (2011) use historical exploration routes as instruments to study how the current U.S. road network affects trade and traffic congestion, respectively.

5 Conclusion

Over the last few years, the use of quantitative models has gained prominence in economic geography. In this article, I have reviewed the part of this literature that studies the economic impact of historical events with a quantitative methodology. I have argued that this methodology is able to bridge the gap between reduced-form empirical work and classical structural modeling. On the one hand, it is suitable to be combined with rich spatial data, thus allowing the data to speak about important historical questions. On the other hand, structural by nature, quantitative modeling can measure the aggregate impact of historical events and distinguish growth from the mere relocation of economic activity. These benefits from quantitative modeling, however, come at a price. Specifically, studying historical spatial questions with quantitative models is associated with three key challenges: the sparse nature of historical data, model tractability, and identification issues. In this article, I have reviewed how the literature has addressed these challenges.

Even though quantitative historical studies have come a long way over the course of a few years in economic geography, there is clearly room for further research in the field. As already mentioned, historical questions are often important for today's economists because similar events are expected to take place in the future. Developing models that can be used by policymakers to assess the economic impact of these future events has the potential to inform these policymakers' decisions. Quantitative models of economic geography are still rarely used in policy work. One reason for this might be the complex structure of these models, which often makes them seem like a black box to researchers outside the literature. Simplifying and clarifying model structure, without giving up too much on model realism, seems a fruitful direction that may allow these models to be adapted to a larger extent in the world of policy.

References

- Acemoglu, D. (2009): Introduction to modern economic growth. Princeton University Press.
- Ahlfeldt, G., Redding, S., Sturm, D. and Wolf, N. (2015): The economics of density: Evidence from the Berlin Wall. *Econometrica*, vol. 83(6), 2127–2189.
- Allen, T., Arkolakis, C. and Li, X. (2020): On the equilibrium properties of network models with heterogeneous agents. Mimeo.
- Allen, T. and Arkolakis, C. (2014): Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, vol. 129(3), 1085–1140.
- Allen, T. and Donaldson, D. (2018): The geography of path dependence. Mimeo.
- Anderson, J. (1979): A theoretical foundation for the gravity equation. *American Economic Review*, vol. 69(1), 106–116.
- Barjamovic, G., Chaney, T., Coşar, K. and Hortaçsu, A. (2019): Trade, merchants, and the lost cities of the Bronze Age. *Quarterly Journal of Economics*, vol. 134(3), 1455–1503.
- Baum-Snow, N. and Ferreira, F. (2015): Causal inference in urban and regional economics. In *Handbook of Regional and Urban Economics*, vol. 5, ed. Duranton, G., Henderson, V. and Strange, W. 3–68.
- Bolstad, P. (2005): GIS fundamentals: A first text on Geographic Information Systems. Eider Press.
- Bolt, J., Inklaar, R., de Jong, H. and van Zanden, J. (2018): Rebasing 'Maddison': New income comparisons and the shape of long-run economic development. *GGDC Research Memorandum* 174.
- Bos, F. (1992): The history of national accounting. *MRPA Paper*, University Library of Munich.
- Brinkman, J. and Lin, J. (2019): Freeway revolts! The quality of life effects of highways. Mimeo.
- Brunet, M. (2010): Two new Mio-Pliocene Chadian hominids enlighten Charles Darwin's 1871 prediction. *Philosophical Transactions of the Royal Society B.*, 365(1556), 3315–3321.

- Carson, C. (1975): The history of the United States national income and product accounts: The development of an analytical tool. *Review of Income and Wealth*, 21, 153–181.

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

- David, P. (1967): The growth of real product in the United States before 1840: New evidence, controlled conjectures. *Journal of Economic History*, 27, 151–197.

- Delventhal, M. (2018): The globe as a network: Geography and the origins of the world income distribution. Mimeo.

- Desmet, K., Nagy, D., and Rossi-Hansberg, E. (2018): The geography of development. *Journal of Political Economy*, vol. 126(3), 903–983.

- Desmet, K. and Rappaport, J. (2015): The settlement of the United States, 1800-2000: The long transition to Gibrat's Law. *Journal of Urban Economics*, forthcoming.

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

- Donaldson, D. (2018): Railroads of the Raj: Estimating the impact of transportation infrastructure.

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

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

- Duranton, G., Morrow, P. and Turner, M. (2014): Roads and trade: Evidence from the US. *Review of Economic Studies*, vol. 81(2), 681–724.

- Duranton, G. and Turner, M. (2011): The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, vol. 101(6), 2616–2652.

- Easterlin, R. (1960): Interregional differences in per capita income, population and total income, 1840–1950. In *Trends in the American economy in the nineteenth century.* Committee on Research in Income and Wealth. Princeton University Press.

- Eaton, J. and Kortum, S. (2002): Technology, geography, and trade. *Econometrica*, vol. 70(5), 1741–1779.

- Ebenstein, A., Fan, M., Greenstone, M., He, G., Yin, P. and Zhou, M. (2015): Growth, pollution and life expectancy: China from 1991-2012. *American Economic Review*, 105(5), 226–231.
- Eckert, F. and Peters, M. (2018): Spatial structural change. Mimeo.
- Fajgelbaum, P. and Redding, S. (2018): Trade, structural transformation and development: Evidence from Argentina 1869-1914. *NBER Working Paper* 20217.
- Fajgelbaum, P. and Schaal, E. (2020): Optimal transport networks in spatial equilibrium. *Econometrica*, vol. 88(4), 1411–1452.
- Fishbein, M. (1973): The Censuses of Manufactures 1810–1890. Reference Information Paper No. 50, National Archives and Records Service.
- Fogel, R. (1964): Railroads and American economic growth: Essays in econometric history. Johns Hopkins University Press.
- Fujita, M. and Thisse, J. (2002): Economics of agglomeration. Cambridge University Press.
- Gallman, R. (1966): Gross National Product in the United States, 1834–1909. In *Output, employment, and productivity in the United States after 1800*, ed. Brady, D. NBER Studies in Income and Wealth, 30. Columbia University Press.
- Galor, O. and Özak, Ö. (2016): The agricultural origins of time preference. *American Economic Review*, vol. 106(10), 3064–3103.
- Heblich, S., Redding, S. and Sturm, D. (2020): The making of the modern metropolis: Evidence from London. *Quarterly Journal of Economics*, vol. 135(4), 2059–2133.
- Heblich, S., Trew, A. and Zylberberg, Y. (2020): East Side Story: Historic pollution and neighborhood segregation. *Journal of Political Economy*, forthcoming.
- Herrendorf, B., Rogerson, R., and Valentinyi, Á (2014): Growth and structural transformation. *Handbook of Economic Growth*, vol. 2, ch. 6, 855–941.
- Kalweit, F. (1937): Die Baustellenwerte in Berlin: Strassen-Verzeichnis, Neuarbeitung abgeschlossen am 30. Juni 1936. Carl Heymanns.
- Klein Goldewijk, K., Beusen, A. and Janssen, P. (2010): Long term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. *The Holocene*, vol. 20, 565–573.
- Krugman, P. (1991): Increasing returns and economic geography. *Journal of Political Economy*, vol. 99(3), 483–499.

- Lee, S. and Lin, J. (2018): Natural amenities, neighborhood dynamics, and persistence in the spatial distribution of income. *Review of Economic Studies*, vol. 85(1), 663–694.

- Manson, S., Schroeder, J. van Riper, D. and Ruggles, S. (2017): IPUMS National Historical Geographic Information System: Version 12.0. *Minneapolis: University of Minnesota. http://doi.org/10.18128/D050.V12.0.*

- Nagy, D. (2018): Trade and urbanization: Evidence from Hungary. Mimeo.

- Nagy, D. (2020): Hinterlands, city formation and growth: Evidence from the U.S. westward expansion. Mimeo.

- Pérez-Cervantes, F. (2014): Railroads and economic growth: A trade policy approach. *Banco de México Working Paper* 2014-14.

- Peters, M. (2019): Market size and spatial growth – Evidence from Germany's post-war population expulsions. Mimeo.

- Ramankutty, N., Foley, J., Norman, J. and McSweeney, K. (2002): The global distribution of cultivable lands: Current patterns and sensitivity to possible climate change. *Global Ecology and Biogeography*, vol. 11(5), 377–392.

- Redding, S. (2016): Goods trade, factor mobility and welfare. *Journal of International Economics*, vol. 101, 148–167.

- Redding, S. and Sturm, D. (2008): The costs of remoteness: Evidence from German Division and Reunification. *American Economic Review*, vol. 98(5), 1766–1797.

- Redding, S. and Sturm, D. (2016): Estimating neighborhood effects: Evidence from war-time destruction in London. Mimeo.

- Redding, S. and Turner, M. (2015): Transportation costs and the spatial organization of economic activity. In *Handbook of Regional and Urban Economics*, vol. 5, ed. Duranton, G., Henderson, V. and Strange, W. 1339–1398.

- Santamaria, M. (2020): The gains from reshaping infrastructure: Evidence from the division of Germany. Mimeo.

- Shaw-Taylor, L., Wrigley, E., Davies, R., Kitson, P., Newton, G. and Satchell, A. (2010): The occupational structure of England c.1710 to c.1871. Mimeo.

- Trew, A. (2014): Spatial takeoff in the first industrial revolution. *Review of Economic Dynamics*, vol. 17: 707–725.

- Trew, A. (2020): Endogenous infrastructure development and spatial takeoff in the first industrial revolution. *American Economic Journal: Macroeconomics*, vol. 12(2), 44–93.
- Ubelaker, D. (2008): Human skeletal remains: Excavation, analysis, interpretation. Transaction Pub.
- U.S. Census (1841): Compendium of the enumeration of the inhabitants and statistics of the United States. *http://usda.mannlib.cornell.edu/usda/AgCensusImages/1840/1840.pdf*.
- U.S. Census (1852): The seventh census of the United States: 1850 - California. *https://www2.census.gov/library/publications/decennial/1850/1850a/1850a-47.pdf*.
- Weiss, T. (1992): U.S. labor force estimates and economic growth, 1800-1860. In *American economic growth and standards of living before the Civil War*, 19–78. University of Chicago Press.

Appendix

This appendix consists of three sections. In Section A.1, I derive the equilibrium conditions of the model of Section 3.1.1. In Section A.2, I provide the proofs of Theorems 1 to 3. Finally, in Section A.3, I show the isomorphisms presented in Section 3.1.3.

A.1 Derivation of the model's equilibrium conditions

Denote total spending on tradables (by workers and landlords) at location r by $X(r)$. Note that this also equals the nominal income of workers residing at r. This is because their income is ultimately spent on tradables, either directly or indirectly (through spending on housing, which is spent on tradables by landlords). As a result, spending on housing can be expressed as

$$P_H(r)\,H(r) = (1-\nu)\,X(r) \tag{1}$$

where $P_H(r)$ denotes the price of housing, and $H(r)$ denotes the quantity of housing at r. Note that $H(r)$ is exogenously given in equilibrium, as the supply of housing is fixed at each location. Rearranging (1) yields the equilibrium price of housing:

$$P_H(r) = (1-\nu)\,\frac{X(r)}{H(r)} \tag{2}$$

Using (2) and the Fréchet distribution of idiosyncratic location tastes, one can express the number of workers choosing to reside at r and work at s as

$$L(r,s)=\frac{\bar{a}(r,s)\,N(r)^{-\lambda}\left[w(s)\,/\left(P(r)^{\nu}\,X(r)^{1-\nu}\,H(r)^{\nu-1}\right)\right]^{1/\eta}\kappa(r,s)^{-1/\eta}}{\sum_{u=1}^{S}\sum_{v=1}^{S}\bar{a}(u,v)\,N(u)^{-\lambda}\left[w(v)\,/\left(P(u)^{\nu}\,X(u)^{1-\nu}\,H(u)^{\nu-1}\right)\right]^{1/\eta}\kappa(u,v)^{-1/\eta}}\bar{L} \tag{3}$$

where $w(s)$ denotes the wage at s, and $P(r)$ denotes the ideal price index of tradables at r. Similarly, the expected utility of a worker can be expressed as

$$\bar{U}=\delta\left[\sum_{r=1}^{S}\sum_{s=1}^{S}\bar{a}(r,s)\,N(r)^{-\lambda}\left[w(s)\,/\left(P(r)^{\nu}\,X(r)^{1-\nu}\,H(r)^{\nu-1}\right)\right]^{1/\eta}\kappa(r,s)^{-1/\eta}\right]^{\eta} \tag{4}$$

where $\delta=\Gamma(1-\eta)$, such that $\Gamma(\cdot)$ denotes the gamma function.

On the production side, perfect competition among firms at location s implies that the local price of tradable good s, $p(s,s)$, equals its marginal cost of production:

$$p(s,s)=\bar{A}(s)^{-1}\,L(s)^{-\alpha}\,w(s)$$

At any other location r, no arbitrage guarantees that the price of the good equals the marginal cost of production and trade:

$$p(s,r)=\bar{A}(s)^{-1}\,L(s)^{-\alpha}\,w(s)\,\tau(s,r) \tag{5}$$

In equilibrium, the ideal price index at location r is given by the equation

$$P(r)^{1-\sigma}=\sum_{s=1}^{S}p(s,r)^{1-\sigma}$$

due to CES preferences for tradables. Using (5), one can rewrite this equation as

$$P(r)^{1-\sigma}=\sum_{s=1}^{S}\bar{A}(s)^{\sigma-1}\,L(s)^{\alpha(\sigma-1)}\,w(s)^{1-\sigma}\,\tau(s,r)^{1-\sigma}\,. \tag{6}$$

Market clearing for tradable good s implies

$$w(s)\,L(s)=\sum_{r=1}^{S}\pi(s,r)\,X(r)$$

where $\pi(s,r)$ denotes the share of good s in total spending on tradables at r. Due to CES preferences for tradables, this share can be obtained as

$$\pi(s,r)=p(s,r)^{1-\sigma}\,P(r)^{\sigma-1}=\bar{A}(s)^{\sigma-1}\,L(s)^{\alpha(\sigma-1)}\,w(s)^{1-\sigma}\,P(r)^{\sigma-1}\,\tau(s,r)^{1-\sigma}$$

where I used equation (5). This allows me to rewrite the market clearing condition as

$$\bar{A}(s)^{1-\sigma} w(s)^{\sigma} L(s)^{1-\alpha(\sigma-1)} = \sum_{r=1}^{S} P(r)^{\sigma-1} X(r) \tau(s,r)^{1-\sigma}. \tag{7}$$

Finally, labor market clearing implies that the nominal income of residents of location r equals

$$X(r) = \sum_{s=1}^{S} w(s) L(r,s)$$

while the population of this location equals

$$N(r) = \sum_{s=1}^{S} L(r,s)$$

and the employment of location s equals

$$L(s) = \sum_{r=1}^{S} L(r,s).$$

Using (3) and (4), these last three equations can be rewritten as

$$H(r)^{-\frac{1-\nu}{\eta}} P(r)^{\frac{\nu}{\eta}} X(r)^{1+\frac{1-\nu}{\eta}} N(r)^{\lambda} = \left(\frac{\bar{U}}{\delta}\right)^{-1/\eta} \bar{L} \sum_{s=1}^{S} \bar{a}(r,s) w(s)^{1+\frac{1}{\eta}} \kappa(r,s)^{-1/\eta}, \tag{8}$$

$$H(r)^{-\frac{1-\nu}{\eta}} P(r)^{\frac{\nu}{\eta}} X(r)^{\frac{1-\nu}{\eta}} N(r)^{1+\lambda} = \left(\frac{\bar{U}}{\delta}\right)^{-1/\eta} \bar{L} \sum_{s=1}^{S} \bar{a}(r,s) w(s)^{1/\eta} \kappa(r,s)^{-1/\eta} \tag{9}$$

and

$$w(s)^{-1/\eta} L(s) = \left(\frac{\bar{U}}{\delta}\right)^{-1/\eta} \bar{L} \sum_{r=1}^{S} \bar{a}(r,s) H(r)^{\frac{1-\nu}{\eta}} N(r)^{-\lambda} P(r)^{-\frac{\nu}{\eta}} X(r)^{-\frac{1-\nu}{\eta}} \kappa(r,s)^{-1/\eta}. \tag{10}$$

(3), (4), (6), (7), (8), (9) and (10) constitute a system of $5S + S^2 + 1$ equations and $5S + S^2 + 1$ unknowns: locations' residential populations $N(r)$, employment levels $L(s)$, wage levels $w(s)$, total spending on tradables $X(r)$ and ideal price index $P(r)$, cross-location commuting flows $L(r,s)$, as well as the economy-wide level of workers' expected utility, $\bar{U}$. These are the equilibrium conditions I use to prove Theorems 1 to 3 in Section A.2 and show the isomorphisms with alternative models in Section A.3.

A.2 Proofs of theorems

Proof of Theorem 1. (6), (7), (8), (9) and (10) and (4) consitute a system of $5S+1$ equations and $5S+1$ unknowns: $P(r)$, $w(s)$, $X(r)$, $N(r)$, $L(s)$ and $\bar{U}$. This system is a special case of the systems considered in Allen et al. (2020):[15]

$$\prod_{h=1}^{H} x_h(r)^{\gamma_{kh}} = \sum_{s=1}^{S} K_k(r,s) \prod_{h=1}^{H} x_k(r)^{\kappa_{kh}} x_h(s)^{\beta_{kh}} \qquad k=1,2,...,H \tag{11}$$

such that $H=5$, $x_1(r)=P(r)$, $x_2(r)=w(r)$, $x_3(r)=X(r)$, $x_4(r)=N(r)$, $x_5(r)=L(r)$,

$$K_1(r,s) = \bar{A}(s)^{\sigma-1} \tau(s,r)^{1-\sigma},$$

$$K_2(r,s) = \bar{A}(r)^{\sigma-1} \tau(r,s)^{1-\sigma},$$

$$K_3(r,s) = K_4(r,s) = K_5(r,s) = \left(\frac{\bar{U}}{\delta}\right)^{-1/\eta} \bar{L}\bar{a}(r,s) H(r)^{\frac{1-\nu}{\eta}} \kappa(r,s)^{-1/\eta},$$

κ_{kh} is the (k,h) entry of a 5×5 matrix $\mathbf{K}$ of zeros, γ_{kh} is the (k,h) entry of the matrix

$$\mathbf{\Gamma} = \begin{bmatrix} 1-\sigma & 0 & 0 & 0 & 0 \\ 0 & \sigma & 0 & 0 & 1-\alpha(\sigma-1) \\ \nu/\eta & 0 & 1+\frac{1-\nu}{\eta} & \lambda & 0 \\ \nu/\eta & 0 & \frac{1-\nu}{\eta} & 1+\lambda & 0 \\ 0 & -1/\eta & 0 & 0 & 1 \end{bmatrix}$$

and β_{kh} is the (k,h) entry of the matrix

$$\mathbf{B} = \begin{bmatrix} 0 & 1-\sigma & 0 & 0 & \alpha(\sigma-1) \\ \sigma-1 & 0 & 1 & 0 & 0 \\ 0 & 1+\frac{1}{\eta} & 0 & 0 & 0 \\ 0 & 1/\eta & 0 & 0 & 0 \\ 1-\frac{\nu}{\eta} & 0 & -\frac{1-\nu}{\eta} & -\lambda & 0 \end{bmatrix}.$$

Theorem 1 in Allen et al. (2020) shows that the solution to (11) exists and is unique under the condition that the largest eigenvalue of the matrix $|\left(\mathbf{B}(\mathbf{\Gamma}-\mathbf{K})^{-1}\right)_{kh}|$ is strictly less than one.[16] As $\mathbf{\Gamma}$ and $\mathbf{B}$ are functions of the model's structural parameters only, the condition guaranteeing existence and uniqueness only depends on these parameters. This proves Theorem 1.

Proof of Theorem 2. Theorem 1 in Allen et al. (2020) also shows that if the largest

[15] In particular, see Remark 3 in Allen et al. (2020).

[16] Remark 2 in Allen et al. (2020) addresses the issue that $\bar{U}$ is an endogenous constant, determined by equation (4).

eigenvalue of $|\left(\mathbf{B}\left(\mathbf{\Gamma}-\mathbf{K}\right)^{-1}\right)_{kh}|$ is strictly less than one, then an algorithm that consists of iteratively applying (11) is guaranteed to converge to the solution of (11). In the context of the model of Section 3.1.1, this implies that the researcher can guess any initial distribution of $P\left(r\right)$, $w\left(s\right)$, $X\left(r\right)$, $N\left(r\right)$ and $L\left(r\right)$, plug them into equations (6), (7), (8), (9) and (10), and update the distributions using the left-hand sides of these equations. Applying this procedure iteratively, the distributions of $P\left(r\right)$, $w\left(s\right)$, $X\left(r\right)$, $N\left(r\right)$ and $L\left(r\right)$ are guaranteed to converge to the equilibrium values of these variables. Once these variables are known, equations (3) and (4) provide commuting flows $L\left(r,s\right)$ and workers' expected utility $\bar{U}$ in closed form. This proves Theorem 2.

Proof of Theorem 3. Using wages $w\left(s\right)$ and the matrix of commuting flows $L\left(r,s\right)$ (including commuting from a location to itself), the researcher can recover each location's residential population, employment and spending on tradables as

$$N\left(r\right)=\sum_{s=1}^{S}L\left(r,s\right),$$

$$L\left(s\right)=\sum_{r=1}^{S}L\left(r,s\right)$$

and

$$X\left(r\right)=\sum_{s=1}^{S}w\left(s\right)L\left(r,s\right)$$

respectively. Plugging $L\left(s\right)$ and $X\left(r\right)$ into equations (6) and (7), $2S$ unknowns remain in these $2S$ equations: namely, the ideal price index $P\left(r\right)$ and the fundamental productivity $\bar{A}\left(r\right)$ of each location r. This system is again a special case of the systems considered in Allen et al. (2020), such that $H=2$, $x_1\left(r\right)=P\left(r\right)$, $x_2\left(r\right)=\bar{A}\left(r\right)$,

$$K_1\left(r,s\right)=L\left(s\right)^{\alpha(\sigma-1)}w\left(s\right)^{1-\sigma}\tau\left(s,r\right)^{1-\sigma},$$

$$K_2\left(r,s\right)=w\left(r\right)^{-\sigma}L\left(s\right)^{\alpha(\sigma-1)-1}X\left(s\right)\tau\left(r,s\right)^{1-\sigma},$$

$\mathbf{K}$ is a 2×2 matrix full of zeros,

$$\mathbf{\Gamma}=\begin{bmatrix}1-\sigma & 0\\ 0 & 1-\sigma\end{bmatrix}$$

and

$$\mathbf{B}=\begin{bmatrix}0 & \sigma-1\\ \sigma-1 & 0\end{bmatrix}.$$

It is straightforward to show that the largest eigenvalue of $|\left(\mathbf{B}\left(\mathbf{\Gamma}-\mathbf{K}\right)^{-1}\right)_{kh}|$ equals one. In this case, Theorem 1 in Allen et al. (2020) implies that the solution to the system exists and is unique up to scale, although iterating on the system is not guaranteed to converge to the solution. Thus, the set of fundamental productivities $\bar{A}(s)$ rationalizing the observed data is unique (up to scale). Next, plugging the price indices $P(r)$ just recovered, as well as the observed $L(r,s)$, $w(s)$ and $N(r)$, into equation (3), one can obtain the fundamental amenity level of any location pair (r,s) up to scale by simply rearranging this equation. As price indices are uniquely identified by the previous step, these fundamental amenity levels are also unique (up to scale). This proves Theorem 3.

A.3 Proofs of isomorphisms

Land use in production. The representative firm's first order conditions imply that land rents are given by

$$R(s) = \frac{1-\mu}{\mu}\frac{w(s)L(s)}{\Lambda(s)}$$

where $\Lambda(s)$ equals the exogenous supply of production land at s in equilibrium. As a result, the firm's marginal cost of production equals

$$\begin{aligned} &\mu^{-\mu}(1-\mu)^{-(1-\mu)}\bar{A}(s)^{-1}L(s)^{-\alpha}w(s)^{\mu}R(s)^{1-\mu} \\ &= \left(\mu\bar{A}(s)\Lambda(s)^{1-\mu}\right)^{-1}L(s)^{-[\alpha-(1-\mu)]}w(s) \end{aligned} \tag{12}$$

which has the same form as the marginal cost of production in the model of Section 3.1.1 once fundamental productivities and the agglomeration parameter are redefined as

$$\tilde{\bar{A}}(s) = \mu\bar{A}(s)\Lambda(s)^{1-\mu}$$

and

$$\tilde{\alpha} = \alpha - (1-\mu).$$

Finally, the income of an employee at s, $y(s)$, equals her wage income plus a $1/L(s)$ fraction of rents,

$$y(s) = w(s) + \frac{R(s)\Lambda(s)}{L(s)} = \mu^{-1}w(s) \tag{13}$$

and total spending on tradables at residential location r is given by

$$X(r) = \sum_{s=1}^{S} y(s)L(s) = \mu^{-1}\sum_{s=1}^{S} w(s)L(s). \tag{14}$$

Redefining $\tilde{X}(r) = \mu X(r)$, the equilibrium conditions yield equations (3), (4), (6), (7), (8), (9) and (10) in the redefined variables, except that workers' expected utility is multiplied by a constant $\mu^{-\nu}$. Thus, this model is isomorphic to the model of Section 3.1.1.

Endogenous specialization, driven by comparative advantage. Following the same steps as in Eaton and Kortum (2002), one can show that the share of location-s goods in total spending on tradables at r equals

$$\pi(s,r) = \frac{\bar{A}(s)^{\theta} L(s)^{\alpha\theta} w(s)^{-\theta} \tau(s,r)^{-\theta}}{\sum_{u=1}^{S} \bar{A}(u)^{\theta} L(u)^{\alpha\theta} w(u)^{-\theta} \tau(u,r)^{-\theta}}$$

and the ideal price index at r equals

$$P(r) = \Gamma\left(1 - \frac{\sigma-1}{\theta}\right)^{\frac{1}{1-\sigma}} \left[\sum_{s=1}^{S} \bar{A}(s)^{\theta} L(s)^{\alpha\theta} w(s)^{-\theta} \tau(s,r)^{-\theta}\right]^{-1/\theta}.$$

Redefining $\tilde{P}(r) = \Gamma\left(1 - \frac{\sigma-1}{\theta}\right)^{\frac{1}{\sigma-1}} P(r)$ and rearranging, we obtain the equation

$$\tilde{P}(r)^{-\theta} = \sum_{s=1}^{S} \bar{A}(s)^{\theta} L(s)^{\alpha\theta} w(s)^{-\theta} \tau(s,r)^{-\theta} \tag{15}$$

while goods market clearing yields

$$\bar{A}(s)^{-\theta} w(s)^{1+\theta} L(s)^{1-\alpha\theta} = \sum_{r=1}^{S} \tilde{P}(r)^{\theta} X(r) \tau(s,r)^{-\theta}. \tag{16}$$

Equations (15) and (16) are identical to (6) and (7), except that they include the redefined price index variable, and θ everywhere instead of $\sigma - 1$. The remaining equilibrium conditions are unchanged, except that workers' expected utility is multiplied by a constant $\Gamma\left(1 - \frac{\sigma-1}{\theta}\right)^{\frac{\nu}{\sigma-1}}$. Thus, this model is isomorphic to the model of Section 3.1.1.

Endogenous specialization, driven by increasing returns. Due to monopolistic competition with a linear production technology, CES preferences and iceberg trade costs, firms in the model with increasing returns set a price that is a constant markup over their marginal cost:

$$p_{\omega}(s,r) = p(s,r) = \frac{\sigma}{\sigma-1} \bar{A}(s)^{-1} L(s)^{-\alpha} w(s) \tau(s,r) \tag{17}$$

Plugging this price into firms' profits, one can show that the profit-maximizing output of a firm equals

$$q_{\omega}(s) = q(s) = (\sigma - 1) f \bar{A}(s) L(s)^{\alpha} \tag{18}$$

while the firm's employment equals

$$\ell_{\omega}(s) = \ell(s) = \frac{q(s)}{\bar{A}(s) L(s)^{\alpha}} + f = \sigma f.$$

Plugging this into the local labor market clearing condition

$$L(s) = \int_{\omega \text{ produced at } s} \ell_{\omega}(s)\, d\omega$$

pins down the mass of tradables produced at s as

$$N(s) = \frac{L(s)}{\ell(s)} = \frac{L(s)}{\sigma f}. \tag{19}$$

In equilibrium, the ideal price index at location r is given by the equation

$$P(r)^{1-\sigma} = \int_0^N p_{\omega}(r)^{1-\sigma}\, d\omega = \sum_{s=1}^{S} N(s)\, p(s,r)^{1-\sigma}.$$

Combining this with equations (17) and (19) yields

$$P(r)^{1-\sigma} = \sigma^{-\sigma}(\sigma-1)^{\sigma-1} f^{-1} \sum_{s=1}^{S} \bar{A}(s)^{\sigma-1} L(s)^{1+\alpha(\sigma-1)} w(s)^{1-\sigma} \tau(s,r)^{1-\sigma}. \tag{20}$$

By CES preferences, market clearing for a good produced at s implies

$$q(s) = \sum_{r=1}^{S} p(s,r)^{-\sigma} P(r)^{\sigma-1} X(r).$$

Combining this with equations (17) and (18) and rearranging yields

$$\bar{A}(s)^{1-\sigma} w(s)^{\sigma} L(s)^{-\alpha(\sigma-1)} = \sigma^{-\sigma}(\sigma-1)^{\sigma-1} f^{-1} \sum_{r=1}^{S} P(r)^{\sigma-1} X(r)\, \tau(s,r)^{1-\sigma}. \tag{21}$$

Normalizing the fixed cost of production to $f = \sigma^{-\sigma}(\sigma-1)^{\sigma-1}$ and redefining the agglomeration parameter of the model of Section 3.1.1 as

$$\tilde{\alpha} = \alpha + \frac{1}{\sigma-1},$$

equations (20) and (21) become identical to equations (6) and (7). The remaining equilibrium conditions are unchanged. This shows the isomorphism between this model and the model of Section 3.1.1.