Essays in the economics of immigration and language in Canada

Ibrahim Bousmah

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Economics

Department of Economics
Faculty of Social Sciences
University of Ottawa

Dedication

This thesis is dedicated to my mother and my father, Mina Mallis and Mohamed Bousmah, for their unconditional support, love and sacrifices; to my wife Sarah and my children Salim and Riyad for their unlimited love and kindness.

Table of Contents

List of figures

List of Tables

Abstract

This thesis consists of three essays on the economics of immigration and language in Canada.

The first essay, entitled "Linguistic Distance, Languages of Work and Wages of Immigrants in Montreal", explores whether the distance between an immigrant's mother tongue and a Canadian official language (English or French), evaluated with the use of the Levenshtein distance measure, has an impact on his/her economic integration into the labour market. Using microdata from the master files of the 2001 and 2006 Canadian censuses and from the 2011 National Household Survey, I investigate the relationship between linguistic distance and the intensity of use of English and French at work in the Montreal metropolitan area. That region is characterized by the presence of sizeable French and English-speaking communities, as well as by a large number of immigrants from a wide variety of linguistic backgrounds. Those elements of linguistic diversity interact in the context of English being the international *lingua franca.* I find that linguistic distances between immigrants' mother tongues and English and French have an important impact on the relative intensities of use of the two Canadian official languages at work. I further investigate the role of the languages used at work on the earnings of immigrants by estimating earnings functions. I find that the use of both French and English are remunerated in the labour market, but that using English at work has a larger impact on earnings.

The second essay, entitled "Labour Shortages and Immigration: The Case of the Agriculture Sector", examines the role played by immigration in filling labour shortages in rural areas. Reliable access to labour is an ongoing key concern for many employers, in particular for those in regions. As an attempt to help mitigate the effects of labour shortage on the labour market, immigration has always been deployed as a key strategy, but most immigrants are concentrated in large cities. Immigration programs that try to modify the regional distribution of immigrants by attracting and retaining immigrants to rural areas all have in common that they try to fill labour shortages in those regions. A sector that represents an interesting case in point, in the dual context of attracting and retaining immigrants in rural areas and in filling a labour shortage, is the agriculture sector. I use a novel longitudinal micro-database for the years 2001-2013 from the Canadian Employer-Employee Dynamic Database (CEEDD) to identify the factors that have an impact on the recruitment and retention of Canadian and immigrant workers in this sector. In particular, in response to the efforts to explore permanent residence pathways, whether or not former temporary foreign workers (TFW) with Canadian experience are most likely to stay in the sector after entering remains a key question for policy makers that I investigate.

The third essay, "Immigrant Assimilation in a Multicultural and Multilingual Context", expands on the work of the first essay by looking at languages used at home. I investigate whether linguistic distance between an immigrant's mother tongue and a Canadian official language (English or French) has an impact on his/her assimilation to the host country's official languages. Using microdata drawn from the master files of the 2001, 2006 and 2016 Canadian censuses and the 2011 National Household Survey, I investigate the relationship between linguistic distance and the intensity of use of English and French at home for immigrants in the Montreal metropolitan area. I find that linguistic distances between immigrants' mother tongues and English and French have an important impact on the relative intensities of use of the two Canadian official languages at home. I further investigate the role of spousal characteristics and other variables on the assimilation of immigrants. The results suggest that the home environment is an important factor contributing to the linguistic assimilation of immigrants. Individuals exposed to an official language at home with their spouse have significantly higher rates of assimilation.

Acknowledgement

First and foremost, I would like to show my gratitude and thanks to my Creator who gave me health, patience and other uncountable blessings to be able to pursue and finish this degree. I would also like to express my special appreciation to my supervisors Dr. Gilles Grenier and Dr. David Gray for their guidance, valuable advice, and support throughout the past years. They have helped me walk in the research world and become a better researcher. I wish to thank my committee members, Dr. Louis-Philippe Morin, Dr. Jean François Tremblay, Dr. Jose Galdo and Dr. François Vaillancourt for their valuable comments and suggestions that helped improve the quality of this thesis. My profound appreciation to my parents, Mina Mallis and Mohamed Bousmah, for their love, financial support and encouragements. Finally, I would like to thanks my beloved wife and my children, Sarah Affany, Salim Bousmah and Riyad Bousmah for their understanding, love and presence in my side during these past years.

Introduction to the thesis

Le marché du travail canadien a bien performé par rapport aux normes internationales au cours du 21e siècle. Malgré un profil économique favorable par rapport à de nombreux pays de l'OCDE, le Canada fait toutefois face à d'importants défis futurs, tels que le vieillissement de sa population, un taux de fécondité relativement bas et des pénuries de main-d'œuvre dans certains métiers. L'une des solutions pour réduire l'impact de certains de ces problèmes est de s'appuyer sur l'immigration, qui a pour objectif de contribuer de manière positive à l'économie. Toutefois, le processus d'intégration des immigrants peut être difficile. Il est donc important d'étudier différents aspects de l'intégration des immigrants et les coûts d'ajustement associés. Cette thèse s'intéresse particulièrement à l'assimilation linguistique des immigrants ainsi qu'à leur intégration dans les régions rurales.

Le premier chapitre de cette thèse cherche à déterminer si la distance entre la langue maternelle d'un immigrant et une langue officielle canadienne (anglais ou français) a une incidence sur son intégration économique dans le monde du travail. Avec l'utilisation de la mesure de distance linguistique de Levenshtein, et à l'aide des microdonnées des fichiers confidentiels des recensements canadiens de 2001 et 2006 et de l'Enquête nationale auprès des ménages de 2011, j'étudie la relation entre la distance linguistique et l'intensité d'utilisation de l'anglais et du français au travail dans la région métropolitaine de Montréal. Cette région se caractérise par la présence d'importantes communautés francophones et anglophones, ainsi que par un grand nombre d'immigrants issus de milieux linguistiques divers. J'étudie en outre le rôle des langues utilisées au travail sur les salaires des immigrants en estimant des fonctions de gains. Je trouve que l'utilisation du français et de l'anglais est rémunérée sur le marché du travail, mais que l'utilisation de l'anglais au travail a un impact plus important que le français sur les revenus

de travail.

Le deuxième chapitre examine le rôle joué par l'immigration pour combler le manque de main-d'œuvre dans les régions rurales ou petites villes. Un accès fiable à la main-d'œuvre est une préoccupation majeure pour de nombreux employeurs, en particulier dans les régions. Pour tenter d'atténuer les effets de la pénurie de main-d'œuvre sur le marché du travail, l'immigration a toujours été déployée en tant que stratégie clé, mais la plupart des immigrants sont concentrés dans les grandes villes. Les programmes d'immigration qui tentent d'améliorer la répartition nationale des immigrants en attirant et en retenant des immigrants dans les zones rurales ont tous un objectif principal en commun, celui d'essayer de combler les pénuries de main-d'œuvre dans ces régions. Le secteur de l'agriculture est un secteur qui représente un exemple intéressant dans le double contexte d'attraction et de rétention d'immigrants dans les zones rurales et dans le traitement d'une pénurie de main-d'œuvre. J'utilise la base de données longitudinale sur la dynamique canadienne entre employeurs et employés de 2001 à 2013 afin d'identifier les facteurs qui ont une incidence sur le recrutement et la rétention de travailleurs canadiens et immigrants dans ce secteur. En particulier, en réponse aux efforts politiques visant à faciliter la résidence permanente pour les travailleurs étrangers, je cherche à comprendre si le fait d'avoir de l'expérience canadienne avant de devenir permanent résident contribue à de meilleurs niveaux de rétention dans les secteurs avec pénurie de main-d'œuvre.

Le troisième chapitre s'inspire du travail du premier chapitre en cherchant à déterminer si la distance linguistique entre la langue maternelle d'un immigrant et une langue officielle canadienne (anglais ou français) à une incidence sur son assimilation aux langues officielles du pays hôte en termes de la langue utilisée à la maison. À l'aide de micro- données tiré des fichiers confidentiels des recensements canadiens de 2001, 2006 et 2016 et de l'Enquête nationale auprès des ménages de 2011, j'étudie la relation entre la distance linguistique et l'intensité de l'utilisation

de l'anglais et du français à la maison dans la région métropolitaine de Montréal. Je constate que les distances linguistiques entre les langues maternelles des immigrants et l'anglais et le français ont une incidence importante sur l'intensité relative de l'utilisation à la maison des deux langues officielles du Canada. De plus, j'étudie aussi le rôle des caractéristiques des époux et d'autres variables dans l'assimilation des immigrants. Les résultats suggèrent que l'environnement familial est un facteur important qui contribue à l'assimilation linguistique des immigrants. Les personnes exposées à une langue officielle à la maison avec leurs conjoints/conjointes ont des taux d'assimilation beaucoup plus élevés.

L'immigration est devenue l'un des principaux moteurs de la croissance démographique et de la croissance de l'emploi au Canada. Les immigrants constituent une importante proportion de la population canadienne. De là l'importance d'étudier différents aspects de leur intégration dans leur pays d'accueil. Les recherches sur les différents enjeux d'intégration des immigrants permettent d'appuyer le développement de politiques d'immigration ainsi que de fournir les éléments empiriques pour de futures initiatives.

Chapter One

Linguistic Distance, Languages of Work and Wages of Immigrants in Montreal

1.1 INTRODUCTION

The Canadian labour market has performed well by international standards over much of the 21st century, exhibiting a relatively high employment-to-population ratio, a relatively high labour force participation rate, and a relatively low long-term unemployment rate. Despite a favourable economic profile relative to many OECD countries, however, Canada is also facing important future challenges, such as an aging population, a fairly low fertility rate, higher health-care spending, labour market shortages for certain occupations, affordable housing shortages, and the erosion of competitiveness in international trade. One of the remedies to reduce the impact of some of those challenges is to rely on immigration, which hopefully benefits the host country and contributes positively to the economy by augmenting the total supply of workers, increasing productivity, creating new businesses and jobs, improving innovation performance, and improving fiscal balances (OECD, 2016).

Having mentioned the benefits of immigration, one must also address the potential costs. Indeed, it is important to consider the fact that many of those newcomers are ethnically, linguistically, and culturally different from the host population, and that without adequate social

and economic support from the receiving country, the process of integration can be costly for both the immigrants and the host country. It is therefore important to investigate different aspects of the economic integration of immigrants into the Canadian labour market and the associated adjustment costs. This study will focus on the linguistic aspect of immigrants' integration and its effect on their economic well-being. Immigrants' capacity to communicate in either one of Canada's official languages is essential in order to succeed socially and economically, as linguistic barriers constitute one of immigrants' principal barriers to integration. To the extent that they are not addressed, immigrants are expected to experience lower earnings and productivity profiles. It is therefore relevant to conduct research on linguistic characteristics of immigrants in the Canadian labour market.

The metropolitan area of Montreal is an interesting case for research purposes, as it is characterized by the presence of important French-and English-speaking communities, as well as by a large number of immigrants from a wide variety of linguistic backgrounds. Unlike most other destination areas, where only one major language is used, immigrants in Montreal are exposed to two languages that are commonly used in the labour market, and they usually have to make choices between the two. Using microdata from the master files of the 2001 and 2006 Canadian censuses and from the 2011 National Household Survey, I investigate the impact of the linguistic distances between an immigrant's mother tongue and both English and French on the intensity of use of those languages at work in the Montreal metropolitan area. I also explore the role of the languages used at work in affecting the earnings of immigrants. As a proxy for linguistic distance, I employ an index developed by the *Max Planck Institute for Evolutionary Anthropology*, which is derived from an algorithm based on the Levenshtein distance.[1] I find that

[1] I describe how the index is calculated in section 1.4.

linguistic distances between immigrants' mother tongues and English and French have an important impact on the relative intensities of use of the two Canadian official languages at work. I also find that immigrants in the greater Montreal area gain a higher payoff from using English at work than from using French.

This paper is structured as follows. The second section consists of a review of selected studies in the existing literature regarding the effect of language skills on the labour market in Canada. The third section lays out the conceptual framework. The fourth section presents some general background information and a description of the dataset. It also explains the sample restrictions that are imposed and presents some descriptive statistics. The fifth section presents the econometric model and the various specifications. The sixth section contains the discussion of the results of the empirical work, and the last section contains the conclusion.

1.2 LITERATURE REVIEW

Many studies have focused on the effect of language skills and characteristics on labour market outcomes. The traditional line of research focused on trying to estimate the wage premium received from knowing or using an official language. A more recent line of research upon which I draw involves the introduction of the concept of linguistic distance that is accompanied by a quantitative measure of the distance between languages.

Vaillancourt (1980), one of the earliest influential studies, investigates the role of the languages known by an individual in explaining earnings in the labour market of the province of Quebec. Drawing on the public use microdata file of the 1971 Census of Canada, Vaillancourt excluded from his estimating sample women, individuals who are neither francophones nor anglophones, non-whites, and those without positive earnings in 1970. The estimation method employed is a semi-logarithmic earnings equation that includes a set of three dichotomous

language variables as the key independent variables: Unilingual Anglophone, Bilingual Anglophone, and Bilingual Francophone, with Unilingual Francophone as a reference group. That scheme for categorizing the language groups was adopted in many subsequent studies. The main empirical findings are that knowing English brings higher earnings to males in Quebec and in Montreal, but that bilingualism brings no monetary returns to Anglophones. In contrast, bilingualism does afford wage premiums of ten percent to francophones in Quebec and six percent in Montreal. That paper is one of the first to show evidence of the importance of possessing English skills in the predominantly French-speaking province of Quebec.

Carliner (1981) analyzes wage differences among language groups in the French-speaking and English-speaking parts of Canada, specifically examining the supply and demand for language skills and the factors that can influence them. The data source is the 1971 Canadian Census Public Use Microdata Files (PUMF). One interesting result relates to the link between education and language skills; unilingual Anglophones have 2.7 more years of education on average than the unilingual Francophones and have 0.2 more years than the bilingual Francophones. In addition, one of the findings regarding Montreal indicates that the group whose native tongue is not an official language receives substantial economic rewards for learning French or English. Furthermore, unilingual Francophones in this region earn less than the unilingual English workers, suggesting that the possession of English skills for a native French speaker in Montreal is more beneficial than is the case for English native speakers possessing French skills. In contrast, the results for the English-speaking part of Canada show that the bilingual English workers are the group earning the highest income, followed closely by the unilingual English speakers. Carliner (1981) thus concludes that in English Canada, the economic rewards for learning French are lower than they are in the province of Quebec, where

learning English as a second official language leads to significantly higher wages.

Shapiro and Stelcner (1997) examine linguistic earnings disparities over twenty years among Francophones, Anglophones, and Allophones situated in Quebec. They draw on data from the 1991 Census and compare their results with those obtained from previous studies from 1971 and 1981 census data. Restricting their samples to full-time workers aged 18 to 65 years, they estimate earnings regressions including controls for human capital, personal characteristics, labour supply patterns, and indicators for the type of occupation and industry. The main findings suggest that since the implementation of Bill 101 in 1977, the earnings gap between bilingual and unilingual Anglophones and bilingual Francophones has decreased, but that the relative earnings of Allophones and unilingual Francophones have fallen. With unilingual Anglophones as a reference group in all their regression equations, they found that the earnings disadvantage for unilingual Francophone men increased to approximately 8 to 9 percent in 1990, compared to a range from 5 to 7 percent in 1980. The results for Allophones show that men who speak only French had a wage penalty of about 18 percent in 1990 compared to approximately 15 percent in 1980. Allophone men who speak only English had a penalty of 13 to 15 percent in 1990, up from 8 to 10 percent in 1980, and the earnings disadvantage experienced by bilingual Allophones was 6 to 8 percent in 1990 compared to 4-6 percent a decade earlier.

Albouy (2008) focuses on the wage gap between Francophones and Anglophones in Canada from 1970 to 2000 based on a sample of 20 to 59-year-old males born in Canada who speak French or English fluently. The data sources are the Canadian censuses of 1971, 1981, 1986, 1991, 1996 and 2001. The results suggest that from 1970 to 2000, the wage premium for the Quebecois Francophones for learning English fell from 11 percent to 8 percent. In the case of Anglophones in Quebec, the returns to learning French increased from 0 to 5 percent. The

author suggests that this evolution is the result of the large number of laws and regulations implemented in Quebec. When examining the English-speaking part of Canada, Albouy (2008) did not find significant returns for Anglophones to acquiring French skills.

A common feature of the literature cited above is the focus on the attribute of knowledge of a language as opposed to a focus on the premium of using French or English in the workplace. This emphasis was mainly due to the fact that data on the language used at work were not collected before the 2001 Census. This relatively new variable is used by Li and Dong (2007), Christofides and Swidinsky (2010) and Grenier and Nadeau (2016). Li and Dong (2007) use the language-of-work variable, but in a different context than the preceding studies. They employ it as a proxy for participation in an enclave labour market, defined as a geographic area with a high ethnic concentration. The main purpose of their paper is to test if Chinese immigrants in the Canadian labour market who participate in the enclave economy earn a lower return than their counterparts who are working in the mainstream economy. The authors examine this wage gap and find (as expected) that earnings are substantially lower for immigrants in the enclave economy. They explain this key finding by the difference in the types of jobs concentrated in the enclave (compared to those jobs outside of the enclave), and by the fact that Chinese immigrants in the enclave economy are more likely not to speak the official languages.

Christofides and Swidinsky (2010) estimate the earnings advantage acquired by learning a second official language in Canada and using it at work, focusing on these effects for Francophones in Quebec and Anglophones in the rest of Canada. Their data source is the individual file of the 2001 Census Public Use Microdata File. They omit immigrants from their sample and include only Canadian-born individuals of 15-64 years of age who worked full-time, full year and had at least a high school certificate. Their results reveal that in Canada outside

Quebec, the wages of men who are Anglophones are 3.8 percent lower than those of bilingual men who work only in English, 5.4 percent lower than those of bilingual men who frequently use French in the workplace, and 8.4 percent higher than the small number of bilingual men who work either equally, mostly, or exclusively in French. In the case of Anglophone women, the outcomes are quite similar, but with a few differences. Compared to unilingual, English-speaking women, bilingual women who use French frequently at work earn 9.3 percent more, and the women who are fluent in French but use only English at work earn a premium of 6 percent.

In comparison, Christofides and Swidinsky's findings for the province of Quebec show that bilingual men who use only French at work earn a premium of 7 percent relative to the unilingual Francophones. Moreover, bilingual Quebec residents who regularly use English at work earn 20.9 percent more than their unilingual counterparts. These results again confirm that knowledge and use of the English language is an important factor for success in the Canadian labour market irrespective of the official language of the province.

Grenier and Nadeau (2016) focus on the effects of the languages used at work on wages in the metropolitan area of Montreal while treating the mother tongue as the principal variable of interest, which is divided into three categories: French, English, and others. Their specification allows for the estimation of three distinct effects on wages of working in a second language: i) the effect of using French at work for Anglophones, ii) the effect of using English at work for Francophones, and iii) the effect of using either official language at work for those with other mother tongues. Their data source is the 2006 Census master file. They find that Anglophones do not benefit from the use of French at work. By comparison, however, when considering Francophones and the group with other mother tongues, they discern an important positive effect for the usage of English in their workplace. A further interesting finding is that French-speaking

immigrants increase their use of English at work as they spend more time in Canada, whereas immigrants in the other-mother-tongues group seem to use English at first and then to move gradually to using French. Overall, their findings indicate that knowledge and command of English is an important element for success in the Montreal labour market, and that financial incentives are an important determinant of the decision to learn English for Francophones and for members of the other-mother-tongues group.

More recently, an emerging literature has developed a new approach that introduces the concept of linguistic distance into the process that involves estimating the effects of linguistic differences on the labour market outcomes of immigrants. This variable is accompanied by a quantitative measure of the distance between languages. I will briefly describe the three most popular techniques employed to measure these linguistic distances. For a more detailed explanation of those techniques, one can consult Ginsburgh and Weber (2016).

Several measures of linguistic distances have been proposed. One popular metric is based on the Levenshtein distance, which is an algorithm that measures the distance between words in two different languages that have the same meaning by determining the minimum number of transformations required to convert the word from one language into the corresponding word in the other language.

Isphording and Otten (2013) apply the Automated Similarity Judgment Program (ASJP), a measure based on the Levenshtein distance and developed by the *Max Planck Institute for Evolutionary Anthropology*, to analyze the economic success of immigrants in the labour market of some host countries. In order to carry out an international comparison, they use data from the 2000 U.S Census, the German SocioEconomic Panel, and the National Immigrant Survey of Spain. They estimate a probit regression whose dependent variable is equal to one if the

individual reported having “good” or “very good” language abilities in the host-country language. The key independent variable is the linguistic distance between the immigrant’s native language and the language of the host country, which is entered in the specifications as a percentile measure in order to be able to compare the results with the test score's approach (described below). The other control variables are those typically included in the immigration literature, such as age at migration, years since migration, years of education, marital status, and number of children. In some specifications they also include the geographical distance between national capitals in kilometers as a proxy for migration costs. For all specifications their results show a significant negative effect of immigrants’ linguistic distance on the probability of reporting either good or very good ability in the host country language. In the U.S., for example, an individual situated in the first percentile of the Levenshtein distance distribution has an increased probability of 20 percentage points of reporting good or very good linguistic skills compared to another individual coming from the highest percentile. In Germany and Spain, these increases in probability reach 40 and 20 percentage points, respectively.

Gunduz (2017) employs the Levenshtein distance measure in order to investigate the degree of immigrant-native substitutability in production for Canada by applying methodologies proposed by Borjas (2003) and Ottaviano and Peri (2012). Immigrants’ skill types are allowed to vary by language skills, for which linguistic difference serves as the proxy. She divides her sample into three principal groups (low, medium and high) based on the value of their linguistic distance from English or French. Her results show that when the analysis is conducted separately by language groups, the estimates for low-language-skill immigrants suggest an imperfect degree of immigrant-native substitutability, and the estimates for medium and high-language-skill immigrants suggest a perfect degree of substitutability.

Another measure of linguistic distances is the method of learning scores, which consists of following over time a group of people who learn a language, and then measuring their progress at the end of some learning period. This measure is used by Chiswick and Miller (2005), who developed a quantitative measure of the distance between English and other languages based on the difficulty that a typical American would have in learning those languages within a fixed period of time. It was based on a standardized proficiency test developed by the U.S. State Department School of Language Studies that teaches foreign languages to English-speaking Americans in preparation for placement as diplomats. A test was designed and conducted according to which a higher or lower score indicates a lower or higher linguistic distance from English. A limitation of this index is that it only considers the distance between English and other languages, which poses a problem when applying it to analyze the Canadian labour market, a task for which one also needs a measure of the distance between French and other languages.

A further interesting distance technique that has been used is one based on linguistic trees, which computes these distances by dividing language into branches as represented by a tree diagram. The closer the branches are from one language to another, the lesser the distance between them. Adsera and Pytlikova (2015) investigate how linguistic distance influences migration choices by using this particular measure as a proxy for how easy or difficult it is to learn the language of the destination country. They created an index ranging from 0 to 1 in value that reflects the total number of levels of the linguistic family tree that the languages of the destination and the source country share in common. The estimating sample was on immigrants in 30 OECD destinations coming from 223 source countries during the years 1980-2009. The data were collected from the OECD International Migration Database, and for some countries the

data were obtained directly by writing to national statistical offices. Their results suggest that migration rates tend to increase with linguistic proximity, and that it is more important for migrants moving to non-English-speaking destinations than to English-speaking countries. They conjecture that this pattern could be due to the likely higher English proficiency level of the average migrant compared to proficiency levels in other languages.

In a recent study, Adsera and Ferrer (2015) contributed to the analysis of the labour market integration of Canadian immigrants. They used the confidential micro-data of the Canadian censuses of 1991, 1996, 2001 and 2006. Their sample is restricted to males aged 18 to 60 years with the exclusion of Aboriginals. They selected a 25 percent random sample of Canadian-born individuals from each census plus all immigrants who arrived in Canada at age 18 or older. Using the measure of linguistic tree proximity of the immigrant mother tongue to the host country language developed in Adsera and Pytlikova (2012), they examined the role that it plays in the labour market performance of immigrants compared to their native-born counterparts. They estimate a Mincerian earnings equation that is augmented with indicators for job skill requirements and for linguistic proximity. Their results show that immigrants with closer linguistic proximity receive higher weekly wages than those with more distant proximity. In fact, immigrants whose languages share no branch with either English or French earn weekly wages that are 32 percent lower than otherwise similar Canadian-born workers. Another interesting result is that immigrants whose mother tongue is close to French and who are settling in Quebec have similar or better labour market outcomes than immigrants whose mother tongue is close to English, and who are settling outside Quebec.

My work expands on the above-cited literature by first investigating the effect of immigrants' linguistic distances on the choice of the language utilised at work. Specifically, the

outcome variables are the relative intensities of using each of the two official languages given the worker's host language. Furthermore, this research aims to compare the economic returns of using French and English at work for immigrants in the Montreal metropolitan area based on wage equations. This facet of my paper extends the work of Grenier and Nadeau (2016) by including the linguistic distance variable.

1.3 Conceptual Framework

Learning a new language (and using it at work) is an investment in human capital that depends on the benefits and costs (Chiswick and Miller, 2015; Grenier and Nadeau, 2016). The benefits of obtaining this skill can be directly related to the wage premium received from applying it, and the costs are related to the difficulty of learning that language. The linguistic distance from an immigrant mother tongue and an official language is an important element of the cost of learning and using the languages. The first hypothesis that I will attempt to test in my empirical analysis is that immigrants whose mother tongue is close to French/English will have relatively low difficulty in learning and eventually using French/English in the workplace.

The second component of my empirical analysis involves the economic returns to using these languages at work. Conditional on learning the language, immigrants are expected to use it at work, and the choice of using French, English or both should have a direct impact on their earnings. In a multilingual community such as Montreal, one language might yield a higher economic return than another, an effect that could be represented by a supply and demand framework. On the supply side, there are those workers who are able and willing to use at least one of the official languages at work, and on the demand side there are the employers who are looking to fill positions requiring these skills. A decrease in supply and/or an increase in demand of labour will *ceteris paribus* raise wages; and a decrease in demand and/or an increase in supply

of labour will *ceteris paribus* reduce wages. As an illustration, consider a group of workers called 'X' who can only use French at work, and another group of workers called 'Y' who can only use English at work. Other things held constant, if a change such as the implementation of a new immigration policy increases the supply of workers in group X and reduces the supply of workers in group Y, we would expect a decrease in the wage rate for group X and an increase in the wage rate of group Y.

In addition to direct impacts on worker productivity in the local labour market, wage differentials among language groups could be generated depending on the relative values of languages in the world. In the twenty-first century, English is perceived across the world as the major *lingua franca* (Brutt-Griffler, 2002; Jenkins, 2007; Grenier, 2015). With the increasing globalization of the world's economy and the intercommunication that it involves, English is now commonly used as the principal vehicle of communication for facilitating international trade, business relations, scientific research, and tourism interactions. This special status and value of the English language over French could also be a potential explanation of any estimated difference in economic returns to their utilization in the Montreal area. My econometric specifications do not allow us to determine specifically the relative explanatory powers of these two behavioural mechanisms.

1.4 DATA AND SUMMARY STATISTICS

1.4.1 DATA AND MAIN INDICATORS

The datasets used for this paper are the micro-data master files of the 2001 and 2006 Canadian Censuses and the 2011 National Household Survey (NHS) from Statistics Canada. Censuses prior to 2001 are not exploited in this analysis because the variables on languages used at work are not available. Those databases provide a rich source of information on labour market

characteristics, immigrant status, and language features. They also contain detailed information about the economic, social and demographic characteristics of the Canadian population. Another important advantage is the large sample size covering 20 percent of the Canadian population and containing more than 200 variables. To compute the measure of linguistic distance between languages, I use the database of the ASJP developed by linguists at the Max Planck Institute for Evolutionary Anthropology. From these sources of data, two key indicators are employed in the analysis: *Linguistic distance* and *Relative intensity of use of English and French at work*.

In regard to the measure of *linguistic distance*, I draw on the database of the ASJP developed by the Max Planck Institute for Evolutionary Anthropology, and which is based on the Levenshtein distance. The measure is constructed from a list of words in pairs of languages with similar meanings and from the number of edits that are necessary to transform a word from one language to the other. It is normalized to account for differences in lengths of words and for word lexical similarities resulting from coincidence. The details of the calculations are provided in the Appendix 1.1. I obtain the linguistic distance between both of Canada's official languages (French and English) and all other foreign mother tongues reported by the respondents in the datasets (about 150 different languages). The linguistic distances from French and English have values ranging from 0 (for the same language) to approximately 104 for the greatest distance between two languages.

The measure of the relative intensity of use of English and French at work is derived from a main question and sub-question posed in the censuses and in the NHS regarding the languages used at work, which are worded in the following way: 1) In this job, what language did this person use most often? 2) Did this person use any other languages on a regular basis in this job? To those questions the respondent can answer French, English, both, or another

language. Since this research focuses on the use of both of the official languages, I remove immigrants who reported using a non-official language at work in either the main question and the sub-question; this restriction left us with approximately 90 percent of our original sample of immigrants. Based on the various combinations of answers to the above two questions, I define the English-to-French intensity (**EtoF**) index, which assumes values between zero and one hundred. Relatively low (high) values are associated with a pattern of English (French) predominance. Specifically, the values refer to typical patterns and are assigned as follows:

- 0 if English is used most often and French is never used on a regular basis
- 25 if English is used most often, but French is also used on a regular basis
- 50 if both official languages are equally used most often
- 75 if French is used most often, but English is also used on a regular basis
- 100 if French is used most often and English is never used on a regular basis.

An intuitive interpretation of this index is that it approximates the proportion of time an immigrant uses French at work as opposed to English in the context of the dual linguistic nature of the labour market in the Montreal metropolitan area. Given the somewhat arbitrary assignment of values to this index, a qualitative version that considers only the order of the choices will also be used in this study.

1.4.2 SUMMARY STATISTICS

I restrict my analysis to immigrants aged between 25 and 65 years old who are employed as full-time, full-year workers (30 hours or more per week, 48 weeks or more per year) and who became landed immigrants after the age of 18. I follow previous studies by focusing on individuals who reported having worked full-time in order to obtain consistent annual wages estimates that are not affected by workers who work few hours.

Table 1.1 presents the mean values of the key variables. The Consumer Price Index (CPI) has been utilized to convert 2001 and 2006 wages into 2011 constant dollars. The summary statistics indicate that immigrants in my sample earn on average $47,640 a year. Two thirds of them are married, and the average age is 45 years. For education, I divided my sample into five categories based on the highest diploma and degree attained. Immigrant workers in Montreal have high levels of education, as the proportion without a high school degree is the lowest for any category. Approximately a third of them obtained some post-secondary college education, and two out of five received a university bachelor's degree or a graduate degree. Table 1.1 also shows that there is a higher proportion of men than women, and that immigrants have been in Canada for an average of 24 years.

A main variable of interest the linguistic distance from English and French. We see that on average immigrants in the greater Montreal area have a mother tongue that is significantly more distant from English (85.9) than from French (71.6). To have a better idea of the distribution of mother tongues in our sample, table 1.2 reports the five most common mother tongues and their respective linguistic distances from French and English. Besides French and English, Arabic, Spanish and Haitian Creole are the main immigrant mother tongues, and that they are all closer to French than to English. Those five mother tongues account for almost six out of ten immigrants in the total sample.

The other linguistic variable is the English-to-French intensity index defined above. Figure 1.1 consists of a histogram of the proportion of immigrants corresponding to each level of intensity, and it shows a fairly high degree of dispersion. The distribution confirms the presence and the importance of both official languages at work in the Montreal metropolitan area. In fact, about half of the workers used both languages at work.

1.5 EMPIRICAL METHODOLOGY

1.5.1 MODEL SPECIFICATIONS

My model consists of two regression equations, the first of which estimates the effect of linguistic distance and other regressors on the relative intensity of use of the two official languages at work. The second one estimates the economic return of that relative intensity. The model consists of the following equations:

$$EtoF_{i,t} = \beta_0 + \beta_1 LDF_i + \beta_2 LDE_i + X_{i,t}\delta + \lambda_t + [Z_{i,t}\theta + \partial_{i,c}] + \epsilon_{i,t} \quad (1)$$

$$\ln(wage)_{i,t} = \gamma_0 + \gamma_1 EtoF_{i,t} + X_{i,t}\eta + \lambda_t + [Z_{i,t}\zeta + \partial_{i,c}] + \nu_{i,t} \quad (2)$$

In equation (1), the (limited) dependent variable, $EtoF_{i,t}$, is the measure of the intensity of using French as opposed to English at work for individuals (i) in the period (t). As explained above, it increases (decreases) with the use of French (English). The two key independent variables for this model are the linguistic distances between the immigrant's mother tongue and French (LDF_i) and English (LDE_i) respectively. The sign of the estimated coefficient of β_1 is expected to be negative, since the further the language is from French, the less likely it is to be used at work. The sign of the estimated coefficient of β_2 is expected to be positive, since the further the language is from English, the less (more) likely that English (French) is used at work. The matrix $X_{i,t}$ contains the socioeconomic variables, which include age, age squared, gender, education, marital status, years since migration, years since migration squared, and an indicator variable for the presence of at least one child. The argument for including those variables in our model is that based on our conceptual framework, the costs involved in an immigrant learning a new language would be influenced by several factors, and the variables on the right hand side of

equation (1) are all assumed to influence the cost of learning and using an official language at work. The vector λ_t includes year-specific dummies to control for aggregate trends. An augmented specification includes the matrix $Z_{i,t}$ containing 15 industry-specific dummies, a set of dummies to control for the location of work within the Montreal region (working in the west, center, east and other subdivisions), and an indicator for immigrants whose previous residence was outside Quebec one or five years earlier.

A potential problem with this regression equation is that the identification of the impact of linguistic barriers on the language use intensity might not be reliable due to the correlation between unobservable cultural differences and the linguistic distance variables (Isphording and Otten, 2014). To address this source of potential bias, I add a specification that includes detailed country-of-birth fixed effects$(\partial_{i,c})$.[2] The additional terms are written in brackets. The last term labeled $\epsilon_{i,t}$ is the error term, which varies across individuals and time periods.

Equation (2) is a standard human capital earnings equation that is augmented to capture the impact of the intensity of using French as opposed to English at work on earnings. The dependent variable is the logarithm of annual wages, and the independent variables are the official language intensity$(EtoF_{i,t})$, the socioeconomic controls $(X_{i,t})$ and the year dummies (λ_t). The additional controls specified as $Z_{i,t}$ and $\partial_{i,c}$ (defined above) are included in some specifications. The error term labeled $\nu_{i,t}$. The sign of the coefficient estimates of the variable $EtoF_{i,t}$ is our primary focus. If working in French (English) is more highly remunerated in the labour market than is working in English (French), we expect a positive (negative) sign.

[2] I examined the correspondence between mother tongues and country of birth, and I found that 89% of the countries in our sample have more than one mother tongue reported. Given the absence of a one-to-one correspondence, I also conducted the regression analysis including region-of-birth indicators and additional variables designed to capture cultural differences. The results are pretty similar to those generated from our primary specifications.

One concern with equation (2) is the common endogeneity problem, which in our case can arise from measurement errors or from omitted variables. The intensity index is derived from self-reported responses to a primary question and a sub-question that could reflect reporting errors. Certain individuals could take more or less time answering the sub-question, consequently introducing some unknown influences in the reported value of the index. In addition, similarly to the well-known ability bias that affects estimates of the economic return to schooling, the ordinary least squares estimates might overestimate the value of the true return on earnings of the intensity of using one of the official languages at work. For instance, immigrants with higher levels of innate ability tend to be those who perform better in the labour market and earn higher wages. Without controlling for individual ability, the estimated coefficient of the language-of-work variable could in part capture part this effect. To address this problem, I employ the instrumental variable methodology (IV), which will be discussed in section 6.3.

1.5.2 ROBUSTNESS CHECKS

I provide a set of sensitivity analyses to test the robustness of the results. The values of the English-to-French intensity index were arbitrarily assigned to 0, 25, 50, 75 and 100 according to the answers to the two relevant survey questions. Given the ordered nature of the language intensity index, as a first check, I verify the robustness of the results by estimating an ordered probit model as an alternative specification to equation (1). Let $EtoF^*$ denote the latent variable of the "real intensity of use of French as opposed to English at work", with higher (lower) values representing a higher intensity of use of French (English) at work. This variable is then modelled as a multi-variate function of the same set of regressors as those included in equation (1). This model is expressed as follows:

$$EtoF^{*} = \phi_1 LDF_{i,t} + \phi_2 LDE_{i,t} + X_{i,t}\pi + Z_{i,t}\varrho + \lambda_t + \partial_{i,c} + \varepsilon_{i,t}\,, \qquad (3)$$

where ϕ_1 and ϕ_2 are the coefficients associated with the effects of linguistic distance on the intensity of using an official language at work. The latent dependent variable $EtoF^{*}$, corresponds to the observed variable $EtoF$, which must be recoded such that the values of the language of work index are 0, 1, 2, 3, and 4[3]. It is assumed that $EtoF^{*}$ is related to the observable, ordinal $EtoF$ variable as follows:

$y_i = 0$ if $-\infty < y_i^{*} < \mu_1$ (English only)

$y_i = 1$ if $\mu_1 < y_i^{*} < \mu_2$ (English first and French second)

$y_i = 2$ if $\mu_2 < y_i^{*} < \mu_3$ (English and French equally)

$y_i = 3$ if $\mu_3 < y_i^{*} < \mu_4$ (French first and English second)

$y_i = 4$ if $\mu_4 < y_i^{*} < \infty$ (French only),

Where $\mu_i's$ are threshold parameters $(\mu_1 < \mu_2 < \mu_3 < \mu_4)$ to be estimated along with the coefficients of equation (3). After obtaining the parameter values that maximize the likelihood functions, I derive the marginal effects showing the changes in probabilities of immigrants of being at different intensity levels occasioned by one-unit changes in the linguistic distance variables.

As a second check of the robustness of the results of our first equation, I investigate whether immigrants who have one of the official languages as a mother tongue are the group that drives the findings. I explore this possibility by omitting from the sample immigrants with a linguistic distance of zero from French or English. These individuals account for about 28% of our original sample.

For the third robustness check, I verify if the results are sensitive to the specification of

[3] This mapping is positive and monotonic such that higher (lower) levels represent greater use of French at work and vice versa.

the linguistic distance variable by using an alternative indicator, namely a dummy variable that is equal to 0 if the distance between an immigrant's mother tongue is closer to English than to French, and equal to 1 if the distance is closer to French than to English. The motivation behind this specification is the assertion that immigrants will tend to choose the official language that is closer to their mother tongue. This set of regressions also forms the first stage of the instrumental variables earnings regression that I will present below.

The last robustness check investigates whether the main findings of the earnings regression are affected by using an alternative measure of the official language intensity variable. I repeat the regression analysis by defining the indexes as a group of indicator variables, allowing to derive separate estimates for the impact for each group of language use intensity. This is similar to the first robustness check described above, but in this instance the intensity index enters as an independent variable rather than the dependent variable. The results of all of these sensitivity analyses are presented below along with the main findings.

1.6. THE RESULTS

1.6.1 IMPACT OF LINGUISTIC DISTANCE ON THE USE OF OFFICIAL LANGUAGES AT WORK

The OLS results for the equation modelling the impact of the linguistic distance variables on the official language used at work are presented in table 1.3. The first three columns display the results for the full sample, and results in the last column are generated from a sample that excludes immigrants whose mother tongue is English or French (which refers to the second robustness check described above). The specification listed in column (1) includes the basic controls for age, age-squared, gender, education, marital status, years since migration, years since migration-squared, the presence of children, and year dummies. The specification in column (2) adds controls for the location of work within the census metropolitan area of

Montreal, an indicator for the region of residence outside the province of Quebec one year or five years earlier, and industry fixed effects.[4] The results for the fullest specification, which includes the detailed country-of-birth effects, are listed in column (3).

For all of the specifications, the sign of the estimated coefficients for the distance from French variable are negative and highly statistically significant, and thus one of my key findings is that the relative intensity of using French at work as opposed to English decreases with the linguistic distance from French. Similarly, the results for the estimates related to the linguistic distance from English variable are positive and highly statistically significant, indicating that the relative intensity of using French at work as opposed to English increases with the linguistic distance from English.[5] We note, however, that the coefficients' estimates decline in magnitude in the preferred third specification when I add the detailed country-of-birth effects. This pattern suggests that in the first and second specifications, from which those controls are omitted, these estimates of the impact of linguistic distance captured the effect of cultural differences and could also capture the quality of teaching English or French in the home countries of immigrants.

Column (4), in which the results based on omitting immigrants with a linguistic distance of zero from one of the official languages are listed, shows similar empirical patterns for both linguistic distance variables.

To illustrate the economic magnitude of the effect of the linguistic distance from French, we can take as an example two immigrants, one with Vietnamese as a mother tongue and another with Haitian Creole as a mother tongue. The linguistic distance from French for the Vietnamese

[4] The census contains two questions regarding past residences. One is a flag for residence outside of Quebec one year ago or earlier. Another is a flag for residence outside of Quebec five years ago or earlier. I include a single indicator that is the union of these two variables.

[5] In order to address multicollinearity concerns, I calculated the coefficient of correlation between two variables: the linguistic distance between French and a given mother tongue and the linguistic distance between English and the same mother tongue. The value of this correlation is -0.097, indicating that there is a very weak linear relationship between the variables.

immigrant is equal to 102.42, and is equal to 49.06 for the Haitian Creole, a difference of 53.36. The estimated coefficient in column (3) of table 1.3 implies that the intensity of using French at work will be lower by 3.62 percentage points for the immigrants with a Vietnamese mother tongue than for the immigrants with a Haitian Creole mother tongue. To also illustrate the economic magnitude of the effect of the linguistic distance from English, we take as an example two other immigrants, one with Dutch as a mother tongue, and another with Finnish as a mother tongue. The linguistic distance from English for the immigrant with Dutch as a mother tongue is equal to 60.73, and is equal to 104.24 for the immigrant with Finnish as a mother tongue, a difference of 43.51. The estimated coefficient in column (3) of table 1.3 implies that the intensity of using English at work will be lower by 4.35 percentage points for the immigrants with a Finnish mother tongue than for the immigrants with Dutch as a mother tongue.

I repeated the analysis by replacing the linear equations with an ordered probit model (for the first robustness check noted above), and the results are presented in table 1.4. The estimated coefficients for the linguistic distance variables are all statistically significant and follow the same empirical pattern as the prior model for all four specifications, confirming that immigrants whose mother tongue is more distant from French/English have a lower probability of using French/English at work. Table 1.5 presents the corresponding marginal effects of a unit change in the linguistic distance from French and English (respectively) on the probability of being in one of the language-at-work intensity levels, holding the value of other covariates constant at their means. For all specifications, we discern similar patterns: a unit increase in the distance from French (English) decreases the probability for the event of the intensity levels of using French (English) more often at work.

I also extended my analysis by estimating the predicted probabilities of the trait of having

different intensity levels of using an official language at work corresponding to the range of values of the variables of the distances from French and English. The results are estimated separately by specification and are presented in Figures 1.2 and 1.3, respectively, for the distances from French and English. The results from these graphs support the previous findings. For instance, the graphs in Figure 1.2 show that immigrants with a closer linguistic distance to French have a higher predicted probability of realizing the outcome of using French only at work, and a lower predicted probability of realizing the outcome of not using French at work.[6] The pattern is similar for the effects of the linguistic distance from English on the intensity of using the two official languages at work (Figure 3). Note that the slope of the outcomes profile decreases in the case of the fullest specification that includes the country-of-birth fixed effects. This again shows the importance of controlling for these indicators and omitting them can overestimate the true impact of the linguistic distance variables.

Table 1.6 presents the results of the effect on English to French intensity of the linguistic distance variable when it is specified as a single binary regressor (the third robustness check). The variable "Closer to French" assumes a value of 1 when French is closer to the immigrant's mother tongue than English and a value of 0 otherwise. The estimated coefficients presented in table 1.6 are consistent with prior expectations. The positive coefficient estimates indicate that immigrants whose linguistic distance is closer to French are more likely to use French at work. Again, this effect decreases in magnitude in the fullest specification, but the estimates remain statistically significant.

The results for the socioeconomic control variables are consistent across all models (tables 1.3, 1.4, 1.6). Age has a positive effect on the use of French at work, with the square of

[6] Moving from left to right, the distance of the native tongue from French increases, the predicted probability of the outcome of using French only at work decreases, and the predicted probability of the outcome of not using French at work increases.

age indicating a diminishing effect as age increases. This results might in part suggest that as an immigrant becomes older, he or she becomes less attracted by the additional value that English brings in the labour market. Male immigrants are less likely to use French at work than female immigrants, and the effect of marital status is not consistent across specifications. The coefficient for the presence of children is positive in all specifications, suggesting that immigrants with children are more likely to use French. This is an expected result, as French schooling is mandatory for children of immigrants in Quebec, and therefore immigrants with kids will have higher exposure to learning and eventually using French at work. The results for the years since migration variable are negative across all specifications but are not statistically significant. The negative signs of the estimated coefficients for the educational indicators are highly statistically significant in all specifications, suggesting that immigrants with higher levels of education are more likely to use English than French at work. A potential explanation for this result is that the use of English becomes more important for a job that requires a higher level of education. Immigrants with higher levels of education will generally have to operate in a broader labour market, and the economic return from learning and using the international language is expected to be more important as the level of education increases. In contrast, those with less education might be more likely to restrict activities to the local labour market and use the local official language at a higher intensity at work.

1.6.2 THE IMPACT OF THE LANGUAGES USED AT WORK ON EARNINGS

Table 1.7 presents the regression results of the impact on earnings of the intensity of working in an official language. The coefficient estimates of the English-to-French intensity index are negative in all specifications. They are not statistically significant in the first two specifications, but they are in our fuller specifications (columns 3 and 4) that include countries-

of-birth fixed effects. This interesting result suggests that immigrants in Greater Montreal gain a higher payoff from using English at work than from using French, despite the fact that French is the sole official language of Quebec. For instance, in the third column, the annual earnings disadvantage associated with a change from 0 to 100 in our English to French index (i.e. from working only in English to working only in French) is estimated at 7.3 percent. The corresponding wage penalties in the specifications whose results are contained in the first two columns are much lower in magnitude. This empirical pattern could suggest that the wage premium for using English in the workplace is partly correlated with immigrants' foreign qualifications, such as the education, experience, and skill sets obtained from their source countries. Holding these characteristics fixed across immigrants from different source countries increases the expected return of using the *lingua franca* relative to French. Altogether, those results provide evidence that immigrants in the metropolitan area of Montreal gain higher economic returns from using English at work than from using French.

A potential explanation for this key finding is the importance of English as an international language and also as the main language in the rest of Canada, which gives it an additional value in the workplace. The result suggests that learning and using English in today's globalized economy is important for the economic success of immigrants who work in places where it is necessary to communicate with people outside of the province of Quebec.

The regression results contained in Table 1.8 are based on a set of indicators for each level of the index (i.e. the fourth robustness check) for the following categories: English only (reference), English first and French second, English and French equally, French first and English second, and French only. The results are consistent with the previous findings, but they provide further insights. For all the specifications, immigrants using English first and French

second (i.e. English approximately 75 percent of the time and French approximately 25 percent of the time) are the groups receiving the highest wage premium, and immigrants who use only French at work are the ones receiving the highest wage penalty. The results of the fullest specification show that immigrants who use English at work most of the time and French occasionally benefit from a wage premium of 4.3 percent compared to immigrants who use only English at work. In contrast, immigrants who use only French at work are the group receiving the largest wage penalties, earning 6 percent less than the reference group that uses only English. Those findings confirm that it pays more to use English at work, but that immigrants who are able to alternate to French enjoy the highest wage premium.

In regard to the impacts of the other socioeconomic variables, the results are as expected. For instance, in all the specifications, education has an important effect on earnings; immigrants with higher education levels earn significantly higher wages. The results suggest that English becomes increasingly necessary as one does a job that requires more education, which is likely due to its role as an international language. Age and years since migration have positive effects on earnings, while both variables have negative estimated coefficients for the squared variables, indicating a concave pattern. Male immigrants earn significantly higher wages than their female counterparts, and the effect of marital status is positive.

1.6.3 RESULTS OF THE INSTRUMENTAL VARIABLE ESTIMATION OF THE EARNINGS EQUATION

One important empirical challenge arising from the earnings equation is the potential endogeneity of the language-of-work variable, which could be affected by measurement errors and/or ability bias. To address this problem, I employ the instrumental variable technique. As is well-known, one must select an instrument that is correlated with the language-of-work variable

but also uncorrelated with the error term of the wage equation.

Several variables have been used as instruments in the literature (see Chiswick and Miller, 2015, for a brief survey). This research has shown that the IV estimates of the effects of language skills on earnings are usually higher than their OLS counterparts, suggesting that an attenuation bias stemming from measurement errors dominates the potential ability bias, which one would expect to work in the other direction.[7] However, Chiswick and Miller (2015) also note that in some regression equations estimated by IV, "the coefficient on the instrumented language variable is extremely large, far too large to be believed" (page 242). The large variability of results that are found in the literature are not uncommon for IV estimates.

In my context, candidates for instruments could be the linguistic distance indicators from English and French that I defined earlier. In the analysis presented above, I verified that those linguistic distances from French/English are correlated with the intensity of using French and English at work (i.e. the instrumented variable), which validates the first condition required for identification for the IV estimates. However, since our dependent variable takes two values (English and French), immigrants in this particular case will be likely to choose the language that is closer to their mother tongue, even if both of their linguistic distance are far or close from both official languages. In this case, what would have a higher impact on the choice of the language used at work is which official language is closer to the immigrants' mother tongue. With this possibility in mind, I attempted a series of different regressions using the distance from French and the distance from English as instruments, and I obtained erratic results across specifications. In an attempt to address this challenge, I employ as an instrument my previous indicator, which assumes a value of 1 if the mother tongue of an immigrant is closer to French than to English,

[7] To the extent that able individuals are more likely to invest in learning a second language, the OLS estimates of the impact of the language-at-work variable would be upwardly biased.

and a value of 0 otherwise.

The second required condition is that the instrument (the linguistic distance between an immigrant's mother tongue and a Canadian official language) has no direct impact on earnings. This assumption seems to be satisfied, as we expect linguistic distance to impact immigrants' labour market outcome only indirectly through the channel of the language-of-work variable.

These results are presented in Table 1.9. The coefficient estimates of the English-to-French intensity variable are all statistically significant and negative. The estimated magnitude of the impact on earnings for using French instead of English at work exceeds the magnitude of the one obtained from the OLS regression for all four specifications. Those results show a larger earnings disadvantage for the change from working only in French to working only in English (i.e. the full range), estimated at 66 percent in the specification that includes place-of-birth effects based on the entire sample. Because of the two potential sources of endogeneity bias mentioned above, my prior expectations of the magnitudes of the IV results compared to the OLS ones are not obvious. If we expect that more able immigrants choose to learn English and work in it for whatever reason, the OLS coefficient estimates would overstate the returns to working in English. Since the IV estimates, however, indicate stronger effects than the OLS estimates, the ability bias hypothesis would apply only if the more able immigrants choose to learn and work in French. My estimates militate more to the presence of measurement errors causing attenuation bias. This pattern is consistent with the findings of most of the earlier research - e.g., Dustmann and Van Soest (2001), Bleakley and Chin (2004), and Ginsburgh and Prieto-Rodriguez (2011).

Several diagnostic statistics are presented at the bottom of Table 1.9. To confirm my suspicion that our language-of-work variable is endogenous, I carry out the Wooldridge's score

test, which is designed to account for robust standard errors. The resulting test statistics are significant for all of our specifications, confirming that the language-of-work variable is indeed endogenous.

I investigate the possibility that my instrument is weak by reporting the F-statistic of the first stage regression. The values of the F-statistics across the first, second and third (our preferred) specifications are larger than the conventional benchmark of 10, verifying that the instrument is strong. However, when I omit Francophone and Anglophone immigrants in the fourth specification, the value of the F-statistic falls below 10. It is not surprising that the instrument loses its strength in this last specification, as those individuals comprise 28 % of the estimating sample. The range of the instrument and its coefficient of variation are decreased as a result of the omission of individuals whose linguistic distances from one of the official languages are zero. The results also suggest that the linguistic distance of 0 has an important effect on the choice of using English or French at work. This is in line with our conceptual framework, for instance the assertion that immigrants with a linguistic distance of 0 from French will have no cost of learning and using French at work and will have a high probability of choosing to use French at work instead of English, even though the latter can provide higher economic return. In summary of my analysis of the earnings equations, the OLS and IV results are qualitatively similar; immigrants in the Greater Montreal area reap a higher payoff of using English at work, even though French is the official language of the province of Quebec.

1.7 CONCLUSION

The linguistic dimension of immigrants' integration in the Canadian labour market is an important factor in determining their economic success, and their capacity to use the host country's languages plays an important role. The linguistic dynamics in the region of Greater Montreal represent an interesting case of an environment where both Canadian official languages are widely used, and where immigrants are an increasingly important component of the labour force. Based on data drawn from the 2001 and 2006 Canadian censuses, from the 2011 National Household Survey, and from the ASJP database, I have investigated the relationships between linguistic distance and the intensity of use of English and French at work by immigrants in the Montreal metropolitan area. I found that the linguistic distances between immigrants' mother tongues and both English and French have an important impact on the language choices at work. I also investigated the role of the languages used at work on the earnings of immigrants. I found that these immigrants gain higher payoffs from using English at work than from using French, and that the higher payoff occurs for jobs where English is used most often, and French is used on an occasional basis. The results confirm that the command of the international *lingua franca* is an important factor in the economic success of immigrants in Canada's largest primary French-speaking city, as reflected in their wage levels are presumably through their marginal products.

In regards to policy repercussions, favouring the selection of immigrants who have a good knowledge of French while also knowing some English could be a suitable strategy to ensure the sustainability of the French language, while at the same time facilitating their assimilation. Over the past decades, various legislations (Bill 63, Bill 22 and Bill 101) helped the province of Quebec promote the preservation of the French language, and many of those laws

impacted the integration of immigrants directly (such as the mandatory schooling in French). Nevertheless, the presence of English as an international language remains an important ingredient for success in the Montreal labour market. The challenge for the province is to find an appropriate good balance between the preservation of the French language and the economic benefits that its citizens can gain by being able to learn and use English.

References 1

Adserà, Alicia and Ana Ferrer. 2015. "The Effect of Linguistic Proximity on the Occupational Assimilation of Immigrant Men in Canada." CreAM Discussion Paper Series 1517, Centre for Research and Analysis of Migration (CreAM), Department of Economics, University College London.

Adserà, Alicia and Mariola Pytlikova. 2015. "The Role of Language in Shaping International Migration." *Economic Journal*, 125 (586): F49–F81.

Albouy, David. 2008. "The Wage Gap Between Francophones and Anglophones: a Canadian Perspective, 1970-2000." *Canadian Journal of Economics*, 41(4): 1211-1238.

Bleakley, Hoyt and Aimee Chinn. 2004. "Language Skills and Earnings: Evidence from Childhood Immigrants." *The Review of Economics and Statistics*, 86 (2): 481-496.

Borjas, George J. 2003. "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labour Market." *The Quarterly Journal of Economics*, 118(4): 1335-1374.

Brutt-Griffler, Janina. 2002. "World English: A Study of its Development." Clevedon: Multilingual Matters.

Carliner, Geoffrey. 1981. "Wage Differences by Language Group and the Market for Language Skills in Canada." *Journal of Human Resources*, 16(3): 384-399.

Chiswick, Barry R. and Paul W. Miller. 2005. "Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages." *Journal of Multilingual and Multicultural Development*, 26 (1): 1-11.

Chiswick, Barry R. and Paul W. Miller. 2015. "International Migration and the Economics of Language." In Barry R. Chiswick and Paul W. Miller, *Handbook of The Economics of International Migration*, Volume 1A, The Immigrants, Elsevier, Oxford, chapter 5, 211-269.

Dustmann, Christian. and Arthur van Soest. 2001. "Language Fluency and Earnings: Estimation with Misclassified Language Indicators" *The Review of Economics and Statistics*, 83(4): 663-674.

Christofides, Louis N. and Robert Swidinsky. 2010. "The Economic Returns to the Knowledge and Use of a Second Official Language: English in Quebec and French in the Rest-of-Canada." *Canadian Public Policy*, 36(2): 137-158.

Ginsburgh Victor A. and Juan Prieto-Rodriguez. 2011. "Returns to Foreign Languages of Native Workers in the EU." *Industrial and Labor Relations Review* 64 (3): 599–618.

Ginsburgh, Victor and Shlomo Weber. 2016. "Linguistic Distances and Ethnolinguistic Fractionalization and Disenfranchisement Indexes." In Victor Ginsburgh and Shlomo Weber, The *Palgrave handbook of economics and language*, Palgrave Macmillan, New York, chapter 5: 137-173.

Grenier, Gilles. 2015. "The value of language skills." *IZA World of Labor*, (205):1-10.

Grenier, Gilles and Serge Nadeau. 2016. "English as the *Lingua Franca* and the Economic Value of Other Languages: the Case of the Language of Work in the Montreal Labour Market." In Michele Gazzola and Bengt-Arne Wickström, *The Economics of Language Policy*, The MIT Press, Cambridge, Mass, chapter 8, 267-312.

Gunduz, Seda. 2017. "The Impact of Immigrant Language Skills on Canadian Wages" Doctoral dissertation, University of Ottawa, retrieved from https://ruor.uottawa.ca/bitstream/10393/36624/1/Gunduz_Seda_2017_thesis.pdf

Isphording, Ingo E. and Sebastian Otten. 2013. "The Costs of Babylon—Linguistic Distance in Applied Economics." *Review of International Economics*, 21(2): 354-369.

Isphording, Ingo E. and Sebastian Otten. 2014. "Linguistic Barriers in the Destination Language Acquisition of Immigrants." *Journal of Economic Behavior & Organization*, 105(1): 30-50.

Jenkins, Jennifer. 2007. "English as a Lingua Franca: Attitude and Identity." Oxford University Press.

Li, Peter S. and Chunhong Dong. 2007. "Earnings of Chinese Immigrants in the Enclave and Mainstream Economy." *Canadian Review of Sociology*, 44(1): 65–99.

OECD. 2016. "Canada – Economic Forecast Summary." http://www.oecd.org/canada/canada-economic-forecast-summary.html

Ottaviano, Gianmarco I.P. and Giovanni Peri. 2012 "Rethinking the Effect of Immigration on Wages." *Journal of the European Economic Association*, 10 (1): 152–197.

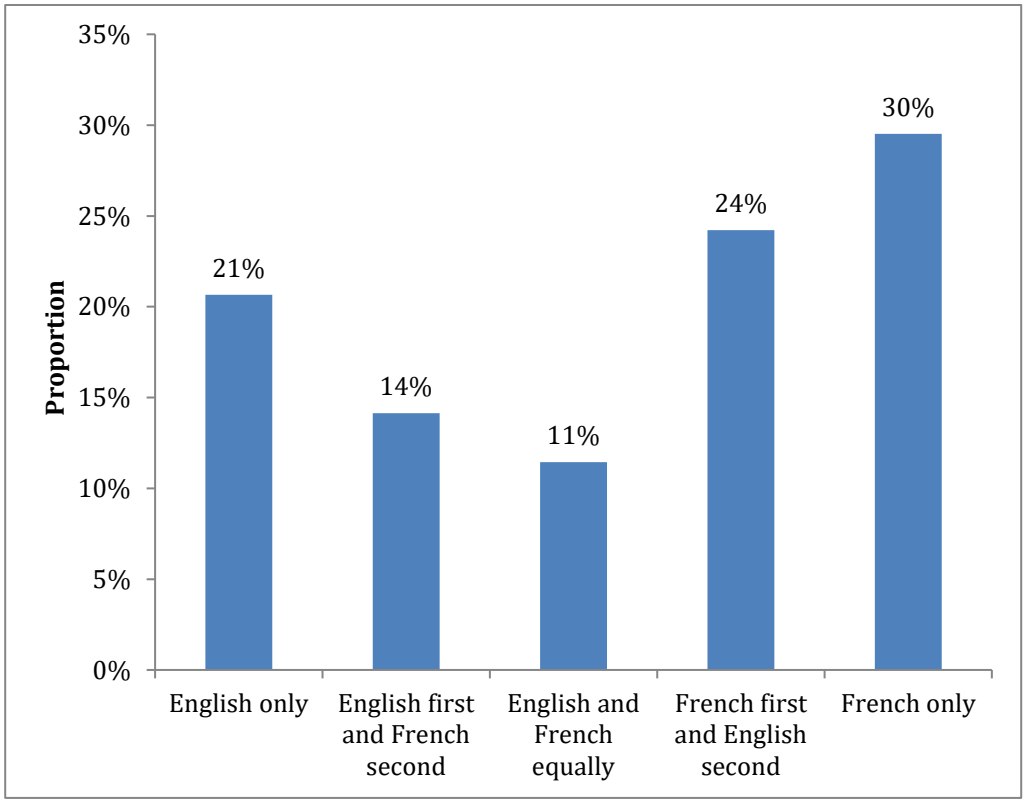
Petroni, Filippo. and Maurizio Serva. 2010 "Measures of Lexical Distance Between Languages." *Physica A: Statistical Mechanics and its Applications*, 389 (11): 2280–2283.

Shapiro, Daniel and Morton Stelcner. 1997. "Language and Earnings in Quebec: Trends Over Twenty Years, 1970-1990." *Canadian Public Policy*, 23 (2): 115-140.

Swadesh, Morris. 1952. "Lexicostatistical Dating of Prehistoric Ethnic Contracts." *Proceedings of the American Philosophical Society*, 96:121–137.

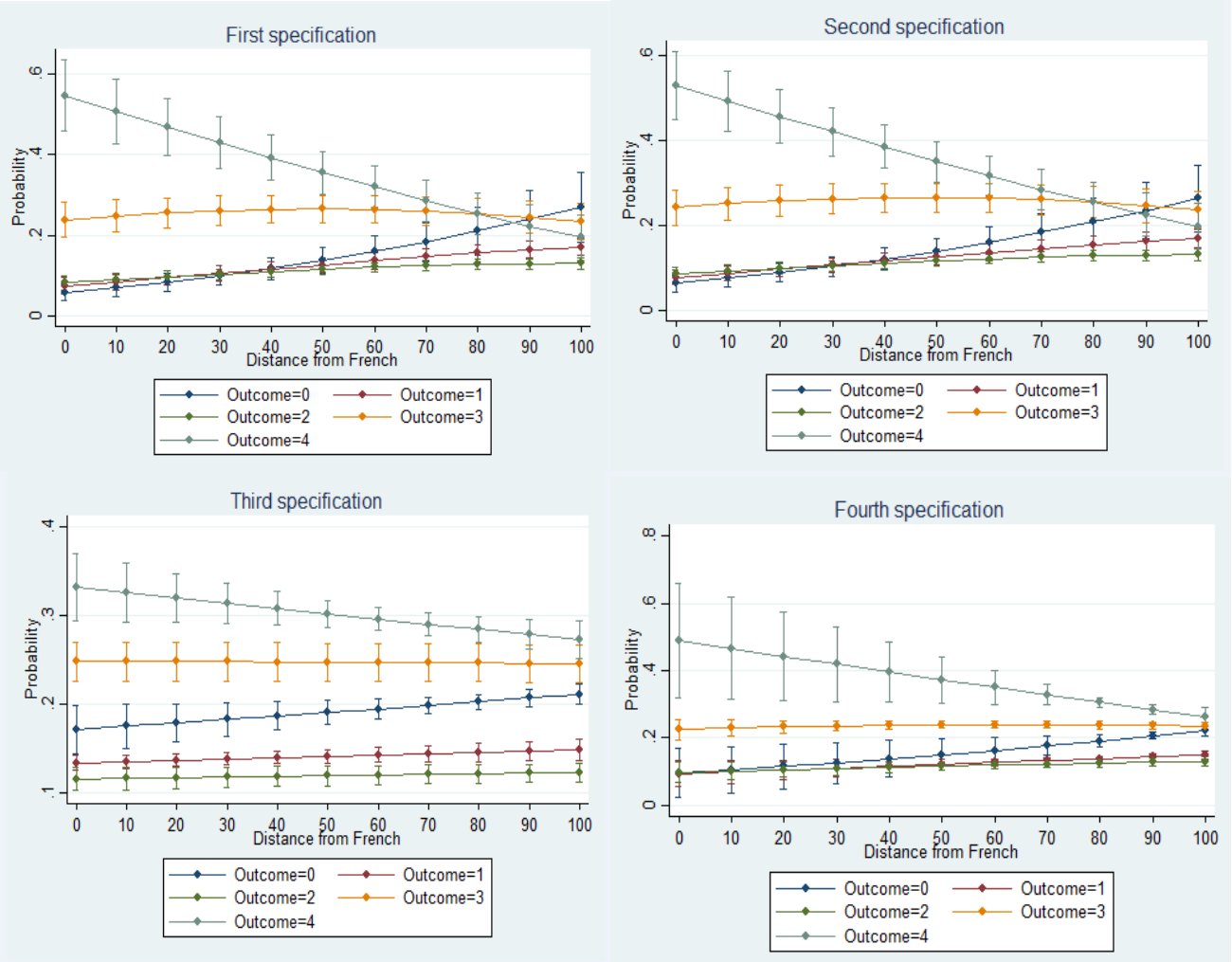
Vaillancourt, François. 1980. "Differences in Earnings by Language Groups in Quebec, 1970: An Economic Analysis." International Center for Research and Bilingualism.

Figure 1.1: Proportion of immigrants using English and French at work, Montreal CMA, 2001, 2006 and 2011



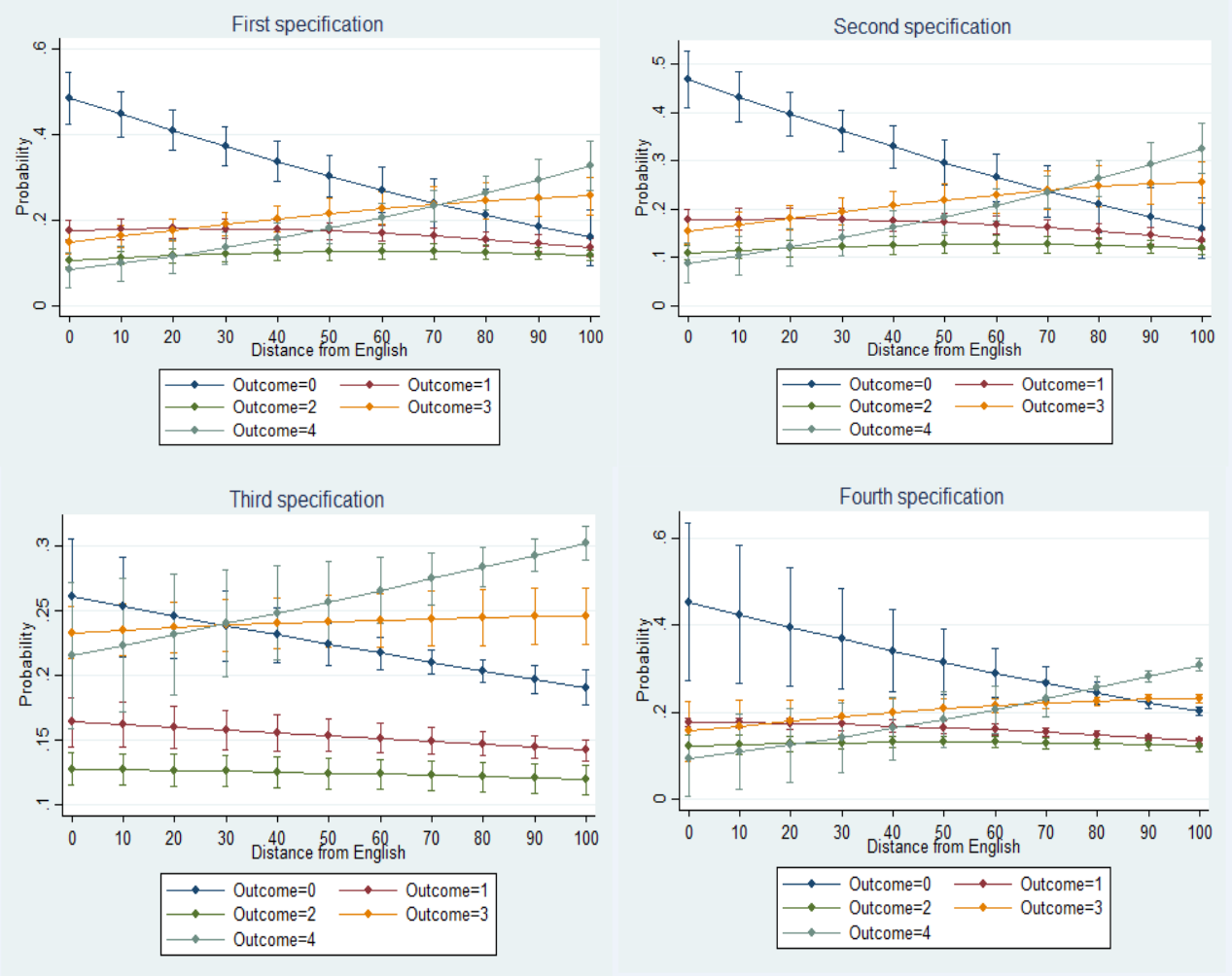
Note: Figure 1.1 shows the proportions of immigrants using English and/or French at work for the five nodes. This is a graphical illustration of our English-to-French intensity index, which is interpreted as the approximate amount of time that an immigrant uses an official language (English or French) at work. English only corresponds to a value of 0, English first, French second corresponds to a value of 25. English and French equally maps to a value of 50. French first-English second maps to a value of 75. French only corresponds to a value of 100.

Figure 1.2: Predicted probabilities generated by the ordered probit model for the events of immigrants being at different intensity levels of using French relative to English as the linguistic distance varies from French (left to right means further from French)



Notes: Figure 1.2 shows how the probabilities of being at different intensity level of using an official language at work change as the linguistic distance from French varies while holding all other covariates at their means. **Outcome 0 is when only English is used at work, outcome 1 is when English is used first and French second, outcome 2 is when English and French are used equally, outcome 3 is when French is used first and English second and outcome 4 is when French is used only at work.** The first specification includes controls for age, age-squared, presence of a child, marital status, gender, years- since- migration, years-since-migration squared, education, census year. The second specification adds controls for location of work within Montreal, previous residence outside Quebec 1 or 5 years earlier and industry-fixed effects. The third specification adds place-of-birth fixed effects. The fourth specification is similar to the third specification but excludes immigrants whose mother tongue is French or English.

Figure 1.3: Predicted probabilities generated by the ordered probit model for the events of immigrants being at different intensity levels of using French relative to English as the linguistic distance varies from English (left to right means further from English)



Note: Figure 1.3 shows how the probabilities of being at different intensity levels of using an official language at work change as the linguistic distance varies from English, while holding all other covariates at their means. **Outcome 0 is when only English is used at work, outcome 1 is when English is used first and French second, outcome 2 is when English and French are used equally, outcome 3 is when French is used first and English second, and outcome 4 is when French is used only at work.** The first specification includes controls for age, age-squared, presence of a child, marital status, gender, years- since- migration, years-since-migration squared, education, census year. The second specification adds controls for location of work within Montreal, previous residence outside Quebec 1 or 5 years earlier and industry-fixed effects. The third specification adds place-of-birth fixed effects. The fourth specification is similar to the third specification but excludes immigrants whose mother tongue is French or English.

Table 1.1

Descriptive Statistics of major variables, Immigrants in Montreal CMA, 2001, 2006 and 2011

	Means (Standard deviations)
Linguistic distance from French	71.6
	(36.9)
Linguistic distance from English	85.9
	(28.1)
Age	45.2
	(9.7)
Annual wage	47640
	(41382)
Years since migration	24.5
	(11.6)
At least one child	0.65
Married	0.68
Male	0.59
Education:	
High school dropout	0.13
High school degree	0.16
Some post-secondary	0.33
Bachelor's degree	0.19
Graduate degree	0.20
Sample Size	67,440

Note: The Linguistic distance from English ranges from 0 to 104.2; the linguistic distance from French ranges from 0 to 104.0. The linguistic distances have been calculated using version 2.2 of the ASJP distance matrices. Annual wage include wages and salaries and do not include self-employment income and are converted to 2011 constant dollars using the CPI for Canada. Standard deviations are in parentheses.

Table 1.2

Top 5 mother tongues, Immigrants in Montreal CMA, 2001, 2006 and 2011

Mother tongues	Proportion	Distance from French	Distance from English
French	0.1888	0	91.6
Arabic	0.1211	97.8	98.6
Spanish	0.0966	84.0	93.3
English	0.0936	91.6	0
Haitian Creole	0.0769	49.1	93.6

Note: The linguistic distances have been calculated with the ASJP distance matrices, version 2.2.

Table 1.3: Ordinary least squares regression of the English-to-French (at work) intensity: impact of linguistic distance variables

	(1)	(2)	(3)	(4)
Linguistic distance from French	-0.338***	-0.320***	-0.068**	-0.239***
	(0.0587)	(0.0538)	(0.0270)	(0.0904)
Linguistic distance from English	0.337***	0.320***	0.100***	0.295***
	(0.051)	(0.049)	(0.034)	(0.094)
Demographic and socioeconomic controls:				
Male	-2.246	-0.968	-3.163***	-3.876***
	(1.550)	(1.262)	(0.450)	(0.517)
Children	1.454*	1.023	0.877***	1.032***
	(0.838)	(0.713)	(0.320)	(0.356)
Age	1.717***	1.573***	0.763***	0.742***
	(0.413)	(0.401)	(0.186)	(0.218)
Age squared	-0.017***	-0.015***	-0.006***	-0.005**
	(0.004)	(0.004)	(0.002)	(0.002)
Years since migration	-0.053	-0.110	-0.133	-0.193
	(0.370)	(0.367)	(0.133)	(0.164)
Years since migration squared	0.00028	0.00059	0.00064	0.00096
	(0.006)	(0.006)	(0.002)	(0.003)
Married	-2.024	-1.508	0.024	0.671
	(1.366)	(1.254)	(0.422)	(0.471)
High school degree	-8.158***	-7.744***	-2.804***	-3.458***
	(2.471)	(2.325)	(0.986)	(0.920)
Some post-secondary	-4.639	-4.939	-2.725**	-3.085**
	(3.210)	(3.024)	(1.311)	(1.276)
Bachelor's degree	-9.442**	-9.453**	-5.712***	-6.008***
	(3.827)	(3.668)	(1.671)	(1.490)
Graduate degree	-12.545***	-12.917***	-8.555***	-8.718***
	(3.905)	(3.583)	(1.632)	(1.654)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Country-of-birth fixed effects	NO	NO	YES	YES
Observations	67,440	67,440	67,440	48,050
R-squared	0.202	0.241	0.45	0.414

Note: The dependent variable is the English-to-French language of work intensity. All regressions are weighted using Census weights. Columns (1), (2) and (3) are based on the entire sample. Column (4) excludes immigrants whose mother tongue is either French or English. Robust standard errors in parentheses are clustered at the country-of-birth level. Significance levels: *** at 1%, ** at 5%, * at 10%

Table 1.4: Ordered probit regression of the English-to-French intensity (at work): Impact of Linguistic Distance Variables

	(1)	(2)	(3)	(4)
Linguistic distance from French	-0.010***	-0.009***	-0.002*	-0.007*
	(0.001)	(0.001)	(0.0009)	(0.003)
Linguistic distance from English	0.010***	0.010***	0.003**	0.010**
	(0.002)	(0.002)	(0.001)	(0.003)
Demographic and Socioeconomic controls:				
Male	-0.093	-0.049	-0.139***	-0.165***
	(0.052)	(0.042)	(0.019)	(0.022)
Children	0.054	0.040	0.042**	0.048***
	(0.029)	(0.026)	(0.014)	(0.013)
Age	0.053***	0.049***	0.027***	0.025**
	(0.013)	(0.012)	(0.007)	(0.008)
Age squared	-0.0005***	-0.0004***	-0.0002*	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Years since migration	-0.002	-0.004	-0.005	-0.007
	(0.012)	(0.012)	(0.005)	(0.006)
Years since migration squared	0.00001	0.00002	0.00003	0.00003
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Married	-0.065	-0.050	-0.0002	0.023
	(0.043)	(0.040)	(0.017)	(0.018)
High school degree	-0.281***	-0.272***	-0.133**	-0.159***
	(0.083)	(0.080)	(0.049)	(0.042)
Some post-secondary	-0.175	-0.191	-0.134*	-0.156*
	(0.109)	(0.106)	(0.066)	(0.061)
Bachelor's degree	-0.333*	-0.346**	-0.253**	-0.268***
	(0.132)	(0.131)	(0.084)	(0.072)
Graduate degree	-0.440**	-0.473***	-0.375***	-0.375***
	(0.137)	(0.132)	(0.082)	(0.075)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Country-of-Birth fixed effects	NO	NO	YES	YES
Threshold 1	0.216	0.0025	-0.995***	-1.04*
	(0.305)	(0.315)	(0.178)	(0.436)
Threshold 2	0.710*	0.51	-0.353	-0.445
	(0.310)	(0.316)	(0.181)	(0.434)
Threshold 3	1.054***	0.866**	0.0844	-0.002
	(0.316)	(0.320)	(0.182)	(0.435)
Threshold 4	1.767***	1.603***	0.932***	0.792
	(0.326)	(0.321)	(0.190)	(0.442)
Observations	67,440	67,440	67,440	48,050
Pseudo R-squared	0.067	0.083	0.173	0.16

Note: The dependent variable is the English- to-French language at work intensity, and all the regressions are weighted using Census weights. Columns (1), (2) and (3) are based on the entire sample. Column (4) excludes immigrants whose mother tongue is French or English. Robust standard errors in parentheses are clustered at the countries of birth level. Significance levels: *** at 1%, ** at 5%, * at 10%.

Table 1.5: Ordered probit regression of the different levels of language intensity (at work): impact of linguistic distance variables

	English Only	English first, French second	English and French equally	French first, English second	French only
Specification 1					
Distance from French	0.00264***	0.00106***	0.00036***	-0.00069***	-0.00337***
	(0.0006)	(0.0001)	(0.0001)	(0.0002)	(0.0005)
Distance from English	-0.00266***	-0.00107***	-0.00036**	0.00070***	0.00339***
	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0007)
Specification 2					
Distance from French	0.00250***	0.00108***	0.00037***	-0.00071***	-0.00323***
	(0.0006)	(0.0001)	(0.0001)	(0.0002)	(0.0005)
Distance from English	-0.00254***	-0.00109***	-0.00037**	0.00073***	0.00328***
	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0007)
Specification 3					
Distance from French	0.00045**	0.00031**	0.00010**	-0.00023**	-0.00063**
	(0.0002)	(0.0001)	(0.0000)	(0.0001)	(0.0003)
Distance from English	-0.00071***	-0.00049***	-0.00016***	0.00036**	0.00099***
	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0004)
Specification 4					
Distance from French	0.00170**	0.00101**	0.00033**	-0.00075**	-0.00229**
	(0.0007)	(0.0004)	(0.0001)	(0.0003)	(0.0010)
Distance from English	-0.00228***	-0.00136***	-0.00044***	0.00101***	0.00307***
	(0.0008)	(0.0005)	(0.0002)	(0.0004)	(0.0010)

The first specification includes control for age, age-squared, marital status, gender, year-since-migration, years-since-migration squared, education, and census year. The second specification adds controls for location of work within Montreal, previous residence outside Quebec 1 or 5 years earlier, and industry-fixed effects. The third specification adds country-of-birth fixed effects. The fourth specification is similar to the third specification, but native speakers of English or French are omitted. The dependent variable is the English-to-French language of work intensity, and the regressions are weighted using Census weights. Robust standard errors in parentheses are clustered at the country-of-birth level (186 clusters). Significance levels: *** at 1%, ** at 5%, * at 10%.

Table 1.6: Ordinary least squares regression of the English-to-French intensity (at work); linguistic distance measured as dummy variable

	(1)	(2)	(3)	(4)
Closer to French	28.942***	26.982***	9.342***	3.876*
	(5.329)	(5.113)	(2.774)	(2.332)
Demographic and Socioeconomic controls:				
Male	-2.483	-1.057	-3.206***	-3.877***
	(1.940)	(1.522)	(0.449)	(0.518)
Children	1.841	1.401	0.878***	1.035***
	(1.287)	(1.081)	(0.318)	(0.358)
Age	1.514***	1.388***	0.742***	0.745***
	(0.524)	(0.507)	(0.183)	(0.219)
Age-squared	-0.015***	-0.013***	-0.006***	-0.005**
	(0.005)	(0.005)	(0.002)	(0.002)
Years-since-migration	-0.362	-0.424	-0.129	-0.191
	(0.364)	(0.369)	(0.131)	(0.164)
Years-since-migration squared	0.005	0.005	0.0007	0.0009
	(0.006)	(0.006)	(0.002)	(0.003)
Married	-5.811***	-4.957**	-0.070	0.680
	(2.197)	(2.006)	(0.442)	(0.470)
High school degree	-7.652**	-7.422***	-2.659***	-3.429***
	(2.966)	(2.792)	(0.990)	(0.920)
Some post-secondary	-0.891	-1.779	-2.521*	-3.074**
	(3.875)	(3.667)	(1.317)	(1.274)
Bachelor's degree	-7.304	-8.163*	-5.561***	-6.006***
	(4.545)	(4.411)	(1.672)	(1.489)
Graduate degree	-7.903	-9.535**	-8.311***	-8.720***
	(5.207)	(4.698)	(1.645)	(1.650)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Country-of-Birth fixed effects	NO	NO	YES	YES
Observations	67,440	67,440	67,440	48,050
R-squared	0.10	0.15	0.449	0.414

Note: The dependent variable is the English-to-French language intensity (at work), and all regressions are weighted using Census weights. Columns (1), (2) and (3) are based on the entire sample. Column (4) excludes immigrants whose mother tongue is either French or English. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance levels: *** at 1%, ** at 5%, * at 10%.

Table 1.7: Ordinary least squares regression of the wage equation: Impact of the English-to-French Intensity (at Work) variable

	(1)	(2)	(3)	(4)
English to French intensity	-0.00027	-0.00014	-0.00073***	-0.00064***
	(0.0004)	(0.0004)	(0.0002)	(0.0002)
Demographic and Socioeconomic controls:				
Male	0.236***	0.234***	0.231***	0.243***
	(0.015)	(0.013)	(0.012)	(0.013)
Children	-0.0057	-0.004	0.012*	0.009
	(0.009)	(0.009)	(0.007)	(0.008)
Age	0.041***	0.041***	0.043***	0.038***
	(0.004)	(0.004)	(0.005)	(0.004)
Age squared	-0.0004***	-0.0004***	-0.0004***	-0.0004***
	(0.000)	(0.000)	(0.000)	(0.000)
Years since migration	0.019***	0.014***	0.018***	0.019***
	(0.003)	(0.002)	(0.002)	(0.002)
Years since migration squared	-0.00007	-0.00002	-0.0001***	-0.0001**
	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.004	0.003	0.014*	0.015*
	(0.015)	(0.015)	(0.008)	(0.009)
High school degree	0.131***	0.112***	0.096***	0.094***
	(0.016)	(0.014)	(0.010)	(0.013)
Some post-secondary	0.339***	0.286***	0.245***	0.242***
	(0.022)	(0.023)	(0.013)	(0.015)
Bachelor's degree	0.607***	0.492***	0.455***	0.440***
	(0.024)	(0.024)	(0.021)	(0.023)
Graduate degree	0.851***	0.699***	0.638***	0.621***
	(0.031)	(0.031)	(0.023)	(0.028)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Country-of-Birth fixed effects	NO	NO	YES	YES
Observations	67440	67440	67440	48050
R-squared	0.2	0.24	0.27	0.26

Note: The dependent variable is annual earnings converted to 2011 constant dollars (with the CPI), and all the regressions are weighted using Census weights. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 1.8: Ordinary least squares regression of wage equation: impact of language intensity (at work) specified as set of categorical variables

Official Language used at work (ref: English only)				
	(1)	(2)	(3)	(4)
English first and French second	0.083***	0.087***	0.043**	0.015
	(0.027)	(0.023)	(0.020)	(0.021)
English and French equally	-0.003	0.014	-0.017	-0.017
	(0.037)	(0.031)	(0.018)	(0.021)
French first and English second	0.03	0.053	-0.011	-0.023
	(0.045)	(0.039)	(0.019)	(0.021)
French only	-0.013	-0.004	-0.060***	-0.058***
	(0.040)	(0.036)	(0.018)	(0.019)
Demographic and Socioeconomic controls:				
Male	0.235***	0.232***	0.231***	0.243***
	(0.015)	(0.013)	(0.012)	(0.013)
Children	-0.005	-0.003	0.013*	0.01
	(0.009)	(0.009)	(0.007)	(0.008)
Age	0.042***	0.042***	0.043***	0.039***
	(0.004)	(0.004)	(0.005)	(0.004)
Age squared	- 0.0001***	-0.0001***	-0.0001***	-0.0001***
	(0.000)	(0.000)	(0.000)	(0.000)
Years since migration	0.019***	0.014***	0.018***	0.019***
	(0.003)	(0.002)	(0.002)	(0.002)
Years since migration squared	-0.0001	-0.0001	-0.0001***	-0.0001**
	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.006	0.004	0.015*	0.015*
	(0.015)	(0.015)	(0.008)	(0.009)
High school degree	0.126***	0.106***	0.093***	0.092***
	(0.016)	(0.014)	(0.010)	(0.012)
Some post-secondary	0.332***	0.276***	0.239***	0.239***
	(0.022)	(0.022)	(0.014)	(0.015)
Bachelor's degree	0.595***	0.477***	0.446***	0.436***
	(0.025)	(0.025)	(0.021)	(0.023)
Graduate degree	0.839***	0.682***	0.628***	0.616***
	(0.029)	(0.029)	(0.023)	(0.028)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Country-of-Birth fixed effects	NO	NO	YES	YES
Observations	67,440	67,440	67,440	48,050
R-squared	0.20	0.24	0.27	0.26

Note: The dependent variable is annual earnings converted to 2011 constant dollars using the CPI. All the regressions are weighted using Census weights. Robust standard errors in parentheses are clustered at the country-of-birth level (186 clusters). Significance levels: *** at 1%, ** at 5%, * at 10%.

Table 1.9: Instrumental variables regression of wage equation: impact of English-to-French (at work) intensity with "closer to French" indicator employed as instrument

	(1)	(2)	(3)	(4)
English to French intensity	-0.0031*	-0.0028*	-0.0066*	-0.0152*
	(0.002)	(0.002)	(0.004)	(0.009)
Demographic and Socioeconomic controls:				
Male	0.232***	0.233***	0.212***	0.187***
	(0.0178)	(0.0139)	(0.0139)	(0.0340)
Children	0.0009	0.0005	0.018**	0.024**
	(0.008)	(0.008)	(0.008)	(0.012)
Age	0.045***	0.045***	0.047***	0.049***
	(0.004)	(0.004)	(0.004)	(0.007)
Age squared	-0.0005***	-0.0004***	-0.0005***	-0.0005***
	(0.000)	(0.000)	(0.000)	(0.000)
Years since migration	0.018***	0.014***	0.017***	0.016***
	(0.003)	(0.003)	(0.002)	(0.003)
Years since migration squared	-0.00006	-0.00001	-0.00010***	-0.00010*
	(0.000)	(0.000)	(0.000)	(0.000)
Married	-0.010	-0.008	0.013	0.025**
	(0.021)	(0.019)	(0.009)	(0.013)
High school degree	0.105***	0.088***	0.080***	0.043
	(0.026)	(0.024)	(0.014)	(0.034)
Some post-secondary	0.332***	0.277***	0.229***	0.197***
	(0.031)	(0.031)	(0.014)	(0.035)
Bachelor's degree	0.581***	0.466***	0.421***	0.352***
	(0.033)	(0.032)	(0.027)	(0.058)
Graduate degree	0.821***	0.668***	0.587***	0.493***
	(0.047)	(0.046)	(0.035)	(0.082)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Country-of-Birth fixed effects	NO	NO	YES	YES
IV Test:				
Wooldridge's score test adjusted for clusters	3.74*	3.49*	5.55**	7.71***
F-statistic from first stage adjusted for clusters	29.50***	27.85***	11.34***	2.76*
Observations	67,440	67,440	67,440	48,050
R-squared	0.18	0.22	0.22	.

Note: The dependent variable is annual earnings converted to 2011 constant dollars with the CPI; all the regressions are weighted using Census weights; Robust standard errors in parentheses are clustered at the country-of-birth level (186 clusters). Significance levels: *** at 1%, ** at 5%, * at 10%

APPENDIX 1.1

Calculation of the Levenshtein Linguistic Distance

The following explanation of the computation of the linguistic distance is based on the work of Petroni and Serva (2010). The ASJP linguistic distance is computed by using a list of 40 words in each language having similar meanings. The list includes, for example, words describing body parts, animals, plants, nature, verbs, adjectives, and pronouns that are used universally across languages. It was originally based on the 100-item "Swadesh list" (Swadesh, 1952), but was reduced to 40 items that were shown to suffice. To calculate the distances, the lexical similarities of all pairings of languages are compared using an algorithm called the "Levenshtein distance" (LD), which is calculated as the minimum number of edits (deletions, substitutions or insertions) required to transform a word from one language into another. To provide a very simple illustration, the Levenshtein distance between the French word "allo" to its corresponding English word "hello" is equal to two, the transformation of one word into the other cannot be effectuated with fewer than two edits.

1. allo hllo (substitution of 'a' with 'h')
2. hllo hello (insert 'e' after 'h')

A normalized measure of the Levenshtein distance (LDN) needs to be provided in order to account for the word lengths, because longer words inherently require more edits to be executed. The normalization is performed by dividing the LD between similar words in two different languages by the number of characters of the longer of the words in whichever language applies. The LDN between the words with meaning *i* in two languages labelled Q and W is equal to:

$$LDN(Q_i, W_i) = \frac{LD(Q_i, W_i)}{L(Q_i W_i)}$$

where $LD(Q_i, W_i)$ is the Levenshtein distance between Q_i and W_i, and $L(Q_i W_i)$ is the number of

characters of the longer word. The total linguistic distance (involving all words) between a pair of languages is then calculated by measuring the average distance of all n words for those languages as follows.

$$LDN(Q,W) = \frac{1}{n}\sum_{i=1}^{n} LDN(Q_i, W_i)$$

where Q_i and W_i correspond to the word i in languages Q and W. Finally, to account for word lexical similarity resulting from merely pure coincidence (as opposed to pure etymology), the program provides a further normalized measure labelled the Levenshtein distance normalized divided (LDND) between pairs of languages. It is obtained by dividing the $LDN(Q,W)$ by the "global distance", where the "global distance" (GD) is the average distance between two languages using only pairs of words with different meanings. This quantity is expressed as:

$$GD(Q,W) = \frac{1}{n(n-1)}\sum_{i\neq k}^{n} LDN(Q_i, W_k)$$

The LDND is the final measure of linguistic distance, which is obtained by dividing the $LDN(Q,W)$ between pairs of languages with their respective values of $GD(Q,W)$. It is employed in our empirical analysisand is written as:

$$LDND(Q,W) = \frac{LDN(Q,W)}{GD(Q,W)}$$

APPENDIX 1.2

Table A.1: Detailed Results for Industry and Country-of-Birth Fixed Effects

	(1)	(2)	(3)	(4)
Dependent variables	**EtoF index**	**EtoF index**	**Ln (wage)**	**Ln (wage)**
Distance from French	-0.06855**			
	(0.02699)			
Distance from English	0.10024***			
	(0.03421)			
Closer to French		9.34298***		
		(2.77411)		
English to French Intensity			-0.00073***	-0.00661*
			(0.00020)	(0.00370)
Industry (ref: Manufacturing)				
Agriculture,forestry,fishing and hunting	12.67562***	12.73996***	-0.25191***	-0.17686**
	(2.64799)	(2.63717)	(0.08488)	(0.08763)
Mining and oil and gas extraction	-8.37298*	-8.33560*	0.63461***	0.58859***
	(4.67710)	(4.67106)	(0.11077)	(0.11449)
Utilities	13.95556***	14.02572***	0.39521***	0.47810***
	(2.40204)	(2.37520)	(0.04385)	(0.06718)
Construction	6.51504***	6.54915***	-0.01747	0.02031
	(1.14879)	(1.15466)	(0.02869)	(0.03946)
Wholesale trade	-4.17248***	-4.09330***	0.02406	-0.00042
	(0.69202)	(0.69452)	(0.01790)	(0.02168)
Retail trade	4.41908***	4.46739***	-0.23028***	-0.20358***
	(1.16918)	(1.16855)	(0.02906)	(0.02970)
Transportation and warehousing	-4.30481***	-4.30108***	-0.02600	-0.05180
	(0.97105)	(0.97702)	(0.02843)	(0.03214)
Information and cultural industries	-2.71143**	-2.56498**	0.18339***	0.16807***
	(1.23777)	(1.24500)	(0.02296)	(0.02093)
Finance and insurance	-1.41520	-1.32513	0.17648***	0.16862***
	(1.14321)	(1.16616)	(0.02605)	(0.02536)
Real estate and rental leasing	1.33878	1.44964	-0.07762**	-0.06949*
	(1.38378)	(1.38738)	(0.03721)	(0.03922)
Professional, scientific and technical services	-4.16519***	-4.08257***	0.17154***	0.14736***
	(0.73699)	(0.74422)	(0.02643)	(0.02522)
Management of companies and enterprises	-2.99369	-2.94578	0.21244***	0.19393***
	(2.66476)	(2.70459)	(0.07089)	(0.06807)
Administrative and support, waste management and remediation services	2.33279**	2.39020**	-0.22287***	-0.20939***
	(1.06568)	(1.08203)	(0.02409)	(0.02555)
Educational services	2.03615	2.07945	0.04711*	0.05892*
	(1.85727)	(1.87810)	(0.02738)	(0.03412)

Health care and social assistance	5.06820***	5.09043***	-0.00788	0.02186
	(1.36073)	(1.36488)	(0.04110)	(0.04233)
Arts, entertainment and recreation	6.92555***	7.07114***	-0.16982***	-0.12856**
	(1.67608)	(1.69913)	(0.04606)	(0.05735)
Accommodation and food services	4.71900***	4.74106***	-0.30143***	-0.27377***
	(1.34806)	(1.34113)	(0.03051)	(0.03934)
Other services (except public administration)	0.83890	0.83600	-0.25324***	-0.24873***
	(1.15035)	(1.14795)	(0.02672)	(0.02672)
Public administration	10.26980***	10.33378***	0.14274***	0.20378***
	(1.78796)	(1.79917)	(0.02668)	(0.04895)
Place of birth (ref: Italy)				
Saint Pierre and Miquelon	17.49431***	22.86188***	0.27645***	0.41213***
	(2.01307)	(0.68977)	(0.01124)	(0.08879)
United States of America	-24.90785***	-24.63274***	0.22403***	0.03339
	(2.84771)	(2.67613)	(0.01722)	(0.12104)
Belize	-42.71505***	-43.36118***	0.09753***	-0.20524
	(3.02003)	(2.67960)	(0.02772)	(0.18936)
Costa Rica	-1.07621	-1.05313	-0.21721***	-0.22453***
	(0.96141)	(0.84377)	(0.01605)	(0.01662)
El Salvador	5.29580***	5.27700***	-0.12886***	-0.09753***
	(0.92508)	(0.82158)	(0.01494)	(0.03070)
Guatemala	7.68359***	7.68596***	-0.13073***	-0.08522**
	(0.90693)	(0.79999)	(0.01420)	(0.03786)
Honduras	9.49815***	9.63300***	-0.05507***	0.00194
	(0.99314)	(0.89519)	(0.01598)	(0.04643)
Mexico	1.39484	1.47207	-0.06694***	-0.05812***
	(1.12012)	(1.00160)	(0.01927)	(0.02088)
Nicaragua	8.78158***	8.93547***	-0.10752***	-0.05488
	(0.99985)	(0.93449)	(0.01641)	(0.04255)
Panama	-0.62334	-0.67178	-0.04022**	-0.04383**
	(1.03115)	(0.95935)	(0.01726)	(0.01782)
Anguilla	-52.40313***	-53.04424***	0.47394***	0.10778
	(4.01069)	(3.67228)	(0.03867)	(0.22995)
Antigua and Barbuda	-17.28974***	-18.16728***	0.11348***	-0.04698
	(3.27999)	(2.85154)	(0.01167)	(0.10180)
Aruba	-27.73280***	-27.90451***	0.03176**	-0.18553
	(3.24253)	(3.07854)	(0.01417)	(0.14292)
Bahamas	-14.57414***	-13.40533***	0.02641	-0.08613
	(2.30808)	(2.14010)	(0.02111)	(0.07316)
Barbados	-42.73462***	-43.60664***	-0.09488***	-0.40486**
	(3.26742)	(2.86120)	(0.01671)	(0.19621)
Bermuda	-28.11976***	-28.93936***	0.11129***	-0.11241
	(3.67446)	(3.25238)	(0.02352)	(0.14185)
Cuba	7.44370***	7.50876***	-0.10503***	-0.06019
	(1.14868)	(1.03867)	(0.01921)	(0.03729)
Dominica	-46.97085***	-47.75552***	-0.13493***	-0.46964**
	(3.21763)	(2.86359)	(0.02192)	(0.20977)
Dominican Republic	12.84160***	12.85243***	-0.12439***	-0.04895
	(0.99233)	(0.87889)	(0.01497)	(0.05635)
Grenada	-47.35252***	-48.05891***	-0.14667***	-0.48350**
	(3.17011)	(2.79858)	(0.02445)	(0.20917)
Guadeloupe	9.85635***	14.52016***	-0.10021***	-0.01391
	(2.12716)	(1.08418)	(0.02008)	(0.05796)
Haiti	14.41366***	17.46361***	-0.10018***	0.00306
	(1.29236)	(0.78614)	(0.01260)	(0.06987)

Jamaica	-42.23341***	-42.97284***	-0.08710***	-0.39306**
	(3.17878)	(2.78369)	(0.01978)	(0.19111)
Martinique	11.14055***	16.21408***	0.23794***	0.33425***
	(2.04320)	(0.79525)	(0.01476)	(0.06173)
Montserrat	-49.21337***	-50.01102***	-0.07024***	-0.41557*
	(3.14964)	(2.74862)	(0.01794)	(0.21691)
Netherlands Antilles	-37.55429***	-38.45387***	0.54307***	0.26331
	(3.54411)	(3.13591)	(0.02344)	(0.17328)
Puerto Rico	-14.99319***	-14.94703***	-0.43428***	-0.52144***
	(1.34040)	(1.19714)	(0.02543)	(0.05470)
Saint Kitts and Nevis	-44.30724***	-45.16917***	-0.07350***	-0.39272*
	(3.40784)	(2.98880)	(0.01726)	(0.20148)
Saint Lucia	-38.75047***	-39.11875***	-0.22930***	-0.50882***
	(2.92024)	(2.63122)	(0.02611)	(0.17355)
Saint Vincent and the Grenadines				
	-45.29242***	-45.94840***	-0.16252***	-0.48567**
	(3.05227)	(2.70047)	(0.02254)	(0.20073)
Trinidad and Tobago	-40.26233***	-41.04624***	-0.09657***	-0.39155**
	(3.25997)	(2.85856)	(0.01735)	(0.18489)
Turks and Caicos Islands	-53.23613***	-53.84767***	0.01679	-0.35384
	(3.81701)	(3.33062)	(0.03170)	(0.24126)
Argentina	-2.33775**	-2.29160***	0.00016	-0.01290
	(0.96347)	(0.85489)	(0.01692)	(0.01723)
Bolivia	0.86349	0.81321	-0.07662***	-0.07172***
	(0.94238)	(0.89172)	(0.01568)	(0.01720)
Brazil	-7.47763***	-6.80443***	0.13624***	0.09565***
	(1.15234)	(1.07829)	(0.02081)	(0.03134)
Chile	14.18097***	14.18910***	-0.02294*	0.06108
	(0.86020)	(0.81194)	(0.01268)	(0.05748)
Colombia	5.49245***	5.54421***	-0.05523***	-0.02281
	(1.06918)	(0.96285)	(0.01712)	(0.02917)
Ecuador	8.91422***	8.91145***	-0.08682***	-0.03438
	(0.86894)	(0.80826)	(0.01358)	(0.03916)
French Guiana	2.52041	5.25744***	0.00969	0.02559
	(1.86637)	(1.46238)	(0.02364)	(0.02287)
Guyana	-40.52399***	-41.21869***	-0.10027***	-0.39608**
	(3.17105)	(2.78884)	(0.01831)	(0.18289)
Paraguay	-12.99155***	-11.91973***	-0.47389***	-0.55392***
	(1.30495)	(1.34887)	(0.02284)	(0.05065)
Peru	8.80351***	8.90960***	-0.11029***	-0.05788
	(1.02459)	(0.94640)	(0.01624)	(0.04108)
Suriname	-10.56924***	-5.31703*	-0.07278***	-0.15805***
	(1.54156)	(2.94355)	(0.01892)	(0.05612)
Uruguay	13.47748***	13.60508***	0.08045***	0.16034***
	(0.86371)	(0.76402)	(0.01408)	(0.05578)
Venezuela	0.40451	0.59212	0.00201	0.00506
	(1.13730)	(1.03695)	(0.01932)	(0.01964)
Austria	-26.82648***	-21.73962***	0.10663***	-0.07126
	(1.23641)	(2.74768)	(0.01335)	(0.11316)
Belgium	5.52104***	10.56161***	0.17790***	0.23552***
	(1.82697)	(0.85399)	(0.01264)	(0.03895)
France	10.80960***	16.10113***	0.16904***	0.26418***
	(2.11216)	(0.89859)	(0.01481)	(0.06153)
Germany	-15.54530***	-10.44816***	0.15191***	0.04712
	(1.12126)	(2.50125)	(0.01258)	(0.06799)
Luxembourg	-6.33352***	-0.72445	0.33361***	0.30641***
	(1.15913)	(1.59975)	(0.01468)	(0.02052)
Monaco	15.74085***	21.05574***	-0.58703***	-0.46221***
	(3.01506)	(1.66168)	(0.03693)	(0.08086)
Netherlands	-21.28034***	-16.30330***	0.22314***	0.07444
	(1.62194)	(3.04886)	(0.01584)	(0.09669)
Switzerland	-1.61635	3.52503**	0.23161***	0.23392***

	(1.32705)	(1.40323)	(0.01311)	(0.01254)
Bulgaria	-0.96353	7.20072**	-0.11855***	-0.12838***
	(1.31625)	(2.99180)	(0.02220)	(0.02139)
Czech Republic	-27.89646***	-19.93546***	0.06668***	-0.10344
	(0.99900)	(3.01720)	(0.01758)	(0.10697)
Slovakia	-16.13949***	-8.25315***	-0.00566	-0.10686*
	(0.97513)	(2.97763)	(0.01552)	(0.06335)
Czechoslovakia	-18.78102***	-11.75250***	0.08902***	-0.02992
	(1.14581)	(2.92490)	(0.01704)	(0.07396)
Hungary	-32.30561***	-24.36518***	0.02193*	-0.17435
	(0.74932)	(2.93980)	(0.01231)	(0.12512)
Poland	-18.44618***	-19.03364***	-0.00148	-0.11347
	(1.01306)	(0.84976)	(0.01735)	(0.06925)
Romania	1.41688	1.57830	0.02359	0.03060
	(1.14018)	(1.07620)	(0.02089)	(0.02176)
Estonia	-14.89584***	-7.08409**	0.04636**	-0.04489
	(1.34615)	(2.94640)	(0.02246)	(0.06024)
Latvia	-34.26025***	-27.44203***	0.03105	-0.17420
	(1.29648)	(2.68475)	(0.02592)	(0.12740)
Lithuania	-12.46898***	-10.50336***	-0.37652***	-0.45413***
	(1.34439)	(1.54930)	(0.02526)	(0.04811)
Belarus	-11.20785***	-2.88213	-0.10862***	-0.17867***
	(1.33522)	(2.99538)	(0.02388)	(0.04577)
Moldova	5.87145***	9.29948***	-0.12252***	-0.08900***
	(1.31243)	(1.68055)	(0.02297)	(0.03203)
Russia	-17.99107***	-9.69833***	-0.05775**	-0.16807**
	(1.28003)	(3.03681)	(0.02289)	(0.06914)
Ukraine	-12.85802***	-6.44976**	-0.05261**	-0.13186***
	(1.27937)	(2.48386)	(0.02298)	(0.05085)
Ireland	-25.68691***	-26.12952***	0.18208***	-0.02526
	(3.41499)	(3.11037)	(0.01818)	(0.13276)
Denmark	-32.13391***	-26.84249***	-0.12809***	-0.33895***
	(1.45238)	(2.92919)	(0.02039)	(0.13146)
Finland	-37.66946***	-37.57748***	0.06378***	-0.16027
	(1.13567)	(0.90589)	(0.01990)	(0.14168)
Iceland	-38.24351***	-31.21868***	0.68907***	0.45189***
	(2.06351)	(3.50083)	(0.03520)	(0.15121)
Norway	-18.89369***	-13.22649***	-0.07384***	-0.20125**
	(1.39533)	(2.94982)	(0.01814)	(0.08572)
Sweden	-18.88618***	-13.69791***	0.41335***	0.28083***
	(1.64023)	(3.03061)	(0.02176)	(0.08531)
United Kingdom	-33.25977***	-33.90658***	0.37535***	0.12452
	(3.29986)	(2.94430)	(0.01662)	(0.15886)
Albania	2.15208	1.59926	0.14958***	0.15990***
	(1.47042)	(1.20712)	(0.02270)	(0.02701)
Greece	-30.31247***	-31.06191***	-0.16690***	-0.34943***
	(0.60842)	(0.30108)	(0.00946)	(0.11622)
Malta	-21.82264***	-22.43431***	-0.66019***	-0.82760***
	(2.37884)	(2.17320)	(0.01999)	(0.10509)
Portugal	9.00695***	9.63591***	0.07911***	0.13609***
	(0.65609)	(0.54825)	(0.00918)	(0.04228)
Spain	0.67351**	0.59585**	0.11565***	0.11961***
	(0.32932)	(0.28373)	(0.00494)	(0.00544)
Bosnia	-2.07743*	5.92390**	-0.14617***	-0.16417***
	(1.17820)	(2.85821)	(0.02054)	(0.02050)
Croatia	-28.11509***	-20.60652***	0.04986***	-0.12291
	(0.65659)	(2.76132)	(0.01246)	(0.10819)
Macedonia	-16.84890***	-9.66851***	0.26384***	0.16756***
	(1.23272)	(2.52663)	(0.02288)	(0.05875)
Serbia	-6.52746***	0.86798	-0.08443***	-0.12516***
	(1.00765)	(2.50287)	(0.01781)	(0.02436)
Slovenia	-35.35098***	-28.08891***	0.21022***	-0.00548

	(0.63299)	(2.68378)	(0.01552)	(0.13445)
Yugoslavia	-7.83695***	-1.15795	0.08648***	0.03468
	(0.86704)	(2.55104)	(0.01353)	(0.03536)
Benin	10.35610***	16.78019***	0.05233**	0.13513**
	(1.62763)	(1.46747)	(0.02109)	(0.05550)
Burkina Faso	13.65491***	19.85223***	-0.09703***	0.00453
	(1.66842)	(1.40485)	(0.02025)	(0.06761)
Cape Verde	18.93351***	19.47376***	0.34619***	0.46116***
	(1.24937)	(1.18475)	(0.02091)	(0.08087)
Côte d'Ivoire	7.53419***	13.49129***	-0.09293***	-0.02303
	(1.71128)	(1.20829)	(0.01954)	(0.04968)
Gambia	-39.63013***	-31.43650***	-0.30084***	-0.54027***
	(1.19415)	(2.63488)	(0.02490)	(0.14711)
Ghana	-38.98397***	-37.80344***	-0.16128***	-0.43413***
	(2.46154)	(2.68376)	(0.02443)	(0.16850)
Guinea	20.27276***	27.76665***	-0.15714***	-0.03047
	(1.15531)	(2.14038)	(0.01813)	(0.08399)
Guinea-Bissau	-1.86549	-1.11010	-0.77961***	-0.78515***
	(1.30177)	(1.22854)	(0.02286)	(0.02374)
Liberia	-34.91563***	-35.52603***	-0.09052***	-0.34866**
	(2.98355)	(2.64344)	(0.02214)	(0.15804)
Mali	-1.15911	5.12591***	0.07485***	0.08112***
	(1.09817)	(1.67186)	(0.01833)	(0.01931)
Mauritania	18.67664***	22.52280***	0.01108	0.12736*
	(1.06888)	(1.27741)	(0.02222)	(0.07699)
Niger	11.57252***	17.63389***	-0.03537*	0.05922
	(1.82244)	(1.28105)	(0.02116)	(0.06196)
Nigeria	-28.33340***	-25.64929***	-0.09551***	-0.28955**
	(1.88776)	(2.52245)	(0.02294)	(0.12244)
Senegal	8.47255***	13.55223***	-0.07570***	-0.00821
	(1.40220)	(1.28081)	(0.01890)	(0.04800)
Sierra Leone	-3.60872***	-0.93396	-0.11400***	-0.14206***
	(1.12272)	(1.54726)	(0.01842)	(0.02624)
Togo	7.25395***	13.91837***	0.06762***	0.12604***
	(1.20594)	(1.63152)	(0.01827)	(0.04324)
Burundi	12.84191***	20.30729***	-0.03387*	0.04363
	(1.21454)	(2.39185)	(0.01861)	(0.05504)
Djibouti	-6.66306***	-4.09924***	0.05117**	0.01584
	(1.25995)	(1.24124)	(0.02201)	(0.03204)
Eritrea	-30.94634***	-29.70282***	-0.00240	-0.18625*
	(1.10284)	(0.88937)	(0.02097)	(0.11039)
Ethiopia	-26.27013***	-19.91826***	-0.17156***	-0.33239***
	(1.14473)	(2.40232)	(0.01939)	(0.09868)
Kenya	-29.57357***	-28.98162***	0.03501**	-0.15952
	(1.40877)	(1.56742)	(0.01703)	(0.12047)
Madagascar	12.46387***	14.52339***	0.03297**	0.11861**
	(1.10565)	(0.92802)	(0.01425)	(0.05882)
Malawi	-24.22570***	-24.88779***	0.59108***	0.39046***
	(3.67963)	(3.22961)	(0.02156)	(0.12859)
Mauritius	-3.95074***	-1.12071	0.12644***	0.11865***
	(1.23531)	(0.85532)	(0.01457)	(0.01447)
Mozambique	2.67226***	3.42630***	0.03124**	0.04936**
	(0.82092)	(0.80229)	(0.01381)	(0.02023)
Réunion	20.29991***	25.80221***	0.19586***	0.34829***
	(2.34185)	(1.22037)	(0.02332)	(0.09719)
Rwanda	9.71818***	16.51279***	-0.12940***	-0.06609
	(1.09930)	(2.01091)	(0.01833)	(0.04579)
Seychelles	-19.14634***	-17.13412***	0.01970	-0.08794
	(0.95607)	(0.94256)	(0.02078)	(0.06285)
Somalia	-18.59733***	-18.29705***	-0.08821***	-0.20148***
	(1.20600)	(0.94926)	(0.01890)	(0.06856)
Tanzania	-24.65315***	-21.90512***	0.03891**	-0.11150

	(0.82253)	(1.37796)	(0.01520)	(0.09256)
Uganda	-34.96367***	-35.11080***	-0.11450***	-0.35343**
	(1.99613)	(1.89421)	(0.01277)	(0.15068)
Zambia	-33.19458***	-33.60332***	-0.09973***	-0.32580**
	(2.02359)	(1.87135)	(0.02680)	(0.13951)
Zimbabwe	-49.66131***	-47.96681***	-0.02194	-0.35839*
	(2.56527)	(2.96559)	(0.02359)	(0.21199)
Algeria	13.75911***	14.72464***	-0.03386*	0.05330
	(1.17750)	(1.08154)	(0.01983)	(0.06112)
Egypt	-8.91239***	-7.95579***	0.07574***	0.02774
	(0.83970)	(0.86289)	(0.01352)	(0.03264)
Libya	-14.51762***	-14.53009***	0.16783***	0.08140
	(1.15356)	(0.94095)	(0.01873)	(0.05668)
Morocco	8.54111***	9.71403***	-0.02885*	0.02901
	(0.99236)	(0.93190)	(0.01621)	(0.04160)
Sudan	-16.68146***	-16.83466***	-0.07898***	-0.18204***
	(1.17539)	(1.10282)	(0.02047)	(0.06499)
Tunisia	10.85210***	11.15412***	-0.08338***	-0.01735
	(1.09398)	(0.97649)	(0.01725)	(0.04602)
Angola	10.77087***	11.41787***	0.02022	0.08695*
	(0.88482)	(0.81063)	(0.01266)	(0.04886)
Cameroon	1.31180	6.72510***	-0.02506	0.00270
	(1.76419)	(1.35094)	(0.02097)	(0.02769)
Central African Republic	4.39453**	10.28959***	0.23004***	0.28356***
	(1.70443)	(0.87264)	(0.01658)	(0.03675)
Chad	16.14510***	18.54569***	-0.18865***	-0.08334
	(1.27013)	(1.20447)	(0.02244)	(0.07212)
Congo	13.61371***	17.41838***	-0.09460***	0.00455
	(1.42852)	(1.08732)	(0.01790)	(0.06800)
Equatorial Guinea	31.20119***	36.81156***	-0.17013***	0.04736
	(2.13810)	(1.16043)	(0.03488)	(0.15230)
Gabon	8.27795***	13.97923***	-0.03719*	0.04305
	(2.05134)	(1.13242)	(0.01971)	(0.05466)
Botswana	13.92876***	13.37617***	-0.62724***	-0.60146***
	(4.07385)	(3.72032)	(0.03410)	(0.03579)
South Africa	-40.71355***	-41.45291***	0.27234***	-0.02221
	(3.19277)	(2.82994)	(0.01841)	(0.18510)
Afghanistan	-5.51372***	-5.91813***	-0.09373***	-0.12850***
	(1.27310)	(0.97008)	(0.02005)	(0.02518)
Cyprus	-43.18273***	-43.90777***	0.05623***	-0.20906
	(1.06328)	(0.92267)	(0.01626)	(0.16995)
Iran	-26.43463***	-26.95187***	-0.15483***	-0.31455***
	(1.18653)	(1.03610)	(0.02175)	(0.10004)
Bahrain	-63.16594***	-63.96290***	0.29504***	-0.09948
	(1.71024)	(1.56743)	(0.02241)	(0.25275)
Iraq	-22.89407***	-23.34873***	-0.03788**	-0.17941**
	(1.00936)	(0.91160)	(0.01666)	(0.08850)
Israel	-29.33720***	-29.12365***	0.07547***	-0.10035
	(0.85681)	(0.85090)	(0.01719)	(0.10880)
Jordan	-30.95585***	-31.47707***	-0.11666***	-0.30494***
	(1.25101)	(1.10829)	(0.02029)	(0.11767)
Kuwait	-28.10433***	-28.47013***	0.03902*	-0.13617
	(1.32190)	(1.21615)	(0.02138)	(0.10800)
Lebanon	-1.88102*	-1.74465*	-0.04701***	-0.05738***
	(1.01659)	(0.93966)	(0.01660)	(0.01614)
Palestine	-21.30964***	-21.57475***	0.03213*	-0.09588
	(0.97258)	(0.80337)	(0.01827)	(0.08287)
Qatar	-39.93131***	-40.36603***	-0.19177***	-0.42856***
	(1.34354)	(1.10078)	(0.02305)	(0.14310)
Saudi Arabia	-25.40304***	-25.51046***	-0.10796***	-0.25837***
	(1.24914)	(1.03729)	(0.01977)	(0.09576)
Syria	-6.74882***	-7.00225***	-0.10426***	-0.14520***

	(0.98829)	(0.81045)	(0.01655)	(0.02682)
United Arab Emirates	-16.43586***	-17.03493***	-0.06420**	-0.18089**
	(1.80728)	(1.71191)	(0.02685)	(0.08096)
Yemen	-9.05474***	-9.53263***	-0.36180***	-0.42448***
	(1.31671)	(1.19326)	(0.02866)	(0.03832)
Turkey	-12.96192***	-13.02648***	-0.09317***	-0.17023***
	(1.02767)	(0.71595)	(0.01606)	(0.04564)
Kazakhstan	-11.14859***	-3.02362	-0.15005***	-0.21949***
	(1.29012)	(2.98215)	(0.02289)	(0.04482)
Kyrgyzstan	-18.76489***	-10.06371***	0.03473	-0.07886
	(1.46841)	(2.92846)	(0.02504)	(0.07010)
Tajikistan	-6.75087***	0.75337	0.30036***	0.26921***
	(1.05068)	(2.25157)	(0.02602)	(0.03287)
Turkmenistan	-12.91376***	-5.37847*	-0.26448***	-0.35055***
	(1.41542)	(3.10737)	(0.02638)	(0.05448)
Uzbekistan	-19.56401***	-11.18588***	0.09381***	-0.02604
	(1.37500)	(3.16109)	(0.02406)	(0.07432)
Armenia	-10.55942***	-9.60819***	-0.13215***	-0.19569***
	(1.20663)	(0.99260)	(0.01848)	(0.03904)
Azerbaijan	8.70622***	17.39924***	-0.24414***	-0.19608***
	(1.42815)	(2.84650)	(0.02379)	(0.04332)
Georgia	-21.08219***	-13.74264***	-0.18246***	-0.30992***
	(1.29574)	(2.64742)	(0.02081)	(0.07899)
China	-38.00294***	-38.24716***	-0.16808***	-0.39305***
	(1.36574)	(1.02605)	(0.02479)	(0.13933)
Hong Kong	-45.17807***	-45.61860***	-0.06748***	-0.33701**
	(1.04038)	(0.80229)	(0.02086)	(0.16638)
Macau	-49.57166***	-50.17406***	0.25179***	-0.04183
	(1.05653)	(0.92718)	(0.01961)	(0.18253)
Japan	-30.22307***	-22.97939***	0.07563***	-0.11019
	(1.26280)	(2.87986)	(0.02380)	(0.11711)
South Korea	-29.14580***	-20.76172***	-0.17220***	-0.34769***
	(1.25691)	(3.05732)	(0.02617)	(0.10732)
Taiwan	-35.64840***	-35.99802***	-0.03843	-0.24987*
	(1.27489)	(0.97381)	(0.02370)	(0.13048)
Cambodia	7.14477***	13.34557***	-0.05339***	-0.01144
	(1.04510)	(1.83365)	(0.01310)	(0.03585)
Indonesia	-37.69088***	-37.91154***	0.02705	-0.19657
	(1.34651)	(1.05024)	(0.02561)	(0.13839)
Laos	13.36659***	13.40820***	-0.12036***	-0.04121
	(0.99704)	(0.70889)	(0.01341)	(0.05855)
Malaysia	-46.33694***	-46.80320***	-0.02307	-0.30807*
	(1.23774)	(1.07144)	(0.02210)	(0.17660)
Myanmar	-74.51438***	-74.92534***	0.44175***	0.00133
	(1.17410)	(0.94470)	(0.03552)	(0.26279)
Philippines	-50.23519***	-50.67536***	-0.20480***	-0.50834***
	(1.26070)	(1.02282)	(0.02703)	(0.18695)
Singapore	-36.52793***	-37.13720***	-0.06352***	-0.30507**
	(1.93938)	(1.81005)	(0.02283)	(0.15285)
Thailand	11.39330***	14.50098***	-0.01920	0.04198
	(1.13979)	(1.59095)	(0.02017)	(0.04403)
Vietnam	-2.73727***	-2.68673***	-0.05165***	-0.06776***
	(0.97593)	(0.75718)	(0.01291)	(0.01328)
Bangladesh	-47.76828***	-48.30016***	-0.31696***	-0.60243***
	(1.27262)	(1.03936)	(0.02594)	(0.17607)
India	-47.11026***	-47.29125***	-0.12530***	-0.40943**
	(1.04218)	(0.85891)	(0.02117)	(0.17524)
Pakistan	-42.60637***	-42.98800***	-0.16124***	-0.41866***
	(1.10199)	(0.89833)	(0.02209)	(0.15887)
Sri Lanka	-50.90802***	-50.49429***	-0.19829***	-0.50166***
	(1.28380)	(0.95941)	(0.02633)	(0.18508)
Australia	-33.02860***	-33.46775***	0.38496***	0.13798

	(3.10708)	(2.81707)	(0.02048)	(0.15577)
Fiji	-40.84344***	-41.05052***	-0.30624***	-0.54720***
	(1.13037)	(0.86714)	(0.02350)	(0.14502)
French Polynesia	-9.73872***	-4.20525***	0.34428***	0.32057***
	(1.88043)	(1.13154)	(0.02438)	(0.02752)
New Caledonia	19.04988***	23.48274***	0.16331***	0.30242***
	(1.73817)	(0.83976)	(0.01556)	(0.08951)
New Zealand	-31.19085***	-31.98226***	0.38895***	0.14730
	(3.43440)	(3.04156)	(0.01892)	(0.15330)
Others	-56.55655***	-57.01124***	0.27339***	-0.08537
	(1.52113)	(1.36656)	(0.02901)	(0.22694)
Observations	67440	67440	67440	67440
R-squared	0.45	0.45	0.27	0.22

All the regressions are weighted using Census weights; All regressions include the full set of controls variables: age, age squared, gender, education, years since migration, years since migration squared, marital status and an indicator variable for the presence of children, year-specific dummies and location and city-specific controls. Robust standard errors in parentheses are clustered at the countries of birth level. Significance levels: *** at 1%, ** at 5%, * at 10%

Chapter Two

Labour Shortages and Immigration: The case of the Agriculture Sector

2.1 INTRODUCTION

With the shifting demographics in the Canadian workforce, mostly driven by the falling fertility rates and the ageing of the baby boom generation, the ability to attract and retain workers is becoming more crucial for improving the competitiveness of Canadian industries. As an attempt to help mitigate these effects on the labour market, immigration has been deployed as a key strategy. Immigration is the main driver of Canada's population growth and employment growth, with just over one fifth of the Canadian population being born outside Canada in 2016.[8] However, the majority of immigrants are concentrated in few major cities, with 63.4% of immigrants in Canada living in Montreal, Toronto and Vancouver.[9] At the same time, Canadian regions and rural communities have been experiencing persistent labour challenges with a more rapid ageing of their population, a declining and lower fertility rate compared to urban regions, and youth out-migration to urban areas. As a consequence, rural communities are faced with more persistent labour shortages and slower economic growth. Over the past decades, there have been a few policy initiatives aimed to help with a more balanced national distribution of immigrants. For instance, the Provincial Nominee Program has enabled provinces to choose

[8] https://www150.statcan.gc.ca/n1/dai-quo/ssi/homepage/rel-com/theme43-eng.htm
[9] https://www150.statcan.gc.ca/n1/pub/89-657-x/89-657-x2016002-eng.htm

immigrants with specific demographics, skills, and work experience required to meet their local labour market needs. In addition, a recent Rural and Northern Immigration Pilot program was created to help attract and retain immigrants to smaller communities. The program is driven by local communities with the aim of filling labour market shortages and improving regional economic development.

Immigration programs that are trying to diversify the national distribution of immigrants by attracting and retaining immigrants to rural areas all have in common that they are trying to fill labour shortages in those regions. They are intended to help employers fill labour market needs by finding permanent immigrant workers. A sector that represents an interesting case in point in the dual context of attracting and retaining immigrants in rural areas and of filling a labour shortage is the agriculture sector. In fact, this sector is concentrated in rural areas and is also one of the industries that shows the most significant need for labour. Over the past few years (2015-2018), the agriculture sector has persistently shown the highest vacancy rate among all other industries.[10] Labour migration has been one of the key policy measures to maintain the economic development of this sector. A widely-used approach by employers is to rely on temporary foreign workers (TFW) to fill short-term labour demand requirements and seasonal vacancies. An alternative option that has been explored is to facilitate the pathway to permanent residency for foreign workers with Canadian work experience who are filling a year-round need. In September 2016, the Standing Committee on Human Resources, Skills and Social Development and the Status of Persons with Disabilities (HUMA) produced a detailed report on the Temporary Foreign Workers Program (TFWP) containing recommendations for the Government of Canada (HUMA, 2016). One of the key recommendations was to facilitate pathways to permanent

[10] Statistics Canada Job Vacancy and Wage Survey, author calculation with Cansim Table: 14-10-0326-01. Job vacancies data for the agriculture sector where only available since 2015.

residency for foreign workers filling long-term labour market needs. The Government's response, released January 2017, welcomed the recommendation to review the current pathways to permanent residence for all temporary workers. The government also mentioned its willingness to continue to support and facilitate the pathway to permanent residency for TFW (Government Response, 2017). In addition, recently (July 2019) the Government of Canada has announced that it is launching a new agri-food immigration pilot project that will offer a permanent residency pathway to temporary foreign workers to help fill labour shortages in Canada. The program aims to accept approximately 16,500 new immigrants over the 3-year duration of the pilot.[11]

As the agriculture sector is one of the industries that exhibits the most significant need for labour and is also concentrated in rural areas, and is also one of the main users of the TFWP, the main objective of this paper is to identify how immigration can help filling this shortage by first investigating the factors that are likely to affect the attraction of immigrants compared to Canadian-born workers in the agriculture sector, and then to investigate the factors that impact the retention after entering in the sector. In particular, in response to the efforts to facilitate permanent residency pathways, whether former foreign workers with Canadian experience who became permanent residents are most likely to stay in the sector after entering remains a key question for policy makers that this paper will attempt to investigate. I will also examine if the Provincial Nominee Program has been successful in attracting and retaining immigrants in rural areas for filling labour shortages by looking at the case of the agriculture sector. This analysis is based on a rich longitudinal micro-database for the years 2001-2013 developed by Statistics Canada, the Canadian Employer-Employee Dynamic Database (CEEDD).

[11] https://www.canada.ca/en/immigration-refugees-citizenship/news/2019/07/new-agri-food-immigration-pilot.html

The paper is organized as follows. The second section consists of a review of some selected studies on the labour supply decisions in the agriculture sector and on the labour market impact of immigration in Canada. The third section provides some description of the dataset, and presents some descriptive statistics. The fourth section presents the econometric models followed by the clarification of the specifications. The fifth section consists of a discussion of the results, and the last section concludes the paper.

2.2 Literature Review

The review of the literature consists of two parts. The first part reviews existing studies that could provide us with insights on factors contributing to the recruitment and retention of agriculture and agri-food workers. To my knowledge, there have been no studies examining the choice of entering the agriculture and agri-food sector among the entire workforce. The existing line of research exclusively focuses on labour allocation between farm and off-farm activity of those who are already farmers. Therefore, I present pertinent papers of this existing line of research in the first part of the review. The second part of the review delves into papers that investigate the labour market outcome of immigrants in Canada.

2.2.1 Labour Supply Allocation Decisions in the Agriculture Sector

Among the earliest influential studies on work decisions of farmers, Huffman (1980) examines the effect of investments in education on the off-farm labour supply of farmers in the United States. He models the labour supply decision as the result of a household utility maximization subject to constraints on farm production, time, and income. He uses a logistic equation to study the relationship between the participation rate and education. Other control variables include off-farm wage rates, size of the farms, average number of children under age five, share of farm households that are non-white, average age of farm operators and state

dummy variables. The main finding of the research suggests a positive relationship between farmers' education and their off-farm labour supply. Following Huffman (1980), many other studies have investigated the factors explaining the choices to work on or off the farm. For instance, Sumner (1982) argues for the importance of controlling for age to capture the employment experience, which increases the marginal value of time in each activity. In addition, Goodwin and Holt (2002) and Hennessy and Rehman (2008)[12] have also found empirical evidence of a positive relationship between education and off-farm labour supply, suggesting that the indirect effect of a potential increase in the off-farm wage due to higher levels of education dominates the direct effect of an increase in productivity in farm operation as a result of higher levels of education.

Other relevant individual characteristics that have been used in this literature include gender, marital status and the presence of children. For instance, Juvančič and Erjavec (2005) explore the employment decisions of agricultural households in Slovenia and find that males have a lower probability of off-farm participation than females; they explain that this is likely related to men having a lower level of mobility of labour supply.

To summarize this line of research, variables used in the literature to explain choice of on-farm employment can mainly be divided into three main characteristics: (1) individual characteristics such as age, education, gender, race (2) family characteristics such as marital status, the number of children, household size, and (3) location and labour market characteristics.

2.2.2 LABOUR MARKET OUTCOME OF IMMIGRANTS

Most immigration-related studies have concentrated on the national level, and fewer

[12] See among others Huffman and Lange (1989), Gould and Saupe (1989), Kimhi (1994), Benjamin and Kimhi (2006).

studies have analyzed the labour market of immigrants in regions. Among them, Akbari (2011) studies the labour market performance of immigrants in Atlantic Canada. The analysis uses census data for the period between 1981 and 2006 and is based on descriptive statistics. The results of the paper show that labour market outcomes of immigrants are better than those of the Canadian-born in Atlantic Canada. The author finds that immigrants have higher labour force participation rates, higher earnings, and lower unemployment rates than non-immigrants. Those results are not in agreement with what the literature has found for the economic performance of immigrants at the national level for Canada. The author argues that this may in part be explained by problems of credential recognition as a barrier to labour market integration, which may be less important in smaller provinces and regions where there is more opportunity for lower-skill employment. He also further explains that the smaller immigrant communities in regions lead to more interaction and networking opportunities with the native-born. Finally, he argues that the selection programs of immigrants in regions such as the Provincial Nominee Program are generally aimed at meeting labour market shortages, and the fact that those new immigrants have often a job lined for them at their arrival may in part explain the observed better economic outcomes.

Warman and Worswick (2004) analyze the decline in the labour market outcomes of recent immigrants in Canada at the regional and the CMA levels. They employ the 1981, 1986, 1991, 1996 and 2001 Canadian Census micro data files with their sample restricted to all Canadian-born and immigrant workers between 25 and 59 years of age. The analysis is based on simple means. Their results suggest that immigrants who reside in non-CMAs (rural areas) have better economic outcomes relative to the Canadian-born in non-CMAs. They also found that when the earnings of immigrants living in CMAs are compared to those of the Canadian-born in

CMAs, the performance of immigrants falls even more relatively to the national level. In general, their findings suggest that the difference between urban and rural areas is an important element to consider in the analysis of earnings differences between immigrants and Canadian-born. However, it is important to note that their definition of rural can be controversial as someone can argue that not all non-CMAs are rural. Their results might be better interpreted in the context of large cities (CMAs) compared to smaller cities (non CMAs).

Bernard (2008) compares the economic integration of immigrants who decide to settle in urban centres to those who settle in rural centres. Urban areas in this paper are defined as census agglomerations with population exceeding 15,000, and rural areas as all other locations. He uses the Longitudinal Administrative Databank (LAD), which contains a 20 % sample of T1 Family tax files. The T1 Family tax files is a combination of the individual T1 file with the T4 tax file and with a file from the Federal Child Benefits, which are aggregated at the family level. The sample is restricted to individuals of 20 years of age or over. The methodology consists of an ordinary least squares regression model with the dependent variable being a measure of the individual's income advantage and controlling for the province of residence, type of family, age group and sex; Bernard (2008) also includes a vector of explanatory variables only for immigrants (admission class of the immigrant, prior knowledge of official language, level of education at the time of arrival, and the immigrant's country of origin). The author's key finding is that the economic integration of immigrants in rural areas is significantly more rapid than in urbanized areas, and that the economic advantage also increases over time. The author suggests that immigrants in rural areas benefit from better formal and informal networks with non-immigrants because of the smaller proportion of immigrants living in those regions. The author also argues that immigrants in smaller cities are more likely to learn one of the official languages

quickly because they live in an area with a higher proportion of native Francophones or Anglophones.

On the other hand, studies that have focused on the national level have documented low economic performance of recent entry cohorts of immigrants in Canada (Bloom, Grenier, and Gunderson (1995), Beaudry and Green (2000), Green and Worswick (2012)). Aydemir and Skuterud (2005) suggest that the cause of this deterioration is in part related to low returns of foreign labour market experience. Schaafsma and Sweetman (2001) also find that foreign work experience yields no return in the Canadian labour market. These results were also similar to the finding of Ferrer and Riddell (2008), who studied how the human capital of immigrants is rewarded in Canada. They found that immigrants' work experience obtained before landing is valued much less than Canadian experience.

Following those findings, an interesting question to ask is whether we should expect immigrants with pre-landing Canadian work experience (P.L.C.E) to exhibit similar outcomes to those with foreign work experience. Sweetman and Warman (2014) used the Longitudinal Survey of Immigrants to Canada (LSIC) to compare the economic outcomes of skilled worker principal applicants with former TFW and former international students. They found that former TFW and international students have higher employment rates and earnings than immigrants without P.L.C.E. They suggested that prior work experience in Canada could be used as an important signal for policymakers. Hou and Picot (2016) used the Longitudinal Immigration Database to examine the earnings trends among immigrants over the 1990s and 2000s. They also found that the large increase in the share of immigrants with high pre-landing Canadian earnings during the 2000s was one of the key factors that put upward pressure on entry earnings of immigrants.

A related group of studies investigated the impact of TFW on the Canadian labour market. Gross and Schmitt (2012) explored whether wide regional disparities in unemployment rates' persistence was in part caused by the expansion of the TFW program in 2002 to all low-skill occupations. Their results suggest that the change in the program made more persistent the pattern of regional disparities in unemployment rates for Canada. They argued that the cost to hire low-skill TFW was set too low, and that employers preferred to hire TFW instead of seeking local employees from high-unemployment provinces. Brochu, Gross and Worswick (2017) tried to understand the reasons behind why Canadian employers might prefer to hire TFW despite the associated hiring costs related to the program. They found evidence that TFWs exert more effort than domestic workers, working longer hours, and also exhibiting lower levels of absenteeism from the job.

As mentioned earlier, immigrants coming to Canada have not dispersed evenly across the countries and are mostly concentrated in large cities. Over the past years, there have been some immigration policies aimed at attracting and retaining immigrants to rural areas to help with a more balanced national distribution of immigrants across Canada. As rural immigration policies are mainly designed to fill labour shortages, an interesting case in point is the agricultural sector. This sector is concentrated in rural areas, exhibits a significant need for labour, and is also the main user of the TFWP.[13] Therefore, this paper aims to extend the Canadian literature on immigration by identifying the factors that are likely to affect the attraction and retention of immigrants compared to Canadian-born workers in rural areas to fill a labour shortage by looking at the case of the agriculture sector. This paper also investigates whether former foreign workers with Canadian experience who became permanent resident have better retention rates

[13] Source: https://open.canada.ca/data/en/dataset/360024f2-17e9-4558-bfc1-3616485d65b9#wb-auto-6

than other immigrants. This could inform policy makers with evidence for the ongoing revision of the pathway to permanent residency. Finally, this research also examines whether the Provincial Nominee Program has been successful in attracting and retaining immigrants in rural areas for filling labour shortages by looking at the case of the Agriculture sector. For comparison purposes, I will also repeat my analysis for the agri-food sector (food and beverage processing), as this industry has a lot of similarities with the primary agriculture sector and follows the agriculture sector in the food value chain.

2.3 Database and Descriptive Analysis

2.3.1 Data

This analysis is based on a novel micro-database for the years 2001-2013 developed by Statistics Canada, the Canadian Employer-Employee Dynamic Database (CEEDD). It is the ideal longitudinal database to study the issue of recruitment and retention rates for Canadian-born and immigrant workers across industries, since it provides rich information on employer-employee characteristics over time. The data files are drawn from the T1 Personal Master File, the T4 Statement of Remuneration Paid File, the longitudinal immigration database (IMDB), and the Longitudinal Employment Analysis Program (LEAP). The CEEDD was constructed using data from tax filers, and so its frequency is annual.

The T1 Personal Master File provides information on incomes, taxes paid and characteristics of individuals who file taxes in a given year. The longitudinal immigration database (IMDB) offers detailed information on characteristics of immigrants at the time of the landing, such as source countries, immigrant class, landing years, gender, education level, Canadian official language ability, etc. The database also provides information on pre-landing work experience in Canada, provincial mobility, and family composition. The pre-landing work

experience variable is an indicator of whether the immigrant had experience in the Canadian labour market before becoming a permanent resident; there is no information on the sector in which immigrants acquired pre-landing experience. An immigrant is included in the IMDB if he/she has obtained permanent resident status since 1980 and has filed at least one tax return since 1982. The information available in the IMDB is critical to this project, since I need immigrant status information and pre-landing employment information.

The CEEDD provides longitudinal data on the labour market of both Canadian-born and immigrant workers. Other interesting databases with labour market information, for example, the Census and the Labour Force Survey, have a lack of employer information. However, the CEEDD suffers from some drawbacks, such as a limited amount of information on socio-demographic characteristics. For instance, there is no information on education for Canadian-born subjects; the variable related to education is only available for immigrants from the IMDB, which consists of the level of education at time of landing. Another limitation of this database is the impossibility of determining if an employee is working full-time or part-time. Furthermore, there is no information on length of working time, and thus precise hourly wage rates cannot be obtained.

The T1 and the IMDB were linked together using social insurance numbers (SINs) as identifiers and the employers' information were linked to employees' individual-level records using the business register identification number.[14]For employees who are multiple jobholders, I only kept the highest earnings obtained in each given tax year. Approximately 27 percent of our sample have multiple jobs; this figure is 29 percent among immigrants. This discrepancy is a potential caveat of the study since I need to link employee information to only one employer for

[14] The linkage using the social insurance numbers was completed by Statistics Canada, and the linkage of the employees' information using the business register identification number was performed by the author.

every tax year to be able to use panel regression analysis. I also restrict my sample to Canadian-born and immigrant employees aged between 18 and 64 years.

2.3.2 Descriptive Analysis

Two main models are used in this paper to analyze the likelihood of working and staying in the industry. First, I investigate the individuals' characteristics that affect the likelihood of working in the sector (the recruitment model) with the use of a generalized estimating equation (GEE). This model will include the entire workforce. After investigating the recruitment effect, I then turn to analyzing the retention of an employee conditional on entering the sector in a discrete survival model framework (the retention model). I have two principal samples in the retention model; one that only includes those who have worked for at least one year in the primary agriculture sector, and another sample that only includes those who have worked in the agri-food sector for at least one year. The next section will present general descriptive statistics of the main samples used and a non-parametric estimation of the retention model. The models, justifications, and their specifications are explained more in detail in section 2.4.

2.3.3 Descriptive Analysis of the Recruitment Model

I present the means of key variables that are used in the recruitment model in Table 2.1. The summary statistics indicate that the majority of workers are Canadian-born (85 percent) compared to immigrants without P.L.C.E (11 percent) and immigrants with P.L.C.E (4 percent). Half of them are men, and the average age is 39 years. I can compare those statistics for workers in the agriculture and the agri-food sector in tables 2.2 and 2.3. The proportion of immigrants is approximately similar for the primary agriculture sector, but there is a higher representation of immigrants in the food and beverage-processing sector. Another difference is that there is a higher proportion of men than women in the agriculture and in the food and beverage processing

sector compared to the all-industry average. The second column of table 2.1 also provides summary statistics pertaining only to immigrants based on the rich information available in the IMDB at the time of the landing. I see that immigrants have been in Canada on average for 11 years, and that most of them come from the family immigration class (principal applicant) and the skilled workers' immigration class (principal applicant). I also observe that the proportion of immigrants without a high school degree at the time of landing is larger (48.5 percent) than those in all other categories. It is also important to mention that employees in the agriculture sector represent only approximately 1 percent of the workforce, while agri-food workers represent approximately 2 percent. However, farmers are a minority group that face a number of labour challenges and are also the beginning and the base of whole food value chain that account for one in eight jobs in Canada.[15] In addition, the labour market of the agriculture has almost never been studied in Canada because of its small sample size in other datasets. The CEEDD database provides the opportunity to have an important number of observations for this industry and to analyze the labour market of this important minority group.

2.3.4 DESCRIPTIVE ANALYSIS OF THE RETENTION MODEL

Table 2.2 shows the means of key variables for those who have worked for at least a year in the primary agriculture sector. I see that 3.8 percent of them are immigrants with P.L.C.E, and there is a higher proportion of men than of women. The summary statistics also indicate that two fifths of them are married, and the average age is 36 years. The majority of immigrants in this sector have not obtained a high school degree at the time of the landing (69.7 percent), and only 45.7 percent of them had knowledge of an official language at the time of the landing.

Table 2.3 shows the descriptive statistics for the food processing retention sample, i.e.

[15] Agriculture and Agri-Food Canada: https://www.canada.ca/en/agriculture-agri-food/news/2016/05/report-highlights-agriculture-sector-s-strong-contributions-to-the-canadian-economy.html

those who have worked for at least a year in the food processing sector. I observe a high proportion of immigrants, with 7 percent of the whole sample having P.L.C.E. This is higher than those observed in the primary agriculture sector and higher than the average for all industries. The proportion of immigrants without a high school degree is still larger (55 percent) than those in all other categories, and 57 percent of them had knowledge of an official language at the time of the landing.

2.3.5 NON-PARAMETRIC ANALYSIS FOR THE RETENTION MODEL

I first estimate a non-parametric survival function using the Kaplan-Meier estimator for both samples in my retention model:

$$\widehat{S}(j) = \prod_{k=1}^{j} \frac{n_k - d_k}{n_k} \quad (1)$$

Where d_k is the number of observed outcomes for the event of quitting the sector that happened at time k, and n_k is the number of individuals known to survive (i.e. stay in the sector) at time k. Figure 2.1 shows the Kaplan-Meier survival function for our primary agriculture sample. I see that the survival probability shows the most significant drop after one year of working, with only 63% of the spells being observed to last for more than one year, and only 36% to last over five years. After five years, the curve flattens but does not completely stabilize. It suggests that during the first five years, there is a high probability of the individuals quitting the sector and those who remained up to 5 years are more likely to stay afterward. Figure 2.2 shows the Kaplan-Meier survival function for the groups of immigrants without P.L.C.E, immigrants with P.L.C.E, and the Canadian-born in the primary agriculture sector. The findings suggest that immigrants with P.L.C.E have higher survival rates than immigrants without P.L.C.E and then

Canadian-born workers over the whole period after entering the sector, and I do not see any important difference of survival rates for immigrants without P.L.C.E and Canadian-born workers. The log rank test indicates that the differences between those three groups are significant nonetheless.

The Kaplan-Meier survival function for the food processing sample is presented in Figure 2.3. The results show similar patterns to those of the primary agriculture sector, characterized by a very high probability of quitting the sector after one year of entering. In fact, 65% of the spells are observed for more than one year and 36% are over five years. After five years, the slope of the survival curve decreases and becomes flatter, suggesting again that those who stayed in the sector for at least five years are less likely to quit afterward. Figure 2.4 shows the Kaplan-Meier survival functions by immigration status. The findings suggest that immigrants with P.L.C.E have higher survival rates than immigrants without P.L.C.E and then Canadian-born workers over the whole period after entering the sector. In addition, the results show that immigrants without P.L.C.E also have higher survival rates than the Canadian-born workers for the whole period after entering. The log rank test also indicates that the differences between those three groups are significant.

2.4 METHODOLOGY

2.4.1 MODEL 1- RECRUITMENT MODEL

In our dataset, each individual has multiple observations across tax years, and as a result, observations are correlated. Using cross-sectional models such as OLS will be less efficient because responses within subjects are not independent, and these within-subject correlations are not considered. Therefore, I take advantage of the longitudinal nature of the dataset to model the probability of working in the sector with a panel data model technique. The advantage of the

panel model is that I can account for correlation of responses within subjects, and unobserved individual characteristics that do not change across time. The choice of the model depends in part on whether it is more appropriate to use a population average (PA) or subject specific (SS) models. The fixed effects model and the random effects model are commonly used to estimate subject-specific effects models, and the generalized estimating equations (GEE) can be used to estimate population-averaged effects models.

The fixed-effects model will give consistent estimators regardless of whether or not the unobserved effect is correlated with the explanatory variables. However, one limitation with the fixed effects model is that I cannot estimate the effects of time-invariant covariates. For this reason, I avoid using this model in my research because my variable of interest, immigrant P.L.C.E, does not vary over time. An alternative SS approach is the random effects model, which assumes that the unobserved individual characteristics are uncorrelated with the observed explanatory variables. In this model I can include time-invariant variables that could explain the probability of working in the sector. However, to obtain consistent estimates, I have to assume that the omitted variables are uncorrelated with the included explanatory variables. Finally, another alternative to those models is the PA model, also known as a generalized estimating equation (GEE). The next section describes the GEE model with its underlying assumptions and how it is being used in the recruitment model.

2.4.2 THE GEE MODEL

The GEE model is an extension of the standard generalized linear models (GLM) using quasi-likelihood estimation and first introduced by Liang and Zeger (1986). This method has been widely used in medical sciences such as in epidemiology and biology research. The use of this model to date has been more limited in social science research. An advantage of the GEE

model is that it relaxes the assumption of independence of the unobserved effects with the observed covariates and provides consistent and asymptotically normal estimates, and time invariant covariates can also be included. The GEE involves a marginal or a population average model, and the model gives an average response for a unit increase in the explanatory variables across the population. The assumptions for the GEE model are the following:

1. The outcome variables (Y_{it}) are correlated.
2. The model assumes a linear relationship between the independent variables and a transformation of the outcome variable (Y_{it}), which is transformed by a link function $\boldsymbol{g}$.
3. The structure of the within-subject covariance needs to be defined. It can be either: (1) exchangeable (assuming that observations over time have similar correlation). (2) independent (observations over time are independent). (3) auto-correlated to the first order (decreasing correlation depending on how far apart in time two observations are), or (4) unstructured (correlation between different time periods can be different).

Before estimating the model, I have to choose the appropriate link function, a distribution of the response variable and a covariance structure. To model the expected value of the dependent variable as a linear combination of the covariates, I need first to choose an appropriate link transformation function. For a binary dependent variable that takes value 0 or 1, the standard link function is the logit formulation. The second step consists of specifying the distribution of the dependent variable. I follow the literature, which uses the binomial distribution when the responses are binary. Finally, the last step is to specify the form of correlation within subjects. I use the exchangeable correlation structure, which assumes that the within-subject observations are equally correlated over time, which is also the default covariance structure that is used with the population averaged logit model. One advantage is that the GEE estimated coefficients are

based on the first moment, and they will thus be robust to misspecification of the correlation structure (Liang and Zeger, 1986). The GEE model only requires specification of the first moment (mean structure) and the second moment (covariance structure) without any distributional assumptions of the population.

Therefore, the main GEE regression equation for the recruitment model takes the following form:

$$E(Y_{i,t}/IMS_i, X_{it}, \vartheta_{it}, [I_{i,t}], \lambda_t) = g(\beta_1 IMS_i + X_{it}\delta + \lambda_t + \vartheta_{itp} + [I_{i,t}\emptyset]) \qquad (2)$$

where g(.) is the link function, and the dependent variable $Y_{i,t}$ is the likelihood of an employee working in the sector corresponding to the individual *i* and to the year *t*. The likelihood of working in the agriculture or agri-food sector depends on pre-determined characteristics. More precisely, the key variable of interest is IMS_i, a set of dummy variables for immigration status indicating if the employee is a Canadian-born, an immigrant without P.L.C.E, or an immigrant with P.L.C.E. The reference group is the Canadian-born workers. X_{it} is a vector of socio-demographic characteristics, with both time-dependent and time-independent covariates, which includes age, marital status, and gender. Age may impact the marginal value of time in each sector. If the marginal value of an increase in age is higher in another sector, I would expect the employee to be more attracted to this particular sector. Finally, I incorporate a province-of-employment fixed effect ϑ_{itp} that aims to capture difference in employment opportunities, and λ_t reflects the impact of year dummies that control for aggregate macroeconomics trends such as the stage of the business cycle. In additional specifications, I omit Canadian-born workers and include a vector of immigrant characteristics denoted as $I_{i,t}$. This vector includes the years since

migration, immigration class, education at time of the landing, and knowledge of an official language at time of landing. I use robust standard errors clustered at the individual level to correct for the correlation of error terms across individuals.

2.4.3 MODEL 2- RETENTION MODEL

After investigating the likelihood of working in the sector, an important question is to ask which factors influence the likelihood of retention. I am interested in the time elapsed before an employee who entered the sector quits the sector. Since the focus is on the duration of time until an event occurs, a survival analysis framework is the best-suited methodology. The CEEDD was constructed using data from tax filers, and so its frequency is annual. Consequently, the unit of observation for this analysis is the employment spell; defined as the number of consecutive years an employee worked in the industry before quitting. Because of the nature of the data, I decide to treat time as a discrete variable and base my analysis on a discrete survival model. The first step before conducting the estimation is the reorganization of the data into the following format, taking as examples two employees identified as individuals 1 and 2:

Identifier	Tax Year	Spell	Binary dependent variable (censoring variable=1 if quits, 0 if stays in the sector)
1	2001	1	0
1	2002	2	0
1	2003	3	0
1	2004	4	1
2	2011	1	0

2	2012	2	0
2	2013	3	0

Individual 1 has four years of observation, entering the sector in 2001 and exiting it in 2004. In comparison, individual 2 started working in 2011 and had not left the sector at the end of the sample period, thus generating a right-censored spell. The right censoring problem can be easily handled with survival analysis techniques. However, the more difficult problem of left censoring arises when constructing our dataset. This challenge can be illustrated using our above example; I see that individual 1 worked in the sector the first year of our sample period (2001), making it impossible to know exactly when this individual started working in the sector. This problem cannot be easily resolved, and without trying to correct for it, our results would be biased. Therefore, I follow the empirical literature related to discrete survival analysis by excluding the left-censored employment spells from the sample, and I include only employees who started working in the sector after 2001 and for which I observe starting date. In the primary agriculture approximately, 17 percent of individuals are discarded due to left censoring; this figure is 28 percent for the food and beverage processing sector.

Another problem that I face in the restructuring of our discrete survival sample involves employees who have multiple spells of employment in the sector. For instance, individual 1 could have stopped working in 2003 for a few years and started working again in the sector from 2008 to 2013. This is an issue that cannot be easily handled, and for simplicity in this analysis I limit my analysis to single spells and to the first spell for employees with multiple spells. In primary agriculture 8.6 percent of individuals have a multiple spells, and 8.4 percent in the food processing sector.

The survivor function for the discrete time model is given by the following equation:

$$S_i(j) = \Pr(T_i > j) = (1 - h_{i1})(1 - h_{i2})(1 - h_{i3}) \ldots\ldots \left(1 - h_{ij}\right) = \prod_{k=1}^{j}(1 - h_{ik}) \qquad (3)$$

The survivor function gives the probability of survival beyond the period *j*. $T_i = \min\{T_i^*, C_i^*\}$ with T_i denoting some latent failure time, and C_i some latent censoring time. In other words, it is the number of years before an employee working in the sector quits; it can be either a complete spell (denoted by $c_i = 1$) or a right-censored spell (denoted by $c_i = 0$). h_k is the discrete hazard rate of quitting the sector in period *j*, conditioned on survival up to period *j-1*. For instance, the probability of surviving up to 3 years will be equal to the probability of survival in period 1 multiplied by the probability of survival in period 2 (based on the survival in period 1); multiplied by the probability of survival in period 3, (based on survival in period 2). The discrete hazard rate of employee *i* is given by

$$h_i(j) = \Pr\left(T_i \leq j \middle| T > j - 1, (IMS_i, X_{ij}, Z_{it}, \vartheta_{itp}, \left[I_{i,t}\right], \lambda_t, \gamma_j)\right) \qquad (4)$$

The explanatory variables that affect the likelihood of staying after entering the sector include the independent variables used in the recruitment model. I also add a time-varying matrix Z_{it} containing firm characteristics where the individual works, which includes the logarithm of the number of employees to proxy for firm size and a variable indicating whether a firm, is a low, medium, or high-paying firm. For the classification of this variable, I borrow Ci and Hou (2017)'s formulation, by first calculating the median annual employment earnings offered by firms and then separating them into three groups at the industry level (all firms in the industry for Canada as a whole). Firms with median employment earnings at or under the 25th percentile are considered to be low-paying firms, those with median employment earnings from the 25th to 50th percentiles are medium-paying firms, and finally high-paying firms are the ones with median

employment earnings from the 50th to 100th percentiles. However, the interpretation of this variable should be made with caution, since firms' classification into these three groups can be influenced by the labour intensity choices of their workers. For instance, firms with a larger number of employees who work part time might be classified as a lower-paying firm. By taking the median instead of the average, I try to mitigate this concern. Lastly, I include a variable γ_j capturing the structure of duration dependence, which is modelled by the log of the spell elapsed. In other words, this variable represents the baseline hazard function, which is function of survival time *t* per person.

The discrete time representation of an underlying continuous time model is obtained when the hazard rate is assumed to follows a complementary log-log distribution (Prentice and Gloeckler, 1978). Then equation (4) can be rewritten as:

$$h\left(IMS_i, X_{ij}, Z_{it}, \vartheta_{itp}, \left[I_{i,t}\right], \lambda_t, \gamma_j\right) = 1 - \exp\left[-exp\left(\beta_1 IMS_i + X_{it}\delta + Z_{it}\theta + \left[I_{i,t}\sigma\right] + \vartheta_{itp} + \lambda_t + \gamma_j\right)\right] \quad (5)$$

After simplifying and taking logs, I obtain the following implicit form for the hazard rate:

$$\log\left(-log\left[1 - h\left(IMS_i, X_{ij}, Z_{it}, \vartheta_{itp}, \left[I_{i,t}\right], \lambda_t, \gamma_j\right)\right]\right) = \beta_1 IMS_i + X_{it}\delta + Z_{it}\theta + \left[I_{i,t}\sigma\right] + \vartheta_{itp} + \lambda_t + \gamma_j \quad (6)$$

In the above model, I only used observed explanatory variables that are available in our dataset and ignore unobserved individual effects. The problem of unobserved individual effects can arise from omitted variables or measurement errors. Without correcting for this issue, our results could be biased and either over or underestimate the parameters. Therefore, I estimate additional specifications by adding unobserved individual heterogeneity terms (u_i) which is a random variable assumed to be normally distributed.[16] Our main equation is now given by:

$$\log\left(-log\left[1 - h\left(IMS_i, X_{ij}, Z_{it}, \vartheta_{itp}, \left[I_{i,t}\right], \lambda_t, \gamma_j\right)\right]\right) = \beta_1 IMS_i + X_{it}\delta + Z_{it}\theta + \left[I_{i,t}\sigma\right] + \vartheta_{itp} + \lambda_t + \gamma_j + u_i \quad (7)$$

[16] A detailed explanation of the use of unobserved heterogeneity for discrete time survival analysis is presented by S.Jenkins and is available at: https://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968st7.pdf

2.5 THE RESULTS

2.5.1 RESULTS OF THE RECRUITMENT MODEL

The main GEE results for the equation modelling the likelihood of working in the primary agriculture and food processing sector are presented in table 2.4 and table 2.5. Table 2.4 estimates the participation model by considering first the full sample of immigrants (15% of the sample) versus Canadian-born individuals (85% of the sample). After having obtained a better understanding of differences in participation in the agriculture and agri-food sectors between immigrants and Canadian-born, the results in table 2.5 are estimates of the participation model for immigrants with P.L.C.E (4%) and immigrants with non-P.L.C.E (11%). The first four columns pertain to the primary agriculture sector, and the last four columns are for the food processing sector. I estimate two principal specifications: one that includes only indicators for immigration status to capture the effect without covariates, and another that includes all the explanatory variables that I presented in the previous section. I also estimate the respective marginal effects of the explanatory variables.

2.5.2 RESULTS OF THE RECRUITMENT MODEL FOR THE PRIMARY AGRICULTURE SECTOR

The results of table 2.4 for the primary agriculture sector suggest that immigrants are less likely than native-born Canadian to work in the agriculture sector. The marginal effects of 0.16 percentage points (column 4) need to be interpreted relative to the average probability of working in the agriculture sector, which is equal to approximately 1%. This result suggests that immigrants are approximately 16% less likely to enter the agriculture sector than the Canadian-born.

Among immigrants, the estimated coefficients for immigration status, when divided into

two groups (table 2.5, column 4), indicate that immigrants with P.L.C.E are less likely to work in the sector than immigrants without P.L.C.E. The p-value of the equality test shows significant differences between those groups. This is likely related to the fact that most immigrants with P.L.C.E had their experience from sectors other than the agriculture sector, and this is also consistent with the fact that the current immigration program is directed more toward facilitating the pathway to high-skill foreign workers than to low-skill foreign workers. In fact, the difference between immigrants with P.L.C.E and immigrants without P.L.C.E become very small and not economically significant (0.01 percentage points) after removing the effects of all immigrants pre-landing characteristics that capture some of the effect mentioned below, such as education and immigration class (table 2.6, column 4).

In general, the results for the primary agriculture sector suggest that immigrants are less attracted to work in this industry than Canadian-born workers, but that there is no important difference between immigrants with P.L.C.E and those without P.L.C.E. One potential explanation is that immigrants are more attracted and concentrated in cities and less attracted to regions where jobs related to primary agriculture are located.

The findings for the other explanatory variables for the primary agriculture sector confirm many of our initial expectations. Males are more likely to work in the sector than females. The marginal effects related to age decline as age increases, suggesting that young workers are more likely to work in the sector than older workers. This could be in part explained by the fact that working in the agricultural sector often involves physical work and more demanding working conditions. I also observe that being married has a positive impact on the likelihood of working in the primary agricultural sector. In terms of the work location, I see that workers in Manitoba and Saskatchewan are those with the highest likelihood of working in the

agricultural sector. This may reflect better agricultural job opportunities and the importance of that sector size in those provinces.

In addition, I repeated the above analysis by omitting Canadian-born workers, and adding a rich set of immigrant landing characteristics available in the IMDB. The results are presented in table 2.6. An interesting result is that when considering only immigrants, I find that women are more likely than men to work in the sector, but that the difference is not economically significant (0.01 percentage points, approximately 1% difference). The result suggests that there is no important difference between men and women immigrants in entering the agricultural sector. On the other hand, in the prior analysis for the sample including Canadian-born workers, I found that men are significantly more likely to work in the agricultural sector than women. The difference between immigrant and Canadian-born women might be related to the fact that immigrant women have less mobility with fewer opportunities to choose to work in other sectors that require less demanding working conditions due to language barriers, networking, and credentials.

The estimated coefficients for years since migration are negative and statistically significant, indicating that immigrants become less attracted to the agricultural sector as they spend more time in Canada. The findings also suggest that immigrants who come through the Provincial Nominee Program are more likely to enter the primary agricultural sector when compared to all other immigration classes. This could be related to the fact that the Provincial Nominee Program is driven by provinces with the aim of filling labour market shortages and improving regional economic development. The agricultural sector has exhibited labour shortages and is one of the sectors that the PNP is trying to assist with new immigrants. Finally, I also find a negative relationship between immigrants' levels of education at time of the landing

and the likelihood of entering this sector, which is consistent with the literature mentioned above on off-farm employment.

2.5.3 Results of the Recruitment Model for the Agri-Food Sector

For the specification modelling the likelihood of working in the food processing sector, I found distinct and interesting results. The results are presented in columns (5) to (8) of table 2.4, 2.5 and 2.6. Immigrants are significantly more likely to work in this sector relative to the native-born. The marginal effect in column 8 of table 2.4 indicates that immigrants are 0.83 percentage points more likely to enter the food and beverage processing sector. In the context of the average probability of working in the sector being equal to approximately 2 percent, the marginal effect suggests that immigrants are approximately 45% more likely to enter the agri-food sector than the native-born. These results indicate that immigrants are on average more attracted to this type of work. I again do not find any significant difference between immigrants without P.L.C.E than those with P.L.C.E. The estimates of other covariates also show that males are more likely to work in this sector than females. The marginal effects related to age increases with age, which suggests that older workers are this time more likely to work in the sector than younger workers. This might reflect less demanding working conditions when compared to the primary agriculture sector. The findings also show that workers in the Atlantic regions are significantly more likely to work in the food processing sector compared to all other provinces. This can be in part explained by the high concentration of seafood processing in this region.

I again repeated the analysis by omitting Canadian-born workers, and adding a rich set of immigrant landing characteristics available in the IMDB. The results are presented in columns (5) to (8) of table 2.6. I still find no significant difference between the attraction of immigrants with P.L.C.E and immigrants without P.L.C.E. Estimated coefficients for the years since

migration variable are again negative and statistically significant. Similarly, to the primary agricultural sector, immigrants are more likely to enter the sector through the Provincial Nominee Program. In terms of the education variable, I still find a negative and significant relationship between the levels of education and attraction to this sector.

2.5.4 Results of the Retention Model for the Primary Agriculture Sector

The discrete survival results for the equation modelling the likelihood of quitting the primary agricultural sector are presented in columns (1) to (4) of table 2.7, 2.8 and 2.9. The displayed estimated coefficients are the hazard ratios. An estimated hazard ratio smaller (greater) than 1 is interpreted as longer (shorter) level of retention compared to the reference group and a hazard ratio equal to 1 suggests no difference in duration relative to the reference group. The first column includes the baseline hazard, immigration status and year fixed effects. The second column adds a vector of socio-demographic characteristics, which consists of age, province of employment, marital status, and gender. The third column is the richest specification, with the addition of firm characteristics, and the fourth column is similar to the third column but accounts for individual unobserved heterogeneity through an individual random effect.

Table 2.7 estimates the retention model by considering first the full sample of immigrants versus Canadian-born individuals. The results of our baseline specification (column 1) indicate that immigrants have higher retention than Canadian-born workers, but the difference is not economically significant (1.8% lower risk of quitting the sector). However, if I remove the effects of socio-demographic characteristics and firm characteristics, the results suggest that Canadian-born workers have now higher retention in the sector than immigrant workers. The preferred specification (column 4) that account for individual unobserved heterogeneity suggest

that Canadian-born workers have approximately a 23% lower risk of quitting the sector after entering than immigrant workers.

Table 2.8 contains the estimates of the retention model for immigrants with P.L.C.E and immigrants with non-P.L.C.E separately. As anticipated from the descriptive analysis, the results of our baseline specification (column 1) indicate that immigrant workers with P.L.C.E stay longer in the sector than Canadian-born workers and other immigrant workers without P.L.C.E. However, when I control for socio-demographic characteristics (column 2) and firm characteristics (column 3), this higher level of retention disappears, and Canadian-born workers are now the group that stays the longest in the sector. Finally, in the fourth specification, when I account for unobserved heterogeneity, the findings indicate that there is no significant difference between Canadian-born and immigrant workers with P.L.C.E. In this last specification, the likelihood ratio test suggests the existence of statistically significant unobserved heterogeneity and confirms that the frailty has expected effects on the estimated coefficients. Therefore, the model with unobserved heterogeneity is the preferred model, and I conclude that there is no significant difference between the Canadian-born and immigrant workers with P.L.C.E in terms of retention for this sector, but that the latter group has greater retention than immigrant workers with only foreign work experience.

The results for the impact of socio-demographic characteristics on the retention of agricultural workers are quite interesting. The findings in the specification that control for socio-demographic characteristics suggest that males stay longer in the sector than females. However, as I control for firm characteristics, I find that females are now the ones with higher retention levels. These results imply that if males and females work for similar firms, females will have higher retention levels than their male counterparts. In particular, column 4 shows that males

have 26.9%[17] higher hazard (risk) of quitting the sector than females. The results for age are also consistent across specifications, indicating that older workers have higher retention rates than younger workers. I also find that married workers are significantly more likely to stay longer than non-married workers. The findings for the impact of firm characteristics indicate that workers working in high-paying firms have significantly higher levels of retention than other groups and those working in smaller firms tend to stay longer than those working in larger firms. Finally, the results for ln (spell) that capture the effect of duration show an important and significant impact in all our specifications. This suggests that an additional year of working in the sector increases significantly the likelihood of quitting the sector. This is in agreement with our non-parametric analysis that showed a steep slope of the Kaplan-Meier curve in the first to 5 years of working in the sector. However, this result needs to be interpreted with caution, because in the context of this study, the first spell is the year of entry of the workers in the sector, and we can observe whether he exits the sector or not only in the second spell. Therefore, there is no exit in the first spell, and this is likely why we observe very large coefficients in some of the specifications.

I also repeated the analysis by omitting Canadian workers and including a rich set of immigrant landing characteristics (table 2.9). I still find that immigrants with P.L.C.E have a higher likelihood of staying in the sector. In particular in the richest specification (column 4), immigrants with P.L.C.E have 37.8% lower risk of quitting the sector than immigrants without P.L.C.E. For other control variables, the results show similar empirical patterns to the above findings, with female immigrants being more likely to stay longer than male immigrants, older immigrants being more likely to stay longer than younger immigrants, and married immigrants

[17] This percentage and other percentages for the retention model are obtained by subtracting the hazard ratio from 1.

being more likely to stay longer on farms than non-married immigrants. The estimated coefficients for firms' characteristics also show that immigrants working in high-paying firms and smaller farms have higher likelihoods of staying in the sector.

Immigrants' landing characteristics provide interesting further information. The years since migration variable has a positive impact on the retention of agricultural workers. One additional year in the host country decreases the hazard of quitting the sector by 2.4%. Immigrants entering the country through the Provincial Nominee Program is the group that on average stays the longest in the sector. I also find a negative significant relationship between the level of education and retention and that immigrants with knowledge of an official language at the time of landing have lower levels of retention than other immigrants.

2.5.5 Results of the Retention Model for the Agri-Food Sector

The hazard ratios for the main results of our food processing sample are presented in columns (5) to (8) of table 2.7, 2.8 and 2.9. The results are in agreement with the preliminary evidence previously discussed in the summary statistics section. The coefficient estimates in table 2.7 show that immigrants remain longer in the sector than Canadian-born workers. The results are all highly significant and consistent across specifications. More precisely, the results of the preferred specification, column (4), indicate that immigrant workers face an 18.1% lower hazard rate of quitting the sector than the Canadian-born workers.

Table 2.8 provide further interesting evidence when immigrants are divided into those with P.L.C.E and those without. The coefficient estimates in table 2.8 show that immigrants with P.L.C.E remain longer in the agri-food sector than immigrants without P.L.C.E and Canadian-born workers. The results are also all highly significant and consistent across specifications. The results of the preferred specification (column (8)) indicate that immigrants with P.L.C.E face a

30.5% lower hazard rate of quitting the sector than the Canadian-born workers, and immigrants without P.L.C.E face a 13% lower hazard rate of quitting the sector than Canadian-born workers.

The findings for the socio-demographic characteristics and firms' characteristics are generally similar to the results obtained from the primary agriculture sector with a few differences. I find that women remain longer than men, and older workers remain longer than younger workers. Being married has also a positive impact on the likelihood of staying in the food processing sector. In regard to the firm characteristics, workers in high-paying firms stay significantly longer than other workers, and this time there is a positive relationship between the size of the business and the retention of workers. Those findings are also robust across specifications and statistically significant.

Finally, I once more repeated the analysis with immigrants only (table 2.9). The results confirm that immigrants with P.L.C.E stay longer than other immigrants. The results also still indicate that immigrants coming through the Provincial Nominee Program are more likely to stay longer in the sector when compared to all other immigration classes. Again, I find a negative relationship between the level of education and the retention of workers, and newcomers with knowledge of an official language at the time of landing have a higher likelihood of quitting the sector than those who do not.

2.6 CONCLUSION

Over recent decades, agriculture and agri-food stakeholders have raised concerns about the difficulty in recruiting and retaining workers. Labour migration has been utilized as one of the key policy measures to address this problem, with the use of TFWs to fill short-term labour demand requirements and immigration inflows for longer-term labour needs. A key strategy that has been proposed lately is to facilitate pathways to permanent residency for foreign workers

who are deemed to be filling long-term labour market needs. Drawing on data from the Canadian Employer-Employee Dynamic Database for the years of 2001 to 2013, I have investigated the role of immigrants P.L.C.E in the choice of entering and remaining in the agriculture and agri-food sectors. I also further investigated the role of individual socio-demographic characteristics and firm characteristics on the workers' decisions of entry and exit in those sectors.

In regards to the decision to enter these sectors, my findings suggest that immigrants are less attracted to the primary agricultural sector, but more attracted to the food processing sector than native Canadians. I further found a negative relationship between immigrants' level of education and the degree of attraction to both the agriculture and agri-food sectors. Finally, another interesting result for the attraction to these sectors is that immigrants coming through the Provincial Nominee Program are more likely to enter the agriculture and the agri-food sectors.

The key results for the retention of employees suggest that immigrants with P.L.C.E stay longer in the agricultural sector when compared to immigrants with only foreign work experience; I found no statistically significant difference between immigrants with P.L.C.E and native Canadians. In regard to the food processing sector, I found that immigrants with P.L.C.E are the group that remains significantly the longest in this sector, followed by other immigrants without P.L.C.E and native Canadians. Another interesting result is that I found a negative relationship between the level of education and the retention rate for both sectors. I also found for both sectors that female workers stay longer than male workers, married workers longer than non-married workers, older workers longer than younger workers, and those employed in high-paying firms longer than other workers.

In general, the results confirm that immigrants with P.L.C.E have higher retention rates than other immigrants in the agriculture and agri-food sectors. This provides support for the current

government effort to facilitate pathways to permanent residency for foreign individuals with Canadian work experience. The results also suggest that resorting to the Provincial Nominee Program is a good strategy for the agricultural and agri-food sectors. According to the findings, the Provincial Nominee Program seems to be an effective immigration program in terms of recruitment and retention in the agricultural and agri-food sectors. Provincial and territorial governments are the best informed about the particular employment needs of their province or regions, and better outcomes can be obtained by letting them nominate the individuals who wish to immigrate and settle in their province. Another key result is the negative relationship between the level of education of immigrants and attraction and retention in those sectors. This provides support in reducing the importance of education for immigration programs that focus on attracting and retaining immigrants to regions for filling a labour shortage. There is more opportunity for lower-skilled jobs in regions than in large cities.

Some immigration programs are already doing this. For instance, the Provincial Nominee Program (which is not based on the points system) have streams that give lower importance to education in terms of filling labour shortage in those provinces. A good example is the Provincial Nominee Program in Saskatchewan, which started in 2019. It consists of a Semi-Skilled Agriculture Workers Category (which is under the Saskatchewan Experience Category) that enables producers to recruit both low-skilled General Farm Workers and Nursery and Greenhouse Workers on a permanent, year-round basis. In general, facilitating a pathway to permanent residency for lower-skilled immigrants in regions can have a number of benefits, such as filling a labour shortage and also improving the retention of immigrants in regions, and therefore helping with a more balanced national distribution of immigrants in Canada. However, this approach is not perfect and may also involve some risks, such as the possibility of immigrant

workers losing their jobs and moving to larger cities where their labour market outcome might be weaker compared to other immigrants and the native-born.

Furthermore, the results suggest that there is a significant positive impact on retention for workers in high-paying firms. Given this finding, a case can be made for the sector to offer higher wages in efforts to increase the level of retention. However farm operators are often price takers and have limited influence on the market prices in the output market, and increasing wages in this environment could negatively impact their competitiveness.[18]

Another concern from a long-run efficiency perspective is that one can point out that relying on international immigration to address excess demand for labour in agriculture is not desirable because one can argue that the sector is not competitive and that it might be better to let the sector shrink and shift production factors to sectors that are more productive. However letting the agriculture sector shrink can have some important consequence for a country. A good example is the case of the food policy in Russia from 1992-99, where the agriculture sector was left to market forces and which led to the unprofitability of most of the farms and to significant increases in food price inflation and food insecurity in the country (Wegren, 2016).

[18] According to an analysis conducted by the Conference Board of Canada (2016), "Many of the products grown by the sector are commodities, meaning that producers have only a limited ability to distinguish their products from others. As a result, prices for their products are generally set with a price ceiling that is decided by regional or global markets."

REFERENCES 2

Akbari, Ather H. 2011. "Labor Market Performance of Immigrants in Smaller Regions of Western Countries: Some Evidence from Atlantic Canada." *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale* 12(2): 133-154.

Aydemir, Abdurrahman, and Mikal Skuterud. 2005. "Explaining the Deteriorating Entry Earnings of Canada's Immigrant Cohorts, 1966–2000." *Canadian Journal of Economics/Revue canadienne d'économique,* 38 (2): 641-672.

Beaudry, Paul and David A. Green. 2000. "Cohort Patterns in Canadian Earnings: Assessing the Role of Skill Premia in Inequality Trends." *Canadian Journal of Economics/Revue canadienne d'économique,* 33(4): 907-936.

Beine, Michel, Serge Coulombe, and Wessel N. Vermeulen. 2015. "Dutch Disease and the Mitigation Effect of Migration: Evidence from Canadian Provinces." *The Economic Journal* 125(589): 1574-1615.

Benjamin, Catherine, and Ayal Kimhi. 2006. "Farm Work, Off-Farm Work, and Hired Farm Labour: Estimating a Discrete-Choice Model of French Farm Couples' Labour Decisions." *European review of agricultural economics,* 33(2): 149-171.

Bernard, André. 2008. "Immigrants in the Hinterlands." Ottawa: Statistics Canada.

Bloom, David E., Gilles Grenier and Morley Gunderson. **1995**. "The Changing Labour Market Position of Canadian immigrants." *Canadian Journal of Economics/Revue canadienne d'économique*, 28(4):987–1005.

Brochu, Pierre, Till Gross and Christopher Worswick. 2016. "Temporary Foreign Workers and Firms: Theory and Canadian Evidence" *Canadian Journal of Economics/Revue canadienne d'économique, forthcoming*.

Ci, Wen and Feng Hou. 2017. "Immigrants' Initial Firm Allocation and Earnings Growth." *Canadian Studies in Population*, 1(2):42-58.

Conference Board of Canada. 2016. "Sowing the Seeds of Growth. Temporary Foreign Workers in Agriculture." http://www.conferenceboard.ca/press/newsrelease/16-12 01/Agriculture_Sector_s_Reliance_on_Temporary_Foreign_Workers_to_Fill_Labour_Shortages _is_Growing.aspx.

Ferrer, Ana, and W. Craig Riddell. 2008. "Education, Credentials, and Immigrant Earnings." *Canadian Journal of Economics/Revue canadienne d'économique*, 41(1): 186-216.

Goodwin, Barry K., and Matthew T. Holt. 2002. "Parametric and Semiparametric Modelling of the Off-Farm Labor Supply of Agrarian Households in Transition Bulgaria." *American Journal of Agricultural Economics,* 84(1): 184-209.

Gould W. Brian and William E. Saupe. 1989. "Off-Farm Labor Market Entry and Exit" *American Journal of Agricultural Economics,* 71(4):960–969.

Government Response. 2017. Retrieved from Parliament of Canada Website: https://www.ourcommons.ca/DocumentViewer/en/42-1/HUMA/report-4/response-8512-421-86

Green, David A. and Christopher Worswick. 2012. "Immigrant Earnings Profiles in the Presence of Human Capital Investment: Measuring Cohort and Macro Effects." *Labour Economics* 19(2): 241-259.

Gross, Dominique M., and Nicolas Schmitt. 2012. "Temporary Foreign Workers and Regional Labour Market Disparities in Canada." *Canadian Public Policy,* 38(2): 233-263.

Hennessy, Thia C., and Tahir Rehman. 2008. "Assessing the Impact of the 'Decoupling' Reform of the Common Agricultural Policy on Irish Farmers' Off Farm Labour Market Participation Decisions." *Journal of Agricultural Economics,* 59(1): 41-56.

Hou, Feng, and Garnett Picot. 2016. "Changing Immigrant Characteristics and Pre-Landing Canadian Earnings: Their Effect on Entry Earnings Over the 1990s and 2000s." *Canadian Public Policy,* 42(3): 308-323.

Huffman, Wallace. 1980. "Farm and Off-Farm Work Decisions: The Role of Human Capital." *The Review of Economics and Statistics,* 62(1):14-23.

Huffman, Wallace and Mark D. Lange. 1989. "Off-Farm Work Decisions of Husbands and Wives: Joint Decision Making." *The Review of Economics and Statistics,* 71(3): 471-480.

HUMA Committee Report. 2016. Retrieved from Parliament of Canada Website: https://www.ourcommons.ca/DocumentViewer/en/42-1/HUMA/report-4/

Juvančič, Luka, and Emil Erjavec. 2005. "Intertemporal Analysis of Employment Decisions on Agricultural Holdings in Slovenia." *Agricultural Economics,* 33(2): 153-161.

Kimhi, Ayal. 1994. "Participation of Farm Owners In Farm And Off Farm Work Including The Option of Full Time Off Farm Work." *Journal of Agricultural Economics,* 45(2): 232-239.

Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika,* 73 (1): 13-22.

Prentice, Ross L., and Lynn A. Gloeckler. 1978. "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data." *Biometrics,* 34(1): 57-67.

Schaafsma, Joseph, and Arthur Sweetman. 2001. "Immigrant Earnings: Age at Immigration Matters." *Canadian journal of Economics/Revue canadienne d'économique,* 34(4):1066-1099.

Sumner, A. **Daniel.** 1982. “The Off-Farm Labor Supply of Farmers.”American Journal of Agricultural Economics, 64(3):499–509.

Sweetman, Arthur, and Casey Warman. 2014. “Former Temporary Foreign Workers and International Students as Sources of Permanent Immigration.” *Canadian Public Policy*, 40(4): 392-407.

Warman, Casey R. and Christopher Worswick. 2004. “Immigrant Earnings Performance in Canadian Cities: 1981 through 2001.” *Canadian Journal of Urban Research*, 13(1): 62-84.

Wegren, Stephen K. 2016. "Food policy in Russia." *Reference Module in Food Science*, 1(1): 15-20.

Figure 2.1: Kaplan-Meier survival estimates for the full sample in Primary Agriculture

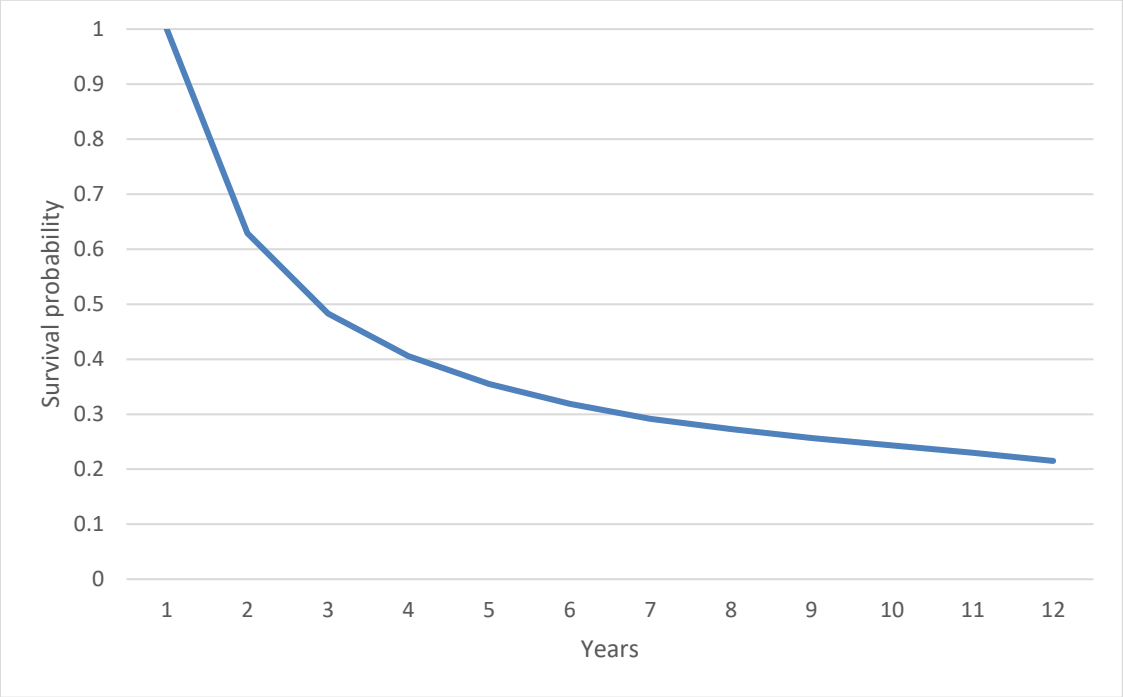


Figure 2.2: Kaplan-Meier survival estimates in Primary Agriculture, by immigration category

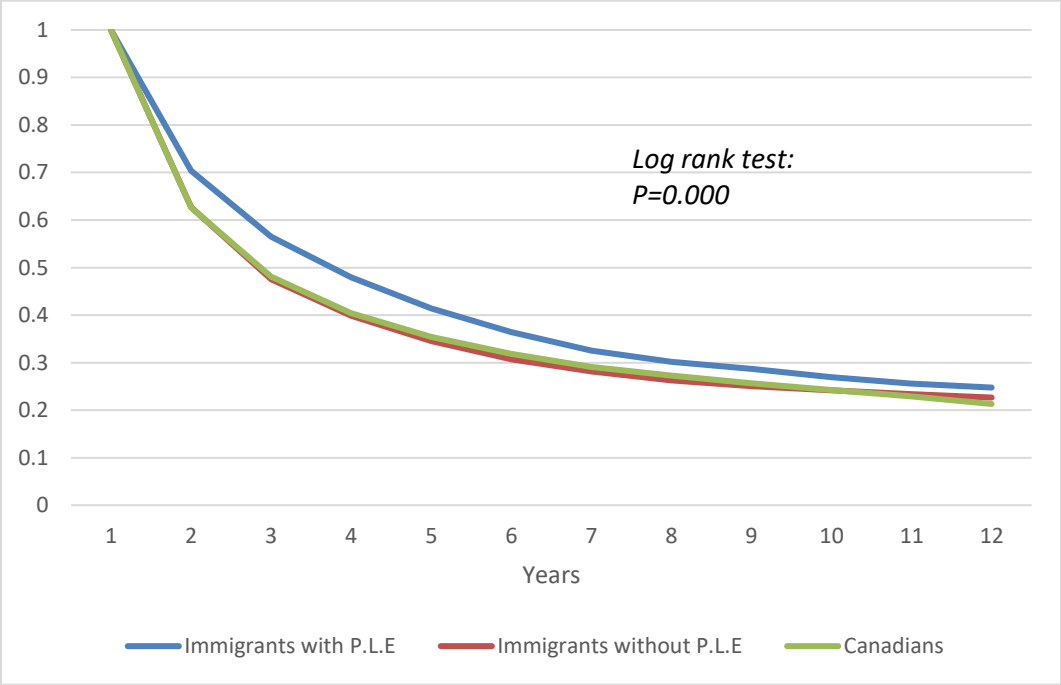


Figure 2.3: Kaplan-Meier survival estimates for the full sample in Food processing

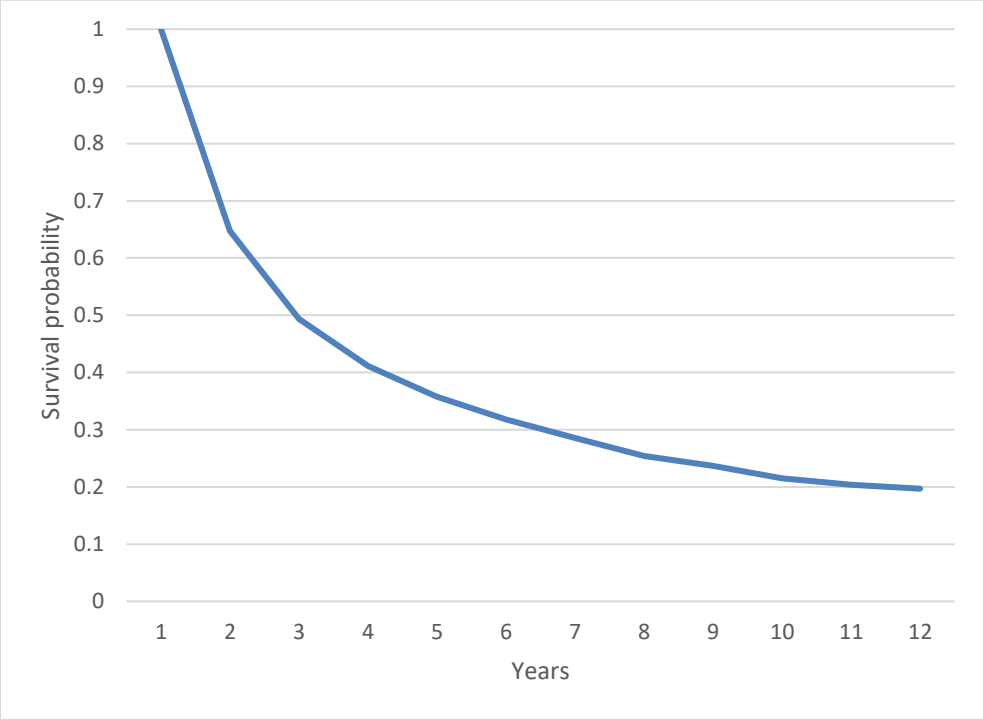


Figure 2.4: Kaplan-Meier survival estimates in Food processing, by immigration category

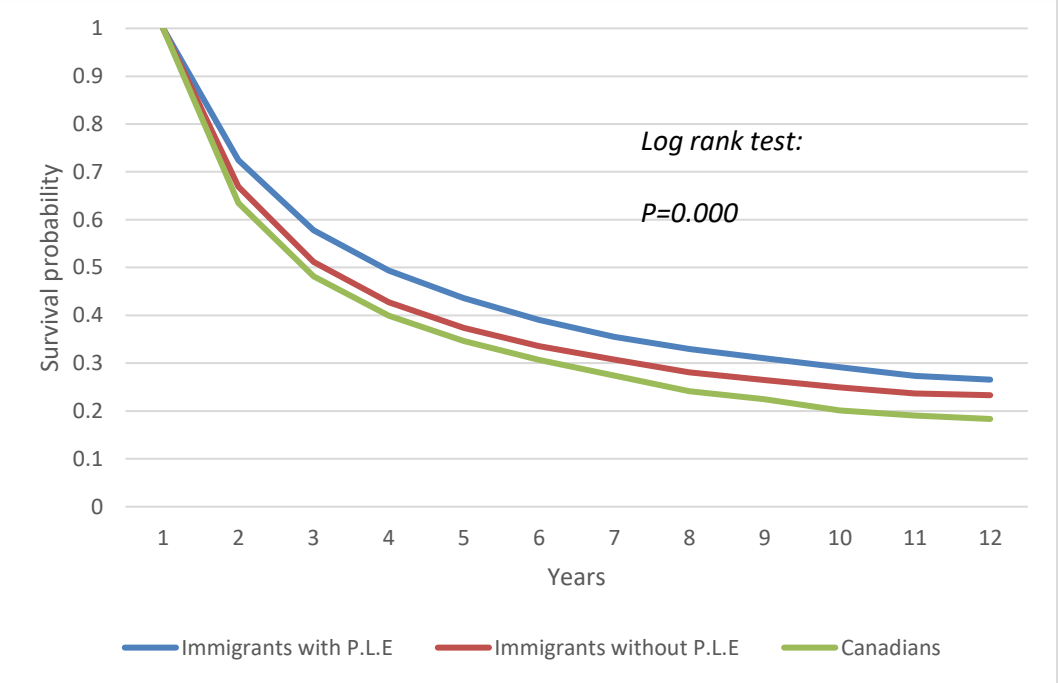


Table 2.1: Descriptive Statistics of major variables in the recruitment model

	Full sample	Only Immigrants
	Means (Std.dev)	Means (Std.dev)
Canadian-born	0.849	
Immigrants without P.L.C.E	0.112	0.740
Immigrants with P.L.C.E	0.039	0.260
Male	0.502	0.499
Age	38.94	38.36
	(12.53)	(11.04)
Atlantic	0.073	0.010
Quebec	0.238	0.146
Ontario	0.381	0.535
Manitoba	0.036	0.027
Saskatchewan	0.029	0.010
Alberta	0.115	0.107
British Columbia	0.124	0.162
Married	0.447	0.595
Year since migration		10.99
Knowledge of an official language at time of landing		0.655
Immigration class:		
Canadian Experience Class (principal applicant)		0.005
Family class (principal applicant)		0.306
Skilled workers (principal applicant)		0.195
Business class (principal applicant)		0.005
Provincial nominee (principal applicant)		0.017
Refugee (principal applicant)		0.146
Others		0.325
Education at time of landing:		
High school dropout		0.485
High school		0.081
Trade or college		0.156
Bachelor		0.209
Graduate degree		0.068
Last place of residence:		
Europe		0.209
Africa		0.084
Asia		0.528
Central America		0.056
South America		0.111
Oceania		0.013
Observations	57,049,066	8,604,111
Individuals	6,971,740	1,222,067
Average number of years a person is observed	8.183	7.041

Table 2.2: Descriptive Statistics of major variables in the retention model for the Primary Agriculture sector

	Full sample	Only immigrants
	Means (Std.dev)	Means (Std.dev)
Canadian-born	0.872	
Immigrants without P.L.C.E	0.090	0.704
Immigrants with P.L.C.E	0.038	0.296
Male	0.623	0.532
Age	35.984	39.298
	(13.49)	(12.28)
Atlantic	0.122	0.024
Quebec	0.279	0.112
Ontario	0.381	0.475
Manitoba	0.032	0.038
Saskatchewan	0.027	0.013
Alberta	0.068	0.075
British Columbia	0.091	0.263
Married	0.399	0.621
Low paying firm	0.238	0.168
Medium paying firm	0.505	0.536
High-paying firm	0.257	0.296
Ln(employment)	3.855	4.500
	(1.74)	(1.75)
Years since migration		8.530
		(7.993)
Knowledge of an official language at time of landing		0.457
Immigration class:		
Canadian experience class (principal applicant)		0.004
Family class (principal applicant)		0.450
Skilled workers (principal applicant)		0.076
Business class (principal applicant)		0.010
Provincial nominees (principal applicant)		0.045
Refugee (principal applicant)		0.210
Others		0.206
Education at time of landing:		
High school dropout		0.697
High school		0.056
Trade or college		0.114
Bachelor		0.107
Graduate studies		0.027
Last place of residence:		
Europe		0.173
Africa		0.061
Asia		0.523
Central America		0.125
South America		0.113
Oceania		0.005
Observations	744,644	94,974
Individuals	267,715	34,168
Average number of years a person is observed	2.781	2.780

Table 2.3: Descriptive Statistics of major variables of the retention model for the Food processing sector

	Full sample	Only immigrants
	Means (Std.dev)	Means (Std.dev)
Canadian-born	0.749	
Immigrants without P.L.C.E	0.180	0.720
Immigrants with P.L.C.E	0.070	0.280
Male	0.581	0.551
Age	36.631	38.146
	(12.57)	(10.74)
Atlantic	0.160	0.011
Quebec	0.285	0.171
Ontario	0.407	0.595
Manitoba	0.022	0.035
Saskatchewan	0.012	0.008
Alberta	0.053	0.085
British Columbia	0.061	0.096
Married	0.394	0.615
Low paying firm	0.239	0.152
Medium paying firm	0.507	0.621
High-paying firm	0.254	0.227
Ln(employment spell)	6.075	6.069
	(2.04)	(2.10)
Years since migration		8.768
		(7.64)
Knowledge of an official language at time of landing		0.571
Immigration class:		
Canadian experience class (principal applicant)		0.003
Family class (principal applicant)		0.340
Skilled workers (principal applicant)		0.149
Business class (principal applicant)		0.005
Provincial nominees (principal applicant)		0.037
Refugee (principal applicant)		0.202
Others		0.264
Education at time of landing:		
High school dropout		0.550
High school		0.080
Trade or college		0.151
Bachelor		0.176
Graduate studies		0.043
Last place of residence:		
Europe		0.159
Africa		0.111
Asia		0.547
Central America		0.053
South America		0.121
Oceania		0.009
Observations	1,687,721	422,701
Individuals	532,855	131,689
Averaged number of years a person is observed	3.167	3.210

Table 2.4: GEE RESULTS OF THE LIKELIHOOD OF WORKING IN THE AGRICULTURE SECTOR AND THE FOOD PROCESSING SECTOR (FULL SAMPLE)

VARIABLES	(1) Primary agriculture Coefficients	(2) Primary agriculture Marginal effects	(3) Primary Agriculture Coefficients	(4) Primary agriculture Marginal effects	(5) Food processing Coefficients	(6) Food processing Marginal Effects	(7) Food processing Coefficients	(8) Food processing Marginal Effects
Immigration status (ref: Canadian-born)								
Immigrants	-0.217***	-0.0017***	-0.227***	-0.0016***	0.427***	0.0069***	0.539***	0.0083***
	(0.010)	(0.000)	(0.010)	(0.000)	(0.006)	(0.000)	(0.006)	(0.000)
Male			0.576***	0.0041***			0.395***	0.0061***
			(0.007)	(0.000)			(0.005)	(0.000)
Age			-0.013***	-0.0001***			0.0055***	0.00009***
			(0.000)	(0.000)			(0.000)	(0.000)
Married			0.113***	0.0008***			0.0076*	0.0001*
			(0.005)	(0.000)			(0.004)	(0.000)
Provinces (ref: Quebec)								
Atlantic			0.349***	0.0025***			0.812***	0.012***
			(0.013)	(0.000)			(0.009)	(0.000)
Ontario			0.048***	0.0003***			-0.098***	-0.0015***
			(0.009)	(0.000)			(0.007)	(0.000)
Manitoba			0.485***	0.003***			0.186***	0.0028***
			(0.017)	(0.000)			(0.016)	(0.000)
Saskatchewan			0.464***	0.0033***			-0.120***	-0.0018***
			(0.020)	(0.000)			(0.019)	(0.000)
Alberta			-0.218***	-0.0015***			-0.309***	-0.0048***
			(0.013)	(0.000)			(0.011)	(0.000)
British Columbia			0.286***	0.002***			-0.170***	-0.0026***
			(0.012)	(0.000)			(0.010)	(0.000)
Year fixed effects	Yes	Yes	**Yes**	Yes	Yes	Yes	Yes	Yes
Observations				57,049,066				
Number of individuals				6,971,740				

Notes: ***p<0.01, **p<0.05, *p<0.1; Standard errors in parentheses are clustered at the individual level. For the province of employment, the territories are excluded.

Table 2.5: GEE RESULTS OF THE LIKELIHOOD OF WORKING IN THE AGRICULTURE SECTOR AND THE FOOD PROCESSING SECTOR (FULL SAMPLE/IMMIGRANTS DIVIDED INTO TWO GROUPS)

VARIABLES	(1) Primary agriculture Coefficients	(2) Primary agriculture Marginal effects	(3) Primary Agriculture Coefficients	(4) Primary agriculture Marginal effects	(5) Food processing Coefficients	(6) Food processing Marginal Effects	(7) Food processing Coefficients	(8) Food processing Marginal Effects
Immigration status (ref: Canadian-born)								
Immigrants without P.L.C.E	-0.180***	-0.0014***	-0.201***	-0.0014***	0.433***	0.007***	0.547***	0.008***
	(0.011)	(0.000)	(0.011)	(0.000)	(0.006)	(0.000)	(0.007)	(0.000)
Immigrants with P.L.C.E	-0.328***	-0.0025***	-0.307***	-0.0022***	0.411***	0.0067***	0.517***	0.008***
	(0.019)	(0.000)	(0.019)	(0.000)	(0.010)	(0.000)	(0.011)	(0.000)
Male			0.576***	0.004***			0.395***	0.006***
			(0.007)	(0.000)			(0.005)	(0.000)
Age			-0.013***	-0.0001***			0.005***	0.00009***
			(0.000)	(0.000)			(0.000)	(0.000)
Married			0.114***	0.0008***			0.0077*	0.0001*
			(0.005)	(0.000)			(0.004)	(0.000)
Provinces (ref: Quebec)								
Atlantic			0.349***	0.002***			0.812***	0.012***
			(0.013)	(0.000)			(0.009)	(0.000)
Ontario			0.047***	0.0003***			-0.099***	-0.001***
			(0.009)	(0.000)			(0.007)	(0.000)
Manitoba			0.484***	0.0035***			0.186***	0.002***
			(0.017)	(0.000)			(0.016)	(0.000)
Saskatchewan			0.465***	0.003***			-0.120***	-0.001***
			(0.020)	(0.000)			(0.019)	(0.000)
Alberta			-0.218***	-0.001***			-0.309***	-0.004***
			(0.013)	(0.000)			(0.011)	(0.000)
British Columbia			0.285***	0.002***			-0.170***	-0.002***
			(0.012)	(0.000)			(0.010)	(0.000)
Test immigrants without P.L.C.E=immigrants with P.L.C.E (pvalue)	0.000		0.000		0.06		0.01	
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations				57,049,066				
Number of individuals				6,971,740				

Notes: ***p<0.01, **p<0.05, *p<0.1; Standard errors in parentheses are clustered at the individual level. For the province of employment, the territories are excluded.

Table 2.6: GEE RESULTS OF THE LIKELIHOOD OF WORKING IN THE AGRICULTURE SECTOR AND THE FOOD PROCESSING SECTOR (ONLY IMMIGRANTS)

VARIABLES	(1) Primary agriculture Coefficients	(2) Primary agriculture Marginal effects	(3) Primary agriculture Coefficients	(4) Primary agriculture Marginal effects	(5) Food processing Coefficients	(6) Food processing Marginal Effects	(7) Food processing Coefficients	(8) Food processing Marginal Effects
Immigration status (ref: Immigrants without P.L.C.E***)***								
Immigrants with P.L.C.E	-0.130***	-0.0008***	0.049**	0.0001**	-0.018	-0.0004	0.005	0.0001
	(0.021)	(0.000)	(0.025)	(0.000)	(0.011)	(0.000)	(0.013)	(0.000)
Years since migration			-0.064***	-0.0002***			-0.037***	-0.0007***
			(0.001)	(0.000)			(0.001)	(0.000)
Male			-0.056***	-0.0001***			0.149***	0.003***
			(0.019)	(0.000)			(0.010)	(0.000)
Age			0.037***	0.0001***			0.024***	0.0004***
			(0.001)	(0.000)			(0.000)	(0.000)
Married			0.008	0.00003			0.049***	0.001***
			(0.016)	(0.000)			(0.009)	(0.000)
Knowledge of an official language at time of landing			-0.777***	-0.002***			-0.471***	-0.009***
			(0.022)	(0.000)			(0.012)	(0.000)
Provinces (ref: Quebec)								
Atlantic			0.975***	0.003***			0.088	0.0017
			(0.082)	(0.000)			(0.055)	(0.001)
Ontario			0.276***	0.0009***			0.107***	0.002***
			(0.042)	(0.000)			(0.019)	(0.000)
Manitoba			0.985***	0.003***			0.590***	0.011***
			(0.064)	(0.000)			(0.036)	(0.001)
Saskatchewan			0.606***	0.0019***			-0.177***	-0.003***
			(0.108)	(0.000)			(0.064)	(0.001)
Alberta			-0.026	-0.00009			0.260***	0.005***
			(0.060)	(0.000)			(0.025)	(0.000)
British Columbia			1.460***	0.0048***			0.082***	0.0016***
			(0.041)	(0.000)			(0.023)	(0.000)
Immigration class (ref: family class)								
Canadian experience class			-0.584***	-0.0019***			-0.719***	-0.014***
			(0.172)	(0.001)			(0.093)	(0.002)
Skilled workers			-0.935***	-0.003***			-0.249***	-0.004***
			(0.044)	(0.000)			(0.019)	(0.000)
Business class			-0.781***	-0.002***			-0.303***	-0.006***
			(0.094)	(0.000)			(0.056)	(0.001)

Provincial nominees			0.207***	0.0006***			0.600***	0.012***
			(0.053)	(0.000)			(0.030)	(0.001)
Refugee			-0.185***	-0.0006***			0.218***	0.004***
			(0.028)	(0.000)			(0.015)	(0.000)
Others			-0.920***	-0.003***			-0.296***	-0.005***
			(0.025)	(0.000)			(0.013)	(0.000)
Education (ref:high school dropout)								
High school			-0.918***	-0.003***			-0.242***	-0.004***
			(0.043)	(0.000)			(0.020)	(0.000)
Trade or college			-0.826***	-0.002***			-0.268***	-0.005***
			(0.032)	(0.000)			(0.016)	(0.000)
Bachelor			-1.164***	-0.003***			-0.559***	-0.011***
			(0.035)	(0.000)			(0.017)	(0.000)
Graduate studies			-1.391***	-0.004***			-0.967***	-0.019***
			(0.060)	(0.000)			(0.029)	(0.001)
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Last region of residence fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations				8,604,111				
Number of individuals				1,222,067				

Notes: ***p<0.01, **p<0.05, *p<0.1; Standard errors in parentheses are clustered at the individual level. For the province of employment, the territories are excluded.

Table 2.7: DISCRETE SURVIVAL ANALYSIS HAZARD RATIOS OF THE LIKELIHOOD OF QUITTING THE AGRICULTURE AND THE FOOD PROCESSING SECTOR CONDITIONAL ON ENTERING (FULL SAMPLE)

	Primary Agriculture				Food and Beverage manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	Cloglog Hazard ratio	Cloglog Hazard ratio	Cloglog Hazard ratio	Xtcloglog Hazard ratio	Cloglog Hazard ratio	Cloglog Hazard ratio	Cloglog Hazard ratio	Xtcloglog Hazard ratio
Immigration status (ref: Canadian-born)								
Immigrant	0.982*	1.208***	1.058***	1.230***	0.879***	0.983***	0.955***	0.819***
	(0.010)	(0.013)	(0.012)	(0.029)	(0.005)	(0.006)	(0.006)	(0.010)
Male		0.913***	1.142***	1.263***		1.010**	1.293***	1.719***
		(0.006)	(0.009)	(0.020)		(0.005)	(0.007)	(0.017)
Age		0.976***	0.978***	0.949***		0.970***	0.971***	0.935***
		(0.000)	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Married		0.576***	0.641***	0.381***		0.748***	0.839***	0.690***
		(0.005)	(0.006)	(0.007)		(0.004)	(0.005)	(0.008)
Provinces (ref: Quebec)								
Atlantic		1.154***	0.886***	0.811***		0.903***	0.522***	0.297***
		(0.014)	(0.012)	(0.020)		(0.006)	(0.005)	(0.005)
Ontario		0.979**	0.778***	0.602***		0.959***	1.319***	1.610***
		(0.009)	(0.008)	(0.011)		(0.005)	(0.009)	(0.020)
Manitoba		0.948**	0.961*	0.746***		0.822***	1.058***	0.947
		(0.020)	(0.021)	(0.034)		(0.013)	(0.018)	(0.033)
Saskatchewan		1.153***	1.370***	1.662***		1.226***	1.409***	2.149***
		(0.025)	(0.032)	(0.076)		(0.023)	(0.028)	(0.089)
Alberta		1.733***	1.740***	3.363***		1.682***	2.545***	7.864***
		(0.024)	(0.025)	(0.099)		(0.016)	(0.027)	(0.168)
British Columbia		1.091***	0.908***	0.811***		1.208***	1.088***	1.349***
		(0.015)	(0.013)	(0.023)		(0.012)	(0.012)	(0.028)
Low paying firm*(ref: medium paying firm)*			2.342***	5.024***			4.122***	14.646***
			(0.022)	(0.076)			(0.028)	(0.168)
High-paying firm*(ref: medium paying firm)*			0.794***	0.720***			0.375***	0.131***
			(0.009)	(0.011)			(0.003)	(0.002)
Ln (employment)			1.429***	1.916***			0.937***	0.800***
			(0.003)	(0.008)			(0.001)	(0.002)
Ln (spell)	1.537***	1.895***	2.102***	73.587***	1.400***	1.742***	2.151***	51.973***
	(0.006)	(0.009)	(0.012)	(2.779)	(0.004)	(0.005)	(0.009)	(1.128)
Likelihood-ratio test of rho=0				***				***
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	744,644	744,644	744,644	744,644	1,687,721	1,687,721	1,687,721	1,687,721
Number of individuals	267,715	267,715	267,715	267,715	532,855	532,855	532,855	532,855

Notes: ***p<0.01, **p<0.05, *p<0.1; Standard errors in parentheses are clustered at the individual level. For the province of employment, the territories are excluded.

Table 2.8: DISCRETE SURVIVAL ANALYSIS HAZARD RATIOS OF THE LIKELIHOOD OF QUITTING THE AGRICULTURE AND THE FOOD PROCESSING SECTOR CONDITIONAL ON ENTERING (FULL SAMPLE/IMMIGRANTS DIVIDED INTO TWO GROUPS)

	Primary Agriculture				Food and Beverage manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	Cloglog Hazard ratio	Cloglog Hazard ratio	Cloglog Hazard ratio	Xtcloglog Hazard ratio	Cloglog Hazard ratio	Cloglog Hazard ratio	Cloglog Hazard ratio	Xtcloglog Hazard ratio
Immigration status (ref: Canadian-born)								
Immigrants without P.L.C.E	1.012	1.251***	1.068***	1.344***	0.914***	1.005	0.979***	0.870***
	(0.012)	(0.016)	(0.015)	(0.037)	(0.005)	(0.006)	(0.007)	(0.012)
Immigrants with P.L.C.E	0.910***	1.115***	1.034*	0.996	0.789***	0.925***	0.893***	0.695***
	(0.016)	(0.021)	(0.020)	(0.041)	(0.007)	(0.009)	(0.010)	(0.014)
Male		0.915***	1.142***	1.269***		1.011**	1.294***	1.720***
		(0.006)	(0.009)	(0.020)		(0.005)	(0.007)	(0.017)
Age		0.976***	0.978***	0.949***		0.971***	0.971***	0.936***
		(0.000)	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Married		0.577***	0.642***	0.382***		0.748***	0.840***	0.691***
		(0.005)	(0.006)	(0.007)		(0.004)	(0.005)	(0.008)
Provinces (ref: Quebec)								
Atlantic		1.154***	0.886***	0.811***		0.903***	0.521***	0.296***
		(0.014)	(0.012)	(0.020)		(0.006)	(0.005)	(0.005)
Ontario		0.979**	0.778***	0.602***		0.957***	1.317***	1.605***
		(0.009)	(0.008)	(0.011)		(0.005)	(0.009)	(0.020)
Manitoba		0.950**	0.961*	0.749***		0.830***	1.065***	0.958
		(0.020)	(0.021)	(0.034)		(0.013)	(0.018)	(0.033)
Saskatchewan		1.154***	1.370***	1.670***		1.227***	1.409***	2.152***
		(0.025)	(0.032)	(0.076)		(0.023)	(0.028)	(0.089)
Alberta		1.738***	1.741***	3.393***		1.683***	2.545***	7.873***
		(0.024)	(0.025)	(0.101)		(0.016)	(0.027)	(0.168)
British Columbia		1.083***	0.906***	0.798***		1.204***	1.084***	1.338***
		(0.015)	(0.013)	(0.022)		(0.012)	(0.012)	(0.028)
Low paying firm*(ref: medium paying firm)*			2.342***	5.042***			4.121***	14.63***
			(0.022)	(0.077)			(0.028)	(0.168)
High-paying firm*(ref: medium paying firm)*			0.795***	0.720***			0.374***	0.131***
			(0.009)	(0.011)			(0.003)	(0.002)
Ln (employment)			1.429***	1.919***			0.937***	0.800***
			(0.003)	(0.008)			(0.001)	(0.002)
Ln (spell)	1.538***	1.895***	2.101***	75.18***	1.400***	1.741***	2.151***	51.97***
	(0.006)	(0.009)	(0.012)	(2.949)	(0.004)	(0.005)	(0.009)	(1.128)
Likelihood-ratio test of rho=0				***				***
test immigrants= immigrantsexp (pvalue)	0.000	0.000	0.143	0.000	0.000	0.000	0.000	0.000
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	744,644	744,644	744,644	744,644	1,687,721	1,687,721	1,687,721	1,687,721
Number of individuals	267,715	267,715	267,715	267,715	532,855	532,855	532,855	532,855

Notes: ***p<0.01, **p<0.05, *p<0.1; Standard errors in parentheses are clustered at the individual level. For the province of employment, the territories are excluded.

Table 2.9: DISCRETE SURVIVAL ANALYSIS HAZARD RATIOS OF THE LIKELIHOOD OF QUITTING THE FOOD PROCESSING SECTOR AFTER ENTERING (ONLY IMMIGRANTS)

	Primary Agriculture				Food and Beverage manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model	Cloglog Hazard ratio	Cloglog Hazard ratio	Cloglog Hazard ratio	Xtcloglog Hazard ratio	Cloglog Hazard ratio	Cloglog Hazard ratio	Cloglog Hazard ratio	Xtcloglog Hazard ratio
Immigration status (ref: Immigrants without P.L.C.E***)***								
Immigrants with P.L.C.E	0.890***	0.759***	0.842***	0.622***	0.867***	0.915***	0.903***	0.841***
	(0.018)	(0.021)	(0.025)	(0.035)	(0.009)	(0.011)	(0.013)	(0.019)
Years since migration		0.985***	0.987***	0.976***		0.989***	0.999	1.004***
		(0.002)	(0.002)	(0.003)		(0.001)	(0.001)	(0.001)
Male		1.035*	1.230***	1.557***		0.995	1.283***	1.561***
		(0.021)	(0.028)	(0.068)		(0.009)	(0.015)	(0.028)
Age		0.979***	0.980***	0.956***		0.974***	0.975***	0.952***
		(0.001)	(0.001)	(0.002)		(0.001)	(0.001)	(0.001)
Married		0.892***	0.920***	0.918**		0.790***	0.886***	0.807***
		(0.018)	(0.021)	(0.037)		(0.008)	(0.010)	(0.014)
Provinces (ref: Quebec)								
Atlantic		0.787***	0.681***	0.548***		1.058	1.076	1.080
		(0.061)	(0.054)	(0.080)		(0.045)	(0.060)	(0.084)
Ontario		0.813***	0.789***	0.591***		0.807***	1.482***	1.573***
		(0.028)	(0.029)	(0.041)		(0.010)	(0.025)	(0.038)
Manitoba		0.784***	0.964	0.766**		0.561***	1.031	0.834***
		(0.049)	(0.071)	(0.099)		(0.020)	(0.039)	(0.051)
Saskatchewan		1.269***	1.420***	2.065***		1.175***	1.786***	2.645***
		(0.108)	(0.152)	(0.348)		(0.059)	(0.097)	(0.243)
Alberta		1.963***	2.267***	5.418***		1.155***	2.432***	4.198***
		(0.080)	(0.111)	(0.511)		(0.021)	(0.054)	(0.150)
British Columbia		0.888***	0.799***	0.609***		0.967*	1.182***	1.283***
		(0.036)	(0.034)	(0.049)		(0.018)	(0.028)	(0.044)
Immigration class (ref: family class)								
Canadian experience class		0.541***	0.886	0.488**		0.840**	0.825*	0.943
		(0.087)	(0.139)	(0.176)		(0.062)	(0.093)	(0.138)
Skilled workers		1.010	1.077	1.024		1.068***	1.137***	1.353***
		(0.044)	(0.053)	(0.095)		(0.018)	(0.023)	(0.043)
Business class		0.810*	1.003	0.695		1.450***	1.015	1.145
		(0.101)	(0.121)	(0.165)		(0.097)	(0.090)	(0.137)
Provincial nominees		0.684***	0.949	0.546***		0.564***	0.711***	0.504***
		(0.038)	(0.054)	(0.065)		(0.020)	(0.027)	(0.032)
Refugee		1.163***	1.140***	1.240***		1.066***	1.051***	1.079***
		(0.032)	(0.034)	(0.073)		(0.015)	(0.017)	(0.028)
Others		1.089***	1.101***	1.128**		1.186***	1.113***	1.274***
		(0.030)	(0.033)	(0.065)		(0.014)	(0.016)	(0.029)
Education (ref:high school dropout)								
High school		1.126***	1.154***	1.439***		1.063***	1.139***	1.253***

		(0.046)	(0.052)	(0.126)		(0.019)	(0.023)	(0.041)
Trade or college		1.143***	1.150***	1.419***		1.065***	1.076***	1.198***
		(0.037)	(0.042)	(0.099)		(0.016)	(0.019)	(0.032)
Bachelor		1.151***	1.222***	1.530***		1.131***	1.305***	1.664***
		(0.040)	(0.049)	(0.113)		(0.017)	(0.023)	(0.047)
Graduate studies		1.380***	1.246***	1.878***		1.119***	1.351***	1.639***
		(0.080)	(0.082)	(0.244)		(0.028)	(0.038)	(0.077)
Knowledge of an official language at time of landing		1.078***	1.038	1.128**		0.962***	1.049***	1.122***
		(0.026)	(0.028)	(0.058)		(0.010)	(0.013)	(0.023)
Low paying firms*(ref: medium paying firm)*			4.416***	12.686***			8.046***	27.293***
			(0.124)	(0.650)			(0.108)	(0.668)
High paying firms*(ref: medium paying firm)*			0.648***	0.484***			0.367***	0.196***
			(0.020)	(0.021)			(0.007)	(0.005)
Ln(spell)			1.290***	1.467***			1.015***	0.916***
			(0.008)	(0.015)			(0.003)	(0.003)
Likelihood-ratio test of rho=0: chibar2(01) =				***				***
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Last region of residence fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	94,974	94,974	94,974	94,974	422,701	422,701	422,701	422,701
Number of individuals	34,168	34,168	34,168	34,168	131,689	131,689	131,689	131,689

Notes: ***p<0.01, **p<0.05, *p<0.1; Standard errors in parentheses are clustered at the individual level. For the province of employment, the territories are excluded.

Chapter Three

Immigrant Assimilation in a Multicultural and Multilingual context

3.1 INTRODUCTION

Immigrants' assimilation to their host country is an important element determining the success of their economic and social development. Proficiency in languages helps to facilitate employment, trade transactions and enhances the socialization processes between people from different cultures. This is particularly important for countries with high shares of immigrants like Canada. However, the assimilation process may also come at a cost, such as an individual's loss of cultural identity. Whether immigrants decide to assimilate or hold on to their language and culture is an interesting and fundamental question in multicultural societies. This decision might also become more complex for immigrants settling in multilingual countries in which more than one official language is spoken, such as Canada, Luxembourg, Switzerland and Belgium. In these cases, immigrants are likely to end up assimilating to the dominant language of the region where they settle, but many factors may impact on the degree and choice of language assimilation.

The metropolitan area of Montreal is a particularly interesting case for investigating the language assimilation of immigrants. This metropolitan, multilingual city is characterized by the presence of sizeable French-and-English-speaking communities, as well as of a large number of immigrants from a wide variety of linguistic backgrounds. Those elements of linguistic diversity

interact in the context of the French being the official language of the city and the province, and English being the dominant minority language but also the major world *lingua franca*. Immigrants are exposed to these two main languages and have to decide whether to assimilate to French, English, both, or neither. Language acquisition decisions are expected to have an important impact on their labour market outcomes. Assimilating to one of the official languages should improve language proficiency and increase immigrants' potential productivity in the labour market. Language assimilation is also expected to increase networking opportunities with the native population, which could bring benefits from informal contact and improve the chances of accessing job opportunities or promotions.

Using microdata from the master files of the 2001, 2006 and 2016 Canadian censuses and from the 2011 National Household Survey, I investigate the role of the linguistic distances between an immigrant's mother tongue and English as well as French on the choice of linguistic assimilation in the Montreal metropolitan area. As a proxy for linguistic distance, I use an index developed by the Max Planck Institute for Evolutionary Anthropology, which employs an algorithm based on the Levenshtein distance.[19] I find that linguistic distances between immigrants' mother tongues and English and French have important impacts on the relative intensities of utilisation and the choices of assimilation to the two Canadian official languages.

This paper is structured as follows. The second section consists of a review of some selected studies of the literature on the language assimilation of immigrants. The third section presents some general background and a description of the dataset, explains the sample restrictions that are imposed, and discusses some descriptive statistics. The fourth section presents the econometric model followed by the clarification of the specifications. The fifth

[19] Section 3.4 and Appendix 1.1 of chapter 1 describe the details of how the index is calculated.

section consists of a discussion of the results of the empirical work, and the last section concludes the paper.

3.2 LITERATURE REVIEW

Over the past three decades, a number of studies have analyzed the language assimilation of minority language groups. Language assimilation is a key factor of immigrants' integration and is expected to improve their degrees of economic and social participation. Therefore, it is important to understand what the factors are that determine assimilation of immigrants to the host country's official languages. The literature has mainly focused on language assimilation of minority language groups to one official language such as English in the U.S. However, there is also a line of research that investigates language acquisition decisions in multilingual contexts where more than one official language are spoken, such as the case of Canada. This review of the literature will present pertinent papers of those lines of research chronologically, allowing us to observe the changes in language assimilation research over time.

Veltman (1983) examines language shifts for minority language groups in the United States. The data were drawn from the 1976 Survey of Income and Education, which includes detailed information on the mother tongue, current usual language, and current second language of more than 150,000 households in the United States. The analysis is based on descriptive statistics, and the main results suggest that minority language groups show high levels of Anglicization, with the exception of the Navajo and Spanish-language groups, who display some resistance in certain regions. The author also finds that foreign-born individuals with a minority mother tongue exhibit a rapid shift to English - even those who arrived recently in the United States. He concludes that immigrant minority language groups have a strong willingness to not only learn English but also to make English their principal language of use.

Grenier and Vaillancourt (1983) investigate the factors that determine if an individual knows a second language. They analyzed and compared the cases of knowing English for Francophones in Quebec and for Hispanics in the United States. For the sample of American Hispanics, the data were also drawn from the 1976 *Survey of Income and Education*, and the data for the Francophones in Quebec were drawn from a survey of the *Conseil de la Langue Française* in 1979. The estimation method employed is a probit model, and the dependent variable is knowledge of English. The key independent variables were grouped into three main categories: those relating to language skills, non-linguistic human capital, and to the environment. A key finding of this research suggests that the home environment is an important factor contributing to linguistic assimilation. Individuals exposed to more English at home with their spouse, or through their neighbors, are more likely to learn and use this second language.

Grenier (1984) further studies the factors that determine assimilation to English among Hispanics in the United States. The theoretical framework used is based on the concept of human capital. The decision to learn and assimilate to a language can be seen as an investment made by the individuals. The individuals invest in themselves in order to maximize the expected net return (benefit minus costs) of that investment. The author explains that the benefits and costs are driven by three major elements. The first is economic incentive of shifting to another language such as the possibility to earn a higher income. The second element is the individual's environment. Indeed, before an individual assimilates to a language, he/she needs to learn it, and the environment where the individual lives determine in large part the opportunity cost of learning that language. Finally, the last important element is the psychic cost of language shifts. For instance, language shift may be resisted to preserve the cultural identity.

Those benefits and costs cannot be measured directly but can be related to observed

individual characteristics. The dataset used by the author to explore the impact of those characteristics on assimilation is the 1976 *Survey of Income and Education*. The empirical method was a probit model with the dependent variable being whether an individual has retained Spanish or has shifted to English as the language most often spoken. The independent variables are a set of individuals' characteristics, which determine the expected net returns of language shift. The main results suggest that Hispanic Americans are assimilating to English at a relatively fast pace, and that the factors with an important impact on assimilation are the age at migration, years since migration, education level, and the choice of a non-Hispanic marriage partner.

Veltman (1988) builds on his earlier study to explore the language shifts from Spanish to English among Hispanic immigrants in the United States. The source of data is again the 1976 *Survey of Income and Education*. In comparison to Veltman (1983), this study restricts the sample to persons with Spanish mother tongue who were not born in the U.S. To measure the degree of assimilation, the author constructed four main variables based on how often an individual speaks English: Spanish monolingual, Spanish bilingual, English-dominant bilingual, and English mono-lingual. The main results suggest that the age at time of arrival and the length of residence in the U.S. are the two main contributors to language shifts. The results show that the shift to English is more important the younger the immigrant is at time of arrival, and the use of English declines as age at time of arrival increases. One drawback of this study is that it does not account for spousal characteristics, which is expected to be an important factor contributing to immigrants' language assimilation.

Chiswick and Miller (1994, 2001) analyze the language choices of immigrants in Canada, where both English and French are the official languages. They develop a framework that explains language assimilation of immigrants through three main channels: exposure, efficiency

and economic factors. They explain that language skills improve with measures of exposure, such as the duration of time spent in the host country, the prior exposure to the language in the country of origin, and the degree of exposure at home and in the neighborhood where the individual lives. The second channel that also increases language skills is through the efficiency of language acquisition. For instance, the younger the age at migration, and the higher the level of education, the greater this efficiency is anticipated to be. Finally, the last channel explaining the degree of language assimilation involves economic variables. Individuals who expect to reap higher economic returns from a second language are more likely to learn and use this language.

To test this framework, Chiswick and Miller (2001) use the 1991 Census of Canada (PUMF) file. The empirical methodology employed is a multinomial logit model with a tri-chotomous dependent language variable. This variable takes three values corresponding to the following categories: (1) inability to conduct a conversation in either of the official languages, (2) able to conduct a conversation in one of the official languages but usually speak a nonofficial language at home, (3) able to speak one of the official languages and usually speak one of the official languages at home. The main results of their empirical analysis suggest that an older age at migration decreases the use of an official language, and that the duration of time spent in Canada is positively associated with the use of English or French. They also found that higher level of education and greater distances between the country of origin and Canada are associated with greater use of French and English.

Lazear (2007) investigates why immigrants from Mexico assimilate more slowly than other immigrants in the U.S. He develops a theoretical model that suggests that incentives to assimilate are reduced when the proportion of individuals sharing the same culture increases. He further explains that this might be the main cause of the slower assimilation of Mexicans into

American society, as they tend to live in communities with large numbers of other Mexicans. To test this hypothesis empirically, the author uses data from the 5 percent sample of the 2000 U.S. Public Use Census. The dependent variable is a dummy equal to 1 if the respondent claims fluency in English and 0 otherwise. The key independent variable is the proportion of other residents in the area (based on counties) who are born in the respondent's native country. Control variables include the number of years in the United States, the highest grade of schooling completed, a dummy if born in Mexico, a dummy if of Hispanic origin, and interaction variables between years in the U.S. and the proportion variable, as well as between the education and the proportion variable. The analysis is repeated separately for Mexican immigrants and non-Mexican immigrants.

The key results are consistent with the theoretical model and confirm that there is a negative relationship between immigrant assimilation and the proportion of immigrants sharing the same culture. The author asserts that this is explained by two mechanisms working together. The first mechanism assumes that the locational choice is exogenous and suggests that immigrants who are more integrated in the host country will learn English more rapidly. The second mechanism assumes endogenous locational choice and suggests that immigrants who have lower proficiency in English will be more likely to locate in enclaves where more individuals speak their languages.

Kim and Min (2010) examine the effects of marital patterns on the likelihood of mother-tongue use at home. They analyze this relationship for native-born Asian-Americans who are 18 years of age or older and were born in the United States after 1964. The study uses the 2005-2007 *American Community Survey* 3-Year Public Use Microdata, and the estimation method employed is a logistic regression. Their dependent variable is the mother-tongue use at home,

which takes a value of 1 if the respondent speaks a language other than English and 0 otherwise. The key independent variables are marital status, indicators for the race of the spouse, and the generational status of the spouse. After controlling for individual, family and geographical characteristics, the authors find that those with other Asian ethnic partners have a greater likelihood of mother-tongue use at home than those with white partners, suggesting that Asian-Asian intra-marriage provides an advantage in maintaining bilingualism and slowing down the process of linguistic assimilation. They also found that those with first generation co-ethnic spouses have a higher likelihood of using their mother tongue than those with the U.S.-born co-ethnic spouses.

Van Tubergen and Wierenga (2011) study language acquisition outcomes of male immigrants in the multilingual context of Belgium, where French and Dutch are both official languages. They employ data from a national survey for the years 1994-1996, which collected information on Turks and Moroccans, which are the two largest immigrant groups in Belgium. They follow the theoretical framework of Chiswick and Miller (1994, 2001), which explains language acquisition outcomes of immigrants through the three mechanisms of exposure, incentives and efficiency. Their empirical model uses binary and multinomial logit regression specifications with language proficiency as dependent variable. Their independent variables include age at migration, length of stay, education in origin country, education in receiving countries, migration motives, knowledge of the first language, membership in a voluntary organization, ethnicity of respondent co-ethnic partner, ethnic minority concentration, settlement intentions, and the taking of language courses. Their key results suggest that immigrants assimilate more to French than to Dutch. The results also show that this effect is more important for Moroccans; the author explains that this likely due to a greater pre-migration exposure to

French. They explain that French has greater international recognition, and this makes it more attractive than the Dutch language. This is an interesting result that resembles the case of Quebec, except that English is the more powerful international language. They also find that immigrants with higher levels of education, those who emigrated at a younger age, those who have been in Belgium for a longer time, and those who live in a region with a smaller proportion of co-ethnics have better proficiency in both official languages.

Ortega and Verdugo (2015) study the language assimilation of immigrants in Canada by trying to understand if language assimilation into the secondary (French) and the dominant (English) language differ. With the use the Public Use Microdata Files (PUMF) of the 2001 and 2006 Canadian Censuses, they explore whether the presence of dominant-language speakers in a location where the secondary language is the majority language will slowdown immigrant assimilation into the secondary language. They also further explore the mechanism behind the negative relationship between language assimilation and the presence of co-mother tongue speakers by investigating whether this relationship is mainly explained by a lower incentive of learning the location-majority language in the presence of more co-mother tongue speakers (the learning mechanism), or whether this is simply the result of immigrants with less knowledge of the majority language choosing to settle in locations with more co-mother speakers (the sorting mechanism).

Their empirical methodology involves a probit model with the knowledge of the dominant language by minority members as the dependent variable. Minority members refer to allophones and Anglophones for cities where French is the majority language, and allophones and Francophones for cities where English is the majority language. Their independent variables include a set of individual characteristics (age, sex, employment status, immigration status, age

at migration, marital status, presence of children, education, and a dummy indicator for the mother tongue), and city-specific variables (the size of the official minority and the proportion of individuals having the same mother tongue, and interaction terms between the language composition measure and years since migration). Similar to the existing literature for the U.S. (Lazear 1999, 2007), they find a negative relationship between knowledge of the city-majority languages by allophones and the size of the own-mother tongue group. However, their results suggest that the mechanism behind this finding differ for French and English-majority cities. They explain that the evidence for the learning mechanism is limited for English-majority cities, but very strong for the case of the French-majority cities. They also find that the presence of English mother-tongue speakers in French majority cities decreases the knowledge of French for allophones, but in contrast, the presence of Francophones does not alter the knowledge of English for allophones in English-majority cities.

More recently, Lee (2018) examines whether official language (s) use at home in Canada is associated with spousal characteristics and how this association varies by generation status. The study uses the 2002 *Ethnic Diversity Survey* data and is limited to individuals who are married or in common-law relationships. The study compares the results of the first generation, second generation, and third and higher generations of immigrants with the use of a multivariate model. The dependent variable is a dichotomous measure of whether individuals speak one or two official languages only at home versus an unofficial language, either alone or with an official language(s). The key independent variables are spousal characteristics, defined as the spouse's ethnicity, nativity, first language and education. Other control variables include individuals' characteristics and parental characteristics. The results suggest that the three-generation model of linguistic assimilation that anticipates a shift from the mother tongue to an official language in

three generations is significantly impacted by marital patterns. The author finds that the shift to an official language is delayed for individuals in foreign-born and cross-generation marriages and accelerated for those whose spouses are more educated or have the same first language.

To summarize the above literature, major variables that have been used to explain language assimilation can be well summarized with the Chiswick and Miller (2001) theoretical framework. Language acquisitions are determined by variables affecting the exposure (e.g. years since migrations, spouse characteristics, location characteristics), the efficiency (e.g. education, age, gender), and the economic incentives to learn and use the host country language.

This paper aims to extend the literature on language assimilation of immigrants by capturing and incorporating group-specific learning costs of assimilation; specifically, a detailed exogenous measure of linguistic distance between languages developed by the *Max Planck Institute for Evolutionary Anthropology*, which is derived from an algorithm based on the Levenshtein distance. The linguistic distance can have an important impact on the efficiency of an immigrant learning and assimilating to an official language. I expect that an immigrant with a closer linguistic distance will be more efficient in learning and assimilating to an official language. The linguistic distance variable used in this paper is explained in detail in Appendix 1.1 of Chapter 1. In addition, this research also develops a multi-dimensional index of language assimilation that provides measures of the relative assimilation to English and French simultaneously in the multi-lingual and multi-cultural context of Montreal. This measure of the relative assimilation to English compared to French will allow us to identify the factors that have an important impact on immigrants assimilating to the main official language (French in this case) when compared to the minority but important "lingua franca" language (English in this case).

3.3 DATA AND SUMMARY STATISTICS

3.3.1 DATA

The datasets used for this paper are the micro-data master files of the 2001, 2006 and 2016 Canadian Censuses and of the 2011 *National Household Survey* (NHS) from Statistics Canada. Those databases provide a rich source of information on labour market characteristics, immigrant status, and language features. They also contain detailed information about the economic, social and demographic characteristics of the Canadian population. Another important advantage is the large sample size covering 20 percent of the Canadian population. I restrict my analysis to allophone immigrants in the greater Montreal area who became landed immigrants after reaching the age of 18. I only consider immigrants who arrived after the age of 18 because the language abilities of immigrants who arrived as adults differ from immigrants who arrived as children or teens. To compute the values of the variable for measuring the linguistic distance between languages, I use the database of the *Automated Similarity Judgment Program* (ASJP) developed by linguists at the Max Planck Institute for Evolutionary Anthropology.

From these sources of data, three key indicators are used in the analysis: *linguistic distance*, *relative intensity of using an official language at home* (regardless of whether it is English or French), *and relative intensity of using English and French at home* if an official language is used. The second indicator on relative intensity of using an official language at home will allow me to identify what are the key drivers of language assimilation to a the host country's official language. It is in part similar to what was done in the literature by Veltman (1983), Grenier (1984), and Chiswick and Miller (1994, 2001). The third indicator of the relative intensity of using English and French at home is used to account for the particular situation of Montreal, where there are two potential languages of attraction.

I link the linguistic distance between both Canada's official languages (French and English) and all other foreign mother tongues reported by the respondents in the Census (involving about 150 different languages). The linguistic distance from French and English have values ranging from 0 (for the same language) to approximately 104 for the largest distance between two languages.

The relative intensity of using English and French at home is used to proxy for the degree of assimilation. This measure is derived from a main question and sub-question asked in the censuses and in the NHS regarding the languages used at home, which are worded in the following way: 1) What language does this person speak most often at home? 2) Does this person speak any other languages on a regular basis at home? To those questions, the respondent can answer French, English, both, or another language, and they can also answer "none" to the second question. With the various combinations of answers to the above two questions, the following index of the relative intensity of using an official language at home is developed (regardless if it is English or French) for immigrant allophones in the Montreal metropolitan area, which we will refer to as ***OFFhome***. Specifically, the values refer to typical patterns and are assigned as follows:

- 0 if a non-official language is used most often, and an official language is never used
- 25 if a non-official language is used most often, but an official language is also used on a regular basis
- 50 if an official language and a non-official language are equally used most often.
- 75 if an official language is used most often, but a non-official language is also used on a regular basis

- 100 if an official language is used most often, and no non-official language is used.

If an official language is used at home (***OFFhome*** is not equal to 0), I extend my analysis to explore whether that official language tends to be French instead of English. I will refer to this index as ***Fhome***:

- 0 if English is used most often and French is never used / or English is used on a regular basis and a non-official language is used most often, and French is never used.
- 25 if English is used most often, but French is used on a regular basis, and any other combinations of non-official languages are used.
- 50 if both official languages are equally used most often, and any other combinations of non-official language are used / or both official languages are used regularly, and a non-official language is used most often.
- 75 if French is used most often, but English is also used on a regular basis, and any other combinations of non-official language are used.100 if French is used most often and English is never used on a regular basis, and any other combinations of non-official language are used / or if French is used on a regular basis and a non-official language is used most often, and English is never used.

Both indexes take values between zero and one hundred. One possible way to interpret the first index (***OFFhome***) is that it approximates the proportion of time an immigrant uses an official language at home as opposed to a non-official language. For those who do use an official language at home, the second index (***Fhome***) can be interpreted as the proportion of time an

immigrant uses French at home relative to English in the context of the dual linguistic nature of the Montreal metropolitan area.

3.3.2 SUMMARY STATISTICS

Table 3.1 presents the means of some of the key variables that are used in this study. A primary variable of interest in this study is the linguistic distances for immigrants from English and French. As expected, we see that on average, immigrants in the greater Montreal area have a mother tongue that is more distant from English (95.26) than from French (88.69). The summary statistics also indicate that the average age is 51 years, and that immigrants have been in Canada for an average of 20 years. For education, I divided my sample into five categories based on the highest diploma and degree attained. We observe that the proportion of immigrants without a high school degree is larger (28.3 percent) than those with a high school degree (17.1 percent). We also see that approximately 25.8 percent obtained some post-secondary education compared to 15.6 percent with a university bachelor's degree and 13 percent with a graduate degree. We also observe that the levels of education differ between men and women. There is a higher proportion of immigrant women without a high school degree (31.4 percent) than immigrant men (24.9%), and a lower proportion of immigrant women with a graduate degree (10.6 percent) than immigrant men (15.5 percent).

Table 3.1 also presents descriptive statistics on spousal characteristics. These variables have been defined by dividing our sample into 4 distinct categories. The first is an indicator variable if an immigrant has a Francophone partner. The second is an indicator variable if the individual has an Anglophone partner. The third indicator is for those with an allophone partner. These variables are derived from the master file with the use of the family identifiers and other socio-demographic characteristics. The reference group for this variable is the group of single

immigrants. The results suggest that allophone immigrants in our sample are more likely to have a Francophone partner (7.7 percent) than an Anglophone partner (3.1 percent) and the majority have an allophone partner, which is also true for both genders separately. We also observe that there is a higher proportion of men with an allophone partner (67 percent) than women (55 percent).

The other key linguistic variables are the official language intensity index and the French intensity index that where defined above. Tables 3.2 and 3.3 show the proportion of immigrants in each level of intensity for those indexes. From Table 3.2, we observe that non-official languages are widely spoken in Montreal. A plurality of immigrants uses only a non-official language at home (39.7 percent), followed by those who use a non-official language first and an official language second (24.4 percent), while 10 percent use a non-official language and an official language equally. For those who use an official language at a higher intensity than a non-official language, 15.5 percent use an official language first and a non-official language second, and only 10.2 percent use only an official language at home. The results differ slightly in both extremities when analyzed separately by gender, with the proportion that uses a non-official language only being larger for women (41.7 percent) than for men (37.5 percent), and the proportion who use an official language only being lower for women (8.8 percent) than for men (11.8 percent).

Table 3.3 shows the intensity of using French as opposed to English for those who do use an official language at home. We observe an interesting distribution that confirms the importance of utilization of both official languages at home in the Montreal metropolitan area. In fact, on average, 32.9 percent of the sample use only English most of the time with no French, 2.3 percent use English first and French second, 6 percent use English and French equally, 2.9

percent use French first and English second, and finally 55.6 percent use French most of the time with no English. This distribution is approximately similar across genders.

3.4 EMPIRICAL METHODOLOGY

I will base my empirical methodology on the conceptual framework of Chiswick and Miller (1994, 2001), which explains language assimilation based on measures of exposure, efficiency, and economic incentive. The linguistic distance is one of the main drivers of the efficiency in learning and assimilating to a second language. The closer an immigrant's mother tongue is to one of the official languages, the more efficient he/she will be in learning and assimilating to this language. Therefore, the first econometric model consists of regression equations which estimate the effect of linguistic distance and other variables on the relative intensity of assimilation to an official language at home:

$$\boldsymbol{OFFhome_{it}} = \beta_0 + \beta_1 LD1_i + X_{it,}\delta + S_{it,}\alpha + \lambda_t + [Z_{it,}\theta + \tau_{ic}] + \epsilon_{it,} \quad (1)$$

In equation (1), the dependent variable, $\boldsymbol{OFFhome_i}$, is the measure of the degree of assimilation to an official language (regardless of whether it is English or French), as opposed to a non-official language for individuals (i). As explained above, it increases (decreases) with the use of an official language (non-official language) at home. The key independent variable for this model is $LD1$, the linguistic distance, which is defined in three ways: the average linguistic distance between both official languages, a closer-to-French-than-English indicator, and the minimum distance between both official languages. The intuition for the average linguistic distance between the immigrant's mother tongue and both official languages is that the further the immigrant's mother tongue is from the official languages, the higher the cost of assimilation

will be. The motivation for the closer-to-French indicator and for the minimum between the distance from French and the distance from English indicator is the assertion that immigrants will tend to choose the official language which is closer to their mother tongue. The closer-to-French indicator is a dummy variable that is equal to 0 if the distance between an immigrant's mother tongue is closer to English than to French, and equal to 1 if the distance is closer to French than to English.

The matrix X_{it} contains the individual socioeconomic variables, which include age, age squared, employment status, gender, education, years since migration, years-since-migration squared, and an indicator variable for the presence of children aged 0-5 or 6-14. The rationale behind those controls is that they are also variables that determine the amount of exposure, efficiency and economic incentive regarding the official languages. For instance, the years since migration indicators measure the duration of time immigrants have been exposed to the language. The presence of children is also related to the exposure to the official language at home. Immigrants with children will be more likely to use an official language in speaking with them, especially for those of school age.

The language assimilation process can also differ by gender. As we observed in the summary statistics, immigrant men are more likely to be employed and educated than immigrant women, and will therefore spend more time interacting with the natives and have more exposure to the official languages, which will again lower the cost of assimilation. Education and age also can increase language assimilation through efficiency. The efficiency of language acquisition is expected to be better at a younger age as well as for more educated individuals. Employment status could also be related to economic incentive; individuals who expect a higher return of using a language in the workplace are more likely to assimilate to this language.

I also add a matrix S_{it} to account for spousal characteristics, which include an indicator variable if the spouse's mother tongue is French, an indicator variable if the spouse's mother tongue is English, and an indicator variable if the spouse's mother tongue is a non-official language (for which the reference category is immigrants who are single). The rationale for these variables is that immigrants who are married with a Francophone or an Anglophone will have more exposure to one of the official languages. In addition, as explained by Grenier (1984), the psychic cost of loss of cultural identity by assimilating to another language is also expected to be lower if an immigrant is attached to a person from another culture.

The vector λ_t includes year-specific dummies to control for aggregate trends. An augmented specification includes the matrix $Z_{i,t}$ for location and city-specific controls that may have an impact on the degree of assimilation through higher or lower exposure. It contains a variable for the proportion of individuals having the same mother tongue as the respondent at the level of Montreal, a set of dummies to control for the location of residence within the Montreal region (residing in the west, center, east and other subdivisions), and an indicator for immigrants whose previous residence was outside Quebec either one or five years earlier. Finally, I add a specification that includes detailed country-of-birth fixed effects (τ_{ic}) to capture the degree of exposure to English or French prior to immigration and other influences. I cluster standard errors at the country-of-birth level to account for correlation between linguistic distances errors terms within each country of birth.

My second regression model is for individuals who use an official language at home (***Offhome*** is not equal to 0). I extend the analysis to explore whether that official language tends to be French instead of English (***Fhome***) and how their respective linguistic distances from French and English impact their assimilation choice:

$$\boldsymbol{Fhome_{it}} = \rho_0 + \rho_1 LD2_i + X_{it}\gamma + S_{it}\sigma + \lambda_t + [Z_{it}\eta + \tau_{ic}] + \mu_{it} \quad (2)$$

The key independent variable is a vector $LD2_i$, the linguistic distance, which is initially defined by the linguistic distance from French and the linguistic distance from English separately; in different regressions, I also employ the "Closer to French" indicator used above. The sign of the estimated coefficient of the linguistic distance from French is expected to be negative, since the further the language is from French, the less likely the subject it is to assimilate to it. The sign of the estimated coefficient of the linguistic distance from English is expected to be positive, since the further the language is from English, the more likely the subject is to assimilate to French. All other covariates are similar to the ones we used in equation (1).

3.5 THE RESULTS

3.5.1 IMPACT OF LINGUISTIC DISTANCE ON THE OFFICIAL LANGUAGES USED AT HOME

The OLS results for equation (1) modelling the impacts of the different indicators of linguistic distance, and of the partner's language on the degree of assimilation to an official language (regardless of whether it is English or French), as opposed to a non-official language, are presented in the table 3.4, 3.5, and 3.6. Column (1) of each table includes controls for the Census year and for socioeconomic variables, which include age, age-squared, gender, education, years-since-migration, years-since-migration squared, presence of children and employment status. Column (2) adds spousal characteristics, which include indicator variables if the spouse of the respondent is Francophone, Anglophone or Allophone. Column (3) is the richest specification, with the inclusion of location and city-specific controls and detailed place-of-birth fixed effects. I also repeated the richest specification for men and women separately in

columns (4) and (5).

The results of all of the specifications for the average linguistic distance from both official languages involve negative and highly statistically significant estimated coefficients, confirming our expectation that the relative intensity of using an official language relative to a non-official language at home decreases with the linguistic distance. More precisely, in the richest specification (column 3), the estimated coefficient of the average linguistic distance (Table 3.4) suggests that a unit increase in the average linguistic distance from an official language will decrease the ***OFFhome*** index by 0.41 units. To illustrate the economic magnitude of this effect, we can take as an example two immigrants, one with Vietnamese as a mother tongue, and another with Haitian Creole as a mother tongue. The average linguistic distance from French and English for the Vietnamese immigrant is approximately equal to 103, and the average distance of the Haitian Creole immigrant is approximately equal to 71, a difference of 32. The estimate in column (3) implies that the intensity of using an official language at home will be lower by 13 percentage points for the immigrants with a Vietnamese mother tongue than for the immigrants with a Haitian Creole mother tongue.

We also observe that spousal characteristics have a significant impact on the assimilation of immigrants, with the coefficients being consistent across all three tables. In my preferred specification, we see that allophone immigrants with a Francophone partner have approximately a 19 percentage point higher intensity of using an official language at home than the reference group (single allophone immigrants). Those who have an Anglophone partner spend approximately 16 percentage points more of their time using an official language, and those with an allophone partner spend approximately 12 percentage points less of their time using an official language than other single allophone immigrants. When we repeat the analysis separately

by gender, we observe that the magnitude of this effect is larger for men than for women. It suggests that men are more likely to assimilate to their partner's mother tongue than women. This could be because women are more attached to their mother tongue and might use it at a higher intensity than men with their kids or other family members.

I repeated the analysis with alternative indicators of linguistic distance. Table 3.5 shows the results for the measure consisting of the minimum between the distance from French and the distance from English, and Table 3.6 shows the results when I include a dummy variable that is equal to 1 if the distance is closer to French than to English, and 0 otherwise. As mentioned earlier, the motivation behind these specifications is the assertion that immigrants will tend to choose the official language that is closer to their mother tongue. The results for the minimum linguistic distance variables are all economically and statistically significant and follow the same empirical pattern as the results of the prior model with average distance for all specifications, confirming the conjecture that immigrants with a more distant mother tongue have a lower probability of assimilating to an official language.

The negative coefficient estimates in table 3.6 for the variable "Closer to French" indicate that immigrants whose mother tongues are closer to French are less likely to use an official language at home than those whose mother tongues are closer to English. This result is quite interesting and is consistent with the idea that English is a *lingua franca* and is more attractive than French. Immigrants who have a mother tongue closer to English find greater benefits in assimilating to this international language. In comparison, French might seem less attractive and less useful internationally than English. Again, the effect decreases in magnitude as I control for more variables, but the estimates are all statistically significant. When I conduct the analysis separately by gender, I still find that the impact of linguistic distance is larger in magnitude for

women than for men.

The results of the socioeconomic control variables are consistent across all models (Table 3.4, Table 3.5 and Table 3.6). I present the richest specification of each of those tables in columns (1) (2) and (3) of Table 3.9 below. Appendix 3.1 also shows the detailed country fixed effects. I find that age has a positive effect on the use of an official language at home, with the square of age indicating a diminishing effect as age increases. Male immigrants are more likely to use an official language at home than female immigrants, and the effects of years since migration are positive across all specifications and are economically and statistically significant. This is in line with the estimates found in the literature, which suggests that immigrants increase their use of an official language as they spend more time in the host country. The positive coefficient estimates for education and for the attribute of being employed are highly economically and statistically significant in all specifications, showing that immigrants with higher levels of education and who are employed are more likely to assimilate to the host country official languages.

3.5.2 Impact of linguistic distance on the English-to-French intensity at home

I extend my analysis in the second model (equation 2) to immigrants who do use an official language at home to explore whether linguistic distances from English and French have an impact on the choice of the assimilation language. The results with the linguistic distances from French and English are presented in table 3.7, and the results of the linguistic distance specified as a dummy variable for "Closer to French" are presented in table 3.8.

In table 3.7, I find that the estimated coefficients of all specifications for the linguistic distance from French are negative and highly statistically significant, suggesting that the relative

assimilation to French as opposed to English decreases with the linguistic distance from French. Similarly, the results for the linguistic distance from English are positive and highly statistically and economically significant, indicating that the relative assimilation to English as opposed to French decreases with the linguistic distance from English. I also observe that the positive coefficients of the distance to English are smaller in magnitude than the negative coefficients of the distance to French. This suggests again that English is more attractive to immigrants than French, and that the linguistic distance to English has a lower impact on the decision to assimilate to the *lingua franca* as compared to the linguistic distance from French and the decision to assimilate to French. However, we observe that the coefficient estimates decline in magnitude in the third specification when I add the detailed country-of-birth fixed effects. This shows the importance of controlling for those effects. In the first and second specifications, for which I omitted those controls, the coefficient estimates of linguistic distance reflected partly the impact of exposure to English or French in the source country prior to immigration.

To illustrate the economic magnitude of the effect of the linguistic distance from French, we can take again as an example two immigrants, one with Vietnamese as a mother tongue, and another with Haitian Creole as a mother tongue. The linguistic distance from French for the Vietnamese immigrant is equal to 102.42, and is equal to 49.06 for the Haitian Creole, a difference of 53.36. The estimate in column (3) of table 3.7 implies that the intensity of using French at home will be lower by 29 percentage points for the immigrants with a Vietnamese mother tongue than for the immigrants with a Haitian Creole mother tongue. To also illustrate the economic magnitude of the effect of the linguistic distance from English, we can take again as an example two immigrants, one with Dutch as a mother tongue, and another with Finnish as a mother tongue. The linguistic distance from English for the immigrant with Dutch as a mother

tongue is equal to 60.73, and is equal to 104.24 for Finnish, for a difference of 43.51. The estimate in column (3) of table 3.7 implies that the intensity of using English at home will be lower by 19 percentage points for the immigrants with a Finnish mother tongue than for the immigrants with Dutch as a mother tongue.

As expected, the results of spousal characteristics show that allophone immigrants with a Francophone (Anglophone) partner are more likely to assimilate to French (English) than other immigrants. For instance, in column 3, of both tables 3.7 and 3.8, we see that allophone immigrants with a Francophone partner use French approximately 15 percentage points more of their time at home than single immigrants, and those with an Anglophone partner use English approximately 24 percentage points more of their time than single allophones immigrants. Again, the results suggest that English is more attractive than French from the perspective of immigrants, and those immigrants with an Anglophone partner will have higher assimilation rates to English than those with a Francophone partner assimilating to French. As in the previous analysis, when I repeat the analysis by gender, I observe that the magnitude of this effect is greater for men than women.

Table 3.8 indicates that immigrants with a linguistic distance closer to French are more likely to assimilate to French than to English (19.1 to 8.6 percentage points). Again, the effect decreases in magnitude in the richest specification, but is still statistically and economically significant. When I repeat the analysis by gender separately, I do not find an important difference for the effects of the linguistic distance.

The estimated coefficients of the other control variables provide further interesting information. They are presented in Table 3.9 and in Appendix 3.1, which shows the detailed country fixed effects. In columns (4) and (5) of table 3.9, we see that age has a statistically

significant positive effect on assimilating to French relative to English, but this time age squared has a non-significant negative coefficient. This could be partly explained by the fact that the additional value that English brings in the labour market becomes less important for immigrants as they become older, and therefore they might prefer the additional social value that French bring to their social interaction in Montreal, as it is the main language used in the province. The variables of education, years since migration, and being employed have positive effects on assimilating to English rather than to French. This is again an indication of the value of English as an attractive *lingua franca*. A potential explanation is that as the level of education increases, the individual becomes more aware of the value of English as an international language, which makes it more attractive than French and also more valuable within Canada. The result for the years since migration variable suggests that the longer one is in Canada, the more immigrants assimilate to English. However, it could also be related to the fact that more recent immigrants assimilate more to French than older ones.

3.6 CONCLUSION

This paper has analysed the language assimilation of immigrants in a multilingual and multicultural context, namely the Metropolitan area of Montreal. The language dynamic in the region of Greater Montreal represents a very interesting case of an environment where both official languages interact with each other, and where immigrants are an increasingly important component of the population. With the use of data drawn from the 2001, 2006 and 2016 Canadian censuses and from the 2011 National Household Survey, along with the ASJP database, I have investigated the relationships between linguistic distance and the assimilation to an official language by allophone immigrants in the Montreal metropolitan area. I found that the linguistic distances between immigrants' mother tongues and English and French have an

important impact on the assimilation choices. Immigrants tend to assimilate to the language that is closer to their mother tongue, which may be in part explained by lower assimilation costs. In addition, the results for the impact of spousal characteristics on assimilation are consistent with the previous literature, suggesting that the home environment is an important factor contributing to the linguistic assimilation of immigrants. I found that individuals exposed to an official language at home with their spouse have significantly higher rates of assimilation. I also found that the degree of language assimilation increases with the duration of time spent in the host country, with the level of education, and for those who are employed. This paper focused only on the region of Montreal, but the implication of the key findings could also be extended to other multicultural and multilingual countries or regions, and more particularly for countries where one of the official languages is an international language.

REFERENCES 3

Chiswick, Barry R., and Paul W. Miller. 1994. "Language Choice Among Immigrants in a Multi-Lingual Destination." *Journal of Population Economics,* 7 (2): 119-131.

Chiswick, Barry R., and Paul W. Miller. 2001. "A Model of Destination-Language Acquisition: Application to Male Immigrants in Canada." *Demography,* 38 (3): 391-409.

Grenier, Gilles and François Vaillancourt. 1983. "An Economic Perspective on Learning a Second Language." *Journal of Multilingual & Multicultural Development,* 4 (6): 471-483.

Grenier, Gilles.1984. "Shifts to English as Usual Language by Americans of Spanish Mother Tongue." *Social Science Quarterly,* 65(2):537-550.

Kim, Chigon, and Pyong Gap Min. 2010. "Marital Patterns and Use of Mother Tongue at Home Among Native-Born Asian Americans." *Social forces,* 89(1): 233-256.

Lazear, Edward P. 1999. "Culture and Language." *Journal of Political Economy* 107(6): S95-S126.

Lazear, Edward P. 2007. "Mexican Assimilation in the United States." *Mexican Immigration to the United States*. University of Chicago Press.

Lee, Rennie. 2018. "Spousal Characteristics and Language Use at Home: Immigrants and Their Descendants in Canada." *Sociological Perspectives,* 61(6): 874-893.

Ortega, Javier, and Gregory Verdugo. 2015. "Assimilation in Multilingual Cities." *Journal of Population Economics,* 28(3): 785-815.

Van Tubergen, Frank and Menno Wierenga. 2011. "The Language Acquisition of Male Immigrants in a Multilingual Destination: Turks and Moroccans in Belgium." *Journal of Ethnic and Migration Studies,* 37(7): 1039-1057.

Veltman, Calvin. 1983. "Language Shift in the United States. " Amsterdam: Mouton

Veltman, Calvin. 1988. "Modelling the Language Shift Process of Hispanic Immigrants." *International Migration Review,* 22(4): 545-562.

Table 3.1: Descriptive Statistics of major variables, Immigrants Allophones in Montreal - CMA, 2001, 2006, 2011 and 2016

	Means (Std.dev)			
	Total	Men	Women	Difference
Distance from French	88.69	89.06	88.36	0.700***
	(13.89)	(13.34)	(14.37)	
Distance from English	95.26	95.28	95.25	0.030
	(6.17)	(6.09)	(6.24)	
Age	51.61	51.7	51.53	0.170***
	(16.33)	(15.81)	(16.8)	
Child 0-5	0.073	0.075	0.072	0.003***
Child 6-14	0.151	0.146	0.156	-0.010***
Years since migration	19.73	19.92	19.55	0.370***
	(15.6)	(15.61)	(15.58)	
Employed	0.507	0.596	0.426	0.170***
Education:				
Below high school	0.283	0.249	0.314	-0.065***
High school	0.171	0.166	0.176	-0.010***
Some post-secondary	0.258	0.266	0.251	0.015***
Bachelor	0.156	0.161	0.151	0.010***
Graduate	0.130	0.155	0.106	0.049***
Spouse characteristics:				
Francophone partner	0.077	0.076	0.078	-0.002*
Anglophone partner	0.031	0.027	0.034	-0.007***
Allophone partner	0.614	0.677	0.555	0.122***
Observations	331,220	158,560	172,660	

Note: In the sample, the linguistic distance from English ranges from 0 to 104.24, and the linguistic distance from French ranges from 0 to 104.01. The linguistic distances have been calculated with the ASJP distance matrices, version 2.2. Standard deviations are in parentheses.

Table 3.2: Proportion of allophones immigrants using an official language at home, Montreal CMA, 2001, 2006, 2011 and 2016

	Total	Men	Women
Non-official language only	0.397	0.375	0.417
Non-official language first and an official language second	0.244	0.246	0.242
Non-official language and official language equally	0.100	0.101	0.099
Official language first and non-official language second	0.155	0.158	0.152
Official language only	0.102	0.118	0.088
Mean value of the index	0.329	0.348	0.312
Observations	331,220	158,560	172,660

Note: This table shows the proportion of immigrants using official and non-official languages at home from the study sample. This represents the ***OFFhome*** intensity index. It can be interpreted as the approximate amount of time an immigrant uses an official language (English or French) at home. Non-official language only corresponds to the case when the index is equal to 0. Non-official language first and an official language second corresponds to the case when the index is equal to 25. Non-official language and official language equally corresponds to the case when the index is equal to 50. Official language first and a non-official language second corresponds to the case when the index is equal to 75. Official language only is when the index is equal to 100.

Table 3.3: Proportion of immigrants' allophones using English and French at home, Montreal CMA, 2001, 2006, 2011 and 2016

	Total	Men	Women
English first with no French	0.329	0.329	0.328
English first, French second	0.023	0.025	0.022
English and French equally	0.060	0.064	0.057
French first and English second	0.029	0.034	0.025
French first with no English	0.556	0.546	0.566
Mean value of the index	0.619	0.613	0.625
Observations	200,225	99,230	100,995

Note: This table shows the proportion of immigrants using English and French at home from the study sample. This represents our ***Fhome*** intensity index. It can be interpreted as the approximate amount of time an immigrant uses English and French at home. English first with no French corresponds to the case when the index is equal to 0. English first and French second corresponds to the case when the index is equal to 25. English and French equally corresponds to the case when the index is equal to 50. French first and English second corresponds to the case when the index is equal to 75. French first with no English corresponds to the case when the index is equal to 100.

Table 3.4: ORDINARY LEAST SQUARES REGRESSION ON OFFICIAL LANGUAGE AT HOME INTENSITY: The Effect of Average Linguistic Distance

	Both genders (1)	Both genders (2)	Both genders (3)	Men (4)	Women (5)
Average Distance	-0.574***	-0.466***	-0.414***	-0.337**	-0.497***
	(0.008)	(0.008)	(0.146)	(0.137)	(0.164)
Spouse Characteristics (ref: Single Immigrants)					
Francophone partner		21.868***	19.433***	21.107***	16.669***
		(0.263)	(2.501)	(2.403)	(2.43)
Anglophone partner		17.644***	16.545***	17.470***	15.192***
		(0.399)	(0.981)	(1.412)	(0.982)
Allophone partner		-15.068***	-12.643***	-16.383***	-10.104***
		(0.158)	(0.869)	(1.47)	(0.708)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes
Location and city specific controls	No	No	Yes	Yes	Yes
Country-of-Birth fixed effects	No	No	Yes	Yes	Yes
Observations	331,220	331,220	331,220	158,560	172,660
R-squared	0.12	0.23	0.33	0.34	0.33

Note: The dependent variable is the official language at home intensity, and all the observations are weighted using Census weights. Location and city-specific controls include a variable for the proportion of individuals having the same mother tongue as the respondent in Montreal, a set of dummies to control for the location of residence within the Montreal region (residing in the west, center, east and other subdivisions), and an indicator for immigrants whose previous residence was outside Quebec one or five years earlier. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 3.5: ORDINARY LEAST SQUARES REGRESSION ON OFFICIAL LANGUAGE AT HOME INTENSITY: The Effect of Minimum Linguistic Distance

	Both genders (1)	Both genders (2)	Both genders (3)	Men (4)	Women (5)
Minimum Distance	-0.408***	-0.332***	-0.298***	-0.241***	-0.363***
	(0.005)	(0.005)	(0.086)	(0.089)	(0.086)
Spouse Characteristics (ref: Single Immigrants)					
Francophone partner		21.817***	19.422***	21.101***	16.651***
		(0.264)	(2.499)	(2.401)	(2.428)
Anglophone partner		17.857***	16.540***	17.463***	15.188***
		(0.398)	(0.978)	(1.411)	(0.979)
Allophone partner		-14.760***	-12.645***	-16.385***	-10.107***
		(0.158)	(0.870)	(1.471)	(0.708)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes
Location and city specific controls	No	No	Yes	Yes	Yes
Country-of-Birth fixed effects	No	No	Yes	Yes	Yes
Observations	331,220	331,220	331,220	158,560	172,660
R-squared	0.13	0.24	0.33	0.34	0.33

Note: The dependent variable is the official language at home intensity, and all the observations are weighted using Census weights. Location and city-specific controls include a variable for the proportion of individuals having the same mother tongue as the respondent in Montreal, a set of dummies to control for the location of residence within the Montreal region (residing in the west, center, east and other subdivisions), and an indicator for immigrants whose previous residence was outside Quebec one or five years earlier. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 3.6: ORDINARY LEAST SQUARES REGRESSION ON OFFICIAL LANGUAGE AT HOME INTENSITY: The effect of Closer to french Linguistic Distance Indicator

	Both genders (1)	Both genders (2)	Both genders (3)	Men (4)	Women (5)
Closer to French	-7.796***	-6.223***	-5.531**	-4.494*	-6.682***
	(0.240)	(0.218)	(2.470)	(2.594)	(2.411)
Spouse Characteristics (ref: Single Immigrants)					
Francophone partner		22.021***	19.443***	21.124***	16.666***
		(0.263)	(2.501)	(2.404)	(2.426)
Anglophone partner		16.227***	16.535***	17.458***	15.179***
		(0.398)	(0.981)	(1.412)	(0.983)
Allophone partner		-15.778***	-12.627***	-16.363***	-10.097***
		(0.158)	(0.866)	(1.468)	(0.706)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes
Location and city specific controls	No	No	Yes	Yes	Yes
Country-of-Birth fixed effects	No	No	Yes	Yes	Yes
Observations	331,220	331,220	331,220	158,560	172,660
R-squared	0.11	0.22	0.33	0.34	0.33

Note: The dependent variable is the official language at home intensity, and all the observations are weighted using Census weights. Location and city-specific controls include a variable for the proportion of individuals having the same mother tongue as the respondent in Montreal, a set of dummies to control for the location of residence within the Montreal region (residing in the west, center, east and other subdivisions), and an indicator for immigrants whose previous residence was outside Quebec one or five years earlier. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 3.7: ORDINARY LEAST SQUARES REGRESSION ON ENGLISH TO FRENCH AT HOME INTENSITY: The Effect of Linguistic Distances From French and English

	Both genders (1)	Both genders (2)	Both genders (3)	Men (4)	Women (5)
Distance from French	-0.960***	-0.904***	-0.542***	-0.502**	-0.584***
	(0.005)	(0.005)	(0.199)	(0.199)	(0.202)
Distance from English	0.522***	0.485***	0.445***	0.480***	0.414***
	(0.016)	(0.016)	(0.081)	(0.085)	(0.088)
Spouse Characteristics (ref: Single Immigrants)					
Francophone partner		21.354***	15.389***	16.851***	13.518***
		(0.315)	(2.420)	(2.459)	(2.315)
Anglophone partner		-42.866***	-24.422***	-26.887***	-23.013***
		(0.400)	(3.279)	(2.713)	(3.996)
Allophone partner		-0.496*	0.800	-0.017	1.215**
		(0.256)	(0.676)	(1.035)	(0.548)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes
Location and city specific controls	No	No	Yes	Yes	Yes
Country-of-Birth fixed effects	No	No	Yes	Yes	Yes
Observations	200,220	200,220	200,220	99,220	100,990
R-squared	0.15	0.22	0.53	0.52	0.55

Note: The dependent variable is the English to French at home intensity, and all the regressions are weighted using Census weights. Location and city-specific controls include a variable for the proportion of individuals having the same mother tongue as the respondent in Montreal, a set of dummies to control for the location of residence within the Montreal region (residing in the west, center, east and other subdivisions), and an indicator for immigrants whose previous residence was outside Quebec one or five years earlier. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 3.8: ORDINARY LEAST SQUARES REGRESSION ON ENGLISH TO FRENCH AT HOME INTENSITY: The Effect of Closer To French Linguistic Distance Indicator

	Both genders (1)	Both genders (2)	Both genders (3)	Men (4)	Women (5)
Closer to French	19.100***	18.312***	8.640**	9.934***	7.525*
	(0.358)	(0.345)	(3.774)	(3.562)	(3.956)
Spouse Characteristics (ref: Single Immigrants)					
Francophone partner		20.326***	15.379***	16.839***	13.507***
		(0.320)	(2.413)	(2.454)	(2.305)
Anglophone partner		-47.976***	-24.437***	-26.891***	-23.044***
		(0.417)	(3.280)	(2.722)	(3.995)
Allophone partner		-3.685***	0.763	-0.050	1.167**
		(0.266)	(0.675)	(1.034)	(0.546)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes
Location and city specific controls	No	No	Yes	Yes	Yes
Country-of-Birth fixed effects	No	No	Yes	Yes	Yes
Observations	200,220	200,220	200,220	99,220	100,990
R-squared	0.08	0.15	0.53	0.52	0.55

Note: The dependent variable is the English to French at home intensity, and all the regressions are weighted using Census weights. Location and city-specific controls include a variable for the proportion of individuals having the same mother tongue as the respondent in Montreal, a set of dummies to control for the location of residence within the Montreal region (residing in the west, center, east and other subdivisions), and an indicator for immigrants whose previous residence was outside Quebec one or five years earlier. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 3.9: ORDINARY LEAST SQUARES REGRESSION OF RICHEST SPECIFICATIONS FOR ALL MODELS (FULL SAMPLE) PRESENTING THE IMPACT OF SOCIOECONOMIC VARIABLES.

Dependent variable	OFFhome (1)	OFFhome (2)	OFFhome (3)	Fhome (4)	Fhome (5)
Average Distance	-0.414***				
	(0.146)				
Minimum Distance		-0.298***			
		(0.086)			
Closer to French			-5.531**		8.640**
			(2.470)		(3.774)
Distance from French				-0.542***	
				(0.199)	
Distance from English				0.445***	
				(0.081)	
Age	0.065	0.066	0.066	0.363***	0.369***
	(0.089)	(0.088)	(0.089)	(0.137)	(0.138)
Age squared	-0.004***	-0.004***	-0.004***	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	3.455***	3.452***	3.441***	-2.477***	-2.498***
	(0.569)	(0.569)	(0.570)	(0.467)	(0.468)
Child 0-5	-2.198***	-2.209***	-2.200***	-0.144	-0.162
	(0.787)	(0.788)	(0.791)	(0.441)	(0.444)
Child 6-14	0.867	0.864	0.857	2.752***	2.741***
	(0.756)	(0.756)	(0.754)	(0.677)	(0.682)
High school	9.758***	9.771***	9.726***	-3.133***	-3.088***
	(1.306)	(1.304)	(1.314)	(1.100)	(1.114)
Some post-secondary	13.415***	13.430***	13.350***	-1.613	-1.566
	(1.042)	(1.039)	(1.050)	(1.013)	(1.026)
Bachelor	14.537***	14.564***	14.464***	-3.624***	-3.578***
	(1.256)	(1.251)	(1.246)	(1.266)	(1.286)
Graduate	17.755***	17.785***	17.659***	-3.315**	-3.271**
	(1.366)	(1.360)	(1.359)	(1.388)	(1.401)
Years since migration	0.483***	0.485***	0.485***	-0.066	-0.051
	(0.111)	(0.110)	(0.111)	(0.140)	(0.141)
Years since migration squared	0.001	0.001	0.001	-0.006**	-0.006**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Employed	2.818***	2.811***	2.856***	-0.207	-0.201
	(0.315)	(0.316)	(0.317)	(0.423)	(0.427)
Francophone partner	19.433***	19.422***	19.443***	15.389***	15.379***
	(2.501)	(2.499)	(2.501)	(2.420)	(2.413)
Anglophone partner	16.545***	16.540***	16.535***	-24.422***	-24.437***
	(0.981)	(0.978)	(0.981)	(3.279)	(3.280)
Allophone partner	-12.643***	-12.645***	-12.627***	0.800	0.763
	(0.869)	(0.870)	(0.866)	(0.676)	(0.675)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Location and city specific controls	Yes	Yes	Yes	Yes	Yes
Country-of-Birth fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	331,220	331,220	331,220	200,220	200,220
R-squared	0.33	0.33	0.33	0.53	0.53

Note: All the regressions are weighted using Census weights. Location and city-specific controls include a variable for the proportion of individuals having the same mother tongue as the respondent in Montreal, a set of dummies to control for the location of residence within the Montreal region (residing in the west, center, east and other subdivisions), and an indicator for immigrants whose previous residence was outside Quebec one or five years earlier. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

APPENDIX 3.1

Table A.3: Detailed Results for Country-of-Birth Fixed Effects

	(1)	(2)	(3)	(4)	(5)
Dependent variables:	OFFhome	OFFhome	OFFhome	Fhome	Fhome
Average Distance	-0.414***				
	(0.146)				
Minimum Distance		-0.298***			
		(0.086)			
Closer to French			-5.531**	8.640**	
			(2.470)	(3.774)	
Distance from French					-0.542***
					(0.199)
Distance from English					0.445***
					(0.081)
Countries- of-Birth (ref:Italy)					
Greenland	54.160***	53.792***	48.753***	2.779	9.119**
	(5.976)	(5.791)	(5.790)	(3.880)	(3.818)
United States of America	18.124***	16.665***	14.389***	-45.508***	-38.081***
	(3.505)	(3.504)	(3.658)	(3.140)	(3.713)
Belize	72.361***	72.217***	69.095*** (2.937)	-39.233***	-35.283***
	(3.140)	(3.035)		(2.545)	(2.853)
Costa Rica	11.904***	11.675***	10.162*** (2.998)	0.623	2.637
	(3.044)	(3.001)		(2.710)	(2.594)
El Salvador	4.211	3.996	2.453	10.632***	12.629***
	(2.673)	(2.621)	(2.596)	(2.584)	(2.504)
Guatemala	4.849*	4.630*	3.100	14.926***	16.912***
	(2.685)	(2.634)	(2.610)	(2.570)	(2.481)
Honduras	7.870***	7.651***	6.133**	12.803***	14.814***
	(2.786)	(2.741)	(2.715)	(2.627)	(2.518)
Mexico	10.723***	10.488***	9.028***	-0.766	1.219

	(3.261)	(3.231)	(3.222)	(2.892)	(2.781)
Nicaragua	7.852***	7.628**	6.129**	10.829***	12.883***
	(3.009)	(2.963)	(2.953)	(2.742)	(2.644)
Panama	12.958***	12.731***	11.223***	-0.123	1.903
	(3.179)	(3.131)	(3.132)	(2.836)	(2.709)
Antigua and Barbuda	38.512***	35.066***	44.274***	-65.405***	-82.268***
	(3.889)	(4.086)	(3.443)	(2.666)	(6.575)
Aruba	39.831***	37.597***	40.304***	-43.607***	-45.371***
	(2.716)	(2.965)	(2.597)	(2.952)	(3.097)
Bahamas	52.430***	49.007***	58.076***	8.707***	-8.106
	(3.909)	(4.120)	(3.444)	(3.075)	(7.165)
Barbados	72.562***	72.074***	70.153***	-23.323***	-18.719***
	(2.069)	(2.009)	(1.990)	(1.244)	(1.918)
Cuba	10.624***	10.385***	8.948***	6.850**	8.838***
	(3.334)	(3.298)	(3.299)	(2.878)	(2.752)
Dominica	22.969***	22.183***	22.549***	-16.466***	-18.089***
	(2.542)	(2.504)	(2.533)	(2.229)	(2.143)
Dominican Republic	8.217***	7.981***	6.504**	13.107***	15.008***
	(2.800)	(2.749)	(2.738)	(2.593)	(2.473)
Grenada	67.376***	65.315***	62.945***	-50.057***	-38.890***
	(1.424)	(1.640)	(2.527)	(4.262)	(4.559)
Guadeloupe	48.894***	45.500***	54.489***	31.227***	14.343**
	(3.752)	(4.091)	(3.235)	(2.472)	(7.010)
Haiti	21.727***	18.348***	27.285***	28.191***	11.362*
	(3.467)	(3.812)	(2.917)	(2.241)	(6.857)
Jamaica	45.300***	43.448***	45.744***	-57.692***	-65.445***
	(2.172)	(2.279)	(2.141)	(1.534)	(2.912)
Martinique	37.449***	34.070***	43.009***	21.650***	4.698
	(3.836)	(4.265)	(3.225)	(2.152)	(7.029)
Netherlands Antilles	30.315***	27.737***	31.041***	-17.100***	-16.009***
	(3.700)	(3.947)	(3.502)	(2.907)	(2.921)

Puerto Rico	22.649***	22.381***	20.897***	-39.582***	-37.287***
	(2.671)	(2.629)	(2.647)	(2.103)	(2.109)
Saint Lucia	32.753***	29.365***	38.353***	-57.473***	-74.438***
	(3.069)	(3.463)	(2.438)	(1.979)	(6.754)
Saint Vincent and the Grenadines	10.858***	8.715***	13.091***	-46.938***	-58.906***
	(3.248)	(3.344)	(3.170)	(2.309)	(5.013)
Trinidad and Tobago	69.071***	68.540***	66.086***	-51.111***	-48.081***
	(2.213)	(2.036)	(1.923)	(1.731)	(2.313)
Argentina	10.468***	10.238***	8.732***	2.513	4.526*
	(2.969)	(2.940)	(2.926)	(2.621)	(2.577)
Bolivia	13.378***	13.130***	11.623***	6.070**	8.199***
	(2.895)	(2.847)	(2.840)	(2.549)	(2.500)
Brazil	15.100***	13.917***	14.145***	-1.985	-3.583
	(3.995)	(3.966)	(4.001)	(3.075)	(2.998)
Chile	11.497***	11.279***	9.742***	18.413***	20.410***
	(2.787)	(2.739)	(2.725)	(2.476)	(2.402)
Colombia	4.800	4.566	3.096	7.647***	9.645***
	(3.138)	(3.114)	(3.088)	(2.901)	(2.839)
Ecuador	9.327***	9.091***	7.566***	3.988	6.086**
	(2.834)	(2.795)	(2.782)	(2.484)	(2.450)
French Guiana	62.001***	58.577***	67.655***	32.545***	15.793**
	(4.059)	(4.318)	(3.495)	(3.111)	(7.283)
Guyana	62.079***	60.990***	60.104***	-62.139***	-64.458***
	(3.091)	(3.000)	(3.023)	(2.249)	(2.240)
Paraguay	11.843***	11.483***	9.953***	3.154	6.086**
	(3.134)	(3.100)	(3.090)	(2.649)	(2.713)
Peru	7.568**	7.341**	5.841*	11.832***	13.844***
	(3.087)	(3.048)	(3.036)	(2.804)	(2.680)
Suriname	26.506***	24.484***	24.525***	-28.619***	-16.715***
	(3.751)	(3.915)	(3.975)	(4.096)	(4.495)

Uruguay	10.423***	10.182***	8.591***	14.103***	16.207***
	(2.560)	(2.513)	(2.485)	(2.301)	(2.282)
Venezuela	6.736**	6.506**	4.987	-3.910	-1.823
	(3.306)	(3.282)	(3.264)	(2.979)	(2.891)
Austria	37.923***	36.134***	33.516***	-30.123***	-19.931***
	(1.298)	(1.543)	(2.538)	(3.899)	(4.240)
Belgium	38.575***	37.167***	35.769***	4.850	14.296***
	(2.831)	(2.930)	(3.120)	(2.977)	(3.402)
France	40.391***	40.048***	36.641***	19.118***	23.576***
	(2.683)	(2.592)	(2.534)	(1.680)	(2.504)
Germany	35.848***	34.037***	31.453***	-28.502***	-18.032***
	(1.769)	(1.950)	(2.705)	(3.792)	(4.208)
Luxembourg	61.989***	60.166***	57.871***	6.857**	17.102***
	(2.965)	(3.088)	(3.461)	(3.350)	(3.663)
Netherlands	41.505***	39.490***	39.309***	-24.767***	-12.083***
	(2.417)	(2.693)	(2.970)	(3.992)	(4.279)
Switzerland	39.454***	37.708***	35.455***	-5.727	4.139
	(2.614)	(2.739)	(3.167)	(3.555)	(3.920)
Bulgaria	7.276*	8.016*	-0.328	4.265	4.002
	(4.206)	(4.235)	(4.459)	(4.160)	(3.532)
Czech Republic	30.938***	31.352***	23.441***	-24.306***	-23.378***
	(2.735)	(2.702)	(3.287)	(3.953)	(3.370)
Slovakia	31.312***	31.861***	23.319***	-22.645***	-22.669***
	(2.499)	(2.468)	(3.102)	(3.968)	(3.439)
Czechoslovakia	36.863***	36.772***	29.760***	-28.986***	-26.283***
	(1.660)	(1.563)	(2.656)	(4.095)	(3.801)
Hungary	27.834***	27.946***	18.198***	-29.713***	-29.350***
	(2.129)	(1.880)	(2.711)	(4.048)	(4.172)
Poland	17.066***	17.391***	13.084***	-28.924***	-21.645***
	(2.768)	(2.676)	(2.417)	(1.850)	(3.402)

Romania	6.643*	7.344**	7.004*	5.347**	7.945***
	(3.635)	(3.605)	(3.624)	(2.706)	(2.665)
Estonia	25.649***	25.709***	18.738***	-36.070***	-31.947***
	(2.021)	(1.695)	(1.406)	(2.627)	(4.546)
Latvia	17.881***	18.269***	11.838***	-38.692***	-33.398***
	(1.911)	(1.802)	(1.670)	(2.526)	(4.145)
Lithuania	28.933***	29.112***	24.336***	-35.796***	-29.382***
	(1.695)	(1.509)	(1.262)	(2.404)	(3.884)
Belarus	5.169	5.979	-3.060	-10.682**	-11.702***
	(3.868)	(3.903)	(4.178)	(4.129)	(3.506)
Moldova	4.586	5.378	2.354	17.665***	19.089***
	(4.068)	(4.077)	(4.059)	(3.243)	(3.193)
Russian Federation	6.681*	7.493*	-1.539	-11.518***	-12.463***
	(3.876)	(3.902)	(4.178)	(4.064)	(3.374)
Ukraine	5.187	5.836*	-1.400	-16.586***	-14.591***
	(3.475)	(3.490)	(3.424)	(2.875)	(3.261)
Ireland	64.760***	64.639***	54.409***	-34.329***	-33.344***
	(4.093)	(3.875)	(3.965)	(4.209)	(4.572)
Denmark	39.555***	37.916***	35.948***	-30.204***	-19.717***
	(2.451)	(2.615)	(3.072)	(4.004)	(4.229)
Finland	44.003***	42.973***	37.507***	-39.985***	-34.073***
	(3.651)	(3.178)	(2.836)	(2.519)	(4.201)
Iceland	32.339***	31.730***	27.322***	-37.670***	-31.664***
	(4.179)	(4.171)	(4.483)	(4.664)	(4.363)
Norway	38.800***	36.647***	34.513***	-26.939***	-15.266***
	(2.633)	(2.807)	(3.280)	(4.453)	(5.004)
Sweden	36.647***	34.524***	33.484***	-25.366***	-13.047***
	(3.832)	(3.981)	(4.145)	(4.247)	(4.678)
United Kingdom	32.672***	32.105***	28.668***	-37.080***	-31.752***
	(3.176)	(3.114)	(3.135)	(2.700)	(3.384)

Albania	20.038***	20.310***	16.096***	-31.925***	-25.813***
	(3.900)	(3.848)	(3.652)	(2.781)	(3.324)
Greece	1.915	2.339	-2.605**	-49.154***	-42.197***
	(2.082)	(1.896)	(1.183)	(1.427)	(3.374)
Malta	30.820***	29.823***	28.110***	-0.455	-2.244
	(2.150)	(1.956)	(1.965)	(2.828)	(3.093)
Portugal	12.768***	11.616***	11.691***	14.839***	12.905***
	(2.178)	(2.105)	(2.124)	(1.875)	(1.902)
Spain	13.769***	13.595***	11.937***	10.320***	12.139***
	(1.446)	(1.377)	(1.324)	(1.093)	(1.398)
Bosnia	13.159***	13.708***	6.267	-0.119	0.856
	(3.705)	(3.702)	(4.032)	(3.991)	(3.292)
Croatia	20.892***	21.527***	14.348***	-24.792***	-24.187***
	(2.392)	(2.393)	(2.983)	(3.683)	(2.882)
Macedonia	15.302***	16.329***	7.834*	-19.331***	-20.230***
	(3.988)	(4.037)	(4.288)	(3.977)	(3.255)
Serbia	19.043***	19.415***	12.758***	-7.919**	-5.855*
	(3.096)	(3.094)	(3.449)	(3.455)	(2.984)
Slovenia	26.501***	27.462***	18.974***	-34.657***	-35.572***
	(1.614)	(1.642)	(2.539)	(3.955)	(3.158)
Yugoslavia	20.304***	20.177***	14.576***	-16.996***	-14.060***
	(2.030)	(2.001)	(2.628)	(3.457)	(3.086)
Benin	52.118***	51.498***	40.625***	38.447***	39.910***
	(4.882)	(4.645)	(4.697)	(4.100)	(4.693)
Burkina Faso	59.039***	58.254***	48.276***	37.129***	37.457***
	(4.683)	(4.465)	(4.555)	(4.026)	(4.336)
Cape Verde	45.058***	43.785***	44.372***	25.423***	23.079***
	(3.812)	(3.770)	(3.813)	(2.619)	(2.626)
Cote d'Ivoire	57.619***	57.096***	47.161***	35.697***	38.062***
	(4.455)	(4.214)	(4.131)	(3.593)	(4.464)

Gambia	30.584***	29.999***	19.275***	-38.042***	-36.311***
	(4.506)	(4.222)	(4.143)	(4.019)	(4.731)
Ghana	34.953***	34.479***	25.215***	-33.801***	-31.152***
	(4.055)	(3.801)	(3.683)	(3.500)	(4.367)
Guinea	44.662***	44.038***	33.413***	38.285***	39.664***
	(4.512)	(4.260)	(4.271)	(3.955)	(4.595)
Guinea-Bissau	29.847***	28.211***	28.030***	15.360***	13.510***
	(3.648)	(3.637)	(3.630)	(2.692)	(2.571)
Liberia	30.210***	29.937***	22.751***	-18.830***	-13.918***
	(3.814)	(3.638)	(3.248)	(3.203)	(4.650)
Mali	44.972***	44.395***	34.028***	39.962***	41.526***
	(4.302)	(4.051)	(3.977)	(3.780)	(4.532)
Mauritania	24.969***	24.587***	17.304***	32.222***	36.646***
	(3.901)	(3.749)	(3.437)	(2.834)	(4.210)
Niger	38.423***	37.779***	26.677***	45.255***	46.619***
	(4.799)	(4.583)	(4.586)	(4.228)	(5.007)
Nigeria	49.498***	48.965***	39.458***	-34.211***	-31.807***
	(4.258)	(4.032)	(3.922)	(3.926)	(4.836)
Senegal	43.781***	43.328***	35.605***	32.237***	35.660***
	(4.166)	(3.989)	(3.775)	(2.944)	(4.027)
Sierra Leone	25.792***	25.256***	18.941***	-15.167***	-12.928***
	(3.347)	(3.142)	(2.877)	(2.438)	(3.130)
Togo	45.093***	44.462***	33.488***	32.199***	33.581***
	(4.509)	(4.271)	(4.332)	(4.013)	(4.638)
Burundi	46.137***	45.163***	34.599***	34.929***	37.522***
	(4.778)	(4.475)	(4.477)	(4.152)	(4.973)
Comoros	83.388***	83.238***	77.564***	27.934***	34.908***
	(3.725)	(3.687)	(3.234)	(4.215)	(5.136)
Djibouti	34.694***	34.158***	28.385***	31.137***	37.171***
	(4.428)	(4.183)	(3.822)	(2.956)	(4.158)

Eritrea	24.613***	24.772***	19.451***	-39.030***	-32.861***
	(4.171)	(4.037)	(3.770)	(3.039)	(3.842)
Ethiopia	23.490***	23.610***	14.932***	-26.203***	-24.697***
	(4.173)	(4.035)	(3.879)	(3.430)	(3.957)
Kenya	31.570***	31.043***	26.987***	-46.421***	-43.780***
	(3.579)	(3.412)	(3.255)	(2.705)	(3.144)
Madagascar	31.022***	30.760***	25.394***	27.134***	33.620***
	(3.873)	(3.634)	(3.292)	(2.497)	(3.578)
Malawi	24.579***	24.276***	20.080***	-48.722***	-42.703***
	(4.108)	(3.931)	(3.703)	(2.773)	(4.220)
Mauritius	23.721***	20.643***	28.046***	6.971***	-8.112
	(3.663)	(3.896)	(3.333)	(2.656)	(6.311)
Mozambique	29.337***	28.371***	27.699***	-0.280	-1.116
	(3.606)	(3.545)	(3.575)	(2.500)	(2.341)
Reunion	60.671***	57.288***	66.202***	38.685***	21.663***
	(2.683)	(3.277)	(2.089)	(1.484)	(6.167)
Rwanda	48.338***	47.735***	37.285***	29.850***	31.627***
	(4.699)	(4.471)	(4.451)	(4.165)	(4.693)
Seychelles	20.280***	16.899***	25.790***	-25.523***	-42.211***
	(3.728)	(4.033)	(3.230)	(2.621)	(6.952)
Somalia	25.391***	24.715***	18.910***	3.270	9.299**
	(4.493)	(4.159)	(3.790)	(3.062)	(4.272)
Tanzania	33.279***	32.770***	26.673***	-39.653***	-37.065***
	(3.866)	(3.673)	(3.459)	(2.963)	(3.648)
Uganda	24.070***	23.808***	19.320***	-46.984***	-42.585***
	(3.370)	(3.175)	(2.936)	(2.706)	(3.546)
Zambia	3.113	2.386	0.145	-31.642***	-27.296***
	(3.743)	(3.653)	(3.603)	(3.719)	(3.878)
Zimbabwe	56.317***	56.063***	45.678***	-52.868***	-51.525***
	(4.392)	(4.222)	(4.171)	(3.831)	(4.454)

Algeria	33.749***	33.612***	28.023***	26.277***	33.234***
	(3.802)	(3.707)	(3.301)	(3.250)	(4.425)
Egypt	21.465***	21.367***	15.961***	-11.688***	-5.154
	(2.949)	(2.802)	(2.363)	(2.250)	(3.978)
Libya	15.774***	15.659***	10.672***	-39.789***	-33.816***
	(3.272)	(3.198)	(2.808)	(2.925)	(4.197)
Morocco	28.567***	28.434***	22.904***	23.839***	30.600***
	(3.535)	(3.436)	(3.018)	(3.452)	(4.684)
Sudan	24.832***	24.733***	19.140***	-19.751***	-12.915***
	(3.224)	(3.072)	(2.636)	(2.551)	(4.102)
Tunisia	26.622***	26.492***	20.873***	24.592***	31.430***
	(3.583)	(3.498)	(3.075)	(3.474)	(4.831)
Angola	29.832***	28.742***	28.199***	14.859***	14.194***
	(3.559)	(3.483)	(3.513)	(2.686)	(2.493)
Cameroon	64.340***	63.744***	53.428***	32.296***	33.986***
	(4.628)	(4.410)	(4.401)	(3.905)	(4.565)
Central African Republic	56.058***	55.977***	49.048***	35.549***	40.452***
	(4.038)	(3.979)	(3.550)	(3.046)	(4.154)
Chad	36.095***	35.905***	30.088***	26.689***	33.029***
	(3.544)	(3.418)	(3.008)	(3.196)	(4.545)
Congo	46.585***	46.025***	39.549***	30.426***	34.986***
	(4.641)	(4.407)	(4.124)	(3.118)	(4.008)
Equatorial Guinea	38.215***	37.985***	36.494***	32.797***	34.758***
	(3.042)	(3.015)	(3.024)	(2.715)	(2.732)
Gabon	43.993***	43.570***	34.727***	12.184***	15.483***
	(4.050)	(3.881)	(3.696)	(3.731)	(4.778)
Sao Tome	50.913***	49.711***	49.963***	24.970***	23.130***
	(3.560)	(3.504)	(3.547)	(2.839)	(2.787)
Botswana	88.256***	87.642***	76.974***	-54.483***	-52.969***
	(5.090)	(4.980)	(4.920)	(4.426)	(5.166)

Namibia	59.380***	58.168***	51.153***	-39.125***	-33.239***
	(3.699)	(3.539)	(3.983)	(4.120)	(4.524)
South Africa	25.356***	24.905***	21.307***	-29.446***	-25.240***
	(3.351)	(3.240)	(3.131)	(3.586)	(4.044)
Afghanistan	7.031*	7.480*	3.044	-9.738***	-2.551
	(3.987)	(3.948)	(3.698)	(3.036)	(3.648)
Cyprus	8.839***	9.127***	4.215*	-49.234***	-42.257***
	(2.953)	(2.803)	(2.405)	(1.921)	(3.463)
Iran	10.685**	11.093***	6.616*	-35.218***	-28.121***
	(4.129)	(4.093)	(3.901)	(3.056)	(3.819)
Bahrain	16.682***	16.577***	11.206***	-59.858***	-53.193***
	(3.301)	(3.232)	(2.816)	(3.090)	(4.469)
Iraq	10.632***	10.546***	5.035*	-45.097***	-38.339***
	(3.159)	(3.038)	(2.605)	(2.645)	(4.379)
Israel	24.615***	24.399***	20.400***	-38.517***	-32.612***
	(3.598)	(3.477)	(3.397)	(2.922)	(3.775)
Jordan	16.323***	16.198***	10.581***	-51.657***	-44.741***
	(3.554)	(3.469)	(3.035)	(3.403)	(4.843)
Kuwait	14.660***	14.552***	9.008***	-49.917***	-43.056***
	(3.610)	(3.512)	(3.110)	(3.430)	(4.816)
Lebanon	15.309***	15.179***	9.590***	3.008	9.868**
	(3.314)	(3.161)	(2.719)	(3.050)	(4.435)
Oman	37.178***	37.043***	31.449***	-66.875***	-59.989***
	(3.620)	(3.562)	(3.117)	(3.502)	(5.129)
Palestine	11.458***	11.346***	5.705**	-46.356***	-39.467***
	(3.295)	(3.169)	(2.723)	(2.892)	(4.437)
Qatar	2.455	2.362	-3.056	-47.645***	-40.918***
	(3.731)	(3.627)	(3.269)	(3.334)	(4.590)
Saudi Arabia	19.935***	19.818***	14.298***	-46.530***	-39.654***
	(3.673)	(3.599)	(3.184)	(3.490)	(4.881)

Syria	9.601***	9.458***	3.951	-7.953***	-1.139
	(3.292)	(3.147)	(2.701)	(2.861)	(4.324)
United Arab Emirates	18.201***	18.091***	12.566***	-49.673***	-42.889***
	(3.629)	(3.547)	(3.153)	(3.367)	(4.886)
Yemen	21.226***	21.155***	16.057***	-50.042***	-43.590***
	(3.352)	(3.233)	(2.863)	(2.854)	(4.195)
Turkey	13.670***	13.458***	8.102**	-12.946***	-6.270
	(3.837)	(3.599)	(3.230)	(2.589)	(3.828)
Kazakhstan	3.377	4.200	-4.903	0.344	-0.786
	(4.101)	(4.133)	(4.383)	(4.274)	(3.576)
Kyrgyzstan	8.286**	9.035**	0.010	9.050**	8.193**
	(4.188)	(4.215)	(4.458)	(4.358)	(3.731)
Tajikistan	9.267**	9.983**	2.114	12.489***	13.855***
	(4.141)	(4.158)	(4.154)	(3.688)	(3.670)
Turkmenistan	-5.483	-4.628	-13.848***	7.330*	6.050*
	(4.020)	(4.054)	(4.328)	(4.305)	(3.601)
Uzbekistan	4.184	4.966	-3.992	-7.536*	-8.280**
	(4.084)	(4.114)	(4.343)	(4.234)	(3.545)
Armenia	7.410*	7.247*	1.921	-12.969***	-7.251*
	(4.182)	(4.042)	(3.736)	(2.965)	(3.817)
Azerbaijan	3.691	4.349	-4.477	-3.357	-3.585
	(4.125)	(4.120)	(4.263)	(3.956)	(3.589)
Georgia	-1.682	-1.137	-9.151**	2.326	2.015
	(4.025)	(4.011)	(3.955)	(3.881)	(3.481)
China	1.233	0.860	-5.292	-43.580***	-36.190***
	(4.173)	(3.954)	(3.490)	(2.893)	(4.449)
Hong Kong	4.059	3.833	-1.858	-54.092***	-47.004***
	(3.975)	(3.738)	(3.352)	(2.713)	(4.107)
Macau	0.671	0.473	-5.183	-58.918***	-51.854***
	(3.847)	(3.622)	(3.208)	(2.415)	(4.078)

Japan	29.487***	28.736***	17.722***	-28.273***	-26.451***
	(4.691)	(4.367)	(4.442)	(4.171)	(4.726)
North Korea	8.147**	8.114***	-3.242	-34.471***	-34.751***
	(3.413)	(3.111)	(3.559)	(4.786)	(5.220)
South Korea	6.914	6.822	-4.371	-33.748***	-33.801***
	(4.506)	(4.348)	(4.401)	(4.286)	(4.536)
Mongolia	12.540***	13.327***	13.582***	32.718***	35.183***
	(3.847)	(3.842)	(3.885)	(2.864)	(2.803)
Taiwan	6.525	6.126	-0.261	-48.580***	-40.989***
	(4.427)	(4.168)	(3.719)	(2.885)	(4.556)
Brunei	21.551***	21.120***	14.798***	2.283	9.207
	(3.896)	(3.602)	(3.269)	(6.076)	(7.299)
Cambodia	7.707*	7.526**	-2.795	24.322***	25.053***
	(4.009)	(3.697)	(3.337)	(3.522)	(4.230)
East Timor	70.799***	71.695***	62.392***	24.923***	23.502***
	(4.094)	(4.012)	(4.376)	(4.208)	(3.222)
Indonesia	28.382***	27.966***	22.685***	-47.273***	-39.161***
	(4.135)	(3.934)	(3.664)	(2.915)	(4.164)
Laos	9.776**	9.660***	4.350	23.845***	30.713***
	(3.896)	(3.647)	(3.279)	(2.835)	(3.798)
Malaysia	29.113***	28.759***	23.022***	-50.665***	-43.695***
	(4.226)	(3.960)	(3.623)	(3.060)	(4.306)
Myanmar	6.997**	6.785**	1.414	-39.312***	-33.103***
	(3.411)	(3.158)	(2.742)	(2.496)	(3.573)
Philippines	28.406***	28.089***	21.516***	-58.663***	-50.611***
	(4.442)	(4.156)	(3.694)	(3.070)	(4.614)
Singapore	50.785***	50.483***	44.740***	-54.246***	-47.134***
	(4.036)	(3.772)	(3.430)	(2.708)	(4.198)
Thailand	14.735***	14.485***	6.871**	7.098**	10.933***
	(4.109)	(3.820)	(3.382)	(2.756)	(4.033)

Vietnam	2.082	1.529	-5.204	6.613**	14.139***
	(4.172)	(3.770)	(3.210)	(2.544)	(4.490)
Bangladesh	4.998	5.465	0.937	-57.823***	-50.412***
	(4.214)	(4.173)	(3.935)	(3.243)	(3.858)
India	12.266***	12.041***	7.653**	-54.679***	-49.232***
	(3.975)	(3.833)	(3.606)	(3.016)	(3.713)
Nepal	25.269***	24.970***	16.010***	-53.877***	-50.566***
	(4.611)	(4.334)	(3.860)	(3.546)	(4.829)
Pakistan	10.963***	11.090***	6.921*	-56.294***	-50.110***
	(4.082)	(4.006)	(3.802)	(3.168)	(3.658)
Sri Lanka	7.047	6.612	1.208	-56.450***	-51.023***
	(4.268)	(4.043)	(3.715)	(3.013)	(3.916)
Australia	36.099***	35.862***	31.329***	-35.061***	-27.325***
	(3.423)	(3.312)	(3.216)	(2.584)	(3.942)
Fiji	36.251***	36.127***	32.032***	-36.135***	-30.685***
	(3.830)	(3.644)	(3.509)	(2.729)	(3.375)
New Caledonia	15.011***	14.415***	7.534**	25.178***	32.830***
	(4.074)	(3.561)	(2.899)	(2.417)	(4.449)
New Zealand	67.356***	66.749***	63.375***	-48.578***	-39.500***
	(3.915)	(3.857)	(3.820)	(3.280)	(4.324)
Others	24.794***	23.914***	21.417***	-19.782***	-16.035***
	(3.302)	(3.190)	(3.147)	(2.648)	(2.820)
Observations	331,221	331,221	331,221	200,224	200,224
R-squared	0.33	0.33	0.33	0.53	0.53

Note: All the regressions are weighted using Census weights. All regressions include the full set of controls variables: age, age squared, employment status, gender, education, years since migration, years since migration squared, and an indicator variable for the presence of children aged 0-5 or 6-14, spousal characteristics, year-specific dummies and location and city-specific controls. Robust standard errors in parentheses are clustered at the countries of birth level. Significance: *** at 1%, ** at 5%, * at 10%.