June 2017

Essays on Policies Related to Immigration, School Choice, and Crime

Georgy Orlov
The University of Western Ontario

Supervisor
Dr. Chris Robinson
The University of Western Ontario

Joint Supervisor
Dr. Nirav Mehta
The University of Western Ontario

Graduate Program in Economics

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Abstract

This thesis consists of three policy-motivated chapters in the area of applied microeconomics.

In chapter 1, I estimate the impact of English-language courses on the wages of new immigrants. I develop a model of immigrants' investment in language skills which may affect wages directly, as well as change the proportion of pre-immigration skills transferred into the host-country economy. Using unique panel data, LSIC, I find that attending language courses for six months leads to a 0.3 standard deviations gain in language skills, corresponding to an average wage increase of 11.7 percent. The increase in the total return to language skill accounts for 46 percent of this wage growth, while the remaining 53 percent is driven by the transfer of pre-immigration skills.

Chapter 2 examines the determinants of school choice and its effect on student outcomes. For any school choice policy to be successful, parents must select schools based on attributes that improve students' academic achievement. Using ECLS-K data, I find that students who move schools for academic reasons suffer a decline in their math performance. I estimate a random utility model of parental school choice and a test score production function to provide an explanation for this finding. Parents seem to select schools based on their socio-economic attributes while ignoring attributes important for test score production. Potentially, this results in worsened academic performance

of their children.

The Becker (1968) model of crime establishes the importance of the probability of apprehension as a key factor in a rational individual's decision to commit a crime. Most empirical studies based on US data have relied on variation in the number of police officers to estimate the impact of the probability of apprehension or capture. In chapter 3, the probability of apprehension is measured by clearance rates and their effects on crime rates are studied using a panel of Canadian provinces from 1986 to 2005. OLS, GMM, GLS, and IV estimates yield statistically significant elasticities of clearance rates, ranging from -0.2 to -0.4 for violent crimes and from -0.5 to -0.6 for property crimes. These findings reflect the importance of police force crime-solving productivity.

Keywords: immigrants, language, skills, task-based approach, human capital, school choice, value-added, violent crime, property crime, probability of apprehension, clearance rates, Canada

Co-Authorship Statement

This thesis contains material co-authored with Philip A. Curry and Anindya Sen. All the authors are equally responsible for the work which appears in Chapter 3 of this thesis.

Acknowledgments

I am indebted to Chris Robinson and Nirav Mehta for patient and encouraging supervision. I further want to thank Nirav Mehta for the opportunity to work as his research assistant, which greatly helped me develop my skills as an economist and as a researcher in general. I would like to thank Todd Stinebrickner and Salvador Navarro for many illuminating conversations.

I am grateful to my co-authors from the University of Waterloo, Philip Curry and Anindya Sen, for making my first co-authorship experience a truly wonderful one. I want to thank John Burbidge and Matt Doyle for acting as great mentors, even after I graduated from the University of Waterloo. Leigh MacDonald has encouraged my love for teaching and I thank her.

I would also like to thank the participants, referees, and discussants for their insightful comments and recommendations at a number of workshops and conferences, including the 49$^{\text{th}}$ and the 50$^{\text{th}}$ Canadian Economic Association Meetings, the University of Waterloo Economics PhD Conference, and the Western Economics 50$^{\text{th}}$ Anniversary Conference. Seminars in the Department of Economics at the University of Western Ontario have been a constant source of feedback and constructive criticism and I am grateful to the organizers, Jin Zhou and Antonella Mancino, and all the faculty who have attended the student talks over the years, including Audra Bowlus, Lance Lochner, Tim Conley, David Rivers, and Joseph Mullins.

The completion of the dissertation would not have been possible without the assistance of the Statistics Canada Research Data Center staff (Brad Corbett, Nathalie Goodwin, Mphatso Mlotha, Orsolya Gyorgy, Ashley Calhoun, and Glenda Babe). Dale Klassen and Labour and Immigration Manitoba have provided invaluable assistance with collection of EAL providers data for Manitoba. I thank the Economics Department at UWO, Ontario Ministry of Education, and Social Sciences and Humanities Research Council of Canada for financial support.

The economics staff at Western have ensured that my teaching and research activities ran smoothly on the administrative end of things. Yvonne Adams, Sandra Augustine, Karin Feulgen, Jennifer Hope, Leslie Kostal, Debra Merrifield, Maureen O'Connell, and Sharon Phillips have my heartfelt gratitude.

During my time at the University of Western Ontario, I have met many good friends who have offered support and inspiration. I extend my gratitude to my fellow graduate students, Bohan Li, Youjin Choi, Hyeongsuk Jin, Wenya Wang, Peter Foltin, Zinaida Kalinina-Foltin, Galyna Grynkiv, Brandon Malloy, Miguel Cardoso, Diego Salazar, Andrew Naaum, and the late Michael Padbury. I am especially grateful to Hiroaki Mori and Sergii Pypko for comments provided during the writing of the first two chapters of my dissertation.

Finally, I am deeply grateful to my parents, Viacheslav and Elena, for their unconditional love and unwavering support of all my endeavours, including my passion for learning, teaching, and research. From the moment we've met, my wife, Masha, has been a tremendous source of support, encouragement, patience, understanding, and joy. I would not have made it this far without the three of them.

To my parents, Viacheslav and Elena,

and to my love, Masha.

Table of Contents

List of Tables

List of Figures

List of Appendices

Introduction

My doctoral dissertation consists of three chapters each of which can be considered a separate essay. All three essays are motivated by policy-relevance and contribute to different areas of applied microeconomics. The first chapter, *"The Impact of Language Training on the Transfer of Pre-Immigration Skills and the Wages of Immigrants"*, contributes to the field of labour economics with the focus on investment in and returns to human capital. The second chapter, *"Understanding Traditional School Choice and Student Achievement: Evidence from the United States"*, contributes to the field of education economics focusing on school choice. The third chapter, *"Crime, Apprehension and Clearance Rates: Panel Data Evidence from Canadian Provinces"*, written in co-authorship with Philip A. Curry and Anindya Sen, contributes to the field of the economics of crime examining the connection between clearance rates and crime rates.[1]

The first chapter estimates the impact of English-language courses on the wages of new immigrants. These estimates are the first of the kind in the literature. I develop a model of immigrants' investment in language skills.

[1]Curry et al. (2016) has been published in the Canadian Journal of Economics, 49(2).

Two mechanisms through which language may affect immigrants' wages are modeled explicitly. First. there is a "direct" channel: a higher total return to language skills. Second, there is an "indirect" channel (higher language proficiency allows the immigrant to use a higher percentage of pre-immigration skills in the labour market). There are no market frictions and the immigrant allocates time between working and participating in language courses. The model serves two purposes. First, it clearly demonstrates the endogenity issues which have to be addressed when estimating the parameters of the model. Specifically, time spent in language courses depends on the immigrant's work and language learning abilities which are unobservable to the econometrician. The model further indicates that course participation cost-shifters (distance to the nearest course provider and refugee status) can serve as instrumental variables (IV). Second, the model provides the structure for the return to time spent in language courses. I use a unique panel data set, the Longitudinal Survey of Immigrants to Canada (LSIC), to estimate the four equations of the model: language skill evolution, log-wage, and cognitive and manual skill transfer equations. LSIC uses self-reported measures of English proficiency, and the data patterns indicate that these variables may be subject to measurement error (as in Dustmann and Van Soest (2002)). I use IV to estimate the language evolution equation and a combination of IV and fixed effects to consistently estimate the log-wage and skill transfer equations. I find that attending language courses for six full-time equivalent months leads to a 0.3 standard deviations gain in language skills, corresponding to an average wage increase of 11.7 percent. The increase in the total return to language skill accounts for 5.5 percent of wage growth. The remaining average gain of 6.2 percent is driven by the transfer of pre-immigration cognitive skills into the host-country economy. The standard deviation of the return from the transfer of cognitive skills is 2.5 percent. Given the declining labour market performance of successive cohorts of immigrants to Canada, the immigration policy which selects individuals with high cognitive skills, and the recent cuts to language program budgets, the findings have significant policy relevance.

The second chapter aims to understand what school characteristics drive parental school choice and whether the school moves lead to an improvement in the students' cognitive achievement. The essay focuses on *traditional school choice*: residential moves that are motivated by desire to enroll a child at a specific school.[2] As noted in Caetano and Macartney (2014), while understanding this type of school choice is important for the effective implementation of school choice policy, research on the matter is very limited. I study students in public elementary schools in the United States using the data from the Kindergarten Cohort of the Early Childhood Longitudinal Study (ECLS-K). The data contain responses to surveys conducted with students, their parents, teachers, and school administrators. Using the parental answers I am able to identify which of them exercised *traditional school choice*. I find that over thirty percent of residential moves are motivated by exercising school choice. I further

[2]Following Hoxby (2003).

find that students who move schools underperform in math and reading after the move. To pinpoint which school inputs have a significant effect on cognitive achievement, I estimate test score production functions for mathematics and reading. I find that teacher experience, qualifications, and effort (measured as the amount of homework assigned and average time the teacher spends on the subject in-class), as well as peer quality at the school (measured as the school average in the subject) all play a significant role. I then estimate a random utility model in which school inputs that affect test scores and school amenities that have no impact on test scores both enter the parents' utility function. I find that schools' socio-economic characteristics, such as the proportion of college educated households and the proportion of white students, play a much bigger role in choosing a school than the school inputs into academics. Reconciling these findings with previous the literature, I conjecture that parents may not be well-informed about which school characteristics signal the school's ability to improve the cognitive achievement of their children.

The third chapter is motivated by Becker's (1968) theory that establishes the importance of the probability of apprehension in a rational individual's decisions to commit crime. The majority of research on the effect of police on crime focuses on the number of police officers. In this paper, we use panel data for Canadian provinces (1986 to 2005) to complement the existing literature by examining the effects of clearance rates on crime while controlling the number of police officers. A police-reported incident of crime is said to be "cleared" if an individual associated with the specific criminal act is apprehended. A clearance rate within a jurisdiction is then defined as the number of cleared crimes divided by the number of reported incidents. The deterrence effects associated with an increased probability of apprehension may be more precisely captured through clearance rates than the per capita number of police officers or arrests. Our empirical strategy is motivated by an extension of the theoretical model developed by Polinsky and Shavell (2000), which allows us to link spending on police services to corresponding changes in clearance rates and crime. We then use per capita provincial expenditures on police among the instrumental variables used to address the bias in the coefficient of interest. Additional instruments which are used in the reduced form estimation include political party variables meant to reflect the variation in the policy focus with respect to crime within provinces. We find statistically significant elasticities of crime rates with respect to clearance rates, ranging from 0.2 to 0.4 for violent crimes and from 0.5 to 0.6 for property crimes. This highlights the importance of the effectiveness of police force, not just its size, for crime reduction.

Bibliography

BECKER, G. S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76 (1968), 169–217.

CAETANO, G. AND H. MACARTNEY, "Quasi-Experimental Evidence of School

Choice through Residential Sorting," *Working Paper* (2014).

CURRY, P. A., A. SEN AND G. ORLOV, "Crime, apprehension and clearance rates: Panel data evidence from Canadian provinces," *Canadian Journal of Economics* 49 (2016), 481–514.

DUSTMANN, C. AND A. VAN SOEST, "Language and the Earnings of Immigrants," *Industrial and Labor Relations Review* (2002), 473–492.

HOXBY, C. M., "School Choice and School Productivity. Could School Choice Be a Tide That Lifts All Boats?," in *The Economics of School Choice* (University of Chicago Press, 2003), 287–342.

POLINSKY, A. M. AND S. SHAVELL, "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature* 38 (2000), 45–76.

Chapter 1

The Impact of Language Training on the Transfer of Pre-Immigration Skills and the Wages of Immigrants

1.1 Introduction

While there is a vast literature documenting the returns to immigrants' proficiency in host-country languages, the literature on the returns to host-country language training has been limited by data availability.[1] Immigrants constitute approximately a fifth of the Canadian labour force and Canada welcomes approximately 250,000 new immigrants every year.[2] Considering these

[1]Commonly, "host-country" refers to the immigrant's target country of immigration, while "source-country" – to the point of origin.

[2]According to the Labour Force Survey (LFS) in 2011 immigrants accounted for approximately 21.1 percent of the Canadian labour force as indicated on the Citizenship and Immigration Canada website http://www.cic.gc.ca/english/resources/research/

numbers, the declining labour market performance of successive cohorts of new immigrants reported in the literature [e.g., Aydemir and Skuterud (2005)], and the significant budget cuts to language programs offered to new immigrants in Canada in the recent years[3], the evaluation of the impact of ESL training on immigrants' economic outcomes is important for Canada's policies aimed at assisting with new immigrants' settlement.

In this paper I use the unique data set, Longitudinal Survey of Immigrants to Canada (LSIC), to estimate the effect of time spent in ESL courses on wages. I find that attending ESL courses full-time for six months (with a full-time monthly attendance of 126 hours) causes an 11.7 percent increase in an immigrant's wages. The increase in the total return to language skill accounts for 5.5 percent of wage growth, while the remaining 6.2 percent of the wage increase is driven by the change in the proportion of cognitive skills transferred into the Canadian economy. These are the first estimates of the kind in the literature.

Language skills can be thought of as a part of a person's human capital portfolio. Studying the return to and investment in these skills for native-born workers is problematic, as they possess "baseline" proficiency in the language. Immigrants, however, may not have "baseline" proficiency in the host-country

---

`2012-migrant/sec02.asp`. The approximate annual number of new immigrants is based on the annual immigration numbers from 1990 to 2014 found at `http://www.cic.gc.ca/english/resources/statistics/facts2014/permanent/01.asp`

[3]For example, a 22 million dollar cut to the funding of colleges offering free English as a Second Language (ESL) courses to immigrants in British Columbia in 2014 (Shen, 2014).

official language, providing an environment to study the returns to and investment in language skills. Moreover, the level of investment in host-country language skills can be clearly measured as it comes in the form of time spent in language classes. As noted by Chiswick and Miller (2014), immigrants' investment in host-country language skills has been given a very limited treatment in the literature due to data limitations. Specifically, very few data sets contain a record of immigrants' participation in host-country language courses. Among the data sets containing such information, several are cross-sectional which makes it more problematic to study changes in language skills due to investment. Furthermore, of all data on immigrants, to my knowledge, only LSIC contains a detailed record of host-country language course participation including the start and end dates and weekly in-class hours.

I construct a two-period model of investment in host-country language skills in which the immigrant decides on the optimal time spent on ESL training. Both the language evolution and the wage determination are modeled explicitly. Furthermore, I explicitly model the mechanism through which pre-immigration skills are transferred into the host-country economy. Together, the language skill return to the time spent in ESL courses, wage returns to the language skill and cognitive skills, and the pre-immigration skill transfer equation describe the returns to language course participation. The solution to the model provides two insights regarding the optimal ESL course participation. First, the time spent on ESL training depends on the immigrant's language learning ability and her work-related ability. Both of these abilities are unobserved; hence, the coefficients of the language evolution equation cannot be consistently estimated by OLS due to endogeneity. Second, variables that shift the cost of participation in the ESL courses, such as distances to ESL course providers and refugee status, affect the optimal time spent on ESL training while not being correlated with the language learning ability. This allows me to to use such cost-shifters as instrumental variables when estimating the effect of time spent in ESL courses on language skill gains.

A further complication in estimating the key equations of the model arises from the presence of measurement error in the language skill variable. While this does not impact the estimation of the coefficients in the language evolution equation, it prevents obtaining consistent estimates of the wage equation and skill transfer equation coefficients on language skill by using OLS. This issue is addressed by using instrumental variables combined with a fixed effects estimation. While fixed effects address the time-invariant component of the measurement error, the instrumental variables mitigate the correlation between the language skill and the idiosyncratic component of the measurement error. Moreover, in the case of the wage equation, using fixed effects estimation addresses the potential correlation between occupational skills and the unobserved work ability which would also serve as a threat to identification.

Unlike the majority of previous studies which use speaking ability-based binary variables to measure being "proficient" and "not proficient" in the host-country language, I use a continuous measure of language skill.[4] This measure is constructed by performing principal component analysis (PCA) on self-reported measures of speaking, reading, and writing ability and the responses to five additional questions regarding the immigrant's speaking and comprehension capabilities recorded in LSIC. To enable the estimation of the model, I augment LSIC with data from the Career Handbook, the Canadian 2001 Census of Population, and hand-collected data on the 2001-2003 ESL providers in English-speaking Canadian provinces. The hand-collected data on ESL providers contains six-character postal codes for providers' class locations which, paired with the immigrants' residential postal codes, allows me to calculate the distance between each immigrant and the nearest ESL provider. The Career Handbook data on the level of nine aptitudes and complexity of three tasks corresponding to four-digit National Occupational Classification (NOC) codes together with occupational weights corresponding to the 2001 Canadian labour market calculated using the 2001 Census are used to obtain two occupational skills using PCA. An advantage of using the Career Handbook data for the Canadian labour market versus the Dictionary of Occupational Titles (DOT) or Occupational Information Network (O*NET) data is that it obviates the assumption that the occupations within the US use the same skills as those in

[4]To be more precise, the measure is discrete but is defined on a fine grid.

Canada.[5]

I find that attending ESL courses for six months increases the English language skill by 0.291 standard deviations. The number of months spent in Canada between the interviews and the number of household members who speak English also serve as significant predictors of the language skill acquisition. I find that a one standard deviation increase in language skill results in a 18.99 percent increase in wages, which is consistent with previous literature finding significant returns for immigrants to becoming proficient in the host-country language. I also find significant returns to cognitive and manual skills, as well as work experience acquired in Canada. In line with previous studies, I find returns to cognitive skills, which encompass analytical and communication skills, to be larger than returns to manual skills which, in this paper, correspond to dexterity-related and hand-eye coordination tasks performed on the job. Finally, estimates from the skill transfer equations for cognitive and manual skills indicate that host-country language skills play an important role in transferring pre-immigration cognitive skills into the host-country economy. On the other hand, I find no evidence that language skills affect the transfer of pre-immigration manual skills.

The remainder of the paper is organized as follows. Section 1.2 provides a review of relevant literature. Section 1.3 presents the theoretical model of

[5]Imai, Stacey and Warman (2016) note the necessity of this assumption as one of the caveats in employing O*NET data in their analysis using LSIC.

investment in language skills. Section 1.4 describes the data sources and provides descriptive statistics for the estimation sample. Section 1.5 discusses the identification and estimation of the language skill evolution, wage determination, and skill transfer equations. Section 1.6 discusses the estimates from these equations and calculates the effect of ESL course attendance on wages. Section 1.7 concludes.

1.2 Review of Relevant Literature

My paper draws on and contributes to three main strands of literature: research on immigrants' post-arrival investment in human capital, papers on the return to immigrants' human capital, especially the host-country language skills, and the work on the transferability of the source-country human capital into the host-country economy. I will focus my review of the prior work on post-immigration human capital investment on papers which examine the host-country language acquisition.

The literature on the effect of host-country language courses on immigrants' outcomes is very limited, largely due to lack of data on the subject, as noted in the literature review by Chiswick and Miller (2014).[6] Gonzalez (2000) and

[6] Other examinations of language acquisition do not use data on language courses in their models. For example, Chiswick, Lee and Miller (2004) use the Longitudinal Survey of Immigrants to Australia (LSIA) to estimate a model of language acquisition focusing on the role of factors such as age at migration, visa status, and birthplace. To see whether unobservable characteristics play a role in language proficiency, Chiswick and Miller estimate bivariate

Hayfron (2001) investigate this topic using cross-sectional data. Gonzalez (2000) uses the 1992 National Adult Literacy Survey (NALS) to investigate the factors linked to immigrants' proficiency in speaking, understanding, reading, and writing in English as well as the impact of these proficiencies on earnings. Hayfron (2001) uses Norwegian data on immigrants from Morocco, Pakistan, and Chile to estimate the contribution of Norwegian language courses on language proficiency. He also estimates the effect of Norwegian proficiency on wages, but does not find a statistically significant effect even after instrumenting for the potential measurement error in the recorded proficiency.[7] Gonzalez (2000) and Hayfron (2001) use instrumental variables to address the endogenity of language course participation. Gonzalez utilizes the use of another language at work or while shopping as proxies for residing in an ethnic enclave (which should reduce incentives to participate in language courses). Hayfron uses dummies for receiving unemployment and other social benefits. Both studies use cross-sectional data with only the measure of present language abilities which are coded into binary proficiency indicators.

The lack of a pre-language-training measure of proficiency makes interpreting the probit estimates of the effect of language courses on present proficiency

probit models with English proficiency at different interviews as the dependent variable arguing that the positive correlation between the disturbance terms would indicate that the same unobservables are relevant for proficiency in the two time periods.

[7]A potential caveat is that the *ethnicity of wife* variable that Hayfron (2001) uses as one of the instruments for language proficiency does not have a statistically significant effect as a determinant of language proficiency.

problematic. Furthermore, both data sets include immigrants with a long residency in the host-country (e.g., NALS includes immigrants who have been in the US for over 40 years) and the language course questions identify only the participation and completion of the course at some point since arriving in the country.

Beenstock (1996) uses the Immigration Absorption Survey (IAS) data from Israel to examine how new immigrants acquire Hebrew. He finds a positive effect of attending a language school, but the study does not correct for potential self-selection into language courses. Beiser and Hou (2000) use panel data on Southeast Asian refugees who arrived in Vancouver (British Columbia, Canada) in 1981. The refugees were interviewed at arrival, and again two and then ten years after arrival in Canada. The study does not find a statistically significant effect of ESL courses on language acquisition in the first two years in Canada using an ordered logistic regression.[8] However, the authors do not test or correct for self-selection into ESL courses. Akresh (2007) uses the New Immigrant Survey-Pilot (NIS-P) panel data covering the immigrants' first year after receiving the US permanent resident status. Regression of log earnings on covariates including an ESL course participation dummy with fixed effects introduced to address selection into ESL courses does not yield a statistically or economically significant effect of language courses on earnings. Finally, Kaida

[8]Beiser and Hou (2000) find an effect limited only to female refugees after ten years in Canada, though there is no statistically significant effect when the interaction is not present in the model.

(2013) uses the Longitudinal Survey of Immigrants to Canada (LSIC) to examine the role of language course completion, rather than participation, and the host-country education in immigrants' exit from poverty. She estimates a bivariate probit of an indicator for language course completion and an indicator for poverty exit which has the former as an independent variable. She finds a statistically significant effect of ESL courses on poverty exit.

There are several areas in which the present paper expands the knowledge on the effects of ESL courses. First, none of the preceding papers estimate the effect of intensive margin of ESL courses. It is very likely that immigrants can benefit from attending ESL courses even if they do not officially complete the prescribed program. From the policy standpoint, knowing the returns to a fixed unit of ESL course participation time (e.g., full-time equivalent months) is useful for budget planning. Second, obtaining the returns to time spent in ESL courses in terms of wages allows cost-benefit assessment of existing and planned language programs for immigrants.

The seminal work by Chiswick (1978) that proposed that English language proficiency can play an important role in the economic assimilation of immigrants prompted the development of a vast literature concerning the returns to immigrants' language proficiency. The majority of literature covers immigrants in the US, Canada, UK, Australia, Germany, and Israel. A vast majority of these studies find that immigrants enjoy significant returns to becoming proficient in the host-country language [for examples, see Chiswick (1991), Dustmann (1994), Dustmann and Fabbri (2003), Berman et al. (2003), Bleakley and Chin (2004), Gonzalez (2005), Skuterud (2011)]. Chiswick and Miller (2014) provide a detailed overview of the literature for all of these countries.

The majority of studies of returns to immigrants' host-country language proficiency use self-reported proficiency. Immigrants may under- or over-report their level of fluency which may bias the estimates of returns to host-country language proficiency. Dustmann and Van Soest (2001, 2002) address the problem of misclassification of host-country language proficiency within the German Socio-Economic Panel (GSOEP). Dustmann and Van Soest (2001) use a model of language acquisition to correct for the probabilities of misreporting the speaking proficiency level. The authors simultaneously estimate the monthly earnings equation for full-time workers and a model of language acquisition which includes variables that are constant over time (country-of-origin dummies, year of entry, age at entry, father's education) and vary as time passes (years of education, family composition and marital status, years since migration). The estimates of the latent speaking fluency are used as a measure of language skill in the earnings equation. Dustmann and Van Soest (2002) treat the misclassification as an additive measurement error and use an instrumental variable approach to deal with the issue. They use parental education to deal with both idiosyncratic and time-independent measurement errors in language. Both studies find that misclassification results in underestimation of returns to German speaking proficiency (driven by immigrants overstating their speaking ability). In the current paper, I perform an IV estimation which includes fixed effects to address the potential sources of bias in the coefficients; however, my set of instruments differs from Dustmann and Van Soest (2002).

Some of the more recent studies have utilized data sets with objective measures of literacy (e.g., document literacy). Ferrer, Green, and Riddell (2006) estimate the returns to immigrant literacy skills using the Ontario Immigrant Literacy Survey (OILS) and the International Adult Literacy Survey (IALS). They find that there is a significant return to literacy as well as lack of difference in returns between immigrant and native-born workers. Clarke and Skuterud (2014) come to similar conclusions using data from the Adult Literacy and Life Skills Survey (ALLS) on US, Canada, and Australia.

Work on the economic assimilation of immigrants suggests a U-shaped pattern of occupational mobility: an initial downgrade from the pre-immigration job followed by a recovery, which may be attained through human capital investment. For example, Duleep and Regets (1999) develop a two-period model of immigrants' investment in human capital with an exogenously determined incomplete transfer of immigrants' source-country skills into the host-country economy. The paper offers evidence to support the predictions of the model such as higher levels of human capital investment by immigrants who are subject to lower skill transferability. Duleep (2014) provides an excellent discussion of work on occupational mobility and skill transferability of immigrants. While, until recently, the literature on the transferability of pre-immigration skills has not explicitly considered the role of host-country language proficiency in the skill transfer mechanism, the idea that returns to host-country language skills may be occupation-specific was discussed in Kossoudji (1988). Chiswick and Miller (2003) use Canadian 1991 Census data, providing evidence that higher host-country language proficiency enhances the gains from schooling and pre-immigration work experience. Imai et al. (2016) develop a theoretical model of occupational choice where the proportion of pre-immigration skills useable in the host-country depends on the level of the language skill. O*NET data is used in conjunction with LSIC and the Canadian Census of Population to obtain occupational skill vectors. The paper presents descriptive statistics supporting the mismatch between source-country and host-country occupations, with lower levels of mismatch associated with higher language proficiency levels. The authors further show that speaking proficiency (measured on a discrete three-level scale) leads to a reduction in skill gaps between pre-immigration and post-immigration occupational skills, measured as the difference in the ordinal rank of the pre- and post-immigration occupation based on the particular skill.

1.3 The Model of Immigrant's Investment in Language Skills

This section presents a model of investment in language skills for an individual i. There are two time periods; let t denote time. In each period t the immigrant is endowed with one unit of time. At $t = 1$, the immigrant can allocate one unit of time between work $wk_{i1} \in [0, 1]$ and an ESL course $esl_{i1} \in [0, 1]$. At $t = 2$, she can only work, that is, $esl_{i2} = 0$ and $wk_{i2} = 1$.[9] In what follows I denote scalars with lower case letters (e.g., l_{it}) and 1-by-k vectors with upper case letters (e.g., X_{it}^{j}, where k is j-specific) with associated coefficients being considered k-by-1 vectors. Consistently with previous research, working earns a wage described by a Mincer-style function

$$w_{it} = \exp\{l_{it}\beta_l + xp_{it}\beta_{xp} + s_{it}\beta_s + \alpha_i^W\}, \tag{1.1}$$

where l_{it} denotes the host-country language proficiency, xp_{it} is the work experience, s_{it} refers to a measure of human capital, the skill individual i uses at work, and α_i^W is the individual's work ability.[10] Equation (1.1) lends itself to the linear-in-parameters log-wage form commonly used in the literature:

$$\ln w_{it} = l_{it}\beta_l + xp_{it}\beta_{xp} + s_{it}\beta_s + \alpha_i^W\,. \tag{1.2}$$

[9]This is not a restriction, as the optimal time allocation in a two-period problem has $esl_{i2} = 0$

[10]In the present section the measure of skills is kept as a scalar for the ease of exposition. When estimating the model, I distinguish between cognitive and manual skills.

Let s_{i0} denote the individual's pre-immigration skill. The skill is transferred to the host-country economy using the technology

$$s_{it} = s_{i0}\exp\{-\tau_l(\bar{l} - l_{it})\}, \tag{1.3}$$

where $\bar{l}$ denotes the near-native language skill level (the highest language skill level for immigrants). Previous work suggests that immigrants' host-country language proficiency affects the amount of pre-immigration skills which are useable in the host-country; τ_l measures the skill transfer "penalty" imposed for not having a near-native proficiency level.[11]

New immigrants live in an English-language environment; hence, it is unlikely that their language skills would depreciate over time. Even if some immigrants reside in ethnic enclaves, this is more likely to deter acquisition of further language skills rather than cause the deterioration of existing language skills, at least in the relatively short two-year period considered in this paper. Therefore, I assume that language skills do not depreciate over time and host-country language skill evolution can be described by

$$l_{i2} = l_{i1} + esl_{i1}\varphi_{esl} + X_{i1}^L\varphi_x + \alpha_i^L\,. \tag{1.4}$$

The coefficient on l_{i1} is assumed to be one, reflecting zero depreciation of the language skill over time, X_{i1}^L represents other variables affecting language

[11]While I do not make s_{i0} explicitly depend on α_i^W, it does not change any insights obtained from the solution for optimal ESL participation. Furthermore, I address the endogeneity of s_{it} caused by s_{i0} being dependent on the unobservable α_i^W when discussing the identification of the wage equation in Section 1.5.

learning (the number of other household members who are proficient in English and the individual's age), and α_i^L is the individual's language learning ability. Without loss of generality, assume that $xp_{i1} = 0$ and $xp_{i2} = wk_{i1}$. Participating in an ESL course involves a cost $X_{i1}^C \gamma_x$. X_{i1}^C consists of the distance to the ESL provider, which accounts for travel costs in attending the course, and refugee status, which may account for some monetary costs being alleviated for refugees who attend ESL courses paired with the impact of case workers directing refugees to attend courses. The immigrant has preferences represented by the utility function

$$u_{it} = wk_{it} \cdot \ln w_{it} - esl_{it} \cdot X_{it}^C \gamma_x, \quad t \in \{1, 2\}. \tag{1.5}$$

This means that the individual extracts utility from total per-period income (all of the income gets consumed within the period), while experiencing disutility from paying costs connected to attending ESL courses. I assume that there is no psychic cost of attending ESL courses. Let ρ be the discount factor. The immigrant's problem is then

$$\max_{esl_{i1},\, wk_{i1}} \left\{u_{i1} + \rho u_{i2}\right\} = \max_{esl_{i1},\, wk_{i1}} \left\{wk_{i1} \cdot \ln w_{i1} - esl_{i1} \cdot X_{i1}^{C} \gamma_x + \rho \cdot wk_{i2} \cdot \ln w_{i2}\right\}$$

subject to

$$esl_{i1} + wk_{i1} = 1$$

$$esl_{i1} \geq 0, \quad wk_{i1} \geq 0$$

$$esl_{i2} = 0, \quad wk_{i2} = 1$$

$$l_{i2} = l_{i1} + esl_{i1}\varphi_{esl} + X_{i1}^{L}\varphi_x + \alpha_i^L$$

$$xp_{i2} = wk_{i1} + xp_{i1}$$

$$l_{i1},\ xp_{i1},\ s_{i0},\ X_{i1}^{L},\ X_{i1}^{C},\ \alpha_i^W,\ \alpha_i^L \qquad \text{given.}$$

Solving this problem for optimal investment in language skills, $esl_{i1}^{\star}$, I obtain a result that is a function of both language learning ability α_i^L and the vector of cost-of-investment variables X_{i1}^{C} in addition to other variables.

$$esl_{i1}^{\star} = \frac{1}{\tau_l \varphi_{esl}} \left(\tau_l(\bar{l} - l_{i1} - X_{i1}^{L}\varphi_x - \alpha_i^L) + \ln\left(l_{i1}\beta_l + s_{i1}\beta_s + \alpha_i^W + X_{i1}^{C}\gamma_x + \rho\beta_{xp}\right) - \ln\left(\rho\tau_l s_{i0}\varphi_{esl}\right)\right) \tag{1.6}$$

There are two insights relevant for the estimation of the equations of the model from the above solution. First, $esl_{i1}^{\star}$ is not independent from the language learning ability α_i^L and the work ability α_i^W which are unobserved in the data. Since the language skill evolution equation contains esl_{i1}, OLS estimation will yield biased results. Taking a derivative of (1.6) with respect to α_i^L we obtain

$$\frac{\partial\, esl_1^{\star}}{\partial\, \alpha_i^L} = -\frac{1}{\varphi_{esl}}\,, \tag{1.7}$$

meaning that higher language learning ability will be associated with less time spent in ESL courses. This is not surprising, since those immigrants who are better at picking up language skills through day-to-day interactions will choose to spend less time in ESL courses. Second, the cost variable vector X_{i1}^C enters $esl_{i1}^{\star}$, but does not enter the wage or the language skill evolution equations. Differentiating (1.6) with respect to X_{i1}^C we obtain

$$\frac{\partial\, esl_1^{\star}}{\partial\, X_{i1}^C} = \left(\frac{1}{\tau_l \varphi_{esl}}\right)\left(\frac{\gamma_x}{l_{i1}\beta_l + s_{i1}\beta_s + \alpha_i^W + X_{i1}^C\gamma_x + \rho\beta_{xp}}\right). \tag{1.8}$$

That is, an increase in costs associated with attending ESL courses reduces the optimal time spent in ESL courses. For example, living further away from the nearest ESL provider would decrease the optimal time spent in ESL courses. On the other hand, as being a refugee may involve reductions of these costs (e.g., subsidized public transit tickets), refugees are likely to attend ESL courses for longer periods of time. At the same time, neither of these variables would have a direct effect on language acquisition. Hence, variables in X_{i1}^C can serve as instruments for ESL_{i1} in order to address the aforementioned endogeneity issue.

1.4 Data

To estimate the parameters of the model, I use data from three sources: the Longitudinal Survey of Immigrants to Canada (LSIC), the Career Handbook paired with the 2001 Canadian Census of Population, and a personally collected data set containing the locations of LINC/EAL/ELSA providers from 2001 to 2003 paired with the Postal Code Conversion File (PCCF).

1.4.1 Longitudinal Survey of Immigrants to Canada (LSIC)

LSIC is a panel data set collected and created jointly by Statistics Canada and Citizenship and Immigration Canada under the Policy Research Initiative. The data include immigrants who arrived in Canada between October 1, 2000 and September 30, 2001. Surveys were administered to the same individuals in three waves: approximately six months (Wave 1), two years (Wave 2), and four years (Wave 3) after arrival in Canada. The target population of the survey were immigrants who were 15 years or older at the time of landing, applied for immigrant status from abroad (and were not asylum refugee claimants). This included economic immigrants (both primary applicants and their spouses and dependents), sponsored family members, and convention refugees. A sample of 20,300 immigrants representative of 165,000 immigrants arriving in this time period was chosen. 12,040 immigrants responded to Wave 1 survey. LSIC continued following only those immigrants that responded to the previous waves (monotonic design), and Wave 3 file contain data on approximately 7,500 immigrants. The target population of Wave 3 are immigrants who have resided in Canada for four years.[12]

[12]Due to the attrition in the sample, I caution that results are applicable to the new immigrants who have remained in Canada for at least four years. It is possible that immigrants who have participated in ESL courses, but are no longer followed in LSIC at later waves, could have

LSIC is a restricted-access data set and is available only through the Statistics Canada Research Data Centers. LSIC is an extremely rich data set and is uniquely suited for this study. I use data from LSIC to obtain measures of l_{i1} and l_{i2}, esl_{i1}, w_{i1} and w_{i2}, xp_{i1} and xp_{i2}, X^L_{i0} and X^L_{i1}, and two of the variables in X^C_{i1}.

LSIC contains extremely rich data on immigrants' language ability. Self-reported English-language speaking, reading, and writing ability are further complemented by responses to five questions on speaking and comprehension competence (e.g., "How easy is it for you to understand a message in English over the telephone?"). The answers to these eight questions permit me to construct a continuous measure of language proficiency using principal component analysis (PCA).[13] The continuous factor for Waves 1 and 2 of LSIC obtained by performing PCA is used as l_{i1} and l_{i2} respectively.[14]

Responses on the methods of improving English language proficiency contain participation in English as a Second Language (ESL) courses. Further, to my knowledge, LSIC is the only panel data set that contains the history of participation in ESL courses which includes weekly hours and the start and end

experienced higher gains in language and wages. I thank Michael Haan and Mikal Skuterud for their comments on the matter.

[13]The majority of previous studies use self-reported speaking proficiency to derive binary ("proficient/not proficient") or three-level categorical ("low/medium/high proficiency") measures of proficiency which are not suitable for examining the gradual increase in host-country language skills over time.

[14]Details can be found in Appendix A.2.1.

dates of the course attendance. One important limitation of LSIC is that no detailed records exist for ESL courses that were started between Wave 1 and Wave 2 interviews. Further, the record of courses from Wave 2 to Wave 3 interviews makes it impossible to separate hours of ESL and hours of French as a Second Language (FSL) training if the immigrant participated in both activities. In this paper I focus on the investment in language skills which occurred between six months and two years in Canada. I, therefore, use the available records of immigrants participating in ESL courses between Wave 1 and Wave 2 interviews as my measure of esl_{i1}.

I use data on immigrants' age and the number of household members who are proficient in English as the two measures in X_{i1}^{L}. I assume that the number of individuals in the household who speak English reported in an interview served as one of the inputs in the language evolution equation for the language proficiency reported in that interview (e.g., X_{i1}^{L} that determines l_{i2} includes the number of English-proficient household members reported in the Wave 2 interview). X_{i1}^{L} also includes the lag of age at Wave 2 (i.e., age at Wave 1 interview). I use the visa status of the immigrant to identify refugees and sponsored family members. The two visa status indicators are used as components of X_{i1}^{C}. I use the six-character postal codes of immigrants' residences in calculation of distances to ESL providers in Section 1.4.3.

LSIC contains a detailed employment history for each immigrant for the first four years in Canada from which I obtain the information on wages, Canadian work experience, and occupations worked in after the arrival in Canada. I divide the provided weekly wages by weekly hours and the provincial CPI at the start date of the job to obtain real hourly wages. The hourly wages earned between Wave 1 and Wave 2 interviews are used as a measure of w_{i1} and wages earned between Wave 2 and Wave 3 interviews - as w_{i2}.[15] Days worked between arrival in Canada and the Wave 1 interview are used to measure xp_{i1}, days worked between arrival and the Wave 2 interview - to measure xp_{i2}. I convert the number of days worked into six month-equivalent intervals.[16] The occupational information is used to match the skills, obtained from the Career Handbook in the next section, with immigrants in LSIC. Further, I use the pre-immigration occupation reported in the Wave 1 interview to obtain measures of pre-immigration skills. Data on the pre-immigration occupation is one of the unique features of LSIC not found in other data sets.

Figure 1.1 shows the timing of the data; variables under the braces were obtained from the employment and language course histories corresponding to the time between the interviews. No retrospective questions regarding the language proficiency at arrival in Canada were asked in LSIC, so the measure of l_{i0} is not available. Based on the timing, I assume that wages between Waves

[15]I use wages for the job which Statistics Canada identifies as the "main job."

[16]This choice is one of convenience: both the time spent in ESL courses and work experience are measured in the same units, as it makes interpreting and comparing the coefficients in Section 1.6 easier.

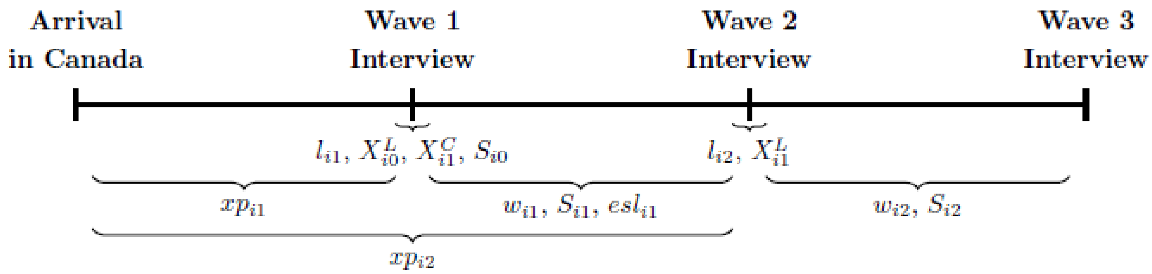


Figure 1.1: Data Timeline

1 and 2 (w_{i1}) and Waves 2 and 3 (w_{i2}) are a function of language proficiency recorded at the Wave 1 (l_{i1}) and Wave 2 (l_{i2}) interviews respectively. This leaves me with two time periods worth of data for the estimation of wage equation. I use the data for waves 1 and 2 (l_{i1}, l_{i2}, esl_{i1}, X^L_{i1}, X^C_{i1}) to estimate the language skill evolution equation.

1.4.2 Career Handbook

In Section 1.3, an individual's pre- and post-immigration skills are represented by scalars for the ease of exposition. For the empirical portion of the paper, I use two separate variables to measure cognitive and manual skills. Let S_{it} denote a two-by-one vector $\{s^c_{it}, s^m_{it}\}$. To obtain S_{i1} and S_{i2} represented by measures of cognitive and manual skills used on the job and the vector of pre-immigration skills S_{i0}, I use the Career Handbook paired with the 2001 Canadian Census of the Population. The Career Handbook is the counseling component of the National Occupational Classification (NOC) system. It contains information on 923 occupations with corresponding attributes, task complexity, and physical and environmental requirements. The attributes have five levels, with “1” corresponding to the ability within the lowest 10 percent of the population, “2” corresponding to the lowest third of the population excluding the lowest 10 percent, “3” corresponding to the middle third, “4” corresponding to the top third excluding the top ten percent of the population, and “5” corresponding to the the top ten percent of the population. Tasks in working with data, people, and machinery (things) are ranked based on their complexity on an eight-level scale. I have combined some of the levels due to the similarity of tasks performed to create four-level measures of task complexity (where “0” corresponds to the task not playing a significant role for the occupation and the rest of the levels correspond to “low”, “medium”, and “high” levels of complexity).

Many of these attributes and task measures are highly correlated. To reduce the dimensionality of data and improve its interpretation, I obtain two skill measures underlying these variables using principal component analysis (PCA). Two approaches to reducing the dimensionality of attribute and task vectors which come from the Dictionary of Occupational Titles (DOT) and Occupational Information Network (O*NET) have been discussed in the recent literature. The first approach, used by Ingram and Neumann (2006) and Poletaev and Robinson (2008), does not assume an *a priori* knowledge of the skills underlying the multitude of task and attribute measures and relies on using a set of factors which are orthogonal to each other to represent skills. The second approach, used in Yamaguchi (2012), assumes that a subset of attributes and tasks measures one skill. There is no clear advantage of using one methodology over the other. However, the interpretation of skills obtained using the second approach may be easier as the resulting factors are constructed from a group of conceptually similar or related variables. In the case of the first approach, the several seemingly unrelated variables may have high factor loadings on the same factor (such as numerical ability and finger dexterity) making clear interpretation of this factor more problematic.

I follow the second approach in performing my analysis of the skills in the Career Handbook. I assume that there are two skills, cognitive and manual, underlying the nine attributes and data, people, and things tasks. I use general learning ability, numerical ability, verbal ability, clerical perception, and the aforementioned task complexity in working with data and people to construct a composite measure of occupational cognitive skills. I use motor coordination, manual dexterity, finger dexterity, form perception, spatial perception and the task complexity in working with machines and tools ("things") to construct the measure of occupational manual skills reflecting hand-eye coordination.

Each occupation appears within the Career Handbook only once. When performing PCA, this will result in all occupations being assigned the same weight. However, in the labour market, occupations are not uniformly distributed (e.g., there are more retail salespersons than computer programmers). Following the literature, I use the native-born workers of the 2001 Canadian Census of Population for appropriate weighting when obtaining factors. Therefore, zero corresponds to the mean of cognitive and manual skills for the Canadian native-born workers. The resulting skill vectors are then matched with the immigrants in LSIC using the NOC and SOC 1991 codes. Details on the factor loadings for both skills can be found in Appendix A.2.2.

1.4.3 LINC/EAL/ELSA Data

To complete the data used in the estimation of the language skill evolution and wage determination equations, I obtain the distance between the immigrant's residence and the nearest ESL course provider. This variable is the third component of X_{i1}^{C}. In the period of interest, the Canadian Federal Government directly financed the Language Instruction for Newcomers (LINC) program in Ontario, Nova Scotia, Prince Edward Island, New Brunswick, Newfoundland, Alberta, and Saskatchewan. Both Manitoba and British Columbia received transfers from the government but managed their respective programs, English as an Additional Language (EAL) and English Language Services for Adults (ELSA), on the provincial level.[17] All these programs were

[17]As the focus of the paper is on English language proficiency, I do not include residents of Quebec in the analysis.

provided free of charge to individuals who were over 18 years of age and had permanent resident (economic or family) or refugee (convention or government claimant) status. As there was no readily available data set including the addresses of ESL providers to match the LSIC, I hand-collected the data (details are provided in Appendix A.1). The resulting data set contains the names and postal codes of 368 ESL providers. I have also made note of programs which provided ESL training exclusively to women. Combining this data with the Postal Code Conversion File (PCCF) and information on the immigrants' postal codes of residence from LSIC, I calculated the distances from each immigrant's residence to the nearest ESL provider.[18]

1.4.4 Estimation Sample and Descriptive Statistics

The data set resulting from combining all of the above sources contains a sample representative of immigrants who have lived in Canada for approximately four years. I impose a set of restrictions on these data to obtain my estimation sample. I use the age reported at the first interview to measure the age at immigration and restrict the sample to contain immigrants who were between 25 and 55 years old at arrival. This ensures that the sample includes individuals who have completed the major part of their human capital investment, excluding the investment in the host-country language skills, in their

[18]These are geodetic "straight-line" distances. While individuals generally do not travel in straight lines, these distances are nonetheless indicative of the travel costs potentially incurred by immigrants.

source-countries. I further trim the top and bottom 1 percent of observations based on their hourly real wages to remove any outliers. I do not restrict the sample to males, in contrast to previous studies. This is backed by the argument that, at least in Canada, female immigrant workers follow an economic assimilation pattern similar to that of men, presented in Adserà and Ferrer (2014). I exclude immigrants who lived in Canada for six or more months at any point prior to immigration, as their economic integration pattern is likely to be very different from the majority of immigrants who have not resided in Canada for any significant amount of time prior to their arrival. Because I focus on acquisition of and returns to English language skills I exclude immigrants with native ability in English, those residing in Quebec and those who cross the linguistic border between Quebec and the rest of Canada.

Table 1.1 provides the descriptive statistics for Waves 1 and 2 of the data. These statistics were obtained using the weights provided for LSIC by Statistics Canada. The "Number of Observations" column shows the total weight for the individuals in the relevant subsample.

There is an increase in the average immigrants' language skill from 0.095 in Wave 1 to 0.234 in Wave 2 along with a decrease in the standard deviation, as more immigrants become proficient in the language. 71.3 percent of immigrants in the sample have a record of ESL hours (including zero hours for no participation). Out of these individuals, 23.8 percent of immigrants report participating in an ESL course between the first and the second interview. The

ESL hours have been converted to full-time equivalent six months with the assumption of 126 hours of ESL per month. Immigrants who participated in ESL courses between the interviews on average attended them for approximately 6.8 full-time equivalent months.

When estimating the language skill gains equation in Section 1.6, I use the distance to the nearest ESL course provider as one of the instrumental variables for the time spent in ESL courses.[19] There are multiple factors that are likely to affect the immigrant's choice of residence, including availability of housing, rent levels, and access to amenities such a stores and public transit. Therefore, I treat the residential location as chosen independently from the existing locations of ESL course providers, which would make it a credible instrument.[20] On average immigrants lived approximately 3.8 kilometers away from the nearest ESL course provider. Furthermore, the standard deviation is 23.9 kilometers, implying a high amount of variation in the data, further supporting the use of the distance to the nearest ESL provider as an instrument for ESL course participation time. I also use the visa status as an instrument for time spent in ESL courses as it may affect the costs of participating in ESL courses. While government-funded ESL courses are provided free of charge to all eligible immigrants, participants are still responsible for daily travel costs

[19] I am grateful to Mikal Skuterud for our informal discussion of the use of distance to LINC centers as a potential instrument for course participation.

[20] If lower language learning ability immigrants were to choose to intentionally settle near ESL providers it would make the estimate of the effect of ESL courses on language gain be biased toward zero.

incurred to reach ESL providers. Refugees are eligible for subsidized public transit tickets which reduces these travel costs for them. Around 5.1 percent of immigrants are refugees and 17 percent arrived in Canada as sponsored family members.

Since zero corresponds to the average occupational skill for the native-born workers, an average immigrant in the sample is employed in an occupation which uses cognitive skills 0.1775 standard deviations below and manual skills 0.1779 standard deviations above the native-born average in Wave 1. Therefore, on average, immigrants are employed in occupations with greater use of manual skills. However, in Wave 2 these numbers change to 0.0256 and 0.2209 standard deviations above the native-born average, respectively, which may indicate the continuing economic assimilation of immigrants including the transfer of their pre-immigration skills. Looking at the gaps between pre-immigration skills and current occupational skills used on the job we can see that the mean gap decreases from 1.1221 to 0.9113 standard deviations for cognitive skills. For manual skills the gap changes from 0.084 to 0.042 standard deviations on average from Wave 1 to Wave 2. This reduction in the skill gaps over time together with the increase in host-country language skill and work experience may indicate that more of the pre-immigration skills transfer over time, which is consistent with the U-shaped occupational mobility of immigrants proposed in the earlier literature. On average, the immigrants' real wage is 13.82 dollars per hour in Wave 1 rising to 16.32 dollars per hour in Wave 2. This real wage growth of 18.1 percent is driven by the increase in language skills, Canadian work experience (which increased from 4.1 to 21 months on average), and cognitive and manual skills used on the job, as becomes evident from the results in Section 1.6.

Table 1.1: Descriptive Statistics for the Estimation Sample

Variable Name	Wave 1 Mean	Wave 1 Std. Dev.	Wave 2 Mean	Wave 2 Std. Dev.	Number of Obs.
Language Skill	0.0947	0.8761	0.2340	0.7833	78545
Record of Hours in an ESL Course	0.7131	–	–	–	78545
Participation in an ESL Course	0.2380	–	–	–	56017
Time in an ESL Course	1.1361	0.9744	–	–	13331
English-proficient HH Members	2.3324	1.1819	3.0645	1.3647	78545
Months Between Interviews	–	–	19.8279	1.2112	78545
Age	35.75	7.15	37.41	7.15	78545
Refugee Status	0.0507	–	0.0507	–	78545
Family Visa Status	0.1700	–	0.1700	–	78545
Distance to the Nearest ESL Provider	3.8403	23.9195	–	–	78545
Real Hourly Wage	13.8163	6.6021	16.3234	7.4713	34577
Cognitive Skill	−0.1775	1.0784	0.0256	1.0833	34577
Manual Skill	0.1779	0.9660	0.2209	0.9994	34577
Cognitive Skill Gap	1.1221	1.1838	0.9113	1.1258	32430
Manual Skill Gap	0.0840	1.1756	0.0424	1.2153	32430
Canadian Work Experience	0.6944	0.3203	3.5005	0.9405	34577
Male	0.6167	–	0.6167	–	34577
Interview Conducted in English	0.5927	–	0.6649	–	78545

Note 1: All variables are weighted using the weights provided with LSIC by Statistics Canada.

Note 2: The number of observations is the total weight reported for the relevant statistic.
Unweighted descriptive statistics cannot be disclosed under the Statistics Canada RDC regulations for LSIC.

1.5 Model Identification and Estimation

The goal of this paper is to estimate the return to investment in language skills, which is driven by the language skill increase due to ESL participation, φ_{esl} from equation (1.4), the wage return to language skills, β_l, and occupational skills, β_s, from equation (1.1), and the language skill-driven transfer of pre-immigration skills, τ_l, from equation (1.3). This section discusses the empirical model and its identification.

First, using the assumption that language does not depreciate, I rewrite (1.4) as a language skill gains equation:

$$\Delta l_{it} = esl_{it-1}\varphi_{esl} + X^L_{it-1}\varphi_x + \alpha^L_i \tag{1.9}$$

Since language skill l_{it} is constructed based on self-reported language proficiency it may be subject to measurement error. Table 1.2 shows that a sizable portion of the sample reports a decline in language proficiency across multiple measures. A decline of English language proficiency while living in an English speaking country is not likely as noted earlier, in Section 1.3.

As discussed by Dustmann and Van Soest (2002) this pattern of reported

Table 1.2: Changes in Reported English Proficiency (Wave 1 to Wave 2)

Reported Change	Speaking	Reading	Writing
"Declined"	16.6%	18.9%	20.0%
"Unchanged"	49.8%	55.7%	53.2%
"Improved"	33.6%	25.4%	26.8%

language change indicates the presence of the measurement error. The measure of language skill I use in this chapter is a weighted average of eight different proficiencies. Following Dustmann and Van Soest (2002), each of the underlying variables, $l_{it}^{k,rep}$, contains an individual time-invariant error $\eta_i^{k,L}$ which reflects that some individuals may always understate their language ability, while others - always overstate it, and idiosyncratic error $\upsilon_{it}^{k,L}$, which reflects interview-specific measurement error. The measure of the overall language skill, l_{it}^{rep}, can then be modeled as

$$l_{it}^{rep} = \sum_{k=1}^{8} \omega_k (l_{it}^k + \eta_i^{k,L} + \upsilon_{it}^{k,L}) = l_{it} + \eta_i^L + \upsilon_{it}^L \tag{1.10}$$

where ω_k is a weight attached to the underlying language proficiency k. Substituting (1.10) into (1.9) we obtain

$$\Delta l_{it}^{rep} = esl_{it-1}\varphi_{esl} + X_{it-1}^L \varphi_x + \alpha_i^L + \varepsilon_{it}^L \tag{1.11}$$

where I rewrite $\upsilon_{it}^L - \upsilon_{it-1}^L$ as ε_{it}^L. I assume that ε_{it}^L is independent from esl_{it-1} and X_{it-1}^L. That is, the measurement error does not affect the estimation of the language evolution equation. However, as evident from solution for the optimal time spent in ESL courses, esl_{it-1} is not independent from α_i^L which is unobserved by the econometrician. To address the endogeneity of esl_{it-1}, I use the distance to the nearest ESL provider and refugee and family visa status as instrumental variables.[21] These variables can be thought of as X_{it}^C in the

[21]Distance to one's ESL provider affects travel costs in terms of time and money. Refugees may receive additional benefits such as public transit tickets to cover travel expenses. Further, they may experience pressure to attend ESL classes from their case-workers.

solution for the optimal ESL attendance. Variables in X_{it}^{C} affect the optimal ESL attendance but are not correlated with the unobserved language learning ability α_i^L. Without the use of instrumental variables, we expect the coefficient for esl_{it-1} to be biased towards zero, since having higher α_i^L corresponds to a lower esl_{it-1}.[22]

Second, suppose that hourly wages are measured with error, that is $w_{it}^{rep} = w_{it}\exp\{\upsilon_{it}^{W}\}$.[23] The log-wage equation which can be obtained from equation (1.2) by introducing measurement error and substituting the scalar s_{it} for the vector of cognitive and manual skills S_{it} is then:

$$\ln w_{it}^{rep} = l_{it}\beta_l + S_{it}\beta_s + xp_{it}\beta_{xp} + \alpha_i^W - \upsilon_{it}^W \tag{1.12}$$

where υ_{it}^W is measurement error independent from l_{it}, S_{it}, and xp_{it}. Further, substituting (1.10) into (1.12) we obtain

$$\ln w_{it}^{rep} = l_{it}^{rep}\beta_l + S_{it}\beta_s + xp_{it}\beta_{xp} + \alpha_i^W - \eta_i^L\beta_l + \varepsilon_{it}^W \tag{1.13}$$

where $\varepsilon_{it}^W = -\upsilon_{it}^W - \upsilon_{it}^L\beta_l$. As l_{it} is not independent from η_i^L and υ_{it}^L and, hence, the unobservable $-\eta_i^L\beta_l + \varepsilon_{it}^W$ of (1.13), using OLS will not yield consistent estimates. S_{it} may be correlated with α_i^W as it depends on both l_{it} (α_i^W and α_i^L may be correlated) and S_{i0}, which in turn likely depends on α_i^W. I use individual fixed effects to address the $\alpha_i^W - \eta_i^L\beta_l$ portion of the unobservables and instrumental variables to address the time-varying portion of the measurement error

[22]Here $\hat{\varphi_{esl}} = \varphi_{esl} + \mathtt{Cov}(esl_{it-1}, \alpha_i^L)/\mathtt{Var}(esl_{it-1})$ and $\mathtt{Cov}(esl_{it-1}, \alpha_i^L) < 0$ according to the model.

[23]Since the record of weekly wages and weekly hours is based on the self-reported values this assumption is rather reasonable.

in language, $-\upsilon_{it}^{L}\beta_l$.[24] Specifically, I use an indicator for whether the interview was conducted in English and the number of the household members that are proficient in English as instruments. While the latter is only one of the variables in X_{it-1}^{L}, it is the only variable which is not correlated with other explanatory variables in (1.13). The direction of bias on β_l is difficult to sign *ex ante*: unobserved ability can cause an *over*estimation of the the effect, if α_i^W and α_i^L are positively correlated, while the measurement error in the language skill - an *under*estimation. Dustmann and Van Soest (2002) find that the returns to language a subject to a significant downward bias.

The final two equations which I need to estimate in order to obtain the returns to time spent in ESL courses are the skill transfer equations for cognitive and manual skills. Taking a log of equation (1.3) and rearranging terms, we obtain

$$\ln s_{i0}^{j} - \ln s_{it}^{j} = \tau_l^{j}(\bar{l} - l_{it}) \qquad , \tag{1.14}$$

where j indexes cognitive and manual skills (i.e., $j \in \{c, m\}$). Since language is measured with error described in equation (1.10), the equation which will be estimated is

$$\ln s_{i0}^{j} - \ln s_{it}^{j} = \tau_l^{j}(\bar{l} - l_{it}^{rep}) + \tau_l^{j}\eta_i^{L} + \tau_l^{j}\upsilon_{it}^{L} \tag{1.15}$$

Since l_{it}^{rep} is correlated with the error term $\tau_l^{j}\eta_i^{L} + \tau_l^{j}\upsilon_{it}^{L}$, estimation of this equation using OLS will yield biased estimates of τ_l^{j} for $j \in \{c, m\}$. As with the

[24] In this case, IV do not have to address whether the immigrant always under- or overstates their language ability (i.e., η_i^L), but only the idiosyncratic shock, υ_{it}^{L}, which is treated as classical measurement error.

wage equation, to address this problem, I use the combination of fixed effects and instrumental variables (using an indicator for whether the interview was conducted in English and the number of the household members that are proficient in English as IV).

1.6 Estimation Results

1.6.1 Language Skill Evolution, Wage, and Skill Transfer Equations

In this section I present and discuss the estimates of parameters in equations (1.11), (1.13), and (1.15).

First, I estimate the language gains equation. In the full specification, X^L_{it-1} is represented by three variables: months between the interviews, the number of other household members who can speak English, and the immigrant's age at $t-1$. Months between the interviews relates to the "passive" learning outside of the ESL classroom due to exposure to English through media and day-to-day interactions. I expect the sign on this variable to be positive. The number of household members able to speak English is another measure of exposure to English. There are two ways in which household members' English ability can affect language acquisition. On one hand, having access to a household member who is proficient in English and is able to assist one with learning should have a positive effect on language acquisition. On the other hand, English-proficient individuals could act as translators for other members of the household reducing the exposure to the host-country language and the effort they may put into learning from daily activities. Therefore, no statement can be made regarding the sign of the coefficient prior to obtaining the estimates. Immigrant's age accounts for the fact that over the years people lose brain plasticity and, hence, language acquisition becomes a slower process. We would expect the sign on the lag of age to be negative.

Table 1.3 presents the estimated coefficients for the language evolution equation (1.11). For the ease of interpretation the ESL participation time has been standardized to six month intervals. The results in columns (1) through (4) present the estimates for the OLS models, while columns (5) through (8) cover the estimates obtained by using the generalized method of moments (GMM) procedure where the distance to the nearest ESL provider and the refugee and sponsored family visa status were used as instrumental variables.[25] For all specifications the signs of the coefficients are as expected: positive for ESL participation, the number of family members who can speak English, months between the interviews, and negative for the lag of age. The coefficients on the time spent in an ESL course estimated by OLS are lower than those estimated with the use of instrumental variables, indicating bias towards

[25] I use GMM as there are three variables instrumenting for the time in ESL courses, so the model is over-identified. Furthermore, GMM estimators are a better choice if the model might have heteroscedastic errors. Baum, Schaffer and Stillman (2003) contains a discussion on the subject. First stage estimates can be found in Appendix A.3.1.

zero predicted in the previous section. I focus the discussion on the results for the specification (8), which is the preferred specification for this paper. The coefficient on the ESL course time is 0.291, meaning that increasing ESL participation time by six months yields a language skill gain of a 0.291 of a language skill standard deviation. An additional month's stay in Canada results in a 0.0158 standard deviations gain in the language skill. We can conclude that attending ESL courses is a more effective way of acquiring language skills than learning from day-to-day activities. Family members who can speak English prove to be beneficial for language skill acquisition. While acquisition slows down with age based on the sign of the coefficient, the effect is small and statistically not different from zero.

Now that I have obtained an unbiased estimate of φ_{esl}, I need to obtain an estimate of wage returns to language skills, β_l. In what follows, I discuss the estimates for the equation (1.13) which is estimated using fixed effects and fixed effects with instrumental variables. I use measures of cognitive and manual skills discussed in Section 1.4.

The first column of Table 1.4 shows a pooled OLS model provided for comparison. The next two columns show the model estimated with fixed effects without correcting for the measurement error in language skill and with the use of instrumental variables to account for the measurement error. All coefficients change noticeably when moving from the OLS to fixed effects estimation, reflecting the endogeneity of occupational skills, which are likely correlated with unobserved ability. As anticipated, the coefficient on the language skill is statistically and economically insignificant in the fixed effects specification. This result is driven by the idiosyncratic portion of the measurement error. I use the number of members of the household who speak English which has been shown to be a significant determinant of language skill and an indicator set to one if the interview was conducted in English as instruments. The latter indicates that the individual was sufficiently comfortable with the interview to be conducted in language other than his or her native one.[26] Comparing the FE and FE-IV specifications we can see that the coefficient changes from -0.001 to 0.1899 and becomes statistically significant when instrumenting for language skill.[27] An 18.99 percent increase in wages for a standard deviation increase in language skill is consistent with the previous literature finding significant wage gains from becoming proficient in the host-country language. Returns to cognitive and manual skills are positive and significant, indicating a 12.79 and 2.1 percent gain in real wages for a standard deviation increase in the corresponding skill. Consistently with earlier work, cognitive skill returns are larger than the manual skill returns. Canadian work experience also has a positive effect on wages.

[26]One of the advantages of LSIC is that the interviews were conducted in a language chosen by the interviewee.

[27]First stage estimates can be found in Appendix A.3.1

Table 1.3: Estimates for the Language Skill Evolution Equation

Language Skill Gains	OLS				IV-GMM			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time in an ESL Course	0.1761^{***}	0.1734^{***}	0.1719^{***}	0.1721^{***}	0.2976^{***}	0.2953^{***}	0.2911^{***}	0.2910^{***}
	(0.0171)	(0.0170)	(0.0169)	(.0169)	(0.0401)	(0.0403)	(0.0403)	(0.0403)
Months Between	–	0.0209^{**}	0.0212^{**}	0.0212^{**}	–	0.0151^{*}	0.0157^{*}	0.0158^{*}
Interviews	–	(0.0089)	(0.0089)	(0.0089)	–	(0.0091)	(0.0092)	(0.0092)
English-proficient HH	–	–	0.0213^{***}	0.0224^{***}	–	–	0.0187^{**}	0.0206^{**}
Members	–	–	(0.0083)	(0.0084)	–	–	(0.0083)	(0.0085)
Lag of Age	–	–	–	-0.0009	–	–	–	-0.0014
	–	–	–	(0.0017)	–	–	–	(0.0017)
N	–	–	–	–	–	–	–	–
R^2	0.0568	0.0586	0.0613	0.0614	0.0298	0.0316	0.0355	0.0357

Note 1: “Time in an ESL Course” is measured in full-time six month equivalents.

Note 2: IV-GMM specifications use the distance to the nearest ESL provider, refugee status, and sponsored family visa status as instruments for the “Time in an ESL Course” variable.

Note 3: Robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note 4: All models are unweighted, N is not displayed at present due to Statistics Canada RDC disclosure process.

Table 1.4: Estimates for the Log Wage Equation

Log of Hourly Real Wage	OLS	FE	IV-FE
Language Skill	0.0762***	−0.0010	0.1899**
	(0.0065)	(0.0140)	(0.0970)
Cognitive Skill	0.2104***	0.1299***	0.1279***
	(0.0061)	(0.0122)	(0.0122)
Manual Skill	0.0574***	0.0228**	0.0210*
	(0.0065)	(0.0113)	(0.0115)
Canadian Job Experience	0.0413***	0.0417***	0.0361***
	(0.0038)	(0.0026)	(0.0038)
Individual Effects	No	Yes	Yes
Instruments for Language	No	No	Yes
N	–	–	–
R^2	0.4324	0.4025	0.3646

Note 1: The IV-FE specification uses the number of household members who speak English and the indicator for whether the interview was conducted in English as instrumental variables.

Note 2: Clustered robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note 3: All models are unweighted, N is not displayed at present due to Statistics Canada RDC disclosure process.

Finally, I estimate the effect that immigrants' English-language skills and Canadian work experience have on the transfer of their pre-immigration skills. Table 1.5 shows the skill transfer equation estimation results for cognitive and manual skills. The first column for each skill shows the results using only the fixed effects estimator, while the second column shows the estimates using both fixed effects and instrumental variables estimation. There is a significant change in the coefficient reflecting the penalty for non-native proficiency for cognitive skills (the coefficient changes from 0.0774 to 0.6526 and its statistical significance increases). The coefficient implies that an individual who is 3 standard deviations below the near-native proficiency in English, for example, can use only 14.12 percent of their pre-immigration cognitive skills in the host-country job, and so on. For manual skills I do not find any evidence of language skills assisting in the skill transfer.[28] This is not a surprising finding if one considers the use of each of the skills on the job. The use of cognitive skills may include, for example, understanding and preparing documents, an activity which necessitates the knowledge of the host-country language. An immigrant who knows how to work with the same types of documents in his native language but has poor command of English will be unable to perform the task at hand. The use of manual skills, conversely, involves physical tasks, the successful completion of which does not require host-country language proficiency.

1.6.2 Returns to Attending ESL Courses

Now that all the key equations have been estimated and I have unbiased estimates of φ_{esl}, β_l, and τ_l^c, it is possible to calculate the impact of time spent in ESL courses on immigrant wages. As I find no evidence of language skills affecting the transfer of manual skills, I calculate the returns to ESL training accounting only for the increase in total return to language skills and the

[28] First stage estimates for both cognitive and manual skills can be found in Appendix A.3.3.

Table 1.5: Estimates for the Skill Transfer Equations

Gap in Log-Skills	Cognitive Skills		Manual Skills	
	FE	IV-FE	FE	IV-FE
Language Gap	0.0774*** (0.0291)	0.6526*** (0.1774)	0.0361 (0.0434)	−0.1211 (0.2082)
Individual Effects	Yes	Yes	Yes	Yes
Instruments for Language	No	Yes	No	Yes
N	–	–	–	–
R^2	0.0335	0.0335	0.0002	0.0002

Note 1: The skill levels have been shifted to be positive in order to take logs, with the standard deviation preserved.

Note 2: The IV-FE specification uses the number of household members who speak English and the indicator for whether the interview was conducted in English as instrumental variables.

Note 3: Robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note 4: All models are unweighted, N is not displayed at present due to Statistics Canada RDC disclosure process.

increase in the total return to cognitive skills. Consider the following

$$\frac{\partial \ln w_{it}}{\partial\, esl_{it-1}} = \frac{\partial \Delta \ln w_{it}}{\partial\, esl_{it-1}} = \frac{\partial \Delta \ln w_{it}}{\partial \Delta l_{it}} \cdot \frac{\partial \Delta l_{it}}{\partial\, esl_{it-1}} = (\beta_l + \tau_l^c\, \beta_s^c\, s_{i0}^c \exp\{-\tau_l^c(\bar{l} - l_{it})\}) \cdot \varphi_{esl} \tag{1.16}$$

That is, $\beta_l \cdot \varphi_{esl}$ is the effect of time spent in an ESL course on wage growth generated by the increase in the total return to the English-language skill. The whole expression in (1.16) describes the total effect which includes the return to skills transferred into the host-country economy due to the increase in the host-country language skill.

Given the estimates in Tables 1.3, 1.4, and 1.5, I calculate that attending an ESL course for six full-time equivalent months results in an 11.65 percent wage increase on average, with the standard deviation of 2.51 percent. The total return to language skills accounts for a 5.53 percent wage increase, while 6.12 percent of the increase in wages is the result of a larger proportion of the immigrants' cognitive skills being transferred into the host-country economy. This effect is heterogeneous in pre-immigration skills and the current language skill. To my knowledge, this is the first estimate of the impact of ESL attendance on immigrants' wage in the literature. By comparison, gaining an additional six months of host-country work experience increases the individuals wages by 3.61 percent. In line with the Chiswick and Miller (2003) finding regarding English proficiency acting as a substitute to Canadian work experience, I argue that ESL course attendance can act as a substitute for acquiring Canadian work experience for new immigrants. It is further worth noting that even immigrants with less cognitive skills greatly benefit from ESL courses through the direct channel that presents higher wage gains for equivalent amount of time spent on acquiring Canadian work experience.

The findings have high policy relevance, since the Canadian immigration policy selects individuals based on high cognitive skills but the outcomes of recent cohorts of immigrants have been declining. This decline has been attributed, in part, to the compositional shifts in language ability. As the present paper shows, a decline in language ability could have a two-fold negative effect on the labour market outcomes of immigrants: directly and through a reduction in the proportion of skills transferred to the Canadian economy. ESL courses are a potential channel through which new immigrants' outcomes can be improved; however, language programs across Canada have recently been subject to significant budget cuts. Immigrants are a group which is likely to be financially constrained and, therefore, have limited participation to privately provided language courses despite the wage gains. This makes government funding of ESL courses particularly important.

1.7 Conclusion

In this paper I postulate a two-period model of immigrants' investment in language skills. Motivated by the recent literature on the transferability of immigrants' pre-immigration skills, I allow the language skill to affect the proportion of pre-immigration skills which enters the wage equation. I estimate the key equations of the model, language skill evolution, wage determination, and skill transfer, using the insights gained from the optimal solution to the theoretical model. I use instrumental variables to overcome the endogeneity inherent to the language evolution equation and fixed effects estimation with instrumental variables to address the endogeneity in the wage and skill transfer equations.

I find that time spent in ESL courses has an economically and statistically significant impact on the growth of immigrants' wages. There are three further contributions of the paper. First, I find that immigrants' host-country language skills play and important role in the transfer of pre-immigration cognitive skills into the economy; this is, however, not true for manual skills. Second, I use principal component analysis (PCA) to construct a continuous measure of English language skill from eight self-reported English-language proficiency measures. Previous studies have relied on dichotomous measures of language proficiency based on self-reported speaking proficiency. Such dichotomous variables would not be suitable for use in my study as I focus on the gradual increase in language skills. The third contribution is the use of Career Handbook data paired with the Canadian 2001 Census of Population in constructing occupational skills using PCA. Previous studies focusing on the returns to occupational skills note the use of US data, such as DOT and O*NET, as a potential caveat for examining the Canadian labour market outcomes if the skills within occupations with the same or similar names (hence, SOC 1991 to NOC code correspondences) differ between Canada and the US (Imai et al., 2016).

Immigrants with higher pre-immigration cognitive skills experience higher wage gains from six months spent in full-time ESL courses due to the transfer of these skills into the Canadian labour market. However, even immigrants with lower pre-immigration cognitive skills receive a 5.53 percent wage increase through the direct effect of higher language skills on wages. This gain is much larger than 3.61 percent wage gain received from six additional months of Canadian work experience.

One of the limitations of the present study is the assumption that the return to a unit of time spent in an ESL course in terms of language skill gains is the same for all individuals. It is possible that ESL instructors may distribute their in-class help according to students' ability, dedicating more resources to students who lag behind. However, it is also likely that students with higher language learning ability α_i^L progress through the course material faster, reaping higher returns per unit of time spent in the course. Unfortunately, estimation of the model in which the time spent in ESL courses is augmented by the unobserved language learning ability is precluded by data availability, as more than three periods of observations on language and ESL courses will be needed for identification. It may, admittedly, be possible to estimate marginal treatment effects (MTE) in a model with dichotomous ESL participation with essential heterogeneity in language skill returns. However, estimates from such framework would be unable to distinguish between ability and training intensity gains. I intend to explore this in future work.

1.8 Bibliography

ADSERÀ, A. AND A. M. FERRER, "The Myth of Immigrant Women as Secondary Workers: Evidence from Canada," *The American Economic Review* 104 (2014), 360–364.

AKRESH, I. R., "US Immigrants Labor Market Adjustment: Additional Human Capital Investment and Earnings Growth," *Demography* 44 (2007), 865–881.

AYDEMIR, A. AND M. SKUTERUD, "Explaining the deteriorating entry earnings of Canada's immigrant cohorts, 1966–2000," *Canadian Journal of Economics / Revue canadienne d'économique* 38 (2005), 641–672.

BAUM, C. F., M. E. SCHAFFER AND S. STILLMAN, "Instrumental variables and GMM: Estimation and Testing," *Stata Journal* 3 (2003), 1–31.

BEENSTOCK, M., "The Acquisition of Language Skills by Immigrants: The Case of Hebrew in Israel," *International Migration* 34 (1996), 3–30.

BEISER, M. AND F. HOU, "Gender Differences in Language Acquisition and Employment Consequences Among Southeast Asian Refugees in Canada," *Canadian Public Policy / Analyse de Politiques* (2000), 311–330.

BERMAN, E., K. LANG AND E. SINIVER, "Language-Skill Complementarity: Returns to Immigrant Language Acquisition," *Labour Economics* 10 (2003), 265–290.

BLEAKLEY, H. AND A. CHIN, "Language Skills and Earnings: Evidence from Childhood Immigrants," *Review of Economics and Statistics* 86 (2004), 481–496.

CHISWICK, B. R., "The Effect of Americanization on the Earnings of Foreign-Born Men," *The Journal of Political Economy* (1978), 897–921.

———, "Speaking, reading, and earnings among low-skilled immigrants," *Journal of Labor Economics* (1991), 149–170.

CHISWICK, B. R., Y. L. LEE AND P. W. MILLER, "Immigrants' language skills: The Australian experience in a longitudinal survey," *International Migration Review* 38 (2004), 611–654.

CHISWICK, B. R. AND P. W. MILLER, "The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada," *Economics of Education Review* 22 (2003), 469–480.

———, "International migration and the economics of language," *IZA Discussion Paper No. 7880* (2014).

CLARKE, A. AND M. SKUTERUD, "Immigrant Skill Selection and Utilization: A Comparative Analysis of Australia, Canada, and the United States," *Working Paper* (2014).

DULEEP, H. O., "The Adjustment of Immigrants in the Labor Market," *Handbook of the Economics of International Migration* (2014), 105.

DULEEP, H. O. AND M. C. REGETS, “Immigrants and human-capital investment,” *The American Economic Review* 89 (1999), 186–191.

DUSTMANN, C., “Speaking Fluency, Writing Fluency and Earnings of Migrants,” *Journal of Population Economics* 7 (1994), 133–156.

DUSTMANN, C. AND F. FABBRI, “Language Proficiency and Labour Market Performance of Immigrants in the UK,” *The Economic Journal* 113 (2003), 695–717.

DUSTMANN, C. AND A. VAN SOEST, “Language Fluency and Earnings: Estimation with Misclassified Language Indicators,” *Review of Economics and Statistics* 83 (2001), 663–674.

———, “Language and the Earnings of Immigrants,” *Industrial and Labor Relations Review* (2002), 473–492.

FERRER, A., D. A. GREEN AND W. C. RIDDELL, “The Effect of Literacy on Immigrant Earnings,” *Journal of Human Resources* 41 (2006), 380–410.

GONZALEZ, A., “The Acquisition and Labor Market Value of Four English Skills: New Evidence from NALS,” *Contemporary Economic Policy* 18 (2000), 259–269.

GONZALEZ, L., “Nonparametric bounds on the returns to language skills,” *Journal of Applied Econometrics* 20 (2005), 771–795.

HAYFRON, J. E., “Language Training, Language Proficiency and Earnings of Immigrants in Norway,” *Applied Economics* 33 (2001), 1971–1979.

HUMAN RESOURCES DEVELOPMENT CANADA, “Career Handbook, Second Edition (Revised),” (2003).

IMAI, S., D. STACEY AND C. WARMAN, “From Engineer to Taxi Driver? Language Proficiency and the Occupational Skills of Immigrants,” *Working Paper* (2016).

INGRAM, B. F. AND G. R. NEUMANN, “The Returns to Skill,” *Labour Economics* 13 (2006), 35–59.

KAIDA, L., “Do Host Country Education and Language Training Help Recent Immigrants Exit Poverty?,” *Social science research* 42 (2013), 726–741.

KOSSOUDJI, S. A., “English language ability and the labor market opportunities of Hispanic and East Asian immigrant men,” *Journal of Labor Economics* (1988), 205–228.

POLETAEV, M. AND C. ROBINSON, “Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000,” *Journal of Labor Economics* 26 (2008), 387–420.

SHEN, A., “ESL Students in B.C. Face Funding Uncertainty,” *The*

Globe and Mail. Available: http://www.theglobeandmail.com/news/british-columbia/esl-students-in-bc-face-funding-uncertainty/article20295740/ [Last accessed: July 9, 2016] (2014).

SKUTERUD, M., "Language Skills in the New Economy and the Deteriorating Labour Market Performance of Canada's Immigrant Workers," *Report Prepared for Human Resources and Skills Development Canada* (2011).

YAMAGUCHI, S., "Tasks and Heterogeneous Human Capital," *Journal of Labor Economics* 30 (2012), 1–53.

Chapter 2

Understanding Traditional School Choice and Student Achievement: Evidence from the United States

2.1 Introduction

The past decade has seen an increased interest in policies aimed to improve the academic performance of schoolchildren. One such policy is offering parents greater school choice. This policy instrument can take on different forms. However, for any school choice policy to be successful, parents have to select schools based on characteristics that have a positive effect on the students' academic achievement. Therefore, understanding how parents select schools for their children in the absence of reforms is essential for implementing school choice policy changes. Residential moves that are motivated by desire to enroll a child at a specific school are referred to as the *traditional school choice*.[1]

Caetano and Macartney (2014) note that surprisingly little is known about the traditional school choice and factors which drive it. Moreover, unexpectedly, the economics literature on the effects of school and residential moves on academic performance is rather limited. My paper contributes to the understanding of the determinants of traditional school choice and its impact on students' academic achievement. I focus on students in public elementary schools in the United States using data from the Kindergarten Cohort of the Early Childhood Longitudinal Study (ECLS-K). Most of research on school choice has been focused on high school students. However, a growing literature on the development of cognitive and noncognitive skills points to the importance of early investments in human capital.[2]

The use of ECLS-K data has several advantages. First and foremost, I use responses to the parent questionnaire to identify the residential moves which occurred in order to place the student at a specific school (Caetano and Macartney (2014) rely on a quasi-experimental design to determine which families exercise the traditional school choice). Moreover, ECLS-K data contains information on the reasons for residential moves aside from changing schools. Second, the data is extremely rich, containing information from parent, teacher, and school administrator questionnaires; furthermore, the data is longitudinal. Hence, I can use the same data set to estimate the production function for

[1]Following Hoxby (2003)

[2]For example, see Cunha and Heckman (2010)

cognitive achievement and the parental utility function which will allow me to examine how school moves affect academic outcomes. As the standardized tests for reading and mathematics were administered as a part of the ECLS-K data collection, the results are comparable across the states and schools. The teacher and administrator questionnaires provide information on a wide variety of school characteristics which can act as determinants of parental school choice overcoming data limitations faced by earlier studies. The majority of previous studies which examined school choice looked at only a few potential school choice determinants such as distance to school and school test score averages.

Approximately thirty-four percent of students in the ECLS-K change schools at least once after kindergarten and up to grade five. Over thirty percent of residential moves are driven by parents exercising traditional school choice. On average, following a residential move, students experience lower gains and larger losses to their performance on standardized tests in reading and mathematics when compared to those students who do not change schools. Surprisingly, students whose move was due to parents exercising traditional school choice do not outperform those students who remained at their old schools and suffer a decline in scholastic achievement in mathematics. This decline is substantial, equalling 9 percent of standard deviation in the standardized gain score. In the current paper I argue that this decline in academic performance is a result of parents choosing schools based on characteristics that do not affect scholastic achievement while paying little attention to school inputs which can improve the students' performance. I obtain parameter estimates for the schools' value-added test score production function and the parental utility as a function of school characteristics. For the former, I employ a panel data regression with measures of parental and teacher effort along with student-specific fixed effects to account for unobserved heterogeneity. For the latter, I estimate a random utility model of school choice using the conditional logit framework proposed by McFadden (1973). I treat each school as a collection of attributes some of which serve as inputs into the test score production function. The remaining attributes describe the school's socio-economic composition and amenities. I find that socio-economic composition rather than the school inputs determined to be important for the cognitive achievement production process are the main contributors to the probability of a school being chosen by parents. Considering that these results extend to households who base their residential moves on access to a specific school, I conclude that this poor decision-making is a result of limited information about schools and the education process.

The paper is organized as follows: section 2.2 discusses the relevant literature, section 2.3 describes the data and stylized facts, section 2.4 presents the theoretical framework, section 2.5 provides empirical findings and discussion, section 2.6 concludes.

2.2 Review of Relevant Literature

My paper draws on three strands of literature: research on school choice, studies of the effect of residential moves and school choice on student achievement, and the literature on the production of cognitive skills and cognitive achievement.

Caetano and Macartney (2014) are among the first to study the role of traditional school choice in residential decisions. They note that despite the importance of understanding traditional school choice for policy purposes, we know surprisingly little about the matter. Caetano and Macartney use the age cutoff for entry into kindergarten to identify residential moves occurring for schooling reasons. Employing data from North Carolina coupled with a dynamic framework for residential move decisions, they find that 20 percent of families exercise traditional school choice. Whites exercise greater amount of school-related residential moves across districts at earlier ages, while blacks tend to move within districts at later ages.

There is a small set of papers that are concerned with estimating parental preferences for school characteristics. Lankford and Wyckoff (1992) develop and estimate a model of school choice between public and private alternatives. Using a binomial logit they find that parents choose schools based on the academic performance of students at that school as well as the socio-economic characteristics such as income of other students. Hastings, Kane and Staiger (2006) use data from the implementation of a district-wide public school choice plan in Mecklenburg County, North Carolina to estimate a mixed-logit discrete choice model of parental preferences for schools characteristics. Their estimation focuses on six characteristics: distance to school, mean standardized test score at the school, the racial composition at the school, previous attendance by the student, an indicator for being the neighbourhood school, and an indicator for the guaranteed transportation to the school. They find that parents greatly value proximity to school. Preferences for the school test score average show a significant degree of heterogeneity and are increasing in student's academic ability and family income. Rothstein (2006) uses the National Education Longitudinal Study (NELS) data combined with SAT data and examines the role of peer quality in high school choice. He finds that school quality measured in this fashion is not highly valued by parents.

In this paper I focus on traditional school choice discussed in Caetano and Macartney (2014) with the advantage of residential moves committed for schooling reasons being identified by parents themselves. When compared to the previous studies by Hastings et al. (2006) and Rothstein (2006), I use a wider variety of school characteristics when estimating preferences.

Understanding how information about schools affects parental school choice decisions can also be revealing about parental preferences. Bast and Walberg (2004) survey the literature regarding parental school choice to argue that parents are better at choosing schools for their children than government experts. They cite evidence of parents being well-informed and placing academic achievement among their top priorities when choosing a school. Briggs et al. (2008) discuss the reasons why the Moving-to-Opportunity (MTO) experiment did not succeed at placing the children of participants in better schools. The authors argue that one of the main reasons for this failure is that the program did not account for the limited resources and and the logic used by the participants to choose schools. The evidence provided indicates that a lot of families participating in MTO faced limitations with respect to the information available to them. Hastings and Weinstein (2008) use a natural experiment and a field experiment to examine the effect of information availability and accessibility on the parental school choice.[3] Information provided to parents as mandated by the No Child Left Behind act induced a significant fraction of parents to choose schools with better academic performance as an alternative for their current school. Moreover, Hastings and Weinstein find that for disadvantaged families simplified information on school academic achievement made this effect stronger. Azmat and Garcia-Montalvo (2012) employ the Learning and Education Achievement in Punjab Schools (LEAPS) data to study information gathering and school choice by parents. The main finding of the paper is that parents who are aware of and have visited schools in their neighbourhoods form assessments of school rankings based on test scores. Parents beliefs about

[3]Starting from 2002 the Charlotte-Mecklenburg Public School District (CMS) provided parents with the option to submit their three top choices based on the information provided in the school choice guide and school-by-school internet search on school performance on the CMS website, but starting from 2004 CMS provided the parents at the No Child Left Behind (NCLB) sanctioned schools with an alphabetically sorted school list with test score results.

school quality have an important effect on school choice. Friesen et al. (2012) estimate the effect of information about school performance on school choice decisions. They develop a model in which parents learn about school quality based on the reported average standardized test scores for the school. The data come from a natural experiment in which the Ministry of Education in British Columbia (BC) in Canada started the provision of information about the test performance in BC schools to the public beginning from the fall of 2000. Using a difference-in-difference approach they find that parents respond to the new information about schools by moving their children to schools with relatively higher perceived quality.

Residential and school decisions can have a significant effect on the academic performance of students. Pribesh and Downey (1999) are among he first to study the the impact of moving on high school student performance. They look at residential moves, school moves occurring without the change in residence, and combined residential and school moves using NELS data which contains data on high school students in 1988 and 1992. They find that combined residential and school moves negatively impact students' mathematics test scores. They argue that two components are responsible for this decline. First, within the sample, there are significant pre-existing differences in family characteristics between the movers and non-movers. Second, residential and school moves cause the destruction of *social capital* which negatively impacts students' achievement.[4] An interesting finding is that school-only moves are not linked to worse performance in math. Cullen, Jacob and Levitt (2006) exploit randomized lotteries which determine high school admissions under *open enrollment* within the Chicago Public School system. While students who win the lotteries attend high schools which have better characteristics, no systematic gains in terms of regular measures of academic achievement such as test scores and graduation rates are observed. The previously discussed study by Rothstein (2006) also finds no positive effect of parental school choice on academic achievement for high school students. His study, however, focuses on district-level choice without identifying the reasons for changes in the districts. Lavy (2010) looks at behavioural and academic outcomes for students affected by the programme that granted them free public school choice in Tel-Aviv, Israel. He finds evidence of sizeable positive effects on cognitive achievement and reduced drop-out rates.

All of the above studies focus on the effects of school choice and residential moves on high school students. However, the growing literature on cognitive and noncognitive skill development summarized in Cunha and Heckman (2010) emphasizes the importance of early investment in these skills. School choice exercised at elementary school level may, therefore, have a significant effect on the child's human capital and I focus my study on children in grades

[4]Pribesh and Downey (1999) do not offer an explanation of the mechanism through which this occurs.

one through five.

When estimating the production function for cognitive achievement, I rely on the vast literature which studies the value added approach to measuring the impact of school inputs on student test scores. Todd and Wolpin (2003) discuss the methodology for modelling and estimating production functions for cognitive achievement. Dewey et al. (2000) survey the literature on school inputs and caution against the use of parental income as one of the inputs into the test score production function. They run a variety of misspecification tests finding commonly used school inputs to have positive statistically significant marginal effects. Coates (2003) and Clotfelter, Ladd and Vigdor (2007) assess the effectiveness of schooling inputs such as instructional time and measures of teacher quality. Additionally, both of these papers discuss the use of panel data for estimating test score production functions using fixed effects models. Harris and Sass (2011) study the effectiveness of teachers using the “value-added” approach measuring teacher output as their students’ standardized test scores. They find that for elementary school teachers there are gains to productivity in their early years of teaching, though there are still marginal positive effects at 10 years of experience. I use these papers for guidance when selecting the school inputs which enter the test score production function.

2.3 Data and Descriptive Statistics

2.3.1 Data

The data come from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K) which was conducted by the US Department of Education. The survey included children from both public and private schools, attending either full-day or part-day kindergarten. The collected information from a nationally representative sample of children, their parents, teachers, and schools within the US. On average, 23 randomly selected children were sampled from each ECLS-K school. In smaller schools, the number sampled may have been lower. The ECLS-K is a longitudinal study that followed the same children from kindergarten through the 8th grade. Information collection was conducted in the fall and the spring of kindergarten (1998-99), the fall and spring of 1st grade (1999-2000), the spring of 3rd grade (2002), the spring of 5th grade (2004), and the spring of 8th grade (2007). Trained evaluators assessed children in their schools and administered surveys to parents over the telephone. Teachers and school administrators were contacted directly at their schools with request to complete questionnaires. The ECLS-K administered tests in mathematics and reading, rather than using school grade reports, which allows comparison of cognitive achievement across students independent of their location in the US.

I focus on children in elementary schools which corresponds to survey waves four (2000, the first-grade year), five (2002, the third-grade year), and six (2004 the fifth-grade year). Furthermore, I limit the sample to public schools. This sample restriction alleviates the issue of tuition being a part of the cost of moving to a private school as well as the potential differences in the test score production technology between public and private schools (for example, see Altonji, Elder and Taber (2005)).

The longitudinal nature of data permits me to observe school moves throughout the elementary school in order to assess the importance of diverse school characteristics and inputs on the school choice decision. Moreover, longitudinal data permits me to estimate the test score production function using fixed effects to account for inherent student ability.

The parental questionnaire contains the question

PIQ.006 *Did you {or {CHILD}'s parents} choose where to live so that {CHILD} could attend {his/her} current school?*

The answers to this question permit me to identify which parents exercise traditional school choice.[5]

The public-use data available to researchers outside of the United States does not contain information on the location of schools beyond the census region. In section 2.5, I discuss the algorithm used to identify the students' school

[5] An important caveat is that a move can be initiated by a reason different than TSC, such as a divorce, but the parent would still put effort into placing their child in the best school which is available to him/her (this school can be worse than the original school the child attended). I thank Todd Stinebrickner and Elizabeth Caucutt for pointing this out.

choice sets using observed moves. Students crossing regional borders considerably affect the size of the school choice sets which poses issues for the model estimation. For the purposes of computational feasibility I exclude these students from the sample. As they comprise only 0.2 percent of the sample, this restriction should not affect my results in a significant way.

2.3.2 Descriptive Statistics

There are three groups of students who are of particular interest to the current study: students who move schools due to parents exercising traditional school choice, students who move schools due to a residential move undertaken by the family, and those students who remain at the schools from which the above moves have occurred. In what follows I will refer to these students as *school movers*, *residential movers*, and *non-movers* respectively. School movers are identified as those students whose parents reported choosing the residence so that the child could attend their current school or those parents who reported that they moved to get their child to a better school. In this section I discuss and compare the academic performance and family background of these students as well as their respective schools. Approximately 34 percent of parents move at least once while their children are in elementary school. Out of these movers, over thirty percent report choosing their residence to place their child at a specific school. I do not observe any significant difference in the proportions of movers based on their performance on first year standardized test scores when compared to the sample median.[6]

Test Scores

Standardized tests in reading and mathematics were administered as a part of the ECLS-K data collection. I use these tests to calculate test gain scores which serve as a measure of change in academic performance from survey wave $t-1$ to survey wave t and have become commonplace in the literature. These scores are available for students in grades three (gain from grade 1 to grade 3) and five (gain from grade 3 to grade 5). In addition, for students who moved between grades one and three I can examine the gain scores at the new school (labeled as "Gr. 5" for both types of movers) . The gain scores reported in tables 2.1, 2.2, and 2.3 have been calculated as $Y_{i,t} - Y_{i,t-1}$ and standardized to have mean zero and standard deviation of one to make them comparable across grades and simplify the interpretation of differences between groups. I use gain scores since I want to examine the impact of moving on academic performance, which is better reflected by improvement or decline in the test scores compared to earlier performance.

For both reading and mathematics gain scores I perform mean comparisons between the students who moved and students who remained in the schools from which the movers originated. In the time period following the move reading gain scores do not exhibit any statistically discernible differences across

[6] I use this as an informal comparison of ability between groups.

the groups. However, residential movers suffer a decline in their reading performance from grade three to grade five following the move prior to the third grade. This difference constitutes 9 percent of the standard deviation (with associated p-value of 0.0133) when compared to non-movers from the same schools. When comparing residential and school movers, there is no statistically significant difference between the respective next period gain scores of -0.082 and -0.044.

Mathematics gain scores follow a different pattern. Surprisingly, school movers suffer a decline of 9.4 percent of a standard deviation with results holding at 1 percent significance, while residential movers experience a gain which, however, is statistically indistinguishable from zero. When the comparison is conducted just between the movers there is a significant difference of 12 percent of the standard deviation in favour of residential movers. In both cases the next period's gain scores are similar across the groups. These results are rather unexpected since one would intuitively assume that the school moves associated with parents exercising school choice should result in the improvement of academic performance. Instead, we observe that non-movers fare best of all, followed by the residential movers. I investigate the causes of these outcomes further in section 2.5.

Table 2.1: Test Gainscores: Did Not Move

Variable	Mean	Std. Dev.	Min	Max	Obs.
Gain score (Read) Gr. 3	0.012	0.999	-3.443	6.333	6961
Gain score (Read) Gr. 5	0.008	0.995	-4.141	4.247	4315
Gain score (Math) Gr. 3	0.013	0.982	-4.12	7.071	7142
Gain score (Math) Gr. 5	-0.002	0.997	-4.145	5.364	4344

Table 2.2: Test Gainscores: Moved for Residence

Variable	Mean	Std. Dev.	Min	Max	Obs.
Gain score (Read) Gr. 3 & 5	0.000	1.035	-3.277	4.432	1641
Gain score (Read) Gr. 5	-0.082	1.023	-4.741	3.416	728
Gain score (Math) Gr. 3 & 5	0.033	1.074	-3.294	7.453	1703
Gain score (Math) Gr. 5	-0.011	1.03	-3.439	2.898	738

Table 2.3: Test Gainscores: Moved for School

Variable	Mean	Std. Dev.	Min	Max	Obs.
Gain score (Read) Gr. 3 & 5	-0.011	0.942	-3.002	2.979	710
Gain score (Read) Gr. 5	-0.044	1.013	-3.821	2.605	342
Gain score (Math) Gr. 3 & 5	-0.091	0.972	-3.515	2.986	732
Gain score (Math) Gr. 5	-0.015	1.014	-2.617	2.887	343

Family Background

Tables 2.4, 2.5, and 2.6 present the descriptive statistics for family characteristics. To understand if there are significant underlying differences in the backgrounds of the groups I perform a set of mean and proportion comparison tests. The comparison is conducted across a set of socio-economic characteristics such as being a single parent, having a below median income, being non-hispanic white or black, family education background and occupational prestige of parents.[7]

Residential movers are 6.62 percent less likely to have at least one college educated parent, but do not differ in occupational prestige or income from non-movers. They differ significantly by racial composition, being 7.35 percent more likely to be black. Residential movers also have a higher, by 2.3 percent, proportion of single parent households. School movers are 6.48 percent more likely to come from single parent households and 5.18 percent more likely to be lower income. However, they do not significantly differ in racial composition, education levels or parental occupational prestige from the non-movers.

When comparing the two types of movers, the difference in the proportion of below median income households is still noticeable (school movers are approximately 3.5 percent more likely to be lower income). Moreover, school movers

[7]I use the national income median of $30,000 for the 2001-2004 time period.

are more likely to be in the single parent households. There is a sizeable difference of 8.5 percent in the chance of being a higher educated household conditional on being a school mover. School movers are 7.5 percent more likely to be white than their residential move counterparts.

There are noticeable differences in the socio-economic and demographic

Table 2.4: Household Characteristics: Did Not Move

Variable	Mean	Std. Dev.	Min	Max	Obs.
Single Parent	0.199	0.399	0	1	7315
Siblings	1.618	1.194	0	11	6447
Parental Education	0.274	0.446	0	1	6513
Mom Occup. Prest.	32.491	20.411	0	77.5	6283
Dad Occup. Prest.	39.61	13.981	0	77.5	5076
Low Income HH	0.326	0.469	0	1	6735
White	0.485	0.5	0	1	7315
Black	0.138	0.345	0	1	7315

Table 2.5: Household Characteristics: Moved for Residence

Variable	Mean	Std. Dev.	Min	Max	Obs.
Single Parent	0.225	0.418	0	1	1800
Siblings	1.641	1.262	0	10	1345
Parental Education	0.2	0.4	0	1	1437
Mom Occup. Prest.	32.181	19.609	0	77.5	1311
Dad Occup. Prest.	39.183	13.338	0	77.5	953
Low Income HH	0.341	0.474	0	1	1696
White	0.404	0.491	0	1	1800
Black	0.217	0.412	0	1	1800

characteristics between the groups. It is, therefore, worthwhile to consider them when estimating the preferences for school attributes.

Table 2.6: Household Characteristics: Moved for School

Variable	Mean	Std. Dev.	Min	Max	Obs.
Single Parent	0.263	0.441	0	1	786
Siblings	1.625	1.151	0	6	786
Parental Education	0.286	0.452	0	1	777
Mom Occup. Prest.	32.51	19.85	0	77.5	770
Dad Occup. Prest.	40.651	14.23	0	77.5	583
Low Income HH	0.376	0.485	0	1	739
White	0.477	0.5	0	1	786
Black	0.141	0.348	0	1	786

School Characteristics

Finally, I examine the differences between the characteristics of schools which residential and school movers choose and those schools from which they leave. These characteristics are present on either school report cards available to parents through the state-level Department of Education websites[8] or could be observed by visiting the school. I use the questionnaire filled out by school administrators along with the student demographic and test score data to construct these variables. The statistics are school-weighted, rather than student weighted, as I am interested in the characteristics of schools irrespective of the number of people who move to them and their overall enrollment. The statistics are presented in tables 2.7 and 2.8 for residential movers and tables 2.9 and 2.10 for school movers.

The index for school facilities is the average of the reported quality of the cafeteria, gym, computer room, playground, library, classrooms, and the school

[8] Such as the Michigan Department of Education `http://www.michigan.gov/mde`

auditorium. The reported qualities range from "1" corresponding to "Never Adequate" facilities to "4" corresponds to "Always Adequate" facilities. Zeroes indicate that the facilities are not present at the school. There is a statistically significant but economically small increase in the quality of the facilities associated with moving to another school for both groups of movers. I use the answers to the questions which describe the state of the neighbourhood: the presence of graffiti, litter, boarded up buildings, and congregations of people. Higher value of the index is associated with a better quality of the neighbourhood. Schools to which students move are located in marginally better neighbourhoods, but the difference is economically not significant.

Both residential and school movers' new schools are larger with respectively 24 and 35 more students on average. Residential movers move to schools with lower proportions of white students but approximately the same proportion of black students. In the case of school movers both proportions decrease. There is a sizeable difference between school and residential moves when it comes to the proportion of college educated households at the new schools. There is a 0.04 increase for the residential movers, but for school movers this number is three times larger. I use the proportion of students who receive a free school lunch as a measure of poor households at the school, which is common in the literature. I do not observe any statistically significant differences in these proportions. It is also worth having a look at whether the school receives Title I funding since schools which do have to have a certain proportion of students from low income households to qualify. The proportion of schools which receive additional funds under Title I is lower by 0.07-0.08 for the schools movers go depending on the group.

Finally, I discuss the differences in the measures of output. There are four such measures: percentage of students performing above the nationally set standards, test score averages within grades, test score averages within school (across elementary school students), and the school gain scores describing the recent performance of the schools students. With the exception of the percentage above national standards, the other variables have been standardized across schools to make the comparison easier. There is not much difference in the measures of being above the national standards in mathematics or reading for both groups: in both cases the economic difference is close to one percent. School level reading test score averages do not differ statistically between old and new schools. However, there is a small difference in the mathematics averages which are higher at the new schools. Similarly, I observe no statistical difference between grade test score averages. Though, the new schools of school movers have slightly (economically insignificant) higher averages. The most interesting observation is that the school-level gain scores for mathematics are negative at new schools and there is a statistically significant difference between old and new schools for the residential group. The gain scores are negative for the new schools of school movers, but the magnitude is smaller in absolute value. This is curious given that on average residential movers do better than school movers in mathematics, yet, this is not what one would expect looking at the school gain scores.

Table 2.7: Residential Move: Old Schools

Variable	Mean	Std. Dev.	Min	Max	Obs.
Facilities	2.609	0.687	0	4	818
School Enrollment	518.856	226.543	4	900	951
Title 1 Funding	0.774	0.418	0	1	819
Prop. White	0.422	0.392	0	1	954
Prop. College HH	0.124	0.181	0	1	954
Neighbourhood	3.721	0.463	1	4	848
Prop. Free Lunch	0.408	0.295	0	0.950	954
School Avg. in Read.	-0.136	0.8	-4.153	2.177	943
School Avg. in Math.	-0.121	0.774	-3.956	2.343	946
Above National Read	58.926	23.325	1	100	473
Above National Math	59.634	22.807	1	100	465
School gain score (Math)	0.036	0.675	-3.322	2.833	229
School gain score (Read)	0.009	0.759	-2.324	3.215	229
Grade Avg. in Read	-0.198	0.853	-4.378	2.413	941
Grade Avg. in Math	-0.186	0.812	-4.175	2.347	944

Table 2.8: Residential Move: New Schools

Variable	Mean	Std. Dev.	Min	Max	Obs.
Facilities	2.724	0.778	0	4	871
School Enrollment	543.759	202.997	2	1000	1136
Title 1 Funding	0.706	0.456	0	1	871
Prop. White	0.38	0.438	0	1	1147
Prop. College HH	0.166	0.313	0	1	1147
Neighbourhood	3.783	0.396	1.25	4	1002
Prop. Free Lunch	0.416	0.284	0	0.950	1143
School Avg. in Read.	-0.169	0.937	-3.498	3.014	1116
School Avg. in Math.	-0.104	0.952	-3.205	2.923	1121
Above National Read	60.716	22.562	2	100	539
Above National Math	60.922	23.054	2	100	536
School gain score (Math)	-0.107	1.397	-6.314	6.691	283
School gain score (Read)	-0.126	1.417	-6.928	6.548	279
Grade Avg. in Read	-0.242	1.004	-3.742	2.689	1102
Grade Avg. in Math	-0.19	1.026	-3.398	2.578	1112

Table 2.9: School Move: Old Schools

Variable	Mean	Std. Dev.	Min	Max	Obs.
Facilities	2.637	0.716	0	4	469
School Enrollment	518.899	223.496	3	900	526
Title 1 Funding	0.767	0.423	0	1	467
Prop. White	0.449	0.388	0	1	528
Prop. College HH	0.14	0.199	0	1	528
Neighbourhood	3.733	0.439	1.667	4	469
Prop. Free Lunch	0.397	0.292	0	0.950	527
School Avg. in Read.	-0.084	0.737	-3.94	2.621	521
School Avg. in Math.	-0.038	0.749	-3.7	2.542	521
Above National Read	61.298	21.969	5	100	295
Above National Math	61.329	22.461	4	100	292
School gain score (Math)	0.067	0.715	-1.422	3.68	139
School gain score (Read)	0.004	0.738	-2.324	4.391	138
Grade Avg. in Read	-0.149	0.786	-4.158	2.635	519
Grade Avg. in Math	-0.097	0.797	-3.91	2.554	519

Table 2.10: School Move: New Schools

Variable	Mean	Std. Dev.	Min	Max	Obs.
Facilities	2.771	0.776	0	4	441
School Enrollment	554.128	201.374	3	1000	564
Title 1 Funding	0.686	0.465	0	1	442
Prop. White	0.41	0.447	0	1	569
Prop. College HH	0.264	0.382	0	1	569
Neighbourhood	3.795	0.408	1.625	4	495
Prop. Free Lunch	0.37	0.272	0	0.950	568
School Avg. in Read.	-0.025	0.973	-3.884	2.526	547
School Avg. in Math.	0.007	0.946	-3.668	2.746	549
Above National Read	61.376	23.609	3	100	282
Above National Math	61.947	23.536	3	100	282
School gain score (Math)	-0.077	1.48	-5.679	6.691	132
School gain score (Read)	-0.1	1.441	-6.928	3.988	130
Grade Avg. in Read	-0.09	1.027	-3.744	2.537	540
Grade Avg. in Math	-0.065	1.046	-3.877	2.707	543

2.4 Theoretical Framework

In this section I briefly discuss the theoretical framework for the empirical section.

2.4.1 School Choice

A household i can choose any school s in the household's choice set S_i. At any time t a school which student i attends possesses a vector of attributes which serve as inputs into cognitive achievement $I_{is,t}$ and attributes which do not enter test score production, $X_{is,t}$. Households can, therefore, infer how much value-added to cognitive achievement the school will provide based on

$I_{is,t}$. I assume that utility is linear in school attributes but may have random elements to reflect that all households do not have the same tastes. Household i's utility from attending school s at time t can then be written as:

$$U_{is,t} = X_{is,t}\theta_X + I_{is,t}\theta_I - c_{is,t} + \varepsilon_{is,t}\,. \tag{2.1}$$

Here θ_X and θ_I are vectors of parameters, $c_{is,t}$ is the cost of moving to school s for household i, and $\varepsilon_{is,t}$ is the taste shock. $X_{is,t}$ can include such attributes as racial and income composition of the school, and the quality of school facilities. θ_I captures parental preferences for academic achievement, as $I_{is,t}$ includes measures of teacher and peer quality, which serve as inputs in student i's test score production. The household chooses school j if and only if

$$U_{ij,t} \geq U_{is,t} \quad \forall\, j \neq s \quad \text{such that} \quad j,\, s \in S_i \tag{2.2}$$

I estimate this discrete choice model using the conditional logistic model in section 2.5.

2.4.2 Test Score Production

Since I am interested in explaining the decline in academic performance following a school move, I need to understand the role different school attributes play in the cognitive achievement production function. In this paper I use standardized test scores to measure cognitive achievement. I use the value-added model common in the literature. Andrabi, Das, Khwaja and Zajonc (2011) find coefficients of about 0.2 to 0.5 for lagged tests scores in the production functions for cognitive achievement (i.e., imperfect persistence). The use of a more restricted gain score specification (with the perfect persistence assumption, under which β below is fixed to be one) leads to biased estimates of the effect of other inputs into test score production. I follow Andrabi et al. (2011) and use the lagged test score specification of the value added model in the present paper. Moreover, I introduce student-level time-invariant effect, α_i to account for unobserved heterogeneity and family characteristics. The production function is then written as

$$Y^k_{is,t} = Y^k_{is,t-1}\beta + P^k_{i,t}\delta + I^k_{is,t}\gamma + \alpha_i + \eta^k_{i,t}\,, \tag{2.3}$$

where $Y^k_{is,t}$ and $Y^k_{is,t-1}$ are the current and lagged test scores, $P^k_{i,t}$ represents parental inputs, $I^k_{is,t}$ represents a vector of those school attributes which participate in the test score production for subject k, α_i is the household-specific effect, and $\eta^k_{i,t}$ is a random shock to the student's performance.

I describe and motivate the inputs in $P_{i,t}$ and $I_{s,t}$ and estimate this production function in section 2.5. Obtained γ coefficients combined with θ_X and θ_I coefficients from the discrete choice estimation enable me to offer an explanation for the declining performance of students who move schools.

2.5 Estimation and Results

In this section I construct school choice sets, use them to estimate the random utility parental school choice model, and estimate the test score production function. I then provide a discussion of the results.

2.5.1 Estimating Test Score Production Function

I use panel data for grades 1, 3, and 5 to estimate value added test score production functions for reading and mathematics. To make the test scores comparable across time and grades and simplify the interpretation of the regression results, I normalize the test scores to have mean zero and standard deviation of one within each time period in the estimation sample. I estimate each function using balanced panel data. In what follows I define the parental and school inputs relevant for the estimation.

Houtenville and Conway (2008) find evidence that the parental effort has a positive effect on student achievement. For my estimation I measure the parental effort as the number of times per week the parent helps the child with homework in the subject. For reading, on average parents help their children 2.6 days out of possible 5. For mathematics this number is similar at 2.45 days out of 5. These results are similar for all mover and non-mover groups. Let

$$P_{i,t} = \{\text{Parental HW Help}_{i,t},\}$$

The schooling input set includes school-level averages. Let

$$\begin{aligned} I_{is,t} = \quad & \{\text{Expected Hours of HW}_{i,st},\ \text{Class Time}_{is,t},\ \text{Prop. of Teachers with MA or PhD}_{is,t}, \\ & \text{Prop. of Teachers with Regular Certification}_{is,t},\ \text{Teacher Experience}_{is,t},\ \text{Class Size}_{is,t}, \\ & \text{Weekly Administered Tests}_{is,t}\ \text{Average Test Scores (Peer Quality)}_{is,t}\} \end{aligned}$$

Both expected hours of homework and hours of instruction (class time) in the subject per week act as proxies for the teachers' effort. Coates (2003) finds instruction time to have a significant positive impact on test scores for both reading and mathematics using data on schools in Illinois. Years of teaching experience, possession of advanced degrees, and possession of regular (as opposed to provisional and temporary) certification are measures of teacher quality commonly included in the test score production functions in the literature. Clotfelter et al. (2007) evaluate these measures using administrative data from North Carolina and find both years of experience and certification to be significant determinants of student achievement. The information on proportions of certified and MA holding teachers, as well as average years of teaching experience within the school can be found on most "school report cards". Parents may treat these proportions as the probability of their child being matched with a certified teacher and/or teacher holding an advanced degree. The effect of class size on student achievement has been the topic of many research papers. Meghir and Rivkin (2011) survey the literature and provide a conceptual model for the effect of class size. They summarize findings from six papers ranging from no effect to positive effect of smaller class sizes on student achievement. I also add the number of weekly administered tests (local/state standardized tests and teacher-made tests), as writing standardized tests often can make the students more comfortable with the format and approach to answering questions.

As school test score averages for reading and mathematics are easily observable by parents via the “school report cards”, I use them as a measure of peer quality at the school. As I am not able to match the school data to the Common Core of Data, I calculate the averages using the information for grades 1, 3, and 5.

Tables 2.11 and 2.12 present the summary statistics for old and new schools of residential movers. Tables 2.13 and 2.14 present these statistics for school movers. There are no significant differences in the average years of teacher’s experience or proportions of certified teachers between these groups of schools. However, the proportion of teachers with advanced degrees is 0.12 lower in the new schools for school movers and 0.15 lower for residential movers. For both subjects assigned homework hours go up with the move, but the increases are not large. Class times for reading are approximately half an hour lower at the schools which residential and school movers choose. For mathematics the difference is statistically equal to zero for school movers and residential movers experience a small 0.1 hour increase. There is a statistically significant difference between the number of tests administered per week at the new and old schools; however, economically this difference is negligible. Average class size goes up by one student for both groups of movers.

The test score production function for each subject is written as:

Table 2.11: Residential Move: Inputs at Old Schools

Variable	Mean	Std. Dev.	Min	Max	Obs.
Avg. Teacher Exp.	14.14	7.273	1	35	776
Prop. Teachers with MA+	0.457	0.377	0	1	916
Prop. Cert. Teachers	0.891	0.22	0	1	773
Avg. HW Hours	1.465	0.455	0	2.5	779
Avg. Class Size	20.56	3.575	12	35	777
Tests Weekly	0.943	0.819	0	6	954
Avg. Read HW	1.663	0.481	0	2.5	778
Avg. Math HW	1.265	0.528	0	2.5	778
Read. Class Time	6.560	1.024	0.125	7.5	765
Math Class Time	4.497	1.108	0.125	7.5	762

Table 2.12: Residential Move: Inputs at New Schools

Variable	Mean	Std. Dev.	Min	Max	Obs.
Avg. Teacher Exp.	14.101	8.9	1	35	777
Prop. Teachers with MA+	0.303	0.4	0	1	1072
Prop. Cert. Teachers	0.903	0.255	0	1	773
Avg. HW Hours	1.752	0.513	0	2.5	770
Avg. Class Size	21.724	4.583	10	36	765
Tests Weekly	0.812	0.853	0	4.5	1147
Avg. Read HW	1.853	0.542	0	2.5	767
Avg. Math HW	1.621	0.6	0	2.5	719
Read. Class Time	5.993	1.438	0.375	7.5	764
Math Class Time	4.606	1.389	0.125	7.5	730

Table 2.13: School Move: Inputs at Old Schools

Variable	Mean	Std. Dev.	Min	Max	Obs.
Avg. Teacher Exp.	13.915	7.001	1	35	433
Prop. Teachers with MA+	0.465	0.372	0	1	509
Prop. Cert. Teachers	0.907	0.193	0	1	433
Avg. HW Hours	1.49	0.452	0	2.5	434
Avg. Class Size	20.689	3.376	11	31.6	433
Tests Weekly	0.945	0.745	0	3.375	528
Avg. Read HW	1.681	0.465	0	2.5	434
Avg. Math HW	1.295	0.517	0	2.5	432
Read. Class Time	6.523	1.024	0.375	7.5	431
Math Class Time	4.520	1.132	0.125	7.5	431

Table 2.14: School Move: Inputs at New Schools

Variable	Mean	Std. Dev.	Min	Max	Obs.
Avg. Teacher Exp.	13.873	8.141	1	35	397
Prop. Teachers with MA+	0.344	0.403	0	1	517
Prop. Cert. Teachers	0.914	0.236	0	1	394
Avg. HW Hours	1.747	0.502	0.415	2.5	393
Avg. Class Size	21.939	4.401	10	36	389
Tests Weekly	0.830	0.868	0	6	569
Avg. Read HW	1.855	0.513	0.415	2.5	390
Avg. Math HW	1.597	0.602	0	2.5	363
Read. Class Time	5.907	1.407	0.875	7.5	386
Math Class Time	4.565	1.317	0.125	7.5	375

$$
\begin{aligned}
\text{Test Score}_{i,t} &= \beta \text{ Test Score}_{i,t-1} \\
&+ \delta \text{ Parental HW Help}_{i,t} \\
&+ \gamma_1 \text{ Expected Hours of HW}_{is,t} + \gamma_2 \text{ Class Time}_{is,t} \\
&+ \gamma_3 \text{ Prop. of Teachers with MA or PhD}_{is,t} \\
&+ \gamma_4 \text{ Prop. of Teachers with Regular Certification}_{is,t} \\
&+ \gamma_5 \text{ Teacher Experience}_{is,t} + \gamma_6 \text{ Class Size}_{is,t} \\
&+ \gamma_7 \text{ Weekly Administered Tests}_{is,t} \\
&+ \gamma_8 \text{ Average Test Scores}_{is,t} + \alpha_i + \eta_{i,t} \qquad (2.4)
\end{aligned}
$$

where $\eta_{i,t}$ is assumed to be identically and independently distributed with mean zero and variance σ_η^2 and represents random shocks to the student's performance. α_i captures time-invariant characteristics specific to the student and provides a control for unobserved heterogeneity. The parameters of particular interest for the present paper are the estimates of γ's, as they reflect the importance of the school-level inputs for grade production.

Since lagged test scores and parental inputs are likely not independent from unobserved student ability, α_i, it would be impossible to consistently estimate the parameters of equation (2.3) using OLS. To address this, I take first differences to eliminate α_i from the equation, obtaining

$$\Delta Y^k_{is,t} = \Delta Y^k_{is,t-1}\beta + \Delta P^k_{i,t}\delta + \Delta I^k_{s,t}\gamma + \Delta\eta^k_{i,t} \,. \qquad (2.5)$$

However, using first differences (FD) estimation is still insufficient to obtain consistent estimates as $\Delta Y^k_{is,t-1} = Y^k_{is,t-1} - Y^k_{is,t-2}$ is, by construction, not independent from $\Delta\eta^k_{i,t} = \eta^k_{i,t} - \eta^k_{i,t-1}$ ($Y^k_{is,t-1}$ depends on $\eta^k_{i,t-1}$). I use the second lag of the test score, $Y^k_{is,t-2}$, as an instrumental variable for $\Delta Y^k_{is,t-1}$ when estimating equation (2.5).[9]

Table 2.15 contains the results for both the FD and IV estimation. For both reading and mathematics going from first differences estimated by OLS to the first differences with the use of an instrumental variables corrects the bias in the coefficient for the lagged score. The F-statistic for the first stage for both reading and mathematics is large, indicating that $Y_{i,t-2}$ acts as a strong instrument for $\Delta Y_{i,t-1}$. In line with Andrabi et al. (2011), I find that lagged test scores do not have a unit coefficient associated with them and only approximately a fifth of the previous test score gets "carried over". This can reflect that material covered in preceding grades is only partially built upon or that there is a significant depreciation of previously accumulated knowledge.

I find that the coefficient on the parental homework help is significant only for reading. One additional time the parent helps his/her child with reading homework per week yields slightly over one percent increase in the standardized test score.[10] The effect for mathematics is statistically indistinguishable

[9]This approach is a form of an Arellano-Bond estimator, as discussed in Arellano and Bond (1991). Here, I have only one second lag of the dependent variable to use as an instrument, due to the data set length.

[10]ECLS-K did not provide measures of the amount of time spent on parental help, beyond the number of times help was provided per week.

Table 2.15: Estimates of the Test Score Production Functions

Test Score	Reading		Math	
	FD	IV	FD	IV
Lagged Test Score	-0.2106^{***}	0.2002^{***}	-0.2240^{***}	0.1887^{***}
	(0.0113)	(0.0379)	(0.0116)	(0.0436)
Parental HW Help	0.0056^{*}	0.0116^{***}	0.0034	0.0052
	(0.033)	(0.0038)	(0.0032)	(0.0035)
Avg. Teacher Exp.	0.0020^{***}	0.0017^{*}	-0.0008	-0.0008
	(0.0009)	(0.0011)	(0.0008)	(0.0009)
Prop. Teachers with MA+	-0.0260	-0.0527^{***}	0.0384^{**}	0.0336
	(0.0194)	(0.0217)	(0.0188)	(0.0214)
Prop. Cert. Teachers	0.0433	0.0578	0.0496	0.0152
	(0.0318)	(0.0363)	(0.0318)	(0.0354)
Avg. Assigned HW	0.0357^{***}	0.0276^{*}	-0.0041	-0.0168
	(0.0135)	(0.0156)	(0.0107)	(0.0121)
Avg. Class Time	0.0063	0.0126^{***}	0.0212^{***}	0.0252^{***}
	(0.0044)	(0.0050)	(0.0048)	(0.0055)
Avg. Class Size	-0.0006	0.0003	-0.0036^{***}	-0.0047^{***}
	(0.0014)	(0.0016)	(0.0014)	(0.0016)
Tests Weekly	0.0048	-0.0112	0.0003	-0.0115
	(0.0095)	(0.0109)	(0.0095)	(0.0109)
School Avg. in Subject	0.2183^{***}	0.1680^{***}	0.2189^{***}	0.1867^{***}
	(0.0237)	(0.0267)	(0.0284)	(0.0318)
Observations	5,906	5,906	5,473	5,473
First Stage F-stat	-	53.46	-	42.90

Robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

from zero.

For reading an additional year of experience yields a small economically insignificant increase in the test score (it is close to 0.2 percent). The coefficient on average teacher experience for mathematics is not estimated precisely. I find the coefficients on the proportion of teachers with Master's or Doctoral degrees or regular certification to be statistically indistinguishable from zero, which is similar to the findings of the earlier literature.

Teacher effort has a statistically significant effect on the test scores. For reading this manifests in the hours of homework with an additional hour increasing the score by approximately 0.03 standard deviations. Additional hour spent on reading per week increases test scores by 1.3 percent of a standard deviation. For mathematics the coefficient on in-class instruction is significant and is 0.0252 standard deviations of the mathematics test score. The amount of assigned homework does not affect cognitive achievement in mathematics. The frequency of standardized tests seems not to have no effect on the test scores.

For mathematics average class size carries a negative impact of 0.0047 standard deviations for an additional class member but for reading the effect is not precisely estimated. For both reading and mathematics the coefficients on proxy variables for peer quality at the school are positive and statistically significant. It is 0.1680 standard deviations for reading and 0.1867 standard deviations for mathematics. In general these estimates are consistent with those found in the literature. I conclude that school peer quality, average class size, and teacher effort measured as class time spent and assigned homework act as important inputs into the test score production process.[11]

2.5.2 School Choice Sets

The public-use version of ECLS-K does not include any locational identification beyond the four census regions in the US.[12] To obtain the choice sets for individuals in my sample I use the observed moves from and into each school. I apply the following algorithm to construct the choice sets[13]:

1. For each school observe students who moved out of the school and record the school identifier of new schools.

2. For each school observe students who moved into the school and record the school identifier of their previous schools.

3. Choice sets that have shared elements are merged together into one larger choice set.

[11]The present paper focuses on finding out which school inputs matter for test score production. I, therefore, pool students from all public school independently of whether they move schools or stay. However, in future research I would like to estimate the impact of the school moves motivated by different reasons on cognitive achievement of students. Estimating such effects would require controlling for selection in the "mover" groups.

[12]The restricted-use data is available exclusively to researchers within the United States

[13]The process is similar to building the adjacency matrix for a connected graph with the exception of setting all non-zero elements of the matrix to one

4. Parents are then assigned a choice set for the school their child is enrolled at prior to the move.

For example, suppose I observe two students move out of school 1 and go to schools 2 and 4. I also observe a student move from school 5 to school 4. So, on first iteration, the choice sets are $S_{i1} = \{1, 2, 4\}$ and $S_{i4} = \{1, 4, 5\}$ for any household i at these schools. After the implementation of step (3) in the algorithm the choice sets become $S_{i1} = \{1, 2, 4, 5\}$ and $S_{i4} = \{1, 2, 4, 5\}$. Note that the school choice sets are time invariant.

I compute the choices for each school within the sample and then assign these choice sets to each individual at the school. As the estimation of conditional logit requires that the individual has more than one alternative, I am forced to exclude all student-school combinations with singleton choice sets. If schools with singleton choice sets are located in a way which makes moving in and out of them too costly, removing these schools will not impact the estimated parameters. However, if these schools are high quality schools which causes me to observe no moves out of the school paired with no nearby schools observed in the ECLS-K which would cause me to observe no moves into the school. There are indeed differences between these singleton choice set schools and the rest of schools which may signal that they are of higher quality. This may bias my estimates of parental preferences.[14] Using all schools within a census region

[14]The direction of the bias in preferences would be difficult to predict without knowing more about these singleton schools.

as alternatives for each individual may potentially address this, but it is highly computationally costly. Obtaining access to the restricted version of ECLS-K and combining it with the Common Core of Data using geographic information may be the only feasible solution to the problem.

When using my algorithm the maximum choice set sizes are 21 schools for the US Midwest, 51 for the Northeast, 51 for the Southern US, and 38 for the Western US. The resulting number of schools within the sample is 2635.

2.5.3 The Determinants of Parental School Choice

Following equation 2.2 the condition for choosing school j can be written as

$$X_{ij,t}\theta_X + I_{ij,t}\theta_I - c_{ij,t} + \varepsilon_{ij,t} \;\geq\; X_{is,t}\theta_X + I_{is,t}\theta_I - c_{is,t} + \varepsilon_{is,t}\,. \tag{2.6}$$

We can rewrite this as

$$(X_{ij,t} - X_{is,t})\theta_X + (I_{ij,t} - I_{is,t})\theta_I \;\geq\; (\varepsilon_{is,t} - c_{is,t}) - (\varepsilon_{ij,t} - c_{ij,t})\,. \tag{2.7}$$

Assuming that $(\varepsilon_{is,t} - c_{is,t})\,\forall\, i,\ s,\ t$ are identically and independently distributed with an extreme value distribution, we can obtain a conditional logit functional form for the probability of household i choosing school s at time t

$$\mathrm{Pr}_{ij,t} = \frac{\exp\{X_{ij,t}\theta_X + I_{ij,t}\theta_I\}}{\sum_{s\in S_i}\exp\{X_{is,t}\theta_X + I_{is,t}\theta_I\}} = \frac{1}{\sum_{s\in S_i}\exp\{(Z_{ij,t} - Z_{is,t})\theta\}}\,. \tag{2.8}$$

This is the conditional logit model proposed by McFadden (1973), sometimes referred to as McFadden's choice model. Long (2004) uses the conditional logit approach to examine college choice behaviour. She argues that the conditional logit is more suitable to analyze school choice behaviour as it focuses on the attributes of alternatives available to the individual rather than individual-specific characteristics. Moreover, unlike the multivariate logit used in prior analysis, conditional logit does not require the restrictive aggregation of the characteristics of alternatives.[15] Hoffman and Duncan (1988) provide a comparison of the multinomial logit and conditional logit and discuss their use in social science research. They point out that the set up for conditional logit has the advantage of easily adjusting the choice sets available to each individual within the data.

The data are organized as triplet-wise combinations of household i with school s and time t. Under the assumption that $(\varepsilon_{is,t} - c_{is,t}) \forall s \in S_i$ are independent across t, I treat combinations of i, t as separate individuals within the data. I use alternative-wise deletion of observations in cases when they are missing relevant information. As I do not observe school closures and openings, this permits me to remove schools with information missing at time t without dropping the individual from the data.

I include a set of variables which reflect the socio-economic conditions at the school and the facilities quality and the variables which participate in the test score production. The first set of variables includes the index for facilities, the neighbourhood quality, the proportion of white students and students with college-educated parents, and a variable describing the income distribution at

[15] Such as Manski, Charles and David A. Wise (1983). *College Choice in America*. Harvard University Press, Cambridge

the school. The second set of variables represents the school-level inputs into the test score production function from section 2.5.1 which were found to play a role in the production process.

For the estimation I focus on the school choices that were made after grade one. This follows from the assumption that parents know the test score production technology for elementary schools, learning about it while their children are enrolled in grade one. I first estimate the model for all movers, then separately for school movers and, finally, residential movers. While I am unable to observe moving costs, I add an indicator for changing schools to account for a fixed cost of moving. The results are presented in Table 2.16.

When looking at the set of all movers it is evident that the socio-economic characteristics of the school play an important role in the school choice process. Households are more likely to move to schools that have a higher proportion of college-educated parents and less likely to move to schools with higher proportions of lower income households. People also display preference for larger schools, though this effect is not economically large. In addition to the preference for schools with higher-educated parents, school movers have a preference for schools with larger proportions of white students. These results are consistent with the findings of Lankford and Wyckoff (1992) for private-public school choice.

A rather surprising outcome is that the inputs into test score function are not estimated precisely for school movers. From what we can see based on the sample of all movers, both reading and mathematics class times have small negative coefficients significant at five percent. This could indicate that parents do not take class times into consideration. As residential movers do not necessarily take schools into account, it is not surprising that I do not obtain precise estimates for school attributes with the exception of the neighbourhood index. The negative coefficient attached to this variable is unexpected, though it is possible that the index does reflect the quality of the neighbourhoods they actually move to.

The coefficients for peer quality measures, the school test score averages in mathematics and reading, are estimated precisely only for the sample of all movers. The coefficient is positive for the reading average but negative for the mathematics score, indicating that all else equal, parents are less likely to choose schools with better math performance. For school movers the signs on the statistically insignificant coefficients follow the pattern similar to that of the overall mover sample.

In the next section I discuss the implication of these findings for the effect of school choice on academic achievement.

Table 2.16: Estimates of the Parental Utility Function

Selected School	School Movers		Residential Movers		All Movers	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Facilities	-0.0841	(0.1052)	-0.0767	(0.0677)	-0.0297	(0.0493)
Enrollment	0.0017^{***}	(0.0004)	0.0002	(0.0002)	0.0009^{***}	(0.0002)
Prop. White	0.3811^{*}	(0.2129)	-0.0742	(0.1516)	0.1539	(0.1013)
Prop. College HH	0.5111^{**}	(0.2418)	0.0903	(0.1801)	0.5978^{***}	(0.1146)
Neighbourhood	-0.3112	(0.1984)	-0.3272^{**}	(0.1459)	-0.3779^{***}	(0.1016)
Prop. Low Income HH	-0.2553	(0.2275)	-0.2214	(0.1543)	-0.3455^{***}	(0.1092)
Avg. Teacher Exp.	-0.0124	(0.0081)	-0.0007	(0.0052)	-0.0075^{**}	(0.0036)
Prop. Teachers with MA+	0.0042	(0.1997)	0.1177	(0.1291)	0.1376	(0.0863)
Read. Class Time	-0.0586	(0.0559)	-0.0788^{**}	(0.0381)	-0.0417^{*}	(0.0236)
Math Class Time	0.0523	(0.0615)	0.0258	(0.0411)	-0.0630^{**}	(0.0278)
Avg. Read HW	0.0236	(0.1704)	0.2021^{*}	(0.1081)	0.0560	(0.0681)
Avg. Math HW	-0.1130	(0.1493)	-0.0420	(0.0952)	-0.0140	(0.0622)
School Avg. in Read.	0.1198	(0.1422)	0.0398	(0.0899)	0.1397^{**}	(0.0650)
School Avg. in Math.	-0.0823	(0.1443)	-0.1363	(0.0921)	-0.2134^{***}	(0.0654)
Cost	18.0247^{***}	(0.1006)	18.5724^{***}	(0.0634)	19.1757^{***}	(0.0467)
Observations	2,837		7,386		14,493	
N of i, t pairs	457		984		2109	
Log-likelihood	-600		-1499		-3010	

Robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.5.4 Discussion

In section 2.5.1 I have established that peer quality, class time spent on mathematics and reading, and the amount of homework for reading act as important components of the test score production process. Class size is found to have a negative impact on mathematics performance. At the same time parental school choice seems to be primarily driven by the values of school characteristics that do not participate in the test score production. The low gain scores observed in section 2.3.2 may be a consequence of this. While on average we observe improvements in school attributes that serve as inputs into the test score production as seen in section 2.5.1 these increases are small. It is likely that these increases are insufficient to overcome the negative impacts of moving noted in Pribesh and Downey (1999).

There are several explanations that can be offered concerning lack of evidence that parental school choice is driven by the school characteristics which are important academically. First, it can be suggested that parents do not place a significant weight on the achievement of their children. This, however, contradicts the surveys cited in Bast and Walberg (2004). Another explanation comes from the availability of information standpoint. It can be difficult for parents to observe the homework requirements and instruction time for the school prior to moving. As Briggs et al. (2008) point out, parents are often faced with limited information regarding the schools they end up choosing. However, the larger concern arises from the observation that school averages in mathematics and reading that signal peer quality at the school do not serve as important determinants of parental school choice when it comes to school movers. Even in the sample of all movers, I observe that parents are less likely to choose a school with a higher mathematics average.

The explanation which can address the above findings is that parents are faced with limited information not only regarding the characteristics of schools they choose to move to but also lack information about the relative importance of these characteristics in the educational process. This results in the parental school choice that does not lead to improvement in academic performance and even has detrimental effect on performance in mathematics. While this explains the poor performance of school movers compared to non-movers, the causes of discrepancy between residential movers and school movers are less clear and will warrant further investigation. It is possible that "purely residential moves" actually lead to children being placed in better neighbourhoods with schools which are better on a dimension unobserved in my data.

2.6 Conclusion

According to Caetano and Macartney (2014) very little is known about residential moves driven by desire to enroll one's child at a specific school (referred to as *traditional school choice*). Using data on public elementary school students from the ECLS-K, I analyze the academic performance of children who move due to their parents exercising school choice and those whose moves are purely residential. Descriptive statistics reveal a surprising observation that school movers suffer a decline in their mathematics performance while residential movers and non-movers do not.

To examine the potential reasons for this disparity in academic performance, I obtain the estimates of parental preferences for school attributes as well as the test score production function to analyze the relative importance of these attributes for academic achievement. I find that parents who move for school reasons place most value on school characteristics which do not enter test score production. While one can argue that parents value the socio-economic properties of the school and neighbourhood more than the academic success of their children, it is very likely that parents have limited information regarding school attributes and further, may have limited information on what attributes add value to the academic outcomes. Paired with the cost that the move has for the child, such school choice may result in worsened academic performance.

The results of the present paper should be treated with caution. The way in which the school choice sets are constructed may potentially bias the results. Obtaining more detailed data which includes geographic information for the attended schools would allow me to test the sensitivity of results to the method which is used to define school choice sets. In future work, I also intend to further examine the reasons for the discrepancy in mathematics performance between residential and school movers.

2.7 Bibliography

ALTONJI, J. G., T. E. ELDER AND C. R. TABER, “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools,” *Journal of Political Economy* 113 (2005), 151–184.

ANDRABI, T., J. DAS, A. I. KHWAJA AND T. ZAJONC, “Do Value-Added Estimates Add Value? Accounting for Learning Dynamics,” *American Economic Journal: Applied Economics* 3 (2011), 29–54.

ARELLANO, M. AND S. BOND, “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *The Review of Economic Studies* 58 (1991), 277–297.

AZMAT, G. AND J. GARCIA-MONTALVO, “The Role of Awareness, Information Gathering and Processing in School Choice,” *Working Paper* (2012).

BAST, J. L. AND H. J. WALBERG, “Can Parents Choose the Best Schools for Their Children?,” *Economics of Education Review* 23 (2004), 431–440.

BRIGGS, X. D. S., K. S. FERRYMAN, S. J. POPKIN AND M. RENDON, “Why Did Moving to Opportunity Experiment Not Get Young People into Better Schools?,” *Housing Policy Debate* 19 (2008), 53–91.

CAETANO, G. AND H. MACARTNEY, "Quasi-Experimental Evidence of School Choice through Residential Sorting," *Working Paper* (2014).

CLOTFELTER, C. T., H. F. LADD AND J. L. VIGDOR, "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects," *Economics of Education Review* 26 (2007), 673–682.

COATES, D., "Education Production Functions Using Instructional Time as an Input," *Education Economics* 11 (2003), 273–292.

CULLEN, J. B., B. A. JACOB AND S. LEVITT, "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica* 74 (2006), 1191–1230.

CUNHA, F. AND J. J. HECKMAN, "Investing in Our Young People," *NBER Working Paper No. 16201* (2010).

DEWEY, J., T. A. HUSTED AND L. W. KENNY, "The Ineffectiveness of School Inputs: a Product of Misspecification?," *Economics of Education Review* 19 (2000), 27–45.

FRIESEN, J., M. JAVDANI, J. SMITH AND S. WOODCOCK, "How Do School 'Report Cards' Affect School Choice Decisions?," *Canadian Journal of Economics/Revue canadienne d'économique* 45 (2012), 784–807.

HARRIS, D. N. AND T. R. SASS, "Teacher Training, Teacher Quality and Student Achievement," *Journal of Public Economics* 95 (2011), 798–812.

HASTINGS, J. S., T. J. KANE AND D. O. STAIGER, "Parental Preferences and School Competition: Evidence from a Public School Choice Program," *NBER Working Paper No. 11805* (2006).

HASTINGS, J. S. AND J. M. WEINSTEIN, "Information, School Choice, and Academic Achievement: Evidence from Two Experiments," *The Quarterly journal of economics* 123 (2008), 1373–1414.

HOUTENVILLE, A. J. AND K. S. CONWAY, "Parental effort, school resources, and student achievement," *Journal of Human resources* 43 (2008), 437–453.

HOXBY, C. M., "School Choice and School Productivity. Could School Choice Be a Tide That Lifts All Boats?," in *The Economics of School Choice* (University of Chicago Press, 2003), 287–342.

LANKFORD, H. AND J. WYCKOFF, "Primary and Secondary Choice Among Public and Religious Alternatives," *Economics of Education review* 11 (1992), 317–337.

LAVY, V., "Effects of free choice among public schools," *The Review of Economic Studies* 77 (2010), 1164–1191.

LONG, B. T., "How Have College Decisions Changed Over Time? An Application of the Conditional Logistic Choice Model," *Journal of Econometrics* 121 (2004), 271–296.

MANSKI, C. F. AND D. A. WISE, *College Choice in America* (Harvard University Press, 1983).

MCFADDEN, D., “Conditional Logit Analysis of Qualitative Choice Behaviour,” *Frontiers in Econometrics* (1973), 105–142.

MEGHIR, C. AND S. RIVKIN, “Econometric methods for research in education,” *Handbook of the Economics of Education* 3 (2011), 1–87.

PRIBESH, S. AND D. B. DOWNEY, “Why Are Residential and School Moves Associated with Poor School Performance?,” *Demography* 36 (1999), 521–534.

ROTHSTEIN, J. M., “Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions,” *American Economic Review* 96 (2006), 1333–1350.

TODD, P. E. AND K. I. WOLPIN, “On the Specification and Estimation of the Production Function for Cognitive Achievement,” *The Economic Journal* 113 (2003).

Chapter 3

Crime, Apprehension and Clearance Rates: Panel Data Evidence from Canadian Provinces

3.1 Introduction

Becker's (1968) seminal theory of crime hypothesizes that criminals rationally evaluate the benefits of crime against the probability of being caught (apprehension) and the severity of punishment. Early empirical studies often used police-reported clearance rates as a measure of the probability of apprehension.[1] As noted by Chalfin and McCrary (2013), however, most recent econometric studies have focused on the effects of the number of police officers

[1]See, for example, Carr-Hill and Stern (1973), Ehrlich (1973), Thaler (1977), and Wolpin (1978).

on crime. From a policy standpoint, using the number of police officers is appealing because policy-makers cannot choose the probabilities of apprehension directly, but they can choose the size of the police force. Given that labour is an input into the apprehension production function, understanding the effect of police on crime, even in a reduced form, is valuable.

From a researcher's standpoint, the number of police officers is easy to observe. Further, if changes to the size of a jurisdiction's police force are exogenous and not an artefact of trends in crime rates or unobserved shocks, then OLS estimates will yield unbiased and consistent impacts that can be used in policy evaluation. A statistically significant correlation between more police officers and less crime also yields some very clear options for policy-making.

However, even in such ideal circumstances for empirical research, a correlation between more police officers and less crime does not offer insight into how deterrence is achieved. Presumably, the probability of apprehension is the "output" produced by the police force and the number of police officers is the labour input for this production function. However, an increase in the number of police officers may result in reduced crime simply because of higher visibility rather than an increased probability of apprehension. Furthermore, as discussed in the data section below, the relationship between the number of police officers and clearance rates is murky at best. As such, the exact channel through which the number of police officers affects crime rates is not obvious.

We attempt to complement the existing studies on police and crime by focusing on the effects of clearance rates on crime using panel data for Canadian provinces (1986 to 2005) while including the number of police officers. In the majority of cases, a police-reported incident of crime is said to be "cleared" if an individual associated with the specific criminal act is apprehended by the police. There is a reasonable chance that the deterrence effects associated with an enhanced probability of apprehension will be more precisely captured through clearance rates than the per capita number of police officers or arrests. Further, the importance of clearance rates has been acknowledged by criminologists for some time.

The primary objective of our research is to empirically evaluate the relationship between changes in the probability of apprehension – as captured through clearance rates – and corresponding trends in violent and property crime rates while controlling for the number of police officers. Our empirical strategy is motivated by an extension of the theoretical model developed by Polinsky and Shavell (2000), which allows us to link spending on police services to corresponding changes in clearance rates and crime and serves as a foundation for constructing plausible instruments. The model also offers some insight on the relative benefits of focusing on clearance rates as opposed to the number of per capita police officers.

There are very few studies based on panel data that have investigated the effects of apprehension on crime rates in Canada.[2] From a general perspective, a study of Canadian trends should be of interest to policy-makers in the US, given similarities in movements in crime rates over time. Specifically, Canada experienced the same dramatic decline in crime during the 1990s that also occurred in the US and that has continued to baffle academics and policymakers.[3] Figure 3.1a shows that per capita violent crime rates are higher in Canada and that violent crime fell in both countries from the early 1990s onward. Figure 3.1b suggests a closer correspondence in property crime rates between the two countries, with a similar persistent drop from the early 1990s. Using Canadian data also allows a rather clean identification of the effects of the probability of apprehension through clearance rates because legislative penalties for violent crime (and most property crimes) are at the federal level and thus can be accounted for through the use of year-specific dummies. On the other hand, there are rather significant and complex state-specific differences in penalties in the US. The presence of unobserved state-varying and time-specific determinants of crime, such as the well-documented crack cocaine epidemic from the mid-1980s to the early 1990s, makes it difficult to ensure unbiased coefficient estimates of clearance rates based on US data. In the absence of proper controls, empirical estimates based on US data may then be confounded if variation in

[2]Early empirical work on apprehension and Canadian crime is based primarily on time-series data. Examples are Avio (1973), Avio and Clark (1976), Avio and Clark (1978) and Avio (1979).

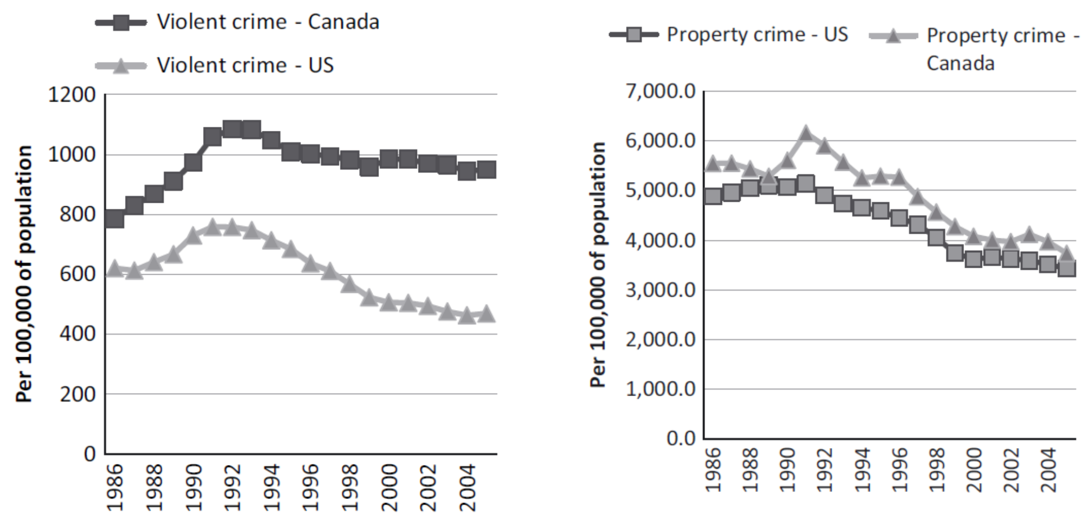
[3]See Levitt (2004) for further details.

clearance rates coincides with amendments to penalties implemented by state legislatures or unobserved factors such as the crack cocaine epidemic.[4]

OLS estimates yield statistically significant elasticities of clearance rates, ranging from -0.2 to -0.4 for violent crimes and from -0.5 to -0.6 for property crimes. These estimates are robust to the use of a wide array of controls, province and year fixed effects and province-specific linear trends. Comparable results are obtained from generalized least squares (GLS), first difference, generalized method of moments (GMM) and instrumental variables (IV) regressions. In contrast, coefficient estimates of the per capita number of police officers are not always significant or possess the wrong sign. We think that these findings reflect the importance of police force productivity in terms of solving crimes and apprehending criminals linked to specific crimes. However, we note that our instruments are weakly correlated with clearance rates. Therefore, appropriate caution should be used in interpreting the results of this study.

The remainder of the paper is organized as follows. Section 3.2 contains a review of the relevant literature. Section 3.3 describes the data and trends in key variables. The theoretical model is presented in section 3.4. Our empirical model is outlined in section 3.5. Empirical estimates are detailed in section 3.6. Section 3.7 concludes with a summary of our main findings and associated

[4]Cheung and Erickson (1997) suggest that crack cocaine use in Canada was quite insignificant relative to corresponding trends in the US. Fryer et al. (2005) provide a detailed account of the crack cocaine epidemic in the US along with associated costs.



(a) Violent Crime

(b) Property Crime

Figure 3.1: A Comparison of Crime in Canada and the US

Source: Canadian data are from CANSIM tables 252-0013 and 252-0051. US data are from the Federal Bureau of Investigation website (`www2.fbi.gov/ucr/05cius/data/table_01.html`).

policy implications.

3.2 Previous Literature

As noted, most empirical research has focused on the effects of the number of police officers on crime.[5] Chalfin and McCrary (2013) offer a detailed review

[5]Recent studies employ varying strategies to address the simultaneity bias of coefficient estimates of police on crime, which occurs when the size of the police force increases as a response to a corresponding upward trend in local crime rates. Evans and Owens (2007) study increases in the number of police officers generated by grants from the Community Oriented Police Service (COPS) program. Di Tella and Schargrodsky (2004), Klick and Tabarrok (2005) and Draca et al. (2011) identify the impacts of police on crime based on sudden exogenous terror events that require enhanced police presence. Levitt (1997) analyzes data on 59 large US

of recent studies.We focus on papers that study the effects of arrest or clearance rates. In this respect, the empirical literature on the effects of apprehension is much thinner, and very few of these studies have accounted for potential simultaneity bias in coefficient estimates of arrest rates with respect to crime through the use of instrumental variables or structural models. Key results and measurement of arrest rates from previous studies are detailed in table 3.1. For the sake of brevity and direct relevance, we restrict our discussion to studies based on US data and that have relied on panel data across jurisdictions and over time in order to evaluate the impacts of arrest or clearance

cities with directly elected mayors, 1970 to 1992. His instrumental variables are constructed from mayoral and gubernatorial elections. Levitt's estimates suggest that an increase in the size of the police force reduces violent crime but does not significantly impact property crime. However, McCrary (2002) finds that Levitt's IV estimates are much less precise once some specific programming errors are corrected. In a rejoinder, Levitt (2002) uses a different dataset and obtains results that are comparable to his original 1997 AER paper. Chalfin and McCrary (2013) note that simultaneity between crime rates and police numbers is much weaker than assumed in previous literature, at least at the municipal level (due to institutional constraints). They correct measurement error by using data from both Uniform Crime Reports (UCR) and Annual Survey of Government (ASG) for 242 US cities with populations over 50,000 during the period from 1960 to 2010. The authors use the ASG measure of police as an IV for UCR data and models using UCR data as IV for ASG measures. Like Levitt (1997), their results suggest an array of elasticities across different categories, with violent crime rates being more responsive than property crime rates to changes in police force size.

rates.[6,7]

Early empirical work on crime and deterrence focused on the effects of clearance rates as a measure of the probability of apprehension. Thaler (1977) uses 1972 census-tract-level data from Rochester, New York, to estimate the effect of deterrence on neighbourhood crime. He emphasizes the importance of measuring the probability of apprehension through an "arrest clearance rate." On page 41, he specifically states "[f]or this measure a crime is only considered

[6]There are, of course, studies that have not relied on panel data. Lochner (2007) uses the National Longitudinal Survey of Youth 1997 Cohort and the National Youth Survey to examine which factors affect an individual's perceived probability of arrest. His results suggest that a 10% increase in the perceived probability of arrest for auto theft reduces auto theft by 7% and major theft by close to 4%. In terms of panel-based Canadian studies, Sen (2007) employs clearance rates as a control, but the emphasis of the study is on the effects of abortion and fertility on crime. Similarly, the focus of Joyce (2009) is on the effects of abortion on US crime. He replicates the results from Donohue and Levitt (2008) and finds that higher arrest rates have a significant and negative impact on some types of violent crime.

[7]A number of panel-based studies focus on the effects of the existence of a death penalty on murder rates. Zimmerman (2004, replicated in 2009) uses state-level panel data from 1978 to 1997. Dezhbakhsh et al. (2003) employ a panel of county-level data covering the years from 1977 to 1996. The same data are used by Shepherd (2005). Shepherd (2004) relies on panel data at the state level and with monthly observations. Durlauf et al. (2010) and Durlauf et al. (2012) employ similar data to these studies but rely on structural econometric models. Another strand of literature focuses on imprisonment rates. Spelman (2005) uses Texas county-level data to examine the effect of prison population on crime rates. Similar to Corman and Mocan (2005), Spelman finds evidence that public order arrests reduce property crime (counties with zero-tolerance and community-policing policies substantially decreased their crime rates). Ehrlich (1973), Levitt (1996) and Marvel and Moody (1994, 1996, 1998) are other important references on crime and imprisonment.

cleared if the criminal was arrested specifically for that crime." Carr-Hill and Stern (1973) also rely on clearance rates in their study of crime in police districts in England and Wales in 1961 and 1966, and they are quite clear on the emphasis that should be placed on clearance rates relative to the number of per capita police officers. In particular, they state (pp. 289–290):

> Deterrence theories indicate that this offence rate should depend on the proportion of crimes 'cleared-up' (or clear-up rate), if this reflects perceived probabilities of apprehension. Such theories might also focus on the number of policemen per capita and a measure of the equipment available to each officer.

Craig (1987) uses 1972 data from Baltimore and obtains a -0.57 elasticity of crime with respect to actual clearance rates (generated from 3 SLS). Wolpin (1978) obtains comparable results (with respect to clearance rates) based on time-series data from 1955 to 1971 for England, Japan and California.

However, studies from the 1990s focus on the relationship between police reported crimes and per capita arrests (relative to either the number of reported crimes or population). Cornwell and Trumbull (1994) exploit variation across counties in North Carolina and obtain OLS elasticities of arrest rates (arrests/crimes) with respect to FBI Index crimes ranging from -0.35 and -0.68. On the other hand, 2SLS estimates with fixed effects are statistically insignificant. Lott and Mustard (1997) employ county-level data between 1977 and 1992 and focus on the effects of legislation pertaining to the right to carry concealed handguns. Their reduced form estimates suggest that a higher probability of arrest is linked with lower crime rates. 2SLS estimates of arrest rates (measured by arrests divided by crimes) instrumented by lagged crime rates also result in negative and significant coefficient estimates. However, in their analysis of the Mustard-Lott dataset, Black and Nagin (1998) are unable to replicate the 2SLS findings and question their credibility. Dezhbakhsh and Rubin (1998) note that Mustard and Lott emphasize OLS findings, which do not correct for the endogeneity of the arrest rate. In contrast, their research suggests coefficient estimates of arrest rates that are less statistically precise and smaller in magnitude than those suggested by Lott and Mustard.[8]

Levitt (1998) exploits data across 59 cities from 1970 to 1992 and studies the effects of arrest rates on the following crimes: murder and non-negligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny and motor vehicle theft. His results suggest elasticities between -0.03 and -0.319.[9] However, he cautions on his inability to control for potential endogeneity bias and concludes that, especially for property crime, deterrence is the most likely factor behind the observed negative relationship between crime and arrest rates, as opposed to incapacitation and measurement error.

Shepherd (2002) uses panel data based on all 58 California counties from

[8] In a much-cited paper, Duggan (2001) uses county- and state-level data over time and finds that the recent decline in gun ownership explains about one-third of the decline in gun homicides (relative to non-gun homicides). His results suggest that a 10% increase in gun ownership translates into a 1.42% increase in the homicide rate the following year.

[9] Calculated as arrests divided by the number of crimes.

1983 to 1996 to explore the deterrent effects of that state's two- and three strike legislation. Her 2SLS estimates reveal that arrest rates (number of arrests/crimes) have a consistently negative impact on all categories of violent and property crime. Gould et al. (2002) focus on the effects of labour market conditions on crime rates, employing aggregated data at the county level between 1979 and 1997, but do include arrest rates in some specifications. In most cases, an increase in the arrest rate is significantly correlated with reductions in different types of crime. Like Levitt (1998), they also specifically acknowledge the difficulty of finding instruments for arrest rates. Mustard (2003) employs county-level data from New York, Oklahoma, Oregon and Washington from 1977 to 1992 and finds that sentence length has little effect on crime. However, conviction rates have a statistically significant negative effect on crime rates, with coefficients on arrest rates ranging from -0.0016 to -0.012, implying a 0.16% to 1.2% drop in crime rates in response to a 1% increase in the corresponding arrest rates.

Corman and Mocan (2005) analyze the effect of economic conditions and deterrence measures on crime as well and verify the validity of the "broken windows" hypothesis, which suggests that reduced tolerance towards minor misdemeanours leads to an overall reduction in crime. They employ monthly time-series data on crime levels (for murder, assault, robbery, burglary, motor vehicle theft, grand larceny and rape) in New York City from 1974 to 1999. Besides obtaining evidence in support of the "broken-windows" hypothesis, the authors find that the size of the police force has an effect only on auto theft and grand larceny, while the number of arrests has a statistically significant impact across all types of crime. The authors note that the use of monthly data reduces the possibility of simultaneity bias of coefficient estimates of the number of police officers, as it usually takes six months to hire more police officers.

In summary, we attempt to complement the existing literature in the following ways. First, despite early studies that clearly emphasized the importance of clearance rates as a measure of the probability of apprehension, we have not been able to locate any recent papers that econometrically investigate the effects of clearance rates on crime based on data across jurisdictions and over time.[10] We think that using clearance rates is important for reasons that are more fully discussed in the next section. Second, employing Canadian data is interesting given the similarities to trends in US violent and property crime observed over the sample period. Third, we assess the sensitivity of our estimates with instrumental variables. As discussed, this is a challenge that has been noted by previous studies, and very few have actually attempted to instrument arrest or clearance rates.[11]

---

[10]Mastrobuoni (2013) uses detailed micro-level data on robberies and deployment of two police forces in the city of Milan, but focuses on the effects of local police presence on the likelihood of clearing cases.

[11]Numerous studies (Ehrlich 1973; Thaler 1977; Mathur 1978; Craig 1987) from the 1970s and 1980s did use either 2SLS or 3SLS in order to correct for simultaneity bias. However, it is difficult to assess the success of their efforts given the absence of F-test statistics and first-stage regression results.

Table 3.1: Crime Elasticities with Respect to Arrests

	All	Property Crimes				Violent Crimes					Measurement
		All	Burglary	Auto Theft	Larceny	All	Robbery	Murder	Rape	Assault	
Thaler (1977)[a]		−0.0017 (2.24) −0.00064 (0.165)									Police proportion of solved crimes; arrests over number of crimes
Mahur (1978)[b]			0.055 (0.310) −0.256 (−2.36)	−0.541 (−2.72) −0.505 (−3.06)	0.626 (1.73) 0.486 (2.18)	−3.07 (−2.03) −1.58 (−2.54)	−1.09 (−1.07) −0.094 (−0.217)	−3.21 (−1.34) −1.10 (−1.28)	−0.909 (−1.56) −0.91 (−1.83)		Admissions to prison for the offence divided by reported crimes
Craig (1987)[c]		−0.57									Clearances as a proportion of crime
Cornwell and Trumbull (1994)[d]	-0.455 (0.618)										Arrests to offences ratio
Gould et al. (2002)[f]	-0.002 (0.002)	-0.01 (0.002)	-0.01 (0.003)	-0.01 (0.001)	-0.01 (0.001)	-0.004 (0.0003)	-0.006 (0.0005)	-0.002 (0.0002)	-0.004 (0.001)	-0.003 (0.0003)	Arrests to offences ratio
Levitt (1998)[e]			-0.272 (0.036)	-0.087 (0.028)	-0.204 (0.030)		-0.339 (0.053)	-0.071 (0.072)	-0.119 (0.030)	-0.201 (0.047)	Arrests divided by crimes
Mustard (2003)[g]			-0.0123 (0.00267)	-0.0052 (0.00108)	-0.0072 (0.00178)		-0.0016 (0.00036)	-0.0035 (0.00025)	-0.0026 (0.00091)	-0.0019 (0.00049)	Arrests divided by offences
Mustard (2003)[h]			-0.0102 (0.00304)	0.0003 (0.0008)	-0.0046 (0.00166)		-0.0035 (0.00076)	0.0000 (0.00034)	-0.0031 (0.00075)	-0.0038 (0.00049)	Lagged arrests over lagged offences
Corman and Mocan (2005)[i]			-0.32 -0.27	-0.51 -0.50	-0.14 -0.10		-0.57 -0.59	-0.40 -0.39	-0.32 -0.30	-0.20 -0.24	Number of arrests
Spelman (2005)[j]		-0.1206 (0.0504)				-0.2376 (0.1057)					Arrests per 1000 people (UCR dta)
Agan (2011)									-0.066 (0.024)		Arrests divided by incidents
Garett and Ott (2011)[k]			-0.009	-0.002	-0.175		-0.070	-0.557	-0.037	-0.006	Number of arrests (UCR data)

Notes: [a]t-statistics are in parentheses. First row reports the result for the police clearance rate. Third row reports the coefficient for the arrests over the number of crimes. [b]t-statistics in parentheses. The first row of elasticities is for 1960, the second is for 1970. [c]Urban type 1 crimes are the focus of the study. [d]From a log-linear specification. [e]First difference estimates with lag of arrest rates included. [f]Marginal effects reported instead of elasticities. [g]Semi-elasticities from the regression of the natural log of crime rates on arrest rates. [h]Semi-elasticities from the regression of the natural log of crime rates on lagged arrest rates. [i]Second row estimates use the average year-to-year growth of arrests in NYC (reported in C&M (2005) Table 3). [j]These elasticities are with respect to prison populations by metropolitan area. [k]Pooled city elasticities are reported. Only robbery is significant at 5%.

3.3 Trends in Crime Rates

Most of our data were downloaded from CANSIM, Statistics Canada's database of socio-economic variables that are publicly available. Table 3.2 details the table numbers, sources and summary statistics for each variable employed in this study. The key variables are the number of police-reported incidents of different types of crimes and the number of incidents cleared for each of these categories. An incident is cleared when a suspect linked to the crime has been identified by the police. Accordingly, the empirical measure of the probability of apprehension that we employ is the number of incidents cleared divided by the number of police-reported incidents.[12] We view the total proportion of incidents cleared as the relevant measure of apprehension because a suspect needs to be identified, irrespective of whether an incident is cleared by charge or otherwise.

[12]An incident can be cleared by charge or otherwise. A suspect needs to be arrested in order for an incident to be cleared by charge. There are many reasons why an incident may be cleared otherwise. Examples include the death of a suspect or the dropping of a charge to a less serious offence. We are indebted to Peter Carrington for some very insightful discussions. Please, see Carrington and Schulenberg (2008) for further discussion on clearance rates.

Table 3.2: Data Sources and Summary Statistics

Variable	CANSIM Table	Mean	Standard Deviation	Min.	Max.
Median income of welfare recipients	2020404	44,111	4,918.2	37,000	58,300
Minimum wage	*(See Note)*	5.51	1.02	3.65	8.00
Average transfers to poorest quintile	2020301	7,738.5	994.59	5,300.0	10,300
Per capita number of new immigrants	510011	428.69	354.54	60.45	1385.7
Incarceration rate per 100,000 adults	2510005	94.77	33.72	48.00	183.00
Proportion of males aged 1524	510001	0.08	0.01	0.07	0.10
Police officers per 100,000 of population	2540002	174.94	17.995	136.80	209.20
Provincial population	510001	0.29248E+07	0.34000E+07	0.12841E+06	0.12565E+08
Property crimes per 100,000 of population	2520013	4749.7	1,642.0	2,342.4	9,007.8
Clearance rate for property crimes	2520013	0.26	0.06	0.13	0.42
Employment rates 15 years and over	510001	89.89	3.75	79.94	96.07
Violent crimes per 100,000 of population	2520013	1,040.7	316.19	478.33	2,059.4
Clearance rate for violent crimes	2520013	0.71	0.062	0.52	0.89
Homicides per 100,000 of population	2520013	1.9967	0.97674	0 4.33	
Clearance rate for homicides	2520013	0.85	0.29	0	2.53
Attempted murders per 100,000 of population	2520013	83.69	102.52	0	408.00
Clearance rate for attempted murder	2520013	0.82	0.26	0	2.00
Sexual assaults per 100,000 of population	2520013	107.78	40.07	28.39	227.40
Clearance rate for sexual assaults	2520013	0.66	0.15	0.41	1.34
Physical assaults per 100,000 of population	2520013	908.66	274.83	374.53	1753.2
Clearance rate for physical assaults	2520013	0.75	0.07	0.53	0.89
Robberies per 100,000 of population	2520013	78.46	53.18	4.87	196.59
Clearance rate for robberies	2520013	0.41	0.15	0.24	1.41
Motor vehicle thefts per 100,000 of population	2520013	427.21	266.93	82.860	1,359.5
Clearance rate for motor vehicle thefts	2520013	0.22	0.099	0.047	0.52
Breaking & entering per 100,000 of population	2520013	1,157.4	421.98	515.39	2,097.0
Clearance rate for breaking & entering	2520013	0.19740	0.062216	0.077497	0.54948
Real police expenditures per capita of population	2540002	155.03	34.643	80.826	235.67
Real police expenditures per 100,000 of police	2540002	87,931	13,776	57,981	126370

Note: Minimum wage data were extracted from the Government of Canada minimum wage database, available at `srv116.services.gc.ca/dimt-wid/sm-mw/menu.aspx?lang=eng`.

An alternative strategy is to use the arrest rate, which is the number of arrests per 100,000 of population. As discussed above, at some point, most empirical studies began to use such arrest rates – or the number of arrests divided by the number of crimes – as measures of the probability of apprehension. However, per capita arrest rates do not clearly yield a probability of apprehension because states or provinces with more crime are mechanically more likely to have higher arrests per 100,000 of population. Further, unlike clearance data, the number of arrests per capita of population is not linked to specific police-reported crime, and therefore, in our opinion, is a weaker empirical measure of the probability of apprehension than the clearance rate.[13,14]

Statistics Canada collects data on police-reported incidents and the number of crimes cleared through the Uniform Crime Reporting (UCR) Survey, which was established in 1962. The scope of these data is comprehensive; they include all Criminal Code offences and other federal statutes that have been reported to all federal, provincial and municipal police services in Canada and that have been substantiated through investigation by these services. As noted on the Statistics Canada website, "[c]overage of the UCR aggregate data reflects virtually 100% of the total caseload for all police services in Canada." It is important to note that the number of incidents is based upon severity, and

[13]However, an arrest rate defined as the number of arrests divided by number of crimes obviously does have a direct connection to crime trends.

[14]A relevant concern is the calculation of clearance rates when crimes committed during a specific year are cleared during subsequent years. Our understanding from discussions with criminologists is that roughly 95% of all crimes are solved within the calendar year.

therefore, the most serious offence.[15] Applying this concept to clearance rates means that, for example, the clearance of a homicide, robbery or breaking and entering receives a higher weight than the clearance of less serious offences such as minor theft, mischief and disturbing the peace.[16]

It is useful to note some differences between US and Canadian crime trends. As discussed, figure 3.1a demonstrates that per capita violent crime rates are much higher in Canada than in the United States. However, it is important to note that violent crime is not defined similarly across both countries. For example, sexual assault in the US requires forcible intercourse by a male against a female. In comparison, the same offence in Canada does not require sexual penetration and is not gender-specific. That is one reason why Canadian crime rates seem higher in figure 3.1a.[17] Further, there are considerable differences in the distribution of crimes by offence. As shown in figures 3.2 and 3.3, murder rates in the US were roughly three times higher than in Canada during the 2000s and robbery rates were one and a half times higher in the US during the same period. On the other hand, although not directly comparable, forcible rape rates in the US (figure 3.4) were much lower than sexual assault rates in Canada. As discussed, this is because of the relatively broad

[15] "Police-reported crime statistics in Canada, 2011," by Shannon Brennan, available at `statcan.gc.ca/pub/85-002-x/2012001/article/11692-eng.htm`.

[16] See table 9-11, Police personnel in municipal police services – Yukon, 2011, in "Police resources in Canada" (Statistic Canada catalogue no. 85-225-X), available at `statcan.gc.ca/pub/85-225-x/2011000/t031-eng.htm`.

[17] For further details please see Gannon (2001).

definition of sexual assault in Canada. Finally, figures 3.5a and 3.5b offer scatterplots of violent and property crime rates against corresponding clearance rates. The graphs depict a visible negative relationship between crime and clearance rates, with a steeper slope for property crime.

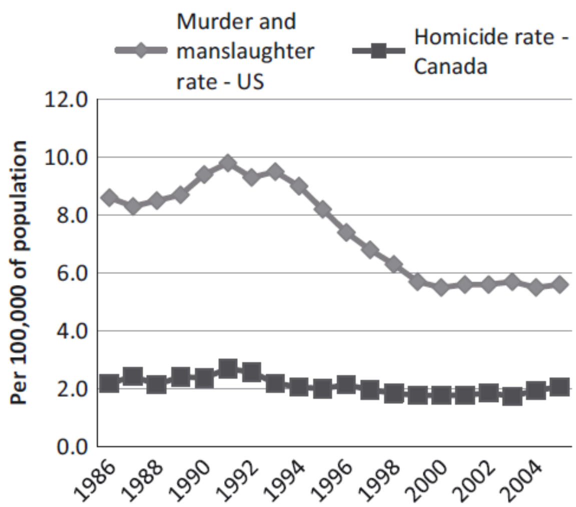


Figure 3.2: A Comparison of Murder Rates in Canada and the US

Source: Canadian data are from CANSIM tables 252-0013 and 252-0051. US data are from the Federal Bureau of Investigation website (`www2.fbi.gov/ucr/05cius/data/table_01.html`).

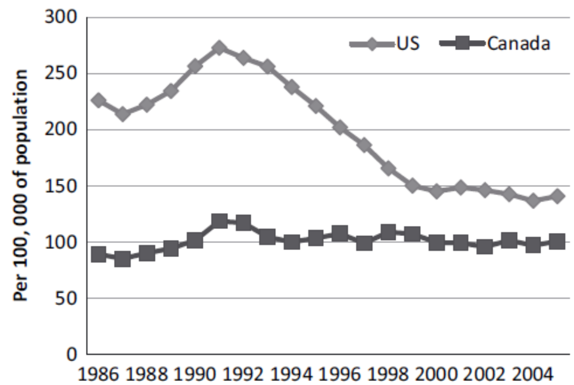


Figure 3.3: A Comparison of Robbery Rates in Canada and the US

Source: Canadian data are from CANSIM tables 252-0013 and 252-0051. US data are from the Federal Bureau of Investigation website (`www2.fbi.gov/ucr/05cius/data/table_01.html`).

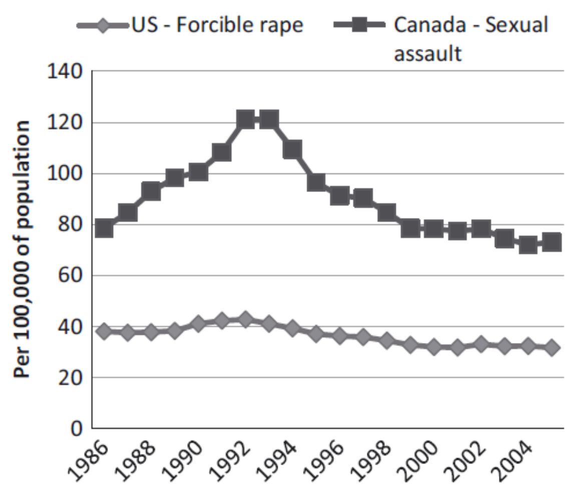
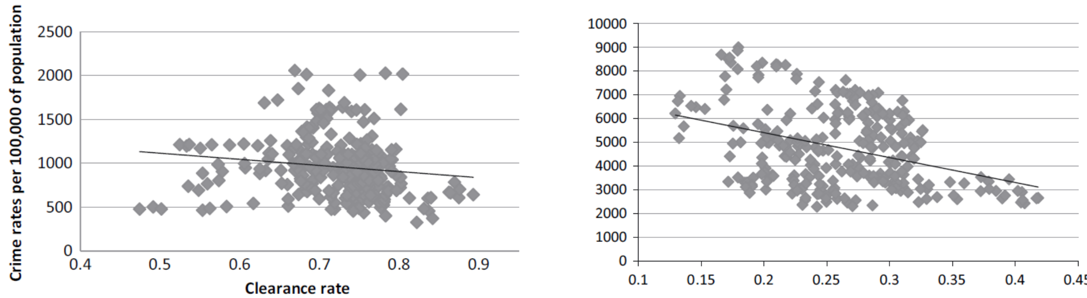


Figure 3.4: A Comparison of Sexual Assault Rates in Canada and the US

Source: Canadian data are from CANSIM tables 252-0013 and 252-0051. US data are from the Federal Bureau of Investigation website (`www2.fbi.gov/ucr/05cius/data/table_01.html`).



(a) Violent Crime (b) Property Crime

Figure 3.5: A Scatterplot of Crime Rates and Clearance Rates for Canada

Source: Canadian data are from CANSIM tables 252-0013.

Are correlations between clearance rates and crime rates different from correlations between clearance rates and the number of per capita police officers? Table 3.3 contains relevant Pearson correlation coefficients as well as the correlation coefficient between clearance rates and the number of per capita police officers, with all variables in natural logarithms. The correlation coefficients are consistent with the above graphs, revealing a -0.122 correlation coefficient between violent crime rates and corresponding clearance rates and a stronger correlation (-0.43) with respect to property crimes and clearance rates for such crimes. On the other hand, Pearson correlation coefficients between the number of per capita police officers and crime rates are positive (from 0.2 to 0.5). Further, correlation coefficients between the number of per capita police officers and clearance rates for violent and property crimes are quite different, with a value of 0.19 for violent crime clearance rates and -0.34 for property crime clearance rates. At the very least, these simple statistics suggest that the variation in clearance rates across provinces and over time is different from corresponding movements in the number of per capita police officers. Therefore, assessing whether clearance rates have different deterrence rates in comparison to the number of police officers becomes a worthwhile exercise.

Table 3.3: Pearson Correlation Coefficients

Violent crime rate	1.00000				
Property crime rate	0.55690	1.00000			
Clearance rate (violent crime)	−0.12262	0.06996	1.00000		
Clearance rate (property crime)	−0.29707	−0.43087	0.51906	1.00000	
Per capita police officers	0.29754	0.50426	0.19985	−0.33855	1.00000

Note: All variables are in natural logarithms.

3.4 Theoretical Model

In the Polinsky and Shavell (2000) model of crime, potential criminals make their decision to commit crime based on expected benefits and costs. The expected costs are determined by policy variables p and s, where p is the probability of being apprehended and convicted and s is the sanction that comes with conviction. Individuals will then commit crime if their assessment of the benefits, b, outweighs the expected costs (i.e., if $b > ps$ for risk-neutral individuals).

If we suppose that, for any given criminal opportunity, the benefits associated with crime are determined by a draw from a distribution $f(b)$, then the probability of a crime occurring is $1 - F(ps)$, where $F(\cdot)$ is the associated cumulative density function. The number of crimes committed in a given time period, therefore, would be a function of this probability and the number of criminal opportunities per time period, N. In addition, the distribution of benefits or the number of criminal opportunities may depend on various demographic and economic variables, such as the proportion of young males, population size, economic conditions and welfare transfers, among others. If we denote the vector describing these variables by X, then the expected number of crimes in a given period is $N(X)[1 - F(ps|X)]$, which we call the supply of crime, $S(ps, X)$.

We build on this model by assuming the probability of apprehension to be a function of labour, l, and capital, k, so that we have $p(l,\, k)$. This allows for the following comparative statics. First, the effect of an increase in labour on the expected number of crimes is given by $-Nf(ps)\partial p/\partial ls < 0$, while the effect of an increase in the probability of apprehension is $-Nf(ps)s < 0$. The model offers some insight into differences between examining the effect of an increase in labour versus the effect of an increase in the probability of apprehension. Specifically, it is worth noting that the difference between the two is just one additional step. When looking at the effects of more labour, there is an additional term in the comparative static $-\partial p/\partial l$. Intuitively, an increase in labour should result in a corresponding rise in the probability of apprehension.

A key question is: How do we measure l ? Policing is a skilled profession, and police officers vary in the amount of human capital that they have. As a result, the number of police officers is a noisy measure of what l is trying to capture in the model. An alternative is to use spending on wages and salaries paid to police personnel. If the labour market for police officers operated efficiently, then there would be reason to believe that an individual's salary would reflect their human capital, and expenditures would be a good measure. However, police officers are generally unionized, and salaries generally reflect seniority rather than productivity, so expenditures are also a noisy measure of l. Thus, while there may be measurement issues associated with looking at clearance rates as a proxy for the probability of apprehension, p, in the model, such measurement issues also exist when considering the effect of an increase in labour, l (measured through police officers). In addition, there are reasons to believe that incomplete data imply that changes in l are not holding all else fixed, meaning that any empirical analysis will not capture the desired partial derivative. However, this would not be the case when looking at the probability of apprehension.

Consider the following example of how changes in l may occur with changes in other variables that are unobserved in the data. Police budgets are set on an annual basis, with expenditures made on both labour and capital. Therefore, once the budget has been set, an increase in expenditures on labour necessarily implies a decrease in expenditures on capital.[18] As a result, changes in

[18]This is in fact how the budgetary process works in Ontario. Future budgets for police services are set by the municipal police services (after consultation with the chief of police services) and must then be approved by the municipality.

expenditures on police officers in the data do not necessarily capture the partial derivative expressed above.

There is supportive evidence that this is the case. Recent media coverage has documented the significant number of Toronto police officers who earn more than a $100,000 in annual income, primarily because of overtime pay.[19] Therefore, it is unsurprising that roughly 90% of the police services budget for the City of Toronto is consumed by salaries and benefits.[20] It is also important to note these trends may not be exclusive to Toronto. Figures 3.6 and 3.7 plot, respectively, the number of police officers per 100,000 of population and police expenditures (in real dollars) on salaries and wages, benefits, accommodation costs, fuel and maintenance across provinces and over time.[21] While there are differences across provinces, time-series movements are comparable. The number of police officers declined with the observed drop in crime rates through most of the 1990s, but then started to increase during the following decade. On the other hand, per capita police expenditures (in real dollars) on

[19]See "Sunshine List: More than a third of Toronto's police officers earned $100,000 in 2013," by Jennifer Pagliaro, published by the Toronto Star on Friday, March 7, 2014, and available at `thestar.com/news/gta/2014/03/07/sunshinelistmorethanathirdoftorontospoliceofficersearned100000in2013.html`.

[20]See "Rising police budget draws few questions from councillors," by Betsy Powell, published by the Toronto Star on January 30, 2014, and available at `thestar.com/news/cityhall/2014/01/30/risingpolicebudgetdrawsfewquestionfromcouncillors.html`.

[21]The data do not allow us to specifically isolate wages and expenditures.

salaries and wages have steadily increased over time for all provinces. These trends suggest that the crowding out of capital expenditures by an increase in wages and salaries is a distinct possibility.

The model suggests other benefits from focusing on clearance rates as opposed to the number of police officers. Specifically, the data on the number of officers must be aggregated across all crimes, while clearance rates are crime specific. Suppose that there are J different crimes, or classes of crimes, in the data. Further suppose that the probability of solving crimes of type j is a function of the resources devoted specifically to them, l_j and k_j, so that there exists a crime-specific probability of apprehension function, $p_j(l_j,\, k_j)$. Finally, suppose that the distribution of benefits also varies across crimes, so that there are J distributions, $f_j(b)$. The effect of an increase in expenditures devoted to solving crime j, would therefore be $-Nf_j(p_j(l_j,\, k_j)s)\partial p_j/\partial l_j$, while the effect of an increase in the probability of apprehension for crime j is $-Nf_j(p_j(l_j,\, k_j)s)$. Since data are not available for the amount of resources devoted to each type of crime, it is not possible to cleanly disaggregate crime rates into groups of crimes or individual crimes for regression analysis. However, this can be done with the probability of apprehension because we do have data for p_j. Simply put, employing clearance rates is informative because data are available for different categories of crime, which may (with appropriate caveats) capture deterrence effects associated with resources devoted to reducing different types of crimes as opposed to estimating the effects of the same number of police officers with respect to different types of crimes.

Finally, there may be an argument that police expenditures directly impact crime rates. Indeed, some papers have used reduced form regressions to estimate the impact of police expenditures on crime rates.[22] However,we think that relying on estimates of expenditures on crime is not very informative and inconsistent with the classic Becker (1968) and subsequent models of crime, which clearly specify that individually rational criminals respond to incentives generated by changes in the probability of apprehension and severity of penalties. Changes in police expenditures should have an indirect effect on crime rates, conditional on how such spending is allocated and the marginal deterrence impacts of such policies.

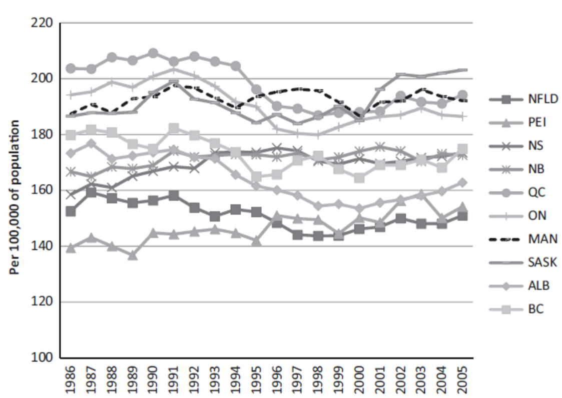


Figure 3.6: Numbers of Police Officers by Province, by Year

Source: Canadian data are from CANSIM tables 252-0002.

[22]Pogue (1975) is an early example. Ajilore and Smith (2011) is a more recent study. Shoesmith and Klein (2012) offer a nice summary of these papers.

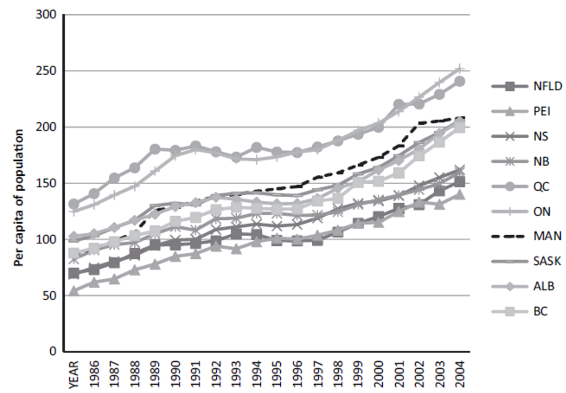


Figure 3.7: Per Capita Police Expenditures by Province, by Year in Real Dollars

Source: Canadian data are from CANSIM tables 252-0002.

3.5 Econometric Model

We test the effects of clearance rates on different types of crime through a parsimonious reduced form specification comparable to recent US studies:

$$\ln Y_{it} = \beta_0 + \beta_1 \ln X_{it} + \beta_2 \ln \psi_{it} + \beta_3 \ln \theta_{it} + \alpha_i + \gamma_t + \varepsilon_{it}\,, \tag{3.1}$$

where $\ln Y_{it}$ is the log of the annual crime rate per 100,000 of population of province i at time t, X_{it} is the measure of the probability of apprehension, ψ_{it} a vector of government policies that might plausibly impact trends in crime rates, θ_{it} is a vector of other time-varying demographic and province-specific factors, α_i are province fixed effects and γ_t is a vector of year dummy variables. The error term, ε_{it}, is assumed to be independently and identically distributed. Data from all 10 provinces from 1986 to 2005 is used for estimation.

The focus of our study is estimating β_1, which is the coefficient estimate of $\ln X_{it}$. Hence, β_1 is the elasticity of crime with respect to apprehension and X_{it}

is measured by the number of incidents cleared by the police divided by the number of police reported incidents. We evaluate the sensitivity of coefficient estimates by including the number of police officers per 100,000 of population. Deterrence may also be captured by incarceration rates, and we measure these through the number of prisoners per 100,000 of population. In such regressions, observations for British Columbia are dropped because data on imprisonment rates are unavailable for that province.

Province-specific policies denoted by ψ_{it} are the hourly minimum wage (in real dollars) and average annual government transfers (in real dollars) to the poorest quintile of population. The intuition is that an increase in the minimum wage or average government transfers acts as an income effect, which reduces the incentive to engage in illegitimate activities. The use of Canadian data offers some pronounced cross-province and time-series variation in order to identify the impacts of social assistance transfers on crime.[23] There also exists significant time-series and cross-province variation in the Canadian minimum wage laws relative to US legislation.[24]

[23]Sen and Ariizumi (2013) report that some Canadian provinces implemented significant reductions to social assistance transfers during the 1990s. In its February 1994 budget, the Conservative Government of Alberta specifically outlined a 19.3% cut in social services. The Ontario Progressive Conservative government followed Alberta's lead, slashing welfare benefits by roughly 22% in 1996. Perhaps more importantly, the amendments implemented through the Ontario Works Act (enacted in 1996) not only reduced the generosity of welfare income but also increased the costs to welfare participation.

[24]As documented by Sen et al. (2011), 1992 to 2005 witnessed several significant legislative changes, with 11 amendments to the minimum wage enacted by Quebec, nine by Nova Scotia,

θ_{it} consists of covariates for the employment rate for prime-aged adults aged 15 and over, the province population, males aged 15 to 24 years as a proportion of the total population, the median income of social assistance recipients and the total number of new immigrants per 100,000 of population. Among these controls, the employment rate and the proportion of young adults have been identified as important determinants of crime rates. Controlling for all else, a more prosperous economy with a higher probability of employment reduces the incentive to commit crimes in order to earn income. Most crimes are committed by young males. Therefore, a province with a higher proportion of young males may experience an increase in crime levels. We think that the use of these covariates results in empirical specifications that are comparable to models used by Levitt (1997, 1998).[25]

In summary, we are identifying the effects of clearance rates, welfare transfers and the minimum wage by relying on within-province time-series variation while controlling for province-specific differences that are constant through time and year-specific shocks that are common across jurisdictions. We also

seven by British Columbia and Manitoba, six by New Brunswick, five by Saskatchewan and Alberta and four by Ontario. As discussed, over the sample period of our study, the minimum wage set by the federal government supersedes the wage set by state government for many states (in the US), resulting in mostly time-series variation.

[25]At the Municipal Statistical Area (MSA) level, Levitt (1997, 1998) uses percent of black households, percent of female-headed households and percent of population aged 15 to 24. At the state level, he employs state unemployment rates and real per capita state and local spending on education and public welfare.

evaluate the sensitivity of our results by running first differences regressions where all variables are transformed by subtracting lagged values from corresponding current values, resulting in growth rates. The estimable framework then becomes:

$$\ln Y_{it} - \ln Y_{i,t-1} = \beta_0 + \beta_1(\ln X_{it} - \ln X_{i,t-1}) + \beta_2(\ln \psi_{it} - \ln \psi_{i,t-1})$$
$$+\beta_3(\ln \theta_{it} - \ln \theta_{i,t-1}) + \tilde{\gamma_t} + \tilde{\varepsilon_{it}} \tag{3.2}$$

As noted by Chalfin and McCrary (2013), estimating these first-differences specifications is typical in the literature in order to remove noise and unobserved jurisdiction-specific characteristics, which are time-invariant. Differencing the data removes the between-jurisdiction variation but does not eliminate the potential confounding effects of unobserved national specific shocks, which are soaked up through year dummies. Our benchmark estimates of equations (3.1) and (3.2) are based on OLS regressions. Following Chalfin and McCrary (2013), we also use GMM to estimate equation (3.2). The benefit of relying on GMM is that we are able to assess the sensitivity of our findings because GMM does not restrict the empirical model to a single specific parametric specification. The kernel and the bandwidth are chosen using the methods proposed by Newey et al. (1987). Finally, given the long time period of the data, we also GLS in order to account for serial correlation. The GLS estimates are based on the cross-sectionally heteroskedastic and time-wise autoregressive model for pooled cross-sections of time series initially developed by Parks (1967). None of these methods accounts for endogeneity bias or measurement error. Our IV strategy to tackle the corresponding bias in OLS estimates is discussed in the next section.

3.6 Results

3.6.1 Baseline Estimates

Table 3.4 contains empirical estimates of the effects of violent crime (columns (1) to (3)) and property crime (columns (4) to (6)) clearance rates on corresponding crime rates, controlling for other factors. Columns (1) and (3) contain results conditioned on the use of covariates and province and year fixed effects; columns (2) and (5) evaluate the effects of adding the number of per capita police officers; and columns (3) and (6) include the number of per capita police officers and province-specific trends. Panel A reports estimates based on 10 provinces from 1986 to 2005, panel B consists of results from the same provinces but from 1988 to 2005 in order to accommodate one- and two-year lagged clearance rates and panel C contains estimates from nine provinces between 1986 and 2005, which are conditioned on the use of per capita incarceration rates. Given the long time-series of the data, we focus on issues that are a consequence of autocorrelation and heteroskedasticity. Therefore, standard errors of coefficient estimates are White and Newey-West corrected for second-order autocorrelation and unknown heteroskedasticity.[26] For the sake of brevity, we report only coefficient estimates of clearance rates, the number of per capita police officers (when employed) and incarceration rates.

First, coefficient estimates of clearance rates are negative and statistically significant at the 1% or 5%levels across almost all columns. With respect to violent crime, estimates of clearance rates are significant in column (1) across all panels. The addition of the number of police officers in column (2) removes the statistical significance of clearance rates in panels A and B. However, the inclusion of province-specific trends in column (3) results in statistically significant (at the 1% level) estimates of clearance rates across all panels with an elasticity of roughly -0.3. Coefficient estimates of per capita police remain positive and are significant in all panels. Lagged values of clearance rates are, in most cases, statistically insignificant.[27] Finally, while the coefficient estimate of incarceration rates is negative and statistically significant (at the 10% level) in column (1), it becomes insignificant with the inclusion of police officers and trends in column (3).

[26]Another option would be to cluster the standard errors by province. However, the number of provinces (10) – and therefore clusters – would be quite small. Conversations with Jeff Wooldridge suggest that, in such cases, using a Newey-West correction for autocorrelation may be a better strategy (econometrics course offered by the Canadian Economics Association Meetings, Ryerson University, May 26 to May 28, 2009).

[27]The inclusion of one- and two-year lagged values of clearance rates is intended to evaluate the possibility that information on actual clearance rates, and therefore the likelihood of apprehension, may take some time to reach criminals before they take them into account in their cost-benefit decisions.

Table 3.4: OLS Estimates with Respect to Violent and Property Crime Rates: ProvinceYear Data

	Violent Crime						Property Crime					
	Base (1)		Police (2)		Police and Province-Specific Trends (3)		Base (1)		Police (2)		Police and Province-Specific Trends (3)	
A. 10 provinces, 1986–2005												
Clearance rate (per incident)	−0.245**	(0.116)	−0.075	(0.072)	−0.334***	(0.101)	−0.592***	(0.086)	−0.563***	(0.073)	−0.516***	(0.048)
Per capita police officers per 100,000 of population			1.505	(0.199)a	0.843	(0.544)			0.542**	(0.214)	−0.345*	(0.196)
Adjusted R-squared	0.9298		0.9497		0.9695		0.9639		0.9657		0.9807	
B. 10 provinces, 1988-2005												
Clearance rate (per incident)	−0.300**	(0.119)	−0.134	(0.109)	−0.306***	(0.08)	−0.498***	(0.06)	−0.486***	(0.056)	−0.431***	(0.043)
One-year lagged clearance rate (per incident)	0.0157	(0.116)	0.0484	(0.112)	−0.0297	(0.075)	−0.074	(0.052)	−0.054	(0.050)	−0.097**	(0.044)
Two-year lagged clearance rate (per incident)	−0.067	(0.136)	0.0184	(0.126)	−0.209**	(0.094)	−0.0002	(0.073)	0.003	(0.067)	−0.065*	(0.049)
Per capita police officers per 100,000 of population			1.437***	(0.247)	0.870***	(0.209)			0.432**	(0.223)	−0.304	(0.182)
Adjusted R-squared	0.9275		0.9446		0.9723		0.9686		0.9696		0.9847	
C. 9 provinces, 1986-2005												
Clearance rate (per incident)	−0.371***	(0.119)	−0.232**	(0.095)	−0.362***	(0.097)	−0.543***	(0.075)	−0.536***	(0.067)	−0.510***	(0.053)
Per capita police officers per 100,000 of population			1.552***	(0.183)	0.923***	(0.279)			0.469***	(0.219)	−0.189	(0.206)
Incarceration rates per 100,000 of population	−0.125*	(0.069)	−0.002	(0.057)	−0.0575	(0.077)	−0.148**	(0.063)	−0.109	(0.066)	−0.083	(0.056)
Adjusted R-squared	0.9302		0.9516		0.9660		0.9583		0.9590		0.9598	
Province fixed effects	Yes		Yes		Yes		Yes		Yes		Yes	
Year fixed effects	Yes		Yes		Yes		Yes		Yes		Yes	
Province linear trends	No		No		Yes		No		No		Yes	

Note: Estimates in columns (1) to (3) are with respect to violent crime, and results in columns (4) to (6) are with respect to property crime. Standard errors are White and NeweyWest corrected for second-order autocorrelation. Other covariates that are not reported but are included in all regressions are the minimum wage, average government transfers, employment rates, population, the proportion of young males aged 15 to 24, the number of new immigrants per 100,000 of population and average income of welfare recipients.With the exception of fixed effects, all variables are in natural logarithms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In summary, the important result is the relative stability of coefficient estimates of clearance rates. Clearance rates are negatively correlated with violent and property crime rates across all columns and also statistically significant (in most cases). When included in tandem with clearance rates, per capita police rates either do not have the correct sign or are statistically insignificant. However, coefficient estimates of clearance rates remain statistically significant and possess negative signs. These findings suggest that previous studies, which focus exclusively on the number of per capita police officers to capture the impacts of the probability of apprehension, may understate the overall deterrence effects of enforcement by police.

Table 3.5 explores the deterrent effects of different measures of apprehension in some more detail. Specifically, column (1) contains the results of focusing on the effects of the per capita number of police officers on property crime in isolation from other deterrence measures. Columns (2) and (3) contain results on the impacts of arrest rates per 100,000 of population and police expenditures per capita of population, respectively.[28] Column (4) assesses the effects

[28]Data on provincial expenditures on police services are available from the Police Administration Survey conducted by Statistics Canada. For further details of this survey, see `www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3301`. The Police Administration Survey collects data on police personnel and expenditures from each municipal, provincial and federal (RCMP) police service in Canada. As detailed by Hutchins (2014), police expenditures are actual operating expenditures on salaries and wages, benefits, accommodation costs, fuel, maintenance and so forth. Unfortunately, capital expenditures are not included.

of including all these measures. Columns (5) to (8) are similarly organized and contain estimates of apprehension measures with respect to violent crime.

Broadly speaking, the results in table 3.5 are similar to corresponding estimates in table 3.4. The coefficient estimate of the number of police officers with respect to property crime is negative and statistically significant (at the 10% level) in column (1) but becomes insignificant in column (4) with the inclusion of other measures of apprehension. The arrest rate covariate is statistically significant (at the 1% level) in columns (2) and (4) but positive. Per capita police expenditures are negatively associated with property crime but statistically insignificant. On the other hand, the coefficient estimate of clearance rates (in column (4)) possesses a negative sign and is statistically significant at the 1% level. The coefficient estimate of -0.49 is comparable to results in table 3.4. Higher clearance rates are also correlated with a reduction in violent crime rates (at the 1% level). Coefficient estimates of arrest rates and the number of police officers are statistically significant but possess counterintuitive signs. The per capita police expenditure covariate is significant in column (7), but becomes statistically insignificant with the inclusion of other measures of apprehension.

Table 3.5: OLS Estimates of Different Measures of Apprehension: Province-Year Data

	Property Crime				Violent Crime			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Clearance rate				−0.493***				−0.441***
(per incident)				(0.062)				(0.080)
Per capita police officers	−0.390*			−0.255	0.925 (0.267)**			0.449***
per 100,000 of population	(0.239)			(0.174)	(0.267)			(0.131)
Arrest rate per		0.409***		0.379***		0.474***		0.486***
100,000 of population		(0.070)		(0.057)		(0.038)		(0.029)
Police expenditures per			−0.170	−0.070			0.368***	0.0278
capita of population			(0.164)	(0.110)			(0.143)	(0.098)
Other exogenous covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province/year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province linear trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.9703	0.9766	0.9701	0.9863	0.9667	0.9821	0.9648	0.9884

Note: Estimates in columns (1) to (4) are with respect to property crime, and results in columns (5) to (8) are with respect to violent crime. Standard errors (in parenthesis) are White and Newey-West corrected for second-order autocorrelation. Other covariates that are not reported but are included in all regressions are the minimum wage, average government transfers, employment rates, population, the proportion of young males aged 15 to 24, the number of new immigrants per 100,000 of population and average income of welfare recipients. With the exception of fixed effects, all variables are in natural logarithms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A relevant question is whether these estimates are comparable with the US based estimates. Our results with respect to the effect of arrest rates and clearance rates on violent crime cannot be compared because many US studies obtain coefficient estimates with negative signs while our regressions reveal arrest rates with positive coefficients. However, our coefficient estimates of the number of police officers with respect to property crime are consistent with the -0.5 elasticity obtained by Levitt (2002) and a bit larger than the preferred estimate of -0.17 reported by Chalfin and McCrary (2013). Further, our estimates of the effects of an increase in clearance rates are comparable to those obtained by Craig (1987) with respect to all crimes. As discussed, there is very little empirical research on the effects of clearance rates.

In summary, coefficient estimates of clearance rates remain robust and statistically significant even after the inclusion of other plausible measures of apprehension. Given the presence of either weak statistical significance or counterintuitive signs, we do not use per capita arrest rates, the number of police officers or per capita police spending as covariates in further regressions. This strategy allows us to focus on obtaining robust estimates of the impacts of clearance rates. However, we acknowledge that omitting these plausible measures of apprehension might induce some bias in coefficient estimates of clearance rates, the magnitude of which is a function of the correlation between these measures and clearance rates. Therefore, coefficient estimates of clearance rates should not be interpreted as causal relationships and should be treated with appropriate caveats. Further, we do not employ province-specific linear trends in the remaining regressions because coefficient estimates of clearance rates remain relatively stable after their inclusion (with other covariates and province and year fixed effects). This allows the other covariates to be identified by time-series variation within provinces.

Table 3.6 evaluates the sensitivity of coefficient estimates of clearance rates through alternative estimation strategies. Columns (1) to (3) contain estimates with respect to violent crime. Column (1) contains OLS first differences estimates based on equation (2); column (2) reports the results of GMM estimation based on equation (2); and column (3) contains results from a GLS regression based on equation (1). Columns (4), (5) and (6) are organized similarly, but with respect to property crime rates. Broadly speaking, the estimates are quite comparable to results in the previous table. The coefficient estimate of the clearance rate with respect to violent crime from the first differences model is -0.126 and statistically significant at the 10% level. Corresponding GMM and GLS estimates are -0.241 and -0.244, respectively, and statistically significant at the 5% and 1% levels, respectively. Coefficient estimates of the effects of clearance on property crime rates are also significant (at the 1% level) and range from -0.3 to -0.6. In terms of other covariates, levels and GLS estimates of the minimum wage with respect to violent crime are negative and statistically significant at the 1% and 5% levels. While GMM and GLS estimates of the minimum wage with respect to property rates are significant, they possess counterintuitive positive signs. First differences and GMM estimates of employment rates are negative and significant with respect to property crime but possess implausibly large coefficient estimates.

Table 3.6: First Differences, GMM, and GLS Estimates With Respect to Violent and Property Crime Rates: Province-Year Data 1986-2005

	Violent Crime			Property Crime		
	FD	GMM	GLS	FD	GMM	GLS
	(1)	(2)	(3)	(4)	(5)	(6)
Clearance rate	-0.126^{*}	-0.241^{**}	-0.244^{***}	-0.339^{***}	-0.611^{***}	-0.583^{***}
	(0.074)	(0.115)	(0.084)	(0.049)	(0.08)	(0.058)
Minimum wage	-0.086	-0.269	-0.276^{**}	-0.076	0.390^{***}	0.325^{***}
	(0.117)	(0.169)	(0.117)	(0.101)	(0.130)	(0.103)
Average government transfers	0.092^{*}	-0.130	-0.138	-0.013	0.026	-0.029
	(0.058)	(0.116)	(0.108)	(0.059)	(0.077)	(0.091)
Employment rate	1.100^{**}	0.227	0.530	-1.01^{***}	-2.176^{***}	-0.533
	(0.447)	(1.024)	(0.550)	(0.540)	(0.959)	(0.469)
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared/Log likelihood	0.4402		247.325	0.5775		280.825

Note: Results are based on data for 10 provinces from 19862005 (200 obs.). Estimates in columns (1) to (3) are with respect to violent crime, and results in columns (4) to (6) are with respect to property crime. Standard errors of coefficient estimates of first difference regressions are White and Newey-West corrected for second-order autocorrelation. For GMM regressions, the kernel and the bandwidth are chosen using the methods proposed by Newey and West (1987). The GLS estimates are based on the cross-sectionally heteroskedastic and time-wise autoregressive model for pooled cross-sections of time series initially developed by Parks (1967). Other covariates that are not reported, but included in all regressions, are population, the proportion of young males aged 15 to 24, the number of new immigrants per 100,000 of population and average income of welfare recipients. With the exception of fixed effects, all variables are in natural logarithms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6.2 Endogeneity Bias and Instrumental Variables

Reduced form estimates are premised on the assumption that changes in clearance rates exogenously impact crime trends. However, as discussed above, OLS estimates are likely to be biased because of reverse causality. It might very well be the case that increases in crime result in public pressure for rapid arrests of perpetrators. In that case, OLS estimates of clearance rates will be biased and confounded. There are other possible sources of bias in estimates of the true impacts of clearance rates. First, the number of police-reported incidents enters the left- and right-hand side of the econometric model. This is similar to the classic measurement error noted by Borjas (1980) with respect to estimating the effects of average weekly (or annual) wages on weekly (or annual) hours of work. The problem is that hours of work enter both the right- and left-hand side of the equation, resulting in a downward bias in coefficient estimates of the effects of changes to average wages. Second, another type of measurement error arises from the fact that some crimes are not reported to the police, which means that clearance rates overstate the true probability of apprehension. Therefore, estimates of the effects of the clearance rate on crime will be upward biased with regards to the true crime rate, but accurate for the recorded crime rate. Third, as explained by Cook (1979), coefficient estimates of clearance rates may also be biased if rational criminals respond to an increase in clearance rates by committing crimes that are more difficult to solve.

In order to evaluate the magnitude of bias in OLS estimates, we construct political party dummy variables as well as instruments based on the proportion of seats held by different parties, which can arguably identify trends in clearance rates and not be correlated with the right-hand side error term of equation (3.1). The intuition is that changes in the governing party at the province level might impact the allocation of resources to law enforcement agencies. This approach is consistent with Besley and Case (2000), who suggest that variation in political variables can exogenously identify trends in policy variables and not have any impact on the outcome of interest. The use of political variables as instruments is also comparable to the strategy used by Levitt (1997), who relied on variation in mayoral and gubernatorial electoral cycles to instrument police rates.

The presence of similar parties allows us to create the same political party fixed effects across provinces. The major political parties in most Canadian provinces are similar: the Liberal Party, the Conservative Party and the New Democratic Party. In addition, British Columbia has the Social Credit Party, and the Parti Quebecois is a major political force in Quebec.[29] Therefore, we construct three dummy variables that take a value of 1 if one of the major political parties (Liberals, Conservatives or NDP) is in power. Hence, the omitted category is the presence of a ruling political party that does not have a national presence. As mentioned, we also construct instruments based on the

[29]In terms of ideology, the Conservatives are considered by most to be on the right end of the political spectrum, the Liberals are positioned at the centre left, and the NDP is on the far left.

proportion of seats held by each party.[30] These variables are intended to reflect the ease with which governing parties can implement policy reforms and corresponding changes in government spending. A greater proportion of legislature seats might imply that a political party with a "tough on crime" agenda will have greater flexibility in increasing government spending on specific anti-crime policies at the expense of reduced expenditures on other items. We also use three- and four-year lagged clearance rates that may impact trends in clearance rates but do not share a statistically significant relationship with current crime rates.We interact these lagged clearance rates with each of the political party dummy variables as a crude proxy for variation in clearance rates generated by changes to political regimes that might result in shifts in provincial funding. Finally, consistent with our theoretical model, we employ police expenditures as another instrument that should impact trends in clearance rates.[31] Given the ambiguities of empirically defining per capita police expenditures, we define per capita expenditures in terms of population and the number of police officers.[32]

[30]Information on the governing political party and the number of seats held by each party was obtained from the websites of the legislative assemblies of each province, the Social Credit Party and the Parti Quebecois.

[31]We are grateful to an anonymous referee for recommending this.

[32]Employing police expenditures as an alternate instrument is also a useful sensitivity check, given the possibility that changes in the governing political party or the seats held by it could indirectly affect crime rates because a different party might result in significant shifts in areas that affect crime, such as employment, public housing and education policies. We thank an anonymous referee for pointing this out.

Table 3.7 contains first- and second-stage estimates from a variety of instrumental variables regressions. Columns (1) to (5) document estimates with respect to violent crime while columns (6) to (10) report corresponding results for property crime. Panel A (B) contains first-stage (second-stage) results. Columns (1) and (6) in table 6 contain estimates of the effects of clearance rates using political party dummies as instruments. Results in columns (2) and (7) are from an estimable model in first differences with second stage IV estimates based on political party dummies and political party dummies interacted with three- and four-year lagged clearance rates. Columns (3) and (8) report results from a regular log-log model with second-stage results identified by political party dummies and police expenditures per capita of population as instruments. Columns (4) and (9) contain results obtained from adding the proportion of seats held by political parties as instruments (in addition to political party dummies and police expenditures per capita of population). Finally, estimates from employing political party dummies and police expenditures per 100,000 of police officers as instruments are detailed in columns (5) and (10).

The availability of multiple instruments allows us to conduct tests of overidentifying restrictions and evaluate the sensitivity of findings to the use of different instrumental variables. Results from standard Sargan and Hansen tests of over-identifying restrictions yield test statistics that do not reject the null hypothesis of over-identification.[33] With the exception of property crime clearance rates instrumented by political party dummies (column (6)), we can

[33]These are available on request.

Table 3.7: Instrumental Variables Estimates With Respect to Violent and Property Crime Rates: Province-Year Data

	Violent Crimes				
	Violent Crime (1)	Violent Crime and First Differences (2)	Violent Crime Police exp. per capita of pop. (3)	Violent Crime Police exp. per capita of pop. (4)	Violent Crime Police exp. per capita of police (5)
A. First Stage (See Instruments Below)					
Per capita police expenditures			-0.421^{***} (0.142)	-0.382^{***} (0.147)	-0.405^{***} (0.173)
F-stat, p-value	5.215, 0.0019	3.036, 0.0027	7.272, 0.0000	5.355, 0.0000	4.508, 0.0001
Adj. R^2	0.5694	0.6094	0.6140	0.6205	0.6099
B. Second Stage					
Clearance rate	-0.519^{**} (0.245)	-0.540^{*} (0.320)	-0.689^{***} (0.202)	-0.726^{***} (0.194)	-0.496^{***} (0.191)
Adj. R^2	0.9262	0.3036	0.9203	0.9186	0.9268
	Property Crimes				
	Property Crime (6)	Property Crime and First Differences (7)	Property Crime Police exp. per capita of pop. (8)	Property Crime Police exp. per capita of pop. (9)	Property Crime Police exp. per capita of police (10)
A. First Stage (See Instruments Below)					
Per capita police expenditures			-0.561^{***} (0.150)	-0.568^{***} (0.150)	-0.658^{***} (0.150)
F-stat, p-value	0.621, 0.6030	4.929, 0.0000	5.386, 0.0000	3.648, 0.0010	4.047, 0.0004
Adj. R^2	0.6137	0.8966	0.8980	0.8980	0.8959
B. Second Stage					
Clearance rate	-0.268 (0.667)	-0.599^{***} (0.202)	-0.831^{***} (0.200)	-0.725^{***} (0.180)	-0.544^{***} (0.188)
Adj. R^2	0.9585	0.4977	0.9610	0.9630	0.9638
Instruments	Political party dummies	Political party dummies and dummies interacted with 3- and 4-year lagged clearance rates	Political party dummies and per capita police exp. (police exp = per capita of pop.)	Political parties, share of seats and per capita police exp. (police exp = per capita of pop.)	Political party dummies and per capita police exp. (police exp = per 100,000 officers)

Note: Results are based on data for ten provinces from 1986-2005 (200 obs.) for all columns, except (2) and (7), the first-differences, using the same provinces from 1990-2005 (160 obs.). Other covariates that are not reported but are included in all regressions are the minimum wage, average government transfers, employment rates, population, the proportion of young males aged 15 to 24, the number of new immigrants per 100,000 of population, average income of welfare recipients and province and year fixed effects. With the exception of fixed effects, all variables are in natural logarithms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

reject the null hypothesis (at the 1% level) that the coefficient estimates of instruments are equal to zero. However, in all cases, the F statistic from the joint test of statistical significance of all instruments is less than 10 in value.

Instrumental variables estimates are quite similar to the above reduced form estimates. Second-stage coefficient estimates of clearance rates on violent crime (in panel B) range from -0.5 to -0.7 and are statistically significant at the 1% or 10% levels. With the exception of column (6), second-stage coefficient estimates of clearance rates with respect to property crime are roughly from -0.3 to -0.7 and are significant at the 1% level. Another key point is the statistical significance and the negative sign of coefficient estimates of per capita police expenditures across all columns and irrespective of whether police expenditures are measured per capita of population or per 100,000 of police officers. Coefficient estimates of per capita police expenditures range from roughly -0.4 to -0.6 and imply that an increase in such expenditures is associated with lower clearance rates for violent and property crimes.

While a negative relationship may seem odd, we suspect that it is driven by either an inability to measure capital expenditures or by the fact that expenditures may be a poor measure of quality units of labour, as discussed earlier.[34] For example, if a higher proportion of a police force's budget is being devoted

[34]In particular, the latter may cause this negative relationship because our sample period covers the tail end of a 30-year increase in crime levels and most of the current 20-year decline in crime. We thus have, in our sample, police forces that have gone through extensive hiring booms and generally cannot let officers go due to strong unions. The result is ageing police forces that get more expensive without necessarily getting more productive.

to salaries and wages, this implies less spending on capital expenditures. At some point, it is possible that diminishing returns set in, and any incremental increase in the marginal deterrence gained from more spending on wages and salaries is outweighed by the decline in deterrence associated with the corresponding decrease in relative spending on capital expenditures. From an empirical perspective, the result is negative signs of coefficient estimates of the effects of spending on wages and salaries and variable cost items (such as fuel expenses).

There may be concerns that the use of police expenditures as an instrument is invalid because of potential simultaneity between crime or clearance rates and police expenditures. An increase (decrease) in crime (clearance) rates may result in public demand for higher government expenditures on police services. To assess this possibility, we ran three-stage least squares regressions assuming the following: in the third stage, crime rates are functions of clearance rates; in the second stage, clearance rates are identified by police expenditures per capita; and in the first stage, police expenditures are identified by the political party dummies and the share of seats held by political parties. Relative to our earlier instrumental variables approach, we now evaluate whether changes in government and/or in the distribution of seats among political parties impact spending on police services.

Our results remain largely unchanged.[35] In the first stage, the coefficient

[35] Each of the three regressions have the minimum wage, average government transfers, employment rates, population, the proportion of young males aged 15 to 24, the number of new

estimates of the political variables with respect to per capita police expenditures are statistically significant, with an F statistic and p-value (from a joint test of significance) equal to 2.72 and 0.015, respectively. The impact of police expenditures on clearance rates has already been discussed. The third-stage coefficient estimates of clearance rates with respect to violent and property crime are -0.633 and -0.308, respectively, and statistically significant at the 1% and 10% levels, respectively.[36]

In summary, we do not claim that the IV estimates are successfully purged

Table 3.8: OLS Estimates with Respect to Violent and Property Crime Rates: Province-Year Data

	Violent Crime Diff.-in-Diff. (1)	Property Crime Diff.-in-Diff. (2)	Violent Crime 1990–2001 (3)	Property Crime 1990–2001 (4)
Clearance rate (per incident)	-0.215	-0.557^{***}	-0.398^{***}	-0.446^{***}
	(0.133)	(0.097)	(0.148)	(0.060)
Clearance rate (per incident) interacted with a year dummy for 1990-2001	-0.112	-0.066		
	(0.181)	(0.051)		
Province fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Adjusted R-squared	0.9296	0.9641	0.9383	0.9756

Results in columns (1) and (2) are based on data for 10 provinces from 19862005 (200 obs.).
Estimates in columns (3) and (4) are based on data for the same provinces from 19902001 (110 obs.).
Standard errors (in parenthesis) are White and Newey-West corrected for second-order autocorrelation.
Other covariates that are not reported but are included in all regressions are the minimum wage, average government transfers, employment rates, population, the proportion of young males aged 15 to 24, the number of new immigrants per 100,000 of population and average income of welfare recipients.
With the exception of fixed effects, all variables are in natural logarithms.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

immigrants per 100,000 of population, average income of welfare recipients and province and year fixed effects as right-hand side variables.

[36] All these estimates are available on request.

of reverse causality or measurement error. However, as discussed above, this is an extremely difficult accomplishment, and very few papers in the literature have actually attempted some type of correction. The relative similarity between OLS and IV estimates suggests that potential bias in OLS estimates of clearance rates may not be significant. However, it is also important to acknowledge that IV estimates are biased towards OLS estimates when multiple instruments are weak.[37] As a consequence, it is prudent to treat our findings with appropriate caution.

Finally, table 3.8 offers some further sensitivity analyses through models designed to evaluate whether changes in clearance rates during the 1990s possessed different marginal impacts relative to other years in our sample. The motivation of this exercise is to investigate whether clearance rates might be one of the contributing factors behind the significant decline in crime rates observed during the 1990s.[38] Columns (1) (violent crime) and (2) (property crime) contain difference-in-differences regressions based on the entire sample (1986 to 2005), with an additional clearance rate covariate interacted with a dummy that takes a value of 1 for all observations from 1991 to 2000 and is 0 otherwise. The coefficient estimate of this dummy variable reflects the marginal effect of

[37]We are very grateful to an anonymous referee for pointing this out. An excellent exposition of this point is available from "Weak Instruments - EC533: Labour Economics for Research Students" by Jörn-Steffen Pischke, available at `econ.lse.ac.uk/staff/spischke/ec533/WeakIV.pdf`.

[38]Levitt (2004) cites increased police hiring as one of the factors but does not mention clearance rates.

clearance rates during this time period relative to other years. Columns (3) (violent crime) and (4) (property crime) are based on an alternative approach in which regressions are based on a reduced sample (1991 to 2000) during which significant reductions in crime were observed.

Consistent results emerge across columns. The coefficient estimate of the clearance rate interacted with the year dummy with respect to violent crime (column (1)) is statistically insignificant, while the coefficient estimate of the violent crime clearance rate in column (3) is statistically significant but not that different in magnitude from estimates in table 3.4. The interacted term with respect to property crime (in column (2)) is negative but statistically insignificant. The coefficient estimate of the property crime clearance rate in column (4) is negative and statistically significant and quite comparable to previous results. In summary, while these results offer further evidence on the importance of clearance rates, we cannot conclude with certainty that they were a significant determinant of the observed decline in crime during the 1990s.

3.7 Conclusion

Relative to the vast literature on crime and the effects of more police officers, the number of studies that have focused on the effects of clearance rates on crime is quite limited. This is unfortunate, as we think that the clearance rate is a reasonable approximation of Becker's probability of apprehension. Indeed, early empirical studies on crime and deterrence focused on the effects of clearance rates rather than changes to the size of a jurisdiction's police force. We evaluate the importance of clearance rates with respect to crime by using data across Canadian provinces from 1986 to 2005. The use of Canadian data is informative from a general perspective, given the correlation between US and Canadian crime rates over time. Exploiting Canadian data is also useful given that penalties for *Criminal Code* offences are set at the federal level, yielding some reassurance that estimates of the impacts of clearance rates are not biased by variation in local penalties, which reflect changes to the severity of penalty rather than the probability of apprehension. Finally, we note that Canada did not experience the adverse consequences associated with the crack cocaine epidemic that occurred over the sample period.

In terms of other contributions, we develop a simple model that links labour and capital to the probability of apprehension and the incentive to commit crime. The model allows us to construct instruments, such as per capita police expenditures that proxy effective labour units of policing and enable us to assess the sensitivity of OLS estimates. OLS, GMM, GLS and IV estimates yield very comparable results. All else being equal, an increase in the clearance rate is correlated with a reduction in crime; marginal effects are higher with respect to property crime rates. These results are robust to the use of police force size and a wide array of other covariates, fixed effects and province-specific linear trends. However, the similarity between the IV and OLS estimates might be an artifact of the relative statistical weakness of multiple instruments. Hence, our estimates should be treated with suitable caveats.

Further, our IV estimation has been conducted only with respect to clearance rates and ignores potential endogeneity bias in coefficient estimates for other covariates, such as the number of police officers (one measure of labour). Therefore, we cannot say that higher arrest rates or hiring more police officers does not result in lower crime rates. It is possible that these other measures of apprehension are significantly associated with lower crime rates and that we have been unsuccessful in purging coefficient estimates of endogeneity bias. Given these caveats, we interpret the consistent statistical significance of clearance rates as cautious evidence of the importance of the probability of apprehension, but with the possibility of bias, taking into account the results with respect to other measures of apprehension. In future research we hope to better understand the reasons for the positive correlation between crime and per capita police rates, possibly through the use of more structural methods of estimation.

3.8 Bibliography

AJILORE, O. AND J. SMITH, "Ethnic Fragmentation and Police Spending," *Applied Economics Letters* 18 (2011), 329–332.

AVIO, K. L., "An Economic Analysis of Criminal Corrections: the Canadian Case," *Canadian Journal of Economics* 6 (1973), 164–178.

———, "Capital Punishment in Canada: a Time-Series Analysis of the Deterrent Hypothesis," *Canadian Journal of Economics* (1979), 647–676.

AVIO, K. L. AND C. S. CLARK, *Property Crime in Canada: an Econometric Study* (University of Toronto Press, 1976).

———, "The Supply of Property Offences in Ontario: Evidence on the Deterrent Effect of Punishment," *Canadian Journal of Economics* (1978), 1–19.

BECKER, G. S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76 (1968), 169–217.

BESLEY, T. AND A. CASE, "Unnatural Experiments? Estimating the Incidence of Endogenous Policies," *The Economic Journal* 110 (2000), 672–694.

BLACK, D. A. AND D. S. NAGIN, "Do Right-to-Carry Laws Deter Violent Crime?," *The Journal of Legal Studies* 27 (1998), 209–219.

BORJAS, G. J., "The Relationship Between Wages and Weekly Hours of Work: The Role of Division Bias," *Journal of Human Resources* 15 (1980), 409–423.

CARR-HILL, R. A. AND N. H. STERN, "An Econometric Model of the Supply and Control of Recorded Offences in England and Wales," *Journal of Public Economics* 2 (1973), 289–318.

CARRINGTON, P. J. AND J. L. SCHULENBERG, "Structuring Police Discretion: The Effect on Referrals to Youth Court," *Criminal justice policy review* 19 (2008), 349–367.

CHALFIN, A. AND J. MCCRARY, "The effect of police on crime: New evidence from US cities, 1960-2010," *NBER Working Paper No. 18815* (2013).

CHEUNG, Y. W. AND P. G. ERICKSON, "Crack Use in Canada: A Distant American Cousin," in C. Reinarman and H. G. Levine, eds., *Crack in America: Demon Drugs and Social Justice* (University of California Press, 1997), 175–193.

COOK, P. J., "The Clearance Rate as a Measure of Criminal Justice System Effectiveness," *Journal of Public Economics* 11 (1979), 135–142.

CORMAN, H. AND N. MOCAN, "Carrots, Sticks, and Broken Windows," *The Journal of Law and Economics* 48 (2005), 235–266.

CORNWELL, C. AND W. N. TRUMBULL, "Estimating the Economic Model of Crime with Panel Data," *The Review of economics and Statistics* (1994), 360–366.

CRAIG, S. G., “The Deterrent Impact of Police: An Examination of a Locally Provided Public Service,” *Journal of Urban Economics* 21 (1987), 298–311.

DEZHBAKHSH, H. AND P. H. RUBIN, “Lives Saved or Lives Lost? The Effects of Concealed-Handgun Laws on Crime,” *The American Economic Review* 88 (1998), 468–474.

DEZHBAKHSH, H., P. H. RUBIN AND J. M. SHEPHERD, “Does Capital Punishment Have a Deterrent Effect? New Evidence from Postmoratorium Panel Data,” *American Law and Economics Review* 5 (2003), 344–376.

DI TELLA, R. AND E. SCHARGRODSKY, “Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack,” *The American Economic Review* 94 (2004), 115–133.

DONOHUE, J. J. AND S. D. LEVITT, “Measurement Error, Legalized Abortion, and the Decline in Crime: A Response to Foote and Goetz,” *The Quarterly Journal of Economics* 123 (2008), 425–440.

DRACA, M., S. MACHIN AND R. WITT, “Panic on the Streets of London: Police, Crime, and the July 2005 Terror Attacks,” *The American Economic Review* 101 (2011), 2157–2181.

DUGGAN, M., “More Guns, More Crime,” *Journal of Political Economy* 109 (2001), 1086–1114.

DURLAUF, S. N., C. FU AND S. NAVARRO, "Assumptions Matter: Model Uncertainty and the Deterrent Effect of Capital Punishment," *The American Economic Review* 102 (2012), 487–492.

DURLAUF, S. N., S. NAVARRO AND D. A. RIVERS, "Understanding Aggregate Crime Regressions," *Journal of Econometrics* 158 (2010), 306–317.

EHRLICH, I., "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation," *Journal of political Economy* 81 (1973), 521–565.

EVANS, W. N. AND E. G. OWENS, "COPS and Crime," *Journal of Public Economics* 91 (2007), 181–201.

FRYER, R. G., P. S. HEATON, S. D. LEVITT AND K. M. MURPHY, "Measuring the Impact of Crack Cocaine," *NBERworking paper no. 11318* (2005).

GANNON, M., "Crime Comparisons Between Canada and the United States," *Juristat* 21 (2001), 1–12.

GOULD, E. D., B. A. WEINBERG AND D. B. MUSTARD, "Crime Rates and Local Labor Market Opportunities in the United States: 1979-1997," *Review of Economics and Statistics* 84 (2002), 45–61.

HUTCHINS, H., "Police Resources in Canada, 2013," *Juristat: Canadian Centre for Justice Statistics, no. 85-002-X* (2014).

JOYCE, T., "A Simple Test of Abortion and Crime," *The Review of Economics and Statistics* 91 (2009), 112–123.

KLICK, J. AND A. TABARROK, “Using Terror Alert Levels to Estimate the Effect of Police on Crime,” *The Journal of Law and Economics* 48 (2005), 267–279.

LEVITT, S. D., “The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation,” *The Quarterly Journal of Economics* 111 (1996), 319–351.

———, “Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime,” *The American Economic Review* 87 (1997), 270–290.

———, “Why Do Increased Arrest Rates Appear to Reduce Crime: Deterrence, Incapacitation, or Measurement Error?,” *Economic Inquiry* 36 (1998), 353–372.

———, “Using Electoral Cycles in Police Hiring to Estimate the Effects of Police on Crime: Reply,” *The American Economic Review* 92 (2002), 1244–1250.

———, “Understanding Why Crime Fell in the 1990s: Four Factors That Explain the Decline and Six That Do Not,” *The Journal of Economic Perspectives* 18 (2004), 163–190.

LOCHNER, L., “Individual perceptions of the criminal justice system,” *The American Economic Review* 97 (2007), 444–460.

LOTT, J. R. AND D. B. MUSTARD, “Crime, Deterrence, and Right-to-Carry Concealed Handguns,” *The Journal of Legal Studies* 26 (1997), 1–68.

MARVELL, T. B. AND C. E. MOODY, "Prison Population Growth and Crime Reduction," *Journal of Quantitative criminology* 10 (1994), 109–140.

———, "Specification Problems, Police Levels, and Crime Rates," *Criminology* 34 (1996), 609–646.

———, "The Impact of Out-of-State Prison Population on State Homicide Rates: Displacement and Free-Rider Effects," *Criminology* 36 (1998), 513–536.

MASTROBUONI, G., "Police and Clearance Rates: Evidence from Recurrent Redeployments within a City," *Collegio Carlo Alberto Working Paper* (2013).

MCCRARY, J., "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment," *The American Economic Review* 92 (2002), 1236–1243.

MUSTARD, D. B., "Reexamining Criminal Behavior: The Importance of Omitted Variable Bias," *Review of Economics and Statistics* 85 (2003), 205–211.

NEWEY, W. K., K. D. WEST ET AL., "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55 (1987), 703–08.

PARKS, R. W., "Efficient Estimation of a System of Regression Equations When Disturbances Are Both Serially and Contemporaneously Correlated," *Journal of the American Statistical Association* 62 (1967), 500–509.

POGUE, T. F., "Effect of Police Expenditures on Crime Rates: Some Evidence," *Public Finance Quarterly* 3 (1975), 14–44.

POLINSKY, A. M. AND S. SHAVELL, "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature* 38 (2000), 45–76.

SEN, A., "Does Increased Abortion Lead to Lower Crime? Evaluating the Relationship between Crime, Abortion, and Fertility," *The BE Journal of Economic Analysis & Policy* 7 (2007), 1–38.

SEN, A. AND H. ARIIZUMI, "Teen Families, Welfare Transfers, and the Minimum Wage: Evidence from Canada," *Canadian Journal of Economics* 46 (2013), 338–360.

SEN, A., K. RYBCZYNSKI AND C. VAN DE WAAL, "Teen Employment, Poverty, and the Minimum Wage: Evidence from Canada," *Labour Economics* 18 (2011), 36–47.

SHEPHERD, J. M., "Fear of the First Strike: The Full Deterrent Effect of California's Two- and Three-Strikes Legislation," *The Journal of Legal Studies* 31 (2002), 159–201.

———, "Murders of Passion, Execution Delays, and the Deterrence of Capital Punishment," *The Journal of Legal Studies* 33 (2004), 283–321.

———, "Deterrence Versus Brutalization: Capital Punishment's Differing Impacts Among States," *Michigan Law Review* 104 (2005), 203–256.

SHOESMITH, B. AND C. KLEIN, "An Examination of the Impact of Police Expenditures on Arrest Rates," *Explorations (Mathemathics and Economics)* VII (2012), 106–117.

SPELMAN, W., "Jobs or Jails? The Crime Drop in Texas," *Journal of Policy Analysis and Management* 24 (2005), 133–165.

THALER, R., "An Econometric Analysis of Property Crime: Interaction Between Police and Criminals," *Journal of Public Economics* 8 (1977), 37–51.

WOLPIN, K. I., "An Economic Analysis of Crime and Punishment in England and Wales, 1894-1967," *Journal of Political Economy* 86 (1978), 815–840.

ZIMMERMAN, P. R., "State Executions, Deterrence, and the Incidence of Murder," *Journal of Applied Economics* 7 (2004), 163–193.

Appendix A

Chapter 2 Appendices

A.1 The LINC/EAL/ELSA Data Set

As there was no readily available data set including the addresses of LINC (Language Instruction for Newcomers), EAL (English as an Additional Language), and ELSA (English Language Services for Adults) providers for the time period from 2001 to 2003, I hand-collected the data. To collect the information on LINC programs I used the 2001-2002, 2002-2003, and 2003-2004 Public Accounts of Canada, Transfer Payments documents (under Citizenship and Immigration Canada: Language Instruction for Newcomers) for the names of organizations. Using the names of organizations I have recovered their postal codes from their websites where available or by using internet searches (if no website could be located). To obtain the information on ELSA programs, I used multiple 2002 and 2003 imprints of the `www.elsanet.org` using the *Internet Archive: Wayback Machine* (`https://archive.org/web/`) which contained the lists of all ELSA providers in British Columbia including their addresses. Finally, the data on EAL programs was provided at my request by Labour and Immigration Manitoba. The postal codes of EAL providers were obtained using Google Maps when necessary. The result was a data set that included 368 locations of ESL course providers with multiple locations corresponding to the same provider in some cases (e.g., a school board offering ESL courses at several school campuses). The ESL providers' postal code data was then used in conjunction with the Postal Code Conversion File (PCCF) to match the postal codes with the latitude and longitude.

A.2 Principle Component Analysis

A.2.1 Constructing the Measure of the Language Skill

As binary (or even three-level) speaking-proficiency-based measures of host-country language proficiency employed in previous studies are not suitable for examining the gradual acquisition of language skills, I construct a continuous measure of language skill by conducting principle component analysis (PCA) on the self-reported measures of English-language speaking, reading, and writing proficiencies and the responses to five additional questions related to speaking and comprehension abilities. I use the weights provided in LSIC to make the resulting factor representative of the new immigrant population in Canada in 2001. Therefore, the responses to questions related to language ability in Wave 1 are used to conduct the PCA and obtain factor loadings. These loadings are then used to construct the factor (language skill measure) for both Wave 1 and Wave 2. There is an underlying assumption that the factor loadings do not change over time. Given that approximately only a year and a half passes between the interviews the assumption is not unreasonable.

Speaking, reading, and writing proficiencies were reported on a five-point scale. The question in the LSIC survey asked: "How well can you speak/read/write (in) English?". The potential answers were: "cannot speak/read/write (in) this language," "poorly," "fairly well," "well," and "very well." I code the responses as ranging from "0" to "4" respectively. Further five questions expanded on the comprehension and speaking ability:

1. How easy is it for you to tell someone in English what your address is?
2. How easy is it for you to tell someone in English what you did before immigrating to Canada?
3. How easy is it for you to understand a message in English over the telephone?
4. How easy is it for you to tell a doctor who speaks only English what the problem is?

5. How easy is it for you to ask someone who speaks only English to rearrange a meeting with you?

Respondents had a choice between: "cannot do this," "can do this with a lot of help," "can do this with some help," and "can do this easily." I have assigned these responses values from 0 to 3 respectively.

Individuals identified as native speakers of English (those reporting English as their mother tongue and the language most spoken at home) were excluded from the sample. I first calculate a matrix of polychoric correlations (as the eight underlying variables are ordinal categorical variables). This matrix is then used to perform the principal component analysis. Table A.1a presents the Eigenvalues and the proportion of variance of the variables accounted for by the underlying factor. As evident the first factor is sufficient to serve as a measure of English-language skill. Table A.1b shows the factor loadings for the factor used as the language skill.

Table A.1: PCA of English-Language Proficiencies

(a)

Factor	Eigenvalue	Proportion of Variance
Factor 1	6.9927	0.9699
Factor 2	0.2656	0.0368
Factor 3	0.0826	0.0115
Factor 4	−0.0047	−0.0006
Factor 5	−0.0130	−0.0018
Factor 6	−0.0293	−0.0041
Factor 7	−0.0401	−0.0056

(b)

Variable	Factor Loading
Speaking	0.9432
Reading	0.9192
Writing	0.9047
Ability Question 1	0.9391
Ability Question 2	0.9500
Ability Question 3	0.9466
Ability Question 4	0.9257
Ability Question 5	0.9499

A.2.2 Constructing the Measures of Cognitive and Manual Skills

I follow Yamaguchi (2012) by *a priori* assuming which attributes and tasks in the Career Handbook measure the level of the cognitive skill and which - of the manual skill. General learning ability, numerical ability, verbal ability, clerical perception, "data" and "people" task complexities correspond to the cognitive skill. Motor coordination, finger dexterity, manual dexterity, form perception, spatial perception, and the "things' task complexity correspond to the manual skill. All abilities are measured on a five-point scale. "1" corresponds to the ability of the lowest ten percent of the population, "2" - to the lowest third of the population excluding the lowest ten percent, "3" - to the middle third, "4" - to the top third excluding the top ten percent of the population, and "5" - to the the top ten percent of the population. Data, people, and things tasks were rescaled to a four-point scale, with "0" designating no significant use of the task and "1" through "3" designating increasing complexity of performed tasks. Weighted counts of occupations within the 2001 Canadian Census of Population are used to make the generated factors representative of the Canadian native-born workers. Table A.2a shows the eigenvalues and the proportion of variance accounted for by each of the first five factors in the cognitive skill group. Table A.2b does the same for the manual skill group. The first factor for each group is retained to serve as the measure of the cognitive and manual skill respectively. This choice is consistent with the amount of variance in the variables accounted for by the first factor in each group. Table A.3 shows the factor loadings for the variables in the cognitive and manual groups.

Table A.2: PCA of Career Handbook Abilities and Tasks

(a) Cognitive Group

Factor	Eigenvalue	Proportion of Variance
Factor 1	4.0790	0.9770
Factor 2	0.2639	0.0632
Factor 3	0.0929	0.0223
Factor 4	−0.0241	−0.0058
Factor 5	−0.0985	−0.0236

(b) Manual Group

Factor	Eigenvalue	Proportion of Variance
Factor 1	3.0251	0.9101
Factor 2	0.5779	0.1739
Factor 3	0.1415	0.0426
Factor 4	−0.0593	−0.0178
Factor 5	−0.1373	−0.0413

Table A.3: Factor Loadings for the Cognitive and Manual Skills

(a) Cognitive Group

Variable	Factor Loading
General Learning Ability	0.8996
Numerical Ability	0.8520
Verbal Ability	0.9224
Clerical Perception	0.6496
“Data” Tasks	0.8420
“People” Tasks	0.7497

(b) Manual Group

Variable	Factor Loading
Motor Coordination	0.7808
Finger Dexterity	0.6656
Manual Dexterity	0.7076
Spacial Perception	0.6408
Form Perception	0.7100
“Things” Tasks	0.7464

A.3 First Stage Regressions

A.3.1 Language Evolution Equation

Table A.4 presents the estimates from the first stage regressions for the language evolution equation shown in Table 1.3. The corresponding specifications are shown above each column. The coefficients on the instrumental variables (distance to the nearest ESL provider, refugee status, and sponsored family visa status) are highly significant and have expected signs.

Table A.4: Estimates for for First Stage of the Language Skill Evolution Equation

Time in an ESL Course	(5)	(6)	(7)	(8)
Distance to the Nearest ESL Provider	-0.0010^{***} (0.0002)	-0.0010^{***} (0.0002)	-0.0010^{***} (0.0002)	-0.0010^{***} (0.0002)
Refugee Status	1.0362^{***} (0.0595)	1.0293^{***} (0.0593)	1.0285^{***} (0.0593)	1.0273^{***} (0.0592)
Sponsored Family Status	-0.0375 (0.0311)	-0.0365 (0.0310)	-0.0346 (0.0314)	-0.0386 (0.0321)
Months Between Interviews	– –	0.0454^{***} (0.0113)	0.0455^{***} (0.0113)	0.0455^{***} (0.0113)
English-proficient HH Members	– –	– –	0.0046 (0.0097)	0.0012 (0.0100)
Lag of Age	– –	– –	– –	0.0027 (0.0018)
N	–	–	–	–
R^2	0.2016	0.2062	0.2063	0.2069
Robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$				
Note: All models are unweighted, N is not displayed at present due to Statistics Canada RDC disclosure process.				

A.3.2 Wage Equation

Table A.5 presents the estimates from the first stage regression for the fixed effects instrumental variables log wage equation shown in Table 1.4. The coefficients on both instruments, the indicator for whether the interview was conducted in English and the number of household members who can speak English, are economically and statistically significant.

Table A.5: Estimates for First Stage of the Log-Wage Equation

Language Skill	Coefficient	Standard Error
Interview Conducted in English	0.1587***	(0.0517)
English-proficient HH Members	0.0447**	(0.0178)
Cognitive Skill	0.0054	(0.0186)
Manual Skill	0.0095	(0.0199)
Canadian Job Experience	0.0170**	(0.0070)
Individual Effects	Yes	
N	–	
R^2	0.2398	

Robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All models are unweighted, N is not displayed at present due to Statistics Canada RDC disclosure process.

A.3.3 Skill Transfer Equations

Table A.6 presents the estimates from the first stage regression for the fixed effects instrumental variables skill transfer equation shown in Table 1.5. The coefficients on both instruments, the indicator for whether the interview was conducted in English and the number of household members who can speak English, are economically and statistically significant. Note that the regression for cognitive and manual skill transfer share the first stage, as the language gap is the only explanatory variable.

Table A.6: Estimates for First Stage of the Skill Transfer Equations

Language Skill	Coefficient	Standard Error
Interview Conducted in English	-0.1543^{***}	(0.0504)
English-proficient HH Members	-0.0583^{***}	(0.0159)
Individual Effects	Yes	
N	–	
R^2	0.2009	

Robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All models are unweighted, N is not displayed at present due to Statistics Canada RDC disclosure process.

A.4 Bibliography

YAMAGUCHI, S., “Tasks and Heterogeneous Human Capital,” *Journal of Labor Economics* 30 (2012), 1–53.

Currirculum Vitae

Name:	Georgy Orlov
Place of Birth:	Moscow, Russia
Year of Birth:	1987
Post-Secondary Education and Degrees:	The University of Western Ontario London, Ontario, Canada 2011-2017, Ph.D. Economics
	Yale University New Haven, Connecticut, USA 2009-2011, M.A. Economics
	The University of Waterloo Waterloo, Ontario, Canada 2005–2009, B.A. Honours Economics
Honors and Awards:	SSHRC Doctoral Fellowship The University of Western Ontario, Canada 2014-2016
	Ontario Graduate Scholarship (OGS) The University of Western Ontario, Canada 2012–2013, 2013–2014, 2014–2015 (Declined) 2009–2010 (Declined due to study in the US)
	Tutorial Leader of the Year Award The University of Western Ontario, Canada 2012, 2014 (Honourable Mention)
	University of Western Ontario Graduate Fellowship The University of Western Ontario, Canada 2011–2016
	Cowles Foundation Fellowship Yale University, USA 2009–2010
	Yale University Fellowship Yale University, USA 2009–2010

SSHRC Joseph-Armand Bombardier Master's Scholarship
(Declined due to study in the US)
2009–2010

UW Alumni Gold Medal
(Top graduating student in the Faculty of Arts)
The University of Waterloo, Canada
2009

Departmental Award for Distinguished Academic Achievement
(Department of Economics)
The University of Waterloo, Canada
2009

Robin K. Banks Scholarship
The University of Waterloo, Canada
2009

Peter Swann Achievement Award in East Asian Studies
Renison College, The University of Waterloo, Canada
2009

Related Work Experience:

Teaching and Research Assistant
The University of Western Ontario
2011–2016

The University of Waterloo
2007–2009

Instructor
The University of Western Ontario
2013, 2014, 2015, 2016

Huron University College
October 2013

The University of Waterloo
2011

Publications:

"Crime, Apprehension and Clearance Rates: Panel Data Evidence from Canadian Provinces", (with Philip A. Curry & Anindya Sen), Canadian Journal of Economics, 49(2), pp. 481-514