博士論文

論文題目　Ethnicity, Language, and Economy

(エスニシティ・言語と経済)

氏名　中川　万理子

Ethnicity, Language, and Economy

Mariko Nakagawa
Graduate School of Economics, University of Tokyo

January 31, 2016

Contents

Introduction

Ethnically heterogeneous society is inevitably affected by the socio-economic activities of residents in it, whether they are indigenous or immigrant, and whether majorities or minorities. Impacts on the host country given by the existence of multiple ethnic groups are negligible from socio- and politico-economic viewpoints. Ethnic heterogeneity is considered one of the essential factors influencing government institutional quality and public goods provision (Alesina et al., 1999; La Porta et al., 1999; Baldwin and Huber, 2010). Further, Fearon and Laitin (2003) examine ethnic diversity and societal and political stability of a country. As for economic activities, ethnic diversity may give positive effect on productivity through production complementality if integration and coordination of different ethnic and cultural groups are successfully maintained. Especially, advanced countries such as the United States and European nations are more likely to enjoy positive impacts of ethnic diversity on economic success (Alesina and La Ferrara, 2005; Ottaviano and Peri, 2006; Alesina et al., 2013; Bellini et al., 2013).

Ethnicity is not a single concept, but connotes several categories of characteristics of individuals such as language, religion, culture, and genetic race. Because various groups of people with different ethnic backgrounds scatter and spread around the world, they must influence political and economic relationships between countries, not only including the commodity/service flow but also intra- and international people flow or migration. After immigrants of different foreign origins have reached destination countries/regions/cities, they may be considered ethnic minorities there, contrasted with indigenous residents. In such society of ethnic heterogeneity, harmonized integration of various ethnic groups of people is vital for economic success. Hence, this dissertation aims at providing economic analysis in the domain of ethnic heterogeneity and diversity.

Turning our eyes toward social phenomena associated with ethnic heterogeneity, residential separation by ethnic or linguistic groups is observed, known as ethnic segregation, whose spatial scale spans a wide spectrum from a neighborhood level to a regional one. Neighborhood segregation, where minorities' residential distributions are more geographically concentrated, such as Chinatowns, can be seen in cities in the United States, Europe, and elsewhere (Musterd, 2005). Similarly, examples of regional segregation according to ethnolinguistic characteristics are found in many regions, partly as a consequence of the regional (or local) administrative division's choice of the official language,due to historical and political reasons.[1]

Which factors bear ethnolinguistic segregation? Put differently, what are the benefits borne by proximity to the residents who share common ethnic characteristics, particularly for ethnic minorities such as immigrants? The advantage of geographical cluster by ethnicity is that ethnic minorities are more likely to enjoy the benefits of stronger ethnic networks when clustered, which may improve their socio-economic outcomes (Yancey et al., 1976): residentially clustered ethnic minorities tend to perform better in the labor market (Dietz, 1999; Munshi, 2003) and in the housing market (Søholt, 2001). Further, public goods provision reflecting ethnic-oriented preference is another important factor related to ethnic clustering (Boustan, 2007). Residential segregation of ethnic minorities may increase their utility though better access to the ethnicity specific local goods which are provided once a certain threshold of population of the same ethnicity is reached in the neighborhood: as for education, for example, schools for foreign-oriented students may be more easily established when a sufficient number of foreign children reside in the local community.

While the societal aspects of ethnolinguistic heterogeneity mentioned above are such that found within a country, investigation of economic activities associated with it between countries is also worth conducting. In the global economy, it is not rare that individuals with different backgrounds of language, ethnicity, and culture meet and work together, and hence, ethnic heterogeneity is more commonly experienced than in the domestic context. When people of various ethnic backgrounds are in collaboration, communication costs attributed to cultural and linguistic difference necessarily emerge. Especially, difference in language use is a crucial obstacle which hinders smooth interaction and cooperation. Then, English as an internationally widely spoken language overcomes this communication barrier, gluing the linguistically heterogeneous individuals together. Ability to use a globally common language can enhance worldwide connection such as international trade, which contributes to economic growth.

The same can be said to the domestic communication in ethnically diverse countries, because exchanging ideas and cooperative work in those countries are highly costly without shared languages. In those cases, official or national languages should act as connection of different linguistic groups when within-country communication is taken place. Moreover, smooth interaction via domestic central languages could be a bridge between opposed ethnic groups, which would improve political stability and lessen disparities among them, leading to economic success. Considering the coordination costs among ethnolinguistically distant individuals as well as productivity benefits borne by ethnic and cultural diversity must be an important related issue.

Since societies and economies are affected by ethnolinguistic heterogeneity at different

[1] Well-known examples of linguistic segregation by regions in a country are found in Switzerland, Canada, Spain, and the former Soviet Union countries (see Chapter 2).

levels of geographical scales—within and across cities, regions, and countries,— analyses in this dissertation span the following three spatial scales of intra- and international socio-economic activities and phenomena: (i) within a city, (ii) between regions within a country, and (iii) between countries.

For the featured subjects in this dissertation, the following two topics are covered: (a) ethnolinguistic segregation and (b) economic development and the cost of diversity in used languages in a country. Segregation by ethnolinguistic groups, which are observed worldwide, reveals historical persistence as will be shown in Chapter 2. Ethnic segregation is an issue of interest in a research field coping with ethnic heterogeneity from the sociological point of view. Furthermore, investigating the impacts of ethnolinguistic diversity on economic and political activities has been recently attracting academic attention. Ethnic/cultural diversity and economic performance are jointly analyzed in a vast literature, wherein benefits of production complementality stemming from ethnic diversity are expected to improve economic conditions. Unexpectedly, however, existence of the ethnolinguistic heterogeneity shows negative impact on economic success on balance. The reason is simple—hidden behind this production benefits are the impacts on economic development given by the cost of integrating different ethnolinguistic groups. The present focus is on how linguistic heterogeneity affects the cross-national economic income differences, where the communication cost among different linguistic groups is captured as linguistic distance between languages. The use of linguistic distance as between-language communication cost is based on the idea that more distant languages from one's mother tongue may be more difficult to acquire.

From the above-mentioned aspects of spatial scales and topics of socio-economic interests, this dissertation covers the following subjects on ethnicity, language and economy, which are split into three chapters:

1. Ethnic segregation in a city
2. Regional ethnolinguistic segregation and industrial agglomeration in a country
3. Domestic and international linguistic distance and economic development

Each chapter is briefly summarized below.

Chapter 1 analyzes residential segregation by introducing the concept of ethnic clustering externality. In an economy with two areas, namely the center and suburb, households with different ethnic characteristics (termed the majority and minority), both of which have identical skill levels, endogenously choose their residential locations in the long run. By analyzing stable residential equilibria, we show that, because of their ethnic clustering preferences, minority residents are more likely to cluster in one area than majority residents. In addition, when the commuting cost is low, minority residents always cluster, widening the population gap between areas. At the same time, majority households migrate to a less crowded area to avoid residential congestion caused by minority clustering, thus reducing the population gap. In this sense, the majority acts as an equalizer of population sizes between the center and suburb under low commuting costs.

Chapter 2 investigates how regional segregation patterns are affected by industrial agglomeration and ethnic clustering, by adding the externality of ethnicity to the model of agglomeration and trade proposed by Ottaviano et al. (2002). We show that ethnic segregation patterns are persistent, while ethnic mixing distribution appears only when trade costs are intermediate and ethnic clustering preferences are less intense. Further, discrepancies of the social optimum and equilibrium are caused by that the social optimum is less sensitive to a change in trade costs, when the population of farmers (immobile factors affecting ethnic utilities) is sufficiently large.

Chapter 3 considers the impacts of accessibility to domestic and international communication on the economic development of a nation by constructing two indexes of linguistic distance—domestic and international. While the domestic linguistic distance index captures the constraints of nationwide communication among speakers of different mother tongues, the international linguistic distance index captures the constraints of the global communication. The domestic linguistic distance has a negative impact on the economic development of poor countries, while only rich countries enjoy a positive impact on the national income if the international linguistic distance is smaller. Particularly, we show that the capability to use English as the first language is highly advantageous for economic development.

Chapter 1

Segregation Patterns in Cities: Ethnic Clustering without Skill Differences[1]

1.1 Introduction

We often observe the residential separation of two or more ethnic groups into different neighborhoods, also known as segregation. For instance, ethnic areas such as Chinatowns have emerged in many places and immigrants or minorities sometimes reside in clusters. The term "segregation" has been negatively perceived because it often implies a gap between the rich and poor. Originally, however, segregation indicates a situation in which people of various ethnic or racial groups reside in clusters based on ethnic characteristics.

The phenomenon that minorities' residential distributions are more geographically concentrated can be found in cities in the United States, Europe, and elsewhere (Musterd, 2005). Some well-known examples of residential clustering include the Maghreb immigrants in France (Wacquant, 1993), Turkish immigrants in Germany (O'Loughlin, 1980), and Malayans in Singapore (Van Grunsven, 1992). Wacquant (1993) also discusses the difference between segregation in the ghettos of New York and in the suburbs of Paris.

The mechanism of segregation has been of great interest in the socio-economic field. However, because the U.S. inner-city problem has attracted social concerns, researchers of segregation mechanisms have focused on the uneven residential distribution of minorities in U.S. cities. Kain (1968), for example, examines the relationship between residential segregation by race and differences in unemployment rate (or income level) between races from a sociological viewpoint.[2] Indeed, researchers of segregation in U.S. cities have succeeded

[1] I would like to thank Takatoshi Tabuchi for the thoughtful comments and suggestions. I am also grateful to Dan Sasaki, Masahisa Fujita and the seminar participants at the University of Tokyo, ARSC meeting at Kyoto University and JEA meeting at Kanagawa University. Further, I appreciate anonymous referees' comments which have drastically improved this chapter. All remaining errors are on the author's responsibility. This research is partially supported by the Grants-in-Aid for Scientific Research (Research project number: 13J10130) for JSPS Fellows by the Ministry of Education, Science and Culture in Japan.

[2] "Spatial mismatch" is a social phenomenon wherein the area of job offered by firms and the residence of unemployed job applicants geographically differ. Theoretical explanations for the spatial mismatch

in theoretically explaining these mechanisms by associating ethnicity characteristics with income level (or years of education). Many studies have analyzed the segregation mechanism by examining the income levels of multiple types of households, using the traditional bid rent curve analysis of Alonso (1964),[3] analytically explaining U.S. inner-city problems and the tendency for whites to reside in the suburbs under the core assumption that the blacks have lower income levels. In other words, they assume the assumption that majorities are more likely to be better educated and more affluent than minorities.

However, in reality, some minorities are richer than majorities, although we cannot necessarily conclude that ethnicity characteristics are strongly associated with skill levels.[4] Taking this stance, this chapter does not assume that majorities are richer than minorities. Put differently, the analysis of ethnic segregation presented in this chapter is conducted under the assumption of skill/income-level homogeneity among ethnic groups, and assume that the majority and minority have identical skill levels. Discarding the usual assumption that the majority tends to be better educated or earn higher incomes is sometimes reasonable. In the United States, while Hispanics and African Americans are likely to be less educated than whites, the academic performance of Asians is superior still. Moreover, in countries such as Malaysia and South Africa, minorities (Chinese and whites, respectively) have a much stronger influence on the economy. Thus, one cannot assert that the income and skill levels of the majority are greater than those of the minority. What is unusual and possibly characterizes the present analysis is that even without assuming a difference in skill level between majority and minority groups, we still draw a conclusion that ethnic segregation occurs.

To go in this direction, more support for adopting the assumption that the group tendency of skill/income level of the majority and minority is identical is needed. Reardon et al. (2015) show that black and Hispanic middle-class households tend to reside in neighborhoods that contain larger proportions of their same ethnic groups than those of similar earning white households in U.S. cities, and argue that wealth differences alone do not explain the disproportionate residential concentration of black households. Bayer et al.'s (2014) empirical study more clearly shows this segregation tendency of black residents in U.S. cities, highlighting that the increasing proportion of better educated blacks is leading to an expanding set of available neighborhood options. As a result, highly educated black

hypothesis proposed by Kain (1968) are provided by, for example, Brueckner and Martin (1997) and Brueckner and Zenou (2003).

[3]See, for example, Fujita (1989, Chapter 4, Part I) for the segregation mechanism by income level for multiple types of households. As for the theoretical literature on segregation mechanisms, see Fujita and Thisse (2013, Chapters 6 and 7), who consider social interactions in a land market model without assuming a city center as exogenously given. In addition, the body of research from this perspective includes Mossay and Picard's (2013) segregation analysis.

[4]Nevertheless, we do not deny the existence of an ethnicity bias in skill levels. Coulton et al. (1996) suggest that the geographical concentrations of poverty and affluence can be partially explained by racial and ethnic segregation. Clark and Blue (2004) examine the relationship between residential separation and income or education levels.

residents are moving from predominantly white neighborhoods into those of middle-class blacks. This finding suggests that highly educated blacks prefer to live with black neighbors of middle- or high-income class if they are available. Even after eliminating income or skill level difference among ethnicities, which may be one of the reason to bring about ethnic segregation, there still remains preference to their own ethnicity.[5] In other words, even after eliminating the income- or skill-level differences among ethnicities, a preference for own ethnicity remains.

Which factors other than income and skill differences might cause ethnic segregation? The benefit provided by residential proximity to the same minority group, or ethnic residential clustering, especially for minority rather than majority residents, is one important factor. For example, minorities or immigrants are more likely to face language difficulties. As suggested by Yancey et al. (1976), the effect of an ethnic network is stronger among members who are geographically clustered, indicating that common occupational positions and dependence on local institutions and services are important factors bearing ethnic clustering benefits.

As for public goods provision, racial division may reflect different preferences for public goods consumption (Boustan, 2007). Further, Besley et al. (2004) argue that residential proximity and sharing group identity with local politicians play a role when considering public goods provision. Ethnicity clustering is also beneficial from an educational viewpoint. For instance, when a sufficient number of foreign children reside in a community, schools for such students may be more easily established.[6] The mechanism considered to explain the residential segregation of minorities is the fact that once a certain threshold is reached in a neighborhood, specific "community goods" may be provided to the minority, which increases its utility. In other words, better accessibility to "community goods" related to ethnic-oriented preferences is a key factor generating the benefits of ethnic clustering.

In terms of the housing market, minority communities may play important roles. Calomiris et al. (1994) show that community development banks may eliminate search costs by creating an alternative source of funds for minority residents in the mortgage

[5] In addition to these U.S. city examples, in Tokyo, foreign residents from North America or Europe, who are as wealthy as native Japanese residents, tend to reside in clusters in Minato, Setagaya, or Shibuya Wards. See http://www.toukei.metro.tokyo.jp/gaikoku/2015/ga15010000.htm (in Japanese).

[6] Some may deem that the way in which educational factors affect ethnic segregation is a sort of negative externality. For example, in the U.S. context, whites may avoid living close to blacks because blacks are less educated, and thus have children who perform worse at school, which may lead to negative externalities for white children. However, as in the examples of the clustering of American or European residents in Tokyo, these negative externalities caused by minorities' lesser educational attainments are not the only factor creating ethnic segregation, because such a large gap in educational level among foreign residents from rich countries and native Japanese might not exist. Nevertheless, American and European children in Tokyo are likely to choose international schools rather than Japanese public schools, which is indicated by the fact that a large proportion of international schools in Tokyo are located in Minato Ward, where American and European residents are clustered.

market. The benefits of ethnic clustering in the housing market have also been empirically shown. For instance, Søholt (2001) shows that ethnic groups with denser networks enjoy higher home acquisition rates. In addition, networks within the same ethnic group create more favorable consequences for minorities in the labor market. Mexican immigrants in the United States are more likely to be employed when the ethnic network is larger (Munshi, 2003). Dietz (1999) reveals that migrant networks positively affect the labor market performance of ethnic Germans, suggesting that such networks support minority enclaves in Germany.

To highlight the driving forces of ethnic clustering, we introduce the concept of "ethnic clustering preference" in the minority's utility function, since the above-mentioned factors are more related to minority residents than the majority group.[7] Unlike in the extant literature on segregation related to negative externalities such as prejudice against minority residents,[8] this chapter thus emphasizes aspects induced by the residential proximity in the minority group. As noted already, although differences in skill level among ethnic groups have long been considered an important factor of ethnic segregation, elements associated with the minority's preference cannot be ignored when analyzing residential segregation by ethnicity. Indeed, a segregation analysis under the assumption of identical skill-level tendencies by ethnicity can be meaningful, particularly if our primary interest is the effect of minority residents' clustering preferences. This chapter investigates this sorting process on the basis of a minority group's positive externalities rather than the negative externalities induced by residential proximity to other ethnic members, as in Rose-Ackerman (1975) and Yinger (1976). The present segregation analysis uses a model that captures the externalities stemming from residential proximity to the same ethnic community, especially for minority residents.[9]

In the present model, both high- and low-skilled workers are perfectly mobile within a closed city that contains the center (high-skilled intensive area) and suburb (low-skilled intensive area) as workplaces. There are two types of ethnic characteristics, majority and minority, in the economy; thus, in addition to a skill level, each individual is endowed with an ethnic characteristic. Further, we assume that both ethnic groups have the same high-/low-skilled population ratios. Minority residents obtain utility from proximity to residents of the same minority group. As for land consumption, for the sake of tractabil-

[7]Note that in this chapter, clustering does not mean industrial clustering or agglomeration as is used in new economic geography contexts.

[8]When considering the mechanism of segregation by ethnicity, especially in the context of the United States, whites' prejudice against minorities is a key issue. Rose-Ackerman (1975), Yinger (1976), and Courant and Yinger (1977) consider a situation in which white residents hesitate to live close to black residents because of the negative externalities created by such residential proximity.

[9]Kanemoto (1980, Chapter 7) suggests ways in which to express the externalities stemming from proximity to another type of household. Thus, this chapter defines externalities from proximity to the same type of household by employing Kanemoto's (1980) model, in which externalities are defined as negative ones that affect the utility levels of other ethnicity groups when near another community's residence.

ity, we draw on Helpman (1998) and assume that residents in the same area consume the same amount of land. If a minority individual's residence and workplace are in different areas, she faces a trade-off between the benefits of ethnic clustering and distant commuting (as well as disutility of residential congestion). The existence of this trade-off for minority people is empirically investigated by Liu (2009) in the context of Latino workers in the United States.

The main results are as follows. When the commuting cost is low, minority residents always cluster, while majority residents move to a less crowded area to avoid residential congestion. Although this model does not directly tackle the spatial mismatch hypothesis, its key factors are deeply related to some of those in this hypothesis such as commuting costs and access to job centers (Ihlanfeldt and Sjoquist, 1990, 1991), both of which are deemed to be important in the present context.

The remainder of this chapter is organized as follows. Section 1.2 presents the framework of the model. We employ Krugman's (1991) model as the base model.[10] Section 1.3 introduces the notion of a long-run residential equilibrium and classifies the possible equilibrium configurations. Sections 1.4 and 1.5 discuss the residential patterns in response to low and high commuting costs. Section 1.6 briefly discusses efficiency. Section 1.7 compares the population distribution gap between the central and suburban areas on the basis of stable residential equilibrium paths. Section 1.8 concludes.

1.2 Framework

1.2.1 Settings

This section models an economy in which people with various ethnic characteristics reside. As mentioned in Section 1.1, people with the same ethnic characteristics prefer to cluster, ceteris paribus. To define a mechanism that brings about this tendency, we first consider two types of ethnic groups—a majority group with a larger population, L_X, and a minority group with a smaller population, L_x. To express different population sizes, assume $L_X > L_x$. In addition, there are two types of workers: high- and low-skilled. κ_X is defined as the high-skilled share of the majority and κ_x the minority, such that $1 - \kappa_X$ is the low-skilled share of the majority and $1 - \kappa_x$ is that of the minority. In this way, the economy is characterized in terms of population share. There are four types of households in the economy: high-skilled majority, low-skilled majority, high-skilled minority, and low-skilled minority. Thus, the population of each household type is $\kappa_X L_X$ for the high-skilled majority, $(1 - \kappa_X)L_X$ for the low-skilled majority, $\kappa_x L_x$ for the high-skilled minority, and $(1 - \kappa_x)L_x$ for the low-skilled minority. By normalization, we set $L_X = 1$ so that $L_x <$

[10]An attempt was made to adopt a bid rent approach to deal with the topic in question, considering a spatially continuous city in this chapter; however, this makes the analysis cumbersome.

1. Following discussions on the ethnic segregation without skill heterogeneous tendency according to ethnic groups in Section 1.1, we set $\kappa_X = \kappa_x = 1/2$.[11]

Next, we describe the geographic characteristics of the economy. Consider a closed city with a central and suburban area. Since we assume a closed city, there is no migration into or out of the economy (of course, internal migration is allowed in the long run). Both the center and suburb are assumed to compose one unit of land and no space exists between them; thus, the center and suburb play the role of not only production areas but also residential areas. For the production areas, assume that the differentiated good is produced in the center, while the homogeneous good is produced in the suburb. Differentiated goods production needs high-skilled labor as input, whereas low-skilled labor is the sole input in homogeneous good production. We ignore goods transportation costs in the economy.[12] As for the residential areas, both the center and suburb accommodate households and each household incurs an iceberg-type commuting cost, τ, if she is a commuter (i.e., the workplace and residence are located in different areas). If she is a non-commuter (i.e., the workplace and residence are in the same area), there are no commuting costs. Because in some U.S. and European nations, not firms but workers incur commuting costs, it is not unnatural to assume that the income reduces in line with iceberg commuting costs.

1.2.2 Households

Because there are two types of households in terms of ethnic characteristics in the economy (majority and minority), different utility functions are assumed for each ethnic group. As mentioned in Section 1.1, people with the same ethnic characteristics tend to cluster and this tendency is stronger in the case of a minority. Thus, utility stemming from ethnic residential clustering, especially for the minority, should be included in the utility function. In addition, the utility function includes a land consumption term. Following Krugman's (1991) model, we formulate the utility function of a household living in area j ($j = C$ or S, C; C denotes the center and S the suburb) with ethnicity characteristic e ($e = X$ or x; X denotes a majority and x a minority) as follows:

$$U_e^j = \alpha \log M + (1 - \alpha - \beta) \log A + \beta \log(H^j) + \gamma_e \log(X_e^j) \tag{1.1}$$

[11] Some may suspect the existence a certain level of ethnicity bias. For example, in the Unites States, Hispanics and African Americans are likely to be less educated than the Whites, so that $\kappa_X > \kappa_x$ may hold. However, as mentioned in Section 1.1, one cannot necessarily assert that the high-skilled ratio of the majority is greater than that of the minority. In addition, in the present model, skill levels are exogenously given for each ethnicity. For a model in which skill differences endogenously emerge through neighborhood interactions in the presence of peer community influence, see Bénabou (1993) and Bénabou (1996). Further, Billings et al. (2014) empirically examines the effects of racial segregation on educational attainment.

[12] Here, although goods transportation costs are ignored, consumer commuting costs are not; such as the assumption employed in the context of a traditional bid rent curve analysis (Fujita, 1989).

with

$$M \equiv \left[\int_0^n m(i)^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}, \qquad \sigma > 1,$$

where M and A denote the consumption of the CES composite of varieties of the differentiated goods and the homogeneous good, respectively. $m(i)$ is the consumption of variety i, n is the mass of varieties produced in the economy, and $\sigma > 1$ is the elasticity of substitution between any two varieties (Dixit and Stiglitz, 1977). H^j is the amount of land consumption in area j and X_e^j is the composition of households in area j in terms of ethnicity; that is,

$$X_e^j \equiv \frac{N_e^j}{N^j} = \frac{N_e^j}{\sum_{e \in \{X,x\}} N_e^j},$$

where N^j is the total population residing in area j and N_e^j is the population with ethnic characteristic e living in area j.[13] The consumption share parameter α satisfies $0 < \alpha < 1$, and the housing consumption parameter β also satisfies $0 < \beta < 1$. As for the ethnicity clustering parameter γ_e, we assume

$$\gamma_e = \begin{cases} \gamma > 0 & \text{if} \quad e = x \\ 0 & \text{otherwise,} \end{cases}$$

which captures the tendency that minority households more strongly prefer residential proximity to the same ethnic groups than the majority.

With this expression of the ethnicity clustering parameter, minorities derive utility from clustered residences by ethnicity, while majority households are indifferent to it. However, the question remains whether majority households derive utility from clustering with people with the same ethnicity characteristics. In reality, even though a fervent answer cannot be provided, some defense can be offered. Clark and Blue (2004) empirically show that even if ethnicity characteristics differ between groups, their affinity levels toward other ethnic groups are not as low if the education (or income) levels are somewhat the same among groups. With this, consider the assumption of a lack in the majority's clustering preference by ethnicity. Suppose that the social status of the majority is higher than that of the minority (unlike in the present model), so that the income levels widely vary between ethnic groups. In this case, a realistic reason underlying the majority's propensity to reside in a cluster is that it would be safer, that is, richer areas are thought to be safer with lower crime rates. In this model, however, skill levels (high-skilled ratios) are assumed to be the same between the majority and minority, and thus, the anxiety that

[13] This expression of residential externality from different racial households living in a neighborhood is borrowed from Kanemoto (1980). It may be questioned as to why the ethnicity externality term X_e^j is not a level but a share. As mentioned in Section 1.1, local governments' decisions are more likely to be influenced by minority residents, that is, if they are not negligible. In other words, the ethnicity composition of the area may be directly related to the level of ethnic clustering utility, making the employment of a share more rationalizable than a level.

the majority residents feel regarding safety can be disregarded.

Next, consider the budget constraint of households. The wage rate of a household with skill level s ($s = h$ or l; h denotes high- and l low-skilled) is denoted by w_s, and each household is assumed to be endowed with one unit of labor. An individual supplies her labor endowment inelastically, and thus, the household income with skill level s is w_s. In addition, she pays rent for land consumption and we assume that the land in area j is equally owned by the residents in the area, so that each household earns land rent income. The assumption that all land rent is collected and equally redistributed among the residents in the same area is also adopted in Ottaviano et al. (2002). We choose the homogeneous good as the numéraire and denote the price of variety i by $p(i)$. Then, the budget constraint of a household that possesses skill level s, lives in area j, and works in area k is given by

$$\int_0^n p(i)m(i)di + A + r^j H^j = \frac{w_s}{\tau^{jk}} + R^j, \tag{1.2}$$

where

$$\tau^{jk} = \begin{cases} \tau > 1 & \text{if} \quad j \neq k \qquad \text{(commuter)} \\ 1 & \text{otherwise} \quad \text{(non-commuter).} \end{cases}$$

As mentioned above, commuters incur commuting costs, while non-commuters do not. r^j is the rent for land consumption H^j, and R^j is the land rent paid to her. Following Helpman (1998), we assume that the total supply of land in area j is set to unity, and each resident in area j owns and consumes the same amount of land. Then, we get the following utility function and budget constraint:[14]

$$U_e^j = \alpha \log M + (1 - \alpha - \beta) \log A + \beta \log \left(\frac{1}{N^j} \right) + \gamma_e \log \left(\frac{N_e^j}{N^j} \right) \tag{1.3}$$

and

$$\int_0^n p(i)m(i)di + A = \frac{w_s}{\tau^{jk}}. \tag{1.4}$$

In utility function (1.3), the interpretation of β is important. Since the land consumption per person in area j, $1/N^j$, decreases as the total population of area j increases, a larger population in area j (the area where the household resides) means lower utility from housing consumption. This implies an alternate interpretation of the term $\beta \log(1/N^j)$, which is the disutility created by residential congestion. It turns out that in the long-run equilibrium analysis, the congestion parameter β is a key parameter. Each household

[14]Some may oppose the assumption that land is equally consumed and owned by residents in the same area (public land ownership by area). Thus, we introduce certain modifications to the model in terms of land consumption and ownership. First, we remove the assumption that each resident in the same area consumes the same amount of land, and assume public land ownership, where residents obtain the same share of total land rent in the city. Numerical exercises show that the outcome does not contradict the analytical solutions of the original model; thus, the result obtained from the main analysis on the basis of this assumption can be deemed robust. Details of the analysis can be made available upon request.

maximizes the utility function (1.3) subject to the budget constraint (1.4). The first-order conditions yield the following demand functions:

$$m(i) = \alpha \frac{w_s}{\tau^{jk}} p(i)^{-\sigma} P^{\sigma-1}, \qquad M = \alpha \frac{w_s}{\tau^{jk}} P^{-1}, \qquad A = (1-\alpha-\beta)\frac{w_s}{\tau^{jk}}, \tag{1.5}$$

where P is the price index of the differentiated good:

$$P \equiv \left[\int_0^n p(i)^{1-\sigma} di\right]^{\frac{1}{1-\sigma}}.$$

1.2.3 Production

Firms in the differentiated goods sector are monopolistically competitive and employ only high-skilled workers. In addition, we assume free entry and exit. To produce $q(i)$ units of variety i, $f + cq(i)$ units of high-skilled input are required. The fixed requirement of high-skilled labor f exhibits increasing returns to scale, so that each variety i is produced by a single firm. On the other hand, the homogeneous good sector is perfectly competitive and employs low-skilled labor as its only input. To produce one unit of a homogeneous good, one unit of low-skilled labor input is required, so that the technology in the homogeneous good sector exhibits constant returns to scale. Since the low-skilled labor market is perfectly competitive, the wage rate in the homogeneous good sector, or the wage rate of low-skilled workers, is 1 ($w_l = 1$).

Firm i maximizes its profit:

$$\Pi(i) = [p(i) - cw_h]\, D(i) - fw_h, \tag{1.6}$$

where $D(i)$ is the total demand for the variety produced by firm i. By profit maximization, the price of variety i is

$$p(i) = \frac{\sigma}{\sigma-1} cw_h, \tag{1.7}$$

which is a constant markup on the marginal cost. Because of free entry and exit and the zero profit condition, the high-skilled labor demand of a firm is rewritten as

$$f + cq(i) = \sigma f. \tag{1.8}$$

On the other hand, the high-skilled labor supply as a whole[15] is given by

$$\sum_{e\in\{X,x\}} \kappa_e L_e = \kappa_X L_X + \kappa_x L_x = \frac{1}{2}(1 + L_x),$$

[15] Here, we simply assume fixed working hours, such that the high-skilled labor supply equals the total amount of high-skilled labor.

so that for the high-skilled labor market to clear, the equilibrium mass of firms is

$$n^* = \frac{1 + L_x}{2\sigma f}.$$

Thus, we obtain the price index

$$P = \frac{\sigma}{\sigma - 1} c w_h \left(\frac{1 + L_x}{2\sigma f} \right)^{\frac{1}{1-\sigma}}. \tag{1.9}$$

1.2.4 Instantaneous equilibrium (high-skilled wage determination)

In this section, we derive the income of high-skilled workers. First, we examine how many residents in the economy choose to commute. We define λ_{es} as the center-residing ratio of households with skill level s and ethnic characteristic e, so that $1 - \lambda_{es}$ is the suburb-residing ratio of households es. Because the consumption amount of differentiated and homogeneous goods of commuters is less than that of non-commuters owning to commuting costs, defining λ_{es} allows us to derive the net income (in the present context, this is the income remaining after paying commuting costs). The total net income of each group of households es, y_{es}, is then given by

$$y_{Xh} = \frac{w_h}{2} \left(\lambda_{Xh} + \frac{1 - \lambda_{Xh}}{\tau} \right), \quad y_{Xl} = \frac{1}{2} \left(\frac{\lambda_{Xl}}{\tau} + 1 - \lambda_{Xl} \right)$$
$$y_{xh} = \frac{w_h L_x}{2} \left(\lambda_{xh} + \frac{1 - \lambda_{xh}}{\tau} \right), \quad y_{xl} = \frac{L_x}{2} \left(\frac{\lambda_{xl}}{\tau} + 1 - \lambda_{xl} \right).$$

From these equations, we obtain the net income of the economy as a whole:

$$\begin{aligned} Y &= \sum_{e \in \{X,x\}} \sum_{s \in \{h,l\}} y_{es} \\ &= \frac{w_h}{2} \left[\lambda_{Xh} + \frac{1 - \lambda_{Xh}}{\tau} + L_x \left(\lambda_{xh} + \frac{1 - \lambda_{xh}}{\tau} \right) \right] \\ &+ \frac{1}{2} \left[\frac{\lambda_{Xl}}{\tau} + 1 - \lambda_{Xl} + L_x \left(\frac{\lambda_{xl}}{\tau} + 1 - \lambda_{xl} \right) \right]. \end{aligned}$$

Combining this expression with the zero-profit condition (setting $\Pi(i) = 0$ in (1.6)), (1.7), (1.8), and (1.9) with the demand functions in (1.5) yields the equilibrium wage rate for high-skilled workers:

$$w_h^* = \frac{\alpha \sum_{e\in\{X,x\}}(1-\kappa_e)L_e\left(\frac{\lambda_{el}}{\tau}+1-\lambda_{el}\right)}{\sum_{e\in\{X,x\}}\kappa_e L_e\left[1-\alpha\left(\lambda_{eh}+\frac{1-\lambda_{eh}}{\tau}\right)\right]}$$
$$= \frac{\alpha\left[\frac{\lambda_{Xl}}{\tau}+1-\lambda_{Xl}+L_x\left(\frac{\lambda_{xl}}{\tau}+1-\lambda_{xl}\right)\right]}{1+L_x-\alpha\left[\lambda_{Xh}+\frac{1-\lambda_{Xh}}{\tau}+L_x\left(\lambda_{xh}+\frac{1-\lambda_{xh}}{\tau}\right)\right]}.$$

1.3 Residential equilibrium

In this section, we analyze the long-run equilibrium, where each household can migrate within this economy (λ is no longer treated as fixed); that is, households can choose to reside in either the center or suburb. To analyze the long-run behavior of households, we compute indirect utility differentials. For a household with skill level s and ethnicity characteristic e, the indirect utility differential is defined as

$$\Delta V_{es}(\lambda) \equiv V_{es}^C(\lambda) - V_{es}^S(\lambda), \tag{1.10}$$

where $\lambda \equiv (\lambda_X, \lambda_x) \equiv (\lambda_{Xh}, \lambda_{Xl}, \lambda_{xh}, \lambda_{xl})$. In the definition provided in (1.10), $V_{es}^j(\lambda)$ is the indirect utility an individual can attain when residing in area j, so that $\Delta V_{es}(\lambda)$ is the extra utility attained when residing in the center rather than in the suburb. She chooses to reside in the center if $\Delta V_{es}(\lambda) > 0$, and in the suburb if $\Delta V_{es}(\lambda) < 0$, in the long run.

1.3.1 Majority's residential patterns

By definition of the indirect utility differential, we can identify $\Delta V_{Xh}(\lambda)$ and $\Delta V_{Xl}(\lambda)$. Further, by plugging the budget constraint (1.4) into the utility function (1.3), and replacing w_s with $w_h = w_h^*$ and $w_l = 1$, respectively, the indirect utilities of the high- and low-skilled majority living in the center and suburb are

$$\left(V_{Xh}^C(\lambda), V_{Xh}^S(\lambda)\right) = \left(B_h - \beta\log(N^C), B_h - \log\tau - \beta\log(N^S)\right)$$
$$\left(V_{Xl}^C(\lambda), V_{Xl}^S(\lambda)\right) = \left(B_l - \log\tau - \beta\log(N^C), B_l - \beta\log(N^S)\right),$$

where $B_h \equiv \alpha\log(\alpha w_h^* P^{-1}) + (1-\alpha)\log[(1-\alpha)w_h^*]$ and $B_l \equiv \alpha\log(\alpha P^{-1}) + (1-\alpha)\log(1-\alpha)$, respectively, so that the indirect utility differentials of high- and low-skilled majority are given by

$$(\Delta V_{Xh}(\lambda), \Delta V_{Xl}(\lambda)) = \left(\log\tau - \beta\log\left(\frac{N^C}{N^S}\right), -\log\tau - \beta\log\left(\frac{N^C}{N^S}\right)\right). \tag{1.11}$$

Note that indirect utility differentials do not depend on individual income levels because one's place of residence does not affect earned income. In fact, the only factor affecting income is skill level. Thus, after subtracting the indirect utilities when living in areas j and k, incomes cancel each other out, so that there should be no residual terms related to wages. This result means that the indirect utility differential can ignore the wage level.

By analyzing (1.11), we obtain the following lemma.[16]

Lemma 1.3.1. *If the commuting cost is low* ($1 < \tau < \tau_X \equiv (1 + 2L_x)^\beta$)*, then there are two possibilities:*
(M_HD/LS): low-skilled majority households cluster in the suburb, whereas high-skilled majority households reside in both areas ($\lambda^*_{Xh} \in (0, 1)$ *and* $\lambda^*_{Xl} = 0$)*.*
(M_HC/LD): high-skilled majority households cluster in the center, whereas low-skilled majority households reside in both areas ($\lambda^*_{Xh} = 1$ *and* $\lambda^*_{Xl} \in (0, 1)$)*.*
If the commuting cost is high ($\tau \geq \tau_X$)*, then*
(M_HC/LS): high-skilled majority households cluster in the center, whereas low-skilled majority households cluster in the suburb, no majority households commute ($\lambda^*_{Xh} = 1$ *and* $\lambda^*_{Xl} = 0$)*.*[17]

To elaborate on Lemma 1.3.1,[18] when the commuting cost is low, avoiding residential congestion is more important for consumers than low commuting costs; thus, they prefer to live in a less crowded area. Note that not all consumers live in the same residential area to avoid heavy residential congestion. On the other hand, in the case of high commuting costs, consumers choose to reside in the area of their workplace because they can avoid paying expensive commuting costs and are less concerned about residential congestion within their neighborhoods.

1.3.2 Minority's residential patterns

As with the majority's residential patterns, we identify $\Delta V_{xh}(\lambda)$ and $\Delta V_{xl}(\lambda)$. Because the indirect utilities of the high- and low-skilled minority living in the center and suburb are

$$
\begin{aligned}
&\left(V^C_{xh}(\lambda), V^S_{xh}(\lambda)\right)\\
&= \left(B_h - (\beta + \gamma)\log(N^C) + \gamma\log(N^C_x), B_h - \log\tau - (\beta + \gamma)\log(N^S) + \gamma\log(N^S_x)\right)\\
&\left(V^C_{xl}(\lambda), V^S_{xl}(\lambda)\right)\\
&= \left(B_l - \log\tau - (\beta + \gamma)\log(N^C) + \gamma\log(N^C_x), B_l - (\beta + \gamma)\log(N^S) + \gamma\log(N^S_x)\right),
\end{aligned}
$$

[16] See Appendix 1.A for the proof.

[17] Note that, in this model, "cluster" does not mean industrial clustering but residential clustering in terms of ethnicity.

[18] (M_HD/LS) stands for a pattern in which Majority, High-skilled workers Disperse across areas and Low-skilled workers cluster in the Suburb. (M_HC/LS) means that Majority, High-skilled workers cluster in the Center and Low-skilled workers cluster in the Suburb.

respectively, the indirect utility differentials of the high- and low-skilled minority are given by

$$(\Delta V_{xh}(\lambda), \Delta V_{xl}(\lambda)) \\ = \left(\log\tau - (\beta+\gamma)\log\left(\frac{N^C}{N^S}\right) + \gamma\log\left(\frac{N_x^C}{N_x^S}\right), -\log\tau - (\beta+\gamma)\log\left(\frac{N^C}{N^S}\right) + \gamma\log\left(\frac{N_x^C}{N_x^S}\right)\right). \tag{1.12}$$

Analyzing (1.12) yields the following possible cases of minority residential patterns.[19]

- **(m_HC/LC)**:
  $B_x < \tau^{-1} \quad\Rightarrow\quad \Delta V_{xh}(\lambda) > 0$ and $\Delta V_{xl}(\lambda) > 0 \quad\Rightarrow\quad \lambda^*_{xh} = 1$ and $\lambda^*_{xl} = 1$
- **(m_HC/LD)**:
  $B_x = \tau^{-1} \quad\Rightarrow\quad \Delta V_{xh}(\lambda) > 0$ and $\Delta V_{xl}(\lambda) = 0 \quad\Rightarrow\quad \lambda^*_{xh} = 1$ and $\lambda^*_{xl} \in (0,1)$
- **(m_HC/LS)**:
  $\tau^{-1} < B_x < \tau \quad\Rightarrow\quad \Delta V_{xh}(\lambda) > 0$ and $\Delta V_{xl}(\lambda) < 0 \quad\Rightarrow\quad \lambda^*_{xh} = 1$ and $\lambda^*_{xl} = 0$
- **(m_HD/LS)**:
  $B_x = \tau \quad\Rightarrow\quad \Delta V_{xh}(\lambda) = 0$ and $\Delta V_{xl}(\lambda) < 0 \quad\Rightarrow\quad \lambda^*_{xh} \in (0,1)$ and $\lambda^*_{xl} = 0$
- **(m_HS/LS)**:
  $B_x > \tau \quad\Rightarrow\quad \Delta V_{xh}(\lambda) < 0$ and $\Delta V_{xl}(\lambda) < 0 \quad\Rightarrow\quad \lambda^*_{xh} = 0$ and $\lambda^*_{xl} = 0$,

where $B_x \equiv \left(N^C/N^S\right)^{\beta+\gamma}\left(N_x^C/N_x^S\right)^{-\gamma}$.[20] A review of the above five cases shows that the residential patterns with $\lambda^*_{xh} \in (0,1)$ and $\lambda^*_{xl} \in (0,1)$, $\lambda^*_{xh} = 0$ and $\lambda^*_{xl} \in (0,1)$, and $\lambda^*_{xh} \in (0,1)$ and $\lambda^*_{xl} = 1$ are excluded. The excluded residential patterns correspond with the situations in which both high- and low-skilled minority workers commute at the same time.[21]

Since this section lists the possible residential equilibrium patterns for majorities and minorities, in the next section, we examine the combinations of majority and minority residential patterns that are in equilibrium.

[19]The possible residential patterns of the minority are derived in almost the same manner as those of the majority. Details are provided for readers upon request.

[20](m_HC/LS) stands for a pattern where the minority, High-skilled workers cluster in the Center and Low-skilled workers who cluster in the Suburb. (m_HC/LD) stands for a pattern where as for the minority, High-skilled workers cluster in the Center and Low-skilled workers Dispersed across both areas.

[21]If $\lambda_{xh} \neq 1$, some high-skilled minority workers commute because their workplaces are in the center. (Note that if $\lambda_{xh} \neq 1$, some high-skilled minority workers live in the suburb.) Similarly, if $\lambda_{xl} \neq 0$, some low-skilled minorities commute because their workplaces are in the suburb. (Note that if $\lambda_{xl} \neq 0$, some low-skilled minority residents live in the center.) For a further discussion, see Appendix 1.B.

1.4 Equilibrium residential patterns for low commuting costs

Here, we consider the residential equilibrium patterns that may emerge in the case of low commuting costs ($1 < \tau < \tau_X$). When considering the two majority and five minority residential patterns, we do not need to examine all 10 combinations. Adopting the assumptions for population and skill levels, we proved that only (M_HD/LS) (m_HC/LC) and (M_HC/LD) (m_HS/LS) hold.[22]

By analyzing (M_HD/LS) (m_HC/LC), we obtain a stable equilibrium,[23]

$$\lambda_{Xh}^* = \frac{2(\tau^{\frac{1}{\beta}} - L_x)}{\tau^{\frac{1}{\beta}} + 1} \in (1 - L_x, 1), \quad \lambda_{Xl}^* = 0, \quad \lambda_{xh}^* = 1, \quad \lambda_{xl}^* = 1.$$

In this residential equilibrium, λ_{Xh}^* increases as τ increases. Since λ_{Xh}^* is the ratio of high-skilled majority households living in the center, if commuting becomes more expensive, fewer high-skilled majority workers will choose to commute and the number of high-skilled majority households preferring to reside in the center compared to the suburb increases. Next, consider the effect on λ_{Xh}^* by L_x and β, the parameters related to residential congestion. λ_{Xh}^* decreases as L_x and β increase. In this residential equilibrium, the total population in the center is larger than that in the suburb. This implies that if the high-skilled majority's negative response to congestion worsens (i.e., β gets larger), they will tend to increasingly flee from the center to the suburb to escape the congestion. Similarly, because all minority households reside in the center, a larger L_x means an increasing central population. This population increase forces the high-skilled majority to migrate to the suburb.

As in the combination (M_HD/LS) (m_HC/LC), (M_HC/LD) (m_HS/LS) is proved to bear a stable equilibrium,[24]

$$\lambda_{Xh}^* = 1, \quad \lambda_{Xl}^* = \frac{1 + 2L_x - \tau^{\frac{1}{\beta}}}{\tau^{\frac{1}{\beta}} + 1} \in (0, L_x), \quad \lambda_{xh}^* = 0, \quad \lambda_{xl}^* = 0.$$

Examining the residential equilibrium, we see that a larger τ decreases λ_{Xl}^*, whereas an increase in β and L_x increases λ_{Xl}^*. The interpretations of these impacts are almost the same as in (M_HD/LS) (m_HC/LC). To sum up this section, we present the following proposition.

Proposition 1.4.1. *When the commuting cost is low ($1 < \tau < \tau_X$), two types of residential equilibria emerge:*

[22] For the proof, see Appendix 1.C.

[23] The stability notion used here is that of local stability with respect to relocation dynamics to higher utility locations. For the derivation and more detailed explanation of equilibrium stability, see Appendix 1.D.

[24] Since the proof is similar to that in the cases of (M_HD/LS) (m_HC/LC) in Appendix 1.D, it has been omitted.

Pattern Lτ-mC: Minority clustering in the center ($\lambda^* = \left(\frac{2(\tau^{\frac{1}{\beta}} - L_x)}{\tau^{\frac{1}{\beta}}+1}, 0, 1, 1\right)$) *corresponding to (M_HD/LS) (m_HC/LC)*

Pattern Lτ-mS: Minority clustering in the suburb ($\lambda^* = \left(1, \frac{1+2L_x - \tau^{\frac{1}{\beta}}}{\tau^{\frac{1}{\beta}}+1}, 0, 0\right)$) *corresponding to (M_HC/LD) (m_HS/LS).*

Both residential patterns are perfectly mixed in terms of skill levels and imperfectly mixed (or equivalently, partially segregated) in terms of ethnic characteristics.

Proposition 1.4.1 can be restated in a simpler way: If the commuting cost is low, the minority group always clusters and there is perfect skill mixing. In addition, ethnic mixing occurs only in the location where the minority group clusters.[25] Figure 1.1 graphically depicts patterns Lτ-mC and Lτ-mS.[26] In pattern Lτ-mC, the suburb is occupied only by

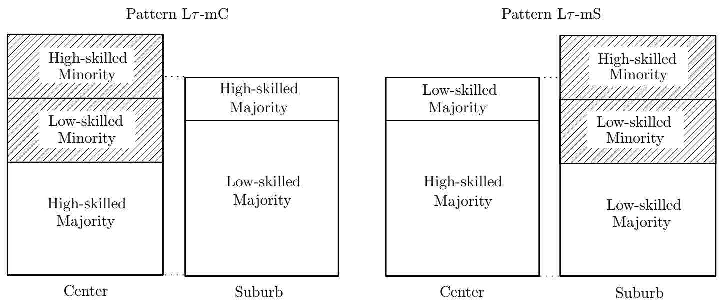

Figure 1.1: Patterns of residential equilibrium in the case of low commuting costs

the majority group, while the center is occupied by both majority and minority residents. Thus, in terms of ethnicity characteristics, this residential pattern is called an imperfectly mixed pattern (featured on the suburb), or equivalently, a partially segregated pattern (featured on the center). As for skill levels, this pattern is interpreted as a perfectly mixed pattern because both areas accommodate high- and low-skilled workers. A corresponding example of this residential pattern is Harlem, New York, where African Americans live in the central area and the Whites live in the suburban area. In Lτ-mS, the center is occupied only by majority residents, while the suburb is occupied by both majority and minority residents. In the real world, Lτ-mS may correspond to the Banlieue suburb in Paris, where immigrants from the Maghreb mainly reside.

We conclude this section with a comment on patterns Lτ-mC and Lτ-mS. In Lτ-mC, the population in the center (the more crowded area in the economy) can be denoted as $N^C = \tau^{1/\beta}(1 + L_x)/(1 + \tau^{1/\beta})$. In L$\tau$-mS, the population in the suburb (the more crowded area) is also represented by $N^S = \tau^{1/\beta}(1 + L_x)/(1 + \tau^{1/\beta})$. Therefore, $N^j =$

[25] I thank an anonymous referee for suggesting this clear restatement of Proposition 1.4.1.

[26] Pattern Lτ-mC stands for a residential pattern where for a Low τ (commuting cost), minority residents cluster in the Center.

$\tau^{1/\beta}(1+L_x)/(1+\tau^{1/\beta})$ is the population size when the majority residents reach the limit of their endurance of disutilities stemming from residential congestion in area j, where they live. In contrast, the less crowded area still has room to accommodate more residents (so that the residents in the crowded area can migrate to the less crowded area). This is the essence of the later discussion on equilibrium paths in terms of skill levels (Section 1.7.2).

1.5 Equilibrium residential patterns for high commuting costs

In this section, the commuting cost τ is so high ($\tau \geq \tau_X$) that no majority workers commute. Therefore, the population distribution of the majority is already given by $\lambda_X^* = (1, 0)$ and what is left to be analyzed is λ_x^*. Before analyzing this, we consider the "sustain points" for high- and low-skilled minority workers. In this chapter, the "sustain point" is the threshold value of the commuting cost such that either high- or low-skilled minority workers start commuting. That is, the sustain point for high-skilled minority workers, τ_{xh}^S, is defined as $\tau_{xh}^S \in \{\tau \mid \Delta V_{xh}(\lambda) \mid_{\lambda=(1,0,1,0)}= 0\}$, and that for low-skilled minority workers, τ_{xl}^S, is defined as $\tau_{xl}^S \in \{\tau \mid \Delta V_{xl}(\lambda) \mid_{\lambda=(1,0,1,0)}= 0\}$. By definition of the sustain points, $\Delta V_{xh}(\lambda) \mid_{\lambda=(1,0,1,0)}> 0$ for $\tau > \tau_{xh}^S$ and $\Delta V_{xl}(\lambda) \mid_{\lambda=(1,0,1,0)}< 0$ for $\tau > \tau_{xl}^S$. Then, from $\Delta V_{xh}(\lambda) \mid_{\lambda=(1,0,1,0)}= 0$ and $\Delta V_{xl}(\lambda) \mid_{\lambda=(1,0,1,0)}= 0$, we obtain $\tau_{xh}^S = \tau_{xl}^S = 1$. Since we now consider the case with $\tau \geq \tau_X = (1 + 2L_x)^\beta$, the commuting cost is always greater than the sustain points for both high- and low-skilled minority workers. This implies the following lemma.

Lemma 1.5.1. *If the commuting cost is sufficiently high ($\tau \geq \tau_X$), the residential pattern such that no one commutes is a stable equilibrium ($\lambda^* = (1, 0, 1, 0)$).*

In other words, if the commuting cost is high, residents hesitate to incur commuting costs so that households are likely to choose the non-commuting option. Lemma 1.5.1 means that combination (M_HC/LS) (m_HC/LS) is a stable equilibrium.

1.5.1 High-skilled minority's residential distribution

We have just seen that the combination (M_HC/LS) (m_HC/LS) shows a stable equilibrium, but for the convenience of further explanation, consider (M_HC/LS) (m_HC/LS), (M_HC/LS) (m_HD/LS), and (M_HC/LS) (m_HS/LS) together. In all of these combinations, the population distribution of low-skilled minority workers is $\lambda_{xl}^* = 0$, while that of high-skilled minority workers remains undetermined. Since the population distribution of low-skilled minority workers (and, of course, that of majorities) can be treated as fixed, we can focus on the residential decision of the high-skilled minority. In the combinations under consideration, the population of each area is given by $N^C = (1 + \lambda_{xh} L_x)/2$, $N^S = [1 + (2 - \lambda_{xh})L_x]/2$, $N_x^C = \lambda_{xh} L_x/2$ and $N_x^S = (2 - \lambda_{xh})L_x/2$, where $\lambda_{xh} \in [0, 1]$.

First, we immediately derive one of the stable equilibria, $\lambda^*_{xh} = 0$, from this population and define $\lambda_{-xh} \equiv (\lambda_{Xh}, \lambda_{Xl}, \lambda_{xl})$ for the convenience of the later discussion. Because $\lim_{\lambda_{xh} \to 0} \Delta V_{xh}(\lambda) \mid_{\lambda_{-xh}=(1,0,0)} = -\infty$, assuming τ is finite, we have $\Delta V_{xh}(\lambda^*) < 0$ when $\lambda^*_{xh} = 0$.

Lemma 1.5.2. *Residential clustering of the minority in the suburb is always a stable equilibrium under sufficiently high commuting costs* ($\tau \geq \tau_X \Rightarrow \lambda^* = (1, 0, 0, 0)$)*.*

According to Lemma 1.5.2, (M_HC/LS) (m_HS/LS) is a stable equilibrium. This lemma is not as intuitive as Lemma 1.5.1. In the case of high commuting costs (Lemma 1.5.2), no majority chooses to commute and they reside in the area where they work. This means that both areas are not prohibitively crowded, so that minority residents do not hesitate to incur disutilities stemming from residential congestion. Instead, they may prefer to gain utility from ethnicity clustering, and thus, residential clustering may occur in one area (in this case, the suburb).[27]

1.5.2 Low-skilled minority's residential distribution

In (M_HC/LS) (m_HC/LC), (M_HC/LS) (m_HC/LD), and (M_HC/LS) (m_HC/LS), the population distribution of high-skilled minority workers is $\lambda^*_{xh} = 1$, while that of low-skilled minority workers remains undetermined. The population of each area is then given by $N^C = [1 + (1 + \lambda_{xl})L_x]/2$, $N^S = [1 + (1 - \lambda_{xl})L_x]/2$, $N^C_x = (1 + \lambda_{xl})L_x/2$ and $N^S_x = (1 - \lambda_{xl})L_x/2$, where $\lambda_{xl} \in [0, 1]$. Similar to that in the previous section, it is shown that $\lambda^*_{xl} = 1$ is always a stable equilibrium because $\lim_{\lambda_{xl} \to 1} \Delta V_{xl}(\lambda) \mid_{\lambda_{-xh}=(1,0,1)} = \infty$, assuming τ is finite.

Lemma 1.5.3. *Residential clustering of the minority in the center is always a stable equilibrium under sufficiently high commuting costs* ($\tau \geq \tau_X \Rightarrow \lambda^* = (1, 0, 1, 1)$)*.*

According to Lemma 1.5.3, (M_HC/LS) (m_HC/LC) is a stable equilibrium. Because the interpretation of this proposition is the same as that of Lemma 1.5.2, it is omitted. As for an interior equilibrium, no solution exhibits stability such as in the analysis of the residential distribution of the high-skilled minority, thus there is no further discussion.[28]

Summing up Lemmas 1.5.1–1.5.3, we obtain the following proposition, although it contains certain repetitive messages.

Proposition 1.5.1. *When the commuting cost is high* ($\tau \geq \tau_X$)*, three types of residential equilibria emerge:*
Pattern Hτ-mD: Minority and majority dispersion across both areas ($\lambda^* = (1, 0, 1, 0)$) *corresponding to (M_HC/LS) (m_HC/LS)*

[27] In this model, there is no stable interior equilibrium for the minority residential distribution. For a further discussion, see Appendix 1.E.

[28] See Appendix 1.E.

Pattern Hτ-mC: Minority clustering in the center and majority dispersion ($\lambda^* = (1, 0, 1, 1)$) *corresponding to (M_HC/LS) (m_HC/LC)*
Pattern Hτ-mS: Minority clustering in the suburb and majority dispersion ($\lambda^* = (1, 0, 0, 0)$) *corresponding to (M_HC/LS) (m_HS/LS)*
Pattern Hτ-mD exhibits a perfectly segregated pattern in terms of skill levels, while patterns Hτ-mC and Hτ-mS exhibit imperfectly segregated (or equivalently, partially mixed) patterns. From the viewpoint of ethnicity characteristics, Hτ-mD exhibits a perfectly mixed pattern, whereas Hτ-mC and Hτ-mS exhibit imperfectly segregated patterns (partially mixed patterns).

Simply put, Proposition 1.5.1 states that if the commuting cost is high, the majority group never commutes, and the minority group does not commute or clusters either in the center or suburb.[29] Figure 1.2 shows the above three equilibrium patterns.[30] In pattern Hτ-mD,

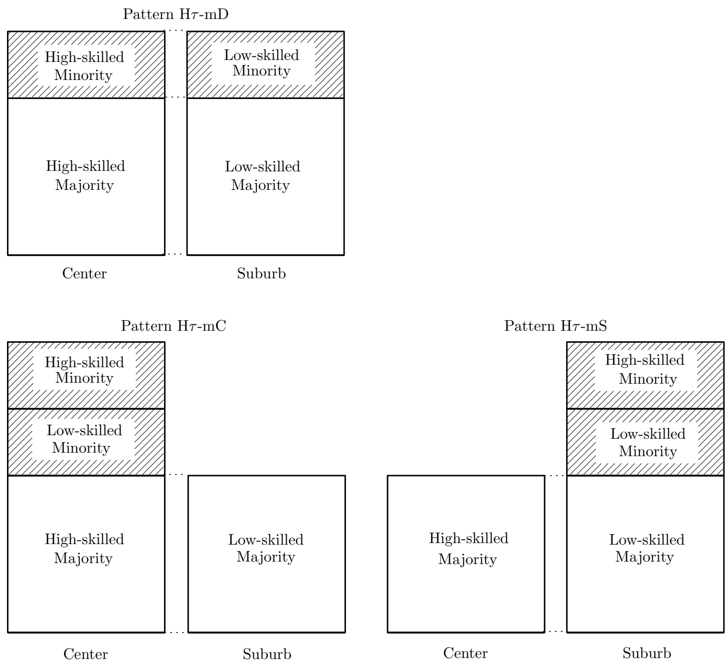

Figure 1.2: Patterns of residential equilibrium in the case of high commuting costs

the commuting cost is sufficiently high, so that no one commutes. This implies that only high-skilled (low-skilled) workers reside in the center (suburb). In this sense, this pattern shows perfect segregation with respect to skill levels. In addition, it shows that both majority and minority households reside in both areas. In fact, this is the most remarkable

[29] I thank an anonymous referee for suggesting this clear restatement.

[30] Pattern Hτ-mD stands for a residential pattern in which for a High τ (commuting cost), minority households Disperse across both areas. Pattern Hτ-mC denotes a residential pattern where for a High τ, minority households cluster in the Center.

feature of this residential pattern because no other residential equilibrium shows perfectly mixed patterns of ethnicity characteristics. In pattern Hτ-mC, the center accommodates both high- and low-skilled residents, while the suburb accommodates only low-skilled residents. Thus, this residential equilibrium is an imperfectly segregated (or partially mixed) pattern of skill levels. From the viewpoint of ethnic characteristics, this pattern also exhibits an imperfectly segregated (partially mixed) pattern because the center is occupied by both majority and minority residents, whereas the suburb is only occupied by the minority group. For pattern Hτ-mS, the center is completely occupied by high-skilled majority residents and the remaining residents are accommodated in the suburb. Hence, this pattern is interpreted as an imperfectly segregated (partially mixed) pattern in terms of ethnicity characteristics, as well as skill levels, as in pattern Hτ-mC.

The main messages of this analysis are contained in Propositions 1.4.1 and 1.5.1. Minority group households always cluster when the commuting cost is low, but they may or may not cluster residentially when the commuting cost is high. We conclude this section with a comment on Propositions 1.4.1 and 1.5.1, which is loosely related to the real world. Combining Propositions 1.4.1 and 1.5.1, it may be asserted that under less costly commuting, ethnic segregation is more likely to occur compared to the case with more costly commuting. In Musterd and De Winter (1998) and Musterd (2005), European cities exhibit less ethnically segregated patterns than U.S. cities. As for commuting, on the other hand, the average journey-to-work travel time is shorter in U.S. cities than in European cities, meaning that commuting is more costly in Europe than in the United States. (Kenworthy and Laube, 1999). Combining these two empirical results, cities with more costly commuting exhibit less segregated ethnic distributions as in Europe, while those with less costly commuting show a high segregation pattern as in the United States, which loosely matches the theoretical results in Propositions 1.4.1 and 1.5.1.

1.6 Discussion on efficiency

Thus far, we have derived equilibrium population distribution by ethnicities and skill levels. Solely investigating equilibrium patterns may sometimes be inadequate, especially in normative viewpoints. Then, hereafter, we briefly discuss Pareto efficiency, i.e., whether equilibrium distribution λ^* is Pareto efficient or not.

By definition of Pareto efficiency, a residential distribution λ is Pareto efficient if someone's utility level cannot be increased without decreasing that of other individuals. In our context, by this, equilibrium population distribution λ^* is *not* Pareto efficient if there exists other population distribution λ such that the utility level of households es living in area j increases and that of other households does not decrease. On the contrary, it can be asserted that λ^* is Pareto efficient if utility level of households es living in area j does not increase or that of some other households decreases when λ differs from λ^*.[31]

By inspecting all equilibrium patterns, none of them bears room for Pareto improvement by changing population distribution λ from the equilibrium distribution λ^*. That is, when the population distribution is in equilibrium λ^*, if a part of residents relocate in the other area, some households decrease (or, at least, do not increase) their utility levels.

1.7 Comparison between the center and suburb

Thus far, we have seen stable equilibrium patterns for high and low commuting costs; however, a simple list of stable equilibria does not lead us to comprehensive insights. Here, we look at these equilibria from a different point of view—the population gap between the center and suburb. As a reminder, the residential equilibrium patterns are

$$\lambda^*_{L\tau-mC} = \left(\frac{2(\tau^{\frac{1}{\beta}} - L_x)}{\tau^{\frac{1}{\beta}} + 1}, 0, 1, 1\right), \quad \lambda^*_{L\tau-mS} = \left(1, \frac{1 + 2L_x - \tau^{\frac{1}{\beta}}}{\tau^{\frac{1}{\beta}} + 1}, 0, 0\right)$$
$$\lambda^*_{H\tau-mD} = (1, 0, 1, 0), \quad \lambda^*_{H\tau-mC} = (1, 0, 1, 1), \quad \lambda^*_{H\tau-mS} = (1, 0, 0, 0).$$

1.7.1 Ethnicity characteristics

First, we examine the extent of the population gap of the minority and majority residents between the center and suburb and measure the population gap between the areas. The simplest index is the center's minority population share of the total minority population:

$$\psi_x \equiv \frac{N_x^C}{\sum_{j\in\{C,S\}} N_x^j}.$$

To calculate ψ_x for each residential pattern, we add a subscript indicating the various residential patterns. For example, $\psi_{x(L\tau-mC)}$ is the center's minority population share of the total minority population in the residential equilibrium pattern Lτ-mC, that is, $\psi_{x(L\tau-mC)} \equiv N^C_{x(L\tau-mC)} / \sum_{j\in\{C,S\}} N^j_{x(L\tau-mC)}$, where $N^j_{x(L\tau-mC)}$ is the minority population in area j under the residential pattern Lτ-mC. Calculating ψ_x for each pattern, we obtain

$$\begin{aligned}\Psi_x &\equiv \left(\psi_{x(L\tau-mC)}, \psi_{x(L\tau-mS)}, \psi_{x(H\tau-mD)}, \psi_{x(H\tau-mC)}, \psi_{x(H\tau-mS)}\right)\\ &= \left(1, 0, \frac{1}{2}, 1, 0\right).\end{aligned}$$

[31] Detailed illustrations of this discussion on Pareto efficiency are in Appendix 1.F.

The center's majority population share of the total majority population is defined as $\psi_X \equiv N_X^C / \sum_{j \in \{C,S\}} N_X^j$. Calculating ψ_X for each pattern yields

$$\begin{aligned}\Psi_X &\equiv \left(\psi_{X(L\tau-mC)}, \psi_{X(L\tau-mS)}, \psi_{X(H\tau-mD)}, \psi_{X(H\tau-mC)}, \psi_{X(H\tau-mS)}\right) \\ &= \left(\frac{\tau^{\frac{1}{\beta}} - L_x}{\tau^{\frac{1}{\beta}} + 1}, \frac{1 + L_x}{\tau^{\frac{1}{\beta}} + 1}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right).\end{aligned}$$

Figures 1.3 and 1.4 show how ψ_x and ψ_X change with commuting cost τ in the stable equilibrium patterns, respectively. Each line expresses a combined set of stable equilibria.

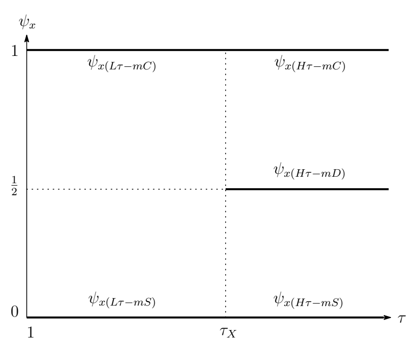

Figure 1.3: Center's share of the total minority population

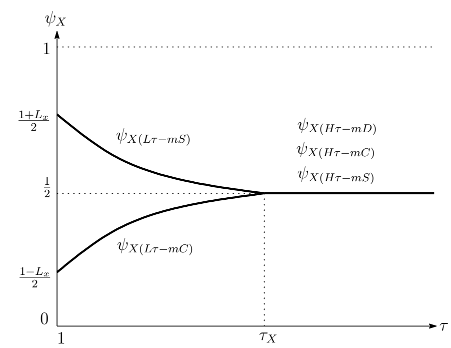

Figure 1.4: Center's share of the total majority population

From Figure 1.3, we immediately realize that there are three stable equilibrium paths for the minority's residential distributions when the commuting cost is high: the paths exhibiting clustering in the center ($\psi_x = 1$), clustering in the suburb ($\psi_x = 0$), and equal dispersion across the two areas ($\psi_x = 1/2$). Explaining the performance of the minority's equilibrium paths is more tractable after examining the majority's equilibrium paths.

Unlike the minorities' equilibrium paths, Figure 1.4 shows that the majority population is equally dispersed between both areas when the commuting cost is high (namely, $\psi_X = N_X^C / \sum_{j \in \{C,S\}} N_X^j = 1/2$). The reason underlying this single stable equilibrium path, where there is an equal distribution of the majority population, is that no majority chooses to commute when the commuting cost is sufficiently high ($\tau \geq \tau_X$). However, once the commuting cost have fallen below τ_X (the level at which the majority workers begin to commute), the majority population gap between the center and suburb widens along with lower commuting costs. Behind the equilibrium path with a smaller ψ_X, which corresponds to pattern Lτ-mC, all the minority households cluster in the center, and thus, some high-skilled majority workers choose to live in the suburb to escape the center's residential congestion, which makes ψ_X smaller as the commuting cost becomes cheaper. Similarly, behind the stable equilibrium path with a larger ψ_X, which corresponds to pattern Lτ-mS, all minority residents reside in clusters in the suburb, so that some low-skilled majority workers migrate to the center to flee the suburban congestion. Hence, the equilibrium path of ψ_X moves upward as the commuting cost decreases.

Returning to the analysis of the minority's equilibrium paths, when the commuting cost is high, no majority workers commute, so that if the minority group clusters in one area then this area becomes heavily crowded and all minority households are accommodated in the more crowded area. If minority workers choose to cluster in one area to gain ethnic utility, the minority commuters have to incur high commuting costs as well as residential congestion disutility in their residential area. If minority workers choose not to commute, they do not incur high commuting costs and residential congestion, although they cannot obtain ethnic utility. In short, both ethnic clustering and ethnic dispersion have advantages for the minority group. Therefore, minority dispersion in the two areas, as well as clustering in one area, can lead to a stable equilibrium.

On the other hand, when the commuting cost is lower than τ_X, some majority workers choose to commute. In this case, even if all minority households reside in a cluster in one area, the area where they cluster cannot become prohibitively crowded because some of the majority households have left the crowded area. Hence, a minority's dispersed residential configuration cannot be a stable equilibrium path, unlike in the case of high commuting costs. This is because the utility losses from commuting costs and residential congestion do not exceed the utility gains from ethnicity clustering when the commuting cost is low. Together, these factors make the equally distributed equilibrium path of the center's minority population share, ψ_x, disappear.

1.7.2 Skill levels

Next, we examine the population gap of high- and low-skilled residents between areas. The adopted measure is the center's high-skilled (low-skilled) population share of the total high-skilled (low-skilled) population:

$$\psi_s \equiv \frac{N_s^C}{\sum_{j\in\{C,S\}} N_s^j},$$

where $s = h$ (high-skilled) or l (low-skilled) and N_s^j is the population of consumers with skill level s living in area j. Calculating ψ_h and ψ_l for each pattern, we obtain

$$\begin{aligned}\Psi_h &\equiv \left(\psi_{h(L\tau-mC)}, \psi_{h(L\tau-mS)}, \psi_{h(H\tau-mD)}, \psi_{h(H\tau-mC)}, \psi_{h(H\tau-mS)}\right)\\ &= \left(\frac{\tau^{\frac{1}{\beta}}(2+L_x)-L_x}{(\tau^{\frac{1}{\beta}}+1)(1+L_x)}, \frac{1}{1+L_x}, 1, 1, \frac{1}{1+L_x}\right).\end{aligned}$$

$$\begin{aligned}\Psi_l &\equiv \left(\psi_{l(L\tau-mC)}, \psi_{l(L\tau-mS)}, \psi_{l(H\tau-mD)}, \psi_{l(H\tau-mC)}, \psi_{l(H\tau-mS)}\right)\\ &= \left(\frac{L_x}{1+L_x}, \frac{1+2L_x-\tau^{\frac{1}{\beta}}}{(1+L_x)(\tau^{\frac{1}{\beta}}+1)}, 0, \frac{L_x}{1+L_x}, 0\right).\end{aligned}$$

How ψ_h and ψ_l change with the commuting cost τ in the stable equilibrium patterns is shown in Figure 1.5. The high-skilled equilibrium paths (solid lines) in Figure 1.5 show

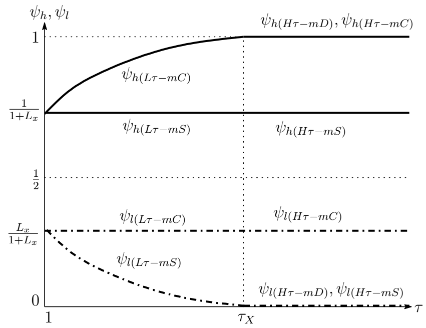

Figure 1.5: Center's share of the total high-skilled (low-skilled) population

that the number of high-skilled residents accommodated in the center is higher than that in the suburb in both of the stable equilibrium paths ($\psi_h \equiv N_h^C / \sum_{j\in\{C,S\}} N_h^j > 1/2$). Since unnecessary commuting never occurs as mentioned in Section 1.3.2 and Appendix 1.B, this is natural.[32]

Moreover, the stable equilibrium path with a larger ψ_h in Figure 1.5, which shows a wider high-skilled population gap between areas, begins to decline as commuting becomes cheaper. This less equally distributed equilibrium path compared to the equilibrium path with a smaller ψ_h originates in the residential Hτ-mD, Hτ-mC, and Lτ-mC. The upper ψ_h path expresses the residential patterns where the center is more crowded than the suburb (patterns Hτ-mD and Hτ-mC). On this path, the high-skilled majority workers suffer from residential congestion caused by a minority's clustered residence in the center. When the commuting cost is sufficiently high ($\tau \geq \tau_X$), commuting is too expensive for the high-skilled majority workers, so that they stay in the center ($\psi_h = 1$). For sufficiently low commuting costs ($\tau < \tau_X$), a slight decline in the commuting cost enables the high-skilled majority households residing in the center to migrate to the suburb. Such majority behavior under low commuting costs leads to a less unequally distributed equilibrium path.

[32]If $\psi_h \equiv N_h^C / \sum_{j\in\{C,S\}} N_h^j < 1/2$, both high- and low-skilled workers commute in vain without reducing congestion, because if no one commutes, then $\psi_h = 1/2$. $\psi_h > 1/2$ means that some low-skilled workers choose to commute when the entire majority group chooses to commute. The possibility that this situation occurs is excluded from the list of five possible residential patterns of the minority group (Section 1.3.2).

In this sense, one can assert that the lower commuting cost acts as an equalizer for the high-skilled population gap between the areas.

As for the lower ψ_h path, the center is not as crowded as the suburb. This implies that the residential congestion level has not yet reached its limit for the high-skilled majority. Hence, even if the commuting cost becomes cheaper, there is no migration of high-skilled majority households to the suburb. Thus, the stable equilibrium path remains constant, that is, $\psi_h = 1/(1 + L_x)$.

1.8 Conclusion

This chapter exploited ethnicity clustering externalities to analyze the mechanism of residential segregation under the assumption of skill-level homogeneity according to ethnicity. The presented findings showed that when the commuting cost is high, the minority group does not necessarily cluster in one area and that their dispersed residential patterns can lead to a stable equilibrium. By contrast, when the commuting cost is low, the minority group always clusters in one area. As for majority group residents, they do not commute when the commuting cost is high and thus suffer from residential congestion caused by the clustered residence of the minority group. Because of this immobility stemming from costly commuting, both the center and the suburb have the same majority population size under high commuting costs. As the commuting cost declines, the majority population gap between areas increases. In short, the majority faces a trade-off between commuting costs and residential congestion, while the minority group faces a trade-off between commuting costs, ethnic clustering, and residential congestion.

Further, because ethnic preferences exist in the minority group, minority households are more likely to migrate to one area, meaning that they always cluster when the commuting cost is low, further widening the population gap between the areas. On the contrary, majority households migrate to the less populated area to avoid the residential congestion caused by minority residential clustering, thus reducing the population gap between areas. In this sense, majority residents are forced to function as adjusters or equalizers of the population sizes in the center and suburb.

Despite this chapter's contribution, it possesses certain limitations, which are left to be addressed by future research. First, this chapter assumed that the skill/income levels are the same for both ethnicities, as it focused on how the minority's clustering preferences affect ethnic segregation. However, we do not deny the existence of an ethnic bias in skill/income levels, and accounting for this income aspect may allow us approaching the reality. In this extension of employing skill/income level difference among ethnicities, an ex post labor market difference among groups after being segregated or mixed should also be considered, because being surrounded by high- or low- skilled residents may lead to having higher or lower skill levels through positive or negative feedback loops. Also, taking consideration on majorities' preference for the proximity to residents sharing the same ethnic characteristics is one direction of the extension. As for geography, a segregation analysis in a city with multiple areas and multiple ethnic groups could be conducted. The results for such geographical conditions (one center and several suburbs, namely a mono-centric city) may be as follows. If the commuting cost is high, the center is ethnically mixed, while the suburbs are completely segregated (i.e., some suburbs are completely occupied by the majority group and the rest are accommodated by minority groups). If the commuting cost is low, all areas (including the center) are completely occupied by one ethnic group and ethnic mixing cannot thus occur. In addition, tackling the analysis in the open city framework unlike in this chapter, where we employed the closed city assumption, would be another promising extension.

Appendix 1.A Proof of Lemma 1.3.1

Lemma 1.3.1 is proved by examining the possible residential patterns of the majority. Before examining Cases 1–9, note that at this point $\frac{N^C}{N^S}$ is not yet determinate.

Case 1 — $\lambda_{Xh} \in (0,1)$ and $\lambda_{Xl} \in (0,1)$
For this residential pattern to be in equilibrium, it is necessary that $\Delta V_{Xh}(\lambda) = \Delta V_{Xl}(\lambda) = 0$. By (1.11),

$$\Delta V_{Xh}(\lambda) = \Delta V_{Xl} = 0 \quad \Leftrightarrow \quad \underbrace{\log \tau}_{(+)} = \underbrace{\beta \log\left(\frac{N^C}{N^S}\right)}_{(+)} \text{ and } \underbrace{\log \tau}_{(+)} = \underbrace{-\beta \log\left(\frac{N^C}{N^S}\right)}_{(-)}$$

But an appropriate τ satisfying this condition does not exist, since $\tau > 1$.
Case 2 — $\lambda_{Xh} \in (0,1)$ and $\lambda_{Xl} = 1$
For this residential pattern to be in equilibrium, conditions $\Delta V_{Xh}(\lambda) = 0$ and $\Delta V_{Xl}(\lambda) > 0$ are necessary. By (1.11),

$$\Delta V_{Xh}(\lambda) = 0 \text{ and } \Delta V_{Xl} > 0 \quad \Leftrightarrow \quad \underbrace{\log \tau}_{(+)} = \underbrace{\beta \log\left(\frac{N^C}{N^S}\right)}_{(+)} \text{ and } \underbrace{\log \tau}_{(+)} < \underbrace{-\beta \log\left(\frac{N^C}{N^S}\right)}_{(-)}$$

Clearly, this is a contradiction.
Case 3 — $\lambda_{Xh} \in (0,1)$ and $\lambda_{Xl} = 0$
For this residential pattern to be in equilibrium, it is necessary that $\Delta V_{Xh}(\lambda) = 0$ and $\Delta V_{Xl}(\lambda) < 0$. By (1.11),

$$\Delta V_{Xh}(\lambda) = 0 \text{ and } \Delta V_{Xl} < 0 \quad \Leftrightarrow \quad \underbrace{\log \tau}_{(+)} = \underbrace{\beta \log\left(\frac{N^C}{N^S}\right)}_{(+)} \text{ and } \underbrace{\log \tau}_{(+)} > \underbrace{-\beta \log\left(\frac{N^C}{N^S}\right)}_{(-)}$$

τ such that $\tau = \left(\frac{N^C}{N^S}\right)^\beta$ satisfies both conditions above. But note that, by the assumptions on population and skill levels, $0 < N^C < \frac{1}{2} + L_x$ and $\frac{1}{2} < N^S < 1 + L_x$ are derived and these imply $0 < \frac{N^C}{N^S} < 1 + 2L_x$. For $\tau = \left(\frac{N^C}{N^S}\right)^\beta$ to exist, it must be that $\tau < (1+2L_x)^\beta$.
Case 4 — $\lambda_{Xh} = 1$ and $\lambda_{Xl} \in (0,1)$
For this residential pattern to be in equilibrium, conditions $\Delta V_{Xh}(\lambda) > 0$ and $\Delta V_{Xl}(\lambda) = 0$ are necessary. By (1.11),

$$\Delta V_{Xh}(\lambda) > 0 \text{ and } \Delta V_{Xl} = 0 \quad \Leftrightarrow \quad \underbrace{\log \tau}_{(+)} = \underbrace{-\beta \log\left(\frac{N^C}{N^S}\right)}_{(+)} \text{ and } \underbrace{\log \tau}_{(+)} > \underbrace{\beta \log\left(\frac{N^C}{N^S}\right)}_{(-)}$$

τ such that $\tau = \left(\frac{N^C}{N^S}\right)^{-\beta}$ satisfies both conditions above. As in Case 3, however, note that $\frac{1}{2} < N^C < 1 + L_x$ and $0 < N^S < \frac{1}{2} + L_x$. This leads to $\frac{1}{1+2L_x} < \frac{N^C}{N^S}$. For $\tau = \left(\frac{N^C}{N^S}\right)^{-\beta}$ to exist, τ must satisfy $\tau < (1 + 2L_x)^{\beta}$.

Case 5 — $\lambda_{Xh} = 0$ and $\lambda_{Xl} \in (0, 1)$

For this residential pattern to be in equilibrium, conditions $\Delta V_{Xh}(\lambda) < 0$ and $\Delta V_{Xl}(\lambda) = 0$ are necessary. By (1.11),

$$\Delta V_{Xh}(\lambda) < 0 \text{ and } \Delta V_{Xl} = 0 \quad \Leftrightarrow \quad \underbrace{\log \tau}_{(+)} = \underbrace{-\beta \log\left(\frac{N^C}{N^S}\right)}_{(+)} \text{ and } \underbrace{\log \tau}_{(+)} < \underbrace{\beta \log\left(\frac{N^C}{N^S}\right)}_{(-)}$$

Obviously, this is a contradiction.

Case 6 — $\lambda_{Xh} = 1$ and $\lambda_{Xl} = 1$

For this residential pattern to be in equilibrium, it is necessary that $\Delta V_{Xh}(\lambda) > 0$ and $\Delta V_{Xl}(\lambda) > 0$. By (1.11),

$$\Delta V_{Xh}(\lambda) > 0 \text{ and } \Delta V_{Xl} > 0 \quad \Leftrightarrow \quad \underbrace{\log \tau}_{(+)} < \underbrace{-\beta \log\left(\frac{N^C}{N^S}\right)}_{(+)} \text{ and } \underbrace{\log \tau}_{(+)} > \underbrace{\beta \log\left(\frac{N^C}{N^S}\right)}_{(-)}$$

τ such that $\tau < \left(\frac{N^C}{N^S}\right)^{-\beta}$ may be a candidate for an appropriate τ, but note that in this case $N^C > 1$ and $N^S < L_x$ so that $\left(\frac{N^C}{N^S}\right)^{-\beta}$ is necessarily less than 1. This contradicts $\tau > 1$.

Case 7 — $\lambda_{Xh} = 1$ and $\lambda_{Xl} = 0$

For this residential pattern to be in equilibrium, it is necessary that $\Delta V_{Xh}(\lambda) > 0$ and $\Delta V_{Xl}(\lambda) < 0$. By (1.11),

$$\begin{aligned}\Delta V_{Xh}(\lambda) > 0 \text{ and } \Delta V_{Xl} < 0 \quad &\Leftrightarrow \quad \log \tau > \beta \log\left(\frac{N^C}{N^S}\right) \text{ and } -\log \tau < \beta \log\left(\frac{N^C}{N^S}\right) \\ &\Leftrightarrow \quad \tau^{-1} < \left(\frac{N^C}{N^S}\right)^{\beta} < \tau\end{aligned}$$

As in the previous cases, $\frac{1}{2} < N^C < \frac{1}{2} + L_x$ and $\frac{1}{2} < N^S < \frac{1}{2} + L_x$ by the restrictions of population and skill level. These imply $(1 + 2L_x)^{-\beta} < \left(\frac{N^C}{N^S}\right)^{\beta} < (1 + 2L_x)^{\beta}$. For an appropriate τ to exist under these restrictions, $\tau \geq (1 + 2L_x)^{\beta}$ must be satisfied.

Case 8 — $\lambda_{Xh} = 0$ and $\lambda_{Xl} = 1$

For this residential pattern to be in equilibrium, it is necessary that $\Delta V_{Xh}(\lambda) < 0$ and $\Delta V_{Xl}(\lambda) >$

0. But these cannot be satisfied at the same time because

$$\Delta V_{Xh}(\lambda) < 0 \text{ and } \Delta V_{Xl} > 0 \quad \Leftrightarrow \quad \underbrace{\log\tau}_{(+)} < \underbrace{\beta\log\left(\frac{N^C}{N^S}\right)}_{(+)} \text{ and } \underbrace{\log\tau}_{(+)} < \underbrace{-\beta\log\left(\frac{N^C}{N^S}\right)}_{(-)}$$

cannot hold simultaneously.
Case 9 — $\lambda_{Xh} = 0$ and $\lambda_{Xl} = 0$
For this residential pattern to be in equilibrium, conditions $\Delta V_{Xh}(\lambda) < 0$ and $\Delta V_{Xl}(\lambda) < 0$ must hold. By (1.11),

$$\Delta V_{Xh}(\lambda) < 0 \text{ and } \Delta V_{Xl} < 0 \quad \Leftrightarrow \quad \underbrace{\log\tau}_{(+)} < \underbrace{\beta\log\left(\frac{N^C}{N^S}\right)}_{(+)} \text{ and } \underbrace{\log\tau}_{(+)} > \underbrace{-\beta\log\left(\frac{N^C}{N^S}\right)}_{(-)}$$

τ such that $\tau < \left(\frac{N^C}{N^S}\right)^\beta$ satisfies both conditions. But note that the population distribution of each area in this case is $0 \leq N^C < L_x$ and $1 \leq N^S < 1 + L_x$. This yields $\left(\frac{N^C}{N^S}\right)^\beta < 1$ and this contradicts with $\tau > 1$.

Combining these cases yields Lemma 1.3.1.

Appendix 1.B Non-occurrence of the simultaneous commute of minority workers

Why do situations wherein both high- and low-skilled minority workers simultaneously commute not occur? A simple example provides us with an intuitive answer.

Consider a situation where a high-skilled minority individual h and a low-skilled minority individual l choose to commute at the same time (situation A), and both individuals h and l do not choose to commute (situation B). Note that here individual h lives in the suburb and individual l lives in the center. (The workplace of individual h is the center and she commutes, which together imply that she lives in the suburb. Further, the workplace of individual l is the suburb and he commutes, which implies that he lives in the center.) When maintaining the population distribution of individuals other than h and l ($N^C = \bar{N^C}$, $N^S = \bar{N^S}$, $N_x^C = \bar{N_x^C}$, and $N_x^S = \bar{N_x^S}$), situations A and B bring about the same population distribution, that is, $N^C|_A = \bar{N^C} + 1 - 1 = \bar{N^C} = N^C|_B$, $N^S|_A = \bar{N^S} - 1 + 1 = \bar{N^S} = N^S|_B$, $N_x^C|_A = \bar{N_x^C} + 1 - 1 = \bar{N_x^C} = N_x^C|_B$, and $N_x^S|_A = \bar{N_x^S} - 1 + 1 = \bar{N_x^S} = N_x^S|_B$, meaning that the residential congestion disutility levels and the ethnic clustering utility levels in situations A and B are the same.

However, from the viewpoint of commuting costs, these two situations—both individuals

h and l commute (situation A) or both choose not to commute (situation B)—are entirely different. In situation A, both individuals h and l incur commuting cost τ, which means that their indirect utility levels are lower by $\log \tau$. On the other hand, in situation B, they do not suffer from this indirect utility loss since none of them commute (so that they do not lose indirect utility levels by $\log \tau$). Thus, a rational individual must prefer situation B (no commuting) rather than situation A (unnecessary commuting).

Appendix 1.C Proof of excluded combinations

(M_HD/LS) (m_HC/LS), (M_HD/LS) (m_HD/LS), and (M_HD/LS) (m_HS/LS):
Since $\left(\frac{N^C}{N^S}\right)^\beta = \tau > 1$ in (M_HD/LS), we have $N^C > N^S$. Also in (M_HD/LS), $N_X^C < N_X^S$ holds. These imply that N_x^C is necessarily greater than N_x^S. By the skill-level assumption ($\kappa_x = \frac{1}{2}$), it is necessary that some low-skilled minority households reside in the center and that the high-skilled majority cluster in the center for N_x^C to be greater than N_x^S. This means only (m_HC/LC) and (m_HC/LD) can hold in (M_HD/LS).

(M_HC/LD) (m_HC/LC), (M_HC/LD) (m_HC/LD), and (M_HC/LD) (m_HC/LS):
Because of the similar discussion as outlined above, the proof is omitted.

(M_HD/LS) (m_HC/LD):
Here the population distribution is $N^C = \frac{1}{2}\left[\lambda_{Xh} + (1+\lambda_{xl})L_x\right]$, $N^S = \frac{1}{2}\left[2 - \lambda_{Xh} + (1-\lambda_{xl})L_x\right]$, $N_x^C = \frac{1}{2}(1+\lambda_{xl})L_x$ and $N_x^S = \frac{1}{2}(1-\lambda_{xl})L_x$. By (m_HC/LD), $\left(\frac{N^C}{N^S}\right)^{\beta+\gamma}\left(\frac{N_x^C}{N_x^S}\right)^{-\gamma} = \tau^{-1}$ must be satisfied. Also by (M_HD/LS), $\left(\frac{N^C}{N^S}\right)^\beta = \tau$ must be satisfied. These together yield $\frac{N^C}{N^S} = \left(\frac{N_x^C}{N_x^S}\right)^{\frac{\gamma}{2\beta+\gamma}}$, and thus,

$$\tau = \left(\frac{N_x^C}{N_x^S}\right)^{\frac{\beta\gamma}{2\beta+\gamma}} = \left(\frac{1+\lambda_{xl}}{1-\lambda_{xl}}\right)^{\frac{\beta\gamma}{2\beta+\gamma}}. \tag{1.13}$$

Since $1 < \tau < (1+2L_x)^\beta$ must hold, we get

$$1 < \left(\frac{1+\lambda_{xl}}{1-\lambda_{xl}}\right)^{\frac{\beta\gamma}{2\beta+\gamma}} < (1+2L_x)^\beta \quad \Leftrightarrow \quad 0 < \lambda_{xl} < \frac{(1+2L_x)^{\frac{2\beta+\gamma}{\gamma}} - 1}{(1+2L_x)^{\frac{2\beta+\gamma}{\gamma}} + 1}.$$

Solving (1.13) for λ_{xl}, we obtain $\lambda^*_{xl} = \frac{\tau^{(2\beta+\gamma)/\beta\gamma}-1}{\tau^{(2\beta+\gamma)/\beta\gamma}+1} \in \left(0, \frac{(1+2L_x)^{(2\beta+\gamma)/\gamma}-1}{(1+2L_x)^{(2\beta+\gamma)/\gamma}+1}\right)$.
Next, consider λ^*_{Xh}. By solving $\left(\frac{N^C}{N^S}\right)^\beta = \tau$ for λ_{Xh} and substituting λ^*_{xl} obtained above,

$$\lambda^*_{Xh} = \frac{L_x(\tau^{\frac{1}{\beta}} - 1) + 2\tau^{\frac{1}{\beta}}}{\tau^{\frac{1}{\beta}+1} - \frac{L_x(\tau^{\frac{2\beta+\gamma}{\beta\gamma}}-1)}{\tau^{\frac{2\beta+\gamma}{\beta\gamma}}+1}}.$$

When $1 < \tau < (1+2L_x)^\beta$, λ^*_{Xh} above is shown to be greater than 1, which is not defined. Thus, (M_HD/LS) (m_HC/LD) does not bear an equilibrium.

(M_HC/LD) (m_HD/LS):
Because of a similar discussion as outlined above, the proof is omitted.

Appendix 1.D Proof of stability of equilibria in (M_HD/LS) (m_HC/LC)

In this combination, the population distribution is given by $N^C = \frac{1}{2}\lambda_{Xh} + L_x$, $N^S = 1 - \frac{1}{2}\lambda_{Xh}$, $N^C_x = L_x$ and $N^S_x = 0$. By (m_HC/LC), $\left(\frac{N^C}{N^S}\right)^{-(\beta+\gamma)} \left(\frac{N^C_x}{N^S_x}\right)^\gamma > \tau$ has to be satisfied. Because the LHS of this inequality goes to infinity under this population distribution, this inequality always holds. By (M_HD/LS), we obtain

$$\tau = \left(\frac{N^C}{N^S}\right)^\beta = \left(\frac{\lambda_{Xh} + 2L_x}{2 - \lambda_{Xh}}\right)^\beta. \tag{1.14}$$

For τ to lie within the interval $(1, (1+2L_x)^\beta)$, it must be that $1 < \left(\frac{\lambda_{Xh}+2L_x}{2-\lambda_{Xh}}\right)^\beta < (1+2L_x)^\beta \Leftrightarrow 1 - L_x < \lambda_{Xh} < 1$. By this, we made sure that there exists an interior solution $\lambda^*_{Xh} \in (1-L_x, 1)$. Solving (1.14) for λ_{Xh} yields $\lambda^*_{Xh} = \frac{2(\tau^{\frac{1}{\beta}} - L_x)}{\tau^{\frac{1}{\beta}}+1}$. Next, we need to check whether $\lambda^* = (\lambda^*_{Xh}, 0, 1, 1)$ is stable. The stability notion used here is that of local stability with respect to relocation dynamics to higher utility locations, i.e., $\dot{\lambda}_{es} = \Delta V_{es}(\lambda)$, where the dot indicates time derivative. First, consider the stability of the corner equilibrium λs. Because $\dot{\lambda}_{Xl} = \Delta V_{Xl}(\lambda_{Xh}, 0, 1, 1) < 0$ for any $\lambda_{Xh} \in (1-L_x, 1)$, $\dot{\lambda}_{xh} = \Delta V_{xh}(\lambda_{Xh}, 0, 1, 1) > 0$ for any $\lambda_{Xh} \in (1-L_x, 1)$, and $\dot{\lambda}_{xl} = \Delta V_{xl}(\lambda_{Xh}, 0, 1, 1) > 0$ for any $\lambda_{Xh} \in (1-L_x, 1)$, what is left to be considered in terms of equilibrium stability is the local stability of λ_{Xh} around λ^*_{Xh}, where $\dot{\lambda}_{Xh} = \Delta V_{Xh}(\lambda^*_{Xh}, 0, 1, 1) = 0$. Because

$$\left.\frac{\partial \Delta V_{Xh}(\lambda)}{\partial \lambda_{Xh}}\right|_{\lambda=(\lambda^*_{Xh},0,1,1)} = -\beta\left(\frac{1}{\lambda^*_{Xh}} + \frac{1}{2-\lambda^*_{Xh}}\right) < 0,$$

this equilibrium is shown to be stable.

Appendix 1.E Nonexistence of stable interior equilibria of the high-skilled and low-skilled minority interior solution

Stability of high-skilled minority interior solution:
We show that no stable interior equilibria emerge in this model. Suppose there exists an interior solution λ_{xh}^*. Then, by $\Delta V_{xh}(\lambda^*) = 0$, λ_{xh}^* must satisfy

$$\tau = \left[\frac{1+\lambda_{xh}^* L_x}{1+(2-\lambda_{xh}^*)L_x}\right]^{\beta+\gamma} \left(\frac{2-\lambda_{xh}^*}{\lambda_{xh}^*}\right)^{\gamma}. \tag{1.15}$$

In addition, for λ_{xh}^* to be a stable equilibrium, it is necessary that λ_{xh}^* satisfies

$$\left.\frac{\partial \Delta V_{xh}(\lambda)}{\partial \lambda_{xh}}\right|_{\lambda_{-xh}=(1,0,0)} < 0. \tag{1.16}$$

By (1.16),

$$\left.\frac{\partial \Delta V_{xh}(\lambda)}{\partial \lambda_{xh}}\right|_{\lambda_{-xh}=(1,0,0)} < 0 \quad \Leftrightarrow \quad \lambda_{xh}^2 - 2\lambda_{xh} + \frac{\gamma(1+2L_x)}{[\beta(1+L_x)+\gamma]L_x} < 0.$$

By analyzing the stability condition (1.16), we get

$$\left.\frac{\partial \Delta V_{xh}(\lambda)}{\partial \lambda_{xh}}\right|_{\lambda_{-xh}=\lambda_{-xh}^*} < 0 \quad \text{if} \quad \frac{\gamma}{\beta} < L_x \quad \text{and} \quad \lambda_{xh} > \underline{\lambda}_{xh} \tag{1.17}$$

$$\left.\frac{\partial \Delta V_{xh}(\lambda)}{\partial \lambda_{xh}}\right|_{\lambda_{-xh}=\lambda_{-xh}^*} \geq 0 \quad \text{otherwise}, \tag{1.18}$$

where $\underline{\lambda}_{xh} \equiv 1 - \sqrt{\frac{(1+L_x)(\beta L_x-\gamma)}{L_x[\beta(1+L_x)+\gamma]}}$. By (1.17) and (1.18), for λ_{xh} such that $\lambda_{xh} > \underline{\lambda}_{xh}$, $\Delta V_{xh}(\lambda)$ decreases as λ_{xh} gets large when λ_{-xh} is fixed under the condition $\gamma/\beta < L_x$. (Note that if $\gamma/\beta > L_x$, $\Delta V_{xh}(\lambda)$ always increases with λ_{xh}, so there is no need to consider this case.) However, from Lemma 1.5.1, $\Delta V_{xh}(\lambda) \mid_{(\lambda_{xh},\lambda_{-xh})=(1,\lambda_{-xh}^*)} > 0$. These together imply that λ_{xh}^* such that $\partial \Delta V_{xh}(\lambda)/\partial \lambda_{xh} < 0$ does not exist. Thus, this model does not exhibit stable interior equilibria (see Figure 1.6).[33] Define the LHS of this inequality $g_h(\lambda_{xh})$. Note that $g_h(0) > 0$ and the axis of symmetry of $g_h(\lambda_{xh})$ is $\lambda_{xh} = 1$. If $\frac{\gamma}{\beta} < L_x$, which is the condition that $g_h(\lambda_{xh})$ has two roots, is satisfied, then the smaller root $\lambda_{xh} = \underline{\lambda}_{xh}$ belongs to the interval $(0, 1)$. Thus, $g_h(\lambda_{xh}) \gtreqless 0$ for $\lambda_{xh} \lesseqgtr \underline{\lambda}_{xh}$. This implies that, if $\frac{\gamma}{\beta} < L_x$ and $\lambda_{xh} > \underline{\lambda}_{xh}$, then $\left.\frac{\partial \Delta V_{xh}(\lambda)}{\partial \lambda_{xh}}\right|_{\lambda_{-xh}=\lambda_{-xh}^*} < 0$.[34]

[33] In fact, this outcome ($\Delta V_{xh}(\lambda_{xh}) \mid_{\lambda=(1,0,1,0)} > 0$) comes from the assumption $\kappa_X = \kappa_x = 1/2$. If this assumption is not given (for instance $\kappa_x < 1/2$), $\Delta V_{xh}(\lambda_{xh}) \mid_{\lambda=(1,0,1,0)}$ can take a negative value. This implies that the stable interior equilibrium λ_{xh}^* can emerge at a certain level of τ.

[34] Drawing the quadratic function $g_h(\lambda_{xh})$ helps us to check this proof.

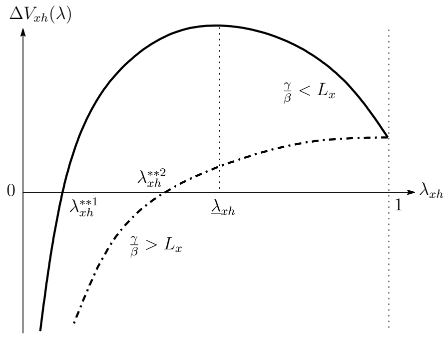

Figure 1.6: Indirect utility differentials of the high-skilled minority

Stability of low-skilled minority interior solution:
The proof is almost the same as above, so it is omitted.

Appendix 1.F Discussion on Pareto efficiency

Followings are the illustration of the procedure of how to discuss Pareto efficiency and improvement in comparison with equilibria.

- Step1: First, we focus on the dispersion of λ_{Xh} from the equilibrium λ^*_{Xh} in Lτ-mC. Because we substitute equilibrium values, $(\lambda_{Xl}, \lambda_{xh}, \lambda_{xl}) = (\lambda^*_{Xl}, \lambda^*_{xh}, \lambda^*_{xl})$, other than λ^*_{Xh} into $V^j_{es}(j \in \{C, S\}, e \in \{X, x\}, s \in \{h, l\})$, so that there are several indirect utility functions expressed as functions of λ_{Xh}.[35]
- Step 2: We investigate whether a dispersion from λ^*_{Xh} would decrease V^j_{es} or not.
- Step 3: If there is *at least one* V^j_{es} values such that would decrease when λ_{Xh} is different from λ^*_{Xh}, then we assert that λ^* is Pareto efficient, because λ_{Xh} different from λ^*_{Xh} cannot achieve Pareto improvement.
- Steps 1-3 are repeatedly applied for λ_{Xh}, λ_{Xl}, λ_{xh}, and λ_{xl}, and if we can find at least one V^j_{es} values such that would be decreased by dispersion from λ^* in *all* four cases, we conclude that Lτ-mC is Pareto efficient.

Above procedures are conducted for all equilibrium patterns, Lτ-mC, Lτ-mS, Hτ-mD, Hτ-mC, and Hτ-mS, respectively.[36]

[35]In Lτ-mC pattern, for instance, the number of utility functions of λ_{Xh} is five ($V^C_{Xh}, V^S_{Xh}, V^S_{Xl}, V^C_{xh}$, and V^S_{xl}), but not eight, because there are no households Xl living in Center, households xh living in Suburb, or households xl living in Suburb.

[36]Showing all results are space consuming and cumbersome. Details are available upon request.

Chapter 2

Which Has Stronger Impacts on Regional Segregation: Industrial Agglomeration or Ethnolinguistic Clustering?[1]

2.1 Introduction

Examples of regional segregation can be found around the world, and many such arrangements are a consequence of the regional (or local) administrative division's choice of official languages made for historical and political reasons. A typical example of regional segregation by daily language preferences is Switzerland. Table 2.1 summarizes the mother tongues of the residents in some selected Swiss Cantons (administrative divisions). Zurich is known as a German dominated area, and its official language is German.[2] French is

Table 2.1: Language distribution in Switzerland (selected cantons)

Canton	Language (%)	1970	1980	1990	2000
Zurich	German	82.9	82.9	82.5	83.4
	French	1.7	1.7	1.5	1.4
	Italian	10.2	8.0	5.8	4.0
Geneva	German	10.9	9.5	5.5	3.9
	French	65.4	64.7	70.4	75.8
	Italian	10.9	9.4	5.3	3.7
Ticino	German	10.5	11.1	9.8	8.3
	French	1.7	1.9	1.9	1.6
	Italian	85.7	83.9	82.8	83.1

[1]I would like to thank Takatoshi Tabuchi for his thoughtful comments and suggestions. I am also grateful to Tomoya Mori and Se-il Mun and seminar participants at Kyoto University. All remaining errors are on the author's responsibility. This research is partially supported by the Grants-in-Aid for Scientific Research (Research project number: 13J10130) for the Japan Society for the Promotion of Science (JSPS) Fellows by the Ministry of Education, Science and Culture in Japan.

[2]For data sources of Tables 2.1-2.5, see Appendix 2.B.

Geneva's official language, and Ticino has predominantly Italian-speaking residents. It is obvious from Table 2.1 that German is the preferred language in the Canton dominated by residents whose mother tongue is German. The same holds true for French and Italian Cantons respectively. This implies that regional segregation by ethnolinguistic characteristics occurs in Switzerland, possibly because ethnolinguistic clustering is beneficial when communicating with other residents.

Another example of regional segregation is found in Quebec, a province in Canada in which French residents (Francophones) are clustered. Table 2.2 compares the percentage of Quebec residents whose mother tongue is French with the residents of Canada as a whole. These figures illustrate the persistence of Quebec's cluster of French residents.

Table 2.2: Share of residents whose mother tongue is French in Quebec and in Canada (%)

	1931	1941	1951	1961	1971	1981	1996	2001	2006	2011
Quebec	79.7	81.6	82.5	81.2	80.7	82.4	81.3	81.2	79.4	78.7
Canada						25.6	24.3	22.8	22.0	21.6

The reasons for this persistence include Quebec's historical and political characteristics. Bill 101, adopted in 1978, made French Quebec's only official language (Laponce, 1984; St-Hilaire, 1997). In contrast, English has been chosen as the only official language in most other areas of Canada, thereby causing residents who speak only French to have difficulties when living in areas other than Quebec (and causing Quebec residents whose only mother tongue is English the same difficulty). In order to avoid struggling in daily communication and facing other language barriers, French-speaking residents have been more likely to establish themselves in Quebec, yielding the necessity of ethnolinguistic clustering. However, what would the ethnolinguistic composition of Quebec residents be if the Quebec government had not decided that the only official language in that province would be French? Put differently, what if both English and French were official languages in Quebec? History does not allow us to imagine "what if" scenarios, but similar cases to this "what if" story can be found in Catalonia in Spain, and South Tyrol in Italy.

In Catalonia, the Franco regime banned the use of Catalan in government-run institutions and during public events. However, after Franco's death in 1975, a democratic Spanish constitution was adopted in 1978, choosing both Catalan and Spanish as official languages in Catalonia (Strubell, 1996). In South Tyrol, on the other hand, where German, Italian, and Ladin were selected as local official languages, the history of triple official language adoption was quite different from that in Catalonia. South Tyrol once belonged to Austria (where the majority of the residents spoke German), but the cessation treaty after WWI granted it to Italy. During the period of Fascism under Mussolini, Italian residents were encouraged to migrate to South Tyrol, which led to the ethnolinguistic mixing of German and Italian residents in this area (Alcock, 1970). Although Catalonia and South Tyrol followed different paths to becoming multilingual areas, language barriers here are not serious impediments when residents communicate with one another, unlike Quebec. Table 2.3 shows how different the level of ethnolinguistic mixing is in the cases of Catalonia and South Tyrol, compared to Quebec. This shows that quite a large portion

Table 2.3: Examples of ethnolinguistic mixing according to different language policies

Language policy	Region	Language	Percentage of residents using as his/her mother tongue (%)
Multilingual	Catalonia (year: 2008)	Catalan	37.2
		Spanish	46.5
	South Tyrol (year: 2013)	German	69.4
		Italian	26.1
Monolingual	Quebec (year: 2011)	French	78.7
		English	7.9

of Catalan residents can understand both Catalan and Spanish, and that most German-speaking residents in South Tyrol can use Italian. In contrast, the number of Quebec residents who can use both English and French fluently is small, making communication between different linguistic groups much harder.

Another striking example of regional distribution is that of Russian residents in former Soviet Union countries. Tables 2.4 and 2.5 show ratios of the Russian population divided by the populations of residents with the dominant ethnic characteristics of a given country. Table 2.4, for example, depicts the ratios of Russian residents by region in Belarus.[3] Table

Table 2.4: Russian residents' distribution in former Soviet Union countries (selected example: Belarus)

Country	Administrative divisions	Population share (Russian/Dominant ethnicity)
Belarus	Brest	0.10
	Homyel	0.13
	Hrodna	0.16
	Mahilyow	0.13
	Minsk	0.10
	Minsk City	0.20
	Vitsyebsk	0.17
	Country	0.14

2.5 illustrates how uneven Russian resident ratios are by country. By studying Table 2.5,

[3]When constructing figures, we used "ethnicity" data rather than "language" data contained in the open data set of Alesina and Zhuravskaya (2011). As is mentioned in Hattori (2000), when asked "what is

Table 2.5: Russian population proportions in former Soviet Union countries (Summary statistics)

Country	Official language	max	min	average	st.dev	Number of administrative divisions
Belarus	Belarusian and Russian	0.20	0.10	0.14	0.04	7
Estonia	Estonian	3.48	0.01	0.33	0.88	15
Latvia	Latvian	3.46	0.04	0.45	0.63	33
Ukraine	Ukrainian	3.20	0.01	0.35	0.74	27

we notice that in Belarus, Russian residents are quite evenly distributed, relative to this distribution in other former Soviet Union countries, because of President Lukashenko's policies. Lukashenko made pro-Russian policies, and adopted Russian as well as Belarusian as one of the official languages of the country (Hattori, 2000). On the other hand, other former Soviet Union countries listed in Table 2.5 did not adopt pro-Russian policies. Those countries did not choose Russian as their official language, even though they realized the importance of their economic and political relationships with Russia. Due to the linguistic policy differences between former Soviet Union countries, Russian residents in Belarus may find it is easier to live in Belarus, and do not feel a strong need to cluster with other residents who share their Russian ethnicity. This example of the former Soviet Union captures how important the language differences are, with respect to the benefits of ethnicity clustering.

By these examples, we predict that different levels of intensity of ethnicity clustering preference bring about different regional population distribution by ethnicity. Further, regional segregation by ethnolinguistic characteristics is revealed to be persistent. In the present model, skilled workers (named "workers") are perfectly mobile and work in the manufacturing sector, while unskilled workers (named "farmers") are immobile between regions and engaged in the agricultural sector. There are two types of ethnolinguistic characteristics in the economy, so that in addition to skill levels, each individual is endowed with an ethnolinguistic characteristic. The economy consists of two regions, and immobile farmers are assumed to distribute separately by ethnicity in each region. Individuals obtain utility from proximity to the residents with the same ethnicity. This type of analysis considering proximity to various ethnic groups is done in Kanemoto (1980). Manufacturing good is differentiated, under increasing returns to scale, but agricultural

your mother tongue?" in a survey, Belarusian people tend to give answers regarding their ethnicity, rather than their mother tongue. As for other former Soviet Union countries, we could not find any evidence of this tendency to confuse the identity of the language with ethnicity, but when comparing the distribution of Russian residents among former Soviet Union countries, the utilization of figures calculated by the mother tongue ratio, instead of the ethnicity ratio, is inappropriate due to the confusing tendency of Belarusians. Actually, "ethnic group" is a slippery concept as is pointed out in Fearon (2003), so that dealing with the arbitrariness of group definitions in terms of ethnolinguistic characteristics itself is an important issue in empirical works (Baldwin and Huber, 2010; Desmet et al., 2012).

good is not, whose production is characterized by constant returns to scale. Under these settings, if complete regional segregation in terms of ethnicity arises, then this implies that industrial dispersion occurs, because both regions accommodate workers.

Our results show that even under low trade costs, the ethnicity segregation/industrial dispersion pattern is in equilibrium. This is consistent with the findings in Fujita et al. (1999, Chapter 7), Helpman (1998), Tabuchi and Thisse (2002), Ottaviano et al. (2002), Puga (1999), and Picard and Zeng (2005), all of which exhibit industrial dispersion patterns in equilibrium at low trade costs. In Fujita et al. (1999, Chapter 7), agriculture transport cost bears industrial dispersed equilibrium even under low trade costs, and taste heterogeneity (Tabuchi and Thisse, 2002) and urban cost (Ottaviano et al., 2002, Section 7) induce dispersion force. The industrial dispersion equilibrium is caused by dispersion force stemming from non-traded goods in Helpman (1998), from immobility of workers in Puga (1999), and from agricultural sector in Picard and Zeng (2005). The essence of the emergence of the industrial dispersion under low trade costs is the immobile elements in the economy. In our model, immobile farmers attached to their home region correspond to the immobile factors to bear industrial dispersion force. Indeed, our results are consistent with this—even when the trade cost is low, industrial dispersion accompanied by ethnic segregation consists of a stable equilibrium. Moreover, our model presents complete segregation equilibria for any levels of the trade costs, so that this is coherent with the examples in the real world shown above. In order to capture ethnolinguistic clustering preferences, we add the ethnolinguistic clustering term to a model in Ottaviano et al. (2002).

Intra- or inter-group social interaction is one of the important determinants of spatial segregation/integration. Segregated urban structure arises in equilibrium when agents interact more with the residents of the same group than those of the other group (Mossay and Picard, 2013). Another sorting mechanism comes from local public goods provision. Boustan (2007) points out that a racial division may reflect different preferences for public goods consumption by income level. In selecting residential locations, individuals choose their preferable bundle of public services, yielding more homogeneous composition of residents in their neighbors.[4] Ethnic segregation is also associated with individual income/education levels. Cutler et al. (2008) finds that first-generation immigrants in the United States exhibit negative selection into ethnic enclaves. Bayer et al. (2014) argues that an increase in educational attainment of black residents in American cities gives a boost to segregation, creating a sufficiently large population of the highly educated black. In addition, partly due to the ability to finance transportation costs, better educated blacks migrate to the North during the Great Migration in the United States (Vigdor, 2002).

[4]Bucovetsky and Glazer (2014) analyze a mechanism in which people sort themselves based on the preferences over the income levels of their neighbors when the cost of local public output is financed by a proportional income tax, using an adverse selection model.

The remainder of the chapter is organized as follows. The following section presents a model description, including an instantaneous equilibrium. Section 2.3 analyzes the effects of long-run regional segregation by ethnolinguistic characteristics. Section 2.4 considers which economic benefits are great enough to break the persistent ethnolinguistic clustering equilibrium. Section 2.5 deals with social welfare analysis, and proposes some linguistic policies to realize the social optimum. Section 2.6 concludes this chapter.

2.2 Model

This economy consists of two geographic regions, labeled 1 and 2, and two types of factors/sectors, named A and L. Factor L is mobile between the two regions, but factor A is not. Sector A represents "agriculture," and sector L represents "manufacturing." The immobile factor, A, is the "farmers," and the mobile factor, L, is the "workers." There are two types of ethnicity, X and x.[5] Combining the classifications in terms of factors and ethnicities, there are four types of individuals: farmers with ethnicity X, farmers with ethnicity x, workers with ethnicity X, and workers with ethnicity x, resulting in total population sizes for the whole economy A_X, A_x, L_X and L_x (including both regions 1 and 2). We assume that the innate natural abilities do not differ across ethnicities, so the farmer-worker ratio is the same for each ethnicity. We further assume that the population size does not differ by ethnicity, and that $A_X = A_x = A$ and $L_X = L_x = L$.[6] For purposes of simplicity, we normalize L to 1, i.e., $L_X = L_x = 1$. Because workers are mobile between regions, $\lambda_X \in [0, 1]$ denotes the share of workers with ethnicity X in region 1, and $\lambda_x \in [0, 1]$ denotes the share of workers with ethnicity x in region 2. In contrast, farmers are immobile between regions. As mentioned in Section 2.1, for historical, political, and/or geographical reasons, immobile agents are left in cluster in one part of a country.[7] Thus, farmers are assumed to distribute separately by ethnicity: all the farmers with ethnicity X (x, respectively) are stuck to region 1 (2, respectively). The (instantaneous) population

[5]As we saw in Section 2.1, linguistic differences in addition to (rather than) ethnic differences may play key roles for ethnolinguistic clustering preference. However, just for notational simplicity, we call ethnicities X and x, instead of ethnolinguistic characteristics X and x, when considering the model.

[6]Cases with different population sizes by ethnicity are considered in Section 2.4.

[7]German speaking populations left in the Tyrolean part of northern Italy after WWI, due to the 1919 Treaty of Versailles, are one example. Also, wrong national borders artificially drawn, which split ethnic groups into neighboring countries, was found to have occurred in many African countries, and resulted in the immobile residents' clustering near the national borders (Alesina et al., 2011). Moreover, French settlements in Canada along Saint Lawrence River gave rise to a high proportion of French residents in Quebec, some of whom are thought to be immobile.

distributions of the total population of region r, N_r $(r \in \{1,2\})$ are

$$N_1 = \sum_{e\in\{X,x\}} N_{e1} = A_X + \lambda_X L_X + (1-\lambda_x)L_x = A + 1 + \lambda_X - \lambda_x$$
$$N_2 = \sum_{e\in\{X,x\}} N_{e2} = A_x + \lambda_x L_x + (1-\lambda_X)L_X = A + 1 + \lambda_x - \lambda_X,$$

where N_{er} is the population size of residents with ethnicity e $(e \in \{X,x\})$ residing in region r.

The utility function of any particular individual consists of two parts: (i) the subutility stemming from consumption of the differentiated and homogeneous goods supplied in the market, and (ii) the ethnicity clustering preference which represents non-economic variables influencing the individual choice of location:

$$U_e(q_0; q(i), i \in [0,n]; N_e) = u(q_0; q(i), i \in [0,n]) + u^E(N_e), \tag{2.1}$$

where

$$u(q_0; q(i), i \in [0,n]) = \alpha \int_0^n q(i)di - \frac{\beta}{2}\int_0^n [q(i)^2]di - \frac{\gamma}{2}\left[\int_0^n q(i)di\right]^2 + q_0 \tag{2.2}$$

and

$$u^E(N_e) = \frac{\delta}{2}N_e. \tag{2.3}$$

First, we consider the subutility related to ethnicity preference (2.3). It is a linear addition to the subutility of goods consumption, whose formula is borrowed from Kanemoto (1980), and Bayer et al. (2014), among others.[8] N_e is the population size of ethnicity characteristic e, and the ethnicity parameter δ measures how important it is for individuals to reside with others sharing their ethnic characteristics.[9][10] Needless to say, the larger δ is, the

[8] In Kanemoto (1980) and Bayer et al. (2014), ethnic clustering utility stems from the share of population of the same ethnic group residents in the area, rather than the population size itself. It is not obvious whether residents obtain utilities from the population size or share of the same group. For analytical tractability, we employ the population size. We also ran numerical simulations with the share type subutility and obtained the similar equilibrium results. Furthermore, adding a quadratic term to the ethnic utility part ($u^E(N_e) = (\delta_1/2)N_e - (\delta_2/2)N_e^2$, where $\delta_1 > 0$ and $\delta_2 > 0$) yields quantitatively the same results. These are provided for readers upon request.

[9] It is natural that residents of the larger population group as well as those of the smaller one obtain ethnic clustering utilities when living in the same region. For example, not only candidates of minority groups but also those of majorities may have advantages in elections.

[10] Related to the interpretation of δ in the model and the data shown in Section 2.1, one comment should be noted. In the present model, δ is captured as the intensity for *within ethnic group* preference (δ_{within}), i.e., how the same ethnic group connection is important. However, in reality, there must be another ethnic preference, namely, *between ethnic group* preference ($\delta_{between}$). The extent to which different ethnic groups have well-established connection may be expressed by $\delta_{between}$. Although, in the exhibited data in Section 2.1, it might be that the exhibited data captures $\delta_{between}$ rather than δ_{within}, these two δs may be interrelated each other and cannot be separated. In the model, we employ δ based on a closer concept of δ_{within} for the analytical tractability.

more important ethnicity clustering is to that population.[11] Ethnicity subutilities for individuals X and x, when located in region 1, are described in the following equation:

$$u_1^E(N_e) = \frac{\delta}{2}N_{e1} = \begin{cases} \frac{\delta}{2}(A+\lambda_X) & \text{if} \quad e = X \\ \frac{\delta}{2}(1-\lambda_x) & \text{if} \quad e = x. \end{cases} \tag{2.4}$$

Similar equations describe the comparable situation in region 2.

Next, we look at the subutility stemming from consumption of the horizontally differentiated good, $q(i)$, and the homogeneous good, q_0, which is chosen as the numeraire in (2.2), following Ottaviano et al. (2002) (we refer Ottaviano et al. (2002) as OTT for convenience hereafter).[12] Preferences for the differentiated good and for the numeraire are identical across individuals. Non-ethnic subutility[13] (2.2) generates a system of linear demands given by a quasi-linear utility with a quadratic subutilities symmetric in all varieties $i \in [0, n]$. $q(i)$ is the quantity of variety i, and q_0 is the quantity of the numeraire. As the function of (2.2) is linear in the numeraire q_0, income effects are absent from individual consumption. As for the parameters, we assume $\alpha > 0$, $\beta > 0$, and $\gamma > 0$. α captures the intensity of preference for the product. β means that consumers have a preference for diversity. Substitutability between varieties is expressed as γ.

Each worker is endowed with one unit of labor and supplies it inelastically. In addition to her labor, she is endowed with $\bar{q}_0 > 0$ units of the numeraire. Her budget constraint is then written as

$$\int_0^n p(i)q(i)di + q_0 = w + \bar{q}_0, \tag{2.5}$$

where w is her wage and $p(i)$ is the price of variety i. We assume that the initial endowment $\bar{q}_0$ is sufficiently large for the equilibrium consumption of the numeraire to be positive.

As in previous literature, the demand function of variety i is obtained from

$$q(i) = a - (b + cn)p(i) + cP, \tag{2.6}$$

where

$$a \equiv \frac{\alpha}{\beta + \gamma n}, \quad b \equiv \frac{1}{\beta + \gamma n}, \quad c \equiv \frac{\gamma}{\beta(\beta + \gamma n)},$$

[11] It is assumed that both groups have the same intensity of ethnic clustering, i.e., $\delta_X = \delta_x = \delta$. However, the extent of ethnic clustering benefits may differ across groups, i.e., one group may benefit from ethnic segregated configuration while agents of other groups do not (Ananat, 2011). Cutler et al. (1999) argue that the extent of intensity to cluster is stronger for whites than for black residents. If we assume $\delta_X > \delta_x$, analytical results basically do not change except that the range of τ in which residents of group x achieve ethnic mixing equilibrium is wider than that of group X.

[12] We employ a simplification version of OTT in terms of coefficients of the utility function, which is adopted in Melitz and Ottaviano (2008), Picard and Tabuchi (2010), and Picard and Tabuchi (2013).

[13] As for land consumption utility, if we assume that all residents in the same region equally own and consume the same amount of land, which captures residential congestion, then the equilibrium configuration results do not change qualitatively according to our numerical simulations.

along with the price index

$$P \equiv \int_0^n p(i)di.$$

The indirect (sub)utility corresponding to the demand system (2.6) is given by

$$v(w; p(i), i \in [0, n]) = \frac{a^2 n}{2b} - a\int_0^n p(i)di + \frac{b + cn}{2}\int_0^n [p(i)]^2 di - \frac{c}{2}\left[\int_0^n p(i)di\right]^2 + w + \bar{q}_0. \tag{2.7}$$

As for production, the agricultural sector is characterized by constant returns to scale and perfect competition, where the homogeneous good is produced using factor A as the sole input. The production of one unit of the homogeneous good requires one unit input of A. Since the homogeneous good is chosen as the numeraire, and can be freely traded between regions, $w_1^A = w_2^A = 1$. In the manufacturing sector, by contrast, the differentiated good is supplied under increasing returns to scale and monopolistic competition. To produce any amount of output of the differentiated good, ϕ units of L are required. For simplicity, we assume $\phi = 1$, so that the fixed labor requirement is 1 (and the marginal one is 0). By this, the labor market clearing condition implies $n_r = \lambda_r$, where

$$\lambda_r = \begin{cases} 1 + \lambda_X - \lambda_x & \text{if} \quad r = 1 \\ 1 + \lambda_x - \lambda_X & \text{if} \quad r = 2. \end{cases}$$

Also, we assume that markets are segmented by firms, i.e., each firm has the ability to set a price specific to the market where the product is sold. Then, the profit of a firm in region r is given by

$$\Pi_r = p_{rr} q_{rr}(p_{rr}) D_r + (p_{rs} - \tau) q_{rs}(p_{rs}) D_s - w_r \quad (r \neq s), \tag{2.8}$$

where $D_r = A + \lambda_r$. p_{rs} (q_{rs}, respectively) is the price (quantity, respectively) of products produced in region r and sold in region s. τ is the trade cost (in order for each variety of the differentiated good to be traded, a positive cost of τ units of the numeraire must be incurred for each unit of the differentiated good transported from one region to the other). By profit maximization with respect to prices for the market in each region, and by the symmetry of firms in the same region, the equilibrium prices are obtained as follows:

$$\begin{aligned} p_{rr}^* &= \frac{2a + \tau c n_s}{2(2b + cn)} \quad (s \neq r) \\ &= \begin{cases} \dfrac{2a + \tau c(1 + \lambda_x - \lambda_X)}{4(b + c)} & \text{if} \quad r = 1 \\ \dfrac{2a + \tau c(1 + \lambda_X - \lambda_x)}{4(b + c)} & \text{if} \quad r = 2, \end{cases} \end{aligned}$$

$$p_{rs}^* = p_{ss}^* + \frac{\tau}{2} \quad (s \neq r).$$

Since $p^*_{12} - p^*_{11} = \tau c(\lambda_X + \lambda_x)/[2(b+c)] + \tau/2 < \tau$ and $p^*_{21} - p^*_{22} = -\tau c(\lambda_X + \lambda_x)/[2(b+c)] + \tau/2 < \tau$, no entry of transportation companies occurs. It is necessary that firms' prices net of trade costs are positive, regardless of the workers' distribution in order for these prices to be meaningful. Thus, we assume

$$\tau < \tau_{trade} \equiv \frac{a}{b+c},$$

which comes from the condition $p^*_{rs} - \tau > 0$ for all $\lambda_e \in [0,1], e \in \{X, x\}$.

Instantaneous equilibrium wage w^*_r is determined by zero-profit condition:

$$w^*_r = p^*_{rr} q^*_{rr} D_r + (p^*_{rs} - \tau) q^*_{rs} D_s. \tag{2.9}$$

Summing up the indirect subutility other than for the ethnic utility, we have

$$v_r(\lambda_X, \lambda_x) = S_r(\lambda_X, \lambda_x) + w^*_r(\lambda_X, \lambda_x) + \bar{q}_0,$$

where $S_r(\lambda_X, \lambda_x)$ is the consumer's surplus of individuals in region r (both for ethnicities X and x), which is given by

$$S_r(\lambda_X, \lambda_x) = \frac{a^2}{b} - a\left(\lambda_r p_{rr} + \lambda_s p_{sr}\right) + \frac{b+2c}{2}\left[\lambda_r (p_{rr})^2 + \lambda_s (p_{sr})^2\right] - \frac{c}{2}\left(\lambda_r p_{rr} + \lambda_s p_{sr}\right)^2 \quad (s \neq r). \tag{2.10}$$

From these, the total indirect utility of the individual with ethnicity characteristic e, when located in region r, is written as

$$V_r(\lambda_X, \lambda_x; e) = v_r(\lambda_X, \lambda_x) + u^E_r(\lambda_e). \tag{2.11}$$

Using this total indirect utility $V_r(\lambda_X, \lambda_x; e)$, we define the indirect utility differential as follows:

$$\begin{aligned}
\Delta V(\lambda_X, \lambda_x; e) &\equiv I(e)\left[V_1(\lambda_X, \lambda_x; e) - V_2(\lambda_X, \lambda_x; e)\right] \\
&= I(e)\left[\Delta v(\lambda_X, \lambda_x) + \Delta u^E(\lambda_e)\right] \\
&= I(e)\left[\Delta S(\lambda_X, \lambda_x) + \Delta w^*(\lambda_X, \lambda_x) + \Delta u^E(\lambda_e)\right],
\end{aligned} \tag{2.12}$$

where

$$I(e) = \begin{cases} 1 & \text{if} \quad e = X \\ -1 & \text{if} \quad e = x, \end{cases}$$

$$\Delta v(\lambda_X, \lambda_x) \equiv v_1(\lambda_X, \lambda_x) - v_2(\lambda_X, \lambda_x).$$

Similar definitions are applied to $\Delta S(\lambda_X, \lambda_x)$, $\Delta w^*(\lambda_X, \lambda_x)$, and $\Delta u^E(\lambda_e)$.[14] Calculating this with (2.4), (2.9), and (2.10), we obtain

$$\Delta V_e(\lambda_X, \lambda_x) = [C^* \tau(\tau^* - \tau) + \delta]\lambda_e - C^* \tau(\tau^* - \tau)\lambda_{-e} + \frac{\delta}{2}(A - 1), \tag{2.13}$$

where

$$C^* \equiv \frac{b + 2c}{4(b + c)^2}[3b^2 + 2bc(3 + A) + 2c^2(1 + A)],$$

$$\tau^* \equiv \frac{2a(3b + 4c)}{3b^2 + 2bc(3 + A) + 2c^2(1 + A)},$$

and λ_{-e} is the ethnicity other than e ($e \neq -e$). τ^* is the critical value of τ at which a stable agglomeration equilibrium emerges instead of a dispersed one in the absence of ethnic clustering preference.

2.3 Industrial agglomeration or ethnic mix

Long-run equilibrium is obtained by allowing workers to move between regions without cost (λ_e's are no longer treated as fixed). For expositional convenience, we define λ is region 1's share of the total number of firms in the economy:

$$\lambda \equiv \frac{n_1}{\sum_{r \in \{1,2\}} n_r} = \frac{1 + \lambda_X - \lambda_x}{2}.$$

This necessarily implies the two scenarios described below:[15]

- **(SD)** Segregation in terms of ethnicity/dispersion in terms of industry:
  Here, $(\lambda_X^*, \lambda_x^*) = (1, 1)$, so that $\lambda^* = 1/2$. Region 1 is populated by residents with ethnicity X, and region 2 is occupied by residents with ethnicity x. In terms of the industrial distribution, both regions accommodate workers.

- **(MA)** Mixing in terms of ethnicity/agglomeration in terms of industry:
  Here, (a) $(\lambda_X^*, \lambda_x^*) = (1, 0)$, so that $\lambda^* = 1$, or (b) $(\lambda_X^*, \lambda_x^*) = (0, 1)$, so that $\lambda^* = 0$. In (a), region 1 accommodates both X and x residents, while region 2 accommodates only x residents. All workers in the economy are in region 1, and region 2 does not attract any of them. Consequently, all firms are agglomerated in region 1, so that it is the core region. As for (b), the same explanation holds.

[14] Put differently, the total indirect utility differential for ethnicity X is defined as $V_1 - V_2$, and that for ethnicity x is as $V_2 - V_1$. In short, the total indirect utility differential for each ethnicity is defined as "the total indirect utility when located in the region where her ethnicity is dominant in terms of population size," minus "the total indirect utility when located in the region in which the other ethnicity is dominant."

[15] This model implies the third scenario: complete ethnic mixing/industrial dispersion ($(\lambda_X^*, \lambda_x^*, \lambda) = (0, 0, 1/2)$). This scenario can be realized when $A < 1$. However, under assumption (2.14), this scenario should be separately considered. Discussion of this third scenario is provided for readers upon request.

Following the procedures in OTT, we calculate the value of A where $\tau^* < \tau_{trade}$:

$$\tau^* < \tau_{trade} \Longleftrightarrow A > \frac{3b^2 + 8bc + 6c^2}{2c(b+c)} > 3 \quad (> 1). \tag{2.14}$$

By assuming (2.14), there are two candidates for long-run equilibria: SD and MA. Because (case 1) $C^*\tau(\tau^* - \tau)$ is positive when $\tau < \tau^*$, (case 2) equal to 0 when $\tau = \tau^*$, and (case 3) negative when $\tau > \tau^*$, we consider the equilibrium configurations by investigating cases 1-3, one by one following the arguments in Combes et al. (2008, Chapter 7) and Fujita et al. (1999, Chapter 14).

Case 1: $\tau < \tau^*$

Since $C^*\tau(\tau^* - \tau) > 0$,

$$\Delta V_e(\lambda_X, \lambda_x) \gtreqless 0 \Longleftrightarrow \lambda_{-e} \lesseqgtr \left[1 + \frac{\delta}{C^*\tau(\tau^* - \tau)}\right]\lambda_e + \frac{\delta(A-1)}{2C^*\tau(\tau^* - \tau)}. \tag{2.15}$$

Notice that the slope is greater than 1 (i.e., $1+\delta/[C^*\tau(\tau^*-\tau)] > 1$) and the intercept is positive (i.e., $\delta(A-1)/[2C^*\tau(\tau^* - \tau)] > 0$). Drawing the lines which satisfy $\Delta V_e(\lambda_X, \lambda_x) = 0$ for $e \in \{X, x\}$ is helpful in capturing which of the various possible long-run equilibria are more likely to emerge. Figures 2.1 and 2.2 help us consider the equilibria when $\tau < \tau^*$ in the long run. The horizontal axis (vertical axis, respectively) represents the fraction of

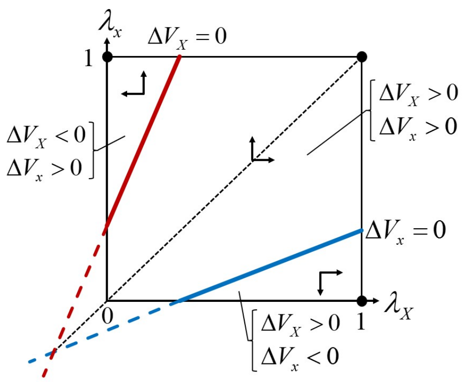

Figure 2.1: Case 1: $\delta \leq \delta^*(\tau)$

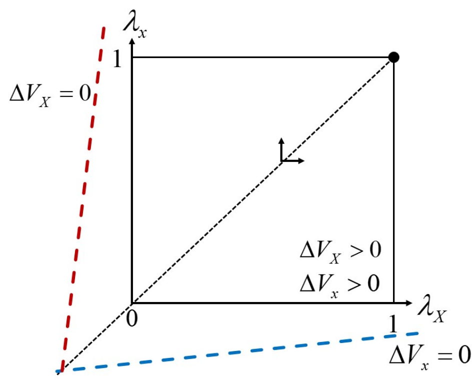

Figure 2.2: Case 1: $\delta > \delta^*(\tau)$

workers with ethnic characteristic X located in region 1, i.e., λ_X (that fraction of workers with ethnic characteristic x located in region 2, i.e., λ_x, respectively). Each point in the positive quadrant corresponds to an instantaneous equilibrium. The line named $\Delta V_X = 0$ ($\Delta V_x = 0$, respectively) depicts all combinations (λ_X, λ_x) which make the instantaneous total indirect utility differential for individuals with ethnicity X (with x, respectively) equal to 0.

First, we investigate the case with $\delta \leq 2C^*\tau(\tau^* - \tau)/(A-1) \equiv \delta^*(\tau)$. In this case, both of the two lines satisfying $\Delta V_X(\lambda_X, \lambda_x) = 0$ and $\Delta V_x(\lambda_X, \lambda_x) = 0$ appear in the positive quadrant in λ_X-λ_x plane. In the right area of line $\Delta V_X = 0$, the total indirect utility differential is such that $\Delta V_X(\lambda_X, \lambda_x) > 0$, which implies that when the fraction of workers with ethnicity X in region 1 (λ_X) is larger than line $\Delta V_X = 0$ given a certain value of λ_x (i.e., given a population distribution of workers with ethnicity x), the total indirect utility when located in region 1 is larger than that in region 2 for individuals X. This implies that more and more workers with ethnicity X relocate to region 1. Thus, the equilibrium λ_X moves away from line $\Delta V_X = 0$ in the long run. On the other hand, when λ_X is smaller than line $\Delta V_X = 0$ given a certain value of λ_x, the indirect utility in region 2 is larger than that in region 1 for individuals X, so they tend to migrate to region 2. Similar discussions can be had in the case of line $\Delta V_X = 0$. These flows of workers are depicted by a horizontal (vertical, respectively) arrow for individual X (x, respectively). In Figure 2.1, two lines $\Delta V_X = 0$ and $\Delta V_x = 0$ never have intersections, so that there is no interior long-run equilibrium. Figure 2.1 exhibits three long-run equilibria $(\lambda_X^*, \lambda_x^*) = (1, 1), (1, 0)$, and $(0, 1)$. Hence, when $\tau < \tau^*$ and $\delta \leq \delta^*(\tau)$, two types of equilibria emerge: (i) segregation/dispersion and (ii) mixing/agglomeration. The interpretation of this emergence will be mentioned after investigating the case with $\delta > \delta^*(\tau)$. For the analysis of the long-run equilibrium case with $\delta > \delta^*(\tau)$, Figure 2.2 is utilized. When $\delta > \delta^*(\tau)$, lines $\Delta V_X = 0$ and $\Delta V_x = 0$ disappear from the positive quadrant in λ_X-λ_x plane, which means that for all $\lambda_X \in [0, 1]$ and $\lambda_x \in [0, 1]$, $\Delta V_X(\lambda_X, \lambda_x) > 0$ and $\Delta V_x(\lambda_X, \lambda_x) > 0$. Then, the only long-run equilibrium is $(\lambda_X^*, \lambda_x^*) = (1, 1)$, so that when $\tau < \tau^*$ and $\delta > \delta^*(\tau)$, the equilibrium configuration is solely the pattern of segregation/dispersion.

Case 2: $\tau = \tau^*$

In this case, $C^*\tau(\tau^* - \tau) = 0$, so that by (2.13), we obtain

$$\Delta V_e(\lambda_X, \lambda_x) \gtreqless 0 \Longleftrightarrow \lambda_e \gtreqless -\frac{A-1}{2}.$$

Since $-(A-1)/2$ is negative, lines $\Delta V_X = 0$ and $\Delta V_x = 0$ never run though the positive quadrant as in Figure 2.3. Also, $\Delta V_X(\lambda_X, \lambda_x) > 0$ and $\Delta V_x(\lambda_X, \lambda_x) > 0$ always hold, in the quadrant where λ_X and λ_x are defined, so that we assert the only stable spatial equilibrium is $(\lambda_X^*, \lambda_x^*) = (1, 1)$ (i.e., SD).

Case 3: $\tau > \tau^*$

Since $C^*\tau(\tau^* - \tau) < 0$, we obtain by (2.13)

$$\Delta V_e(\lambda_X, \lambda_x) \gtreqless 0 \Longleftrightarrow \lambda_{-e} \gtreqless \left[1 + \frac{\delta}{C^*\tau(\tau^* - \tau)}\right]\lambda_e + \frac{\delta(A-1)}{2C^*\tau(\tau^* - \tau)}.$$

By investigating Figure 2.4, it is obvious that the only spatial stable equilibrium is $(\lambda_X^*, \lambda_x^*) = (1, 1)$.[16] Combining cases 2 and 3, it is asserted that when $\tau \geq \tau^*$, only

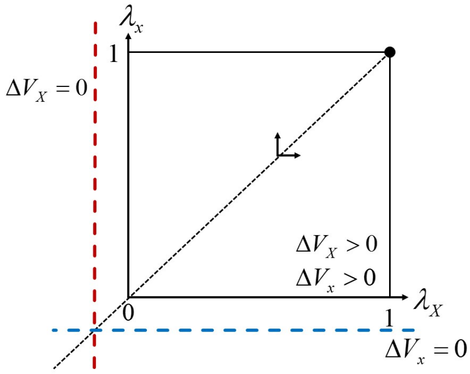

Figure 2.3: Case 2

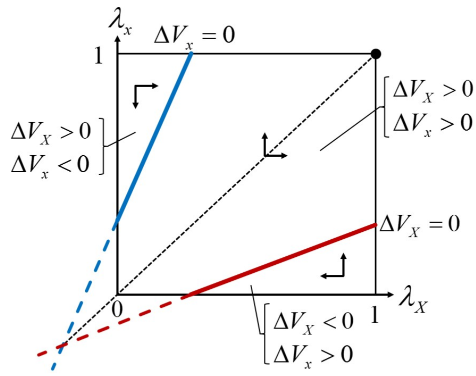

Figure 2.4: Case 3

$(\lambda_X^*, \lambda_x^*) = (1, 1)$ can emerge, i.e., the SD pattern is the only stable spatial equilibrium under high trade costs.

Combining cases 1-3, we have the following proposition.

Proposition 2.3.1. *Assume $\tau < \tau_{trade}$. Regardless of the level of trade costs, SD pattern (segregation in terms of ethnicity/dispersion in terms of industry) is a stable spatial equilibrium. In addition to SD equilibrium, when the trade cost and the ethnicity preference parameter are low ($\tau < \tau^*$ and $\delta \leq \delta^*(\tau)$), MA pattern (mixing in terms of ethnicity/agglomeration in terms of industry) can be stable spatial equilibria.*

Proposition 2.3.1 is interpreted as follows. When the trade cost is high, it is beneficial for firms to disperse the manufacturing sector into two regions because shipping their output is expensive, and they could not enjoy the benefits associated with industrial agglomeration even if they agglomerate in a single region. Industrial dispersion is in equilibrium regardless of the level of δ when $\tau \geq \tau^*$, because export losses caused by high trade costs have strong negative impacts. When $\tau < \tau^*$, the equilibrium configuration depends on the ethnicity parameter δ. If individuals have relatively strong preferences on ethnicity clustering compared with the level of τ ($\delta > \delta^*(\tau)$), gains from ethnicity clustering are larger than those from industrial agglomeration, so that ethnic segregation along with industrial dispersion should be in equilibrium. On the contrary, if δ is relatively low ($\delta \leq \delta^*(\tau)$), enjoying industrial agglomeration benefits without incurring high trade

[16] When δ is sufficiently large, the area such that $\Delta V_X(\lambda_X, \lambda_x) > 0$ and $\Delta V_x(\lambda_X, \lambda_x) < 0$ and the one such that $\Delta V_X(\lambda_X, \lambda_x) < 0$ and $\Delta V_x(\lambda_X, \lambda_x) > 0$ disappear from the quadrant where λ_X and λ_x are defined. However, this does not affect the stable spatial equilibrium in the case in which $\tau > \tau^*$. Of course, unstable equilibria disappear when δ is sufficiently large. Since we are focusing on stable spatial equilibria, we omit the detailed analyses on the cases with high δ.

costs is more important than gaining ethnicity utilities. Hence, ethnic mixing configuration with industrial agglomeration must be a stable equilibrium.

With Proposition 2.3.1, we obtain a diagram depicting a set of the long-run stable spatial equilibria in τ-δ plane under no-black-hole condition (2.14) (Figure 2.5). The parabola dividing the plane into two parts is $\delta^*(\tau)$. The area with $\delta > \delta^*(\tau)$ can realize

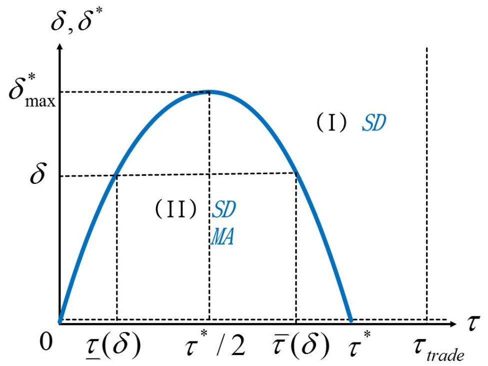

Figure 2.5: Relationship between the set of equilibria, the level of the ethnicity parameter, and trade costs

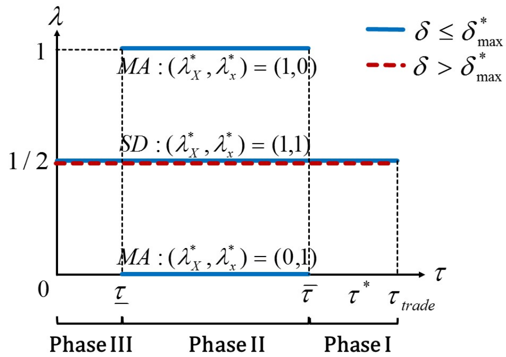

Figure 2.6: Set of stable spatial equilibria (τ changes)

spatial stable equilibrium SD, while in the area with $\delta \le \delta^*(\tau)$, SD or MA can be realized. For δ in the range of $[0, \delta^*_{\max})$, where $\delta^*_{\max} \equiv \max \delta^*(\tau) = \delta^*(\tau)|_{\tau=\tau^*/2}$, there are two values of τ, named $\overline{\tau}$ ($\underline{\tau}$, respectively) for a bigger (smaller, respectively) one, such that $\delta^*(\tau) = \delta$. Given a certain value of $\delta \le \delta^*_{\max}$ (i.e., δ is sufficiently small), if $\tau \in [\underline{\tau}, \overline{\tau}]$ (i.e., τ is intermediate), two types of equilibrium patterns can emerge (SD or MA). If $\tau \notin [\underline{\tau}, \overline{\tau}]$ (i.e., τ is sufficiently high or low), only SD is in equilibrium. For a large value of δ ($\delta > \delta^*_{\max}$), SD is the sole equilibrium configuration for any level of τ. This is depicted in τ-λ plane in Figure 2.6, which asserts the following proposition.

Proposition 2.3.2. *Assume $\tau < \tau_{trade}$, so that $A > 3$.*
When ethnicity clustering is sufficiently important ($\delta > \delta^*_{max}$), ethnic segregation/industrial dispersion (SD) pattern is the only stable equilibrium configuration.
When ethnicity clustering is less important ($\delta \le \delta^*_{max}$), the stable equilibrium configuration takes on three phases, depending on a value of τ:
Phase I ($\tau > \overline{\tau}$): When the trade cost is high, SD pattern (segregation in terms of ethnicity/dispersion in terms of industry) is the only stable equilibrium.
Phase II ($\underline{\tau} \le \tau \le \overline{\tau}$): When the trade cost is intermediate, SD or MA patterns (mixing in terms of ethnicity/agglomeration in terms of industry) can be stable equilibria.
Phase III ($\tau < \underline{\tau}$): When the trade cost is low, SD pattern is again the only stable equilibrium.

Now we briefly interpret Proposition 2.3.2 since it contains some repeated messages. In phase I, due to high trade costs, industrial agglomeration does not weigh much compared with ethnicity clustering utilities. Then, ethnic segregation/industrial dispersion is the only stable equilibrium. In phase II (under intermediate trade costs), MA as well as SD pattern is in equilibrium. This appearance of MA pattern as a stable equilibrium is because firms want to cluster in order to exploit cost and demand linkages. On the other hand, the persistence of the path along which industry is dispersed between regions is due to the existence of the ethnicity clustering preference. When trade costs are at the intermediate level, individuals have two options: (i) enjoying the benefits caused by industrial agglomeration without being anxious about a sharp decrease in exports, and (ii) enjoying the gains of ethnicity clustering and giving up industrial agglomeration benefits. This persistence of the SD path implies that both of the above options are appealing to the mobile individuals in the economy. In phase III, where the trade cost is sufficiently low, the MA path disappears, and only the SD path remains. The disappearance of the MA path under low trade costs will have a "re-dispersion flavor" as in Puga (1999), and Picard and Zeng (2005). When immobile factors are in consideration (workers' immobility in Puga (1999), and agricultural sector in Picard and Zeng (2005)), the economic activity is likely to switch from agglomeration to dispersion if trade costs become sufficiently small. Immobile factors bring about dispersion force, so that disappearance of the industrial agglomeration equilibrium is likely to occur when trade barriers and trade costs vanish.

In the present model, similar statements can be made—(immobile) farmers sharing the same ethnicity play a role of dispersion force, MA pattern cannot be a stable equilibrium when trade costs are sufficiently low. Notice, however, that what characterizes the present model is the persistence of SD equilibrium under the no-black-hole condition (i.e., $A > 3$). Because of this persistence, we say our model has a "re-dispersion flavor" under low trade costs, in that the stable dispersed equilibrium does not vanish. Why the SD path is persistent for any τ such that $\tau < \tau_{trade}$ comes from the assumption that $A > 3$, which is equivalent to $\tau < \tau_{trade}$ itself. Since $A > 3$ means that the number of farmers, who are the immobile factor in this economy, is much larger than that of workers, who are mobile between regions, dispersion force stemming from immobile elements is sufficiently strong. Because mobile individuals are strongly attracted to the region where immobile individuals who share the ethnic attributes reside. Dispersion force made by the large immobile population is so strong that at any level of $\tau < \tau_{trade}$, ethnic segregation (together with industrial dispersion) is always a stable equilibrium.

Now we link this outcome with reality. As we saw in Section 2.1, in Quebec, the ethnicity clustering preference may be stronger than that in Catalonia because of political and historical reasons. In addition, Quebec is almost dominated by French residents. On the contrary, Catalonia is not dominated by Catalans. Rather, population composition in Catalonia exhibits mixture of Spanish and Catalan residents. Given a certain level of τ

with $\tau < \tau_{trade}$, when δ is large, a spatial equilibrium pattern may exhibit segregation by ethnicity as has occurred in Quebec. On the other hand, when δ is small, the spatial equilibrium pattern may show mixing as in Catalonia. Similar arguments and interpretations can be made for South Tyrol. Quebec and Catalonia are captured as examples of different spatial equilibrium patterns, with different levels of intensity to cluster by ethnicity.

2.4 When does the segregation/dispersion equilibrium break?

In the previous section, we saw that at any level of the trade cost, stable spatial equilibrium SD (segregation by ethnicity/dispersion industrially) is persistent. That is, the SD pattern does not break at any level of τ. However, as in much new economic geography literature, a symmetry break is one of the main interests. Does a symmetry break ever occur with our framework? Put differently, with some modification, is it possible for equilibrium pattern SD to break? Actually, our model proposed in Section 2.3 can be thought of as a benchmark, in that the total population for each ethnicity is equal (i.e., the total population of residents with ethnicity X = the total population of residents with ethnicity $x = 1 + A$). This assumption regarding the exogenous population composition should be relaxed to make the model closer to the reality.

To this end, we consider the different population sizes by ethnicity, but the worker-farmer ratio is the same across ethnicities (innate natural ability is the same across ethnicities). In this section, then, we assume the exogenous population composition is as follows: $\sum_{r\in\{1,2\}} N_{Xr} = k(1 + A)$ and $\sum_{r\in\{1,2\}} N_{xr} = (1 + A)$, where $k \geq 1$, so that the population size of residents with ethnicity characteristic X in the economy is greater than that of residents with characteristic x. Also, we define the manufactured worker share $\mu \equiv n/\sum_{r\in\{1,2\}} D_r = 1/(1 + A)$. Then, $A = (1 - \mu)/\mu$. As the exogenous change needed to deal with the symmetry break (SD equilibrium pattern to disappear), we consider an increment of μ instead of a decrease in τ. An increase in the manufactured worker share μ captures the gradual change of time from the past to the future.[17]

Because it is difficult to gain some intuitions analytically from this modification, due to the complicated forms of $\Delta V_X(\lambda_X, \lambda_x)$ and $\Delta V_x(\lambda_X, \lambda_x)$, we rely on numerical analyses in this section. Now, we consider when the SD equilibrium pattern breaks as μ changes, or equivalently, we analyze at which value of μ, a pair of the total indirect utility differentials change from $(\Delta V_X(\lambda_X, \lambda_x), \Delta V_x(\lambda_X, \lambda_x)) = (+,+)$ to $(\Delta V_X(\lambda_X, \lambda_x), \Delta V_x(\lambda_X, \lambda_x)) = (+,-)$ given $(\lambda_X, \lambda_x) = (1, 1)$, where we mean that $(\Delta V_X(\lambda_X, \lambda_x), \Delta V_x(\lambda_X, \lambda_x)) = (+,+)$ indicates $\Delta V_X(\lambda_X, \lambda_x) > 0$ and $V_x(\lambda_X, \lambda_x) > 0$, and so on. In this simulation, we set $a = 3, b = 1, c = 1, k = 1.5, \tau = 0.5$, and $\delta = 1$. With this set of values, when $\mu = 0.06 < 135/2156 \approx 0.063$, $(\Delta V_X(\lambda_X, \lambda_x), \Delta V_x(\lambda_X, \lambda_x))$

[17]Detailed calculations necessary to get the total indirect utility differentials for ethnicities X and x, $\Delta V_X(\lambda_X, \lambda_x)$ and $\Delta V_x(\lambda_X, \lambda_x)$ are provided for readers upon request.

$|_{(\lambda_X,\lambda_x)=(1,1)} \approx (20.815, 0.018) = (+,+)$, so that the equilibrium $(\lambda_X^*, \lambda_x^*) = (1,1)$ is still stable. When $\mu = 0.07 > 135/2156 \approx 0.063$, $(\Delta V_X(\lambda_X, \lambda_x),\ \Delta V_x(\lambda_X, \lambda_x)) \ |_{(\lambda_X,\lambda_x)=(1,1)}$ $\approx (17.961, -0.044) = (+,-)$, so that $(\lambda_X, \lambda_x) = (1,1)$ is no longer a stable equilibrium, and instead, $(\lambda_X^*, \lambda_x^*) = (1,0)$ becomes a new stable equilibrium. From this simulation, we may assert what follows in terms of the relationship between the stability of dispersed equilibrium and industrialization. At first, the industrially dispersed equilibrium (along with ethnically segregated equilibrium) is stable. However, with the advance of industrialization (i.e., as μ increases), the industrial dispersed equilibrium disappears. Instead, industrial agglomeration (accompanied with an ethnicity mixing equilibrium) emerges. This transition from dispersion to agglomeration matches the reality. In this sense, our model does not deny what the new economic geography has built.

As one of the examples in the real world, our numerical result may be consistent with the case of Brussels. In order to enjoy economic benefits, French-speaking residents in the southern area of Belgium migrate to the northern area, where the Dutch language is dominant. The northern part of Belgium used to be poorer than the southern part of that country where French-speaking residents are dominant, but it has economically grown to be a richer area. In Brussels, the language census of 1842 showed that almost two thirds of the population of Brussels spoke Dutch, and one third French. Note that in Brussels, although both French and Dutch are the official languages, Brussels itself belongs to the northern part of Belgium, where Dutch is the only official language. By the 1970s, only about 20% of Brussels' population was Dutch speaking, and the remaining 80% spoke French (Pons-Ridler and Ridler, 1989).

2.5 Social optimum and equilibrium

To deal analytically with the social optimum, we return to the base model we proposed in Section 2.3. Because our settings assume transferable utility, it is further assumed that the social planner will choose (λ_X, λ_x) to maximize the sum of individual indirect utilities over the two regions.[18] Thus, the social welfare function to be maximized is given by

$$\begin{aligned} W(\lambda_X, \lambda_x) =& A[S_1(\lambda_X, \lambda_x) + 1 + u_1^E(\lambda_X)] + \lambda_X[S_1(\lambda_X, \lambda_x) + w_1(\lambda_X, \lambda_x) + u_1^E(\lambda_X)] \\ &+ (1-\lambda_x)[S_1(\lambda_X, \lambda_x) + w_1(\lambda_X, \lambda_x) + u_1^E(\lambda_x)] \\ &+ A[S_2(\lambda_X, \lambda_x) + 1 + u_2^E(\lambda_x)] + \lambda_x[S_2(\lambda_X, \lambda_x) + w_2(\lambda_X, \lambda_x) + u_2^E(\lambda_x)] \\ &+ (1-\lambda_X)[S_2(\lambda_X, \lambda_x) + w_2(\lambda_X, \lambda_x) + u_1^E(\lambda_X)]. \end{aligned} \tag{2.16}$$

[18] As mentioned in Polinsky and Shavell (2007, Chapter 1), if the utility function is quasi-linear as in the present context, the allocation is Pareto efficient if and only if it maximizes social welfare (total surplus). Put differently, our social welfare analysis is also based on Pareto efficiency.

Because all prices are set equal to marginal cost

$$p_{rr}^o = 0, \quad p_{rs}^o = \tau, \quad \text{and} \quad w_r^o = 0 \quad (r \neq s),$$

(2.16) becomes

$$\begin{aligned} W(\lambda_X, \lambda_x) =& [C^o \tau(\tau^o - \tau) + \delta](\lambda_X^2 + \lambda_x^2) - 2C^o \tau(\tau^o - \tau)\lambda_X \lambda_x \\ &+ \delta(A-1)(\lambda_X + \lambda_x) + \text{const}, \end{aligned} \tag{2.17}$$

where

$$C^o \equiv b + c(1+A) \quad \text{and} \quad \tau^o \equiv \frac{2a}{b + c(1+A)}.$$

By solving the optimality conditions (for details, see Appendix 2.A), we obtain the following proposition on social optimum.

Proposition 2.5.1. *When ethnicity clustering is sufficiently important ($\delta > \delta^o(\tau) \equiv C^o\tau(\tau^o - \tau)/A$), SD pattern is the social optimum: $(\lambda_X^o, \lambda_x^o) = (1,1)/\lambda^o = 1/2$. If ethnicity clustering does not weigh much ($\delta < \delta^o(\tau)$), MA pattern is the social optimum: $(\lambda_X^o, \lambda_x^o) = (1,0)/\lambda^o = 1$, $(\lambda_X^o, \lambda_x^o) = (0,1)/\lambda^o = 0$.*

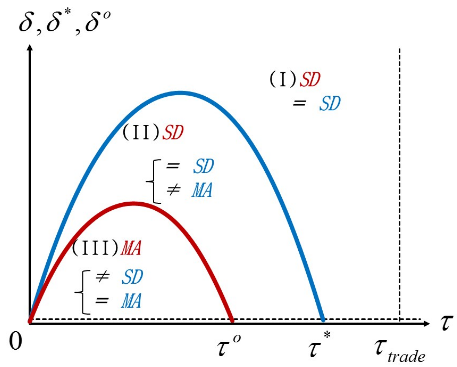

Figure 2.7: Comparison of equilibrium and social optimum

Figure 2.7 is helpful for capturing Proposition 2.5.1 intuitively. When the trade cost is high, it is socially desirable to disperse the manufacturing sector into two regions, because firms could not enjoy the benefits associated with industrial agglomeration even if they agglomerate in a single region, due to export transportation losses. It is socially optimal to disperse into two regions regardless of the level of δ, or, in other words, when $\tau > \tau^o$, export losses due to high trade costs have strong negative impacts. When $\tau < \tau^o$, the social optimal configuration depends on the ethnicity parameter δ. If individuals have relatively strong preferences on ethnicity clustering compared with the level of τ ($\delta > \delta^o(\tau)$), gains from ethnicity clustering are larger than those from industrial agglomeration, so that complete segregation by ethnicity, along with industrial dispersion, should be the social optimum. On the contrary, if δ is relatively low ($\delta < \delta^o(\tau)$), enjoying agglomeration benefits without incurring high trade costs is more important than gaining ethnicity utilities. Hence, partial mixed configuration with industrial agglomeration must be optimum. In addition, when the trade cost τ is low enough to be ignored given a certain value of δ, it is more likely that the SD configuration is the social optimum rather than MA. Because export losses stemming from incurring trade costs are relatively negligible and minute, when compared to ethnicity clustering, complete segregation may be optimum. Ethnic towns such as China-towns found in big cities around the world are thought to be examples of this situation, because within a city, goods' transportation costs are smaller than those between regions far apart from each other.

Finally, we investigate when the equilibrium configuration coincides and with the social optimum and when it does not. Figure 2.5 exhibits the relationship between the set of τ and δ, which bears the social optimum and equilibrium.[19] In area (I), social optimum and equilibrium configurations coincide, unlike in areas (II) and (III), where they do not necessarily. We investigate the reasons for these coincidence and non-coincidence for each area. For expositional convenience, we denote, for example, $(\lambda_X^o, \lambda_x^o) = \{\text{SD}\}$ when the social optimal configuration is a pattern of SD, $(\lambda_X^*, \lambda_x^*) = \{\text{SD, MA}\}$, when the equilibrium configuration is a pattern of SD or MA, and so on.

Area (I) $(\lambda_X^o, \lambda_x^o) = \{\textbf{SD}\} = (\lambda_X^*, \lambda_x^*)$**:**

In area (I), both δ and τ tend to be large compared to other areas (II) and (III). This is so because large δ brings about complete segregation, and large τ makes the firm's distribution more likely to disperse into two regions. Both parameters force the configuration to be completely segregated by ethnicity and dispersed as for industry. Thus, other configurations such as MA have no chance to be either an equilibrium or social optimum. Hence, configurations for equilibrium and social optimum become the same.

Area (II) $(\lambda_X^o, \lambda_x^o) = \{\textbf{SD}\} \neq \{\textbf{SD, MA}\} = (\lambda_X^*, \lambda_x^*)$**:**

In area (II), both δ and τ have intermediate values compared with those in area (I). In particular, a fall in τ affects this discrepancy of $(\lambda_X^o, \lambda_x^o)$ and $(\lambda_X^*, \lambda_x^*)$. As in OTT, the individual demand elasticity is much lower at the optimum (marginal cost pricing) than at the equilibrium (Nash equilibrium pricing), so that regional price indices are less sensitive to a decrease in τ. As a result, the social optimal configuration does not react to the decline of τ (i.e., $(\lambda_X^*, \lambda_x^*) = \{\text{SD}\}$ in area (I) and $(\lambda_X^*, \lambda_x^*) = \{\text{SD, MA}\}$ in area (II)). Thus,

[19]The relationship between $\delta^*(\tau)$ and $\delta^o(\tau)$ depicted in Figure 2.5 can be roughly proved as follows. Simple calculations show that $\tau^* > \tau^o$. Also, it is shown that $d\delta^*(\tau)/d\tau|_{\tau=0} > d\delta^o(\tau)/d\tau|_{\tau=0} > 0$. Combining these with that $\delta^*(\tau)$ and $\delta^o(\tau)$ run through the origin $(\tau, \delta) = (0, 0)$, $\delta^*(\tau)$ runs always above $\delta^o(\tau)$ in the area where τ and δ are defined.

in accordance with this possible reaction, only an equilibrium configuration can show the industrial agglomeration.

Area (III) $(\lambda_X^o, \lambda_x^o) = \{\mathbf{MA}\} \neq \{\mathbf{SD,\ MA}\} = (\lambda_X^*, \lambda_x^*)$**:**
In area (III), both δ and τ have lower values than those in area (II). In this area, a decline in δ affects the non-coincidence of $(\lambda_X^o, \lambda_x^o)$ and $(\lambda_X^*, \lambda_x^*)$, unlike in (II), where τ has an impact on the discrepancy. Consider the farmers left in a periphery region. If δ were high, those farmers would have lost a large amount of ethnicity clustering utility. However, now δ is low, which implies the ethnicity utility losses of immobile farmers in a periphery is so small that it may be ignored, when compared to the industrial agglomeration benefits in a core region. Social optimal configuration captures these two effects and can be more sensitive to a decrease in δ than equilibrium. As a consequence, equilibrium configuration does not react on the decline in δ (i.e., $(\lambda_X^*, \lambda_x^*) = \{\text{SD, MA}\}$ in areas (II) and (III)), while social optimal configuration reacts on a decrease in δ (and also in τ) (i.e., $(\lambda_X^o, \lambda_x^o) = \{\text{SD}\}$ in area (II) and $(\lambda_X^*, \lambda_x^*) = \{\text{MA}\}$ in area (III)).

Since we have analyzed the social optimum, we now consider its influence on policy. As Alesina and Zhuravskaya (2011) pointed out, when several residents with different ethnicity live in the same country, it is difficult for the government to reach consensus between different ethnicities, partly because they may hate each other (Glaeser, 2005). However, there is still a little room for making policy, by choosing the value of δ as a policy variable. Bilingualism may possibly be one of the means of controlling δ. As in Laponce (1984) and St-Hilaire (1997), there are two types of bilingualism: (i) the personality principle chosen in Finland and (ii) the territorial principle chosen in Switzerland and Belgium. With personality principle, the same common official languages are adopted in all areas of a country, which implies that some (or all) of the residents have to learn the two or more languages that are used in the whole country. This makes communication between different ethnolinguistic residents in that country easier, and δ gets smaller. Then, residential mixing may be promoted. However, the cost of learning another language is not negligible. Though becoming fluent in another language may be a form of human capital investment, if residents have to pay the cost of learning a language, attaining a more competitive language such as English is more effective. Indeed, in Switzerland, choosing English as a second language is more popular than choosing German or French (Pap, 1990). In the case of the territorial principle, there are several unilingual areas (only one official language for each area). With this principle, δ should be higher so that segregation is promoted. Besides, the cost of learning another language is smaller in this case, so that this principle is cost saving. Both principles have pros and cons, but for purposes of controlling regional segregation, they are useful if chosen appropriately.

2.6 Conclusions

By adding the ethnic externality into the OTT model, we investigated how regional segregation patterns are affected by industrial agglomeration and ethnic clustering. By setting some assumptions on exogenous population compositions, we showed that segregation by ethnicity is persistent, while ethnically mixed distributions appear only when trade costs are intermediate. Because we can find examples of the persistence of regional ethnicity segregation in places such as Quebec and Geneva, our results can explain the mechanisms that lead to regional segregation. We showed that ethnicity mixing can occur when the preference for ethnicity clustering is less intense, as it did in Catalonia.

With this model, a symmetry break (i.e., transition of industrial distribution from dispersion to agglomeration) does not occur, which sits a little uncomfortable with reality. However, by relaxing the composition of the exogenous population, and we found the possibility of transition from an industrial dispersed equilibrium to an industrial agglomeration equilibrium. Finally, we explained why the social optimum and equilibrium differ in light of trade costs: the social optimum is less sensitive to a change in trade costs than the equilibrium yielded by individuals' utility maximizations.

It is worth addressing the impacts of the feedback given by ethnic segregation/integration in societies. Cutler et al. (2008) argue that segregation has positive mean effects on group average human capital after corrected negative selection biases. Further, ethnic mixing may have positive or negative effect on productivity, and a net positive impact of cultural diversity has been found in the United States (Ottaviano and Peri, 2006). In addition, adopting endogenous δ would let us get closer to the reality than the exogenous δ employed in this chapter, because being surrounded by the residents of the same ethnic group may strengthen the impact of δ. Moreover, constructing a model exhibiting heterogeneity in δ among individuals should be another extension. Tackling these impacts given by inter-related dynamics of segregation and human capital accumulation/productivity improvement as well as segregation/integration or ethnic diversity presents an important issue for the future work.

Appendix 2.A Social optimal configurations

Given the welfare function (2.17), the social planner's problem is

$$\max_{\lambda_X, \lambda_x} W(\lambda_X, \lambda_x) \quad \text{s.t.} \quad 0 \le \lambda_X \le 1, 0 \le \lambda_x \le 1.$$

By KKT optimality, following candidates for the social optimal configuration satisfying the first order conditions arise under condition (2.14):

$$(\lambda_X, \lambda_x) = \begin{cases} (1,1), (1, \lambda_x^{oo}), (\lambda_X^{oo}, 1), (1,0), (0,1) & \text{if } \delta \le \delta^{oo}(\tau) \\ (1,1) & \text{otherwise,} \end{cases} \tag{2.18}$$

where $\delta^{oo}(\tau) \equiv 2C^o\tau(\tau^o - \tau)/(A-1)$, $\lambda_X^{oo} = \lambda_x^{oo} \equiv [2C^o\tau(\tau^o - \tau) - \delta(A-1)]/2[C^o\tau(\tau^o - \tau) + \delta]$. Next, we investigate which set of (λ_X, λ_x) exhibits the largest social welfare value among $(1,1)$, $(1, \lambda_x^{oo})$, $(\lambda_X^{oo}, 1)$, $(1,0)$, and $(0,1)$ in the case of $\delta \le \delta^{oo}(\tau)$.

Comparison between $(\lambda_X, \lambda_x) = (1,1)$ **and** $(\lambda_X, \lambda_x) = (1, \lambda_x^{oo})$**:**

$$W(\lambda_X, \lambda_x)|_{(\lambda_X, \lambda_x)=(1,1)} - W(\lambda_X, \lambda_x)|_{(\lambda_X, \lambda_x)=(1,\lambda_x^{oo})} = (1 - \lambda_x^{oo})\frac{\delta(A+1)}{2} \ge 0.$$

Notice that the equality holds only when $\lambda_x^{oo} = 1$ (i.e., $(\lambda_X, \lambda_x) = (1, \lambda_x^{oo}) = (1,1)$), and when $\lambda_x^{oo} < 1$,

$$W(\lambda_X, \lambda_x)|_{(\lambda_X, \lambda_x)=(1,1)} > W(\lambda_X, \lambda_x)|_{(\lambda_X, \lambda_x)=(1,\lambda_x^{oo})}.$$

Thus, $(\lambda_X, \lambda_x) = (1,1)$ is a candidate of the survivor of social optimum.

Comparison between $(\lambda_X, \lambda_x) = (1,1)$ **and** $(\lambda_X, \lambda_x) = (\lambda_X^{oo}, 1)$**:**
The similar discussion above holds, so that $(\lambda_X, \lambda_x) = (1,1)$ is a candidate of the survivor of social optimum.

Now we know that $(\lambda_X, \lambda_x) = (1,1)$ is the survivor in the comparison of $(1,1)$ and $(\lambda_X^{oo}, 1)/(1, \lambda_x^{oo})$, what is left to be investigated is the comparison between the social welfare values borne by $(\lambda_X, \lambda_x) = (1,1)$ and $(\lambda_X, \lambda_x) = (1,0)/(0,1)$.

Comparison between $(\lambda_X, \lambda_x) = (1,1)$ **and** $(\lambda_X, \lambda_x) = (1,0)$**:**

$$W(\lambda_X, \lambda_x)|_{(\lambda_X, \lambda_x)=(1,1)} - W(\lambda_X, \lambda_x)|_{(\lambda_X, \lambda_x)=(1,0)} = C^o\tau(\tau^o - \tau) + \delta A.$$

Notice that $\delta^{oo}(\tau) \equiv \frac{C^o\tau(\tau^o - \tau)}{A}$ is always below $\delta^o(\tau)$ when $\delta > 0$ (or equivalently in this case, $0 < \tau < \tau^o$). Since the case with $\delta \le \delta^{oo}(\tau)$ is on the present consideration, we need to determine which pair of (λ_X, λ_x) bears the largest social welfare value in accordance with the value of δ and $\delta^{oo}(\tau)$. If $\delta > \delta^o(\tau)$, $(\lambda_X, \lambda_x) = (1,1)$ is the social optimal population distribution (i.e., $W(\lambda_X,\lambda_x)|_{(\lambda_X,\lambda_x)=(1,1)} > W(\lambda_X,\lambda_x)|_{(\lambda_X,\lambda_x)=(1,0)}$). If $\delta < \delta^o(\tau)$, $(\lambda_X,\lambda_x) = (1,0)$ is the social optimal population distribution (i.e., $W(\lambda_X,\lambda_x)|_{(\lambda_X,\lambda_x)=(1,1)} < W(\lambda_X,\lambda_x)|_{(\lambda_X,\lambda_x)=(1,0)}$). If $\delta = \delta^o(\tau)$, $(\lambda_X,\lambda_x) = (1,1)$ and $(\lambda_X,\lambda_x) = (1,0)$ are the social optimal population distribution (i.e., $W(\lambda_X,\lambda_x)|_{(\lambda_X,\lambda_x)=(1,1)} = W(\lambda_X,\lambda_x)|_{(\lambda_X,\lambda_x)=(1,0)}$).

Comparison between $(\lambda_X,\lambda_x) = (1,1)$ and $(\lambda_X,\lambda_x) = (0,1)$:
The similar discussion above holds, so that $(\lambda_X,\lambda_x) = (1,1)$ is a candidate of the survivor of social optimum.

In the end,

$$(\lambda_X^o, \lambda_x^o) = \begin{cases} (1,1) & \text{if } \delta > \delta^o(\tau) \\ (1,1),(1,0),(0,1) & \text{if } \delta = \delta^o(\tau) \\ (1,0),(0,1) & \text{if } \delta < \delta^o(\tau), \end{cases}$$

where $\delta^o(\tau) \equiv C^o\tau(\tau^o - \tau)/A$, which yields Proposition 2.5.1.

Appendix 2.B Data sources

Table 2.6: Data sources

Table	Source
2.1	Swiss Federal Statistical Office (STAT-TAB service)
2.2	Quebec (1931-1981): Bourhis (1984, Chapter 9); Quebec (1996-2011): Census Canada; Total (1981-2011): Census Canada
2.3	Catalonia: IDESCAT (Statistical Institute of Catalonia); South Tyrol: ASTAT (Autonomous Province of South Tyrol Provincial Statistics Institute), South Tyrol in Figures 2013; Quebec: Census Canada 2011. We used the category "identify language" in Catalan data. In South Tyrol data we used "Total number of valid declarations," which equals "the number of declarations of which language group belonged to" plus "the number of declarations of which language group affiliated to." As for Quebec data, we used the category "mother tongue."
2.4 & 2.5	Open data set of Alesina and Zhuravskaya (2011). The year in which the data was obtained in the original data set of Alesina and Zhuravskaya (2011) for each country was 2001 for Belarus, 1994 for Estonia, 1996 for Latvia, and 1998 for Ukraine.

Chapter 3

Linguistic Distance and Economic Development: Costs of Accessing Domestic and International Centers[1]

3.1 Introduction

In the recent years, there has been increasing research interest on the impacts of ethno-linguistic diversity on economic and political activities. Diversity and heterogeneity appear to be the essential factors influencing economic prosperity, quality of governance, public goods provision, and the possibility of civil wars through various channels. With the realization that ethno-linguistic heterogeneity is important when considering various social phenomena, vast literature on diversity has been attracting attention both empirically and theoretically.[2]

Easterly and Levine (1997), Alesina et al. (2003), and Alesina and La Ferrara (2005) investigate the impacts of ethnolinguistic diversity on economic development. Easterly and Levine (1997) show that ethnic diversity negatively affects a country's income level, particularly explaining the low incomes in African nations. According to Alesina et al.'s (2003) cross country comparison, both ethnically and linguistically diverse structures decrease

[1]I would like to thank Takatoshi Tabuchi for his thoughtful comments and suggestions. I also thank Ryo Ito and Marcus Berliant for their comments which have improved this chapter. I am grateful to the seminar participants at the Urban Economics Workshop at the University of Tokyo and the ARSC annual meeting at Ryukyu University for their valuable comments. All remaining errors are the author's responsibility. This study is supported by the Grants-in-Aid for Scientific Research (Research project number: 13J10130) for the Japan Society for the Promotion of Science (JSPS) Fellows by the Ministry of Education, Science and Culture in Japan.

[2]Impacts of ethnolinguistically heterogeneous society on government institutional quality and public goods provision are studied in La Porta et al. (1999), Alesina et al. (1999), and Baldwin and Huber (2010). In particular, La Porta et al. (1999) find that more ethnolinguistically heterogeneous countries tend to have worse government quality. Alesina et al. (1999) certify a negative relationship between ethnolinguistic fragmentation and public good provision. Fearon and Laitin (2003) examine the outbreak of civil wars and ethnic diversity, but they conclude that ethnic diversity alone does not sufficiently explain the outbreak of civil wars.

growth levels, which implies that heterogeneous composition of ethnolinguistic characteristics negatively affect economic development. This at first seems to contradict what can be seen in the cases of highly advanced countries. For example, positive effects of diversity on productivity are found in U.S. cities (Ottaviano and Peri, 2006) and in European regions (Bellini et al., 2013).[3] Further, Alesina et al. (2013) argue that birthplace diversity exhibits a non-linear relationship with long-run income: richer countries have more merits from birthplace diversity than poorer ones do. In other words, cultural diversity may be beneficial only for well-developed nations; for economically backward countries, the effect of diversity might be negative. This matches with the findings of Alesina and La Ferrara (2005) that diversities in ethnicity and language basically have a negative impact on economic success but may have a positive impact for richer countries. It remains to be answered why ethnic diversity is sometimes harmful and not advantageous for improving the economic status and why it results in costs for joining heterogeneous groups. If this between-group harmonizing cost is more crucial than the benefits of diversity, then the effect of cultural diversity appear to be negative.

While empirical research on diversity and productivity is not limited to the country or regional aggregated levels, studies on the relationship between diversity and production improvement have been conducted in micro contexts such as firm/plant-level analysis. For example, using the Japanese patent database, Inoue et al. (2015) investigate how diversity in knowledge stocks between establishments affects the quality of patents. They show that high-quality knowledge creation is achieved efficiently at the moderate diversity level, concluding that not only the benefits of diversity but also a certain level of common knowledge, thus a smaller cost of heterogeneity, is important for high-quality knowledge creation. By using U.K. patent data, Nathan (2015) examines whether ethnicity influences innovation via production complementarities and finds that diversity of inventor communities promotes individual patenting. Moreover, Østergaard et al. (2011) find a positive effect of diversity in education and gender on the firms' likelihood to innovate, while a negative one in the case of diversity in age in Denmark. Further, Marino et al. (2012) find that at the firm level, diversities in educational and cultural backgrounds foster formation of new firms, while those in demographic features such as age and gender discourage entrepreneurship in Denmark. They interpret the negative effect of demographic diversity as the outcome of the communication barrier that hinders productivity.

Diversity and production are also theoretically jointly analyzed. Berliant and Fujita (2012) investigate cultural diversity and productivity in terms of creating new knowledge and find that more culturally heterogeneous societies enjoy higher productivity.[4] Lazear (1999a,b) exhibit models with a trade-off between the benefits from multi-cultural diver-

[3]On how diversity plays important roles in urban context, see Jacobs (1961, Part 2).

[4]In Berliant and Fujita (2012), the notion of culture is slightly different from that of our context, where they capture it as a dynamic and endogenous one rather than an exogenously given one such as ethnicities.

sity borne by production complementarity and the costs of communication barriers, which hinders knowledge transfers. This trade-off between diversity benefits and coordination costs necessarily implies that costs to associate different groups are not negligible. At the country level, maintaining coordination between different ethnolinguistic groups bears some cost, which induces negative effect on economic performance.

When analyzing economic prosperity and ethnolinguistic diversity, costs of jointing various groups such as between-group communication costs always accompany the benefits of diversity. The negative effect caused by the existence of different groups, which is hidden behind in the shadow of positive impact by construction, exceeds the positive effect captured by the diversity index. While the existing literature investigate the positive effects of ethnolinguistic diversity, it is important to study the negative effects of the same, such as the costs involved (e.g., communication costs). To capture the communication cost when a society consists of different linguistic groups, we use linguistic distance between languages. This replacement of linguistic diversity index with linguistic distance is one of the features of this chapter, and we conduct analysis of the impact of linguistic distance on cross-country difference of economic success.

So far, we have not explicitly expressed that the between-language communication cost captured by linguistic distance, in terms of cross-national comparison of economic status, is a concept "within" the nation. At the first glance, this may seem too obvious, since comparison is at the country level. What is to be inspected, induced immediately by the discussions above, would be the intra-national communication difficulty that affects the economic development of the countries. However, explicitly labeling the domestic linguistic distance enables us to identify another important communication cost—international linguistic distance as the global communication barrier. Cultural barriers or differences in use of language are considered in international migration literature (Belot and Ederveen, 2012; Adsera and Pytlikova, 2012). Communication difficulty between different groups is imperative in not only the nationwide but also the global context. For instance, in the Meiji period, the era of Japanese industrialization and modernization, Arinori Mori, who was a diplomat famous for his contribution to founding Japan's modern educational system, emphasized the importance for global communication in order to establish connection with the Western countries. Particularly, he argued the necessity for fluent communication via English (Mori, 1873). Taking this stance, we investigate the impacts of domestic and international linguistic distances on the cross country income difference. We construct two types of linguistic distance indexes—domestic and international linguistic distances. While the domestic linguistic distance index captures within-country communication difficulty, the international linguistic distance, which is the distance between an individual's mother tongue and English (one of the world's most widely used languages). International linguistic distance index is reflective of costs incurred when global communication takes place.

The main findings of this chapter are as follows. The domestic linguistic distance has a negative impact on the economic development of poor countries. Because many African countries have larger domestic linguistic distance, Africa's development tragedy can be partly explained by their problems in nationwide communication. On the other hand, larger domestic linguistic distance positively affects GDP per capita for rich countries, which is due to the diversity benefits to production. In addition, if the sample is the full set of countries instead of the poor and rich subsamples, the cost from within-country linguistic distance outweighs the benefits for improving production borne by heterogeneity, which leads to the negative impact of domestic linguistic distance on national income. Our results are distinct from those in previous literature, in which the negative feature of between-group communication, which is overshadowed by the heterogeneity of the society, is indirectly considered via ethnic diversity indexes; this chapter's focus is on the cost of communication between different language groups rather than the benefits from heterogeneity of society. Furthermore, only rich countries enjoy positive impacts on GDP per capita when international linguistic distance is smaller. This implies that rich countries can improve their economic prosperity if the accessibility to global communication via English is improved.

The remainder of the chapter is organized as follows. Section 3.2 introduces notions of domestic and international linguistic centers and distances, and provides indexes for the linguistic distances. Section 3.3 details the model specification for empirical analysis. Section 3.4 reports the estimation results. Section 3.5 conducts robustness checks, and Section 3.6 concludes this chapter.

3.2 Indexes of linguistic distance

3.2.1 Definition of linguistic distance

Our linguistic distance index data cover a wide cross-section of countries, where indexes for each country are based on weighted averages of linguistic distances. For the first step of index construction, we calculate the linguistic distance for each pair of living languages listed in the 16-th edition of Ethnologue database (Lewis, 2009). In this database, all the world's known living languages are listed[5], and categorized based on the similarities in their linguistic characteristics, and thus, dendrograms (linguistic family tree diagrams) can

[5] Laitin (2000) points out Ethnologue's features and problems, such as the linguistic classification is too detailed especially for non-Christianity nations. However, in a large number of empirical researches focusing on ethnolinguistic diversity and distance, Ethnologue data are utilized to give the quantitative aspects. For example, in the calculation of linguistic distance for examining relationships between international migration and language distance in Adsera and Pytlikova (2012), linguistic data are constructed using Ethnologue.

be drawn using this.[6] Figure 3.1 exhibits a virtual linguistic dendrogram, where languages $A1$ and $A2$ belong to the same linguistic family A, and languages $B1$–$B4$ are categorized into linguistic family B. Among languages $B1$–$B4$, languages $B2$–$B4$ are more similar

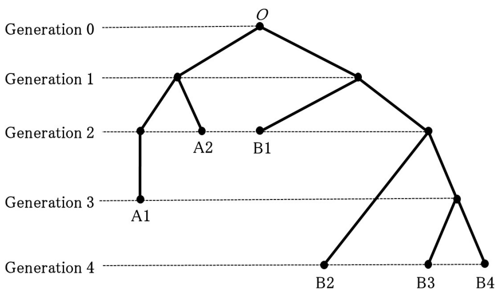

Figure 3.1: Linguistic family tree

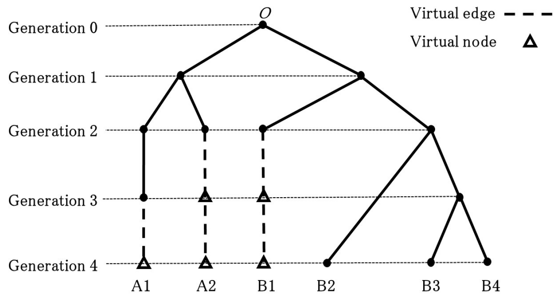

Figure 3.2: Linguistic family tree with virtual edges and nodes

than language $B1$ because the former is of the same sub-category while the latter is not.

As depicted in this example, if two arbitrarily chosen languages belong to the same linguistic family, these are thought to be similar, and thus, they must exhibit shorter linguistic distance. Before introducing the concept of "linguistic distance," consider a notion of "linguistic similarity." In order to give a quantitative value to an abstract notion of linguistic similarity, linguistic dendrograms are utilized. The number of shared edges between languages i and j on a linguistic family tree are denoted by $e(i, j)$. If $e(i, j)$ is large, it implies that languages i and j are categorized into a meta-group, such that they are closer linguistically. $g(i)$ is a generation in which language i belongs, which is used to convert $e(i, j)$ into proportions of cognates between i and j (normalization into the interval $[0, 1]$). For example, language $A1$ in Figure 3.1 is labeled generation 3 ($g(A1) = 3$). As seen in Figure 3.2, languages $A1$–$B4$ have different generation lengths. When quantifying linguistic similarity, we rely on an approach employed in Fearon (2003) and Desmet et al. (2009). By defining $g_{\max}$ as the maximum number of $g(i)$ for all existing languages i in the world, we formulate linguistic similarity as follows:

$$\text{similarity}(i, j) = \frac{e(i, j)}{g_{\max}}. \tag{3.1}$$

In the example of Figure 3.1, $\text{similarity}(A1, A2) = e(A1, A2)/g_{\max} = 1/4$. Figure 3.2 illustrates this approach of the addition of virtual edges and nodes, that is, the extension of the terminal nodes.[7]

Now, we define linguistic distance between languages i and j, $\tau(i, j)$. Higher similarities show shorter linguistic distances, and thus, if $\text{similarity}(i, j)$ increases, $\tau(i, j)$ decreases.

[6] For the research relying on language dendrograms, see Desmet et al. (2012), who compare economic and political effects given by linguistic diversity aggregated at different levels of linguistic cleavages.

[7] Another approach to normalize linguistic similarity is proposed in Desmet et al. (2005).

Further, we assume that $\tau(i, j) = \tau(j, i)$ for all languages i and j.[8] $\tau(i, j)$ is a standardized metric (i.e., $\tau(i, j) \in [0, 1]$), and $\tau(i, i) = 0$ for all i.[9] Under these assumptions on metrics, the linguistic distance between languages i and j, $\tau(i, j)$, is defined as follows:

$$\tau(i,j) = 1 - [\text{similarity}(i,j)]^{\delta} = 1 - \left[\frac{e(i,j)}{g_{\max}}\right]^{\delta} \qquad \text{for all } \; i, j (i \neq j), \tag{3.2}$$

where $\delta \in (0, 1)$ is a parameter determining how fast the linguistic distance declines as the number of shared edges increases. More intuitively, as mentioned in Desmet et al. (2009), δ captures how much more distant two languages from different linguistic families are, compared with languages that belong to the same family.[10][11]

3.2.2 Indexes of linguistic distance

Our linguistic distance indexes are based on the population weighted averages of linguistic distances. Consider country i with a population of $N(i)$ individuals, who are partitioned into $K(i)$ distinct language groups according to their language use. $N_j(i)$ is a population of language group j in country i. We assume that each individual belongs to only one language group, so that no individual is assumed to be a complete multilingual. Then,

$$N(i) = \sum_{j=1}^{K(i)} N_j(i). \tag{3.3}$$

The population share of group j in country i, $s_j(i)$, is defined as

$$s_j(i) = \frac{N_j(i)}{N(i)}, \tag{3.4}$$

[8] This symmetric metric assumption may not apply to the reality. For example, some languages are much more difficult to be acquired than others are because their linguistic properties display linguistic complexity, or they have a smaller portion of foreign-origin vocabularies.

[9] Triangle inequality $\tau(i, j) \leq \tau(i, k) + \tau(k, j)$ for all i, j, and k can be assumed, but this assumption is unnecessary when constructing a linguistic distance matrix.

[10] Consider two pairs of languages (Spanish, English) and (Japanese, English): Spanish and English both belong to the Indo-European language family, while Japanese and English are drawn from completely different language families (i.e., there are no shared edges between them). When $\delta = 0.5$ as in Fearon (2003), τ(Spanish, English)≈ 0.74 and τ(Japanese, English)$= 1$. On the other hand, when $\delta = 0.05$ as in Desmet et al. (2009), τ(Spanish, English)≈ 0.13 and τ(Japanese, English)$= 1$. Lower values increase relative distance between languages that do not have edges in common. In this numerical example, when $\delta = 0.05$, the Japanese-English pair is 6.7 times more distant than the Spanish-English pair. When $\delta = 0.5$, the Japanese-English pair is 0.3 times more distant than the Spanish-English pair.

[11] In the present analysis, δ is treated as exogenous, i.e., the extent to which two languages from different linguistic families are more distant in comparison of the members of the same language group is treated as the exogenously given linguistic characteristics.

so that

$$\sum_{j=1}^{K(i)} s_j(i) = 1 \tag{3.5}$$

for all countries i.

First, we define the domestic linguistic distance index $DLD(i)$ for each country i, which is interpreted as the cost incurred when accessing the "linguistic center" of country i, and by accessing it, the residents can communicate with each other. In other words, without acquiring the national widely spoken language(s), communication among the residents of a nation is impossible. Language $c(i)$ is the central language of country i, and $\tau_{j,c(i)}$ is the linguistic distance between languages j and $c(i)$. $DLD(i)$ is the weighted average of linguistic distances to the domestic central language:

$$DLD(i) = \sum_{j=1}^{K(i)} s_j(i)\tau_{j,c(i)}. \tag{3.6}$$

Next, we define international linguistic distance index $ILD(i)$. Here, English is adopted as the central language of the world, named language C. (By acquiring English, global communication is allowed.)[12] As for the choice of ILD, there are two possibilities, termed ILD_{PC} and ILD_{CC}:

$$ILD_{PC}(i) = \sum_{j=1}^{K(i)} s_j(i)\tau_{j,C} \tag{3.7}$$

$$ILD_{CC}(i) = \sum_{j=1}^{K(i)} s_j(i)\tau_{c(i),C} = \tau_{c(i),C}. \tag{3.8}$$

$ILD_{PC}(i)$, where the subscript PC denotes the linguistic distance from each person to the international center, refers to the linguistic distance cost that each person in group j in country i incurs to access the international center besides the cost to access the domestic center in country i. $ILD_{CC}(i)$ is the linguistic distance between the domestic and international central languages, where the subscript CC stands for the linguistic distance from the domestic center to the international center.

3.2.3 Linguistic center

Now, we have define the linguistic centers, especially the domestic ones. Many countries have several respective official languages authorized by governments. In some cases, both official and national languages, the latter often being a symbolic national language that is not used colloquially, are adopted. Thus, considering only the official language(s) to

[12] In Adsera and Pytlikova (2012), the impact of English as a representative of a widely spoken language on international migration is considered. It may be natural to consider English as a central language of international communication.

determine the domestic central language may be inappropriate.

Here, we introduce the notion of "language status" proposed in Ethnologue 17th edition (Lewis et al., 2014) on the criteria of EGIDS (Expand Graded Intergenerational Description Scale), which ranges from status 0 to 10 according to importance or usage of the languages. For example, languages labeled status 0 (international) are the languages widely used between nations in trade, knowledge exchange, and international policy. Status 1 languages (national languages) are used in education, work, mass media, and government at the national level. The label status 2 (provincial) is put on the language used for education, work, mass media, and government within major administrative subdivisions of a nation. Status 10 (extinct) is ascribed to languages with which no one associates an ethnic identity even symbolically.[13]

As for the international linguistic center, English is considered the most appropriate one, because among languages labeled status 0 (i.e., Arabic, Chinese, English, French, Russian, and Spanish), it has the biggest L2 user population.[14] For the domestic linguistic center(s), national language(s) (status 1 languages) are chosen. Then, some modifications to linguistic distance indexes are necessary. Consider, first, revising DLD index. If there is only one status 1 language in country i, then it is $c(i)$ such that DLD index is the same as proposed in (3.6). If country i has several status 1 languages, DLD index in (3.6) is modified as

$$DLD(i) = \sum_{c(i) \in C(i)} \gamma_{c(i)} \sum_{j=1}^{K(i)} s_j(i) \tau_{j,c(i)}, \tag{3.9}$$

where $C(i)$ is a set of status 1 languages in country i, and

$$\gamma_{c(i)} = \frac{N_{c(i)}}{\sum_{l \in C(i)} N_l}. \tag{3.10}$$

In short, DLD index for a country with multiple national languages is the weighted average of the weighted averages of the domestic linguistic distances to status 1 languages. Further, ILD_{CC} indexes are modified in the same fashion, that is,

$$ILD_{CC}(i) = \sum_{c(i) \in C(i)} \gamma_{c(i)} \sum_{j=1}^{K(i)} s_j(i) \tau_{c(i),C}. \tag{3.11}$$

Because no modification is needed to be applied for ILD_{PC}, in our empirical analysis, the revised DLD and ILD_{CC} given in (3.9) and (3.11), as well as the original ILD_{PC} given in (3.7), are exploited.[15]

[13] For more details of language status definition, see www.ethnologue.com/about/language-status.

[14] While L1 user population uses the given language as the first one, that is, the mother tongue, L2 user population uses that language as a second one.

[15] In most of the countries with several national languages, there is a positive number of populations whose mother tongue belongs to the set of national languages in those countries. In some countries

3.3 Empirical analysis

3.3.1 Model specification

Our main interest lies in empirically investigating the relationship between domestic and international distances and economic success; to measure the economic prosperity of a country, its income (GDP per capita) at real PPP from the Penn World Tables 8.0 (Feenstra et al., 2013b) is adopted, and our empirical model is specified as follows:

$$\ln \mathrm{GDP/capita}_i = \alpha + \beta_D DLD_i + \beta_I ILD_i + x_{\mathrm{control},i}\beta_{\mathrm{control}} + \epsilon_i, \tag{3.12}$$

where ILD is ILD_{PC} or ILD_{CC}. In the PWT 8.0, expenditure- or output-based GDPs are available, and our focus is on the impact of linguistic distances on the economic productivity rather than on the consumption levels. The output-based measure is utilized by following a user guide of PWT 8.0 (Feenstra et al., 2013a).[16] Hereafter, we notate country indicator i as a subscript. In addition, because DLD (ILD, respectively) depends on a given value of linguistic distance parameter δ_D (δ_I, respectively), we denote $DLD(\delta_D)$ ($ILD(\delta_I)$, respectively).

A vector of control variables ($x_{\mathrm{control},i}$) contains variables in terms of education, market size, trade openness, geographic characteristics, institutional qualities, and others. The choice of controls basically follows the perspectives of those in Alesina et al. (2013). For education covariates, years of schooling (Barro and Lee, 2013) is chosen. Market sizes are controlled by population sizes from the Penn World Tables 8.0 (Feenstra et al., 2013b), and land area and a landlocked dummy from CEPII (Head et al., 2010). Further, we control for the standard trade volume measure, trade openness[17] (i.e., exports + imports share of the GDP in real PPP prices) from PWT 8.0, and for the vector of geographical determinants[18] (i.e., absolute latitude of the capital city, ratio of population within 100 km of ice-free coast to total population, average temperature, and average precipitation). For institutional quality, revised combined polity2 score from Polity IV database (Marshall and Jaggers, 2012) is chosen, which measures the extent at which political participation is unrestricted, open, and fully competitive; executive recruitment is elective; and constraints on the chief executive are substantial. From Freedom House database, we also

with multiple national languages, however, one (or more than one) national language(s) has no speaker population whose mother tongue is that language. For example, Ethnologue reports that Cameroon, which is a double national language country of English and French, has no mother tongue population of those two national languages. For such countries, special treatments are needed. Notes for constructing linguistic distance indexes for those ones are provided upon request.

[16] TFP data are also available in the PWT 8.0, but for the purpose of increasing samples, we choose output-based GDPs.

[17] Impacts of trade on a country's income or growth are discussed in Frankel and Romer (1999), Rodriguez and Rodrik (2001), and Yanikkaya (2003).

[18] As for the impacts of geographical determinants of economic development, see Sachs (2003) and Putterman and Weil (2010).

Table 3.1: High correlation between $DLD(\delta_D)$ indexes and Sub-Saharan African countries

Continent	$\delta_D = 0.1$	$\delta_D = 0.2$	$\delta_D = 0.3$	$\delta_D = 0.4$	$\delta_D = 0.5$
Sub-Saharan Africa	0.766	0.757	0.744	0.730	0.716
Latin America	-0.192	-0.219	-0.240	-0.258	-0.272
South-East Asia	-0.128	-0.118	-0.109	-0.100	-0.092
	$\delta_D = 0.6$	$\delta_D = 0.7$	$\delta_D = 0.8$	$\delta_D = 0.9$	
Sub-Saharan Africa	0.702	0.689	0.677	0.667	
Latin America	-0.283	-0.292	-0.300	-0.306	
South-East Asia	-0.085	-0.080	-0.075	-0.071	

use the political rights index, which considers electoral process, political pluralism and participation, functioning of government, and discriminatory political rights, and civil liberalities index, which is defined in view of freedom of expression and belief, associational and organizational rights, rule of law, and personal autonomy and individual rights.[19] Further, following La Porta et al. (1999), legal origin dummy variables are adopted.[20] For other controls, mean and standard deviation of agricultural suitability (Michalopoulos, 2012) are used.[21] For the list of data sources, see Appendix 3.A.

Before running the regressions, some comments should be noted. Occasionally, continent dummy variables are adopted as covariates explaining the extent of economic development. Following the literature, we wanted to include the continent dummies. However, as shown in Table 3.1, DLD index, which is one of the primary interest in our context, has high correlation with Sub-Saharan continent fixed effect in the full range of the values of parameter δ_D, and thus, we are forced to drop continent dummies from our regressions. This necessarily is a caveat under the empirical framework of OLS regression model, because, when explaining why most Sub-Saharan African countries are still be-

[19]Hall and Jones (1999), Acemoglu et al. (2001), Glaeser et al. (2004), and Rodrik et al. (2004) consider how institutional qualities affect economic growth or income difference.

[20]In La Porta et al. (1999), the distinctions among the French, German, and Scandinavian legal origin families, all of which consist of the civil law group, are relatively subtle, while those among civil, common (British legal origin family), and socialist law groups are not. In our empirical estimation part, where we divide country samples into two subsamples (one of which has higher GDP per capita than the median, labeled rich subsample, and the other has lower one, labeled poor subsample), we adopt the categorization of civil, common, and socialist law groups rather than that of British, French, German, Scandinavian, and socialist legal origin groups, because the poor country subsample does not have any sample with German or Scandinavian legal origin.

[21]In addition to the control variables mentioned above, effects caused by "economic activity area based on language use" may play an important role. For example, Spanish or Arabic speaking areas generates the economic activity regions. To account for this effect, we conducted an additional regressions including "Status 0 language" dummy variables. In Ethnologue, Status 0 languages are defined as highly globally used languages, including the following languages: Arabic, Chinese, English, French, Russian, and Spanish. We run regressions including five dummy variables which takes 1 if a country uses Arabic, Chinese, French, Russian, and Spanish as its official languages. The results are more or less the same, but as expected, DLD impact is reported weaker than the results which would appear in this chapter. Results are available upon request.

hind in terms of economic development, we cannot separate the effect given by linguistic distance from African-specific factors. To deal with this pitfall stemming from dropping continent dummy variables, we conduct the empirical analysis in a spatial econometrics framework, which resolves spatial dependences among observations. This resettlement to the empirical field of spatial econometrics is natural since continent dummy variables are considered the simplest way to account for the factors related to spatial characteristics of observations. For instance, Attfield et al. (2000) shed lights on the relationship between continent dummy variable and spatial econometrics model (based on distance among countries) in the cross-country level economic growth context. Moreno and Trehan (1997) and Maurseth (2003) also conduct empirical analysis of economic growth at the cross-country level using spatial econometrics tools, and both conclude the effectiveness of employing those tools to find the clustered economic growth. Further, Romero and Burkey (2011) analyze the impact of debt/GDP ratio on GDP levels with spatial empirics, whose scope is restricted to Eurozone. In addition, cross-country differences and spillover effects of the quality of governance and institution are inspected using spatial econometrics models (Seldadyo et al., 2010; Kelejian et al., 2013). Studies on the political/economic features in the context of spatial empirical analysis are not limited to the country level, but its usefulness is confirmed at the regional-level analysis. Spatial growth regressions in terms of regional analysis are conducted in López-Bazo et al. (2004) and Le Sage and Fischer (2008) for European nations, Seya et al. (2012) for Japan, and Soundararajan (2013) for India.

Following such literature, our national-level growth analysis is done by using spatial econometrics. To be specific, on inspecting Moran's I statistics, the use of spatial econometrics is confirmed by the presence of spatial dependences in the OLS residuals with the controls mentioned above for countries with GDP per capita higher than the median level and with the choice of most of the various combinations of linguistic distance parameters (δ_D, δ_I), both for ILD_{PC} and ILD_{CC}, at the 5% significant level.[22]

3.3.2 Spatial econometrics framework

Based on the discussion on dropping the continent dummy variables in Section 3.3.1, our first possibility of the choice of spatial econometrics model (although this would not be adopted in the empirical part for reasons that will be mentioned later) is spatial Durbin model (SDM) in (3.13):

$$y = \rho W y + \alpha \iota + X\beta + WX\theta + \epsilon. \tag{3.13}$$

The dependent variable y represents an n by 1 observed vector of GDP per capita, where n is the number of observations. The i-th row of the n by k explanatory variable matrix X,

[22]Details of Moran's I statistics are provided upon request.

other than the intercept vector ι (which is n by 1 vector), is $x_i = (DLD_i, ILD_i, x_{\text{control},i})$. The i, j-th entry of the spatial weight matrix W is inverse distance spatial weight matrix without cut-off:[23]

$$w_{ij} = \begin{cases} d_{ij}^{-1} / \sum_{j=1}^{n}(d_{ij}^{-1}) & \text{if } i \neq j \\ 0 \quad \text{otherwise,} \end{cases} \tag{3.14}$$

where d_{ij} is the distance (calculated following the great circle formula) between capital cities of countries i and j. By convention, the principal diagonal elements of the spatial weight matrix are set zero, and in (3.14) the weight matrix is row standardized.[24] The disturbance vector ϵ is an n by 1 normally distributed vector, $\epsilon \sim N(0, \sigma^2 I_n)$. Parameters to be estimated in SDM are $\rho, \alpha, \beta, \theta$ and σ, where β and θ are k by 1 vectors, $\beta = (\beta_D, \beta_I, \beta_{\text{control}})'$ and $\theta = (\theta_D, \theta_I, \theta_{\text{control}})'$. As in Le Sage and Pace (2009, Chapter 2), one of the motivations for the use of SDM is to deal with problems caused by omitted regional variables. It is unlikely that the omitted regional variables are not correlated with at least one of the chosen explanatory variables, which is also the case with our context, since we drop the continent dummy variables because of the high correlation between DLD indexes and the Sub-Saharan Africa dummy variable.[25] However, SDM often suffers from multicollinearity among the variables and the spatial lagged ones (Seya et al., 2012). Indeed, DLD indexes suffer from this problem in our model, which forces us to give up adopting SDM, because our primary interest is on the impact of DLD (and ILD) to the economic prosperity. Then, our second best option is to choose the spatial autoregressive model (SAR) in (3.15):

$$y = \rho W y + \alpha \iota + X\beta + \epsilon \tag{3.15}$$

The definitions of the variable and disturbance vectors (or matrices) are the same as in SDM except for the exclusion of the term, $WX\theta$, in (3.13).[26] Although we realize that SAR model is not a perfect remedy to deal with the problem of dropping continent dummy variables, as it has already been found that OLS model may not be appropriate, we adhere to it. Model (3.15) is estimated maximum likelihood (ML) estimation, following Le Sage and Pace (2009, Chapter 3).

For interpretation of coefficient estimates, estimates of β are no longer valid when SAR

[23] For different types of spatial weight matrix, we conduct estimations in Section 3.5.

[24] Estimates with another type of standardization of the inverse distance spatial weight matrix are shown in Section 3.5.

[25] For this reason, the spatial error model (SEM), which is often used in spatial growth literature is inappropriate, because SEM can only be useful when there are no omitted variables spatially related (Le Sage and Fischer, 2008).

[26] Under this framework, it should be tested whether there is spatial correlation in the residuals of an SAR model. For SAC model, $y = \rho W y + \alpha \iota + X\beta + u$, where $u = \lambda W u + \epsilon$ and $\epsilon \sim N(0, \sigma^2 I)$, if the null hypothesis of $\lambda = 0$ is rejected, SAC rather than SAR should be adopted. However, we conducted LM tests for spatial correlation in the residuals of the SAR model at various combinations of linguistic distance parameters (δ_D, δ_I) to conclude that there may not be spatial correlations in the SAR residual terms. Thus, SAR would be more appropriate. Statistical tests are conducted using the MATLAB routine 'sarlm' posted at Le Sage's website at www.spatial-econometrics.com. Details are provided upon request.

is employed as is argued in Le Sage and Fischer (2008), Le Sage and Pace (2009, Chapter 2), and Seya et al. (2012). Because a change in a single observation with respect to any given explanatory variable affects not only that observation but also other observations due to their spatial relationships, linear regression parameters do not have a straightforward interpretation. Unlike in OLS models ($y_i = \alpha + \sum_{r=1}^{k} x_{r,i}\beta_r + \epsilon_i$), where the impact on y_i given by a change in $x_{r,i}$ is $\partial y_i/\partial x_{r,i} = \beta_r$ for all i and r, a coefficient in SAR, β_r, is not a partial derivative of y_i with respect to $x_{r,i}$. Instead,

$$\partial y_i/\partial x_{r,i} = S_r(W)_{ii} \quad (\neq \beta_r), \tag{3.16}$$

where $S_r(W) \equiv (I_n - \rho W)^{-1}\beta_r$ and $S(W)_{ii}$ is the (i,i)-th element of n by n matrix $S(W)$, and this expression in (3.16) is called the direct effect (from i to i).[27] Although the direct effect (3.16) is an easy way to draw interpretation for each observation i, the direct effect of an explanatory variable differs over all observations in general. Then, a summary measure, average direct effect, is proposed by averaging over the direct effects of all observations i:

$$\bar{M}(r)_{\text{direct}} = n^{-1}tr(S_r(W)). \tag{3.17}$$

The average direct effect is a natural measure following the spirit of OLS regression coefficient interpretations, which represents the averaged response of the dependent variable to the independent variables over the samples, such that this effect in (3.17) is the counterpart of the OLS coefficients. Further, to draw statistical inferences on the significance of average direct effects, we use the variation of 1,000 simulated parameter combinations drawn from the multivariate normal distribution implied by the ML estimates $(\hat{\alpha}, \hat{\beta}, \hat{\rho}, \hat{\sigma})$.[28] In Section 3.4, more weight is placed on inspection of the average direct effects of linguistic distance indexes, DLD and ILD, rather than on the coefficient estimates.

We conclude this section with some treatment for reverse causality. While one of the primary interests of this chapter is the impact of DLD index on GDP per capita, there may be a concern on reverse causality, that is, the possibility of higher GDP per capita countries attracting immigrants, which may affect DLD index. In order to avoid this problem, data on explanatory variables have not been collected for the year 2011 (which is the year in which GDP data were collected) but for the year 2009 for the linguistic distance data and the year 2010 for the data of population size, years of schooling, trade openness, quality of institution, property rights, and civil liberality, following the procedure stated

[27] The formula of the direct effect can be more easily understood if we write $y = \sum_{r=1}^{k} S_r(W)x_r + (I_n - \rho W)^{-1}\iota_n\alpha + (I_n - \rho W)^{-1}\epsilon$. Also, the indirect effects (from j to i) are expressed in a similar way, i.e., $\partial y_i/\partial x_{r,j} = S_r(W)_{ij} \quad (\neq 0)$.

[28] For details, see Le Sage and Pace (2009, Chapter 2) and Elhorst (2014, Chapter 2). This procedure can be conducted using MATLAB routine 'sar' posted at Le Sage's website at www.spatial-econometrics.com.

in Seldadyo et al. (2010), Romero and Burkey (2011), and Adsera and Pytlikova (2012).[29]

3.4 Empirical Results

3.4.1 OLS results with a set of certain values of linguistic distance parameters

The interests lie on the effect of domestic and international linguistic costs, and our first prediction is that both intra- and international linguistic distances have negative impacts on income/productivity levels. If it is difficult to access the domestic linguistic center, it is hard to communicate within countries. Further, acquiring other languages than one's mother tongue is costly, which may reduce economic activity performance levels. The same can be asserted for international linguistic distance. If mastering English is costly so that individuals have difficulty in fluent English communication, they are likely to lose opportunities to create global connections (for international trade, investment, or education).

For illustration, we first run OLS regressions to estimate the model specified in (3.12). In this section, a set of linguistic distance parameters is set equal to $(\delta_D, \delta_I) = (0.8, 0.6)$. In Section 3.4.3, we vary values of the linguistic parameters.[30] Columns (1) and (2) in Tables 3.2 and 3.3 report estimated results for $(DLD(0.8), ILD_{PC}(0.6))$ and $(DLD(0.8), ILD_{CC}(0.6))$ with the full sample of countries, respectively. DLD, as expected, has a negative impact on output-based GDP per capita even for the fully controlled model at the 5% significance level, which coordinates with the results from the literature, where ethnic diversity has negative impacts on economic success. As pointed out in Section 3.1, benefits from socioeconomic diversity and costs of heterogeneous society are two sides of the same coin, it is reasonable that domestic linguistic distance affects GDP levels in a negative way. On the other hand, both ILD_{PC} and ILD_{CC} lose their significance when fully controlled, and in the ILD_{PC} model, although it is insignificant, ILD is estimated to be positive unexpectedly. The reasons behind this unexpected result might be that while in developing countries, only a small number of elites is required to acquire English, the rest of the residents may not. In sufficiently developed nations, on the other hand, most of the residents learn English in elementary, secondary, or higher education. This, in turn, implies that for developing countries, ILD may not have a large impact on income/productivity when the data are aggregated to country levels rather than individual levels. Moreover, in light of firms' activities, it can be predicted that linkages to the international center is more important for already developed countries than for developing ones. For example, firms may prefer to build branches in sufficiently developed nations instead of countries with

[29] In addition, for other explanatory variables than those mentioned, the years of data collection are precede the year 2011.

[30] The OLS results with various values of the linguistic distance parameter, δ, are provided upon request.

Table 3.2: Linguistic distance and economic development, ILD_{PC} international linguistic distance index, OLS results

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Full	Full	Poor	Poor	Rich	Rich
Dependent variable (log)	GDP/capita					
DLD ($\delta_D = 0.8$)	-1.598***	-0.484**	-0.869**	-0.636***	0.190	0.575**
	(0.289)	(0.230)	(0.353)	(0.182)	(0.255)	(0.223)
ILD_{PC} ($\delta_I = 0.6$)	-1.955***	0.021	-2.333	-0.553	-0.864***	-0.613**
	(0.405)	(0.238)	(1.564)	(1.019)	(0.161)	(0.240)
Years of schooling		0.263***		0.192***		0.078**
		(0.037)		(0.028)		(0.030)
Population size (log)		-0.051		-0.060		0.091*
		(0.054)		(0.072)		(0.052)
Land area (log)		0.110**		0.175**		-0.093*
		(0.054)		(0.075)		(0.050)
Landlockedness		-0.125		-0.456**		0.216
		(0.165)		(0.168)		(0.131)
Trade openness		0.229		-0.116		-0.002
		(0.143)		(0.227)		(0.125)
Absolute latitude		0.021***		0.019**		0.007
		(0.006)		(0.009)		(0.006)
Coastal population		0.363		-0.104		0.230
		(0.225)		(0.300)		(0.178)
Mean temperature (log)		0.136		0.361		-0.070
		(0.095)		(0.220)		(0.063)
Mean precipitation (log)		0.004		-0.014		-0.250**
		(0.115)		(0.129)		(0.100)
Agricultural suitability (mean)		0.055		0.547		-0.349*
		(0.255)		(0.548)		(0.180)
Agricultural suitability (std. dev.)		-0.560		-0.341		-0.900
		(0.560)		(0.677)		(0.557)
Institutional quality (Polity2)		-0.051*		-0.042*		-0.013
		0.026		0.021		0.024
Property rights		0.139		0.030		0.112
		(0.137)		(0.105)		(0.122)
Civil liberties		-0.074		0.070		-0.014
		(0.123)		(0.109)		(0.123)
Common law		-0.165		-0.107		-0.264
		(0.130)		(0.132)		(0.159)
Socialist law		-0.768***		-0.249		-0.525***
		(0.174)		(0.205)		(0.138)
Constant	11.134***	4.669***	10.506***	4.066**	10.533***	11.201***
	(0.288)	(1.293)	(1.347)	(1.709)	(0.114)	(1.426)
Observations	108	108	54	54	54	54
Adjusted R-squared	0.460	0.829	0.264	0.688	0.143	0.715

Robust standard errors are in parentheses. Omitted group for legal origin dummy variable: civil law. Subsample of poor countries: <median GDP/capita. Subsample of rich countries: >median GDP/capita.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.3: Linguistic distance and economic development, ILD_{CC} international linguistic distance index, OLS results

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Full	Full	Poor	Poor	Rich	Rich
Dependent variable (log)	GDP/capita					
DLD ($\delta_D = 0.8$)	-2.228***	-0.488**	-1.060***	-0.725***	-0.098	0.431*
	(0.249)	(0.233)	(0.294)	(0.212)	(0.313)	(0.252)
ILD_{CC} ($\delta_I = 0.6$)	-0.662**	-0.016	0.108	-0.161	-0.429**	-0.494**
	(0.314)	(0.244)	(0.263)	(0.264)	(0.205)	(0.214)
Years of schooling		0.262***		0.196***		0.096***
		(0.036)		(0.025)		(0.033)
Population size (log)		-0.049		-0.043		0.060
		(0.058)		(0.074)		(0.051)
Land area (log)		0.109**		0.162**		-0.082*
		(0.053)		(0.076)		(0.047)
Landlockedness		-0.124		-0.459**		0.142
		(0.168)		(0.180)		(0.132)
Trade openness		0.232		-0.150		-0.043
		(0.144)		(0.216)		(0.120)
Absolute latitude		0.021***		0.020**		0.007
		(0.006)		(0.009)		(0.007)
Coastal population		0.362		-0.126		0.178
		(0.226)		(0.298)		(0.186)
Mean temperature (log)		0.134		0.381		-0.061
		(0.092)		(0.226)		(0.063)
Mean precipitation (log)		-0.001		-0.002		-0.274**
		(0.116)		(0.123)		(0.107)
Agricultural suitability (mean)		0.058		0.445		-0.303
		(0.259)		(0.556)		(0.186)
Agricultural suitability (std. dev.)		-0.560		-0.191		-0.661
		(0.577)		(0.779)		(0.505)
Institutional quality (Polity2)		-0.050*		-0.036*		-0.035
		(0.026)		(0.019)		(0.022)
Property rights		0.139		0.006		0.122
		(0.133)		(0.111)		(0.124)
Civil liberties		-0.075		0.086		0.055
		(0.120)		(0.113)		(0.123)
Common law		-0.173		-0.167		-0.313*
		(0.142)		(0.146)		(0.176)
Socialist law		-0.765***		-0.288		-0.607***
		(0.172)		(0.199)		(0.145)
Constant	10.195***	4.756***	8.315***	3.737***	10.242***	10.702***
	(0.264)	(1.182)	(0.312)	(1.199)	(0.158)	(1.268)
Observations	108	108	54	54	54	54
Adjusted R-squared	0.392	0.829	0.228	0.688	0.022	0.708

Robust standard errors are in parentheses. Omitted group for legal origin dummy variable: civil law. Subsample of poor countries: <median GDP/capita. Subsample of rich countries: >median GDP/capita.
*** p<0.01, ** p<0.05, * p<0.1

low economic activity levels, simply because the latter tend to lack social infrastructure or sufficient understanding of instructions even in national languages. Thus, without the domestic economic basement, connection to the rest of the world may be possibly meaningless.

Then, our modified prediction is that difficulty for accessing the international linguistic center has a negative impact solely on sufficiently advanced countries, while it does not have significantly negative effect on less developed ones. We do not deem that accessibility to the international linguistic center is unimportant for economically backward nations, and thus, we still predict a negative sign of the coefficient on ILD. In addition, revision of the prediction for the impact of DLD when samples are split into rich and poor countries (rich: $>$ median GDP/capita, poor: $<$ median GDP/capita) is made. As pointed out in the literature review in Section 3.1, ethnic diversity has a non-linear impact on economic prosperity. Countries with low and high (not medium) income level may tend to show slightly high ethnic diversities. In our context, higher domestic linguistic distance may sometimes imply more diverse nations (i.e., a completely uniform nation in terms of domestic language shows zero within-country language distance, while a nation with several language groups shows positive within-country linguistic distance), such that DLD is predicted to exhibit a significantly negative impact for low income countries and a positive (significant or insignificant) impact on high income ones.

Columns (3) and (4) in Tables 3.2 and 3.3 report the results for poor country subsample for the choice of ILD_{PC} and ILD_{CC}, respectively. Columns (5) and (6) in Tables 3.2 and 3.3 report the results for the rich country subsample. Both estimation results on the adoption of ILD_{PC} and ILD_{CC} exhibit almost expected outcomes: (I) for poor countries, domestic linguistic distance has a significantly negative impact on output-based GDP per capita, (II) for poor countries, international linguistic distance does not have a significant impact, although it still shows negative signs, (III) for rich countries, domestic distance shows a positive effect on GDP per capita, and (IV) for rich countries, international distance significantly decreases GDP per capita. What characterizes our results and distinguishes these from those in the literature is (I), (III), and (IV). Previous literature focuses on the negative effect of ethnolinguistic diversity on economic performance, especially for poor countries, as reviewed in Section 3.1. They interpret that the negative effect of ethnolinguistic diversity on countries at the early stages of economic development is caused by communication costs between different groups. In other words, they indirectly capture communication difficulty caused by difference in language usage as negative impacts made by ethnolinguistic diversity (communication difficulty). Instead, we tackle this negative effect directly from the view of linguistic distance cost, which leads to result (I). For result (III), behind the seemingly positive effect of the domestic linguistic distance for rich countries, the benefits of diversity on production are hidden. It may be argued that basic skill improvement (acquiring national languages) is necessary for developing countries, while intensive skill improvement (such as R & D sector industry, which requires knowledge creation by diversity) is beneficial for richer countries. Another essential determinant of economic prosperity for advanced countries, captured by result (IV), is the proximity and accessibility to the international linguistic center, which measures how costless and smooth communication between different linguistic groups is, or what large portion individuals have in common when exchanging ideas.

3.4.2 SAR results with a set of certain values of linguistic distance parameters

In this section, we investigate SAR results based on the specification of (3.15) along with average direct effects. The set of linguistic distance parameters is set $(\delta_D, \delta_I) = (0.8, 0.6)$, which is the same parameter combination in Section 3.4.1. Columns (1a) and (1b) in Tables 3.4 and 3.5 report estimated results for $(DLD(0.8), ILD_{PC}(0.6))$ and $(DLD(0.8), ILD_{PC}(0.6))$ with the full sample of countries, respectively. Columns (2a) and (2b) in Tables 3.4 and 3.5 report the results for the poor country subsample for the choice of ILD_{PC} and ILD_{CC}, respectively, and columns (3a) and (3b) in Tables 3.4 and 3.5 report the results for the rich country subsample. Both tables report coefficient estimates in columns labeled (a) and average direct effects in (b). Since we have mentioned in Section 3.3.1 that spatial dependences exist among rich country samples, it should be checked whether estimates of the spatial lag ρ are significant, to certify that the adoption of SAR is appropriate. Columns (3a) in Tables 3.4 and 3.5 report the estimates of ρ, both of which show statistical significance.[31]

To see the impacts of DLD and ILD to GDP per capita, we focus on columns labeled (b). As with OLS results in Section 3.4.1, DLD negatively affects GDP levels significantly for the full and poor country (sub)samples, while it has a positive impact on them in the case of the rich subsample. For ILD_{PC} specification, the average direct effect for DLD is significant as in OLS results for rich countries. With the choice of ILD_{CC}, although DLD is insignificant, it still shows a positive sign as expected. In terms of ILD's impact on GDP, ILD is significantly negative only for rich subsample cases, both under the choice of ILD_{PC} and ILD_{CC} as linguistic distance index. This is coherent with our prediction,

[31]As in Tables 3.4 and 3.5, model specification based on ILD_{CC} shows more significant estimates of ρ than the one based on ILD_{PC}. This tendency is common for all parameter sets of (δ_D, δ_I). Although there is slightly weaker significance of ρ in the specification with ILD_{PC}, in all combinations of (δ_D, δ_I), the estimate of ρ is significant at the 5% or 10% levels. For ILD_{CC} specification, ρ is significant at the 1%, 5%, or 10% levels. The estimated results showing only rich countries enjoy proximity to the other rich countries is interpreted as follows. At the early stage of economic development, poor countries do not enjoy proximity to other poor countries, since in those less developed countries, the main industries are characterized by easy and simple tasks. On the other hand, sufficiently developed rich countries can enjoy proximity to other rich countries, because industries in those rich countries are more likely to be characterized by complex and creative tasks which would be improved by introducing other well-developed countries' technologies through interactions.

Table 3.4: Linguistic distance and economic development, ILD_{PC} international linguistic distance index, SAR results, inverse distance spatial weight matrix (row standardized)

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Sample	Full	Full	Poor	Poor	Rich	Rich
Dependent variable (log)	GDP/capita					
	Coefficient	Direct effect	Coefficient	Direct effect	Coefficient	Direct effect
DLD ($\delta_D = 0.8$)	-0.418**	-0.425**	-0.627***	-0.623***	0.522**	0.543**
	(0.196)	(0.195)	(0.207)	(0.207)	(0.236)	(0.255)
ILD_{PC} ($\delta_I = 0.6$)	0.101	0.117	-0.548	-0.531	-0.599**	-0.641**
	(0.326)	(0.324)	(1.088)	(1.083)	(0.235)	(0.280)
Years of schooling	0.254***	0.257***	0.191***	0.193***	0.073**	0.077**
	(0.028)	(0.027)	(0.030)	(0.031)	(0.032)	(0.036)
Population size (log)	-0.064	-0.066	-0.060	-0.060	0.059	0.060
	(0.054)	(0.056)	(0.061)	(0.061)	(0.055)	(0.058)
Land area (log)	0.119**	0.122**	0.174***	0.173**	-0.070	-0.074
	(0.052)	(0.054)	(0.065)	(0.066)	(0.051)	(0.053)
Landlockedness	-0.099	-0.100	-0.456**	-0.463**	0.215	0.220
	(0.154)	(0.151)	(0.187)	(0.181)	(0.134)	(0.146)
Trade openness	0.153	0.162	-0.115	-0.109	-0.032	-0.035
	(0.166)	(0.161)	(0.260)	(0.258)	(0.118)	(0.126)
Absolute latitude	0.017**	0.017**	0.019*	0.018*	0.004	0.004
	(0.008)	(0.008)	(0.010)	(0.010)	(0.006)	(0.007)
Coastal population	0.378*	0.381*	-0.104	-0.111	0.197	0.204
	(0.228)	(0.229)	(0.298)	(0.302)	(0.194)	(0.190)
Mean temperature (log)	0.129	0.132	0.355	0.350	-0.071	-0.074
	(0.115)	(0.118)	(0.254)	(0.263)	(0.073)	(0.082)
Mean precipitation (log)	0.006	0.006	-0.017	-0.025	-0.246**	-0.265**
	(0.104)	(0.103)	(0.137)	(0.136)	(0.101)	(0.116)
Agricultural suitability (mean)	0.054	0.057	0.552	0.579	-0.305	-0.321
	(0.306)	(0.307)	(0.517)	(0.520)	(0.202)	(0.213)
Agricultural suitability (std. dev.)	-0.595	-0.608	-0.341	-0.298	-0.823*	-0.849
	(0.633)	(0.645)	(0.825)	(0.791)	(0.467)	(0.518)
Institutional quality (Polity2)	-0.052***	-0.052***	-0.042**	-0.043**	-0.018	-0.017
	(0.019)	(0.019)	(0.019)	(0.020)	(0.024)	(0.026)
Property rights	0.123	0.122	0.031	0.035	0.159	0.164
	(0.086)	(0.090)	(0.093)	(0.095)	(0.106)	(0.121)
Civil liberties	-0.047	-0.049	0.071	0.068	-0.039	-0.041
	(0.097)	(0.100)	(0.098)	(0.099)	(0.115)	(0.122)
Common law	-0.079	-0.081	-0.104	-0.105	-0.269*	-0.285*
	(0.130)	(0.130)	(0.119)	(0.123)	(0.149)	(0.167)
Socialist law	-0.775***	-0.785***	-0.248	-0.238	-0.537***	-0.564***
	(0.165)	(0.169)	(0.227)	(0.222)	(0.143)	(0.177)
Constant	1.765		3.874*		6.294**	
	(2.207)		(2.310)		(2.910)	
Spatial lag (ρ)	0.327		0.029		0.478*	
	(0.211)		(0.262)		(0.254)	
Observations	108		54		54	
Adjusted R-squared	0.834		0.689		0.732	
Log likelihood	-34.259		-0.167		22.939	

Standard errors are in parentheses. Direct effect means average direct effect. Omitted group for legal origin dummy variable: civil law. Subsample of poor countries: <median GDP/capita. Subsample of rich countries: >median GDP/capita.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.5: Linguistic distance and economic development, ILD_{CC} international linguistic distance index, SAR results, inverse distance spatial weight matrix (row standardized)

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Sample	Full	Full	Poor	Poor	Rich	Rich
Dependent variable (log)	GDP/capita					
	Coefficient	Direct effect	Coefficient	Direct effect	Coefficient	Direct effect
DLD ($\delta_D = 0.8$)	-0.417**	-0.427*	-0.708***	-0.715***	0.348	0.379
	(0.212)	(0.217)	(0.208)	(0.209)	(0.252)	(0.286)
ILD_{CC} ($\delta_I = 0.6$)	-0.004	-0.010	-0.167	-0.164	-0.532**	-0.569**
	(0.223)	(0.223)	(0.265)	(0.268)	(0.212)	(0.273)
Years of schooling	0.253***	0.254***	0.195***	0.194***	0.091***	0.100**
	(0.027)	(0.026)	(0.029)	(0.029)	(0.033)	(0.041)
Population size (log)	-0.061	-0.064	-0.043	-0.044	0.023	0.025
	(0.057)	(0.056)	(0.070)	(0.070)	(0.052)	(0.058)
Land area (log)	0.116**	0.118**	0.161**	0.163**	-0.058	-0.062
	(0.053)	(0.053)	(0.068)	(0.070)	(0.049)	(0.055)
Landlockedness	-0.098	-0.097	-0.457**	-0.463**	0.145	0.159
	(0.155)	(0.148)	(0.184)	(0.181)	(0.128)	(0.143)
Trade openness	0.166	0.164	-0.145	-0.154	-0.076	-0.083
	(0.164)	(0.160)	(0.235)	(0.230)	(0.116)	(0.130)
Absolute latitude	0.016**	0.016**	0.019**	0.019**	0.002	0.002
	(0.008)	(0.008)	(0.009)	(0.009)	(0.006)	(0.008)
Coastal population	0.375*	0.381*	-0.125	-0.131	0.140	0.150
	(0.228)	(0.222)	(0.297)	(0.291)	(0.191)	(0.203)
Mean temperature (log)	0.124	0.126	0.370	0.371	-0.068	-0.072
	(0.115)	(0.116)	(0.256)	(0.260)	(0.073)	(0.084)
Mean precipitation (log)	-0.004	-0.008	-0.009	-0.005	-0.287**	-0.316**
	(0.107)	(0.107)	(0.133)	(0.127)	(0.111)	(0.157)
Agricultural suitability (mean)	0.073	0.074	0.452	0.455	-0.250	-0.270
	(0.301)	(0.300)	(0.522)	(0.506)	(0.200)	(0.219)
Agricultural suitability (std. dev.)	-0.611	-0.596	-0.185	-0.204	-0.577	-0.647
	(0.634)	(0.635)	(0.860)	(0.891)	(0.457)	(0.512)
Institutional quality (Polity2)	-0.052***	-0.053***	-0.036*	-0.037*	-0.039*	-0.041
	(0.020)	(0.020)	(0.019)	(0.019)	(0.021)	(0.024)
Property rights	0.127	0.134	0.007	0.007	0.172	0.189
	(0.086)	(0.086)	(0.101)	(0.100)	(0.105)	(0.127)
Civil liberties	-0.054	-0.058	0.088	0.091	0.026	0.024
	(0.095)	(0.094)	(0.102)	(0.102)	(0.113)	(0.123)
Common law	-0.096	-0.095	-0.163	-0.161	-0.348**	-0.379*
	(0.152)	(0.154)	(0.154)	(0.160)	(0.171)	(0.212)
Socialist law	-0.764***	-0.766***	-0.284	-0.283	-0.621***	-0.675***
	(0.163)	(0.170)	(0.223)	(0.220)	(0.141)	(0.204)
Constant	2.101		3.372*		5.039*	
	(2.135)		(2.023)		(2.572)	
Spatial lag (ρ)	0.313		0.058		0.572**	
	(0.210)		(0.261)		(0.229)	
Observations	108		54		54	
Adjusted R-squared	0.834		0.691		0.728	
Log likelihood	-34.303		-0.094		22.829	

Standard errors are in parentheses. Direct effect means average direct effect. Omitted group for legal origin dummy variable: civil law. Subsample of poor countries: <median GDP/capita. Subsample of rich countries: >median GDP/capita.
*** p<0.01, ** p<0.05, * p<0.1

and only developed countries can enjoy the effect of accessibility to the international communication at the nationally aggregated level.

In concluding this section, we investigate why languages and linguistic distances give such strong impacts on countries' income levels. As pointed out in Acemoglu et al. (2001), long-lasting effects of the quality of early institutions by colonizers determine the post-colonies' current performance. The linguistic distance impact on cross-country income difference is explained by similar persistence of acquired languages. Once languages are mastered, they are sublimated into internalized knowledge, and the ability to use those languages can never be completely separated from the users. If social elites have mastered the colonizers' language, then persistence of their language should be much stronger. In this sense, the colonizers' language bears a lock-in effect. The same thing can be asserted for English's power: successive past and present superpowers of the world (British Empire and U.S.) have had English as national languages, and thus, English has continuously been endowed with the most powerful position among all existing languages of the world. Eventually, English has strongly been locked in.

3.4.3 Results on a full range of the linguistic distance parameters

Since linguistic distance parameters can take various values in the range they are defined, we vary the parameter values and rerun regressions. Tables 3.6-3.13 report average direct effects of domestic and international linguistic distances with different linguistic parameter values.

Tables 3.6-3.8 (Tables 3.10-3.12, respectively) report average direct effects of DLD based on the choice of ILD_{PC} (ILD_{CC}, respectively) as the international linguistic distance index. Tables 3.6 and 3.10 report the average direct effects of DLD for the full sample, Tables 3.7 and 3.11 for the subsample of poor countries, and Tables 3.8 and 3.12 for the subsample of rich countries, respectively. Further, Tables 3.9 and 3.13 report average direct effects of ILD, respectively under the choice of ILD_{PC} and ILD_{CC} for the subsample of rich countries.[32] When investigating the average direct effects on DLD matrices, unilaterally row-by-row (i.e., left to right) comparisons should be made, since given a certain value of δ_I, behavior of DLD direct effects is determined by the change of δ_D. Similarly, unilaterally column-by-column (i.e., top to bottom) investigations of ILD direct effects should be done.[33] For expositional convenience, we first skip interpretation of Tables 3.6 and 3.10, and after investigating other tables, return to them.

In Tables 3.7 and 3.11, direct effects of DLD are significantly negative at the 1% level over the whole range of the matrix. As expected, domestic linguistic distance sharply

[32]In this section, we concentrate on the matrices whose results show significance. The results omitted from the main text, that is, direct effects of ILD_{PC} and ILD_{CC} for the full sample of countries and the subsample of the poor countries are provided for readers upon request.

[33]Wherever it is clear, we refer to "direct effect" rather than "average direct effect" in this section.

Table 3.6: DLD and economic development on a full range of δ_D and δ_I, full sample of countries, ILD_{PC} international linguistic distance index, SAR results with inverse distance spatial weight matrix (row standardized)

$\delta_I \backslash \delta_D$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	-0.398**	-0.423**	-0.420**	-0.417**	-0.411**	-0.404*	-0.407**	-0.406**	-0.405*
	(0.1887)	(0.1981)	(0.2007)	(0.2002)	(0.2036)	(0.2095)	(0.2032)	(0.1932)	(0.2052)
0.2	-0.409**	-0.425**	-0.434**	-0.429**	-0.402**	-0.409**	-0.422**	-0.411**	-0.397**
	(0.1839)	(0.1979)	(0.1931)	(0.1992)	(0.1990)	(0.2040)	(0.2023)	(0.2045)	(0.1945)
0.3	-0.419**	-0.427**	-0.411**	-0.429**	-0.425**	-0.414**	-0.418**	-0.413**	-0.414**
	(0.1880)	(0.1975)	(0.1986)	(0.1986)	(0.1926)	(0.1988)	(0.1881)	(0.2000)	(0.1964)
0.4	-0.430**	-0.429**	-0.439**	-0.447**	-0.429**	-0.428**	-0.426**	-0.423**	-0.421**
	(0.1864)	(0.1916)	(0.1982)	(0.2049)	(0.1983)	(0.1999)	(0.1935)	(0.1931)	(0.2010)
0.5	-0.427**	-0.434**	-0.432**	-0.441**	-0.442**	-0.427**	-0.436**	-0.406**	-0.413**
	(0.1824)	(0.1917)	(0.1942)	(0.1995)	(0.1992)	(0.1907)	(0.1981)	(0.1991)	(0.1931)
0.6	-0.425**	-0.433**	-0.436**	-0.442**	-0.437**	-0.418**	-0.413**	-0.425**	-0.418**
	(0.1904)	(0.1946)	(0.1936)	(0.1930)	(0.2066)	(0.1975)	(0.2011)	(0.1950)	(0.2007)
0.7	-0.423**	-0.433**	-0.440**	-0.431**	-0.432**	-0.436**	-0.422**	-0.415**	-0.411**
	(0.1907)	(0.1899)	(0.1895)	(0.1987)	(0.1956)	(0.1948)	(0.1981)	(0.2014)	(0.1939)
0.8	-0.424**	-0.429**	-0.428**	-0.431**	-0.439**	-0.421**	-0.429**	-0.419**	-0.423**
	(0.1903)	(0.1921)	(0.2041)	(0.1912)	(0.1961)	(0.2014)	(0.1866)	(0.2027)	(0.2000)
0.9	-0.420**	-0.429**	-0.446**	-0.435**	-0.429**	-0.428**	-0.409**	-0.424**	-0.417**
	(0.1882)	(0.1875)	(0.1877)	(0.1951)	(0.1960)	(0.1904)	(0.1999)	(0.1925)	(0.1983)

Table shows average direct effects of DLD on a full range of linguistic distance indexes at different combinations of δ_D and δ_I. GDP/capita as the dependent variable. All results include the full vector of controls. Full country sample. ILD_{PC} international linguistic distance index. SAR model. Inverse distance spatial weight matrix (row standardized). Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.7: DLD and economic development on a full range of δ_D and δ_I, subsample of poor countries, ILD_{PC} international linguistic distance index, SAR results with inverse distance spatial weight matrix (row standardized)

$\delta_I \backslash \delta_D$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	-0.548***	-0.567***	-0.591***	-0.611***	-0.621***	-0.624***	-0.608***	-0.636***	-0.631***
	(0.1808)	(0.1827)	(0.1912)	(0.2059)	(0.2043)	(0.2038)	(0.1953)	(0.2063)	(0.2064)
0.2	-0.532***	-0.563***	-0.588***	-0.591***	-0.610***	-0.628***	-0.631***	-0.634***	-0.639***
	(0.1763)	(0.1821)	(0.1878)	(0.1941)	(0.2136)	(0.2059)	(0.2065)	(0.2037)	(0.2162)
0.3	-0.543***	-0.572***	-0.591***	-0.612***	-0.612***	-0.624***	-0.622***	-0.622***	-0.628***
	(0.1770)	(0.1906)	(0.1915)	(0.1957)	(0.2036)	(0.2115)	(0.2019)	(0.2079)	(0.1972)
0.4	-0.538***	-0.575***	-0.578***	-0.597***	-0.621***	-0.625***	-0.631***	-0.628***	-0.641***
	(0.1822)	(0.1845)	(0.1920)	(0.1967)	(0.2004)	(0.2088)	(0.1975)	(0.2059)	(0.2100)
0.5	-0.539***	-0.562***	-0.599***	-0.591***	-0.610***	-0.621***	-0.634***	-0.628***	-0.632***
	(0.1834)	(0.1879)	(0.1957)	(0.1929)	(0.2049)	(0.2016)	(0.2013)	(0.2110)	(0.2121)
0.6	-0.544***	-0.555***	-0.601***	-0.602***	-0.610***	-0.622***	-0.635***	-0.623***	-0.627***
	(0.1785)	(0.1910)	(0.1924)	(0.1987)	(0.1989)	(0.2087)	(0.2114)	(0.2072)	(0.2088)
0.7	-0.544***	-0.559***	-0.594***	-0.591***	-0.605***	-0.629***	-0.620***	-0.634***	-0.632***
	(0.1774)	(0.1851)	(0.2015)	(0.1951)	(0.1917)	(0.2076)	(0.2032)	(0.2081)	(0.2040)
0.8	-0.521***	-0.568***	-0.579***	-0.604***	-0.606***	-0.614***	-0.629***	-0.645***	-0.634***
	(0.1811)	(0.1957)	(0.1930)	(0.1983)	(0.1980)	(0.2034)	(0.2054)	(0.2105)	(0.2129)
0.9	-0.534***	-0.559***	-0.591***	-0.610***	-0.618***	-0.639***	-0.635***	-0.628***	-0.640***
	(0.1839)	(0.1818)	(0.1936)	(0.2003)	(0.2045)	(0.2057)	(0.2046)	(0.2049)	(0.2051)

This table shows average direct effects of DLD on a full range of linguistic distance indexes at different combinations of δ_D and δ_I. GDP/capita is the dependent variable. All results include the full vector of controls. Subsample of poor countries. ILD_{PC} international linguistic distance index. SAR model. Inverse distance spatial weight matrix (row standardized). Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

drops the output-based GDP level for poor countries. As for the behavior of magnitude and significance of the direct effects of DLD, no general and clear tendency that can be said for all rows in both specifications of ILD_{PC} and ILD_{CC} can be found. In addition, Tables 3.7 and 3.11 correspond to the results of the previous section.

Table 3.8: DLD and economic development on a full range of δ_D and δ_I, subsample of rich countries, ILD_{PC} international linguistic distance index, SAR results with inverse distance spatial weight matrix (row standardized)

$\delta_I \backslash \delta_D$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.551	0.610	0.597*	0.637*	0.619**	0.637*	0.648**	0.617**	0.628**
	(0.3655)	(0.3827)	(0.3235)	(0.3321)	(0.2966)	(0.3223)	(0.2998)	(0.2761)	(0.2815)
0.2	0.543	0.588*	0.595*	0.628*	0.629*	0.614**	0.611**	0.611**	0.617**
	(0.3581)	(0.3441)	(0.3335)	(0.3303)	(0.3184)	(0.2940)	(0.2809)	(0.2861)	(0.2899)
0.3	0.544	0.563*	0.608*	0.594*	0.614**	0.611**	0.617**	0.596**	0.594**
	(0.3573)	(0.3109)	(0.3271)	(0.3123)	(0.2950)	(0.2924)	(0.2786)	(0.2634)	(0.2850)
0.4	0.529	0.555	0.602*	0.594*	0.598**	0.582**	0.575**	0.580**	0.584**
	(0.3459)	(0.3327)	(0.3388)	(0.3019)	(0.2845)	(0.2811)	(0.2582)	(0.2502)	(0.2783)
0.5	0.552	0.539*	0.563*	0.592*	0.568*	0.568**	0.574*	0.567**	0.567**
	(0.3831)	(0.3053)	(0.3201)	(0.3305)	(0.2941)	(0.2742)	(0.2916)	(0.2546)	(0.2539)
0.6	0.497	0.546	0.547*	0.554*	0.555*	0.558*	0.546*	0.543**	0.567*
	(0.3181)	(0.3352)	(0.2985)	(0.3072)	(0.2862)	(0.3086)	(0.2741)	(0.2550)	(0.3037)
0.7	0.508	0.515	0.551	0.528*	0.546*	0.542**	0.552*	0.528**	0.540**
	(0.3470)	(0.3199)	(0.3432)	(0.2887)	(0.2907)	(0.2635)	(0.3014)	(0.2460)	(0.2405)
0.8	0.471	0.523	0.530*	0.534*	0.530*	0.530*	0.533**	0.520**	0.520**
	(0.3152)	(0.3276)	(0.2942)	(0.2867)	(0.2768)	(0.2868)	(0.2570)	(0.2559)	(0.2380)
0.9	0.478	0.480	0.520	0.511*	0.540*	0.529*	0.519**	0.523**	0.521**
	(0.3336)	(0.3028)	(0.3131)	(0.2937)	(0.3022)	(0.2783)	(0.2413)	(0.2394)	(0.2559)

This table shows average direct effects of DLD on a full range of linguistic distance indexes at different combinations of δ_D and δ_I. GDP/capita is the dependent variable. All results include the full vector of controls. Subsample of rich countries. ILD_{PC} international linguistic distance index. SAR model. Inverse distance spatial weight matrix (row standardized). Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.9: ILD_{PC} and economic development on a full range of δ_D and δ_I, subsample of rich countries, ILD_{PC} international linguistic distance index, SAR results with inverse distance spatial weight matrix (row standardized)

$\delta_I \backslash \delta_D$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	-0.126	-0.119	-0.100	-0.101	-0.096	-0.096	-0.084	-0.082	-0.076
	(0.1785)	(0.1717)	(0.1646)	(0.1720)	(0.1636)	(0.1693)	(0.1797)	(0.1661)	(0.1639)
0.2	-0.241	-0.225	-0.214	-0.222	-0.210	-0.205	-0.192	-0.200	-0.199
	(0.2028)	(0.2063)	(0.2081)	(0.2019)	(0.2065)	(0.2041)	(0.1968)	(0.2092)	(0.2042)
0.3	-0.390	-0.350	-0.363	-0.342	-0.357	-0.329	-0.321	-0.330	-0.321
	(0.2491)	(0.2269)	(0.2515)	(0.2228)	(0.2450)	(0.2409)	(0.2208)	(0.2330)	(0.2365)
0.4	-0.531**	-0.512*	-0.492*	-0.461*	-0.463*	-0.459*	-0.451*	-0.452*	-0.461*
	(0.2609)	(0.2690)	(0.2596)	(0.2474)	(0.2473)	(0.2471)	(0.2462)	(0.2379)	(0.2612)
0.5	-0.646**	-0.607**	-0.582**	-0.595**	-0.582**	-0.557**	-0.570*	-0.560**	-0.566**
	(0.3156)	(0.2602)	(0.2650)	(0.2959)	(0.2779)	(0.2599)	(0.2869)	(0.2622)	(0.2684)
0.6	-0.692**	-0.686**	-0.660**	-0.668**	-0.645**	-0.668*	-0.622**	-0.641**	-0.655**
	(0.2815)	(0.2825)	(0.2557)	(0.3021)	(0.2839)	(0.3375)	(0.2729)	(0.2804)	(0.3205)
0.7	-0.771**	-0.748**	-0.743**	-0.697**	-0.704**	-0.688**	-0.705**	-0.678***	-0.676**
	(0.3170)	(0.2871)	(0.3484)	(0.2646)	(0.2923)	(0.2809)	(0.3199)	(0.2519)	(0.2619)
0.8	-0.763***	-0.747**	-0.738**	-0.726**	-0.724***	-0.713**	-0.720**	-0.689**	-0.700***
	(0.2794)	(0.2920)	(0.2942)	(0.2747)	(0.2687)	(0.2707)	(0.2761)	(0.2629)	(0.2602)
0.9	-0.829**	-0.764***	-0.767**	-0.764**	-0.757**	-0.742**	-0.728***	-0.732***	-0.714**
	(0.3422)	(0.2507)	(0.2885)	(0.2968)	(0.3186)	(0.2873)	(0.2611)	(0.2703)	(0.2718)

This table shows average direct effects of ILD_{PC} on a full range of linguistic distance indexes at different combinations of δ_D and δ_I. GDP/capita is the dependent variable. All results include the full vector of controls. Subsample of rich countries. ILD_{PC} international linguistic distance index. SAR model. Inverse distance spatial weight matrix (row standardized). Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Tables 3.8 and 3.12 present positive DLD direct effects for all cells, exhibiting weaker significance under the adoption of ILD_{CC} specification than of ILD_{PC}. Here again, tendencies that are deemed to be common to all rows for both specifications are not very clear. Tables 3.9 and 3.13 are the ones most interesting and instructive, because these

Table 3.10: DLD and economic development on a full range of δ_D and δ_I, full sample of countries, ILD_{CC} international linguistic distance index, SAR results with inversed distance spatial weight matrix (row standardized)

$\delta_I \backslash \delta_D$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	-0.515**	-0.522**	-0.524**	-0.508**	-0.505**	-0.489**	-0.470**	-0.472**	-0.469**
	(0.2131)	(0.2129)	(0.2114)	(0.2159)	(0.2110)	(0.1973)	(0.2015)	(0.2107)	(0.2029)
0.2	-0.521**	-0.532**	-0.515**	-0.514**	-0.506**	-0.488**	-0.469**	-0.458**	-0.453**
	(0.2192)	(0.2161)	(0.2198)	(0.2169)	(0.2092)	(0.2090)	(0.2101)	(0.2131)	(0.2067)
0.3	-0.510**	-0.518**	-0.506**	-0.480**	-0.491**	-0.462**	-0.458**	-0.450**	-0.443**
	(0.2178)	(0.2110)	(0.2172)	(0.2197)	(0.2203)	(0.2169)	(0.2125)	(0.2136)	(0.2087)
0.4	-0.489**	-0.498**	-0.485**	-0.466**	-0.469**	-0.460**	-0.439**	-0.434**	-0.436**
	(0.2137)	(0.2217)	(0.2171)	(0.2210)	(0.2225)	(0.2233)	(0.2187)	(0.2171)	(0.2107)
0.5	-0.478**	-0.473**	-0.470**	-0.464**	-0.460**	-0.444**	-0.439**	-0.424*	-0.409*
	(0.2102)	(0.2163)	(0.2219)	(0.2216)	(0.2158)	(0.2128)	(0.2202)	(0.2219)	(0.2074)
0.6	-0.455**	-0.470**	-0.457**	-0.457**	-0.450**	-0.430*	-0.424**	-0.427*	-0.409*
	(0.2128)	(0.2146)	(0.2180)	(0.2206)	(0.2144)	(0.2181)	(0.2093)	(0.2173)	(0.2114)
0.7	-0.445**	-0.437**	-0.444**	-0.445**	-0.440**	-0.432**	-0.421**	-0.415*	-0.413*
	(0.2165)	(0.2186)	(0.2094)	(0.2177)	(0.2168)	(0.2138)	(0.2071)	(0.2169)	(0.2163)
0.8	-0.441**	-0.451**	-0.440*	-0.442**	-0.425*	-0.423*	-0.414*	-0.414*	-0.399*
	(0.2120)	(0.2066)	(0.2241)	(0.2183)	(0.2209)	(0.2222)	(0.2168)	(0.2142)	(0.2154)
0.9	-0.434**	-0.436**	-0.445**	-0.448**	-0.433**	-0.417*	-0.418**	-0.414*	-0.397*
	(0.2064)	(0.2105)	(0.2171)	(0.2137)	(0.2172)	(0.2165)	(0.2062)	(0.2115)	(0.2091)

This table shows average direct effects of DLD on a full range of linguistic distance indexes at different combinations of δ_D and δ_I. GDP/capita is the dependent variable. All results include the full vector of controls. Full country sample. ILD_{CC} international linguistic distance index. SAR model. Inverse distance spatial weight matrix (row standardized). Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.11: DLD and economic development on a full range of δ_D and δ_I, subsample of poor countries, ILD_{CC} international linguistic distance index, SAR results with inversed distance spatial weight matrix (row standardized)

$\delta_I \backslash \delta_D$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	-0.681***	-0.700***	-0.713***	-0.723***	-0.711***	-0.708***	-0.710***	-0.692***	-0.689***
	(0.1961)	(0.2034)	(0.1969)	(0.2096)	(0.2084)	(0.2106)	(0.2151)	(0.2014)	(0.2081)
0.2	-0.689***	-0.715***	-0.717***	-0.730***	-0.729***	-0.730***	-0.719***	-0.700***	-0.696***
	(0.1958)	(0.2040)	(0.2033)	(0.1940)	(0.2091)	(0.2137)	(0.2147)	(0.2128)	(0.2122)
0.3	-0.706***	-0.719***	-0.726***	-0.737***	-0.733***	-0.726***	-0.719***	-0.721***	-0.692***
	(0.2022)	(0.1958)	(0.2070)	(0.2112)	(0.2096)	(0.2085)	(0.2147)	(0.2131)	(0.2145)
0.4	-0.685***	-0.700***	-0.727***	-0.737***	-0.725***	-0.730***	-0.724***	-0.723***	-0.700***
	(0.2055)	(0.2102)	(0.2078)	(0.2221)	(0.2116)	(0.2016)	(0.2171)	(0.2127)	(0.2127)
0.5	-0.697***	-0.707***	-0.716***	-0.723***	-0.718***	-0.715***	-0.722***	-0.729***	-0.700***
	(0.1965)	(0.2044)	(0.2091)	(0.2131)	(0.2131)	(0.2099)	(0.2209)	(0.2125)	(0.2085)
0.6	-0.657***	-0.689***	-0.709***	-0.714***	-0.723***	-0.713***	-0.717***	-0.715***	-0.704***
	(0.1928)	(0.2040)	(0.2036)	(0.2110)	(0.2053)	(0.2014)	(0.2059)	(0.2086)	(0.2109)
0.7	-0.646***	-0.676***	-0.700***	-0.700***	-0.710***	-0.721***	-0.713***	-0.700***	-0.697***
	(0.1948)	(0.1968)	(0.1982)	(0.2047)	(0.2027)	(0.2066)	(0.2071)	(0.2030)	(0.2101)
0.8	-0.635***	-0.675***	-0.691***	-0.695***	-0.703***	-0.716***	-0.707***	-0.707***	-0.698***
	(0.1888)	(0.1962)	(0.1931)	(0.2124)	(0.2056)	(0.2065)	(0.1956)	(0.2099)	(0.2046)
0.9	-0.641***	-0.666***	-0.679***	-0.689***	-0.708***	-0.715***	-0.714***	-0.692***	-0.695***
	(0.1953)	(0.1937)	(0.2003)	(0.2078)	(0.2056)	(0.1998)	(0.2020)	(0.2126)	(0.2110)

This table shows average direct effects of DLD on a full range of linguistic distance indexes at different combinations of δ_D and δ_I. GDP/capita is the dependent variable. All results include the full vector of controls. Subsample of poor countries. ILD_{CC} international linguistic distance index. SAR model. Inverse distance spatial weight matrix (row standardized). Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

tables display overall tendencies in terms of the behavior of direct effects both in their significance and magnitude according to the change in δ_I. First, as a whole, ILD_{PC} has a negative impact on GDP per capita, matching intuition: high English acquisition level allows global communication and linkages, bearing more chances and opportunities of mer-

Table 3.12: DLD and economic development on a full range of δ_D and δ_I, subsample of rich countries, ILD_{CC} international linguistic distance index, SAR results with inversed distance spatial weight matrix (row standardized)

$\delta_I \backslash \delta_D$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.428	0.499	0.526	0.532*	0.561*	0.548*	0.554*	0.576*	0.547*
	(0.3795)	(0.3427)	(0.3503)	(0.3169)	(0.3218)	(0.3088)	(0.2839)	(0.3122)	(0.2798)
0.2	0.380	0.416	0.497	0.496	0.516	0.515*	0.508*	0.533*	0.533*
	(0.3641)	(0.3394)	(0.3415)	(0.3261)	(0.3202)	(0.2996)	(0.2969)	(0.2759)	(0.2715)
0.3	0.282	0.338	0.416	0.440	0.480	0.455	0.499	0.497*	0.502*
	(0.3611)	(0.3570)	(0.3351)	(0.3097)	(0.3127)	(0.2980)	(0.3003)	(0.2850)	(0.2825)
0.4	0.198	0.276	0.363	0.381	0.405	0.429	0.446	0.440*	0.446
	(0.3884)	(0.3603)	(0.3736)	(0.3099)	(0.3261)	(0.3086)	(0.3084)	(0.2603)	(0.2706)
0.5	0.123	0.215	0.293	0.328	0.353	0.383	0.385	0.420	0.409
	(0.3559)	(0.3722)	(0.3444)	(0.3293)	(0.2968)	(0.3078)	(0.2834)	(0.2801)	(0.2783)
0.6	0.068	0.155	0.221	0.292	0.325	0.323	0.370	0.379	0.398
	(0.3602)	(0.3539)	(0.3290)	(0.3245)	(0.3145)	(0.2868)	(0.2927)	(0.2863)	(0.2787)
0.7	-0.016	0.105	0.180	0.221	0.288	0.306	0.341	0.351	0.369
	(0.3557)	(0.3527)	(0.3435)	(0.3134)	(0.3075)	(0.3148)	(0.3112)	(0.2679)	(0.2841)
0.8	-0.063	0.045	0.143	0.199	0.255	0.276	0.297	0.330	0.344
	(0.3545)	(0.3809)	(0.3376)	(0.3382)	(0.3291)	(0.2798)	(0.3004)	(0.2684)	(0.2962)
0.9	-0.095	-0.007	0.123	0.162	0.218	0.256	0.274	0.300	0.324
	(0.3963)	(0.3879)	(0.3497)	(0.3162)	(0.3206)	(0.3074)	(0.2741)	(0.2666)	(0.2642)

This table shows average direct effects of DLD on a full range of linguistic distance indexes at different combinations of δ_D and δ_I. GDP/capita is the dependent variable. All results include the full vector of controls. Subsample of rich countries. ILD_{CC} international linguistic distance index. SAR model. Inverse distance spatial weight matrix (row standardized). Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.13: ILD_{CC} and economic development on a full range of δ_D and δ_I, subsample of rich countries, ILD_{CC} international linguistic distance index, inverse distance spatial weight matrix (row standardized)

$\delta_I \backslash \delta_D$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	-0.150	-0.132	-0.127	-0.128	-0.125	-0.112	-0.118	-0.123	-0.119
	(0.1573)	(0.1525)	(0.1619)	(0.1552)	(0.1556)	(0.1501)	(0.1487)	(0.1633)	(0.1467)
0.2	-0.239	-0.233	-0.221	-0.206	-0.208	-0.203	-0.205	-0.198	-0.208
	(0.1798)	(0.1865)	(0.1962)	(0.1876)	(0.1808)	(0.1852)	(0.1713)	(0.1663)	(0.1748)
0.3	-0.352	-0.343	-0.327	-0.316	-0.325	-0.307	-0.313	-0.305	-0.299
	(0.2117)	(0.2114)	(0.2140)	(0.2015)	(0.2229)	(0.1988)	(0.2144)	(0.1971)	(0.2065)
0.4	-0.489*	-0.453*	-0.456	-0.429*	-0.419	-0.415	-0.413*	-0.410*	-0.401*
	(0.2522)	(0.2638)	(0.2835)	(0.2315)	(0.2580)	(0.2685)	(0.2419)	(0.2084)	(0.2297)
0.5	-0.576**	-0.559**	-0.526*	-0.507*	-0.524*	-0.501*	-0.498*	-0.507*	-0.505*
	(0.2873)	(0.2748)	(0.2651)	(0.2565)	(0.2653)	(0.2562)	(0.2562)	(0.2734)	(0.2742)
0.6	-0.669**	-0.644**	-0.639**	-0.597**	-0.597**	-0.588**	-0.593**	-0.569**	-0.569**
	(0.2770)	(0.2868)	(0.2854)	(0.2723)	(0.2740)	(0.2658)	(0.2678)	(0.2732)	(0.2636)
0.7	-0.764**	-0.726**	-0.712**	-0.665**	-0.644**	-0.663**	-0.672**	-0.635**	-0.648**
	(0.2877)	(0.2894)	(0.3348)	(0.2757)	(0.2722)	(0.2966)	(0.3334)	(0.2604)	(0.3164)
0.8	-0.861**	-0.806**	-0.759**	-0.730**	-0.724**	-0.666***	-0.693**	-0.679**	-0.711**
	(0.3241)	(0.3576)	(0.3185)	(0.3115)	(0.2970)	(0.2472)	(0.2691)	(0.2713)	(0.3328)
0.9	-0.947**	-0.904**	-0.816**	-0.797**	-0.776**	-0.753**	-0.769**	-0.719**	-0.715**
	(0.4283)	(0.4288)	(0.3381)	(0.3072)	(0.3535)	(0.3327)	(0.3319)	(0.2821)	(0.3267)

This table shows average direct effects of ILD_{CC} on a full range of linguistic distance indexes at different combinations of δ_D and δ_I. GDP/capita is the dependent variable. All results include the full vector of controls. Subsample of rich countries. ILD_{CC} international linguistic distance index. Inverse distance spatial weight matrix (row standardized). Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

chandise dealings, investment, education, and knowledge interaction. The direct effects' behavior shows incremental tendencies in significance and magnitude for all columns. With low values of δ_I, impacts of ILD are negative but insignificant. With high values of δ_I, by contrast, ILD's negative impact on GDP is significant and large in its magnitude. To interpret this tendency of ILD_{PC} direct effects with regard to a change in δ_I, recall what δ_I captures. If δ_I is set high, linguistic distance between English and another language belonging to the Indo-European language family is larger (closer to 1) but that between English and languages outside the Indo-European language family is still 1. This means that only countries whose residents' mother tongue is English have smaller ILD_{PC}, and all other countries (including those whose residents' mother tongue is not a member of the Indo-European language family) have larger ILD_{PC}. If δ_I is set low, on the other hand, linguistic distance between English and another language belonging to the Indo-European language family is smaller (closer to 0) and languages outside the Indo-European language family is still 1. This implies that all countries whose residents' mother tongue belongs to the Indo-European language family have small ILD_{PC}, and all other countries whose residents' mother tongue is not a member of the Indo-European language family have larger (nearly equal to 1) ILD_{PC} values. If easiness of acquiring English for individuals whose mother tongue is an Indo-European language compared to those with non-Indo-European mother tongue were more valuable in explaining GDP differences, lower δ_I should result in ILD that has higher explanatory power. On the contrary, if the benefits enjoyed by individuals whose mother tongue is English, who do not need to devote effort to mastering English as a second language, were more important in explaining them, higher δ_I should lead to ILD of higher explanatory power. With these in mind, consider the tendency shown in Tables 3.9 and 3.13. The results that ILD with larger δ_I is more significant and has stronger impact on GDP per capita mean that the capability of using English as the first language is advantageous.[34]

In these results, it seems that ability to use English as the first language is important, while the benefit of costless learning of English enjoyed by individuals whose mother tongue is Spanish is less important. However, some readers may not be comfortable with these results. They might consider that speakers of Indo-European languages as the first language are much easier to master English as a second language than speakers of non-Indo-European languages. Thus, they deem that the present empirical results showing that access to English for non-English Indo-European speakers has smaller impacts on GDP per capita are due to the impacts of Spanish speaking countries (such as Latin American countries) which are less economically developed. Although this story of the effect of "Latin American countries" seems at the fisrt glance reasonable, it would not be

[34]By checking all log likelihood values reported for each estimation (i.e, 81 log likelihood values), and searching for the pair of (δ_D, δ_I) returning the maximum log likelihood value, we can make discussions on which pair of (δ_D, δ_I) would be the most appropriate one to some extent. (δ_D, δ_I) returning the maximum log likelihood value must be the most appropriate one because the employed estimation method is maximum likelihood estimation. For the full sample and poor subsample cases, we could not find a clear tendency for which pair of (δ_D, δ_I) returns the largest log likelihood values (i.e., we could not find a tendency for the appropriate (δ_D, δ_I) which is common to all the specifications of ILD_{PC}, ILD_{CC}, inverse distance spatial weight matrix, and contiguity spatial weight matrix). For the rich country subsample, on the other hand, it can be safely asserted that the most appropriate pair of (δ_D, δ_I) is $(0.9, 0.9)$.

when thought twice. Because we have already included several aspects of control variables such as political stability, which are thought to be one of the elements bringing about less GDP levels, the effect of "Latin American countries" is not through those channels. Instead, the present results that only English speakers as the first language are highly advantageous and other Indo-European language speakers as the first language are not greatly advantageous come from the construction of linguistic distance, rather than the estimation procedures. When constructing linguistic distances, we have assumed symmetric pairwise distance ($\tau(i,j) = \tau(j,i)$), which would be one of the reasons of the effect of "Latin American countries." In reality, τ(Engish, Spanish) $\neq \tau$(Spanish, English), representing that English would be more powerful than Spanish. The origin of this effect of "Latin American countries" may lie around the assumptions employed under linguistic distance data counstruction.

Finally, we quickly look at Tables 3.6 and 3.10. Domestic linguistic distance still provides negative and significant effects on GDP per capita. Further, even when poor and rich subsamples are merged into a full sample, the negative impact of domestic linguistic distance for poor countries is stronger than the positive impacts of domestic distance on economic activity for rich countries.

3.5 Robustness

We present robustness exercises that check whether the direct effect matrices of linguistic distance indexes show similar tendency and properties to the results provided in Section 3.4.3. Especially, robustness checks focus on whether (i) direct effects of DLD are significantly negative for the sample of full countries, (ii) those of DLD are significantly negative for the subsample of poor countries, (iii) those of DLD are significantly positive for the subsample of rich countries, (iv) those of ILD are significantly negative for the subsample of rich countries, and (v) those of ILD react incrementally to a change in δ_I in terms of their significance and magnitude for the subsample of rich countries.

3.5.1 Robustness in terms of the spatial weight matrix

First, we examine whether the results are sensitive to the choice of the spatial weight matrix. For the first robustness check related to the spatial weight matrix, we use spectral standardization for the inverse distance matrix. This is because row standardization does not use a single normalization factor but a different factor for the elements of each row (Kelejian and Prucha, 2010). This makes it difficult to interpret when the spatial weight matrix is based on distance between observations. The second robustness check for the spatial weight matrix is to employ a contiguity matrix whose elements takes 1 if a pair of countries shares a national border and otherwise zero. The contiguity matrix is row standardized per convention.

Almost all robustness checks related to the choices of spatial weight matrices show preferred results, and five features (i)-(v) are certified. Only in the case of (iii) under the choice of row standardized contiguity spatial weight matrix and ILD_{CC}, our result exhibit slightly weaker robustness. For more details, see Appendix 3.B.

3.5.2 Robustness for other features than the spatial weight matrix

For items other than those related to the spatial weight matrix, we first check for robustness of multi- or single-national-language countries, and simply drop samples that have several national languages. This is because language use in countries with multiple national languages may not show regionally even distribution (that is, regionally segregated) as observed in Canada (Laponce, 1984; Pons-Ridler and Ridler, 1989), Switzerland (Pap, 1990), and so on. In such cases, there are a couple of domestic linguistic centers (for instance, in Canada, the French-speaking areas' domestic linguistic center is French, and that of English-speaking regions is English, and the domestic linguistic distance is not the mid-point of French and English). Thus, we need to exclude samples with ambiguous linguistic centers.

Second, we reconstruct the linguistic distance indexes including the 'immigrants' language and population reported in Ethnologue, because in the main results shown in Section 3.4, linguistic distance indexes are calculated only based on indigenous population for the purpose of harmonizing construction of the linguistic data across countries.[35][36] Note that immigrants reported in Ethnologue refer to a group of people who have not stayed in the given country long enough to be well-established, and these groups are sometimes refugees or transient workers.

Although some of the regressions conducted only show weak robustness related to the above-mentioned items, the results displayed in Appendix 3.B are consistent with those in Section 3.4.3 as a whole. For more details, see Appendix 3.B.

3.6 Conclusion

In this chapter, we have investigated the impacts of domestic and international linguistic distances on the cross-country income difference. First, we constructed pair-wise linguistic distances for about all living languages in the world by using linguistic family trees. Then, we constructed two types of linguistic distance indexes—domestic and international. A domestic linguistic distance index is calculated as a population-weighted average of linguistic

[35]For countries without detailed information of immigrants' languages but with information only on their countries of origin in Ethnologue, national languages of those countries of origin represent their mother tongues.

[36]Ethnologue sometimes reports immigrants' language population along with the residents of a country, and sometimes does not.

distances between mother tongues of residents in a country and the national language. International linguistic distance indexes are calculated in two ways: (i) population-weighted average of linguistic distances between mother tongues of residents in a country and English, and (ii) linguistic distance between national languages and English.

The effects of domestic and international linguistic distance indexes on the output-based GDP per capita are different for rich and poor countries. For poor countries, domestic linguistic distance has a negative impact on economic development, while rich countries' economic output tend to be affected by domestic linguistic distance, owing to the positive impact of diversity, which is hidden behind the cost of domestic linguistic distance. Because many African countries are likely to have larger domestic linguistic distance, Africa's growth tragedy can be partly explained by their worse accessibility to the domestic linguistic center, causing harder nationwide communications. As for international linguistic distance, rich countries enjoy benefits if international linguistic distance is smaller, while poor countries do not. This implies that richer countries can improve their economic prosperity levels with easier access to international linguistic center (global communication without difficulty).

Our results are distinct from those in previous literature, where the negative feature of between-group communication is indirectly considered. We directly focus on the cost of communication between different language groups. As a whole, the negative impact on within-country linguistic distance is certified. For between-country communication, only advanced countries have advantages in accessibility to English usage. Finally, we find incremental impacts on economic productivity by international linguistic distance with a change in values of international linguistic parameters, which capture sensitivity of relative linguistic distance between English and the other Indo-European language family members. We interpret this result as follows: capability of using English as the first language is highly advantageous.

Appendix 3.A Data sources

Table 3.14: Data sources

Variable name	Definition	Source
Income		
GDP/capita	log of GDP/capita in year 2011 (Output-side real GDP at current PPPs (in mil. 2005US$) is chosen)	Penn World Tables 8.0, Feenstra, Inklaar, and Timmer (2013b)
Linguistic distance		
Domestic linguistic distance DLD	Weighted average of linguistic distances to the domestic linguistic center as defined in the main text	Own calculation from Ethnologue 16th edition (Lewis, 2009)
International linguistic distance ILD_{PC}	Weighted average of linguistic distances to the international linguistic center as defined in the main text	Own calculation from Ethnologue 16th edition (Lewis, 2009)
International linguistic distance ILD_{CC}	Weighted average of linguistic distances from the domestic linguistic center(s) to the international linguistic center as defined in the main text	Own calculation from Ethnologue 16th edition (Lewis, 2009)
Language status	Status labeled to each language based on its intra- and international usages and importance	Ethnologue 17th edition (Lewis et al., 2014)
Market size		
Population size	Population size in logs in year 2010	Penn World Tables 8.0, Feenstra, Inklaar, and Timmer (2013b)
Land area size	Country land area size in km^2 in logs	CEPII (2010), Head, Mayer, and Ries (2010)
Landlockedness	Dummy takes 1 if country is landlocked	CEPII (2010), Head, Mayer, and Ries (2010)
Education		
Years of schooling	Years of schooling, population aged over 25 in year 2010	Barro and Lee (2013)
Trade		
Trade openness	Merchandise exports + imports in % of GDP, at PPP in year 2010	Penn World Tables 8.0, Feenstra, Inklaar, and Timmer (2013b)
Geography		
Absolute latitude	Absolute latitude of capital	CEPII (2010), Head, Mayer, and Ries (2010)
Coastal population	Ratio of population within 100 km of ice-free coast to total population in % in year 1995	Gallup, Sachs, and Mellinger (1999)
Mean temperature	Average temperature in celsius in logs, in years 1961-1990	Michalopoulos (2012)
Mean precipitation	Average precipitation/month , in years 1961-1990, in logs	Michalopoulos (2012)
Continent dummy	Sub-Saharan Africa, Latin America, South-East Asia	World Bank (2014)
Agriculture		
Agricultural suitability	Land quality (average of agricultural suitability across regions within a country)	Michalopoulos (2012)
Agricultural suitability	Land quality (standard deviation of agricultural suitability across regions within a country)	Michalopoulos (2012)
Institutions		
Quality of institutions	Combined Polity2 score in year 2010 (-10 for most repressive, 10 for most democratic)	PolityIV database, Marshall and Jaggers (2012)
Property rights index	Inverted index (1 for least rights, 7 for most rights) in year 2010	Freedom House (2014)
Civil liberality index	Inverted index (1 for least liberal, 7 for most liberal) in year 2010	Freedom House (2014)
Legal origin	Common law (Dummy takes 1 if country's legal origin is British law), socialist law (dummy takes 1 if socialist law), civil law (dummy takes 1 if French, German, or Scandinavian law)	La Porta, Lopez-de Silanes, Shleifer, and Vishny (1999)
Spatial weight matrix		
Inverse distance matrix	Inversed distance (calculated following the great circle formula) between capital cities	CEPII (2010), Head, Mayer, and Ries (2010)
Contiguity matrix	Dummy takes 1 if country shares national border	CEPII (2010), Head, Mayer, and Ries (2010)

Appendix 3.B Details for robustness checks

Spectral normalized inverse distance spatial weight matrix (international linguistic distance as ILD_{PC})

(i) Direct effects of DLD are significantly negative for the full sample of countries. 70 out of 81 cells show negative values at the 1% significance level and 11 cells are significantly negative at the 5% level. (ii) Direct effects of DLD are significantly negative for the subsample of poor countries. 66 out of 81 cells (15 out of 81 cells, respectively) are significant at the 5% (10%, respectively) level. (iii) Direct effects of DLD are significantly positive for the subsample of rich countries. 20 out of 81 cells exhibit positive values at the 1% significance level. 52 cells at the 5% level and 9 cells at the 1% level are significantly positive. (iv) Direct effects of ILD tend to be significantly negative for the subsample of rich countries. 16, 31, and 4 out of 81 cells are significant at the 1%, 5%, and 10% levels, respectively. (v) Direct effects of ILD for the subsample of rich countries show incremental tendency in terms of the magnitudes and significance for larger δ_Is.[37]

Spectral normalized inverse distance spatial weight matrix (international linguistic distance as ILD_{CC})

(i) Direct effects of DLD are significantly negative for the full sample of countries. 67 out of 81 cells show negative values at the 1% significance level and 14 cells are significantly negative at the 5% level. (ii) Direct effects of DLD are significantly negative for the subsample of poor countries. All 81 cells are significant at the 5% level. (iii) Direct effects of DLD tend to be significantly positive for the subsample of rich countries. 46 out of 81 cells are significant at least at the 10% level. (iv) Direct effects of ILD tend to be significantly negative for the subsample of rich countries. 49 out of 81 cells are significant at least at the 10% level. (v) Direct effects of ILD for the subsample of rich countries show incremental tendency in terms of the magnitudes and significance for higher δ_Is.[38]

Row standardized contiguity spatial weight matrix (international linguistic distance as ILD_{PC})

(i) Direct effects of DLD are significantly negative for the full sample of countries. All 81 cells show negative values at the 1% significance level. (ii) Direct effects of DLD are significantly negative for the subsample of poor countries. All 81 cells are significant at the 1% level. (iii) Direct effects of DLD exhibit positive signs and are significant for the subsample of rich countries. 80 out of 81 cell at least at the 10% level and 66 cells at least at the 5% level are significant. (iv) Direct effects of ILD tend to be significantly negative for the subsample of rich countries. 54 (28, respectively) out of 81 cells are significant at

[37] More details are provided for readers upon request.

[38] More details are provided for readers upon request.

least at the 10% (1%, respectively) level. (v) Direct effects of ILD for the subsample of rich countries show incremental tendency in terms of the magnitudes and significance for higher δ_Is.[39]

Row standardized contiguity spatial weight matrix (international linguistic distance as ILD_{CC})

(i) Direct effects of DLD are significantly negative for the full sample of countries. All 81 cells show negative significance at least at the 10% level, of which 76 cells are significant at the 5% level. (ii) Direct effects of DLD are significantly negative for the subsample of poor countries. All 81 cells are significant at the 1% level. (iii) Direct effects of DLD exhibit slightly weaker significance but still tend to be positive for the subsample of rich countries. 80 out of 81 cells show positive signs, of which 28 cells are significant at least at the 10% level. (iv) Direct effects of ILD tend to be significantly negative for the subsample of rich countries. 47 (39, respectively) out of 81 cells are significant at least at the 10% (5%, respectively) level. (v) Direct effects of ILD for the subsample of rich countries show incremental tendency in terms of the magnitudes and significance for higher δ_Is.[40]

Results limited to single national language countries (international linguistic distance as ILD_{PC})

(i) Direct effects of DLD exhibit slightly weaker significance but still show negative signs for the full country sample. 55 out of 81 cells are significant at the 10% level. (ii) Direct effects of DLD are significantly negative for the subsample of poor countries. All 81 cells are significant at the 10% level, of which 34 cells are significant at the 5% level. (iii) Direct effects of DLD exhibit slightly stronger significance and all cells are positive for the subsample of rich countries. 73 out of 81 cells are significant at least at the 10% level, and of those, 52 cells are significant at least at the 1% level. (iv) Direct effects of ILD do not show significance for most of the cells but still tend to be negative for the subsample of rich countries. 72 out of 81 cells show negative signs. (v) Direct effects of ILD for the subsample of rich countries show incremental tendency towards negative values.[41]

Results limited to single national language countries (international linguistic distance as ILD_{CC})

(i) Direct effects of DLD still exhibit negative signs in all cells for the full country sample, although they are insignificant. (ii) Direct effects of DLD still exhibit negative signs in all cells for the subsample of poor countries, although most of them are insignificant (iii) Direct effects of DLD exhibit stronger significance and all cells are positive for the

[39] More details are provided for readers upon request.
[40] More details are provided for readers upon request.
[41] More details are provided for readers upon request.

subsample of rich countries. 72 out of 81 cells are significant at least at the 10% level, of which 48 cells are significant at least at the 1% level. (iv) Direct effects of ILD do not show significance but still may tend to be negative for the subsample of rich countries. 44 out of 81 cells show negative signs. (v) Direct effects of ILD for the subsample of rich countries go towards negative values and the magnitudes tend to be larger.[42]

Linguistic distance index including "immigrant" languages (international linguistic distance as ILD_{PC})

(i) Direct effects of DLD are significantly negative for the full sample of countries. All 81 cells show negative significance at the 5% level. (ii) Direct effects of DLD are significantly negative for the subsample of poor countries. All 81 cells are significant at the 1% level. (iii) Direct effects of DLD still exhibit positive signs in all cells for the subsample of rich countries, although they are insignificant. This may be because the "immigrants" reported in Ethnologue may have difficulty in contributing to the improvement of productivity of highly skilled industry. (iv) Direct effects of ILD tend to be significantly negative for the subsample of rich countries. 50 (43, respectively) out of 81 cells are significant at least at the 10% (5%, respectively) level. (v) Direct effects of ILD for the subsample of rich countries show incremental tendency in terms of the magnitudes and significance for higher δ_Is.[43]

Linguistic distance index including "immigrant" languages (international linguistic distance as ILD_{CC})

(i) Direct effects of DLD are significantly negative for the full sample of countries. All 81 cells show negative significance at the 5% level. (ii) Direct effects of DLD are significantly negative for the subsample of poor countries. All 81 cells are significant at the 1% level. (iii) Direct effects of DLD still exhibit positive signs in almost all cells (77 out of 81 cells) for the subsample of rich countries, although they are insignificant. This may be due to the same reason as in Section **??**. (iv) Direct effects of ILD tend to be significantly negative for the subsample of rich countries. 61 (46, respectively) out of 81 cells are significant at least at the 10% (5%, respectively) level. (v) Direct effects of ILD for the subsample of rich countries show incremental tendency in terms of the magnitudes and significance for larger δ_Is.[44]

[42] More details are provided for readers upon request.
[43] More details are provided for readers upon request.
[44] More details are provided for readers upon request.

General Conclusion

The overall concept running through this dissertation has been ethnicity. The aim of this dissertation has been that ethnicity/language and their related topics had been investigated from economic aspects. Chapter 1 analyzed residential segregation according to ethnic characteristics in cities. In the model, the majority faces a trade-off between commuting costs and residential congestion. The minority group, on the other hand, faces a trade-off between commuting costs, ethnic clustering, and residential congestion. The findings in Chapter 1 showed that, due to ethnicity preferences of the minority group, minority residents are more likely to migrate to one area within a city. In addition, minority households always cluster when the commuting cost is low, widening the population gap between the areas, while majority households migrate to the less populated area to avoid the residential congestion caused by minority residential clustering, thus reducing the population gap between areas.

Similarly, in Chapter 2, we have investigated how regional segregation patterns are affected by industrial agglomeration and ethnolinguistic clustering preference. In the model used in Chapter 2, we showed that segregation by ethnicity is persistent, while ethnically mixed distributions appear only when the trade cost is intermediate. This theoretical results are consistent with the real-world examples of regional segregation by language use. Both chapters have considered the impacts of benefits borne by residential clustering of the same ethnic groups, which has been expressed by ethnic externality terms.

On the other hand, in Chapter 3, how accessibility to domestic and international communication affects the national economic development has been investigated, where communication difficulty among speakers of different mother tongues are measured by linguistic distance indexes. The empirical findings exhibited that the effects of domestic and international linguistic distance indexes on the output-based GDP per capita are different for rich and poor countries. For poor countries, difficulty in domestic linguistic communication has a negative impact on economic development, while rich countries' economic output tend to benefit from easier communication in English. Chapter 3 basically considered the cost owing to communication constraints with different linguistic groups.

In a society which consists of several ethnolinguistic communities, investigating internal interactions in an ethnolinguistic group as well as external relationships among different groups is important. Generally, if there are several groups, possibility of looking inside and outside each of them necessarily emerges. In Chapters 1 and 2, our focus was on the benefits within ethno-linguistic communities, while Chapter 3 shed light on the costs between them. This dissertation has dealt with the twofold characteristics associated with ethnolinguistically heterogeneous economy—intra- and inter-group interactions.

Bibliography

ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369–1401.

ADSERA, A. AND M. PYTLIKOVA (2012): "The Role of Language in Shaping International Migration," *IZA Working Paper No. 6333*.

ALCOCK, A. E. (1970): *History of the South Tyrol Question*, London: Michael Joseph.

ALESINA, A., R. BAQIR, AND W. EASTERLY (1999): "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 114, 1243–1284.

ALESINA, A., A. DEVLEESCHAUWER, W. EASTERLY, S. KURLAT, AND R. WACZIARG (2003): "Fractionalization," *Journal of Economic Growth*, 8, 155–194.

ALESINA, A., W. EASTERLY, AND J. MATUSZESKI (2011): "Artificial States," *Journal of the European Economic Association*, 9, 246–277.

ALESINA, A. AND E. LA FERRARA (2005): "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43, 762–800.

ALESINA, A. AND E. ZHURAVSKAYA (2011): "Segregation and the Quality of Government in a Cross Section of Countries," *American Economic Review*, 101, 1872–1911.

ALESINA, A. F., J. HARNOSS, AND H. RAPOPORT (2013): "Birthplace Diversity and Economic Prosperity," *NBER Working Paper No. 18699*.

ALONSO, W. (1964): *Location and Land Use*, Cambridge, MA: Harvard University Press.

ANANAT, E. O. (2011): "The Wrong Side (s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality," *American Economic Journal: Applied Economics*, 3, 34–66.

ATTFIELD, C., E. S. CANNON, D. DEMERY, AND N. W. DUCK (2000): "Economic Growth and Geographic Proximity," *Economics Letters*, 68, 109–112.

Baldwin, K. and J. D. Huber (2010): "Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision," *American Political Science Review*, 104, 644–662.

Barro, R. J. and J. W. Lee (2013): "A New Data Set of Educational Attainment in the World, 1950–2010," *Journal of Development Economics*, 104, 184–198.

Bayer, P., H. Fang, and R. McMillan (2014): "Separate When Equal? Racial Inequality and Residential Segregation," *Journal of Urban Economics*, 82, 32–48.

Bellini, E., G. I. Ottaviano, D. Pinelli, and G. Prarolo (2013): "Cultural Diversity and Economic Performance: Evidence from European Regions," in *Geography, Institutions and Regional Economic Performance*, Berlin: Springer, 121–141.

Belot, M. and S. Ederveen (2012): "Cultural Barriers in Migration between OECD Countries," *Journal of Population Economics*, 25, 1077–1105.

Bénabou, R. (1993): "Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics*, 108, 619–652.

——— (1996): "Equity and Efficiency in Human Capital Investment: The Local Connection," *Review of Economic Studies*, 63, 237–264.

Berliant, M. and M. Fujita (2012): "Culture and Diversity in Knowledge Creation," *Regional Science and Urban Economics*, 42, 648–662.

Besley, T., R. Pande, L. Rahman, and V. Rao (2004): "The Politics of Public Good Provision: Evidence from Indian Local Governments," *Journal of the European Economic Association*, 2, 416–426.

Billings, S. B., D. J. Deming, and J. Rockoff (2014): "School Segregation, Educational Attainment, and Crime: Evidence from the End of Busing in Charlotte-Mecklenburg," *Quarterly Journal of Economics*, 129, 435–476.

Bourhis, R. Y., ed. (1984): *Conflict and Language Planning in Quebec*, Clevedon: Multilingual Matters.

Boustan, L. P. (2007): "Black Migration, White Flight: The Effect of Black Migration on Northern Cities and Labor Markets," *Journal of Economic History*, 67, 484–488.

Brueckner, J. K. and R. W. Martin (1997): "Spatial Mismatch: An Equilibrium Analysis," *Regional Science and Urban Economics*, 27, 693–714.

Brueckner, J. K. and Y. Zenou (2003): "Space and Unemployment: The Labor-Market Effects of Spatial Mismatch," *Journal of Labor Economics*, 21, 242–262.

Bucovetsky, S. and A. Glazer (2014): "Efficiency, Equilibrium and Exclusion When the Poor Chase the Rich," *Journal of Urban Economics*, 81, 166–177.

Calomiris, C. W., C. M. Kahn, and S. D. Longhofer (1994): "Housing-Finance Intervention and Private Incentives: Helping Minorities and the Poor," *Journal of Money, Credit and Banking*, 26, 634–674.

Clark, W. A. and S. A. Blue (2004): "Race, Class, and Segregation Patterns in U.S. Immigrant Gateway Cities," *Urban Affairs Review*, 39, 667–688.

Combes, P.-P., T. Mayer, and J.-F. Thisse (2008): *Economic Geography: The Integration of Regions and Nations*, Princeton, NJ: Princeton University Press.

Coulton, C. J., J. Chow, E. C. Wang, and M. Su (1996): "Geographic Concentration of Affluence and Poverty in 100 Metropolitan Areas, 1990," *Urban Affairs Review*, 32, 186–216.

Courant, P. N. and J. Yinger (1977): "On Models of Racial Prejudice and Urban Residential Structure," *Journal of Urban Economics*, 4, 272–291.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999): "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107, 455–506.

——— (2008): "When Are Ghettos Bad? Lessons from Immigrant Segregation in the United States," *Journal of Urban Economics*, 63, 759–774.

Desmet, K., I. Ortuño Ortín, and R. Wacziarg (2012): "The Political Economy of Linguistic Cleavages," *Journal of Development Economics*, 97, 322–338.

Desmet, K., I. Ortuño Ortín, and S. Weber (2005): "Peripheral Diversity and Redistribution," *CEPR Discussion Paper No. 5112*.

Desmet, K., S. Weber, and I. Ortuño-Ortín (2009): "Linguistic Diversity and Redistribution," *Journal of the European Economic Association*, 7, 1291–1318.

Dietz, B. (1999): "Ethnic German Immigration from Eastern Europe and the Former Soviet Union to Germany: The Effects of Migrant Networks," *IZA Discussion Paper No. 68*.

Dixit, A. K. and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, 297–308.

Easterly, W. and R. Levine (1997): "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 112, 1203–1250.

Elhorst, J. P. (2014): *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*, Berlin: Springer.

Fearon, J. D. (2003): "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 8, 195–222.

Fearon, J. D. and D. D. Laitin (2003): "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 97, 75–90.

Feenstra, R. C., R. Inklaar, and M. Timmer (2013a): *PWT 8.0–A User Guide*, Available for download at www.rug.nl/research/ggdc/data/penn-world-table.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2013b): *The Next Generation of the Penn World Table*, Available for download at www.ggdc.net/pwt.

Frankel, J. A. and D. Romer (1999): "Does Trade Cause Growth?" *American Economic Review*, 89, 379–399.

Fujita, M. (1989): *Urban Economic Theory: Land Use and City Size*, Cambridge: Cambridge University Press.

Fujita, M., P. Krugman, and A. J. Venables (1999): *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge, MA: MIT Press.

Fujita, M. and J.-F. Thisse (2013): *Economics of Agglomeration: Cities, Industrial Location, and Globalization*, Cambridge: Cambridge University Press, 2 ed.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1999): "Geography and Economic Development," *International Regional Science Review*, 22, 179–232.

Glaeser, E. L. (2005): "The Political Economy of Hatred," *Quarterly Journal of Economics*, 120, 45–86.

Glaeser, E. L., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2004): "Do Institutions Cause Growth?" *Journal of Economic Growth*, 9, 271–303.

Hall, R. E. and C. Jones (1999): "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 114, 83–116.

Hattori, M. (2000): "Berarushi-ni-okeru-Kokumin-Ishiki-no-Konton: Tai-Roshia-Tougou-no-Dojou-wo-Saguru (Chaos of the Awareness of Nationalism in Belarus: Searching for the Possibilities of Integration with Russia) (in Japanese)," *Gaimu-shou-Chousa-Geppou (Monthly Report by the Ministry of Foreign Affairs of Japan)*, 4, 43–76.

Head, K., T. Mayer, and J. Ries (2010): "The Erosion of Colonial Trade Linkages After Independence," *Journal of International Economics*, 81, 1–14.

Helpman, E. (1998): "The Size of Regions," in *Topics in Public Economics: Theoretical and Applied Analysis*, Cambridge: Cambridge University Press, 33–54.

Ihlanfeldt, K. R. and D. L. Sjoquist (1990): "Job Accessibility and Racial Differences in Youth Employment Rates," *American Economic Review*, 80, 267–276.

——— (1991): "The Effect of Job Access on Black and White Youth Employment: A Cross-Sectional Analysis," *Urban Studies*, 28, 255–265.

Inoue, H., K. Nakajima, and Y. Saito (2015): "Innovation and Collaboration Patterns between Research Establishments," *RIETI Discussion Paper Series 15-E-049.*

Jacobs, J. (1961): *The Life and Death of Great American Cities*, New York, NY: Random House.

Kain, J. F. (1968): "Housing Segregation, Negro Employment, and Metropolitan Decentralization," *Quarterly Journal of Economics*, 82, 175–197.

Kanemoto, Y. (1980): *Theories of Urban Externalities*, Amsterdam: North-Holland.

Kelejian, H. H., P. Murrell, and O. Shepotylo (2013): "Spatial Spillovers in the Development of Institutions," *Journal of Development Economics*, 101, 297–315.

Kelejian, H. H. and I. R. Prucha (2010): "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," *Journal of Econometrics*, 157, 53–67.

Kenworthy, J. R. and F. B. Laube (1999): "Patterns of Automobile Dependence in Cities: An International Overview of Key Physical and Economic Dimensions with Some Implications for Urban Policy," *Transportation Research Part A: Policy and Practice*, 33, 691–723.

Krugman, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

La Porta, R., F. Lopez-de Silanes, A. Shleifer, and R. Vishny (1999): "The Quality of Government," *Journal of Law, Economics, and Organization*, 15, 222–279.

Laitin, D. D. (2000): "What Is a Language Community?" *American Journal of Political Science*, 44, 142–155.

Laponce, J. (1984): "The French Language in Canada: Tensions between Geography and Politics," *Political Geography Quarterly*, 3, 91–104.

Lazear, E. P. (1999a): "Culture and Language," *Journal of Political Economy*, 107, 95–126.

——— (1999b): "Globalisation and the Market for Team-Mates," *Economic Journal*, 109, 15–40.

Le Sage, J. and R. K. Pace (2009): *Introduction to Spatial Econometrics*, Boca Raton, FL: Chapman &Hall/CRC.

Le Sage, J. P. and M. M. Fischer (2008): "Spatial Growth Regressions: Model Specification, Estimation and Interpretation," *Spatial Economic Analysis*, 3, 275–304.

Lewis, M. P., ed. (2009): *Ethnologue: Languages of the World, 16th edition*, Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Lewis, M. P., G. F. Simons, and C. D. Fennig, eds. (2014): *Ethnologue: Languages of the World, 17th edition*, Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Liu, C. Y. (2009): "Ethnic Enclave Residence, Employment, and Commuting of Latino Workers," *Journal of Policy Analysis and Management*, 28, 600–625.

López-Bazo, E., E. Vayá, and M. Artís (2004): "Regional Externalities and Growth: Evidence from European Regions," *Journal of Regional Science*, 44, 43–73.

Marino, M., P. Parrotta, and D. Pozzoli (2012): "Does Labor Diversity Promote Entrepreneurship?" *Economics Letters*, 116, 15–19.

Marshall, M. G. and K. Jaggers (2012): "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2012," .

Maurseth, P. B. (2003): "Geography and Growth: Some Empirical Evidence," *Nordic Journal of Political Economy*, 29, 25–46.

Melitz, M. J. and G. I. Ottaviano (2008): "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75, 295–316.

Michalopoulos, S. (2012): "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 102, 1508–1539.

Moreno, R. and B. Trehan (1997): "Location and the Growth of Nations," *Journal of Economic Growth*, 2, 399–418.

Mori, A. (1873): *Education in Japan: A Series of Letters Addressed by Prominent Americans to Arinori Mori*, New York, NY: D. Appleton.

Mossay, P. and P. M. Picard (2013): "Spatial Segregation and Urban Structure," *Discussion Papers 13056, Research Institute of Economy, Trade and Industry (RIETI).*

Munshi, K. (2003): "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics*, 118, 549–599.

Musterd, S. (2005): "Social and Ethnic Segregation in Europe: Levels, Causes, and Effects," *Journal of urban affairs*, 27, 331–348.

Musterd, S. and M. De Winter (1998): "Conditions for Spatial Segregation: Some European Perspectives," *International Journal of Urban and Regional Research*, 22, 665–673.

Nathan, M. (2015): "Same Difference? Minority Ethnic Inventors, Diversity and Innovation in the UK," *Journal of Economic Geography*, 15, 129–168.

O'Loughlin, J. (1980): "Distribution and Migration of Foreigners in German Cities," *Geographical Review*, 70, 253–275.

Østergaard, C. R., B. Timmermans, and K. Kristinsson (2011): "Does a Different View Create Something New? The Effect of Employee Diversity on Innovation," *Research Policy*, 40, 500–509.

Ottaviano, G., T. Tabuchi, and J.-F. Thisse (2002): "Agglomeration and Trade Revisited," *International Economic Review*, 43, 409–436.

Ottaviano, G. I. and G. Peri (2006): "The Economic Value of Cultural Diversity: Evidence from US Cities," *Journal of Economic Geography*, 6, 9–44.

Pap, L. (1990): "The Language Situation in Switzerland: An Updated Survey," *Lingua*, 80, 109–148.

Picard, P. M. and T. Tabuchi (2010): "Self-organized Agglomerations and Transport Costs," *Economic Theory*, 42, 565–589.

——— (2013): "On Microfoundations of the City," *Journal of Economic Theory*, 148, 2561–2582.

Picard, P. M. and D.-Z. Zeng (2005): "Agricultural Sector and Industrial Agglomeration," *Journal of Development Economics*, 77, 75–106.

Polinsky, A. M. and S. Shavell (2007): *Handbook of Law and Economics*, vol. 1, Elsevier.

Pons-Ridler, S. and N. B. Ridler (1989): "The Territorial Concept of Official Bilingualism; A Cheaper Alternative for Canada?" *Language Sciences*, 11, 147–158.

Puga, D. (1999): "The Rise and Fall of Regional Inequalities," *European Economic Review*, 43, 303–334.

Putterman, L. and D. N. Weil (2010): "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 125, 1627–1682.

Reardon, S. F., L. Fox, and J. Townsend (2015): "Neighborhood Income Composition by Race and Income, 1990-2009," *Annals of the American Academy of Political and Social Science Direct*, 660, 78–97.

Rodriguez, F. and D. Rodrik (2001): "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence," *NBER Macroeconomics Annual 2000, Volume 15*, 261–338.

Rodrik, D., A. Subramanian, and F. Trebbi (2004): "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development," *Journal of Economic Growth*, 9, 131–165.

Romero, A. A. and M. L. Burkey (2011): "Debt Overhang in the Eurozone: A Spatial Panel Analysis," *Review of Regional Studies*, 41, 49–63.

Rose-Ackerman, S. (1975): "Racism and Urban Structure," *Journal of Urban Economics*, 2, 85–103.

Sachs, J. D. (2003): "Institutions Don't Rule: Direct Effects of Geography on per Capita Income," *NBER Working Paper No. 9490.*

Seldadyo, H., J. P. Elhorst, and J. De Haan (2010): "Geography and Governance: Does Space Matter?" *Papers in Regional Science*, 89, 625–640.

Seya, H., M. Tsutsumi, and Y. Yamagata (2012): "Income Convergence in Japan: A Bayesian Spatial Durbin Model Approach," *Economic Modelling*, 29, 60–71.

Søholt, S. (2001): "Ethnic Minority Groups and Strategies in the Housing Market in Oslo," *European Journal of Housing Policy*, 1, 337–355.

Soundararajan, P. (2013): "Regional Income Convergence in India: A Bayesian Spatial Durbin Model Approach," *MPRA Paper No. 44744.*

St-Hilaire, A. (1997): "North America and the Francophonie: Local and Transnational Movements for the Survival of French-speaking North America," *Language Sciences*, 19, 369–380.

Strubell, M. (1996): "Language Planning and Bilingual Education in Catalonia," *Journal of Multilingual and Multicultural Development*, 17, 262–275.

Tabuchi, T. and J.-F. Thisse (2002): "Taste Heterogeneity, Labor Mobility and Economic Geography," *Journal of Development Economics*, 69, 155–177.

Van Grunsven, L. (1992): "Integration versus Segregation: Ethnic Minorities and Urban Politics in Singapore," *Tijdschrift voor Economische en Sociale Geografie*, 83, 196–215.

Vigdor, J. L. (2002): "The Pursuit of Opportunity: Explaining Selective Black Migration," *Journal of Urban Economics*, 51, 391–417.

Wacquant, L. J. (1993): "Urban Outcasts: Stigma and Division in the Black American Ghetto and the French Urban Periphery," *International Journal of Urban and Regional Research*, 17, 366–383.

Yancey, W. L., E. P. Ericksen, and R. N. Juliani (1976): "Emergent Ethnicity: A Review and Reformulation," *American Sociological Review*, 41, 391–403.

Yanikkaya, H. (2003): "Trade Openness and Economic Growth: A Cross-country Empirical Investigation," *Journal of Development Economics*, 72, 57–89.

Yinger, J. (1976): "Racial Prejudice and Racial Residential Segregation in an Urban Model," *Journal of Urban Economics*, 3, 383–396.