Three Essays on Labour and Health Economics

Three Essays on Labour and Health Economics

By

QING LI, B.A. (HONOURS), M.A.

A Thesis Submitted to the School of Graduate Studies in Partial
Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2014) McMaster University
(Department of Economics) Hamilton, Ontario

Title: Three Essays on Labour and Health Economics

Author: Qing Li, B.A. (University of Western Ontario), M.A. (University of Waterloo)

Supervisors: Professor Arthur Sweetman

Professor Jeremiah Hurley

Professor Lonnie Magee

Number of Pages: xi, 135

Abstract

This thesis comprises three essays in labour economics, econometrics and health economics. The first essay uses educational quality outcomes in immigrants' home countries to explain the variation of immigrants' rates of return to education in Canada. The second essay explores an econometrics technique (i.e., generalized method of moments) that combines macro level data and micro survey data to reduce the bias and/or variance of estimates. The goal of this essay is to address nonresponse and attrition issues that are commonly encountered in surveys in health services and health economics. The last chapter is an empirical investigation on the association between diabetic patients' hospitalizations and their family doctor's payment model.

The first chapter uses international test scores as a proxy for the quality of immigrants' source country educational outcomes to explain differences in the rate of return to schooling among immigrants in Canada. The average quality of educational outcomes in an immigrant's source country and the rate of return to schooling in the host country labour market are found to have a strong and positive association. However, in contrast to those who completed their education pre-immigration, immigrants who arrived at a young age are not influenced by this educational quality measure. Also, the results are not much affected when the source country's GDP per capita and other nation-level characteristics are used as control variables. Together, these findings reinforce the argument that the quality of educational outcomes has explanatory power for labour market outcomes. The effects are strongest for males and for females without children.

The second chapter explores a technique that combines macro and micro level data. Administrative data in the health sector normally provide censuses of relevant populations but the scope of the variables is often limited and frequently only aggregate summary statistics are publically available. In contrast, survey datasets have a broader set of variables but commonly suffer from nonresponse, attrition and small sample sizes. This paper explores a technique that combines complementary population and survey data using a method of moments technique that matches auxiliary moments of the two data sources in estimating micro-econometric models. We provide Monte Carlo evidence showing that the approach can give appreciable reductions in both bias and variance. We show an example looking at midwife training and another looking at an optometrist's location of work, to illustrate its use in a health human resource context. This approach could have wide applicability in health economics and health services.

The third chapter investigates the impact of a blended capitation model (Family Health Organizations -- FHOs) compared to an enhanced fee-for-service model (Family Health Groups - FHGs) on diabetic patients in Ontario, Canada. Using comprehensive administrative data and primary care reform as a quasi-natural experiment, we construct a panel for diabetic patients and employ a difference-in-differences approach to identify the impact of a change in the general practitioner's (GP's) remuneration model on patients' hospital admissions. We find that on both the intensive and extensive margins, the hospital admissions for senior female patients statistically significantly increased after their GP's remuneration model changed from FHG to FHO. In contrast, the impacts on male patients and younger female patients were small and not statistically significant. The results provide a cautionary message regarding the differences in practice patterns towards senior diabetic patients between GPs as a function of their payment model.

Acknowledgements

I would like to thank my supervisor Professor Arthur Sweetman for providing me with constant guidance and great support throughout the past few years. I feel extremely lucky that he joined McMaster University as a professor in Economics and Ontario Research Chair in Health Human Resources in 2010 when I finished writing the core comprehensive exams and started to make a decision for my research field. I feel lucky not only because his research background perfectly matches my interest, but also because his motivation, enthusiasm, and support leads me through the rest of my Ph.D. program. He has been not only my supervisor, but also a friend who cares about me and my family in China; who invites me and other Ph.D. students to his home for BBQ; who treats me for lunches at the Student Center and discusses my research progress.

I also feel grateful to have the other two thesis committee members Professor Lonnie Magee and Professor Jeremiah Hurley, both of whom are very patient and approachable. Professor Magee recruited me to the program and taught me complex econometrics in an easy and understandable way. My second chapter would not be possible in its present form without his advice and guidance. Professor Hurley was the person who inspired me to pursue health economics as my field. As the coordinator of our Health Economics at McMaster (HEAM) group meeting, he always provided me some insight on the presentations and papers.

In addition, I am thankful to Derek K. Lobb, Associate Professor in the Medical Sciences Graduate Program, for providing the survey and the administrative data for midwifery students, which I use in an application in the second chapter. Special thanks to Professor Mike Veall, Professor Seungjin Han and Professor A. Abigail Payne for offering me research assistant positions. Thanks to Professor Jeffery Racine, Shintaro Yamaguchi, Catharine Cuff and other faculty members who provided comments and advices in my department seminar talks. Moreover, thanks

to my colleagues in the Ph.D. program for coming to my seminars, discussing research ideas, and helping me conquer difficulties I encountered these years.

Finally, I would like to express my deepest gratitude to my family, especially my mother, Ping Yin, for being the strongest support during my doctoral program, as well as all years of education since my childhood. Without her love, encouragement and support, I would never be able to make today's achievement. She picks me up when I fall, encourages me to chase my dreams, but never stopped in believing in me. To my family, I dedicate this thesis.

Preface

The Chapter 1 is co-authored with Professor Arthur Sweetman and it was published in *Labour Economics*, Volume 26, 2014, Pages 81-93. Chapter 2 is co-authored with Professor Arthur Sweetman and Professor Lonnie Magee. Chapter 3 is co-authored with Professor Arthur Sweetman. I was responsible for the empirical analysis, participated in all stages of the research, and wrote the manuscripts.

Table of Contents

Abstract iii
Acknowledgements v
Preface vii
Table of Contents viii
List of Tables x
List of Figures xi
Introduction 1
Chapter 1 The Quality of Immigrant Source Country Educational Outcomes: Do they Matter in the Receiving Country? 9
1.1 Introduction 9
1.2 Data and Descriptive Statistics 13
1.3 Empirical Strategy 17
1.3.1 Heteroscedasticity and Feasible Generalized Least Squares 20
1.4 Regression Estimates 22
1.4.1 First Stage Regression Results 22
1.4.2 Second Stage Regression Results 23
1.5 Discussion and Conclusion 28
References 30
Appendix 43
Chapter 2 Using Generalized Method of Moments to Combine Population and Survey Data 45
2.1 Introduction 45
2.2 Literature Review 47
2.3 Presentation of H&I estimator 49
2.4 Monte Carlo Simulation 51
2.4.1 Baseline Model 51
2.4.2 Extensions to Baseline Model 56

2.4.3 Limitations of the Technique 58
2.4.4 Summary of Monte Caro Simulation 60
2.5 Applications 61
2.5.1 Midwifery Program 61
2.5.2 Optometrists in Ontario 66
2.6 Conclusion 68
References 70
Appendix 91
Chapter 3 Hospitalization of Diabetic Patients, and Family Doctors in Primary Care Mixed Payment Models 94
3.1. Introduction 94
3.2. Literature Review 97
3.3. Data 100
3.3.1 Data Sources and Sample 100
3.3.2 Variable Specification 103
3.4. Empirical Method 104
3.4.1 Identification Strategies 105
3.4.2 Empirical Specifications 105
3.5. Results 109
3.5.1 Main Results 109
3.5.2 Subgroup Analysis 110
3.5.3 Sensitivity Tests 111
3.6. Discussions and Conclusion 112
3.6.1 Causality? 112
3.6.2 Conclusion and Future Research 113
References 116
Conclusion 133

List of Tables

Table 1.1: Descriptive Statistics for Males by Source Country....35
Table 1.2: Descriptive Statistics for Females by Source Country....37
Table 1.3: Test for Heteroscedasticity and Estimates of the Variance of the Error Components....39
Table 1.4: Second Stage Regressions for Male Immigrants with Alternative Weighting Schemes....40
Table 1.5: Second Stage Regressions for Female Immigrants with Alternative Weighting Schemes....41
Table 1.6: Second Stage Regression for Female Immigrants without Children....42
Appendix Table 1.7: Descriptive Statistics of First Stage Regression Variables....43
Appendix Table 1.8: Rate of Return to Education in Canadian Market by Country of Birth....44
Table 2.1: Monte Carlo Designs -- Eight Sample Schemes based on Different Variables....73
Table 2.2: Monte Carlo Designs -- OLS and Five GMM Moment Conditions....74
Table 2.3: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE from Baseline Model....75
Table 2.4: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE by Varying Sample Size....76
Table 2.5: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE by Varying Standard Error of the Error Term....77
Table 2.6: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE by Varying the Size of Heteroskedasticity....78
Table 2.7: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE from Baseline Model....79
Table 2.8: Midwife Application -- Proportions of Survey Participants from the Population....80
Table 2.9: Midwife Application -- Comparison of Different Statistics between Survey Data and Administrative Data....80
Table 2.10: Midwife Application -- Regression Results of OLS and GMM in Different Specifications....81
Table 2.11: Midwife Application -- Test of Equality of the Survey data and the Administrative Data....82
Table 2.12: Midwife Application -- Comparison of Means and Standard Deviations of Key Variables....83
Table 2.13: Optometry Application -- Regression Results of Probit and GMMs....84
Table 3.1: Key Characteristics of FHG and FHO models....123
Table 3.2: Summary Statistics of Both Patients and Physicians by Treatment Status....124
Table 3.3: Summary Statistics for the Dependent Variables....125
Table 3.4: FHO Impact on Diabetic Patients' Hospitalizations....126
Table 3.5: Impact on Male vs. Female Diabetic Patients' Hospitalizations....127
Table 3.6: Impact on Urban vs. Rural Diabetic Patients' Hospitalizations....128
Table 3.7: Impact on Diabetic Patients' Hospitalizations by GPs' Characteristics....129
Table 3.8: Sensitivity Tests on the Dynamics of the Impact....130
Table 3.9: Falsification Tests....132

List of Figures

Figure 1.1: Average Annual Earnings / Return to Education and School Outcome by Source Country..........34
Figure 2.1: Monte Carlo -- Probability Density Function of Estimates in Simple Random Samples and Endogenously Stratified Samples 85
Figure 2.2: Monte Carlo -- Probability Density Function of Estimates in Extended Models with Different Sample Sizes (Sample on y)..........86
Figure 2.3: Monte Carlo -- Probability Density Function of Estimates in Extended Models with More Independent Variables (Sample on y)..........87
Figure 2.4: Application 1 -- Histogram of Sample Weights Constructed from GMM 89
Figure 2.5: Application 1 -- Comparion of Cumulative Age Distributions between Survey Data, Weighted Survey Data and Administrative Data 90
Figure 3.1: Mean of Annual Hospital Admissions by Age..........120
Figure 3.2: Mean of Rate of Annual Hospital Admissions by Age..........121
Figure 3.3: Distribution of Switch Dates 122

Introduction

This thesis addresses policy-relevant questions through the application of economic theory and econometric techniques with a focus on the immigrants' labor market and physician payment models. It comprises of three chapters with three topics: the relationship between quality of immigrants' source country educational outcomes and labor market outcomes in a receiving country, an investigation on a generalized method of moments (GMM) technique to combine population and survey data, and the relationship between a diabetic patient's hospitalization and his or her family doctor's payment scheme. Micro-econometric approaches and techniques adopted in the thesis include parametric econometric methods, GMM, and a difference-in-differences (DiD) technique that are applied to individual level census data, survey data and administrative health data.

Immigration policy has been debated politically for decades, especially in the immigration countries which view immigrants as a key element of population and economic growth. Immigration points systems such as those in Canada and Australia frequently weight a year of education with a same score regardless of its quality. Using the Canadian labour market as a common point of reference, the first chapter empirically investigates the influence of immigrants' source country educational outcome on the return to education of immigrants in Canada. Following chapter one, the next two chapters focus on health economics. Since many surveys undertaken as part of health services and health economics research frequently encounter nonresponse and attrition, parameter estimates derived from such data without proper weights commonly suffer from bias and imprecision. Chapter two explores a GMM technique that combines complementary population and survey data to improve estimation. The effectiveness of a physician's payment scheme in improving patient health as well as incentive mechanisms in payment schemes have been debated over the past two decades. Chapter 3 uses Ontario primary care reform as a background to empirically investigate the relationship between diabetic patients' hospitalizations and their family doctors' remuneration models.

Chapter one relates to both the research literatures on education and immigrants' labour market outcomes. First, the gap in the return to education of immigrants and the domestic born is a long-standing issue and one potential explanation is due to the quality of education (Chiswick, 1978). The investigation of the relationship between immigrants' returns to education and their source country educational outcomes contributes to the literature on immigrants' labour market integration and also provides implications for host countries' immigration selection and settlement policies. Second, the quality of educational outcomes at the national level has a very substantial impact on national productivity and economic growth (Hanushek and Kimko, 2000; Barro, 2001). By using national level average test scores as a proxy for the quality of human capital, this chapter's test of its ability to explain the variation in return to education also contributes to that literature. A third related area is the debate regarding the importance of educational outcomes versus school system resource inputs for labour market productivity. Since mixed evidence are documented in the literature on the link between resource inputs and both cognitive outcomes and labour market outcomes, chapter one provides evidence on the correlation between cognitive skills at national level and the labour market outcomes.

The first chapter uses a merged sample of immigrants from the 1986, 1991, 1996 and 2001 Canadian censuses together with an index of the quality of source country educational outcomes derived from country-level scores in international standardized tests and related information by Hanushek and Kimko (2000). Other country-level data includes GDP per capita from the Penn World Trade tables (Heston, Summers and Aten, 2013) and Gini coefficients obtained from the World Bank database.

The empirical approach pursed is similar to that in Card and Krueger (1992). This approach is sometimes referred to as a random coefficient, or hierarchical linear model, which comprises two-stage regression. The first stage uses individual-level data to estimate returns to educations for each country in Canadian labour market. The second stage regression then uses the return to education slope coefficients from the first stage as the dependent variable and tries to explain its variation using the index of quality of source country educational outcomes conditional on GDP per capita and other country-level variables.

Overall, the results show that differences in the source country average quality of pre-immigration educational outcomes have substantial impacts on the Canadian labour market earnings of immigrants. The observed impact flows through the return to education, with those from source countries with higher test scores having much higher returns to education, so that the gap widens as years of schooling increases. Adding country-level controls, especially source country GDP per capita, does not appreciably alter the relationship so it is not a wealthy-country effect. Further, the return to education observed for immigrants who arrived before age 10 is not a function of his/her source country quality of educational outcomes. This reinforces the idea that it is the quality of educational outcomes, and not source country wealth effects *per se*, that is correlated with the return to education. Notably, the findings for the sample of all women differs somewhat from that for men, especially conditional on source country characteristics. However, in line with the literature on immigrant gender roles, when the sample is restricted to women who are unmarried or without children living in the household, the results are quite similar to those for men.

The second chapter is an investigation on a particular estimation technique. Statistical agencies and researchers who conduct surveys normally need to provide survey weights so that estimates can match those from the population. However, many surveys undertaken as part of health services and health economics research do not provide weights (or provide weights that are poorly designed). More importantly, these surveys usually suffer from nonresponse and attrition, so parameter estimates derived from these survey data without proper weights would result in bias and imprecision. In contrast, administrative data in the health sector normally provide censuses of relevant populations, although the scope of variables may be limited and frequently only aggregate summary statistics are publically available. The second chapter explores a technique that combines complementary population and survey data using a GMM technique that combines information from the two data sources to estimate micro-econometric models. The central feature of this approach is that it simultaneously estimates the parameters of the model of interest and matches the moments available in both the survey and aggregate population data.

The approach was developed by Imbens and Lancaster (1994), and Hellerstein and Imbens (1999), that can be interpreted as using a model-specific set of optimal weights. It extends the literature on data combination. Different from an estimator identification using two individual datasets, their approaches only require marginal information at the aggregate level from the relevant population. Particularly, they estimate micro-econometric models combining survey data with marginal information from a complementary population using a GMM technique that matches the moments of the two data sources and recovers estimated parameters of population.

Although the GMM technique has desirable large sample properties, its empirical performance in finite samples is unknown, and this is relevant since many health survey data have small sample sizes. Health administrative data may also have a relatively small number of observations and only limited information available. Because of these challenges, Monte Carlo simulations are conducted to study the properties of this GMM approach in different scenarios. The results show that the efficiency gain from using this approach can be substantial. It reduces the bias of the estimates in an endogenously stratified sample and the variance of the estimates in a simple random sample or an exogenously stratified sample. The results also show that the advantages of this technique are maintained in the face of variation in the sample sizes, explanatory power of independent variables, and the amount of heteroskedasticity in the error term. In fact, the value of the approach seems to increase with increasing heteroskedasticity.

The Monte Carlo also illustrates the technique's limitations. A practical concern is the amount of overlap in the values of the variables between the population and sample. Although we allow for a difference in distribution between two datasets, the GMM estimator may not do well in a sample that has limited overlap with population data. After exploring the small-sample properties of the technique from Monte Carlo simulations, we provide two applications. One examines student retention in Ontario's midwifery program. The weights constructed by this technique are found to change the estimates and shift sample distribution towards the population distribution. The other application is on optometrists' location of work. Estimation using the technique shows some interesting associations between optometrists' location of work, and their age and gender. We believe this approach could be widely adopted in health economics and health services.

In the third chapter, we identify Ontario patients with diabetes and empirically investigate their hospitalizations as a function of their family doctors' remuneration models. Over the past decade, Ontario has experienced a series of primary care reform in terms of physicians' remuneration models (Hutchison and Glazier, 2013, Sweetman and Buckley, 2014). Two blended funding models have become prevalent: one enhanced fee-for-service (FFS) model called a Family Health Group (FHG) and one mixed capitation model called a Family Health Organization (FHO). Although such mixture of payment mechanisms may balance the contrasting incentives to some degree, there is still little empirical evidence on how different are these models in primary care for chronic disease management. Diabetes is a prevalent and serious chronic condition that not only affects patients in terms of morbidity and early mortality but also imposes a heavy financial burden on government.

The data come from the administrative databases maintained by the Ontario Ministry of Health and Long-Term Care. They cover the full population of the province of Ontario and document all the Ontario Health Insurance Plan claims. Using the algorithm from the Institute for Clinical Evaluative Sciences, we identify diabetic patients in Ontario. Then, using the Corporate Provider Database, the Registered Persons Database, the Client Agency Program Enrolment Database and the inpatient Hospital Discharge Abstract Database, we construct a comprehensive longitudinal dataset for diabetic patients' inpatient hospitalization admissions as well as their GPs' payment models.

The study employs a difference-in-differences approach to control of both selection on observables and selection on unobserved time-invariant fixed effects to avoid the estimation bias in the identification. The treatment group is defined as those diabetic patients whose GPs switched from FHG to FHO and the comparison group is defined as those whose GPs stayed in FHG over the sampling period. The “prior to” and “after” periods depend on the intervention time that is the physicians’ switch dates in this case. Several robustness checks, sensitivity analysis and sub-group analysis are conducted.

The results show that on both intensive and extensive margin, hospital admissions for senior (older than age 65) female patients economically and statistically significantly increased after their GP’s remuneration model changed from FHG to FHO. In contrast, the impacts on male patients and younger female patients were small and not statistically significant. The results provide a cautionary message regarding to the differences in practice patterns towards senior diabetic patients between GPs in the two models.

References

Barro, R.J. (2001), "Human Capital and Growth," *American Economic Review* 91(2), 12–17.

Card, David, and Alan B. Krueger. (1992), "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy* 100(1), 1-40.

Chiswick, B.R. (1978), "The Effect of Americanization on the Earnings of Foreign-born Men," *Journal of Political Economy* 86(6), 897-921.

Hanushek, E.A., Kimko, D.D. (2000), "Schooling, Labor-Force Quality, and the Growth of Nations," *American Economic Review* 90(5), 1185-1208.

Hellerstein, J. K. and G. W. Imbens (1999), "Imposing Moment Restrictions from Auxiliary Data by Weighting," *Review of Economic and Statistics*, Vol. 81, 1-14.

Heston, A., Summers, R., Aten, B. (2013), "Penn World Table Version 7.1, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, October".

Hutchison, B., Glazier, R. (2013), "Ontario's primary care reforms have transformed the local care landscape, but a plan is needed for ongoing improvement," *Health Affairs*, 32, 695–703.

Imbens, G. W., and T. Lancaster (1994), "Combining Micro and Macro Data in Microeconometric Models," *The Review of Economic Studies*, Vol. 61, 655-680.

Sweetman, A. and Buckley, G. (2014), "Ontario's Experiment with Primary Care Reform," University of Calgary, School of Public Policy, *Research Paper* No. 7-11.

Chapter 1

The Quality of Immigrant Source Country Educational Outcomes: Do they Matter in the Receiving Country?

1.1 Introduction

It is increasingly recognized that it is beneficial for economic analyses to differentiate between the quantity of education attained (e.g., years of school or highest degree) and the quality of educational achievement (e.g., test score derived measures of cognitive ability).[1] Understanding the relationships between immigrants' formal schooling and source country-level average cognitive skills, as proxied by an index derived from multiple sets of international standardized tests, on the one hand, and labour market outcomes in the receiving country on the other, is relevant to a variety of topics. One issue involves the labour market integration of immigrants in destination country labour markets (Borjas and Friedberg, 2009; Borjas, 1995; Aydemir and Skuterud, 2005; Dustmann, Fabbri and Preston, 2005; Ferrer and Riddell, 2008). Inasmuch as the quality, or relative quality, of pre-immigration educational outcomes varies across source countries this may affect the labour

[1] Although we view test scores as reflecting the quality of an educational outcome, it is important to note that they derive from a variety of inputs including the formal education system, but also including, for example, family inputs, nutrition, and cultural norms affecting student learning effort. That is, there are many inputs to the education production function. In this paper, we assess cognitive outcomes using math and science test scores. However, 'non-cognitive' skills are increasingly viewed as being important (e.g., Cunha, Heckman, and Schennach 2010).

market integration of immigrants and have implications for receiving countries' immigrant selection and settlement policies. This is a long-standing issue; Chiswick (1978) observed a gap in rates of return to education and hypothesized that educational quality might be at issue. More recently Chiswick and Miller (2010) explore source country school quality using American data. Immigration points systems such as those in Canada and Australia, and those being considered in other countries including the US, assume (either implicitly or explicitly) that a year of education is of the same "quality" regardless of where it is obtained. However, in complementary work to that here, Ferrer, Green and Riddell (2006) use individual-level test scores to explore immigrant labour market outcomes and find that these scores explain the entire immigrant-domestic born gap in the rate of return to education.

Second, research on endogenous growth by, for example, Hanushek and Kimko (2000), Barro (2001), Erosa et al. (2010), and Barro and Lee (2012) suggest that the quality of educational outcomes as proxied by, for example, national-level average test scores, has very substantial impacts on national productivity and economic growth in contrast to measures of educational attainment or inputs – see Hanushek and Woessmann (2008) for a review. In a sub-section of their work exploring causality, using US data Hanushek and Kimko undertake an exercise similar in some aspects to that conducted here and they have broadly similar findings. Manuelli and Seshadri (2010) suggest the quality of human capital varies systematically with the level of development and find that effective human capital per worker varies substantially across countries. In accounting for differences in output per worker across countries, Schoellman (2012) demonstrates that "education quality" is roughly as important as quantity. Hanushek and Woessmann (2012) further explore this association by tracking the cognitive skill distribution within countries and over time. This paper builds on Hanushek and Kimko's index of the quality of national-level educational outcomes. Since their index is found to have predictive power in a context other than that for which it was produced, this increases the credibility of the index and their approach.

A third related area of research focuses on the importance of educational outcomes, in contrast to school system resource inputs, for labour market productivity. One reading of the literature suggests that increased inputs are sometimes associated with improved labour market outcomes, especially when the initial level of inputs is low and/or the variation in inputs is large, but that in many situations the link between resource inputs and both cognitive outcomes (i.e., test scores) and labour market outcomes is tenuous (Hanushek, 1996; Betts, 1996). Card and Krueger (1992), and Heckman, Layne-Ferrar and Todd (1996a, 1996b), use data from the US for the American born to look at the impact of educational inputs on labour market outcomes where identification comes from individuals who migrate across states. They find some evidence that inputs matter, but observe that the connection is weak. In a related vein, Bratsberg and Terrell (2002) find that source country educational inputs impact the return to education observed for immigrants to the US.

It is clear that individual-level measures of educational achievement (i.e., test scores) have very substantial (conditional) correlations with labour market success. For example, Green and Riddell (2003) study individual-level International Adult Literacy Survey (IALS) scores in relation to earnings and find a sizeable effect with these simple test scores accounting for a substantial fraction of the return to education.[2] However, the origin of the correlation is less than clear. Plausibly, individual unobserved ability contaminates both measures' relationship with labour market success. In this paper, by using immigrants' source country average levels of educational outcomes we avoid individual-level cognitive ability capturing the effects of unobserved individual-

[2] The IALS contains standard labour force information along with scores from a literacy test.

specific variables with which they may be correlated. This is also closer to the policy question that may be posed by a government considering investing in improved educational quality. That is, is there a relationship between the nations' average level of test scores (i.e., cognitive skills), and labour market outcomes?

Exploring differences in the return to education of immigrants to Canada as a function of the average quality of educational outcomes in each immigrant's source country is the objective of the present study. Overall, we find that differences in the source country average quality of pre-immigration educational outcomes have substantial impacts on the Canadian labour market earnings of immigrants. The observed impact flows through the return to education, with those from source countries with higher test scores having much higher returns to education, so that the gap widens as years of schooling increases. Adding country-level controls, especially source country GDP per capita, does not appreciably alter the relationship so it is not a wealthy-country effect. Further, the return to education observed for those immigrants who arrive before age 10 is not a function of their source country quality of educational outcome. This reinforces the idea that it is the quality of educational outcomes, and not source country effects *per se*, that is correlated with the return to education. Notably, the findings for the sample of all women differs somewhat from that for men, especially conditional on source country characteristics. However, in line with the literature on immigrant gender roles, when the sample is restricted to women who are unmarried or without children living in the household, the results are quite similar to those for men.

The remainder of this paper is structured as follows. Section II discusses the data and provides an initial descriptive analysis. Section III presents a random coefficient approach, including a test for the form of heteroscedasticity in the second stage and a feasible Generalized Least Squares strategy. Estimates from the core regression analysis are discussed in Section IV, as are those from sub-group analysis that helps in confirming and describing the phenomenon under study. Section V discusses the findings, draws conclusions and suggests options for future work.

1.2 Data and Descriptive Statistics

To undertake this analysis Canadian census data are combined with an index of the quality of source country educational outcomes derived from country-level scores from international standardized tests and related information. Hanushek and Kimko (2000) derived the index to allow international comparisons of economic growth. Their measure of the quality of educational outcomes is for 87 countries, but there are only sufficient numbers of immigrants in the Canadian census data to look at 81 of these for males, and 79 for females, with further reductions in some analyses using subsets of the sample. Further, since GDP per capita is not available for three of the countries the number in the regression analysis is reduced to 78 for males and 76 for females.

A merged sample of immigrants from the 1986, 1991, 1996 and 2001 Canadian census 20% files is employed. Combining the four increases the sample size sufficiently to allow more countries to be included in the analysis than would otherwise be possible. (A sensitivity test is conducted to see how robust the results are to the aggregation.) Census 2006 is excluded because the questions pertaining to education changed so substantially that the measurement of schooling is not comparable to that in previous censuses. The selection rules employed for the sample for analysis are that the immigrants must have been born since 1945, be at least 25 years old, and not currently attending school. Those living in the Territories are omitted, as are those with missing relevant variables. Further, immigrants from source countries with fewer than 60 observations are excluded, as are the domestic born. However, in the subgroup analysis we retain all countries with more than 10 observations, which balances several criteria including the desire to retain as large a set of countries as possible. The sample, however, contains the broadest possible set of people in the labour market; thus anyone with positive weeks of work and earnings in the year is included.[3]

Table 1.1, for males, and Table 1.2, for females, present descriptive statistics by source country. *Years of school* is measured as the sum of years of elementary and high school, university, and post-secondary non-university, and includes years from incomplete and/or multiple certifications; it is top coded at 24. Average years of schooling varies by over five years across countries, which is very substantial – equivalent to the difference between an undergraduate degree and senior high school.

Annual earnings, converted to 2001 dollars using the all goods CPI, are the sum of employment and positive self-employment earnings and are top coded at $150,000. They are presented in the subsequent columns with the averages varying markedly across source countries with the top few being about two and a half times the bottom ones for males, and two times for females. Appendix table 1.1 presents descriptive statistics for the census data and provides a listing of the background variables employed in the regressions. One note is that mother tongue, not current language spoken, is employed as a control variable since this is exogenous and is not influenced by the ability to learn new languages, which may be correlated with the quality of educational outcome variable that is the focus of the research. Also, note that the variables "age at immigration" and "domestic potential labour market experience" are used in the regressions rather than "years since migration" and "total potential labour market experience". The former have more natural interpretations given the context and also fit the data slightly better. However, sensitivity tests were conducted using years since migration instead of age at immigration to ensure robustness

[3] The findings appear to be quite robust across alternative approaches to selecting the sample for analysis. Limited experiments suggest that changing or removing the "born in since 1945" restriction makes little difference. Also, sensitivity tests limiting the sample to those with strong labour force attachment produced remarkably similar results.

and there were no substantive changes in the results.

Turning next to the test score data, the "H&K" column presents Hanushek and Kimko's (2000 - Appendix table C1) preferred QL2 measure. The underlying observed test scores from which this measure is derived are all in math and science and are only available for 37 countries. The tests were administered in the local language, which reduces concerns that QL2 is influenced by English proficiency across different countries. Further, those countries had different participation frequencies in the underlying six rounds of international testing conducted between 1965 and 1991. In particular, there are fewer observations from countries with very low scores, and wealthier countries tend to participate more often. Using these test scores as a base, Hanushek and Kimko use information regarding each country's education system (e.g., the primary school enrollment rate and teacher-pupil ratios) and demographics (e.g., population growth rates) to generate their QL2 measure. For this analysis QL2 is normalized to range from zero to one to facilitate interpretation.[4] For this paper an attempt was made to go beyond an index by mapping the score from each test to those age-specific set of individuals for whom the particular test was likely relevant (by using source country and a several year window around each test). This, however, was not fruitful since the sample sizes were too small. Also, no substantive changes to the results in this paper occurred in several experiments with Hanushek and Kimko's alternative measure, QL1.

Since it is derived from six sets of tests by two different organizations, QL2 provides a better proxy than any individual test. It also has the advantage of having been produced for previous work in the US, so it is independent of the current research and the Canadian labour market data employed. (In fact, significant results here add credibility to the index.) However, it cannot be said to be perfect for the purpose at hand. These scores are for students in grade school (up to the end of

[4] Normalizing implies rescaling the data by subtracting the lowest value from each, and then dividing by the highest.

high school or its equivalent), and postsecondary educational quality may vary differentially across countries although the averages at these two levels are likely correlated. Also, the scores are a weighted average of those for males and females, and there may be appreciable gender gaps in some countries. Finally, there are issues regarding how well the source country average test scores represent the scores of those who immigrate. If immigrants are heavily selected based on unobservables, then they may be from particular parts of each source country's distribution. Of course, if selection is similar across countries the relative scores may still be appropriate measures. In short, although this measure is the best available, it is only a proxy for a broad concept. All of these issues can be thought of as sources of measurement error. Thus, if the quality index contains mostly noise and little signal, it will likely not be correlated with the variables of interest in the Canadian census data, and the coefficients estimated in this study will probably be biased towards zero. Note, however, that the endogenous growth literature discussed above finds that national average test scores have substantial information content and are predictors of a nation's economic and productivity growth. Moreover, Schoellman (2012) argues that differences in immigrants' return to schooling in the US derive from source country education quality, and not selection in immigration or a lack of skill transferability.

One check on the QL2 measure is to compare it to subsequent international tests. In particular, QL2 is not based on the TIMSS (Third International Math and Science Survey). This is relevant since the TIMSS contains data on eight countries for which QL2 uses predictions. Hanushek and Kimko conduct a verification test and find that the measure in Tables 1 and 2 are highly correlated with the TIMSS country averages, even out of sample. This has two important implications: first, the QL2 estimates are reasonable, and second, the test score rankings are relatively stable over time. Substantial stability in rankings across the test years is also observed in the earlier data. Overall, while QL2 measures the underlying concept with error, it appears to be the best available measure of the quality of cognitive aspects of international relative educational outcomes and to contain an appreciable amount of information.

Interestingly, rank order correlations (using Kendall's tau statistic) between the test score and average years of schooling measures show no relationship for either sex (the associated p-values for males, and females, are 0.83 and 0.88, respectively). Therefore, among immigrants this piece of evidence does not suggest that countries with higher average years of completed schooling also have higher average quality as measured by these test scores. In interpreting this correlation, however, keep in mind that the standardized tests are taken in school and not at completion. In contrast, average schooling and the quality of educational outcomes are each positively correlated with average earnings by source country (as measured by Kendall-tau statistics with p-values of less than 1% in all cases). The upper plots of Figure 1.1 illustrate the relationship between QL2 and source country average earnings, demonstrating a substantial economic relationship. A shift from the 25th to the 75th percentile of the normalized QL2 distribution is associated with an approximately \$7,000 increase in unadjusted average annual earnings for the males, and about \$3,500 for the females.

1.3 Empirical Strategy

An approach similar to that in Card and Krueger (1992) – sometimes referred to as a random coefficient, or hierarchical linear, model – is pursued to explore differences in rates of return to education in Canadian labour markets as a function of our proxy for the quality of educational outcomes. It is less restrictive than the Mincer-type earnings equation approach on some dimensions though this comes at a cost. A first stage regression using individual-level data estimates each country's schooling slope coefficient and intercept as seen in equation (1).

$$\ln(w_{ij}) = \alpha' X_i + \sum_j \beta_j C_{ij} S_i + \sum_j \gamma_j C_{ij} + \varepsilon_{ij} \qquad (1)$$

In this specification, a, β and γ are sets of coefficients to be estimated; $\ln(w_{ij})$ denotes the natural logarithm of annual earnings for immigrant i born in country j; C_{ij} is an indicator which is set to unity if immigrant i is born in country j; S_i is immigrant i's years of schoolings, so $C_{ij}S_i$ is a set of country-specific measures of years of schooling; and γ_j captures the country-specific fixed effect. The control variables, *X*, are the natural logarithms of weeks and hours, an indicator for zero hours,[5] marital status, a quartic in post-immigration potential labour market experience, three census indicators, up to nine age at immigration indicators (for certain subsamples some of the age indicators are not relevant), three indicators of mother tongue (English, French, and both, with neither English nor French omitted), nine provincial indicators, and an urban indicator.[6] Statistics Canada's composite weight is used in the estimation of equation (1).

The second stage regression, equation (2),

$$\beta_j = b_0 + b_1 Quality_j + b' Z_j + \mu_j \qquad (2)$$

follows with the return to schooling coefficients from equation (1) serving as the dependent variable. *Quality*, Hanushek and Kimko's (2000) QL2, is an index of mean national educational outcomes that derives from the education system and other inputs such as parental and student effort. It might be argued that the quality indicators are proxying for source country characteristics,

[5] Hours in the Canadian Census refer to the actual number of hours that persons worked for pay or in self-employment at all jobs in the week prior to Census Day. An indicator for "zero hours" is needed since people who were in the labor force may be on vacation, sick leave, temporarily unemployed, etc.

[6] Here and throughout the analysis, the post-immigration experience measure included in the regressions is the minimum of potential experience (age-years of school-5), and years since migration. Much work in the Canadian context, especially Schaafsma and Sweetman (2001), suggests that pre-migration labour market experience has zero or negligible returns. These regressions, therefore, control for Canadian labour market experience. The age at immigration categories are: 0-5, 6-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, and 41-45; 46-51 is omitted.

and in particular its level of income, which may affect both educational outcomes and immigration patterns. To explore this possibility, Z_j, a set of country-specific characteristics are employed, including source country GDP per capita from the Penn World Trade tables (Heston, Summers and Aten, 2013). We use purchasing power parity GDP per capita at 2005 constant prices and calculate the average of GDP per capita by country from 1970 to 1991 (in $US adjusted for inflation) converted into an index with the US equal to 1. We believe an average number is useful since over short periods countries may be at different points in their business cycles or be subject to other short-term fluctuations that introduce measurement error. The ideal is a long-term measure of relative wealth/standards of living. We also explored short-term measures; they did not alter our substantial conclusions but had a lower partial R-squared in the second stage regression. Other country-level variables include: an indicator for the language of education in the source country commonly being English or French; Gini coefficients obtained from the World Bank database;[7] and continent-level indicator variables for Asia and Africa. We explored nonlinear versions of the continuous regressors, but coefficients on the quadratic terms were not statistically significant.

If it is the source country quality of educational outcomes that is driving these results, and not factors such as receiving country racial or ethnic discrimination, then immigrants educated exclusively in their source country should have effects that differ from those educated primarily in the Canadian system. The latter should not be directly affected by the source country quality index. In our base models we look at all immigrants, regardless of where they obtained their education. Extensions looking at where each person's education was obtained are then presented for each sex to increase our confidence in the interpretation of the findings.[8]

[7] http://data.worldbank.org/indicator/SI.POV.GINI/

[8] Place of birth, which is reported in the census, is assumed to be the country in which education is received if the years of schooling (plus 5) are less than the age at immigration. If the years of schooling plus 5 are greater than the age at

Although we focus on the impact of source country test scores on the return to schooling, they could also have a direct effect on earnings as pointed out by, for example, Heckman, Layne-Farrar and Todd (1996a, b), building on work by Behrman and Birdsall (1983). Thus, in principle, it is possible for the effect of our quality measure to enter through variation in the intercept if its variation is (or a component of it is) relevant for earnings but independent of the amount of schooling obtained. We explore this possibility using a regression similar to equation (2) but with γ_j from equation (1) as the dependent variable. Such a relationship might, for example, reflect a selection effect in immigration, or systematic differences in selection within the education systems of source countries with different measured test scores. It could also reflect the differential attainment of basic skills across countries that are not highly correlated with years of schooling.

1.3.1 Heteroscedasticity and Feasible Generalized Least Squares

Heteroscedasticity is a concern in the second stage regression since the countries have different sample sizes and perhaps other unobserved common characteristics. For grouped data regression models, if each error term in the individual-level data in equation (1), regardless of which group it belongs to, is independently and identically distributed (i.i.d.) with mean zero and variance ${\sigma_\varepsilon}^2$(that is, if there is no clustering as in the census), then the errors in the group level regression will have mean zero and variance ${\sigma_\varepsilon}^2/N$. Using this logic, a traditional way to deal with the heteroscedasticity in the second stage (equation 2) is to use cell counts N as weights to generate efficient and unbiased estimates. In contrast, in the presence of country-level clustering this

immigration, then some schooling is inferred to have been received post-immigration. Since gaps in educational attendance exist, but are not observed, some of those who are classified as receiving only source country schooling will have obtained some education post-immigration. This will serve to attenuate the coefficient. Errors in the other direction are probably much less common, though some immigrants who arrive in Canada at a young age undoubtedly go out of the country to receive some of their education.

approach need not be appropriate as pointed out by Dickens (1990). If people within a group share unobserved common characteristics then the error term in equation (1) should have two components as in equation (3),

$$\varepsilon_{ij} = \gamma_j + u_{ij} \qquad (3)$$

where γ_j and u_{ij} are group and individual error components respectively. If they are both i.i.d. (as well as independent of each other) with mean zero and variance σ_γ^2 and σ_u^2 respectively, then the variance of the aggregate level error term follows as:

$$Var(\mu_i) = Var(\bar{\varepsilon}_j) = \sigma_\gamma^2 + \left(\frac{\sigma_u^2}{N_j}\right) \qquad (4).$$

Dickens argues that if σ_γ^2 exists and is sufficiently large, or if group sizes are large enough, then constructing weights exclusively based on cell counts (N_j) may increase heteroscedasticity rather than adjust for it, which generates biased estimates of standard errors. Hence, it may be not worth weighting. Solon et al. (2013) re-emphasize this issue and suggest that the practitioner perform a heteroscedasticity test, in particular, a simple regression of squared OLS residuals on a constant and the inverse group size (i.e., $\frac{1}{N_j}$). If the coefficient on $\frac{1}{N_j}$ is statistically significant, then this is evidence of heteroscedasticity from the grouped data structure and they suggest a feasible Generalize Least Squares (GLS) specification. An estimate of the source-country specific variance can be constructed based on equation (4) since the estimated intercept from the regression testing for heteroscedasticity consistently estimates σ_γ^2 while the coefficient on $\frac{1}{N_j}$ consistently estimates σ_u^2. In a related context Brewer et al. (2013) emphasize the importance of the increased statistical power from feasible GLS.

For completeness, we employ four weighting strategies. The first is simply OLS, which gives equal weight to all countries implying that the second term on the right hand side of equation

(4) is set to zero, which is the preferred specification if the coefficient on $1/N_j$ is not statistically different from zero. Second, as is common in the literature we use source country sample sizes as weights, which emphasizes the second right hand side term in equation (4) and is appropriate if there is no clustering. Third, and in a similar vein to the second, we use as weights the inverse of the sampling variances of the estimated returns to schooling from the first stage, which is sometimes discussed in the literature. Finally, we use Dickens style weights based on the estimated two components in the variance of error terms, which is the preferred approach if the heteroscedasticity tests suggest that both components of equation (4) are different from zero. Also, in accord with Solon et al.'s (2013) recommendations, heteroscedasticity-robust standard error estimates are reported in all cases to account for any remaining heteroscedasticity.

1.4 Regression Estimates

1.4.1 First Stage Regression Results

Country specific returns to education from the first stage are reported in Appendix table 1.2. Models are also estimated for selected subsamples of the data, as discussed below, but only the second stage results are presented for the latter. The range of estimates is clearly quite wide, and there are substantial differences across the sexes with females having larger coefficients 72.5% of the time. The correlation between the male and female coefficients is 0.476, which is statistically different from zero with a p-value of 0.0000, however, there are some source countries, such as Thailand, for which the estimated return to education for males is quite high (0.101), while that for females is quite low (0.020). This may be related to the phenomena observed by Antecol (2000) who found a strong positive correlation between source country male-female wage gaps and those observed in the US for first generation immigrants. Also, in Canada the average return to education for females is larger than that for males.

Table 1.3 presents results from heteroscedasticity tests for the entire sample and various subsets of it. In all cases the evidence is in favor of an error component model with positive weight on both components. Clearly, the modest minimum country-level samples together with the substantial differences in cell sizes as seen in tables 1 and 2 generate appreciable heteroscedasticity, but there is also a group-specific (cluster) component. Compared to that for the entire sample, the coefficient on the $\frac{1}{N_j}$ term grows in magnitude for the subsamples, apparently because the smallest country samples are reduced, making the heteroscedasticity more serious. In the extreme, some countries are even lost when the number of individuals per country falls below our threshold for inclusion in the regressions.

1.4.2 Second Stage Regression Results

Quality's coefficient is positive and highly statistically significant, empirically important in magnitude, and although our preferred specification is (4) using GLS it is robust across specifications in the male sample as shown in Table 1.4. Increasing the quality of source country educational outcomes, via the education system or by other determinants of such outcomes, appears to substantially augment the accumulation of skills that are relevant for earnings across years of schooling. After controlling for the full set of country level variables, many of which are individually statistically significant, the magnitude of the quality of educational outcomes is only slightly reduced, which implies substantial independent variation between earnings in Canada and source country test scores. The fact that the quality relationship does not appear to be diminished by including GDP and other variables strengthens the central finding of this paper. Interestingly, the coefficient on the GDP per capita index is statistically significant, suggesting that, conditional on quality, immigrants from "high income" countries have higher returns to education, which is relevant to Manuelli and Seshadri's (2010) interest in productivity and educational quality across nations. The same is true for immigrants from Asian and African countries, although the results for English and French being used in the home country school system and income inequality in the home country have coefficients that are statistically significant in some specifications, but not in the preferred model (4).[9]

Subgroup analyses of the lower panel of table 1.4 tell an interesting story. These results are from regressions identical to those in the upper panel, except that they are for various subsets of the sample.[10] First, we select those immigrants who completed their education before immigrating. Next are those with mixed Canadian and source country education; this sample is the complement to the first. It explores a result observed by Friedberg (2000), which shows that obtaining domestic education increases wages and "undoes" some of the low return to foreign education in the Israeli context. Finally, those who arrive at a very young age – a subset of the second group – are examined in isolation since they have obtained almost all of their schooling in Canada and should not be directly affected by the source country test scores.

Looking at the results in the bottom half of table 1.4, those immigrants who arrive at an older age (for whom all their education is typically obtained in the source country) have a similar relationship between the quality of source country educational outcomes and Canadian labour

[9] It is worth noting that weighting by 1/N or Var(β) generate smaller standard errors than the preferred specification; plausibly they over-reject by virtue of putting insufficient emphasis on the term "clustering" error component. See the discussion in Dickens (1990) and Solon et al. (2013).
[10] One small difference from the earlier regressions is that some of the age at immigration indicators (which are not presented) are not relevant for some of the subgroups.

market earnings as observed in the upper panel for all immigrants.[11] In contrast, those who have Canadian and/or mixed source education are unaffected by the source country index. Apparently, in accord with Friedberg (2000), obtaining receiving country education seems to reduce or sever the relationship with home country test scores. Of course, there is endogenous selection into post-migration Canadian education. Finally, the earnings of those who arrive in Canada at a very young age are not statistically significantly affected by the index – in fact, the point estimates are frequently negative; source country educational outcome quality does not matter for those not educated in the source country. Also, although not displayed, for these groups the returns to years of schooling are quite large compared to those estimated in earlier regressions. This accords with Schaafsma and Sweetman (2001) who find that immigrants who arrive prior to age 10 have equal or greater returns to schooling than the Canadian born and immigrants who arrive later in life.

For the female sample, the coefficients on the quality measure are markedly larger than those for males unconditionally in all four specifications in table 1.5. However, unlike the males, the introduction of control variables substantially reduces the females' coefficient estimates making them slightly smaller than those for the males but with larger standard errors so that in the preferred specification they are statistically insignificant. It is not certain why there is such an appreciable difference between the conditional and unconditional coefficients across the genders, although it is quite normal for the standard errors in annual earnings regressions for females to be larger than those for males given the differences in their labour supply patterns; for example, many more females work part time. Beyond differences in female labour force attachment, one possibility, commonly discussed in the research and popular literature (e.g., Klasen, 2002; Behrman and Grant,

[11] In order to avoid many decimal points for coefficients on the educational outcome index and GDP/capita in the second stage regression, we rescaled both the dependent variable in equation (2) and the independent variable GDP/capita. Specifically, the dependent variable returns to education (i.e. betas) from the first stage regressions were rescaled by multiplying by 10 (for instance, 0.06 becomes 0.6).

2010) has to do with quite different approaches to education for women compared to men in certain nations, whereby historically females received less and/or lower quality education than males. Additionally, Blau, Kahn, and Papps (2011) focus on married women's labour supply assimilation profiles and find that they are a function of source country gender roles for an extended period post-migration. These gender differences may also have something to do with joint mobility decisions as discussed by Loprest (1992).

To explore the female coefficient gap further, we focus on females less affected by gender roles, Table 1.6 shows results for the subset of females with no unmarried children residing in their household. The unconditional estimates are very much like those for the males, although larger. More interestingly, the conditional coefficient estimates are remarkably similar to the unconditional ones. The patterns across the various subsamples in the lower half of the table are also similar to those for males. Although not shown, the results for females who are not married are broadly similar. Overall, it appears that the sub-samples of females who are either without children or not married yield results very similar to those for males, but once those who are married/with children are included in the sample, the patterns change appreciably, especially the gaps between the conditional and unconditional estimates. For female immigrants with limited marital/childcare responsibilities, the rate of return to education is clearly a function of source country test scores. Understanding more about the source of these findings, which adds an education quality dimension to the gender roles phenomena observed by Antecol (2000), and Blau, Kahn, and Papps (2011), is an interesting area for future research.

Although not shown to save space, we explore correlations between the quality index and the country-specific intercepts from equation (1). That is, we estimate equation (2) using γ_i as the dependent variable. For males the point estimates are small, negative, and not statistically significant in the preferred specification. For females the coefficient in the unconditional regression from the preferred model is negative and statistically significant, whereas conditional on the Z_i variables it is effectively zero and not statistically significant. It is feasible that the best way to interpret these results is to suggest that there is no relationship between the country-specific intercepts and the quality measure. However, since one of the four coefficients is statistically significant, and all the point estimates are negative, an alternative interpretation is to recognize that, combined with the findings in tables 4 through 6, individuals with very low years of schooling from source countries with very high test scores do poorly in terms of earnings. Potentially, the selection into higher years of schooling in high test score countries is such that those with very low years of schooling have low productivity for unobserved reasons.

While they are again not shown to save space, we performed a variety of sensitivity tests. We split the sample according to census year, and into those residing in one of three major cities, and found that the quality of source country educational outcomes has a similar effect on earnings across locations and time periods. Of course, some of these estimates are not very precise. The effects are still present when countries with a large number of observations such as the UK, the USA, or India are excluded. When weekly earnings were used as the dependent variable the statistical significance of the coefficients does not change appreciably, but the magnitude of coefficients increases. Also, restricting the sample to “full-year” workers did not alter the results substantially. Overall, the results with respect to earnings appear to be quite robust. Interestingly, we did not find any relationship between the quality measure and various measures of labour supply such as full-year status or labour force participation.

1.5 Discussion and Conclusion

Immigrants' source country educational quality—measured by Hanushek and Kimko's (2000) index based on six sets of source country test scores in math and science—is seen to be strongly positively correlated with the rate of return to education in the Canadian labour market. The index does not measure the test score, or related ability, of any individual, but reflects the quality of national-level educational outcomes, although those outcomes may have sources beyond the education system. Simple correlations and graphical analyses by source country show a substantial positive relationship between the quality of educational outcomes and average Canadian labour market earnings among immigrants. Regression analysis finds that this measure of quality seems to operate primarily through the return to education (as opposed to having a direct association with earnings). For males, adding a range of country-specific controls, and especially source country GDP per capita, does not attenuate the coefficient on quality very much. In contrast, for females the gap between the conditional and unconditional coefficient estimates is greater, and the conditional estimate is not statistically significant for the entire population. However, when the female sample is restricted to those who are without children residing in their household or who are not married, then the labour market relationship between annual earnings and the quality measure looks very much like that for males. Sensitivity tests find that quality of educational outcomes matters for those educated pre-immigration, but not for those who immigrate at a young age and obtain their education in Canada. Taken together, these extensions reinforce the idea that it is the source country quality of educational outcome that is at issue and not some other source country factors such as the average level of source-country wealth or racial/ethnic discrimination. These findings provide plausible evidence that the quality of education, as measured by test scores, has a causal impact on the rate of return to education and through it earnings. However, it remains possible that some omitted country-level characteristic is correlated with both the country-specific rates of return to education our test score index.

These findings suggest that immigrant economic integration and credential recognition are more complex than is allowed for by many analyses, which impose a common rate of return on years of schooling or some similar measure of educational attainment. In terms of credential recognition, the results in the lower half of tables 4 and 6 paint a picture of a labour market that values not only "years of school" and/or credentials, but the cognitive content of that education as measured by this index of test scores. Moreover, these results add support to Hanushek and Kimko's (2000), and Hanushek and Woessmann's (2008), claims that they have captured aspects of educational outcomes in their index that have substantial impacts in the labour market. In terms of both the endogenous growth and the educational outcomes literatures, this is consistent with the notion that the quality of national educational outcomes is associated with labour productivity.

In the future, expanding the information available on educational outcome quality would be valuable. This might include more refined indexes, for example one for sex-specific school quality, and expanding the countries for which the quality measure is available. It would also be useful to consider other aspects of quality that might affect immigrant labour market earnings. For example, advanced technologies, especially computers, are becoming increasingly important in the labour market. Undoubtedly the degree to which the most current technologies are employed varies across national education systems, even at the post-secondary level, and this may matter for labour market outcomes.

References

Antecol, H., 2000. An examination of cross-country differences in the gender gap in labor force participation rates Labour Economics 7, 409-426.

Aydemir, A., Skuterud, M., 2005. Explaining the Deteriorating Entry Earnings of Canada's Immigrant Cohorts: 1966-2000. Canadian Journal of Economics 38 (2), 641-71.

Barro, R.J., 2001. Human Capital and Growth. American Economic Review 91(2), 12–17.

Barro, R.J., Lee, J.W., 2012. A New Data Set of Educational Attainment in the World, 1950-2010. Journal of Development Economics DEVEC-01721, 1-15.

Behrman, J.R., Birdsall, N., 1983. The Quality of Schooling: Quantity alone is misleading. American Economic Review 73(4),928-946.

Behrman, J.R.. Grant, M.J., 2010. Gender Gaps in Educational Attainment in Less Developed Countries. Population and Development Review 36 (1), 71-89.

Betts, J. R., 1996. Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature. In Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success, eds. G. Burtless. Washington, DC: Brookings.

Blau, F.D., Kahn, L.M., Papps, K.L., 2011. Gender, Source Country Characteristics, and Labor Market Assimilation among Immigrants. Review of Economics and Statistics 93(1), 43–58.

Borjas, G.J., 1995. Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s? Journal of Labor Economics 13 (2), 201-245.

Borjas, G.J., Friedberg, R.M., 2009. Recent Trends in the Earnings of New Immigrants to the United States NBER Working Paper No. 15406.

Bratsberg, B., Terrell, D., 2002. School Quality and Returns to Education of U.S. Immigrants. Economic Inquiry 40 (2), 177-198.

Brewer, M., Crossley, T.F., Joyce, R., 2013. Inference with Difference-in-Differences Revisited Institute for Fiscal Studies, mimeo.

Card, D., Krueger, A.B., 1992. Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. Journal of Political Economy 100 (1), 1-40.

Chiswick, B.R., 1978. The Effect of Americanization on the Earnings of Foreign-born Men. Journal of Political Economy 86 (6), 897-921.

Chiswick, B.R., Miller, P.W., 2010. The Effects of School Quality in the Origin on the Payoff to Schooling for Immigrants, in G. Epstein and I. Gang, eds., Migration and Culture, (Frontiers of Economics and Globalization 8), Emerald Group Publishing, 67-103.

Cunha, F., Heckman, J.J., Schennach, S.M., 2010. Estimating the Technology of Cognitive and Noncognitive Skill Formation. Econometrica 78 (3), 883-931.

Dickens, W. T., 1990. Error Components in Grouped Data: Is It Ever Worth Weighting? The Review of Economics and Statistics 72 (2), 328-333.

Dustmann, C., Fabbri, F., Preston, I., 2005. The Impact of Immigration on the British Labour Market. The Economic Journal 115 (November), 324–341.

Ferrer, A., Green, D.A., Riddell, W.C., 2006. The Effect of Literacy on Immigrant Earnings. Journal of Human Resources 41 (2), 380-410.

Ferrer, A., Riddell, W.C., 2008. Education, Credentials and Immigrant Earnings Canadian Journal of Economics 41(1),186-216.

Erosa, A., Koreshkova, T., Restuccia, D., 2010. How Important Is Human Capital? A Quantitative Theory Assessment of World Income Inequality. Review of Economic Studies 77, 1421-1449.

Friedberg, R., 2000. You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. Journal of Labor Economics 18, 221-251.

Green, D., Riddell, W.C., 2003. Literacy and Earnings: An Investigation of the Interaction of Cognitive and Unobserved Skills in Earnings Generation. Labour Economics 10 (2), 165-84.

Hanushek, E.A., 1996. The Economics of Schooling: Production and Efficiency in Public Schools. Journal of Economic Literature 24 (3),1141-77.

Hanushek, E.A., Kimko, D.D., 2000. Schooling, Labor-Force Quality, and the Growth of Nations. American Economic Review 90 (5), 1185-1208.

Hanushek, E.A., Woessmann, L., 2008. The role of cognitive skills in economic development. Journal of Economic Literature 46 (3), 607-68.

Hanushek, E.A., Woessmann, L., 2012. Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation. Journal of Economic Growth 17 (4), 267-321.

Heckman, J., Layne-Ferrar, A., Todd, P., 1996a. Human Capital Pricing Equations with an Application to Estimating the Effect of School Quality on Earnings. The Review of Economics and Statistics 78 (4), 562-610.

Heckman, J., Layne-Ferrar, A., Todd, P., 1996b. Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship. in G. Burtless ed., Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success (Washington, D.C.; Brookings Institution), 192-289.

Heston, A., Summers, R., Aten, B., 2013. Penn World Table Version 7.1, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, October.

Klasen, S., 2002. Low Schooling for Girls, Slower Growth for All? Cross-Country Evidence on the

Effect of Gender Inequality in Education on Economic Development. World Bank Economic Review 16 (3), 345-373.

Loprest, P.J. 1992. Gender Differences in Wage Growth and Job Mobility. American Economic Review 82 (2), 526-32.

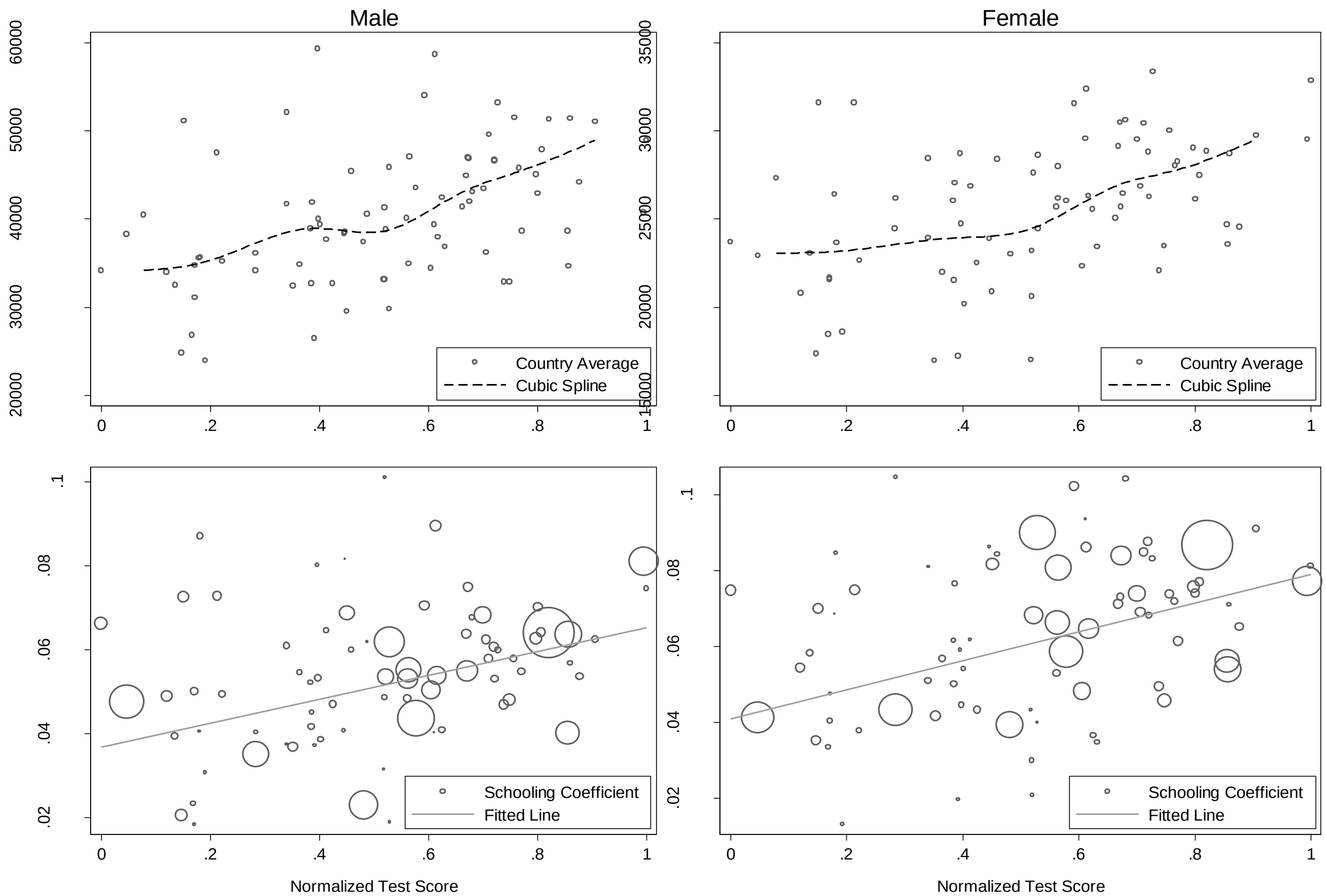
Manuelli, R. E. Seshadri, A., 2010. Human Capital and the Wealth of Nations, Mimeo, University of Wisconsin Working Paper.

Schaafsma, J., Sweetman, A., 2001. Immigrant Earnings: Age at Immigration Matters. Canadian Journal of Economics 34 (4), 1066-1099.

Schoellman, T., 2012. Education Quality and Development Accounting Review of Economic Studies 79 (1), 388-417.

Solon, G., Haider, S.J., Wooldridge, J., 2013. What Are We Weighting For? NBER working paper 18859.

Figure 1.1: Average Annual Earnings / Return to Education and School Outcome by Source Country



Note: In the upper two panels we fit cubic splines using the unadjusted data. In the lower panel we fit a regression line based on the model (4) without controls from tables 4 and 5.

Source: Canadian census data and the Hanushek and Kimko (2000) index of the quality of educational outcomes.

Table 1.1: Descriptive Statistics for Males by Source Country

	Sample Size		Years of School		Mean Earnings		Test Score	
Country	N	%	Mean	Std Dev	Mean	Std Dev	H&K	Norm
Algeria	6355	0.24	16.40	4.06	35608	26440	28.06	0.18
Argentina	9435	0.36	14.07	3.75	40130	26427	48.50	0.56
Australia	13475	0.52	14.57	3.14	49533	31399	59.04	0.76
Austria	9525	0.37	15.34	3.24	51420	33226	56.61	0.71
Barbados	9825	0.38	13.72	3.05	38649	24352	59.80	0.77
Belgium	13880	0.53	14.38	3.41	46717	29095	57.08	0.72
Bolivia	910	0.04	15.07	3.77	34698	23672	27.47	0.17
Brazil	6555	0.25	14.24	3.93	41690	28448	36.60	0.34
Cameroon	500	0.02	18.74	3.20	38610	25098	42.36	0.45
China	110130	4.24	13.64	4.72	34614	26907	64.42	0.86
Columbia	5710	0.22	14.15	3.68	34829	25002	37.87	0.36
Costa Rica	490	0.02	14.12	3.89	33169	24716	46.15	0.52
Cyprus	4370	0.17	13.48	3.87	41301	28808	46.24	0.52
Denmark	12035	0.46	13.69	3.10	47901	29712	61.76	0.81
Dominican R	1710	0.07	12.41	4.13	26526	22019	39.34	0.39
Ecuador	7095	0.27	12.76	3.36	32683	19790	38.99	0.38
Egypt	23210	0.89	16.90	3.16	51060	35374	26.43	0.15
EI Salvador	21050	0.81	12.17	4.05	24832	16490	26.21	0.15
Falkland Is	18835	0.73	14.26	3.29	33980	22838	24.74	0.12
Fiji	16315	0.63	12.63	2.95	32878	20200	58.10	0.74
Finland	8625	0.33	13.54	3.12	45784	26909	59.55	0.77
France	45510	1.75	14.97	3.51	43383	29267	56.00	0.70
Germany	99150	3.82	14.26	3.10	47111	29991	48.68	0.56
Ghana	8670	0.33	14.59	3.66	32551	21747	25.58	0.14
Greece	53950	2.08	11.43	4.22	34481	24578	50.88	0.61
Guyana	58660	2.26	13.67	3.21	37893	23509	51.49	0.62
Honduras	1450	0.06	12.46	4.19	23915	17243	28.59	0.19
Hong Kong	134005	5.16	15.23	3.48	40764	29296	71.85	0.99
Hungary	20425	0.79	14.58	3.18	45040	31315	61.23	0.80
Iceland	315	0.01	13.81	3.27	39396	22677	51.20	0.61
India	182715	7.04	13.91	4.16	38292	26806	20.80	0.05
Indonesia	4410	0.17	15.63	3.01	45444	30303	42.99	0.46
Iran	29325	1.13	15.88	3.30	34199	28632	18.26	0.00
Iraq	9730	0.37	14.07	4.04	31061	26911	27.50	0.17
Ireland	16630	0.64	14.68	3.22	54031	33188	50.20	0.59
Israel	12085	0.47	14.87	3.34	46982	35624	54.46	0.67
Italy	221500	8.53	11.92	3.92	43534	25881	49.41	0.58
Jamaica	70970	2.73	13.08	3.13	34868	22629	48.62	0.56
Japan	8565	0.33	15.11	2.86	44185	28346	65.50	0.88
Jordan	2335	0.09	14.58	3.51	38395	28236	42.28	0.45
Kenya	13280	0.51	15.77	2.95	47500	32203	29.73	0.21
Kuwait	1435	0.06	15.64	2.76	40455	30686	22.50	0.08
Luxembourg	290	0.01	13.42	2.53	40568	23591	44.49	0.49
Malaysia	12070	0.46	15.46	3.32	44890	29294	54.29	0.67
Malta	8485	0.33	12.41	3.29	46589	25018	57.14	0.72
Mauritius	5405	0.21	15.18	3.55	43040	28934	54.95	0.68
Mexico	16440	0.63	10.66	4.79	32420	25182	37.24	0.35
Mozambique	775	0.03	13.85	3.37	35514	23556	27.94	0.18
N Zealand	6795	0.26	14.94	3.16	51027	33228	67.06	0.91
Netherland	73525	2.83	13.71	3.21	46893	27886	54.52	0.67
Nicaragua	4005	0.15	14.29	3.66	26813	18018	27.30	0.17
Nigeria	4095	0.16	17.11	3.26	38913	28793	38.90	0.38

Norway	3220	0.12	14.27	3.14	51375	33002	64.56	0.86
Panama	1195	0.05	15.22	3.16	29785	21414	46.78	0.53
Paraguay	5400	0.21	11.13	3.73	39388	23543	39.96	0.40
Peru	8100	0.31	15.12	3.55	32707	24009	41.18	0.43
Philippines	102295	3.94	14.81	2.98	34190	19576	33.54	0.28
Poland	95310	3.67	14.72	3.09	38651	25026	64.37	0.86
Portugal	138220	5.32	9.49	4.15	37402	20562	44.22	0.48
S Africa	18820	0.72	16.21	3.20	58671	39504	51.30	0.61
S Korea	22205	0.86	15.61	2.78	32934	27203	58.55	0.75
Singapore	4625	0.18	15.67	2.93	48886	32609	72.13	1.00
Spain	7240	0.28	13.78	3.87	42406	27513	51.92	0.62
Sri Lanka	36770	1.42	13.49	3.42	29482	20909	42.57	0.45
Sweden	4920	0.19	15.19	3.05	53201	33850	57.43	0.73
Switzerland	12370	0.48	14.67	3.11	42891	29485	61.37	0.80
Syria	7850	0.30	13.76	4.71	35174	28752	30.23	0.22
Taiwan	13500	0.52	16.13	2.81	36159	30215	56.31	0.71
Thailand	910	0.04	13.86	3.97	33169	23245	46.26	0.52
Trin&Tobago	43855	1.69	14.13	3.07	38800	24339	46.43	0.52
Tunisia	3255	0.13	15.68	4.08	37615	29220	40.50	0.41
Turkey	8805	0.34	13.75	4.83	40040	29977	39.72	0.40
UK	410585	15.81	14.64	2.95	51323	30829	62.52	0.82
Uruguay	4620	0.18	13.33	3.44	36862	23885	52.27	0.63
USA	141655	5.45	15.34	3.43	45870	32118	46.77	0.53
USSR	23470	0.90	15.80	3.28	41907	29571	54.65	0.68
Venezuela	3280	0.13	15.05	3.60	41819	30465	39.08	0.39
Yugoslavia	41520	1.60	13.30	3.14	41379	24204	53.97	0.66
Zaire	2280	0.09	16.91	3.39	36098	26722	33.53	0.28
Zambia	1090	0.04	16.34	3.04	52072	36974	36.61	0.34
Zimbabwe	2275	0.09	16.10	2.86	59330	37495	39.64	0.40

Note: Census derived statistics use weights from Statistics Canada. Earnings are in constant $2001 adjusted using the CPI. QL2 is normalized to range from zero to one to facilitate interpretation. Canada's test score is 54.58, or 0.67 normalized.
Source: The combined 1986, 1991, 1996, and 2001 Canadian censuses, with quality measures from Hanushek and Kimko (H&K - 2000).

Table 1.2: Descriptive Statistics for Females by Source Country

	Sample Size		Years of School		Mean Earnings		Test Score	
Country	N	%	Mean	Std Dev	Mean	Std Dev	H&K	Norm
Algeria	2740	0.12	15.74	3.49	23779	19375	28.06	0.18
Argentina	7145	0.32	13.95	3.71	25623	19305	48.50	0.56
Australia	9560	0.42	14.60	2.90	29668	22358	59.04	0.76
Austria	10590	0.47	13.96	2.90	30366	21767	56.61	0.71
Barbados	10435	0.46	13.62	2.71	28187	16721	59.80	0.77
Belgium	11720	0.52	13.93	3.19	28629	21185	57.08	0.72
Bolivia	670	0.03	14.17	3.56	21788	17665	27.47	0.17
Brazil	5815	0.26	13.99	3.80	23969	17831	36.60	0.34
Cameroon	--	--	--	--	--	--	42.36	0.45
China	98720	4.36	12.47	4.48	23545	18492	64.42	0.86
Columbia	6285	0.28	13.79	3.69	21782	16192	37.87	0.36
Costa Rica	730	0.03	13.58	3.74	16713	12543	46.15	0.52
Cyprus	3445	0.15	12.11	3.64	23156	16792	46.24	0.52
Denmark	9475	0.42	13.30	2.69	27324	20084	61.76	0.81
Dominican R	1405	0.06	11.85	4.34	17306	14133	39.34	0.39
Ecuador	6175	0.27	12.69	3.37	21596	14893	38.99	0.38
Egypt	14390	0.64	11.82	4.05	17367	12678	26.43	0.15
EI Salvador	15605	0.69	15.85	2.95	31437	24374	26.21	0.15
Falkland Is	14245	0.63	13.85	3.24	20807	15845	24.74	0.12
Fiji	15065	0.67	12.02	2.60	22007	14543	58.10	0.74
Finland	8065	0.36	13.74	2.98	27750	21057	59.55	0.77
France	35355	1.56	14.96	3.23	29507	21694	56.00	0.70
Germany	84560	3.73	13.78	2.85	27794	21208	48.68	0.56
Ghana	5910	0.26	13.06	3.18	23169	17795	25.58	0.14
Greece	42260	1.87	10.26	3.96	22296	17946	50.88	0.61
Guyana	56450	2.49	13.10	2.86	26264	16760	51.49	0.62
Honduras	1345	0.06	12.57	4.03	18856	16463	28.59	0.19
Hong Kong	122855	5.42	14.18	3.37	29444	21813	71.85	0.99
Hungary	17425	0.77	14.21	2.95	28975	23721	61.23	0.80
Iceland	370	0.02	14.77	2.51	29510	21001	51.20	0.61
India	148715	6.57	13.22	4.09	22848	18376	20.80	0.05
Indonesia	3870	0.17	14.74	3.08	28155	22723	42.99	0.46
Iran	15305	0.68	15.47	3.02	23680	21041	18.26	0.00
Iraq	4270	0.19	13.51	3.82	21471	19685	27.50	0.17
Ireland	13950	0.62	14.29	2.85	31259	24080	50.20	0.59
Israel	8465	0.37	14.73	3.11	30443	24271	54.46	0.67
Italy	158160	6.98	10.99	3.85	25912	18446	49.41	0.58
Jamaica	81130	3.58	13.21	2.91	26143	17050	48.62	0.56
Japan	9125	0.40	14.82	2.53	24443	19731	65.50	0.88
Jordan	1285	0.06	13.75	2.89	23342	19054	42.28	0.45
Kenya	12720	0.56	14.78	2.80	31652	21451	29.73	0.21
Kuwait	945	0.04	15.69	2.80	27697	22309	22.50	0.08
Luxembourg	--	--	--	--	--	--	44.49	0.49
Malaysia	12415	0.55	14.13	3.34	28812	20631	54.29	0.67
Malta	6295	0.28	11.89	2.99	25923	18126	57.14	0.72
Mauritius	4550	0.20	13.89	2.91	30649	18928	54.95	0.68
Mexico	13055	0.58	11.61	4.60	16989	16198	37.24	0.35
Mozambique	585	0.03	12.91	3.73	26317	17796	27.94	0.18
N Zealand	5550	0.25	14.50	2.69	29376	23049	67.06	0.91
Netherland	52105	2.30	13.22	2.77	25562	20007	54.52	0.67
Nicaragua	3185	0.14	13.82	3.42	18537	15005	27.30	0.17
Nigeria	1850	0.08	15.84	3.14	25850	20426	38.90	0.38

Norway	2320	0.10	13.92	2.50	28538	22315	64.56	0.86
Panama	755	0.03	15.18	2.88	23949	16740	46.78	0.53
Paraguay	3865	0.17	11.03	3.24	20025	16054	39.96	0.40
Peru	7870	0.35	14.47	3.09	22412	17850	41.18	0.43
Philippines	145395	6.42	14.86	2.96	26507	16622	33.54	0.28
Poland	80800	3.57	14.51	2.92	24685	18416	64.37	0.86
Portugal	106635	4.71	9.52	4.18	22913	14672	44.22	0.48
S Africa	16180	0.71	15.10	2.86	32297	25069	51.30	0.61
S Korea	24505	1.08	14.65	2.66	23342	19412	58.55	0.75
Singapore	4745	0.21	14.68	3.08	32749	25346	72.13	1.00
Spain	4640	0.20	13.25	3.96	25604	18604	51.92	0.62
Sri Lanka	21735	0.96	13.45	2.98	20869	16627	42.57	0.45
Sweden	5145	0.23	14.69	2.96	33442	26397	57.43	0.73
Switzerland	8910	0.39	14.31	2.87	25994	21592	61.37	0.80
Syria	4520	0.20	13.30	4.22	22416	20198	30.23	0.22
Taiwan	13775	0.61	15.51	2.87	26664	23256	56.31	0.71
Thailand	2265	0.10	11.99	5.02	20528	17117	46.26	0.52
Trin&Tobago	45025	1.99	13.84	2.86	27638	17425	46.43	0.52
Tunisia	1165	0.05	14.21	3.88	26855	22360	40.50	0.41
Turkey	5140	0.23	12.99	4.62	24591	20838	39.72	0.40
UK	353905	15.63	13.91	2.62	28628	21120	62.52	0.82
Uruguay	3740	0.17	13.57	3.33	23300	16996	52.27	0.63
USA	160815	7.10	14.99	2.93	28486	23897	46.77	0.53
USSR	20745	0.92	15.53	3.21	26436	21135	54.65	0.68
Venezuela	3180	0.14	15.09	3.46	27126	22184	39.08	0.39
Yugoslavia	36205	1.60	12.51	3.33	25052	17859	53.97	0.66
Zaire	1520	0.07	14.94	3.56	24680	22071	33.53	0.28
Zambia	1100	0.05	15.18	2.84	28220	20817	36.61	0.34
Zimbabwe	1755	0.08	15.13	2.67	28408	20132	39.64	0.40

Note: Census derived statistics use weights from Statistics Canada. Earnings are in constant $2001 adjusted using the CPI. QL2 is normalized to range from zero to one to facilitate interpretation. Canada's test score is 54.58, or 0.67 normalized.
Source: The combined 1986, 1991, 1996, and 2001 Canadian censuses, with quality measures from Hanushek and Kimko (H&K - 2000).

Table 1.3: Test for Heteroscedasticity and Estimates of the Variance of the Error Components

	All immigrants	Home Country Education	Mixed Education	Canadian Education
Male				
1/N	2.768*	7.414***	6.123***	7.403***
	(1.532)	(1.858)	(1.353)	(1.861)
Constant	0.022***	0.032***	0.021*	0.040
	(0.005)	(0.010)	(0.012)	(0.033)
R^2	0.041	0.175	0.212	0.178
N	78	77	78	75
Female				
1/N	6.025***	11.109***	2.472*	4.658***
	(2.223)	(2.133)	(1.174)	(0.965)
constant	0.028***	0.034***	0.033***	0.035*
	(0.006)	(0.011)	(0.011)	(0.019)
R^2	0.090	0.268	0.057	0.252
N	76	76	76	73

Note: Standard errors in parentheses. *10% significance; **5% significance; ***1% significance.

Table 1.4: Second Stage Regressions for Male Immigrants with Alternative Weighting Schemes

	(1)		(2)		(3)		(4)	
All Male Immigrants (N=78)								
Test Score	**0.194** (0.078)**	**0.171** (0.070)**	**0.252*** (0.092)**	**0.212*** (0.051)**	**0.285*** (0.099)**	**0.255*** (0.048)**	**0.182** (0.078)**	**0.154** (0.067)**
GDP per Capita		0.199*** (0.059)		0.228*** (0.051)		0.210*** (0.055)		0.218*** (0.055)
English/ French		0.013 (0.035)		0.090*** (0.028)		0.097*** (0.027)		0.011 (0.033)
Gini		0.149 (0.200)		0.283* (0.166)		0.263* (0.151)		0.188 (0.193)
Asia		0.193*** (0.044)		0.172*** (0.042)		0.186*** (0.047)		0.181*** (0.039)
Africa		0.260*** (0.060)		0.276*** (0.075)		0.294*** (0.077)		0.262*** (0.059)
R^2	0.074	0.383	0.202	0.627	0.234	0.673	0.074	0.388
Selected Subsamples								
Only Source Country Education (N=77)								
Test	0.205**	0.226**	0.271***	0.224**	0.299***	0.280***	0.183**	0.149*
Score	(0.091)	(0.095)	(0.102)	(0.087)	(0.093)	(0.081)	(0.087)	(0.080)
R^2	0.046	0.359	0.164	0.544	0.202	0.584	0.050	0.387
Mixed Canadian and Source Country Education (N=78)								
Test	-0.035	-0.000	0.055	0.135**	0.042	0.136**	-0.016	0.060
score	(0.099)	(0.103)	(0.136)	(0.060)	(0.126)	(0.059)	(0.088)	(0.076)
R^2	0.001	0.128	0.009	0.548	0.006	0.554	0.000	0.207
Those who Arrived in Canada at Age 10 or Earlier (N=75)								
Test	-0.072	0.083	-0.116	-0.040	-0.098	-0.038	-0.065	0.085
score	(0.191)	(0.183)	(0.131)	(0.054)	(0.124)	(0.053)	(0.137)	(0.105)
R^2	0.003	0.107	0.027	0.527	0.021	0.487	0.004	0.242
Controls	NO	YES	NO	YES	NO	YES	NO	YES

Notes: Standard errors in parentheses. *10% significance; **5% significance; ***1% significance. Model 1 is unweighted OLS. Model 2 uses source country sample sizes as weights. Model 3 uses as weights the inverse of the sample variances of the estimated returns to schooling from the first stage. Model 4 weights with the estimated error term components from table 1.3. Heteroscedasticity-robust standard error estimates are reported in all cases. The first stage regressions include controls for: the natural logarithms of weeks and hours, an indicator for zero hours, marital status, a quartic in post-immigration potential labour market experience, three census indicators, up to nine age at immigration indicators (for certain subsamples some of the age indicators are not relevant), three indicators of mother tongue (English, French, and both, with neither English nor French omitted), nine provincial indicators, and an urban indicator. Experiments with various specifications for, for example, geography, hours, weeks and the like made little difference.

Table 1.5: Second Stage Regressions for Female Immigrants with Alternative Weighting Schemes

	(1)		(2)		(3)		(4)	
All Female Immigrants (N=76)								
Test score	**0.294***** **(0.083)**	**0.094** **(0.082)**	**0.381***** **(0.103)**	**0.186***** **(0.056)**	**0.376***** **(0.103)**	**0.213***** **(0.058)**	**0.293***** **(0.079)**	**0.115** **(0.080)**
GDP per capita		0.381*** (0.080)		0.387*** (0.061)		0.387*** (0.081)		0.357*** (0.078)
English/ French		0.073* (0.038)		0.087** (0.036)		0.096** (0.042)		0.065* (0.036)
Gini		0.009 (0.186)		0.257 (0.193)		0.202 (0.195)		0.040 (0.176)
Asia		0.102** (0.050)		0.030 (0.045)		0.054 (0.055)		0.089* (0.047)
Africa		0.299*** (0.052)		0.232*** (0.055)		0.275*** (0.065)		0.286*** (0.050)
R^2	0.120	0.482	0.282	0.693	0.282	0.697	0.134	0.462
Selected Subsamples								
Only Source Country Education (N=76)								
Test score	0.233** (0.098)	0.020 (0.122)	0.330*** (0.092)	0.173* (0.087)	0.322*** (0.084)	0.220*** (0.077)	0.240** (0.091)	0.099 (0.106)
R^2	0.048	0.413	0.184	0.563	0.201	0.533	0.062	0.398
Mixed Canadian and Source Country Education (N=76)								
Test score	0.160 (0.103)	0.154 (0.108)	0.280*** (0.053)	0.209*** (0.036)	0.278*** (0.052)	0.214*** (0.037)	0.162* (0.092)	0.190** (0.092)
R^2	0.032	0.056	0.315	0.426	0.323	0.447	0.039	0.061
Those who Arrived in Canada at Age 10 or Earlier (N=73)								
Test score	0.156 (0.163)	-0.019 (0.218)	0.185** (0.087)	0.132** (0.063)	0.185** (0.084)	0.128 (0.082)	0.158 (0.130)	0.080 (0.142)
R^2	0.016	0.065	0.092	0.184	0.087	0.163	0.025	0.078
Controls	NO	YES	NO	YES	NO	YES	NO	YES

Notes: Standard errors in parentheses. *10% significance; **5% significance; ***1% significance. Model 1 is unweighted OLS. Model 2 uses source country sample sizes as weights. Model 3 uses as weights the inverse of the sample variances of the estimated returns to schooling from the first stage. Model 4 weights with the estimated error term components from table 1.3. Heteroscedasticity-robust standard error estimates are reported in all cases. The first stage regressions include controls for: the natural logarithms of weeks and hours, an indicator for zero hours, marital status, a quartic in post-immigration potential labour market experience, three census indicators, up to nine age at immigration indicators (for certain subsamples some of the age indicators are not relevant), three indicators of mother tongue (English, French, and both, with neither English nor French omitted), nine provincial indicators, and an urban indicator. Experiments with various specifications for, for example, geography, hours, weeks and the like made little difference.

Table 1.6: Second Stage Regression for Female Immigrants without Children

	(1)		(2)		(3)		(4)	
All Female Immigrants without Children (N=76)								
Test score	**0.229***** **(0.083)**	**0.269**** **(0.104)**	**0.340***** **(0.083)**	**0.236***** **(0.037)**	**0.332***** **(0.074)**	**0.247***** **(0.029)**	**0.223***** **(0.072)**	**0.231***** **(0.086)**
GDP per capita		0.023 (0.112)		0.185*** (0.049)		0.195*** (0.05)		0.052 (0.094)
English/ French		0.095** (0.040)		0.100*** (0.026)		0.103*** (0.023)		0.095** (0.037)
Gini		0.139 (0.244)		0.103 (0.146)		0.084 (0.130)		0.083 (0.215)
Asia		0.091 (0.055)		0.066** (0.033)		0.070** (0.033)		0.088* (0.048)
Africa		0.152* (0.085)		0.231*** (0.043)		0.242*** (0.045)		0.174** (0.077)
R^2	0.066	0.218	0.334	0.632	0.321	0.642	0.078	0.254
Selected Subsamples								
Only Source Country Education (N=75)								
Test score	0.229* (0.127)	0.304** (0.146)	0.445*** (0.097)	0.310*** (0.045)	0.385*** (0.093)	0.333*** (0.048)	0.256*** (0.096)	0.260** (0.103)
R^2	0.028	0.223	0.312	0.641	0.237	0.592	0.053	0.288
Mixed Canadian and Source Country Education (N=75)								
Test score	0.059 (0.170)	0.170 (0.165)	0.178* (0.097)	0.172** (0.069)	0.123 (0.083)	0.141** (0.059)	0.069 (0.146)	0.189 (0.140)
R^2	0.002	0.064	0.077	0.208	0.042	0.157	0.003	0.072
Those who Arrived in Canada at Age 10 or Earlier (N=68)								
Test score	-0.148 (0.285)	-0.179 (0.414)	-0.016 (0.093)	0.009 (0.083)	-0.029 (0.077)	0.007 (0.089)	-0.114 (0.243)	-0.105 (0.343)
R^2	0.007	0.104	0.000	0.085	0.001	0.054	0.005	0.080
Controls	NO	YES	NO	YES	NO	YES	NO	YES

Notes: Standard errors in parentheses. *10% significance; **5% significance; ***1% significance. Model 1 is unweighted OLS. Model 2 uses source country sample sizes as weights. Model 3 uses as weights the inverse of the sample variances of the estimated returns to schooling from the first stage. Model 4 weights with the estimated error term components from table 1.3. Heteroscedasticity-robust standard error estimates are reported in all cases. The first stage regressions include controls for: the natural logarithms of weeks and hours, an indicator for zero hours, marital status, a quartic in post-immigration potential labour market experience, three census indicators, up to nine age at immigration indicators (for certain subsamples some of the age indicators are not relevant), three indicators of mother tongue (English, French, and both, with neither English nor French omitted), nine provincial indicators, and an urban indicator. Experiments with various specifications for, for example, geography, hours, weeks and the like made little difference.

Appendix

Appendix Table 1.7: Descriptive Statistics of First Stage Regression Variables

	Male		Female	
Variable	Mean	Std. Dev.	Mean	Std. Dev
Annual earnings	41992	28294	26317	19870
ln(earnings)	10.348	0.969	9.806	1.082
Weeks of work in the census year	46.83	10.32	44.26	12.86
Hours of work census week	39.65	17.61	31.05	17.45
Zero hours of work (census week)	0.099	0.299	0.153	0.360
Currently Married	0.750	0.433	0.754	0.431
Age	39.469	7.779	39.267	7.753
Potential Canadian experience	15.605	8.447	15.778	8.488
Immigrant Age at Arrival:				
0 to 5	0.152	0.359	0.142	0.349
6 to 10	0.112	0.315	0.104	0.305
11 to 15	0.092	0.289	0.087	0.282
16 to 20	0.135	0.342	0.158	0.365
21 to 25	0.203	0.402	0.221	0.415
26 to 30	0.153	0.360	0.144	0.351
31 to 35	0.082	0.274	0.077	0.267
36 to 40	0.043	0.202	0.041	0.199
41 to 45	0.022	0.145	0.019	0.136
46 to 50	0.006	0.078	0.005	0.070
51 to 65	0.001	0.025	0.000	0.020
Urban	0.855	0.352	0.859	0.348
BC	0.178	0.382	0.181	0.385
AB	0.090	0.286	0.091	0.287
SK	0.009	0.095	0.009	0.096
MN	0.030	0.171	0.031	0.173
ON	0.567	0.495	0.577	0.494
PQ	0.108	0.310	0.095	0.293
NB	0.005	0.071	0.006	0.074
NS	0.008	0.092	0.008	0.088
PI	0.001	0.029	0.001	0.029
NF	0.002	0.047	0.002	0.042
Mother Tongue:				
English	0.358	0.479	0.378	0.485
French	0.027	0.161	0.024	0.152
Both	0.031	0.174	0.031	0.173
Neither	0.584	0.493	0.568	0.495
Years of school	13.897	3.850	13.451	3.606
Census				
2001	0.343	0.475	0.364	0.481
1996	0.270	0.444	0.249	0.432
1986	0.210	0.407	0.213	0.409
1981	0.177	0.381	0.174	0.379

Notes: The descriptive statistics are weighted and the number of weighted observations for males is 2,596,680, for females 2,264,710. Dollars are in 2001 equivalents.
Source: 1986, 1991, 1996 and 2001 Canadian Censuses.

Appendix Table 1.8: Rate of Return to Education in Canadian Market by Country of Birth

	Males		Females			Males		Females	
Country	Coef	Std error	Coef	Std error		Coef	Std error	Coef	Std error
Algeria	0.087	0.011	0.084	0.010	Kuwait	0.090	0.025	0.084	0.020
Argentina	0.048	0.005	0.053	0.007	Luxembourg	0.062	0.028	na	
Australia	0.058	0.007	0.071	0.007	Malaysia	0.064	0.006	0.068	0.005
Austria	0.058	0.005	0.086	0.007	Malta	0.053	0.006	0.065	0.011
Barbados	0.055	0.006	0.062	0.008	Mauritius	0.068	0.008	0.103	0.013
Belgium	0.061	0.004	0.087	0.006	Mexico	0.037	0.004	0.042	0.005
Bolivia	0.018	0.015	0.047	0.028	Mozambique	0.041	0.020	0.070	0.024
Brazil	0.061	0.007	0.052	0.007	N Zealand	0.062	0.007	0.094	0.010
Cameroon	0.082	0.026	na		Netherland	0.055	0.002	0.083	0.003
China	0.064	0.001	0.053	0.002	Nicaragua	0.023	0.007	0.036	0.014
Colombia	0.055	0.007	0.055	0.008	Nigeria	0.052	0.010	0.060	0.017
Costa Rica	0.031	0.024	0.045	0.025	Norway	0.057	0.010	0.069	0.015
Cyprus	0.049	0.009	0.030	0.011	Panama	0.019	0.021	0.034	0.023
Denmark	0.064	0.007	0.078	0.008	Paraguay	0.039	0.007	0.052	0.010
Dominic R	0.037	0.013	0.020	0.016	Peru	0.047	0.009	0.045	0.008
E Salvador	0.021	0.004	0.035	0.005	Philippine	0.035	0.002	0.045	0.002
Ecuador	0.042	0.008	0.050	0.011	Poland	0.040	0.002	0.057	0.003
Egypt	0.072	0.004	0.069	0.006	Portugal	0.023	0.001	0.039	0.002
Falkland I	0.049	0.006	0.054	0.008	S Africa	0.090	0.004	0.086	0.006
Fiji	0.047	0.007	0.049	0.008	S Korea	0.048	0.005	0.045	0.006
Finland	0.023	0.007	0.071	0.008	Singapore	0.075	0.010	0.080	0.010
France	0.068	0.003	0.074	0.003	Spain	0.041	0.007	0.036	0.009
Germany	0.055	0.002	0.080	0.003	Sri Lanka	0.069	0.004	0.080	0.006
Ghana	0.039	0.007	0.057	0.013	Sweden	0.060	0.009	0.083	0.009
Greece	0.050	0.002	0.048	0.003	Switzerland	0.070	0.007	0.074	0.008
Guyana	0.054	0.003	0.063	0.003	Syria	0.049	0.007	0.037	0.011
Honduras	0.031	0.013	0.011	0.015	Taiwan	0.063	0.007	0.068	0.007
Hong Kong	0.081	0.002	0.078	0.002	Thailand	0.101	0.050	0.020	0.008
Hungary	0.063	0.004	0.079	0.005	Trin&Tobag	0.053	0.003	0.068	0.004
Iceland	0.040	0.020	0.097	0.024	Tunisia	0.065	0.009	0.062	0.019
India	0.048	0.001	0.041	0.002	Turkey	0.053	0.005	0.045	0.008
Indonesia	0.060	0.009	0.083	0.012	UK	0.064	0.001	0.086	0.001
Iran	0.066	0.004	0.074	0.007	Uruguay	0.025	0.007	0.033	0.012
Iraq	0.050	0.008	0.040	0.009	USA	0.062	0.002	0.090	0.002
Ireland	0.070	0.005	0.097	0.007	USSR	0.042	0.005	0.045	0.004
Israel	0.075	0.005	0.073	0.006	Venezuela	0.045	0.017	0.077	0.014
Italy	0.044	0.001	0.058	0.001	Yugoslavia	0.030	0.003	0.045	0.003
Jamaica	0.053	0.002	0.066	0.003	Zaire	0.040	0.013	0.103	0.027
Japan	0.054	0.008	0.066	0.010	Zambia	0.038	0.051	0.081	0.018
Jordan	0.041	0.010	0.084	0.022	Zimbabwe	0.080	0.016	0.060	0.017
Kenya	0.073	0.006	0.074	0.006					
Obs.						2596680		2264710	
R^2						0.274		0.348	

Notes: Also included in the regression are the control variables Appendix Table 1, and a full set of source country intercepts.

Chapter 2

Using Generalized Method of Moments to Combine Population and Survey Data

2.1 Introduction

In health economics and health services research it is common for researchers focusing on specific target populations to conduct surveys or related forms of data collection. Those under study might include, for example, physicians, optometrists, or other healthcare providers in a region, those discharged from hospital meeting particular criteria, graduates from an educational program, or participants in a clinical trial or epidemiological tracking study. Such surveys frequently have small sample sizes and suffer from nonresponse and/or attrition. As a result, parameter estimates derived from such data without proper weights suffer from bias and imprecision.

In many situations, aggregate information regarding the target population is available. This aggregate information can be employed alongside the survey data to help reduce bias and imprecision of parameter estimates using an approach developed by Imbens and Lancaster (1994) (hereafter I&L), and Hellerstein and Imbens (1999) (hereafter H&I) that can be interpreted as using a model-specific set of optimal weights. The central feature of this approach is that it simultaneously estimates the parameters of the model of interest and matches the moments available in both the survey and aggregate population data in a generalized methods of moments (GMM) framework. As a result, the precision of the estimates is improved and their bias is reduced. However, this approach does not address endogenous selection out of the survey sample. (I.e. although it does address nonresponse that is a function of observed characteristics, it does not solve issues related to sample selection as a function of unobserved ones.) In the terminology of Rubin (1976), conditional on observables it makes a missing at random (MAR) assumption.

H&I show that this GMM technique has desirable large sample properties in some contexts, but its small sample empirical performance is unknown. We conduct Monte Carlo simulations to study the properties of this GMM approach in different health-relevant scenarios. In the context of applied health economics and health services research, the combination of information from modest size surveys with aggregate information using the GMM approach allows higher-quality inference than would be feasible using only survey data, as will be seen in a series of Monte Carlo simulations. To illustrate the method, an application studying the propensity of students to drop out of midwifery programs in Ontario is undertaken. Permission was obtained to survey graduates of the program using the administrative data as a survey frame, but each respondent was guaranteed anonymity, therefore the link between the administrative data and the survey responses is lost. In this application, individual-level administrative microdata are not available for privacy/confidentiality reasons, but limited aggregate information from university registration records can be made available. Another application is also given, involving a survey of optometrists in the province of Ontario that was distributed to members by the relevant provincial associations in which each respondent is again guaranteed anonymity. However, aggregate information regarding the population is available from the relevant regulatory authorities. These aggregate data are employed using the technique described here as well.

The remainder of the paper is organized as follows. Section 2 presents a literature review and section 3 summarizes the estimator proposed by H&I. Section 4 summarizes the findings of the Monte Carlo simulation study. Section 5 presents the two applications addressing dropout rates among midwifery students, and whether to choose to work in rural areas among optometrists. Section 6 concludes.

2.2 Literature Review

In economics, this GMM technique is associated with the literature addressing biased survey data from exogenous or endogenous stratification. This dates to Manski and Lerman (1977), who proposed a weighted maximum likelihood estimator for endogenously stratified samples. Manski and McFadden (1981), Cosslett (1981) and Imbens (1992) subsequently devised GMM estimators addressing endogenous stratification. Imbens and Lancaster (1996) proposed a method for other stratified samples such as multinomial sampling, standard stratified sampling, or variable probability sampling. Ramalho and Ramalho (2006) provided a bias-corrected moment-based estimator for endogenous stratified sampling.

Our estimator is particularly related to the econometrics literature on data combination (See Ridder and Moffitt (2007) for a systematic review). Hsieh et al. (1985) provided estimation methods for which a combination of observational data and auxiliary information suffices in principle to identify response probabilities. Lancaster and Imbens (1996) considered a contaminated sampling scheme, which has a random sample with both the dependent variable and covariates available, and second random sample with the dependent variable unknown. Devereux and Tripathi (2009) developed a semiparametric estimator to combine censored and uncensored samples.

The technique is also related to both the econometrics and statistics literature on missing data, which arises from nonresponse or partial response to survey questions, and the literature on combining two datasets, particularly on weighting for nonresponse based on auxiliary information. Rubin (1976) introduced missing at random (MAR) and missing completely at random as benchmark mechanisms. Gourieroux and Monfort (1981) investigated a linear model with missing independent variables. Among others, Griliches, Hall and Hausman (1978) challenged the MAR assumption and addressed the possibility of bias in estimated coefficients with missing data from self-selection, which is of course the motivation for Heckman's (1976, 1979) well-known estimator and the substantial literature that has ensued addressing the identification of causality in observational data (surveys include Heckman, 2008; Imbens and Wooldridge, 2009). These papers are directly relevant since the technique used in this paper assumes MAR conditional on observables. If the missing data are a function of a variable whose marginal information is available, this technique can reduce the bias.

In statistics literature, a post-stratification weighting technique using auxiliary information, was proposed in a seminal paper by Holt and Smith (1979), who demonstrate that this technique improves the precision of estimates. Little (1986, 1993) further investigates post-stratification reweighting and develops a Bayesian model-based theory. Bethlehem (1988) show that proper use of auxiliary information can reduce nonresponse bias studies from studying the characteristics of the Horvitz-Thompson estimator (Horvitz and Thompson, 1952), a generalized least squares (GLS) estimator and a post-stratification estimator. One essential feature of these methods is the requirement of the knowledge on sample sizes (as well as response rates) in specified strata in both the sample and population.

I&L's and H&I's approaches extend the literature on data combination. Their approaches only require marginal information (i.e. macro-level moments) from the population whereas the older estimators obtain identification using two individual level datasets. Specifically, both the approaches estimate micro-econometric models combining survey data with aggregate information from a complementary population using a GMM technique that matches the moments of the two data sources and recovers estimates of the population parameters. While both I&L and H&I allow for biased samples, H&I relax parametric assumptions made by I&L so that additional moment conditions are not necessarily functions of parameters in micro-econometric models. Moreover, I&L assume that the marginal information from the population data has no sampling error, while H&I allow for it.

2.3 Presentation of H&I estimator

Suppose a sample comprises $\{z_1, z_2, \ldots, z_n\}$, which are realizations of the random variable z=(y, X) where y is the scalar dependent variable and X is a vector of independent variable with k elements. In additional to the sample, suppose we have information that reflects the distribution of the population (for example, we might have a mean or know the value of y^* where $y^* = E(y|X > 0)$). As an estimator for comparison, the ordinary least squared (OLS) estimator is unbiased but inefficient since there exist potential efficiency gains from using such macro information that reflects the distribution of the population. In contrast, a more efficient weighted least square (WLS) estimator, which is equivalently expressed as a moment estimator (M-estimator), was proposed by H&I and proved to have desirable asymptotic properties. The WLS estimator solves a Lagrange equation where the estimator's moments are subject to auxiliary moments that effectively reweight the sample. The estimator (β, λ) solves the population moment condition

$$\rho(y,x,\beta,\lambda) = E\begin{bmatrix}\rho_1(y,x,\beta,\lambda)\\ \rho_2(y,x,\lambda)\end{bmatrix} = E\begin{bmatrix}x(y-\beta' x)/\left(1+\lambda' h(y,x)\right)\\ h(y,x)/\left(1+\lambda' h(y,x)\right)\end{bmatrix} = 0. \tag{1}$$

The equation $h(y,x) = h^s(y,x) - h^p$ is the link between $h^s(y,x)$, the vector of moments based on the *sample* data available (e.g., survey data), and h^p, the *population* moments (e.g., from administrative or census sources). For instance, if we have knowledge of the population mean of the explanatory variable age which equals to Age^p, $h(y,x)$ can be constructed as $\sum_{i=1}^{n} h(y_i,x_i) = \sum_{i=1}^{n}[h^s(y_i,x_i) - h^p] = \sum_{i=1}^{n}(Age_i - Age^p) = 0$. More generally, $h(y,x)$ may represent a broad range of complex restrictions including independent and dependent variables, and combination thereof.

This system of equations comprises two types of moment conditions. First, ρ_1 is a vector of the standard OLS moments weighted by a function of the Lagrange coefficients. Second, the vector ρ_2 matches the sample and population moment conditions using weights that are specific to each dataset and specification used in estimation. The first set, ρ_1, can easily accommodate other estimators besides OLS. For instance, I&L use a probit model in their application where the numerator of ρ_1 is replaced by the first order conditions of that estimator's log-likelihood function. Given the formulation as the solution to the Lagrangian, the term $1/\left(1+\hat{\lambda}' h(y_i,x_i)\right)$ serves as a set of weights that reweight the sample distribution to approximate the population distribution. Formally, this model is just identified, so it is a method of moments (MM) rather than GMM estimator. However, since H&I use the GMM terminology, and since MM is a special case of GMM, we refer to it as GMM in this paper.

Unlike the I&L model, an advantage of the H&I technique is that it allows for sampling error in the population moments. But this also introduces limitations. The probability limit of the H&I estimator can only be interpreted as being the population parameter from an artificial population, although the distribution of this artificial population is closest to the true data generating process in an empirical likelihood sense. However, when the population is small or the marginal information is imprecise, the GMM estimator was shown by H&I to increase the variance of the estimates relative to using the relevant simple (e.g., OLS) estimator. Health administrative data may be small and have a distribution dramatically different from survey data. In practice, researchers need to be aware of the quality of the marginal information and the extent of overlap between two datasets. Also, as pointed out by Card, Hildreth, and Shore-Sheppard (2004), the survey data may not actually be drawn from the population it is purported to be from. What marginal information from administrative data is relevant is a judgment call. Facing response bias, one has to make assumptions on what type of information is most relevant in terms of constructing good moment conditions.

2.4 Monte Carlo Simulation

2.4.1 Baseline Model

2.4.1.1 Monte Carlo Design

The data generating process (DGP) for the population in the baseline model is

$$y_i = \beta_0 + \beta_1 d_i + \beta_2 x_i + e_i \text{ , for } i = 1, \dots, N_p. \tag{2}$$

$$\text{where } \beta_0 = 1 \text{ , } \beta_1 = 0.5 \text{ , } \beta_2 = -0.5$$

d is a dummy variable that equals to 0 with probability 0.5 and equals to 2 with probability 0.5; $x \sim iid\ N\ (0, 1)$ and $e \sim iid\ N\ (0, 1)$. A large population with $N_p = 100{,}000$ observations (y_i, d_i, x_i) is generated according to the DGP and is treated as the population data. We set the standard deviations of d_i and x_i to one which equalizes the influence of each regressor on y.[12] The population data is generated based on two schemes: one with uncorrelated regressors and one where they have a correlation of 0.5.

First, the population first moments (means) and second moments (including cross-products) are calculated as:

$$Y = \frac{1}{N_p}\sum_{i=1}^{N_p} y_i\ , D = \frac{1}{N_p}\sum_{i=1}^{N_p} d_i,\ X = \frac{1}{N_p}\sum_{i=1}^{N_p} x_i,$$

$$YY = \frac{1}{N_p}\sum_{i=1}^{N_p} y_i^2, XX = \frac{1}{N_p}\sum_{i=1}^{N_p} x_i x_i,$$

$$DY = \frac{1}{N_p}\sum_{i=1}^{N_p} d_i y_i, XY = \frac{1}{N_p}\sum_{i=1}^{N_p} x_i y_i, DX = \frac{1}{N_p}\sum_{i=1}^{N_p} d_i x_i,\ i = 1, \dots, N_p \tag{3}$$

Second, after these moments are calculated and stored, small samples $\mathrm{N_S} = 500$ are generated from this single simulated large population using eight sampling schemes as described in Table 2.1. The first scheme is simple random sampling (SRS) in which 500 observations are randomly drawn from the population. The remaining schemes capture different forms of nonresponse. Some schemes involve purely exogenous stratification, using unequal probability sampling on d (or x). In this scenario, we randomly draw 100 observations with $\mathrm{d} = 0$ (or $\mathrm{x} < 0$) and 400 observations with $\mathrm{d} = 2$ (or $\mathrm{x} > 0$). Other schemes involve endogenous stratification, following a similar rule, but also involving the variable y. With the first and second population moments calculated, the corresponding sample moment conditions (ρ_2 from equation (1)) are then constructed as

$$\frac{1}{\mathrm{N_S}}\sum_{i=1}^{\mathrm{N_S}} \widehat{w}_i\left(h_1^s(y_i, x_i) - h_1^p\right) = \frac{1}{\mathrm{N_S}}\sum_{i=1}^{\mathrm{N_S}} \widehat{w}_i(y_i - Y) = 0,$$

$$\frac{1}{\mathrm{N_S}}\sum_{i=1}^{\mathrm{N_S}} \widehat{w}_i\left(h_2^s(y_i, x_i) - h_2^p\right) = \frac{1}{\mathrm{N_S}}\sum_{i=1}^{\mathrm{N_S}} \widehat{w}_i(d_i - D) = 0,$$

[12] This is why the dummy variable d is set to be 0 and 2 (rather than 0 and 1) with an equal probability.

$$\vdots \tag{4}$$

$$\frac{1}{\mathrm{N_S}}\sum_{i=1}^{\mathrm{N_S}}\widehat{w}_i\left(h_8^s(y_i,x_i)-h_8^p\right)=\frac{1}{\mathrm{N_S}}\sum_{i=1}^{\mathrm{N_S}}\widehat{w}_i(d_ix_i-DX)=0,$$

$\widehat{w}_i$ is the weight function expressed as $1/\left(1+\hat{\lambda}'h(y_i,x_i)\right)$ in (1). Because the denominator of this weight function is linear combination of $\hat{\lambda}'h(y_i,x_i)$, it may generate negative weights when a sample is very unrepresentative. Hence, we use $\widehat{w}_i=1/e^{\hat{\lambda}'h(y_i,x_i)}$ which is always positive.[13] Using samples generated from the eight sampling schemes and moment conditions obtained from the population, we run OLS and GMM regressions as follows,

OLS: $$0=g(\hat{\beta}_{ols})=\frac{1}{N_S}\sum_{i=1}^{N_S}\left[x_i\left(y_i-\hat{\beta}_{ols}'x_i\right)\right], \tag{5}$$

GMM: $$0=g(\hat{\beta}_{mm},\hat{\lambda})=\frac{1}{N_S}\sum_{i=1}^{N_S}\begin{bmatrix}\widehat{w}_ix_i\left(y_i-\hat{\beta}_{mm}'x_i\right)\\ \widehat{w}_i(h^s(y_i,x_i)-h^p)\end{bmatrix}, \tag{6}$$

where $x_i=(1,d_i,x_i)$ and $\widehat{w}_i$ is defined above.

We compare the results of coefficient estimates of each parameter from six regressions (i.e. OLS and five different GMM specifications, which are described in Table 2.2). GMM6 is not included because so much macro information is added in this model that the coefficients can be perfectly estimated by these moment conditions alone, without using the sample data.

We set up a program and simulate it for 1000 replications, so the sampling distributions and mean squared errors (MSE) of the six estimators are based on these replications. The MSE "for the model" is defined as the sum of the MSEs of $\hat{\beta}_1$ and $\hat{\beta}_2$.[14]

[13] Other candidates of the weight function such as $\widehat{w}_i=1/e^{[1+\hat{\lambda}'h(y_i,x_i)]}$ and $\widehat{w}_i=e^{1/[1+\hat{\lambda}'h(y_i,x_i)]}$ make little difference to the final results.
[14] We also estimated the variances of the simulated MSEs and MSE ratios. They are not reported, but are reassuringly small. The estimated standard deviations are less than 0.01. The formulas are shown in the Appendix.

2.4.1.2 Results for Baseline Model

The performance of each of the five GMM specifications for each of the eight sampling schemes is measured by the ratios of the GMM MSE to the OLS MSE, reported in Table 2.3. The ratio reflects how well each GMM performs relative to unweighted OLS. A clear pattern for each sampling scheme is that the estimators perform better moving from GMM1 to GMM5.[15] (GMM2 and GMM3 are similar.) This shows that including more macro moment conditions can decrease the MSE of a GMM model. GMM1 generates a larger MSE than that of OLS in most sampling schemes except in endogenous sampling on y or sometimes in a sampling scheme that involves y.

Since GMM1 may perform worse than OLS, and any efficiency gain from GMM1 is relatively small, it appears necessary to use a dependent variable in this model when constructing desirable moment conditions unless there are endogenous selections. It is consistent with H&I's finding, that using only independent variables to construct the weight function implies an artificial distribution in which the conditional density of the dependent variable has the same distribution as the unweighted one, whereas using dependent variable in constructing weights can change the distribution of the conditional density. This point relates to Deaton's (1995) observation that when no heterogeneity is present (i.e., no endogenous selection into strata), unweighted regression is preferred. Moreover, GMM2 and GMM3 have similar MSEs, showing that information from the dummy variable and the continuous variable are roughly equally important in our example. The GMM technique provides greater improvement over OLS in a sample that is purely or partially generated by the endogenous variable as compared to a simple random sample or exogenously stratified sample. The results are not sensitive to whether or not a correlation exists between d and x.

[15] It suggests that GMM5 does well because that data provide almost no information.

The probability density functions (*pdfs*) in Figure 2.1 show the sampling distributions of the estimators and provide a different way of seeing the bias and variance of a particular estimator. As the *pdf*s turn out to have similar patterns under different sampling schemes, we only present *pdf*s from two typical samples such as a SRS and an endogenously stratified sample. Also, GMM4 and GMM5 are excluded since *pdf*s of estimates from these two models are very concentrated around the mean. Moreover, GMM4 and GMM5 are the least interesting in the sense that in order to be able to implement them, so much macro information must be available that the parameters could literally be estimated without even using the sample data.

Clearer patterns are observed for samples with no correlation between d and x. In an SRS (or an exogenously stratified sample which is not shown here) as shown in the upper panel in Figure 2.1, OLS is unbiased, so the GMM's efficiency gain comes from variance reduction. There is almost no efficiency gain from GMM1 since the its *pdf* are very close to the *pdf* of OLS. Figure 2.1 shows that in both cases, the OLS *pdf* spreads more widely than the *pdfs* of GMM2 and GMM3. GMM2 (or GMM3), which has information on d (or x), yields more precise estimates for β_1 (or β_2). However, its ability to estimate β_2 (or β_1) is not much better than that of OLS. Hence, it appears that including moment information on a variable can achieve an efficiency gain on estimation of the parameter of this variable, but not necessarily for another variable, when these variables are uncorrelated.

The GMM techinque shows a strong capability for bias reduction in an endogenously stratified sample as shown in the lower panel of Figure 2.1. In this case, OLS is biased while GMM2 and GMM3 are essentially unbiased. In this case, the efficiency gain provided by the GMM2 (or GMM3) estimator for β_1 (or β_2) mainly comes from bias reduction, and little from variance reduction. Although the GMM3 (or GMM2) estimates of β_1 (or β_2) has a large variance, the bias is significantly reduced. Again, GMM2 (or GMM3), which has information on d (or x), is more precise for β_1 (or β_2). The pattern is remarkably similar in a sample with correlated independent variables.

The results from the baseline model show some nice properties of the GMM technique. However, we are unclear about its empirical performance in different health-relevant scenarios, so we extend to baseline model in the section 4.2.

2.4.2 Extensions to Baseline Model

The baseline model from equation (2) has a sample with 500 observations, the standard deviation of the error term set to unity, and no heteroskedasticity. Here we provide extensions with respect to: the sample size (i.e., 150, 1000 and 5000); the explanatory power of the regressors (by changing the standard deviation of error term in the DGP[16]); and multiplicative heteroskedasticity, specified as $\sigma_i^2 = \sigma^2 e^{\tau x_{2i}}$. There is no widely accepted index or measure of the extent of heteroskedasticity and we use Kennedy's (1985) formulation

$$Heteroskedasticity = \frac{st.dev(\sigma_i)}{mean(\sigma_i)}. \tag{7}$$

The amount of heteroskedasticity depends on the ratio of the "standard deviation of the standard deviations of the errors" to the "mean of standard deviations of the errors". The parameter σ^2 is set to 2 and τ is set at 0.3, 0.8, 1.2 and 3.0 resulting in Kennedy's measure equaling 0.15, 0.42, 0.66 and 2.76 respectively.

[16] Note: standard errors are set to be 0.4, 0.8 and 1.2 respectively. OLS regression for the baseline model with the standard error of the error term equal to one in a SRS yields R^2 roughly around 0.3.

In administrative data the scope of variables is often limited, whereas the survey data often includes a wider range of variables. In order to capture this circumstance, we generalize the model by including more variables. Here we assume that macro information is available only on x_1 and x_2, which are called “common” regressors, and the rest are called “survey-only” regressors. All of the regressors are correlated, with correlations equal to 0.3. Since some administrative data may have a small number of observations, we generate a population with a relatively small size at 1500. Based on this single simulated population data, survey samples with size at 500 are generated.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \cdots + e_i \text{ }, for\ i = 1, \ldots, N_L. \tag{8}$$

$$\text{where } \beta_0 = 1 \text{ }, \beta_1 = 0.5 \text{ }, \beta_2 = -0.5, \beta_3 = 0.5 \text{ }, \beta_4 = -0.5, \ldots$$

$$x_1, x_2, x_3, x_4, \ldots \sim N\ (0, 1) \text{ and } e \sim iid\ N\ (0, 1).$$

Only the endogenous sample scheme, sampling on *y*, is considered in this extension.[17]

2.4.2.1 Monte Carlo Results for Extensions

Simulation results from the three extensions are presented in Tables 2.4, 2.5 and 2.6. The MSE ratio is not very sensitive to the sample size or to the explanatory power of the regressors. This indicates that the technique can improve on OLS not only in large samples, but also in a very small sample. The GMM techinque shows a capability for bias reduction in a very small endogenously stratified sample of 150 observations as shown in Figure 2.2, although its efficiency gain is larger when sample size is bigger. Perhaps the most interesting finding is that the technique works well, in fact very well, when heteroskedasticity is introduced in the error terms. Facing a large amount of

[17] We tried different sampling schemes such as sampling on y and some independent variable including both “common” and “survey-only” ones. The results are similar.

heteroskedasticity, the efficiency gain from using GMM is massive in this example. This bodes well for using this approach in applied work as a complement to OLS.

Next, we show cases with one, four, and ten “survey-only” variables. Interestingly, while the coefficient estimates of the two “common” variables from GMM2-5 contain little bias, the biasedness of the “survey only” coefficient estimates are also corrected simultaneously as shown in Figure 2.3. In defining the bias, the population target quantities here are defined as the OLS estimates using the 1500 population members and these are depicted as vertical dotted lines on each plot. (The GMM estimator cannot be expected to do better than the available data.) Efficiency gains in these cases almost exclusively comes from bias reduction. Another interesting finding is that the differences of efficiency gain between GMM2 and GMM5 is small, especially when the model includes more “survey only” variables. It is a useful practical property since it indicates that small amount of the key macro information can do a very good job while the marginal efficiency gain from additional information can be small when a sample is endogenously stratified. Although this is the case when independent variables are correlated, these results still hold in the uncorrelated case.

2.4.3 Limitations of the Technique

The eight sampling schemes are based on some arbitrary thresholds as described in Table 2.1. In reality, however, the sampling probability (or nonresponse) could be a continuous (and probably nonlinear) function of a variable. As a result, building on the baseline model in equation (1), we create sampling schemes with the probability of an observation to be drawn as a smooth function of a variable. Take the continuous variable x in the baseline model as an example. This sampling process is equivalent to the following: For each observation in the population, we generate a value of h, where $h \sim uniform(0,1)$, compute the CDF of the x value, and include those observations for which with $cdf(x) < h$. As a result, a population member in the right tail of the distribution are very unlikely to be drawn. Then 500 observations are randomly drawn from the remaining population. This sampling process is the sampling scheme 1. We also relax such extreme sampling scheme and move it towards a simple random sample with two more sampling schemes. Specifically, the sampling scheme 2 keeps the observations for which with $0.2 + 0.6 * cdf(x) < h$ from the population while the sampling scheme 3 keeps the observations for which with $0.4 + 0.2 * cdf(x) < h$. We also investigate cases where the sampling probability is a smooth function of the on the error term and the dependent variable.

Table 2.7 shows that when a large proportion of observations on the left tail of the distribution in x_2 and e are missing in the sampling scheme 1, the MSEs of the GMM estimators become larger than the MSE of OLS. Because of the lack of observations in right region, the GMM estimators generate large weights for those observations. In this case, a few observations are given extremely large weights resulting in an unstable estimator. This is not as much of a problem when the sampling schemes are endogenously stratified. As shown in row of "sample on y", the GMM estimators are able to generate smaller MSEs.

Although the sampling scheme 2 generates data with a distribution that is quite different from the original one, the probability of an observation from the right tail of the distribution is still larger than roughly 20% (compared with close to 0% in the sampling 1). In this case, GMM2-5 results in smaller MSE than OLS. Sample 3 is closer to a simple random sample although the probability of an observation to be draw from the middle 60 percent of the distribution is a function of the variable that follows a normally cumulative density function. Sampling on the error terms can be considered as pure endogenous sampling or missing is on some unobservables. The results show that in the sampling scheme 1, GMM2-5 estimators generate much larger MSE than OLS. However, when missing observations are not so large from the tail of the distribution (i.e., sampling scheme 2 and 3), GMM2-5 perform better OLS.

2.4.4 Summary of Monte Caro Simulation

We first design a baseline model and study the GMM estimator using Monte Carlo simulation. GMM fares well in comparison with OLS as long as values of the dependent variable are available to construct moments. Information on the independent variables only does not improve efficiency in most of the circumstance that we consider. In an SRS or an exogenously stratified sample, compared with OLS regression that yields unbiased estimates, the efficiency gain of GMM comes from variance reduction. In contrast, in an endogenously stratified sample, the GMM's efficiency gain mainly comes from bias correction. In the extension to the baseline model, we vary the sample size, the explanatory power of the regressors, and create heteroskedasticity. These properties hold to different sample sizes or to the explanatory powers of the regressors. More interestingly, facing a large amount of heteroskedasticity, the efficiency gain from using GMM is massive in this example. When more independent variables are added in GDP, while the coefficient estimates of the "common" variables from the GMM estimator contain little bias, the biasedness of the "survey only" coefficient estimates are also corrected simultaneously. Finally, the GMM estimator improve efficiency when the probability of missing observations is a function of the error terms.

However, this technique has limitations. A practical concern is the amount of overlap in the values of the variables between the population and sample, which is a concern also faced by users of propensity score matching estimators.[18] Although we allow for a difference in distribution between

[18] See Crump, R. K., V. J. Hotz, G. W. Imbens., and O. A. Mitnik (2009).

two datasets, the GMM estimator may not do well in a sample that has limited overlap with population data. Moreover, when some redundant moment conditions are added, the resulting collinearity may cause instabilities in the estimators and/or the convergence algorithm.

2.5 Applications

2.5.1 Midwifery Program

In this section we apply H&I's GMM technique to identify factors that are associated with students' dropout decisions from the Ontario Midwifery Education Program (MEP). The introduction of the MEP in 1993 represented not only the birth of accredited midwifery within the province of Ontario, but also midwifery care coverage under the Ontario Health Insurance Plan. The MEP, which began as a three-year intensive program running in a cooperative module between Laurentian, McMaster and Ryerson Universities with class sizes of under ten students, is now a four-year cornerstone to the development of autonomous women-centered birthing care for families all across the country. The program demands both proficiency in the classroom during the first half of the program, as well as a high level of maturity and professionalism in a clinical setting during its latter half.

Student retention in health professional programs is of increasing importance due to a shortage of health professionals, potential service loss and resource waste. Although a considerable amount of effort was undertaken to educate prospective students in both the difficulties associated with the MEP and the demands of practicing as a midwife, the MEP has experienced high levels of attrition since it was established (Wilson et al., 2013). From 1993 to 2006, at Ryerson and McMaster Universities, 72 midwifery students dropped out - an attrition rate of 25%. This represents an obstacle in the training of highly demanded midwives.

A survey conducted in 2007 was administered to McMaster and Ryerson midwifery students, graduates and dropouts to ascertain the important factors influencing the decisions to leave the MEP.[19] Although the survey questions cover information such as personal background, classroom experience, study habits, and difficulties experienced while enrolled in the program, the small sample size and endogenous non-response create difficulties for econometric estimation.

The survey is endogenously stratified, as individuals were contacted through different channels. Graduates were contacted through provincial Midwifery Association websites such as the Association of Ontario Midwives, so it is unlikely that graduates of the MEP who were not practicing midwifery would have been contacted to participate in the study through this channel. In contrast, individuals who left the program were sent letters to their last known addresses. Web searches using Google and Facebook, as well as gathering information from former classmates and instructors were also utilized to increase the level of dropout participation. Senior students were contacted through their respective universities, and since this is the MEP's primary method of communication most of the students were successfully contacted. Since the ability to locate drop-outs and graduates is different, the sample is likely to be endogenously biased because of the survey design. Table 2.8 describes the participants of the survey with respect to those individuals contacted. Although only one of the 28 contacted dropouts refused to participate in the study, there was still a sizeable missing rate, since only 28 of the total of 72 dropouts were contacted. The underrepresentation of dropouts also suggests the likelihood of endogenous selection bias.

Along with the survey, administrative data were collected by the two universities through the Ontario Universities' Application Centre. Due to confidentiality, our administrative data set is

[19] The survey and protocol were approved by the Research Ethics Boards at both McMaster and Ryerson Universities. Special thanks to Derek K. Lobb, Associate Professor at Medical Sciences Graduate Program, for providing the survey data and the administrative data for midwifery students.

aggregated to the university entry year level. It derives from individual-level data with information on year of admission, age at entry to the program, and current enrolment status. Others variables are irrelevant. For instance, sex is irrelevant since all individuals in our analysis are female. Macro moment conditions are therefore based on year of admission, age group and dropout status for each year.

Individuals who entered the programs from 2001 onwards are excluded from the analysis. By the time that the survey was conducted, students who were admitted in 2000 had either graduated or dropped out, while most students who were admitted in 2001 or later were still in one of the programs. They still might drop out, but the data generating process for those in-course is potentially quite different from those who are beyond the normal completion duration. As a result, we drop them, reducing the number of observations to 85. We drop 3 additional observations in which some information on variables X is missing in equation (9), reducing the final number of observation in our analysis to 82. A comparison of some key variables in Table 2.9 shows inconsistencies between the survey and the administrative data, which indicates the sample is likely to have potential endogenous selection bias. To investigate factors that contribute to the dropout decision, we estimate a simple linear probability model.

$$Dropout_i = \beta_0 + \beta_1 \cdot D_{yr96-00i} + \beta_2 \cdot McMaster_i + \beta_3 \cdot Age_i + \beta_X \cdot X_i + \epsilon_i. \qquad (9)$$

The dependent variable is binary, set equal to one for dropouts. The independent variables include a dummy variable for cohort group by the “years of admission 96-00” (i.e., 93-95 is omitted), a McMaster dummy variable, age, and other variables X. These vector X includes dummy variables: they indicate whether one had children under age of 5 upon admission; whether one’s application required an autobiographical sketch (i.e., personal background); whether one met the time requirements for class preparation (i.e., class experience); whether one questioned career choice, whether one had tutor issues, whether one had difficulties at her placement location, whether one had ongoing financial difficulties, and whether one had psychological health issues (i.e., difficulties experienced while enrolled in the program).

We construct three GMM specifications based on different sets of moment conditions. GMM(A) uses the means of dependent variable "dropout" and independent variables "year of admission 93-96", "McMaster dummy" and "age". Building from GMM(A), GMM(B) adds the means of interaction terms between "dropout" and the three independent variables. Finally, GMM(C) adds the means of interaction terms between the three independent variables.

We form two specifications for both OLS and GMM to address different policy questions. In particular, Model 1 includes only applicants' information before admission. Model 2 adds in-program experience variables. We do not argue for the structural interpretation of specifications since they include potentially endogenous explanatory variables. The first four columns in Table 2.10 compare results from the OLS and the three GMM estimates in specification 1. None of "year of admission 93-96", "McMaster", or "age" OLS estimates are statistically significant. In contrast, all the GMM models find that students from McMaster program had lower dropout rates, which is consistent with the unadjusted administrative data though not the survey data, as seen in Table 2.9. GMM(B) and GMM(C) show students who were admitted during 1993-1996 had higher probability of dropping out. Moreover, the GMM models estimate a negative association between visible minority and dropping out, as well as a positive association between the autobiography requirement before admission and dropout. These associations are statistically significant, unlike OLS. With more in-program variables added, specification 2 suggests that poor academic performance (i.e. class experience) was associated with an increased drop-out rate. Age becomes negatively associated with dropout and its coefficient estimates in both OLS and GMM models are statistically significant. Neither OLS nor the GMM models show that placement issues, having children under age of 5 upon admission, and psychological health problems are related to the dropout decision. Interestingly, GMM estimators generate sizeable changes in both the coefficient estimates and standard errors of some variables such as “year of admission 93-96” in both model 1 and 2, and “McMaster” in model 2. In contrast, they only generate changes in coefficients estimates but little changes in standard errors of some variables such as “Preparation for Class”.

To understand the source of the differences between the OLS and GMM models, we investigate the values of the Lagrange Multipliers on each constraint from the GMM(C) model in Table 2.11. The last row of the Table 2.11 gives the chi-squared test of the hypotheses that all Lagrange multipliers are equal to zero. Clearly, the test rejects the null hypotheses that the two data distributions are equal. Figure 2.4 is a histogram of weights that are constructed by this technique. Some observations which are underrepresented, are given relatively high weights. Figure 2.5 demonstrates that the age distributions of the administrative-weighted survey data and administrative data are much closer than are the unweighted survey data and administrative data.[20] To further explore the influence of the weights, we compare some summary statistics of key variables between survey data, weighted survey data and administrative data in Table 2.12. It is clear that the means and standard deviations are much closer between the administrative data and weighted survey data than the unweighted survey data. All of this suggests that the weighting GMM estimator successfully recovers aspects of the data not observable by OLS but more representative of the underlying distribution by a combination of bias and variance reduction.

[20] We received the administrative age as an anonymized file.

2.5.2 Optometrists in Ontario

The GMM technique fits not only the OLS model but also more complex models such as two-step linear GMM models, exponential regression models, and even Euler equations such as those used in macro structural models. In principle, any model, which can be transformed into a "moment conditions" framework, can apply this technique. Here we demonstrate that the GMM technique can be used in a Probit model in the context of optometrists in Ontario, Canada.

A survey was conducted by researchers affiliated with the Ontario Health Human Resources Research Network, a pan-Ontario network funded by the Ontario Ministry of Health and Long-Term Care (MOHLTC), in collaboration with several universities, to evaluate the process of implementing legislative changes to Bills 171 and 179 of the Ontario Regulated Health Profession Act in 2007. Although 100% of the optometrists were targeted, since participation was voluntary and no funds were provided to increase incentive for participation, we have only 330 observations, which represents about 15% of the population. At the same time, we have access to Health Professions Database (HPDB), which is an administrative registry of all optometrists regulated to work in the province. This dataset was created by the Ontario Ministry of Health and Long-Term Care and the health regulatory College of Ontario, and is a census.

The survey includes a wide range of questions on individuals' demographic characteristics, practice patterns, and opinions about the expanded scope of practices recently adopted in the province. The HPDB includes demographic characteristics such as sex and age, and geographic information on the place of current employment. We are interested in the policy question on the association between "working in a rural area" and age and gender, so we form a Probit Model.

$$Rural_i = Probit(\beta_0 + \beta_1 \cdot Male_i + \beta_2 \cdot Age_i + \beta_3 \cdot Age * Male_i + \beta_X \cdot X_i + \epsilon_i) \qquad (10)$$

The dependent variable *Rural* is binary, equaling 1 if an optometrist works in a rural area and 0 otherwise. This variable is generated from the answer to the survey question "What are the first 3 digits of the postal code of your primary place of employment?" and uses The Rurality Index for Ontario (RIO). The RIO measure was developed by Kralj (2000) and has been widely adopted by MOHLTC and Ontario Medical Association. A nice feature of this index is that it is associated with the 6-digit postal code so that we can measure the rurality of an optometrist's primary place of employment. Since the survey only asks about the first 3 digits of the postal code, we measure the rurality by the mean of the RIO among all with the same first 3 digits. Numerical values of the RIO range between 0 and 100, and an area with an RIO above 40 is commonly defined as rural. Hence, the dependent variable rural is defined to be 1 if RIO is larger than 40. Independent variables include age, male, an interaction term between age and male, and three dummy variables (including "whether one obtained his Optometry Degree in a Canadian University", "whether one's primary place of employment is described as independent optometry practice" and "whether one works full time in his primary place of employment").

The HPDB includes information on rurality, sex and age. We construct two GMM specifications based on different sets of moment conditions. GMM(1) uses the means of the independent variables *Male, Age* and *Male*Age*. Since we show in Monte Carlo simulation that using only independent variables to construct the weight function cannot change the distribution of the conditional density, a comparison of GMM(1) and standard Probit is interesting. In contrast, GMM(2) uses the means of both the dependent and independent variables *Rural, Male, Age, Male*Age, Rural*Age* and *Rural**Male. In GMM specifications, the first type of moment condition involves the usual Probit maximum likelihood first-order conditions multiplied by the weights $\widehat{w}_i$,

Probit:

$$g(\hat{\beta}_{probit}) = \frac{1}{N_S}\sum_{i=1}^{N_S}\left\{x_i\left[y_i\frac{\phi(x_i'\hat{\beta}_{probit})}{\Phi(x_i'\hat{\beta}_{probit})} - (1-y_i)\frac{\phi(x_i'\hat{\beta}_{probit})}{1-\Phi(x_i'\hat{\beta}_{probit})}\right]\right\} = 0 \tag{11}$$

GMM:

$$g(\hat{\beta}_{gmm}, \hat{\lambda}) = \frac{1}{N_S}\sum_{i=1}^{N_S}\left\{\begin{matrix}\hat{w}_i x_i\left[y_i\frac{\phi(x_i'\hat{\beta}_{gmm})}{\Phi(x_i'\hat{\beta}_{gmm})} - (1-y_i)\frac{\phi(x_i'\hat{\beta}_{gmm})}{1-\Phi(x_i'\hat{\beta}_{gmm})}\right] \\ \hat{w}_i(h^s(y_i, x_i) - h^p)\end{matrix}\right\} = 0, \tag{12}$$

$\phi()$ is the standard normal probability density function and $\Phi()$ is the normal cumulative distribution function. The results from Table 2.13 show that the differences between GMM(1) and Probit are small. GMM(2) generates sizeable changes not only in the coefficient estimates and standard errors of these three “common” variables, but also in other “survey only” variables. Specifically, GMM(2) estimates the association between *Rural* and *male*, *age*, and *male**age*. Although standard Probit regression and GMM(1) also find these associations, they are not statistically significant at 5%. Although young male optometrists tend to work in urban areas compared to young female optometrists, this association reverses with age. The rest of the independent variables are not statistically significant in either regression.

2.6 Conclusion

In health economics it is common for researchers to conduct surveys based on a target population, for instance physicians in some regions or patients in a clinical trial study. Such surveys frequently suffer from small samples, nonresponse and attrition. Using longitudinal unweighted survey data can lead to biased estimators. This paper applies a technique developed by Hellerstein and Imbens (1999) to improve the efficiency and reduce the bias of estimates with additional macro-level (marginal) information. In particular, it constructs weights for observations in the survey to force some moments in the sample to approximate the corresponding moments from the population.

Solon et al. (2013) distinguish two purposes of weighted estimation, to estimate population descriptive statistics and to estimate causal effects. This technique is aimed at the first purpose, by making the sample more representative of the target population.

A Monte Carlo simulation study shows that the efficiency gain from the GMM estimators can be substantial in terms of reducing the bias and variance of estimates, especially in an endogenously stratified sample. The study also shows that the advantages of this technique are maintained in the face of variation in the sample sizes, explanatory power of independent variables, and the amount of heteroskedasticity in the error. In particular, the technique performs very well when a high amount of heteroskedasticity is specified. A practical concern of the technique, however, is the amount of overlap in the values of the variables between the population and sample. When the amount of overlap is little, some observations could be given extremely large weights resulting in an unstable estimator. We provide two applications using the technique: one is to examine student retention in Ontario's midwifery program and the other is to investigate the association between the rurality of the Ontario optometrists' work place and their demographic characteristics including age and sex. The weights constructed by this technique are found to change the estimates, and the weights shift the sample distribution towards the population distribution. We believe this approach could be widely adopted in health economics and health services.

References

Bethlehem, J. G. (1988). “Reduction of Nonresponse Bias Through Regression Estimation,” *Journal of Official Statistics*, 4(3), 251-260.

Card, D., A. K. G. Hildreth, and L. D. Shore-Sheppard (2004), “The measurement of Medicaid coverage in the SIPP: Evidence from California, 1990-1996,” *Journal of Business and Economic Statistics* 22.

Cosslett, S. R. (1981), “Maximum Likelihood Estimator for Choice-Based Samples,” *Econometrica*, 49, 1289-1316.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009), “Dealing with Limited Overlap in Estimation of Average Treatment Effects,” *Biometrika* 96(1): 187-199.

Deaton, Angus (1995), “Data and Econometric Tools for Development Analysis,” *Handbook of Development Economics,* in: J. Behrman & T.N. Srinivasan (ed.), Handbook of Development Economics, edition 1, volume 3A, chapter 33, pages 1785-1882.

Devereux, P. J. and G. Tripathi (2009), “Optimally Combining Censored and Uncensored Datasets,” *Journal of Econometrics*, Vol. 151(1), pp17-32.

Griliches, Z., B. H. Hall, and J. A. Hausman (1978), “Missing Data and Self-Selection in Large Panels,” *Annales de l'inséé*, No. 30/31, pp137-176.

Gourieroux, C. and A. Monfort (1981), “On the Problem of Missing Data in Linear Models,” *Review of Economic Studies*, 48, 597-596.

Heckman, J. J. (1976), “The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models,” *Annals of Economic Social Measurement*, 5/4, 475-492.

Heckman, J. J. (1979), “Sample Selection Bias as a Specification Error. *Econometrica*, 47, 153-161.

Heckman, J. J. (2008), “Econometric Causality,” *International Statistical Review*, 76(1): 1-27.

Hellerstein, J. K. and G. W. Imbens (1999), “Imposing Moment Restrictions from Auxiliary Data by Weighting,” *The Review of Economic and Statistics,* Vol. 81, pp. 1-14.

Holt, D. and T. M. F. Smith (1979), “Post Stratification,” Journal of the Royal Statistical Society, Series A, 142, 33-66.

Horvitz, D.G. and Thompson, D.J. (1952), “A Generalization of Sampling without Replacement from a Finite Universe,” Journal of the American Statistical Association, 47, 663-685.

Hsieh, D. A., C. F. Manski, and D. McFadden. (1985), "Estimation of Response Probabilities from Augmented Retrospective Observations," *Journal of the American Statistical Association*, Vol. 80, pp. 651-652.

Imbens, G. W. (1992), "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling," *Econometrica*, 60, 1187-1214.

Imbens, G. W., and T. Lancaster (1994), "Combining Micro and Macro Data in Microeconometric Models," *The Review of Economic Studies*, Vol. 61, pp. 655-680.

Imbens, G. W., and T. Lancaster (1996), "Efficient Estimation and Stratified Sampling," *Journal of Econometrics*, 74, 289-318.

Imbens, Guido W., and Jeffrey M. Wooldridge (2009), "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47(1): 5-86.

Kennedy, P. E. (1985), "A Suggestion for Measuring Heteroskedasticity," *Communications in Statistics, Part B - Simulation and Computation*, 14, 845-851.

Kerm, V. P. (2012), "Kernel-smoothed Cumulative Distribution Function Estimation with akdensity," *The Stata Journal*, 12, 3, pp. 543-548.

Kralj, Boris, 2000. "Measuring 'Rurality' for Purposes of Health-care Planning: an Empirical Measure for Ontario," Ontario Medical Review 67, 33-49.

Lancaster, T., and G. Imbens (1996), "Case-control Studies with Contaminated Controls," *Journal of Econometrics*, 71, 145-160.

Little, R.J.A. (1986), "Survey Nonresponse Adjustments," *International Statistical Review*, 54, 139-157.

Little, R.J.A. (1993). "Post-Stratification: a Modeler's Perspective," *Journal of the American Statistical Association*, 88, 1001-1012.

Manski, C. F., and S. R. Lerman (1977), "The Estimation of Choice Probabilities from Choice-Based Samples," *Econometrica*, 45, 1977-1988.

Manski, C. F., and D. McFadden (1981), "Alternative Estimators and Sample Designs for Discrete Choice Data," *Structural Analysis of Discrete Data*, Cambridge, Mass.: MIT press.

Ramalho, E. A., and J. J. S. Ramalho (2006), "Bias-corrected Moment-based Estimators for Parametric Models under Endogenous Stratified Sampling," *Econometric Reviews*, 25(4), 475-496.

Ridder, G., and R. Moffitt (2007), "The Econometrics of Data Combination," *Handbook of Econometrics*, Vol. 6B, 5470-5543.

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581-592.

Wilson, R., K. Eva, and D. K. Lobb (2013), "Student Attrition in the Ontario Midwifery Education Programme," *Midwifery*, Vol. 29(6), 579-584.

Table 2.1: Monte Carlo Designs -- Eight Sample Schemes based on Different Variables

Sampling Schemes	Threshold	Sampling Probability
Simple Random Sample	N/A	Same
D	$d < threshold$	Low
	$d > threshold$	High
X	$x < threshold$	Low
	$x > threshold$	High
Y	$y < median(y)$	Low
	$y > median(y)$	High
d and x	$d < threshold$	Low
	$x > threshold$	Low
d and y	$d < threshold$	Low
	$y > threshold$	High
x and y	$x < threshold$	Low
	$y > threshold$	High
d, x and y	$d < threshold$	Low
	$xy > \text{threshold}$	High

Table 2.2: Monte Carlo Designs -- OLS and Five GMM Moment Conditions

Models	Macro Moment Conditions	Components
OLS	N/A	N/A
GMM1	D, X, DX, XX	Only independent variables
GMM2	D, Y, DY, YY	Dummy variable and y
GMM3	X, Y, XX, XY, YY	Continuous variable and y
GMM4	D, X, Y, DY, XY, YY	Mixture of both indep and dep variables
GMM5	D, X, Y, DX, DY, XY, YY	Mixture of both indep and dep variables
GMM6	D, X, Y, DX, XX, DY, XY, YY	Mixture of both indep and dep variables

Note: GMM6 is not included because so much macro information is added in this model that the coefficients can be perfectly estimated by these moment conditions alone, without using the sample data.

Table 2.3: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE from Baseline Model

	GMM1	GMM2	GMM3	GMM4	GMM5
$corr(d, x) = 0$					
Simple Random Sample	1.009	0.470	0.432	0.295	0.183
Sample on d	1.229	0.561	0.741	0.378	0.223
Sample on x	1.581	0.605	0.603	0.561	0.374
Sample on y	0.743	0.162	0.157	0.086	0.055
Sample on d & x	1.171	0.601	0.613	0.516	0.310
Sample on d & y	0.957	0.213	0.196	0.123	0.072
Sample on x & y	0.829	0.321	0.305	0.198	0.128
Sample on d, x & y	1.866	0.611	0.138	0.125	0.080
$corr(d, x) = 0.5$					
Simple Random Sample	1.011	0.497	0.486	0.281	0.166
Sample on d	1.284	0.621	0.752	0.380	0.237
Sample on x	1.528	0.792	0.561	0.482	0.390
Sample on y	0.808	0.202	0.192	0.082	0.052
Sample on d & x	1.318	0.720	0.500	0.532	0.449
Sample on d & y	0.981	0.263	0.240	0.128	0.081
Sample on x & y	0.874	0.325	0.296	0.174	0.121
Sample on d, x & y	1.349	0.446	0.116	0.093	0.073

Note: The table entries are MSE_{GMM}/MSE_{OLS} based on the 1000 simulated values. MSE_{GMM} and MSE_{OLS} are defined as the sum of the MSEs of $\hat{\beta}_1$ and $\hat{\beta}_2$. The results are based on the baseline model defined in equation (2) in the text.

Table 2.4: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE by Varying Sample Size

	Model	SRS		Sample on y		Sample on $dx\ y$	
		NoCorr	Corr	NoCorr	Corr	NoCorr	Corr
Ns=150							
	GMM1	1.019	1.042	0.941	0.976	1.620	1.272
	GMM2	0.501	0.511	0.388	0.441	0.569	0.492
	GMM3	0.464	0.512	0.427	0.470	0.303	0.243
	GMM4	0.326	0.310	0.236	0.245	0.290	0.246
	GMM5	0.203	0.195	0.149	0.155	0.199	0.202
Ns=500							
	GMM1	1.009	1.011	0.743	0.808	1.866	1.349
	GMM2	0.470	0.497	0.162	0.202	0.611	0.446
	GMM3	0.432	0.486	0.157	0.192	0.138	0.116
	GMM4	0.295	0.281	0.086	0.082	0.125	0.093
	GMM5	0.183	0.166	0.055	0.052	0.080	0.073
Ns=1000							
	GMM1	1.003	1.008	0.695	0.753	1.959	1.356
	GMM2	0.477	0.493	0.086	0.112	0.614	0.437
	GMM3	0.438	0.471	0.090	0.101	0.076	0.072
	GMM4	0.284	0.274	0.046	0.048	0.070	0.055
	GMM5	0.171	0.163	0.028	0.029	0.045	0.048
Ns=5000							
	GMM1	1.002	1.002	0.627	0.678	1.874	1.361
	GMM2	0.508	0.512	0.025	0.025	0.612	0.436
	GMM3	0.503	0.503	0.025	0.025	0.014	0.021
	GMM4	0.372	0.359	0.014	0.014	0.022	0.021
	GMM5	0.266	0.247	0.010	0.010	0.019	0.026

Note: The table entries are MSE_{GMM}/MSE_{OLS} based on the 1000 simulated values. MSE_{GMM} and MSE_{OLS} are defined as the sum of the MSEs of $\hat{\beta}_1$ and $\hat{\beta}_2$. The results are based on extended baseline model with respective to sample size.

Table 2.5: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE by Varying Standard Error of the Error Term

	Model	SRS		Sample on y		Sample on $dx\ y$	
		NoCorr	Corr	NoCorr	Corr	NoCorr	Corr
Std Err=0.4							
	GMM1	1.009	1.011	0.429	0.362	1.887	1.664
	GMM2	0.598	0.530	0.174	0.109	0.556	0.553
	GMM3	0.421	0.443	0.139	0.089	0.403	0.298
	GMM4	0.467	0.469	0.128	0.088	0.422	0.419
	GMM5	0.262	0.284	0.075	0.058	0.262	0.278
Std Err=0.8							
	GMM1	1.009	1.011	0.648	0.707	2.025	1.446
	GMM2	0.468	0.476	0.137	0.138	0.542	0.375
	GMM3	0.419	0.458	0.131	0.129	0.214	0.135
	GMM4	0.358	0.358	0.085	0.092	0.204	0.158
	GMM5	0.229	0.215	0.053	0.060	0.133	0.126
Std Err=1.0							
	GMM1	1.009	1.011	0.743	0.808	1.866	1.349
	GMM2	0.470	0.497	0.162	0.202	0.611	0.446
	GMM3	0.432	0.486	0.157	0.192	0.138	0.116
	GMM4	0.295	0.281	0.086	0.082	0.125	0.093
	GMM5	0.183	0.166	0.055	0.052	0.080	0.073
Std Err=1.2							
	GMM1	1.009	1.011	0.831	0.886	1.705	1.255
	GMM2	0.478	0.517	0.205	0.263	0.651	0.543
	GMM3	0.444	0.510	0.211	0.270	0.105	0.091
	GMM4	0.241	0.221	0.085	0.087	0.073	0.053
	GMM5	0.145	0.129	0.049	0.054	0.047	0.046

Note: The table entries are MSE_{GMM}/MSE_{OLS} based on the 1000 simulated values. MSE_{GMM} and MSE_{OLS} are defined as the sum of the MSEs of $\hat{\beta}_1$ and $\hat{\beta}_2$. Standard errors of 0.4, 0.8, 1.2 result in big, median, small R^2 in OLS respectively. The results are based on extended baseline model with respective to explanatory power of the regressors.

Table 2.6: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE by Varying the Size of Heteroskedasticity

	Model	SRS		Sample on y		Sample on $dx\ y$	
		NoCorr	Corr	NoCorr	Corr	NoCorr	Corr
$\tau = 0.3$							
	GMM1	1.016	1.014	0.961	0.963	1.650	1.266
	GMM2	0.507	0.535	0.141	0.206	0.788	0.658
	GMM3	0.445	0.524	0.140	0.180	0.076	0.070
	GMM4	0.187	0.168	0.036	0.039	0.037	0.027
	GMM5	0.111	0.101	0.021	0.023	0.020	0.019
$\tau = 0.8$							
	GMM1	1.035	1.031	1.076	1.030	1.558	1.229
	GMM2	0.483	0.507	0.184	0.284	0.781	0.585
	GMM3	0.396	0.461	0.071	0.099	0.068	0.046
	GMM4	0.129	0.126	0.013	0.014	0.024	0.015
	GMM5	0.075	0.079	0.008	0.008	0.012	0.008
$\tau = 1.2$							
	GMM1	1.058	1.050	1.112	1.057	1.470	1.168
	GMM2	0.386	0.436	0.261	0.354	0.550	0.375
	GMM3	0.336	0.367	0.056	0.068	0.061	0.030
	GMM4	0.072	0.082	0.006	0.006	0.015	0.009
	GMM5	0.039	0.052	0.004	0.004	0.006	0.004
$\tau = 3.0$							
	GMM1	1.002	1.083	0.903	0.852	0.950	0.699
	GMM2	0.130	0.229	0.095	0.167	0.056	0.068
	GMM3	0.409	0.263	0.188	0.133	0.224	0.038
	GMM4	0.002	0.009	0.002	0.001	0.002	0.002
	GMM5	0.001	0.007	0.001	0.001	0.001	0.001

Note: The table entries are MSE_{GMM}/MSE_{OLS} based on the 1000 simulated values. MSE_{GMM} and MSE_{OLS} are defined as the sum of the MSEs of $\hat{\beta}_1$ and $\hat{\beta}_2$. $\tau = 0.3$ or $0.5, 0.8$ or $1.2,$ and 3.0 represent small, median and large amounts of heteroskedasticity respectively. The results are based on extended baseline model with respective to multiplicative heteroskedasticity.

Table 2.7: Monte Carlo Results -- Ratio of GMM MSE to OLS MSE from Baseline Model

	GMM1	GMM2	GMM3	GMM4	GMM5
No Correlation					
Sample on x					
Sampling scheme 1	2.897	0.821	1.144	6.533	7.528
Sampling scheme 2	1.372	0.567	0.526	0.531	0.450
Sampling scheme 3	1.024	0.410	0.435	0.253	0.139
Sample on e					
Sampling scheme 1	1.015	8.257	13.64	5.181	2.221
Sampling scheme 2	1.005	0.965	0.888	0.499	0.266
Sampling scheme 3	1.013	0.498	0.474	0.344	0.177
Sample on y					
Sampling scheme 1	1.047	0.609	0.509	0.379	0.224
Sampling scheme 2	1.038	0.230	0.183	0.158	0.078
Sampling scheme 3	1.033	0.375	0.290	0.210	0.138
Correlation					
Sample on x					
Sampling scheme 1	3.456	1.326	1.074	5.226	8.613
Sampling scheme 2	1.356	0.477	0.496	0.372	0.358
Sampling scheme 3	1.014	0.454	0.457	0.339	0.218
Sample on e					
Sampling scheme 1	1.018	8.678	10.87	3.273	1.852
Sampling scheme 2	1.018	0.899	0.954	0.448	0.258
Sampling scheme 3	0.982	0.630	0.540	0.320	0.184
Sample on y					
Sampling scheme 1	1.034	0.603	0.641	0.354	0.177
Sampling scheme 2	1.107	0.237	0.233	0.143	0.085
Sampling scheme 3	1.021	0.376	0.363	0.222	0.143

Note: Take the continuous variable x as an example.

Sampling scheme 1: h~uniform (0,1) keep h>cdf(x), then randomly draw 500 observations.
Sampling scheme 2: h~uniform (0,1) keep h>0.2+0.6*cdf(x), then randomly draw 500 observations.
Sampling scheme 3: h~uniform (0,1) keep h>0.4+0.2*cdf(x), then randomly draw 500 observations.

Table 2.8: Midwife Application -- Proportions of Survey Participants from the Population

	Total Enrolled	Contacted	Participated	Representation of Population (%)
Current Students	131	127	87	66
Graduates	222	119	101	45
Dropouts	72	28	27	38
Total	425	274	215	51

Source: (Wilson et al., 2013)

Table 2.9: Midwife Application -- Comparison of Different Statistics between Survey Data and Administrative Data

		Survey data			Administrative data		
Year of admission		Ryerson	McMaster	Total	Ryerson	McMaster	Total
1993-1995	Dropout	0.30	0.00	0.15	0.20	0.03	0.14
1996-2000	Dropout	0.12	0.18	0.15	0.28	0.20	0.25
All year	Dropout	0.16	0.13	0.15	0.26	0.16	0.21
1993-1995	Age	30.4	25.6	28.0	34.1	28.9	32.3
1996-2000	Age	30.6	28.1	29.5	31.0	28.0	29.6
All year	Age	30.5	27.4	29.1	32.0	28.2	30.4
1993-1995	Obs.	10	10	20	43	22	65
1996-2000	Obs.	34	28	62	92	79	171
All year	Obs.	44	38	82	135	101	236

Note: The figures in the first six rows represent means of variable dropout and age at two groups by year of admission. Administrative data shows both Universities started to expand recruitments in 1996, so year of admission is consequently grouped into two categories (i.e. 93-95 and 96-00).

Table 2.10: Midwife Application -- Regression Results of OLS and GMM in Different Specifications

	(1)				(2)			
	OLS	GMM(A)	GMM(B)	GMM(C)	OLS	GMM(A)	GMM(B)	GMM(C)
Yr_adm96-00	-0.021	0.027	0.084***	0.073***	0.084	0.139	0.179***	0.191***
	(0.097)	(0.112)	(0.027)	(0.024)	(0.089)	(0.084)	(0.048)	(0.047)
McMaster	-0.049	-0.168**	-0.122***	-0.108***	-0.028	-0.112	-0.092	-0.098*
	(0.087)	(0.085)	(0.041)	(0.034)	(0.082)	(0.073)	(0.060)	(0.059)
Age	-0.004	-0.006	-0.001	-0.003	-0.014**	-0.018***	-0.017***	-0.020***
	(0.007)	(0.009)	(0.003)	(0.002)	(0.007)	(0.006)	(0.005)	(0.005)
Visible Minority	-0.143	-0.166***	-0.161***	-0.156***	-0.203	-0.208**	-0.196**	-0.203**
	(0.129)	(0.052)	(0.037)	(0.035)	(0.125)	(0.100)	(0.089)	(0.096)
Required Autobiography	0.138	0.182**	0.173***	0.174***	0.111	0.101	0.091	0.067
	(0.148)	(0.071)	(0.061)	(0.058)	(0.131)	(0.116)	(0.113)	(0.109)
ChildUnder5	-0.03	-0.058	-0.046	-0.020	0.000	-0.042	-0.044	-0.037
	(0.095)	(0.121)	(0.104)	(0.099)	(0.083)	(0.077)	(0.071)	(0.073)
Preparation for Class					-0.085**	-0.116***	-0.120***	-0.124***
					(0.036)	(0.038)	(0.037)	(0.037)
Career choice Issues					0.163*	0.195**	0.185**	0.159*
					(0.089)	(0.092)	(0.088)	(0.082)
Placement Issues					-0.009	0.009	-0.009	-0.01
					(0.093)	(0.100)	(0.092)	(0.088)
Tutor issues					0.319**	0.197	0.255**	0.253**
					(0.150)	(0.155)	(0.130)	(0.126)
Financial Difficulty					0.167	0.184	0.180	0.202*
					(0.126)	(0.121)	(0.111)	(0.119)
Psychological Health					0.154	0.151	0.146	0.146
					(0.092)	(0.106)	(0.099)	(0.106)
Cons.	0.194	0.320	0.137	0.175**	0.608**	0.903***	0.865***	0.988***
	(0.248)	(0.220)	(0.096)	(0.085)	(0.290)	(0.249)	(0.252)	(0.242)
R^2	0.043				0.357			
N.	82	82	82	82	82	82	82	82

*p<0.05, ** p<0.01, *** p<0.001

Table 2.11: Midwife Application -- Test of Equality of the Survey data and the Administrative Data

	Lagrange Multipliers		
Moment	Estimate	Std. error	t-stat
Dropout	-1.10	(1.88)	-0.59
Year of admission 96-00	-0.22	(1.81)	-0.12
McMaster	-2.91**	(1.39)	-2.09
Age	-0.08	(0.06)	-1.39
Dropout×Yr adm96-00	0.21	(1.00)	0.21
Dropout×McMaster	-0.16	(0.91)	-0.18
Dropout×Age	0.01	(0.06)	0.25
Age×McMaster	0.06	(0.04)	1.36
McMaster×Yr adm96-00	1.04	(0.69)	1.50
Age×Yr adm96-00	0.02	(0.05)	0.3
Chi-square tests (dof)		32.903(10)***	
		(P-value=0.000)	

Table 2.12: Midwife Application -- Comparison of Means and Standard Deviations of Key Variables

	Survey Data	Weighted Survey Data	Administrative Data
Dropout-	0.2034	0.1555	0.1481
McMaster	(0.4060)	(0.3655)	(0.3563)
N	39	39	162
Dropout-	0.2097	0.3101	0.2838
Ryerson	(0.4104)	(0.4663)	(0.4524)
N	43	43	148
Year93_95	0.1653	0.2203	0.2323
	(0.3730)	(0.4162)	(0.4230)
Year96_00	0.5372	0.5544	0.5516
	(0.5007)	(0.4991)	(0.4981)
McMaster	0.4876	0.5216	0.5226
	(0.5019)	(0.5016)	(0.5003)
Age	28.165	29.228	29.449
	(5.9838)	(6.4228)	(6.6050)
N	82	82	310

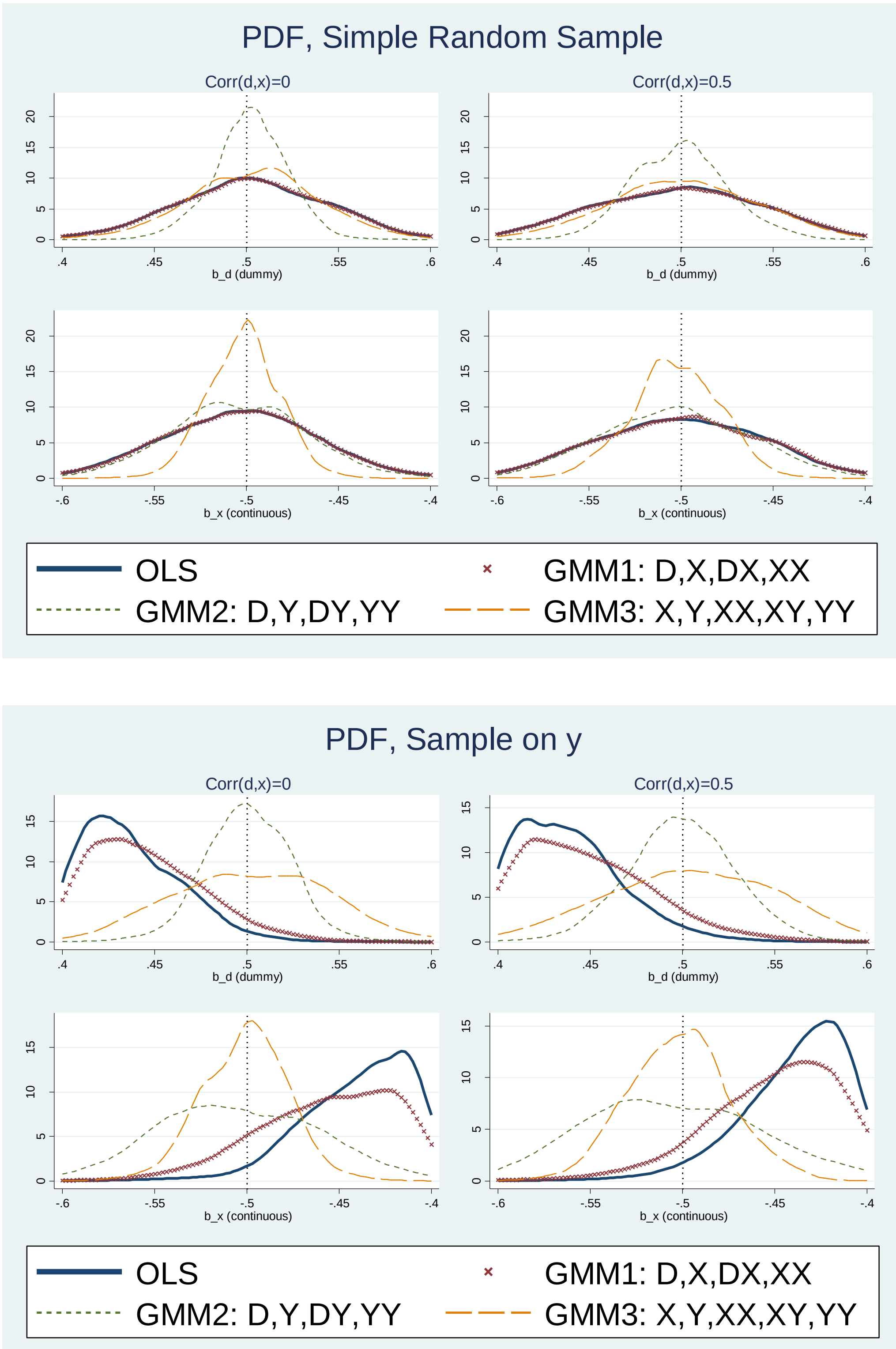
Note: Means and standard deviations (in parentheses) are shown.

Table 2.13: Optometry Application -- Regression Results of Probit and GMMs

	Probit	GMM(1)	GMM(2)
Male	-1.760	-1.801	-1.832*
	(1.013)	(0.984)	(0.803)
Age	-0.003	-0.004	-0.031*
	(0.016)	(0.018)	(0.015)
Age*Male	0.037	0.038	0.049*
	(0.021)	(0.022)	(0.020)
Canadian Education	0.185	0.223	0.096
	(0.309)	(0.280)	(0.252)
Independent practice	0.298	0.317	-0.078
	(0.455)	(0.487)	(0.390)
Fulltime Status	0.031	0.050	0.067
	(0.254)	(0.260)	(0.261)
Constant	-1.780*	-1.788	-0.536
	(0.817)	(0.914)	(0.559)
Pseudo R^2	0.0463		
N	307	307	307

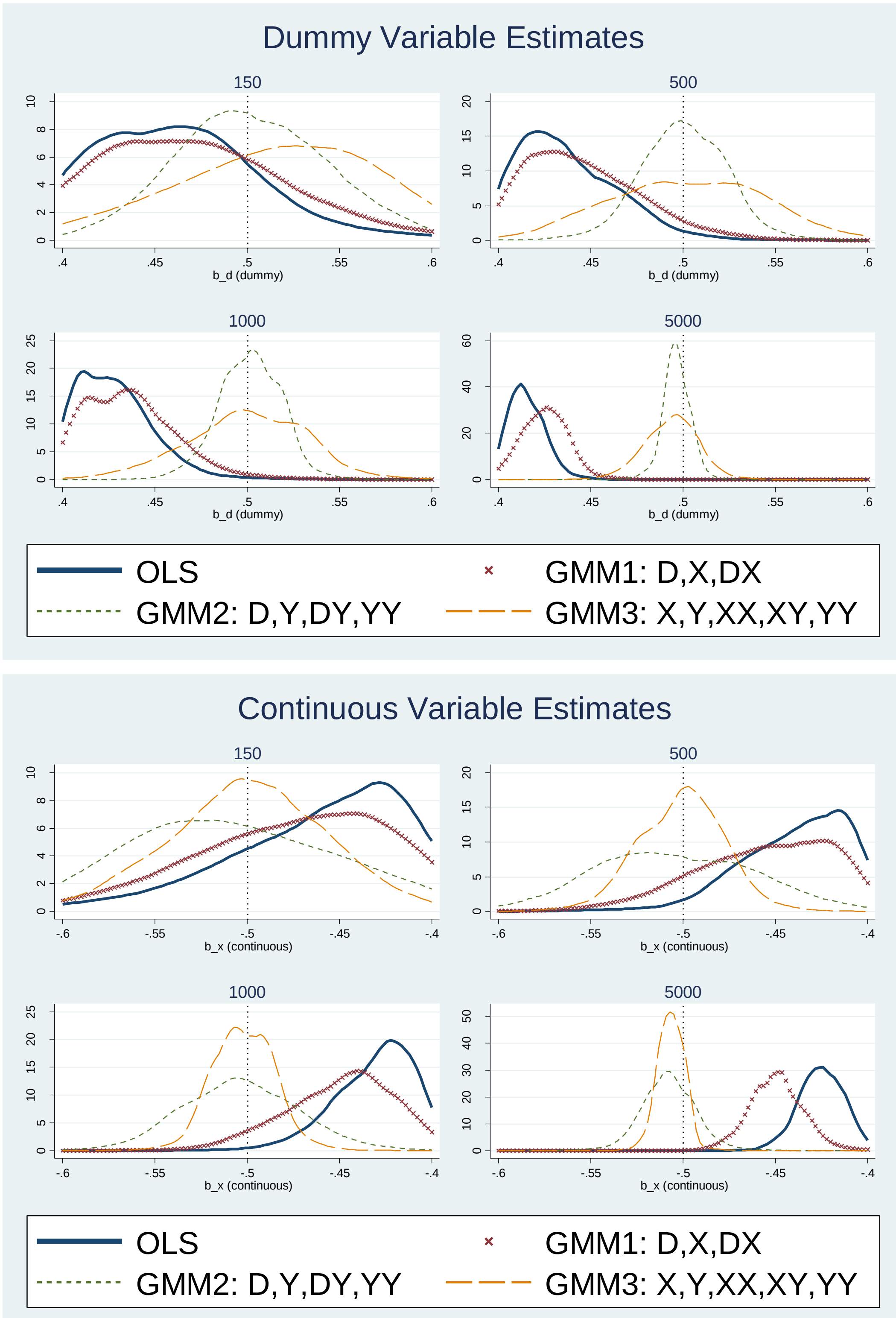
*p<0.05, ** p<0.01, *** p<0.001

Figure 2.1: Monte Carlo -- Probability Density Function of Estimates in Simple Random Samples and Endogenously Stratified Samples



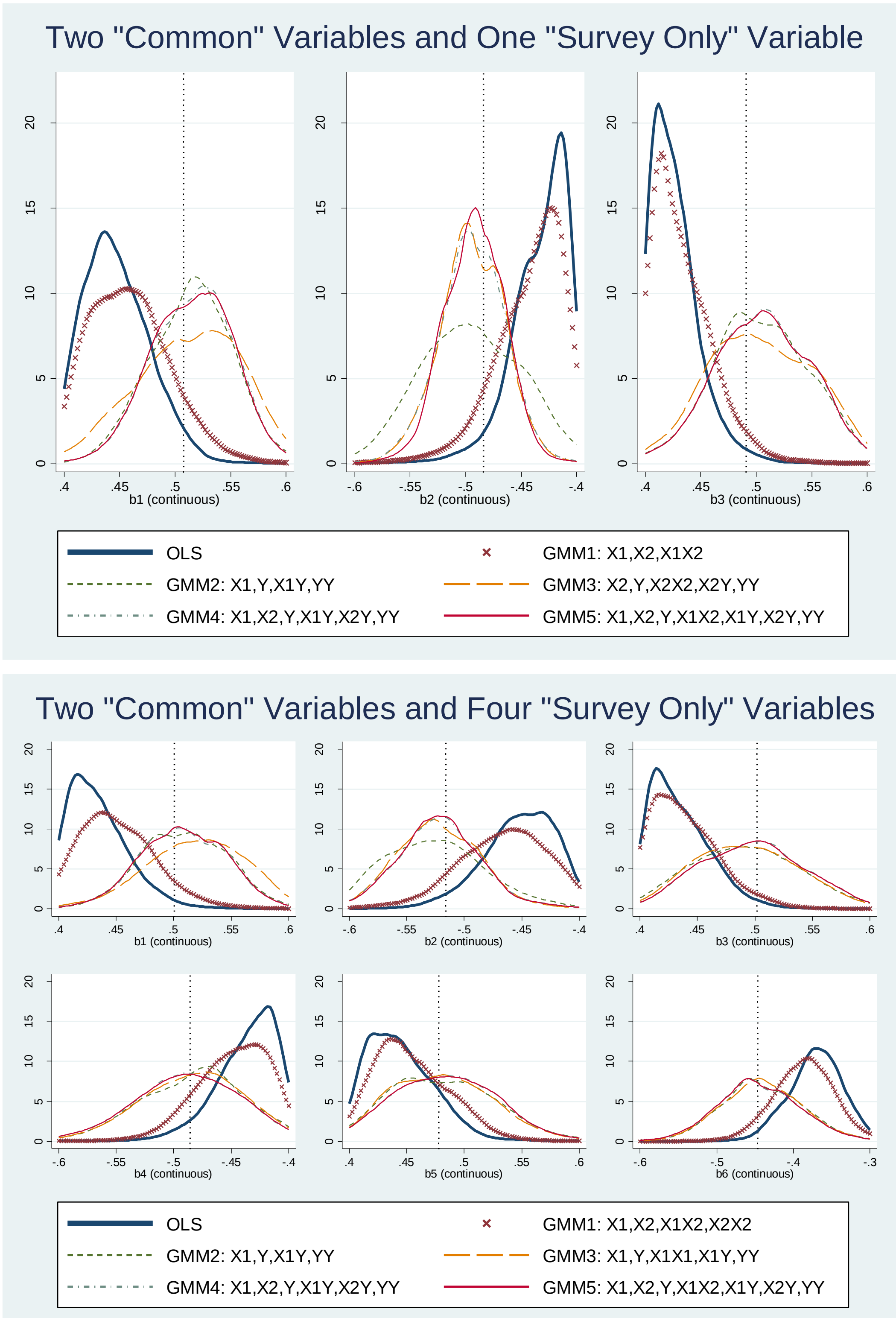
Note: *pdf* is kernel-smoothed cumulative distribution function constructed using Kerm (2012).

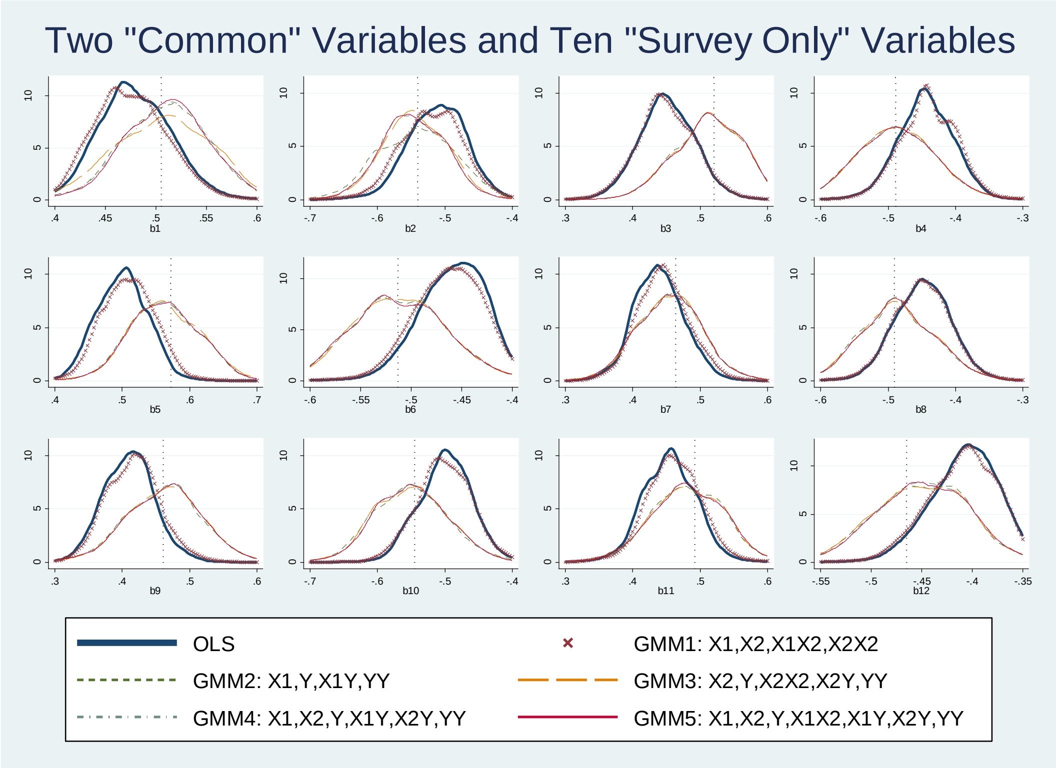
Figure 2.2: Monte Carlo -- Probability Density Function of Estimates in Extended Models with Different Sample Sizes (Sample on y)



Note: *pdf* is kernel-smoothed cumulative distribution function constructed using Kerm (2012).

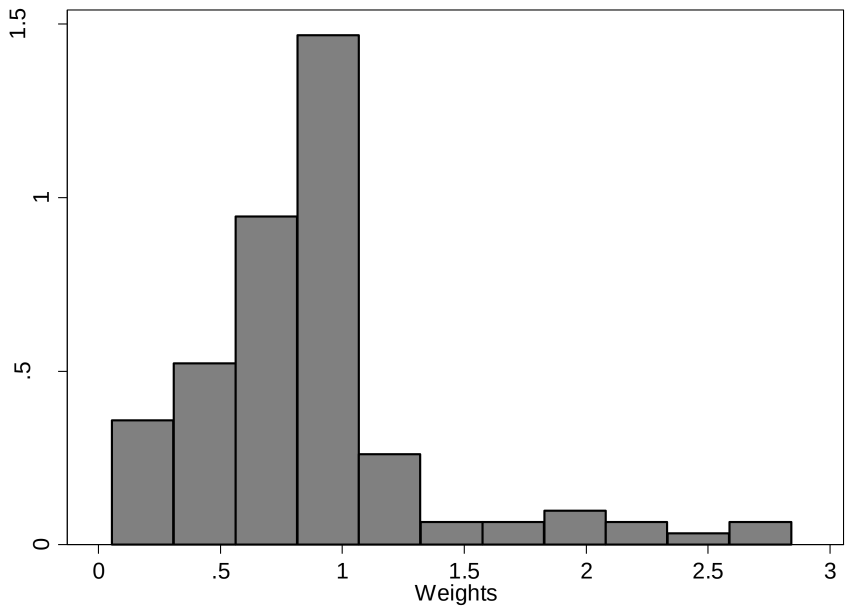
Figure 2.3: Monte Carlo -- Probability Density Function of Estimates in Extended Models with More Independent Variables (Sample on y)





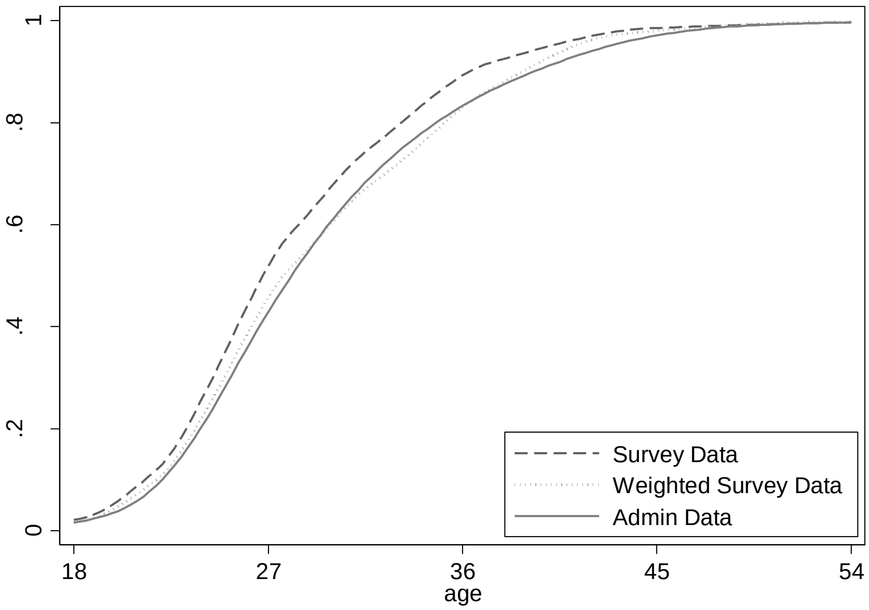
Note: *pdf* is kernel-smoothed cumulative distribution function constructed using Kerm (2012). The results are based on the extend model in equation (8). The population size is 1500 and sample sizes are 500. In defining the bias, the population target quantities here are defined as the OLS estimates using the 1500 population members and these are depicted as vertical dotted lines on each plot.

Figure 2.4: Application 1 -- Histogram of Sample Weights Constructed from GMM



Note: the weights are constructed from the GMM(C) estimator in specification 2 using the midwifery survey data.

Figure 2.5: Application 1 -- Comparion of Cumulative Age Distributions between Survey Data, Weighted Survey Data and Administrative Data



Appendix

Estimating the Variance of Simulated MSEs, their Differences, and their Ratios

1 Simulation and Notation

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two different estimates of θ. We have conducted a simulation where $\hat{\theta}_1$ and $\hat{\theta}_2$ are computed for each of R independently drawn data sets, indexed by $i = 1, \ldots, R$. Then we have $(\hat{\theta}_{1i}, \hat{\theta}_{2i})$, $i = 1, \ldots, R$. Given this setup, we expect $\hat{\theta}_{1i}$ and $\hat{\theta}_{2i}$ to be correlated, but, for example, $\hat{\theta}_{1i}$ and $\hat{\theta}_{2j}$ are not correlated, where $i \neq j$. Similarly, and more relevantly for the current topic, we expect $(\hat{\theta}_{1i} - \theta)^2$ and $(\hat{\theta}_{2i} - \theta)^2$ to be correlated, but not $(\hat{\theta}_{1i} - \theta)^2$ and $(\hat{\theta}_{2j} - \theta)^2$ where $i \neq j$.

$$\widehat{MSE}(\hat{\theta}_1) = R^{-1} \textstyle\sum_{i=1}^{R} (\hat{\theta}_{1i} - \theta)^2 \text{and } \widehat{MSE}(\hat{\theta}_2) = R^{-1} \textstyle\sum_{i=1}^{R} (\hat{\theta}_{2i} - \theta)^2$$

The key ingredient in these variance estimators is called Δ, where $\Delta_{1i} = (\hat{\theta}_{1i} - \theta)^2 - \widehat{MSE}(\hat{\theta}_1)$ and $\Delta_{2i} = (\hat{\theta}_{2i} - \theta)^2 - \widehat{MSE}(\hat{\theta}_2)$. Here, Δ is the difference between the squared error of a particular simulated estimate, and the average squared error of all of those estimates in the simulation. The MSE expressions square these Δs. Since the Δs already involve squared differences, then the MSE variances involve squares of squares. This makes them very sensitive to extreme values. For this reason, it may sometimes take a surprisingly large value of R before the standard errors become acceptably small.

2 Variance Estimators for MSE (Derivations will be provided upon request)

2.1 Variance of $\widehat{MSE}(\hat{\theta}_1)$

$$\widehat{Var}\left(\widehat{MSE}(\hat{\theta}_1)\right) = R^{-1} \textstyle\sum_{i=1}^{R} \Delta_{1i}^2 \text{ and } \widehat{Var}\left(\widehat{MSE}(\hat{\theta}_2)\right) = R^{-1} \textstyle\sum_{i=1}^{R} \Delta_{2i}^2$$

2.2 Variance of $\widehat{MSE}(\hat{\theta}_1)/\widehat{MSE}(\hat{\theta}_2)$

$$\widehat{Var}\left(\frac{\widehat{MSE}(\hat{\theta}_1)}{\widehat{MSE}(\hat{\theta}_2)}\right) = R^{-1} \widehat{MSE}(\hat{\theta}_2)^{-2} \sum_{i=1}^{R} \left(\Delta_{1i} - \left(\frac{\widehat{MSE}(\hat{\theta}_1)}{\widehat{MSE}(\hat{\theta}_2)}\right) \Delta_{2i}\right)^2$$

2.3 Variance of $\widehat{MSE}(\hat{\theta}_1) - \widehat{MSE}(\hat{\theta}_2)$

$$\widehat{Var}\left(\widehat{MSE}(\hat{\theta}_1) - \widehat{MSE}(\hat{\theta}_2)\right) = R^{-2}\sum_{i=1}^{R}(\Delta_{1i} - \Delta_{2i})^2$$

As long as enough moments are finite, we can apply large-R asymptotics to motivate a normal approximation to the shape of the density of the $\widehat{MSE}$'s and their ratios and differences. This enables the use of the usual normality-based confidence intervals and hypothesis tests.

Simulation Code

Thanks to the *gmm* command in Stata12, complex programming for GMM estimation can be easily performed by using the interactive version of the command. Specifically, we request the two-step GMM estimator, which is based on an initial weight matrix. After a new weight matrix is computed with estimates from the first step, the second step reestimates the parameters based on the new weight matrix. For the initial matrix, we use the *unadjusted* option, which proceeds as if the moment equations are independent and identically distributed. The standard errors reported by *gmm* are heteroskedasticity-robust. The standard errors are smaller than those reported from *regress* with a *vce(robust)* option because *regress* makes a small-sample adjustment to the estimated variance matrix while *gmm* does not. Although the difference is tiny, we adjusted it by taking the loss of degrees of freedom due to the number of regressors into account. In order to avoid encountering convergence problems due to a nearly-singular variance matrix, we set convergence maxiter equal to 100. See the *gmm* command in Stata12 manual for details.

Chapter 3

Hospitalization of Diabetic Patients, and Family Doctors in Primary Care Mixed Payment Models

3.1. Introduction

The effectiveness of physician payment schemes in improving patient health has been debated over the past two decades. Funders face an efficiency-selection trade-off (Newhouse 1996). The traditional fee-for-service (FFS) model, which is a retrospective payment model, has been criticized for over-provision of health care (Evans 1974, Hickson et al. 1987, McGuire 2000). In contrast, prospective payment such as capitation has been suspected to create an incentive to under-provide care and cost-shifting behavior (Ellis and McGuire, 1988, 1993). During the past one and a half decades, Ontario has implemented a series of primary care reforms in terms of physician's remuneration models. Detailed information regarding these reforms is well documented (Hutchison and Glazier, 2013, Sweetman and Buckley, 2014). Two blended funding models have become prevalent: one enhanced FFS model called Family Health Groups (FHG) and one mixed capitation model called Family Health Organizations (FHO). Although such a mixture of payment mechanisms may balance the contrasting incentives to some degree, there is still little empirical evidence on how these models affect primary care for chronic disease management.

Diabetes is a common and serious chronic condition that affects not only on patients' morbidity and mortality, but also imposes a heavy financial burden on health payers. According to the report "An Economic Tsunami: The Cost of Diabetes in Canada" by the Canadian Diabetes Association, the number of people diagnosed with diabetes in Canada doubled in the decade after 2000 and will continue to rise from 2010 to 2020, affecting almost 10% of the population. Another report by the Canadian Diabetes Association states that diabetes cost the Canadian healthcare system and economy $11.7 billion in 2010, and it projects the costs to rise to $16 billion by 2020. [21] Estimates suggest that direct costs doubled from 2000 to 2010, and were about 3.5% of public health care spending as of 2010 (Webster et al. 2011). The prevalence of diabetes was also found to increase substantially from 1995 to 2005 due to rising incidence and declining mortality (Lipscombe & Hux 2007). Harris et al. (2005) investigate the disease burden associated with type II diabetic patients in primary care settings in Canada and find that the disease burden is high since a considerable proportion of these patients in Canada are not well controlled.[22] Since FHG and FHO are the dominant payment models in Ontario, including almost 2/3 of the Ontario family physicians, a comparison of the effects of these models on diabetes primary care is very important.

In this paper, we study the impact of physicians switching from a fee-for-service model (FHG) to a capitation model (FHO) on the hospitalization of diabetic patients in Ontario, Canada. We use several administrative datasets to create longitudinal data for each diabetic patient who enrolled with a family doctor (or general practitioner, GP) who switched from FHG to FHO or stayed in

[21] From Canadian Diabetes Association report "Diabetes: Canada at the tipping point-Charting a New Path" page 2.

[22] They found about 50% of diabetic patients' A1C tests failed to satisfy the target threshold 7.0%. The A1C test is a blood test that provides information about a person's average levels of blood glucose over the past three month and is the primary test used for diabetes management and diabetes research.

FHG.[23] The panel includes five years of observations covering one year before and four years after 2007 when the FHO model was introduced. We compare health care utilization including total hospital admissions, hospital admissions by ambulance and emergency hospital admissions on both the extensive and intensive margins, and length of stay (in days) in hospital for acute care treatment.

We find that hospitalization statistically and economically significantly increased for senior female diabetic patients after their family doctor switched from FHG to FHO. Specifically, total annual hospital admissions, hospital admissions by ambulance and emergency hospital admissions increased by 10.6%, 18.2%, 10.0% respectively. On the extensive margin, the likelihood of occurrences in these three hospital admissions also increased by 9.9%, 16.8%, 9.8% respectively. Moreover, the average annual length of stay in hospital for acute treatment increased by 15.0%. There is little or no impact of the switch from a FHG to a FHO on male patients and female patients in other age groups. When physicians switch from FHG to FHO, the incentive to spend time with senior and sick patients may be reduced since the cost of seeing them becomes higher, although it is unclear why the impact is much stronger on senior females than on senior males. The financial risk that a FHO doctor bears may result in a different practice pattern. These results suggest that GP's practicing in the FHO model generates higher secondary care for senior female diabetes. While we have no evidence at this stage, one possibility is that older women with diabetes are higher cost for practices and they are therefore more likely to be hospitalized under capitation than the other age-sex groups. Alternatively and/or additionally, on the margin they may benefit the most from the hospitalization and freed from financial incentives to treat patients themselves. FHO physicians are more likely to facilitate their hospitalization. Future research will need to address these alternative non-exclusive explanations.

[23] If a patient whose enrolment relationship with a primary care physician ended or his/her primary care physician switch between payment models rather than from FHG to FHO during the sample period, he/she is excluded in our analysis.

The remainder of the paper is organized as follows. Section 2 discusses the related literature. Section 3 describes the data and presents summary statistics. Section 4 presents our empirical method. Estimates from the regression analysis are discussed in Section 5, as are those from sensitivity tests and sub-group analysis. Section 6 discusses the findings and concludes.

3.2. Literature Review

As discussed by Thomas G. McGuire in The Oxford Handbook of Health Economics chapter 25 (McGuire, 2010), in the context of patient enrollment with primary care physician, a physician may care about the welfare of her patient because of form of altruism. However, as a utility maximizer, a physician also has incentive to pursue her own self-interest with respect to factors such as income and leisure. Because of the trade-off, the funding mechanism becomes crucial. Since both the traditional FFS and capitation models have been criticized, mixed payment models have been advocated as these models may achieve a socially optimal output of health care with good quality (Léger 2008 and McGuire 2008). Under blended funding, a physician's total funding comprises a mixture of payment mechanisms to optimally balance the contrasting incentives. Physician incentives in the mixed payment models in Ontario Canada have been studied (Kantarevic et al. 2011, Kralj and Kantarevic, 2013, Kantarevic and Kralj, 2014). These studies have investigated, among other things, physician outcomes such as services provided, preventive care bonuses obtained, and patient enrollment. However, little is known about patients' health outcomes or health care utilization as a function of physician incentives in different payment models.

First, as discussed in the last paragraph, this paper is related to diabetes management and physician incentives. Policies on physician fees and incentives have been implemented, and research on the effects of these incentive programs, particularly for diabetes management, have been conducted in many countries. For instance, the effects of a pay-for-performance (P4P) incentive program in a doctor's payment scheme on the diabetes management in primary care in Australia have been studied by Scott et al. (2009). Kantarevic and Kralj (2013) investigate a P4P incentive in two physician payment models in Ontario, where the P4P payment is based on the percentage of enrolled diabetic patients who received Diabetes Management Incentive services. They find that patients enrolled with a doctor in a blended capitation model are more likely to receive diabetes management services than those who enrolled with a doctor in an enhanced fee-for-service model. Kiran et al. (2014) find that patients enrolled in blended capitation models receive higher quality diabetes care measured by the optimal number of three recommended monitoring tests, than those enrolled in blended fee-for-service models. Chen et al. (2011) find that older patients and patients with more co-morbidities or severe conditions are prone to be excluded from a diabetes mellitus P4P program in Taiwan. Given the current policy focus on reducing the medical care cost, it is interesting to investigate hospital admissions of diabetic patients, especially senior diabetic patients, as a function of their GPs' payment models.

Second, this paper also contributes to the literature on incentive mechanisms in physician remuneration models and patients' hospitalization. Lippi Bruni et al. (2009) find that in the Emilia Romagna region of Italy, patients registered with family practitioners who received more income from participation in two financial incentives practice programs, especially pay-for-participation, had a significantly lower probability of hyperglycemic emergencies. However, their results cannot be interpreted as a causal relationship because of some unobserved information about patients such as diet and exercise, which can be strongly correlated with financial incentives. Fiorentini et al. (2011) investigate the impact of three financial incentive programs on the probability of avoidable hospitalization. However, their conclusion was cautionary due to the limitation in the use of cross-sectional data. Dusheiko et al. (2011[b]) avoid this problem by applying panel data methods to examine the association between primary care management and hospital costs and they find only primary care performance in stroke care among 10 chronic diseases is associated with reduced hospital costs. Similar work by Échevin and Fortin (2014) with a focus on the payment mechanisms of specialists shows the length of stay in patients treated in the department with an optional mixed compensation scheme increases on average by 4.2%. Difference-in-differences estimations show that removing incentive does not result in a worse performance of doctors in terms of expenditures for avoidable and total hospital admissions of their patients (Fiorentini et al., 2013) or patient 'dumping' or cost shifting behavior of doctors (Kantarevic and Kralj, 2013).

Third, this paper is also related to the literature on primary care and hospitalization. Hutchison et al. (1996) find that a capitation payment model with an additional payment to encourage low hospital utilization rates does not reduce hospital use. Basu et al. (2004) find that in the USA, private health maintenance organization (HMO) enrollment has fewer preventable hospital admissions than private FFS. In contrast, such difference was not observed among Medicaid adults between HMO and FFS. Kjekshus (2005) find that enhanced interaction between primary and secondary care has no impact on the length of stay in hospital. Parchman et al. (1999) find medicare beneficiaries in primary care shortage areas are more likely to experience preventable hospitalization. Bottle et al. (2007) study the association between quality of primary care and hospitalization for coronary heart disease. The relationship between payment model and hospital referrals has been investigated recently (Ho & Pakes, 2014). They find that patients enrolled with a high-capitation insurer tend to be referred to lower-priced hospitals with no reduction in quality.

While one motivation from the Ontario government for its primary care reforms is to improve quality (Hutchison and Glazier, 2013), it is not yet clear whether the goal has been achieved, so it remains an open question. Although we have no direct measures of quality in this paper, some literature on the association between diabetes care and hospital admissions can help us for future research. Saxena et al. (2006) use a one-year cross-sectional data to find the proportion of family practitioners who were offering health promotion clinics for diabetes in a Primary Care Trust is negatively associated with both emergency and elective admission rates for diabetes. Bottle et al. (2008) found a significant but weak negative association between primary care quality scores and hospital admission for senior patients with age 60 and over. Dusheiko et al. (2011[a]) investigate both cross-sectional and longitudinal association of diabetes care quality and hospital admission, and also find a negative relationship between improvement of quality of diabetic care in a family practice and hospital admissions. However, their finding is based on practice level which does not control for individual patients' characteristics. Purdy et al. (2009) argue that rates of hospital admission for ambulance or primary care sensitive conditions can be used as a measure of the effectiveness of primary care. Iezzi et al. (2014) use the number of yearly hospitalizations for diabetic Ambulatory Care Sensitive Conditions among patients affected by diabetes type II on each GP's list and an indicator of quality of primary care and find that hospitalization is negatively correlated with financial incentives, which is defined as all payments received by GPs, for those activities aimed at improving the delivery of medical services to diabetes patients.

3.3. Data

3.3.1 Data Sources and Sample

The data come from administrative databases maintained by the Ontario Ministry of Health and Long-Term Care. They include Ontario Health Insurance Plan (OHIP) claims data, which documents the medical services provided by every doctor in Ontario; the corporate provider database (CPDB), which documents GPs' payment model and related information; the registered persons database (RPDB), which documents patients' characteristics; the client agency program enrolment database (CAPE), which documents the history of patients' enrolment status; and the inpatient hospital discharge abstract database (DAD), which documents hospitalization admissions. Since all databases can be linked with encrypted unique IDs for every patient and GP, we can create a comprehensive longitudinal data for diabetic patients' hospital admissions as well as their GPs' payment models.

Before the capitation-based FHO model was introduced, the fee-for-service-based FHG was by far the most popular primary care reform model. After the FHO model was implemented, the number of physicians in FHGs started to decline as many switched to FHOs. Both of these payment models are blended payment model and provide the same incentives for patient enrolment and preventative care bonuses. A comparison of FHG and FHO is shown in Table 3.1. The main difference is that physicians in the FHO model receives an age-sex adjusted capitation rate for the core services (over 100 comprehensive care services) provided to their rostered patients. In addition, they receive 15% of the FFS value of core services (i.e. shadow billing premium) provided to their rostered patients. As for non-rostered patients, physicians receive the full FFS value of core services up to $52,883 annually (as of Dec 31, 2013). Both models require evening and weekend clinics and provide incentive payments for immunizations, cancer screening, smoking cessation and chronic disease management such as diabetes care management. The percentage of family physicians in FHG reached a maximum in early 2006. To maximize the number of observations, we use fiscal year 2006 (April 1, 2006 to March 31, 2007) as the base year, and follow the patients and their GPs in the following years. The observation period covers one year before and four years after the FHO model was introduced.

During the five-year period between March 31, 2006 and March 31, 2011, primary care physicians could switch between payment models, unenroll existing patients or start enrolling new ones. CPDB and CAPE data allow us to identify patients who enrolled with a family doctor as of March 31, 2006 through the five-year period to March 31, 2011. Since a comparison between the two models is our focus, we restrict our sample to those diabetic patients who maintained the enrolment relationship with their enrolling doctors during the sampling period. We exclude any physician who switched back and forth between FHG and FHO. In other words, all these GPs were affiliated with a FHG model as of April 1, 2006, and either made only one switch from FHG to FHO, or always stayed in FHG during the period. Since the majority of physicians who joined the FHO model were previously in the FHG model, we capture a sizeable share of primary care physicians and patients.

Summary statistics are presented in Table 3.2. The unit of analysis is the patient. We identify patients with diabetes mellitus through a validated algorithm which has been developed and applied by researchers from the Institute for Clinical Evaluative Sciences (ICES) (Blanchard et al. 1996; Hux et al. 2002; Lipscombe and Hux, 2007; Guttmann et al. 2010). In total, we identify 159,471 diabetic patients who during the period maintained their relationship with 2,999 GPs, among whom 1,354 switched from FHG to FHO and 1,645 stayed in FHG. This captures roughly 30% of the family physicians in Ontario as of April 1, 2006. The enrolment rate was around 25% for all Ontario residence and about 35% of the identified diabetic patients enrolled with these 2,999 GPs at that time. Among these enrolled diabetic patients, 71% and 64% of them maintained their enrolment relationships with their GPs during the sampling period. Moreover, age and sex of the patients who were 'unenrolled' in the treatment group and comparison group are not different. Within each group, age and sex of these 'unenrolled' patients are also not different from those who maintained. Similarly, the average ages of patients in the two groups are not statistically different. The male percentage in the treatment group is only 2% bigger than the comparison group. The patients in the treatment group are less likely to reside in urban areas and their GPs also practice in less urban areas. The degree of rurality of a community is measured by the Rurality Index of Ontario (RIO), which was developed by the Ontario Medical Association (Kralj, 2000). Specifically, it was generated by variables including travel time to nearest basic referral center, travel time to nearest advanced referral center, community population, number of active GPs, population to GP ratio, presence of a hospital and etc. RIO gives each community a score with a range from 0 to 100. Statistics Canada defines the rural population as persons living outside centres with a population of 1,000 and outside areas with 400 persons per square kilometre, so roughly 14% to 15% of the Ontario population between 2006 and 2011 are rural population.[24] We define rural area with a RIO larger than a threshold which includes 14% of our observations. We do not see any noticeable selection on a patient's age and sex by GPs. Although not shown in the table, no significant difference was found between male and female GPs in terms of the percentage of their enrolled patients who maintained the relationship during the sampling period.

3.3.2 Variable Specification

The dependent variables are inpatient hospitalizations[25] including total annual hospital admissions, hospital admissions by ambulance, emergency hospital admissions and the total length of stay in hospital for acute care treatment in the year. An emergency hospital admission is defined as an urgent or emergency admission under the admission category, so elective admissions, newborn, and

[24] See http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo62g-eng.htm

[25] Ontario institutions submit all day surgery abstracts to National Ambulatory Care Reporting System (NACRS), which unfortunately we do not access to. Ontario DAD contains demographic, administrative and clinical data for inpatient hospital discharges (acute, chronic and rehabilitation).

other categories are excluded. This variable is interesting since Dusheiko et al. (2011[b]) show that good primary care management is associated with reduced hospital cost due mainly to reductions in emergency admission, not lower costs per patient treated in a hospital. To avoid some outliers that could potentially bias the results, we top code the annual hospital admissions by 5 visits and length of stay in hospital for acute care treatment by 60 days.[26] On the extensive margin, we create three dependent variables, which are dummy variables indicating at least one hospital admission, one hospital admission by ambulance, and one emergency hospital admission in each year. The summary statistics of the dependent variables are shown in Table 3.3. We also test the hospital admissions associated with diabetes, and the results are similar to total hospital admissions. We define the treatment group as those diabetic patients whose GPs switched from FHG to FHO over the sampling period. A histogram of the switch days is shown in Figure 3.3. As shown, a majority of the GPs switched from FHG to FHO in 2008. We also include a set of variables such as age, sex, location of residence for patients, and age and sex for GPs for subgroup analysis.

3.4. Empirical Method

Since our sample includes patients who maintained a relationship throughout our data period with their GPs, we are concerned about the potential bias from non-random selection of patients by GPs in the treatment group and comparison group and the unobserved patient characteristics that can be confounded with the switch decision by a GP and hospitalizations. Moreover, other policy

[26] Less than 0.1% of the observations are top-coded and some of these outliers have over 25 annual hospitalizations or over 100 days of stay in hospital for treatment. To argue our results are not driven by these outliers, we exclude them. Also, we alter the cut-offs from 5 hospitalizations to 10 hospitalizations or 60 days to 30 days or other numbers, and we do not find any appreciable difference in the estimates in our results.

initiatives targeted at the patients such as pay for performance at the diabetes management incentive may be correlated with the timing of the switch.

3.4.1 Identification Strategies

Observed patient characteristics, such as age and sex, and unobserved ones, may be correlated with a physician's decision to switch from FHG to FHO. Including these types of information can help mitigate bias, but may not eliminate selection bias. Hence, we employ a difference-in-differences (DiD) approach to eliminate any individual patient time-invariant fixed effects that may be correlated with hospitalization and their GP's decision to switch.

The identifying assumption for DiD is that the trends of hospitalization are parallel between the treatment and comparison groups. We could compare hospitalization during the pre-treatment period between the two groups. However, some patients were enrolled prior to April 1, 2006 with GPs other than the one with which they were enrolled during the study period, so their hospitalization would not reflect the primary care from their GP who they enrolled with during the sample period. Although we cannot test our identifying assumption, we argue that the patients in the comparison group and the control group are similar, based on their by age and sex distribution as shown in Table 3.2. Figure 3.1 shows that a patient's hospitalization pattern is a function of age, with a marked increase starting in early 60s. The similarity of patients' characteristics between the comparison and the control group suggests that the change in hospitalization differences over time between the two groups is not driven by the aging of the patients.

3.4.2 Empirical Specifications

Our empirical specification is a standard fixed effects model.

$$Hosp_{it} = \sum_{t=1}^{4} \beta_{T_t} \cdot T_t + \beta_{FHO} \cdot FHO_{it} + (\beta_0 + \sum_{i=1}^{N-1} \beta_i \cdot D_i) + u_{it} \quad (1)$$

where $Hosp_{it}$ represents hospital admission for patient i in year t; $\beta_0 + \sum_{i=1}^{N-1} \beta_i \cdot D_i$ is the set of time-invariant patient fixed effects, where the β_i, which is treated by some researchers as part of the error term, is treated here like a nuisance parameter and is not estimated in accord with the practice for the standard fixed effect model; β_{T_t} is the year fixed effect, which captures any exogenous shock in a particular year and the trend of utilization during the sample period due to aging, especially for the senior patients; FHO_{it} is the treatment indicator that equals to 0 for the years before the patient's physician switched and 1 after. For the year when the switch occurred, this variable measures the proportion of the fiscal year during which their GPs were affiliated with the FHO model. Here we apply a DiD fixed effect estimation to remove individual fixed unobserved characteristics. Differences in hospital admission are calculated for each patient over the sampling period. β_{FHO} identifies the treatment effect from the remuneration model switch. [27] Patients who have the same family doctor may have similar unobserved characteristics and could be correlated within practices. Hence, we use heteroskedasticity and autocorrelation consistent standard errors clustered at the doctor level.

As shown in Figure 3.1 and Figure 3.2, a patient's hospitalization patterns on both intensive and extensive marge, are a function of age, with a marked increase starting in early 60s.[28] For young patients the hospitalization pattern is not as stable as the middle-aged. As a result, we split patients into three age categories: young patients with age less than 35, middle-aged patients with

[27]A falsification test is conducted with only the FHG group. Half of the FHG group is randomly assigned to be a counterfactual group who switched from FHG to FHO. Three sets of switch dates were used for the sensitivity test. We use April 1, 2009, mid-point of the sampling period, and randomly assigned dates which follow a uniform distribution along the sample period for each counterfactual.
[28] For confidentiality, age is bottom-coded at 5 and top-coded at 90 in Chart 1 and Chart 2. The hospital admissions are high and have more variations for those who are older than 90.

age from 35 to 64, and senior patients with age greater or equal to 65 at March 31, 2006. We set up this specification to separate the FHO impact on patients in different age categories. [29]

$$Hosp_{it} = \sum_{t=1}^{4} \beta_{y_{T_t}} \cdot T_t \cdot Age_{yi} + \sum_{t=1}^{4} \beta_{m_{T_t}} \cdot T_t \cdot Age_{mi} + \sum_{t=1}^{4} \beta_{s_{T_t}} \cdot T_t \cdot Age_{si} + \beta_{FHO_y} \cdot FHO_{it} \cdot Age_{yi} + \beta_{FHO_m} \cdot FHO_{it} \cdot Age_{mi} + \beta_{FHO_s} \cdot FHO_{it} \cdot Age_{si} + (\beta_0 + \sum_{i=1}^{N-1} \beta_i \cdot D_i) + u_{it} \quad (2)$$

where Age_{yi}, Age_{mi}, and Age_{si} are dummies for the three age groups, and β_{FHO_y}, β_{FHO_m} and β_{FHO_s} show the impacts on patients in the three age categories.

The interpretation of β_{FHO} in model 1 and β_{FHO_y}, β_{FHO_m} and β_{FHO_s} in model 2 depends on the assumption that FHO_{it}, conditional on the individual fixed effect and year dummies, is exogenous in the sense that it is not correlated with the idiosyncratic error term in any period ($\mathrm{E}[FHO_{is}u_{it}] = 0 \ for \ s,t = 1,2,\ldots,5$). The switch decision is not a choice of a patient, but his GP's unilateral decision. The fact is that many patients are unaware that they have been enrolled or any changes in their GPs' remuneration models (Sweetman and Buckley, 2014). This could cause an endogenous selection bias as a GP's decision making may depend on patients' hospitalization in previous years. Hence, we test whether hospital admissions in the previous period before the switch are correlated with a GP's decision to switch his payment model ($\mathrm{E}[FHO_{it+1}u_{it}] \neq 0 \ for \ t = 1,2,\ldots,4$). In this case, we can test such a possibility by seeing whether FHO_{it+1} is strongly correlated with $Hosp_{it}$ from the following estimation:

$$Hosp_{it} = \sum_{t=1}^{4} \beta_{T_t} \cdot T_t + \beta_{FHO_0} \cdot FHO_{it} + \beta_{FHO_1} \cdot FHO_{it+1} + (\beta_0 + \sum_{i=1}^{N-1} \beta_i \cdot D_i) + u_{it}$$

[29] We alter the cut-offs to less than 30, 30 to 60, and above 60, for the three age groups and find the results are not sensitive to the cut-offs chosen. For senior female patients, the estimates for β_{FHO_s} become slight smaller and less significant.

(3)

If β_{FHO_1} is significant, it suggests that a change of hospitalization may occur one year before the actual switch. If it is not significant while β_{FHO_0} remains significant, we feel confident to accept the null hypothesis that a physician's switch decision is not based on the patients' hospital admissions in the previous year in his rostering list. We can also think of this procedure as testing the assumption that the FHO impact occurs only in years after the switch, but not in the previous years.

Secondly, although the FHO model was implemented in 2007, Figure 3.3 shows that the timing of switch occurred throughout the sampling period. It is doubtful that the FHO impact may come from some unobserved factor in any particular year. We check the assumption that the impact is independent of the timing of switch with the following estimation:

$$Hosp_{it} = \sum_{t=1}^{4} \beta_{T_t} \cdot T_t + \beta_{FHO_0} \cdot FHO_{it} + \beta_{FHO_{yr}} \cdot FHO_{it} \cdot yr_i + (\beta_0 + \sum_{i=1}^{N-1} \beta_i \cdot D_i) + u_{it} \tag{4}$$

where yr includes 2007, 2008, 2009 and 2010, and yr_i is a dummy variable that equals to 1 if patient i's GP switched from FHG to FHO in year yr. $\beta_{FHO_{yr}}$ captures the difference in the impact from switching to FHO when it occurred in year yr compared with the impact when the switch occurred in the other years. If $\beta_{FHO_{yr}}$ is not significant, this test then accepts the null hypothesis that the FHO impact is not due to some potential confounder that happened in these four years. For one example, on April 1, 2009, the value of the diabetes management incentive (DMI) increased from 60 to 75 dollars. Although FHG and FHO are identical with respect to leave for the DMI, this procedure tests whether such a policy initiative in 2009 has different impacts on the hospitalizations of patients in comparison and treatment groups.

3.5. Results

3.5.1 Main Results

In table 3.4 we report estimates for specification 1 and specification 2. Overall, the results from all the sample show that a patient's hospital admissions increased after his GPs switched from the FHG to the FHO model. The marginal effects from the DiD fixed effect estimation shows that the total annual hospital admissions increase by 0.004, which seems small. However, one needs to compare the marginal effects with the mean of hospitalizations in Table 3.3. Specifically, total annual hospital admissions, hospital admissions by ambulance and emergency hospital admissions increased by 3.2%, 8.7%, 4.2% respectively due to their GPs' joining the FHO model. The likelihood of these three hospitalizations to occur also increased by 3.5%, 7.5%, 4.6% respectively. Moreover, the yearly length of stay in hospital for acute treatment increased by 6.6%. All of these estimates are statistically significant at 10%. The coefficients on annual hospital admissions by ambulance on both intensive and extensive margin, and length of stay, are all statistically significant at 1%.

Including the interaction terms between different age dummies and FHO_{it} allows us to further investigate the impact of the switch on patients in these three age categories. As the year dummies capture the trends of hospitalizations, which are different among the three groups, we also include the interaction terms between year dummies with the three group dummies. Interestingly, there is little impact on patients in the middle-aged category. Moreover, the impact on the young patients is small and not statistically significant. In contrast, the impact on senior patients is both substantively and statistically significant. Specifically, total annual hospital admissions, hospital admissions by ambulance and emergency hospital admissions for senior patients increased by 6.0%, 11.1%, 6.9% respectively due to their GPs' joining the FHO model. The likelihood of occurrences for these three hospital admissions also increased by 6.0%, 11.6%, 7.6% respectively. Moreover, the annual length of stay in hospital for acute treatment increased by 8.4%.

3.5.2 Subgroup Analysis

We investigate the impact of joining to FHO on patients in different subgroups by sex and location of residence. As shown in Table 3.5, the coefficient β_{FHO_s} is not statistically significant for males. In contrast, the results are both substantively and statistically significant for senior female patients. Specifically, total annual hospital admissions, hospital admissions by ambulance and emergency hospital admissions increased by 10.6%, 18.2%, 10.0% respectively due to their GPs' joining the FHO model. The likelihood of occurrences for these three hospital admissions also increased by 9.9%, 16.8%, 9.8% respectively. Moreover, the annual length of stay in hospital for acute treatment increased by 15.0%.

Communities with higher rurality index for Ontario (RIO) are classified as more rural (Kralj, 2000). In this paper, we define urban area with RIO equals to 0, which accounts for 46% of the sample; suburban area with RIO larger than 0 and smaller than or equal to 20, which accounts for 40% of the sample; and rural area with RIO larger than 20, which accounts for 14% of the sample. Shown in Table 3.6, the impact of a switch on patients who are in urban areas is small and statistically insignificant. In contrast, when the sample is restricted to suburban area, the FHO coefficients for total annual hospital admissions and hospital admissions by ambulance become at least as three times as big as the ones when the sample is restricted to urban area, and they are statistically significant. However, the FHO impact on the likelihood of these three hospital admission occurrences (i.e., on the extensive margins) for those who are in suburban areas are similar as those who are in urban areas. Interestingly, for senior patients who are in rural areas, the

FHO impact on the three hospital admissions remains relatively big, but statistically insignificant. In contrast, the FHO impact on the extensive margins of the three hospital admissions for rural diabetic patients is big and statistically significant. Overall, the FHO impact is stronger for patients in suburban and rural areas. Finally, we split our sample into groups based on GPs' gender and age categories. As shown in Table 3.7, the impacts from male GPs and GPs with age between 40 and 60 at March 31, 2006 are statistically stronger. The relatively statistically insignificant impact from female GPs and GPs in other age groups could be due to smaller numbers of observations. In fact, the size of FHO impact from female GPs are larger than those from male GPs in most hospital utilizations except "length of stay in hospital" in the last column.

3.5.3 Sensitivity Tests

We explore the exogeneity assumption by performing the tests as discussed using specifications 3 and 4. First of all, we artificially move the switching date a year ahead for the treatment group and see whether the impact is still significant. The results for specification 3 in table 3.8 show that the coefficients of year prior to switch are statistically insignificant while the coefficients of year of switch remain statistically significant. This suggests that the possibility that GP's decision making is based on the patients' hospital utilization in the previous period is low. Specification 4 shows whether the impact of switch in each year is different from the overall impact. This evidence suggests that the impact of a switch is independent of its timing since the coefficients on $FHO_{it} \cdot yr$ are not statistically or economically significant. For instance, as the coefficient on $FHO_{it} \cdot 2009$ is not statistically or economically significant, we accepts the null hypothesis that the FHO impact is not due to the Diabetes Management Incentive (DMI) policy initiative, which could be a confounder that is correlated with both patients' hospitalizations and the switch decisions. As over 40% of GPs in our treatment group switched from FHG to FHO in 2008, FHO_{it} and $FHO_{it} \cdot 2008$ become highly correlated, which causes β_{FHO_0} not significant at 10%. However, it remains economically significant. The results are more robust for senior female samples.

The three falsification tests in Table 3.9 show that the artificial switch dates have no significant impact on the utilizations, supporting the validity of the comparison group. We cannot test the common trend assumption because there is only one year observation prior to the change.

3.6. Discussions and Conclusion

3.6.1 Causality?

We are cautious about drawing inference of causality from our results since the decision of switch from FHG to FHO is a choice of a doctor and the exogeneity of the switch is debatable. There is a potential omitted variable bias since some unobserved characteristics of a GP can be confounders that are correlated with both patients' hospitalizations and the switch decisions. This could be related to the findings in sub-group analysis that the FHO impact is stronger for patients in suburban and rural areas and patients whose GP is male with age between 40 and 60. Although we cannot control the providers' selection into FHO, the panel data allows us to control the individual fixed effects which avoid the omitted variable bias from patients that may be potentially correlated with hospitalization and their GP's decision to switch. The robustness checks by modifying the baseline specification show that the date of switch is not correlated to the hospitalizations in the year prior to the switch (i.e. specification 3) and the FHO impact is not sensitive to the switch date in any particular year (i.e. specification 4).

Our results have some policy-relevant interpretations. Although the treatment group in this paper is not randomly assigned, but is a choice by a family doctor, the estimate of the FHO impact comes from the fact that all physicians have the option to switch (or are offered to be treated), analogous to the "intent to treat" literature (Heckman et al., 1999). We can consider those who did not take up the FHO opportunity as attrition. The experimental mean difference in our case estimates the mean effect of the offer of treatment, so our estimates capture how the availability of a program affects participant outcomes.

3.6.2 Conclusion and Future Research

In this study, we investigate the impact of a GP's remuneration model change on patients' hospital admissions. Using comprehensive administrative data, we construct a panel for diabetic patients and employ a difference-in-differences approach to identify the impact. We find that on both intensive and extensive margin, the hospital admissions for senior female patients significantly increased after their GP's remuneration model changed from FHG to FHO. In contrast, the impact on male patients and other female patients is small. For senior female patients, total annual hospital admissions, hospital admissions by ambulance and emergency hospital admissions increased by 12.4%, 23.6%, 12.3% respectively due to their GPs' joining the FHO model. The likelihood of occurrences for these three hospital admissions also increased by 11.8%, 22.4%, 12.2% respectively. Moreover, the annual length of stay in hospital for acute treatment increased by 18.2%. The prospective payment on each patient gives the fixed benefit to a FHO doctor, but the marginal cost varies differently among patients in different age categories. A FHO doctor may have less incentive to provide a lot of time to sick or senior patients when the marginal cost is too high on the doctor. However, it is not clear why the impact is strong on senior female patients, but not senior male patients.

The results provide a cautionary message regarding to the differences in practice patterns towards senior diabetic patients between GPs in the two models. Although Kantarevic and Kralj (2013) find GPs in FHO are more responsive to the Diabetes Management Incentive (DMI) than GPs in FHG, we find that the hospitalizations of patients whose family doctor switched from FHG to FHO did not decrease, and in fact, increased. If utilization of preventive services in the diabetes management incentive results in lower hospitalization, especially the emergency ambulatory admissions, the incentive should emphasize senior diabetes rather than the percentage of enrolled diabetic patients who received DMI services regardless of their ages. Seshamani and Gray (2004) show that the assessed average hospital cost for females is significantly larger than for males. Given the current policy focus on reducing medical care costs, proper control for avoidable hospital admission for diabetic patients seems crucial.

Rehospitalisation and Ambulatory Care Sensitive Conditions have been used as indicators in quality primary care, but we are cautious about drawing conclusions on the quality of care. First, as Jencks et al. (2009) find that rehospitalisation among Medicare beneficiaries in the USA are prevalent and costly, we also estimate models that have 30-Day, 60-Day and 90-Day rehospitalisation as dependent variables. However, these models give similar results on rehospitalisation, which is consistent with Hansen et al. (2011) for a systematic review of interventions to reduce rehospitalisation. Second, hospital admissions for diabetic Ambulatory Care Sensitive Conditions are identified from hospital records with ICD-9 code 250 and can be used as an indicator of quality on type 2 diabetic patients (Iezzi et al. 2014), but we are not able to identify type 2 diabetes due to lack of patients' pharmaceutical information. Hence, this paper says nothing about the quality of care, but it may shed some light on some practice patterns or service strategies, especially towards to senior patients with chronic conditions.

Theoretically a physician in a capitation model may need to care about his enrolled patients' health in the long term to avoid the high cost from the patients who get sick. Empirically, a capitated doctor has been found to deliver more preventive healthcare services (Kralj and Kantarevic, 2013). However, when the cost of an enrolled patients get so high that a capitated doctor may be more likely to facilitate their hospitalization. Future research will need to address these alternative non-exclusive explanations and investigate the both quantity and quality of primary care services by a FHO doctor compared to a FHG doctor, especially for senior patients in chronical conditions. Forrest et al. (2003) find an increase in the number of discretionary referrals among patients in plans with capitated primary care physician payment and that these referrals were more likely to be for chronic condition those from a FFS primary care physician. Allard et al. (2011) construct a theoretical model to show that capitation induces the most referrals while FFS has the fewest. Recently, Ho and Pakes (2014) investigated the physician payment reform and hospital referral. As cost of senior patients becomes high, further research may investigate see whether the doctors who are in blended capitation model have "patient dumping" behavior or providing more referrals to specialists.

References

Allard, M. Jelovac, I., and Leger, P.T., 2011. Treatment and Referral Decisions under Different Physician Payment Mechanisms. Journal of Health Economics, 30(5), 880-893.

Basu, J., Friedman, B., Burstin, H., 2004. Managed Care and Preventable Hospitalization among Medicaid Adults. Health Services Research, 39, 489–510.

Blanchard, J.F., Ludwig, S., Wajda, A., Dean, H., Anderson, K., Kendall, O., Depew, N., 1996. Incidence and Prevalence of Diabetes in Manitoba, 1986-1991. Diabetes Care, 19, 807–811.

Bottle, A., Gnani, S., Saxena, S., Aylin, P., Mainous, A.G., Majeed, A., 2007. Association between Quality of Primary Care and Hospitalization for Coronary Heart Disease in England: National Cross-sectional Study. Journal of General Internal Medicine, 23, 135–141.

Bottle, A., Millett, C., Xie, Y., Saxena, S., Wachter, R.M., Majeed, A., 2008. Quality of Primary Care and Hospital Admissions for Diabetes Mellitus in England. Journal of Ambulatory Care Manage, 31, 226–238.

Canadian Diabetes Association, 2009. An Economic Tsunami.

Canadian Diabetes Association, 2011. Diabetes: Canada at the Tipping Point-Charting a New Path.

Chen, T. T., Chung, K. P., Lin, I. C., Lai, M. S. 2011. The Unintended Consequence of Diabetes Mellitus Pay-for-performance (P4P) Program in Taiwan: Are Patients with More Comorbidities or More Severe Conditions Likely to Be Excluded from the P4P Program? *Health Services Research*, *46*(1), 47–60.

Dusheiko, M., Doran, T., Gravelle, H., Fullwood, C., Roland, M., 2011[a]. Does Higher Quality of Diabetes Management in Family Practice Reduce Unplanned Hospital Admissions? Health Services Research, 46:1, 27–46.

Dusheiko, M., Gravelle, H., Martin, S., Rice, N., Smith, P.C., 2011[b]. Does Better Disease Management in Primary Care Reduce Hospital Costs? Evidence from English Primary Care. Journal of Health Economics. 30, 919–932.

Ellis, Randall P., and Thomas G. McGuire, 1988. Insurance Principles and the Design of Prospective Payment Systems. Journal of Health Economics. 7:3, 215-237.

Ellis, Randall P., and Thomas G. McGuire, 1993. Supply-Side and Demand-Side Cost Sharing in Health Care. Journal of Economic Perspectives. 7:4, 135-151.

Evans, R. G. 1974. Supplier-induced Demand: Some Empirical Evidence and Implications. In the Economics of Health and Medical Care, M. Perlman (ed.). New York: John Wiley. 162-173.

Échevin, D., Fortin, B., 2014. Physician Payment Mechanisms, Hospital Length of Stay and Risk of Readmission : Evidence from a Natural Experiment. Journal of Health Economics. 36, 112-124.

Fiorentini, G., Iezzi, E., Lippi Bruni, M., Ugolini, C., 2011. Incentives in Primary Care and Their Impact on Potentially Avoidable Hospital Admissions. The European Journal of Health Economics. 12, 297–309.

Fiorentini, G., Lippi Bruni, M., & Ugolini, C., 2013. GPs and Hospital Expenditures. Should We Keep Expenditure Containment Programs Alive? Social Science & Medicine, *82*, 10–20.

Forrest, C.B., Nutting, P., Werner, J.J., Starfield, B., Schrader, V., Rohde, C., 2003. Managed Health Plan Effects on the Specialty Referral Process Results from the Ambulatory Sentinel Practice Network Referral Study. Medical Care, 41, 242–253.

Guttmann, A., Nakhla, M., Henderson, M., To, T., Daneman, D., Cauch-Dudek, K., Wang, X., Lam, K., Hux, J., 2010. Validation of a Health Administrative Data Algorithm for Assessing the Epidemiology of Diabetes in Canadian Children. Pediatric Diabetes 11, 122–128.

Hansen, L.O., Young, R.S., Hinami, K., Leung, A., Williams, M. V., 2011. Interventions to Reduce 30-Day Rehospitalization: A Systematic Review. Annals of Internal Medicine. 155, 520–528.

Harris, S.B., Ekoé, J.-M., Zdanowicz, Y., Webster-Bogaert, S., 2005. Glycemic Control and Morbidity in the Canadian Primary Care Setting (Results of the Diabetes in Canada Evaluation Study). Diabetes Research and Clinical Practice. 70, 90–97.

Heckman, J., LaLonde, R., Smith, J., 1999. The Economics and Econometrics of Active Labor Market Programs," in O. Ashenfelter and D. Card, (eds), Handbook of Labor Economics, North Holland, Vol. 3, 1865-2086.

Hickson, G.B., Altemeier, W.A., Perrin, J.M., Hickson, B., Pernn, M., 1987. Physician Reimbursement by Salary or Fee-for-Service : Effect on Physician Practice Behavior in a Randomized Prospective Study. Pediatrics 80, 344–350.

Ho, K., Pakes, A., 2014. Physician Payment Reform and Hospital Referrals. American Economic Review. 104, 200–205.

Hutchison, B., Birch, S., Hurley, J. Lomas, J. Stratford-Devai, F., 1996. Do Physician-payment Mechanisms Affect Hospital Utilization? A study of Health Service Organizations in Ontario. Canadian Medical Association Journal, 154(5), 653–661.

Hutchison, B., Glazier, R., 2013. Ontario's Primary Care Reforms Have Transformed the Local Care Landscape, But a Plan Is Needed for Ongoing Improvement. Health Affairs. 32, 695–703.

Hux, J.E., Flintoft, V., Ivis, F., Bica, A., 2002. Diabetes in Ontario. Diabetes Care 25, 512–516.

Iezzi, E., Lippi Bruni, M., Ugolini, C., 2014. The role of GP's compensation schemes in diabetes care: evidence from panel data. Journal of Health Economics. 34, 104–120.

Jencks, S.F., Williams, M. V, Coleman, E. A, 2009. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. The New England Journal of Medicine, 360, 1418–1428.

Kantarevic, J., Kralj, B., Weinkauf, D., 2011. Enhanced Fee-for-Service Model and Physician Productivity: Evidence from Family Health Groups in Ontario. Journal of Health Economics. 30, 99–111.

Kantarevic, J., Kralj, B., 2013. Link between Pay for Performance Incentives and Physician Payment Mechanisms: Evidence from the Diabetes Management Incentive in Ontario. Health Economics. 22, 1417–1439.

Kantarevic, J., Kralj, B., 2014. Risk Selection and Cost Shifting in a Prospective Physician Payment System: Evidence from Ontario. Health Policy 115(2-3), 249–57.

Kiran, T., Victor, J.C., Kopp, A., Shah, B.R., Glazier, R.H., 2014. The Relationship between Primary Care Models and Processes of Diabetes Care in Ontario. Canadian Journal of Diabetes 38, 172–178.

Kjekshus, L.E., 2005. Primary Health Care and Hospital Interactions: Effects for Hospital Length of Stay. Scandinavian Journal of Public Health. 33, 114–122.

Kralj, Boris, 2000. Measuring "Rurality" for Purposes of Health-care Planning: an Empirical Measure for Ontario, Ontario Medical Review 67, 33-49.

Kralj, B., Kantarevic, J., 2013. Quality and Quantity in Primary Care Mixed-payment Models: Evidence from Family Health Organizations in Ontario. Canadian Journal of Economics. 46, 208–238.

Léger, Pierre Thomas, 2008. 'Physician Payment Mechanisms', In *Financing Health Care: New Ideas for a Changing Society*, ed. M. Lu and E. Jonsson. New York: Wiley.

Li, J., Hurley, J., Decicca, P., Buckley, G., 2013. Physician Response to Pay-for-Performance: Evidence from a Natural Experiment. Health Economics. Volume 23, Issue 8.

Lippi Bruni, M., Nobilio, L., Ugolini, C., 2009. Economic Incentives in General Practice: the Impact of Pay-for-participation and Pay-for-compliance Programs on Diabetes Care. Health Policy. 90, 140–148.

Lipscombe, L.L., Hux, J.E., 2007. Trends in Diabetes Prevalence, Incidence, and Mortality in Ontario, Canada 1995-2005: a Population-based Study. Lancet 369, 750–756.

McGuire, Thomas. 2000, 'Physician Agency' in *Handbook of Health Economics*, ed. A. J. Culyer and J.P. Newhouse. Vol. 1A. Amsterdam: North-Holland.

McGuire, Thomas. 2008, 'Physician Fees and Behavior: Implications for Structuring a Fee Schedule.' In Incentives and Choice in Health Care, ed. Sloan, F. A., Kasper, H., M.I.T. Press.

McGuire, Thomas, 2010, 'Physician Agency and Payment for Primary Medical Care' in *The Oxford Handbook of Health Economics*, ed. Sherry Glied and Peter C. Smith. Oxford University Press.

Newhouse, J.P., 1996. Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection. Journal of Economic Literature. 34, 1236–1263.

Parchman, M.L., Culler, S.D., 1999. Preventable Hospitalizations in Primary Care Shortage Areas. Annals of Family Medicine, 8, 487-491.

Purdy, S., Griffin, T., Salisbury, C., Sharp, D., 2009. Ambulatory Care Sensitive Conditions: Terminology and Disease Coding Need to Be More Specific to Aid Policy Makers and Clinicians. Public Health 123, 169–173.

Saxena, S., George, J., Barber, J., Fitzpatrick, J., Majeed, A., 2006. Association of Population and Practice Factors with Potentially Avoidable Admission Rates for Chronic Diseases in London: Cross Sectional Analysis. Journal of the Royal Society of Medicine. 99, 81–89.

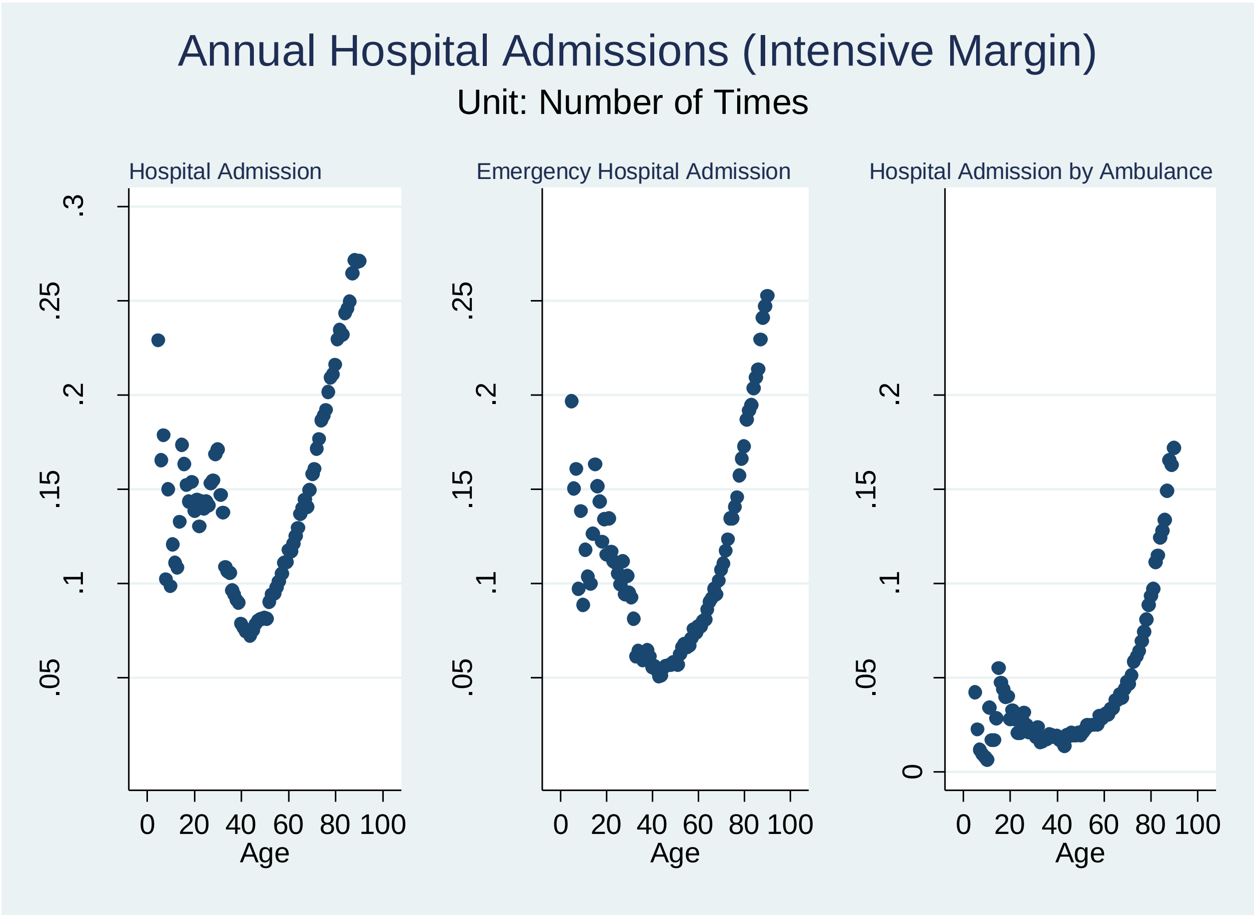
Seshamani, M., Gray, A.M., 2004. A Longitudinal Study of the Effects of Age and Time to Death on Hospital Costs. Journal of Health Economics, 23, 217–235.

Sweetman, A. and Buckley, G. (2014), "Ontario's Experiment with Primary Care Reform," University of Calgary, School of Public Policy, *Research Paper* No. 7-11.

Scott, A. Schurer, S., Jensen, P.H., Sivey, P., 2009. The Effects of an Incentive Program on Quality of Care in Diabetes Management. Health Economics, 18, 1091-1108.

Webster, G., Sullivan-Taylor, P., Terner, M., 2011. Opportunities to Improve Diabetes Prevention and Care in Canada. Healthcare quarterly. 14, 18–21.

Figure 3.1: Mean of Annual Hospital Admissions by Age



Note: We top code the annual hospital admissions by 5 visits. For confidentiality purpose, we bottom code the age at 5 and top code the age at 90. The hospital admissions and rates of hospital admissions in these two age groups are more volatile.

Figure 3.2: Mean of Rate of Annual Hospital Admissions by Age

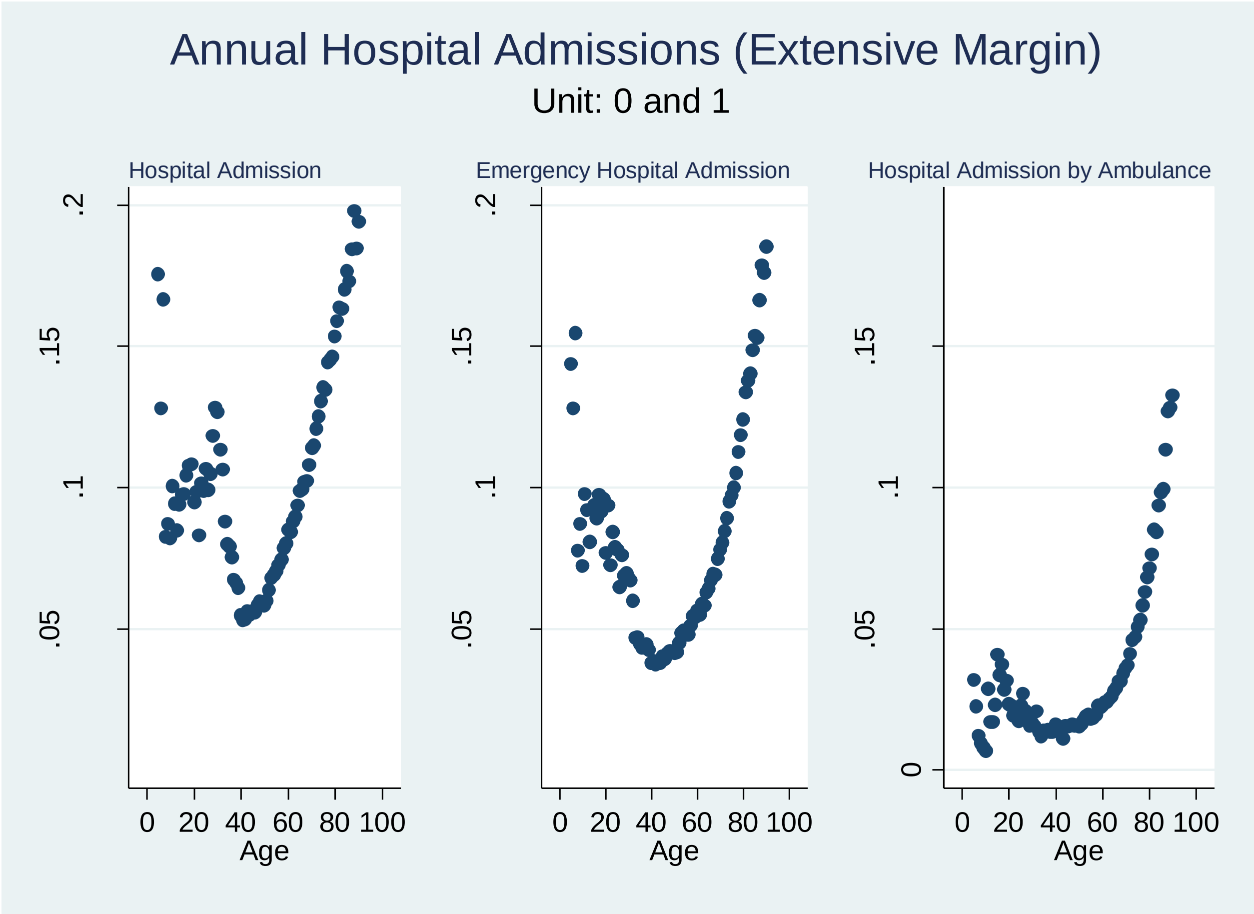
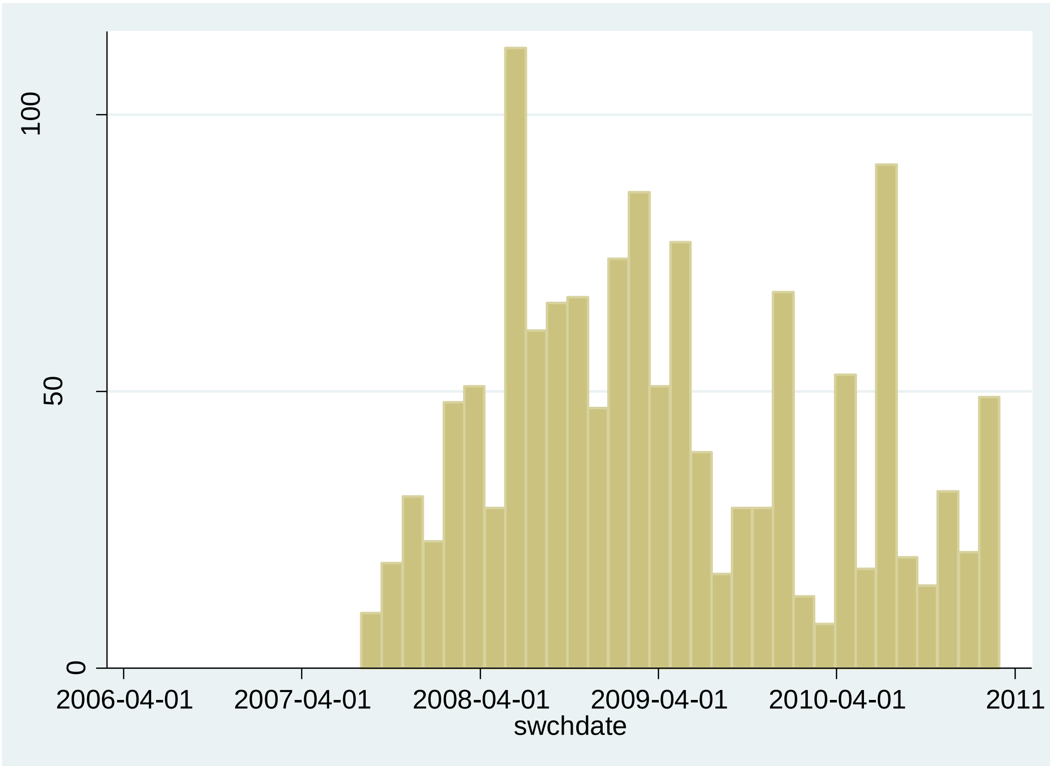


Figure 3.3: Distribution of Switch Dates



Note: 172, 552, 331, and 299 GPs switched from FHG to FHO in 2007, 2008, 2009, and 2010 respectively.

Table 3.1: Key Characteristics of FHG and FHO models

Payment Model	Family Health Group (FHG)	Family Health Organization (FHO)
Year Implemented	2003	2007
Minimum Number of Physicians per Group	At least 3 GPs	Same
Co-operation between Physicians	Share medical records amongst all physicians in the group	Same
After-hours Care Obligations	Physicians must agree to provide after-hours care for a minimum of three hour-blocks per physician in the group	Same
Rostering[1]	Not required	Required
Payment Stream	Blended Fee For Service (FFS)	Blended Capitation
	• 100% FFS as usual • 10% premium on FFS for specified comprehensive care services to rostered patients[2] • Complex/Geriatric-care premium[3]	• Capitation for core services to rostered patients • Access bonus: max 18.59% of the base rate payment • Shadow-billing premium • FFS payments for core services to non-rostered patients

Note: 1. Rostering in FHG is not required but encouraged as most premiums are only paid for enrolled patients. In fact, the size of the roster among FHG doctors is not smaller than FHO doctors.

2. Comprehensive-care premium includes 10 percent premium for 33 codes provided to enrolled patients. Services include, for instance, supportive care, HIV care, diabetes management, palliative care, immunizations, mini-assessment, and home visits.

3. It includes comprehensive-care capitation payments of 15 percent for patients aged 65 or older.

Table 3.2: Summary Statistics of Both Patients and Physicians by Treatment Status

Variable	Total	Comparison group (FHG whole period)	Treatment group (Switched from FHG to FHO)	Difference
Number of physicians	2,999	1,354	1,645	
Physician age (in years) by March 31, 2006	49.79 (9.43)	48.48 (9.06)	50.88 (9.59)	-2.40 (0.34)
Male physician percentage	0.65 (0.48)	0.63 (0.48)	0.66 (0.47)	-0.03 (0.017)
Rurality Index for Ontario (Physician Practice)	6.34 (13.23) (2,985)	9.02 (15.25) (1,345)	4.14 (10.82) (1,640)	4.88 (0.48)
At March 31, 2006				
Enrolled Patients Number of patients	3,393,477	1,623,788	1,769,689	
Patient age (in years) by March 31, 2006	40.65 (22.67)	41.12 (22.79)	40.22 (22.54)	0.90 (0.25)
Male patient percentage	0.45 (0.50)	0.45 (0.50)	0.45 (0.50)	-0.0004 (0.004)
Rurality Index for Ontario (Patient Residence)	7.59 (12.86) (3,349,367)	9.73 (14.27) (1,604,449)	5.63 (11.04) (1,744,918)	4.10 (0.37)
Diabetic Patients with their GPs at April 1, 2006				
Number of enrolled diabetic patients	238,329	109,098	129,231	
Diabetic patient who died during the period	28,838	14,268	14,570	
Patient age (in years) by March 31, 2006	62.56 (14.37)	62.85 (0.53)	62.31 (14.49)	0.54 (0.17)
Male patient percentage	0.52 (0.50)	0.53 (0.50)	0.52 (0.50)	0.01 (0.004)
Rurality Index for Ontario (Patient Residence)	7.35 (12.86) (235,923)	9.79 (14.41) (108,007)	5.29 (10.98) (127,916)	4.50 (0.41)
Final sample for analysis (Diabetic Patients with their GPs from April 1, 2006 till March 31, 2011)				
Number of enrolled diabetic patients	159,471	77,533	81,938	
Patient age (in years) by March 31, 2006	60.96 (13.68)	61.07 (13.76)	60.85 (13.60)	0.22 (1.36)
Male patient percentage	0.53 (0.50)	0.54 (0.50)	0.52 (0.50)	0.02 (0.005)
Rurality Index for Ontario (Patient Residence)	7.20 (12.63) (158,308)	9.56 (14.20) (76,980)	4.96 (10.47) (81,328)	4.60 (0.43)

Note: The entries in the first row for “Patient age”, “Male patient percentage” and “Rurality Index for Ontario” are means. The entries in parentheses in the second row are standard deviation. The entries in the third row for “Rurality Index for Ontario” are number of observations for this variable. There are only less than 1% of our observations with no RIO information due to unmatched postal codes.

Table 3.3: Summary Statistics for the Dependent Variables

Variable	Mean	Std. Dev.	Min	Max
All sample (n=797,355)				
159,471 diabetic patients with 5 years observations				
Intensive Margin				
Total annual hospital admission	0.140	0.490	0	5
Annual hospital admission by ambulance	0.045	0.269	0	5
Annual Emergency hospital admission	0.101	0.417	0	5
Annual length of stay in hospital for acute treatment	0.850	4.158	0	60
Extensive Margin				
Total annual hospital admission	0.100	0.300	0	1
Annual hospital admission by ambulance	0.035	0.184	0	1
Annual Emergency hospital admission	0.073	0.260	0	1
Senior diabetic patients (n=328,405)				
65,681 senior diabetic patients with 5 years of observations				
Intensive Margin				
Total annual hospital admission	0.188	0.560	0	5
Annual hospital admission by ambulance	0.072	0.334	0	5
Annual Emergency hospital admission	0.140	0.483	0	5
Annual length of stay in hospital for acute treatment	1.238	4.962	0	60
Extensive Margin				
Total annual hospital admission	0.133	0.133	0	1
Annual hospital admission by ambulance	0.055	0.229	0	1
Annual Emergency hospital admission	0.102	0.030	0	1
Senior female diabetic patients (n=163,640)				
32,728 senior female diabetic patients with 5 years of observations				
Intensive Margin				
Total annual hospital admission	0.178	0.537	0	5
Annual hospital admission by ambulance	0.073	0.334	0	5
Annual Emergency hospital admission	0.137	0.472	0	5
Annual length of stay in hospital for acute treatment	1.213	4.912	0	60
Extensive Margin				
Total annual hospital admission	0.129	0.335	0	1
Annual hospital admission by ambulance	0.057	0.231	0	1
Annual Emergency hospital admission	0.100	0.301	0	1

Note: On the intensive margin, the unit of "total annual hospital admission", "annual hospital admission by ambulance", and "annual emergency admission by ambulance" is number of times in a year. The unit of "annual length of stay in hospital for acute treatment" is number of days in a year, not number of days per admission in a year.

Table 3.4: FHO Impact on Diabetic Patients' Hospitalizations

	Intensive Margin			Extensive Margin			
	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Length of Stay in Hospital
Specification 1							
yr_0607	-0.038***	-0.028***	-0.039***	-0.019***	-0.019***	-0.021***	-0.340***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.017)
yr_0708	-0.037***	-0.025***	-0.036***	-0.019***	-0.016***	-0.020***	-0.315***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.017)
yr_0809	-0.028***	-0.019***	-0.027***	-0.015***	-0.012***	-0.015***	-0.226***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.016)
yr_0910	-0.018***	-0.013***	-0.017***	-0.009***	-0.008***	-0.009***	-0.156***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.014)
β_{FHO}	0.004*	0.004***	0.004*	0.004**	0.003***	0.003**	0.056***
	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.021)
Specification 2							
β_{FHO_y}	0.010	0.004	0.010	0.000	0.003	0.002	0.040
	(0.011)	(0.005)	(0.010)	(0.007)	(0.004)	(0.006)	(0.062)
β_{FHO_m}	-0.001	0.001	0.000	0.000	0.000	0.000	0.020
	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.023)
β_{FHO_s}	0.011**	0.008***	0.010**	0.008***	0.006***	0.008***	0.103***
	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.038)
N	797355	797355	797355	797355	797355	797355	797355

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficients are defined in equation (1) and (2).

Table 3.5: Impact on Male vs. Female Diabetic Patients' Hospitalizations

	Intensive Margin			Extensive Margin			
	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Length of Stay in Hospital
Male							
β_{FHO_y}	0.005	0.004	0.002	0.000	0.005	-0.002	0.062
	(0.014)	(0.007)	(0.013)	(0.009)	(0.005)	(0.008)	(0.078)
β_{FHO_m}	-0.003	0.000	-0.002	-0.001	-0.001	-0.001	0.004
	(0.004)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.031)
β_{FHO_s}	0.003	0.003	0.005	0.003	0.003	0.006	0.023
	(0.007)	(0.004)	(0.006)	(0.004)	(0.003)	(0.003)	(0.057)
N	420235	420235	420235	420235	420235	420235	420235
Female							
β_{FHO_y}	0.012	0.004	0.017	0.001	0.001	0.005	0.019
	(0.018)	(0.007)	(0.015)	(0.011)	(0.005)	(0.008)	(0.093)
β_{FHO_m}	0.001	0.001	0.001	0.002	0.000	0.002	0.040
	(0.004)	(0.002)	(0.004)	(0.003)	(0.001)	(0.002)	(0.034)
β_{FHO_s}	0.019***	0.013***	0.014***	0.013***	0.010***	0.010***	0.182***
	(0.006)	(0.004)	(0.005)	(0.004)	(0.003)	(0.003)	(0.054)
N	377120	377120	377120	377120	377120	377120	377120

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficients are defined in equation (2).

Table 3.6: Impact on Urban vs. Rural Diabetic Patients' Hospitalizations

	Intensive Margin			Extensive Margin			
	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Length of Stay in Hospital
$\text{RIO} = 0$							
β_{FHO_y}	0.01	0.005	0.013	0.003	0.006	0.005	0.096
	(0.018)	(0.008)	(0.016)	(0.011)	(0.005)	(0.010)	(0.103)
β_{FHO_m}	-0.003	0.000	-0.002	-0.002	-0.002	-0.002	0.005
	(0.004)	(0.002)	(0.004)	(0.003)	(0.001)	(0.002)	(0.037)
β_{FHO_s}	0.004	0.004	0.007	0.005	0.004	0.008**	0.061
	(0.006)	(0.004)	(0.005)	(0.004)	(0.003)	(0.003)	(0.055)
N	363585	363585	363585	363585	363585	363585	363585
$0 < RIO \leq 20$							
β_{FHO_y}	0.021	0.002	0.011	0.008	-0.000	0.002	0.030
	(0.017)	(0.007)	(0.014)	(0.011)	(0.006)	(0.009)	(0.093)
β_{FHO_m}	0.004	0.005**	0.004	0.004	0.003*	0.003	0.052
	(0.005)	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)	(0.036)
β_{FHO_s}	0.015**	0.012***	0.011**	0.005	0.007**	0.005	0.102
	(0.007)	(0.004)	(0.006)	(0.004)	(0.003)	(0.004)	(0.062)
N	320170	320170	320170	320170	320170	320170	320170
$20 < RIO$							
β_{FHO_y}	-0.020	0.009	0.001	-0.019	0.003	-0.003	-0.037
	(0.031)	(0.011)	(0.028)	(0.017)	(0.008)	(0.015)	(0.152)
β_{FHO_m}	-0.010	-0.009**	-0.006	-0.003	-0.005*	-0.001	-0.048
	(0.008)	(0.004)	(0.007)	(0.004)	(0.003)	(0.004)	(0.061)
β_{FHO_s}	0.017	0.007	0.010	0.021***	0.010**	0.010**	0.170*
	(0.013)	(0.007)	(0.011)	(0.007)	(0.004)	(0.004)	(0.101)
N	113600	113600	113600	113600	113600	113600	113600

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficients are defined in equation (2).

Table 3.7: Impact on Diabetic Patients' Hospitalizations by GPs' Characteristics

	Intensive Margin			Extensive Margin			
	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Length of Stay in Hospital
Male Doctor (633,045 patients with 1,951 GPs)							
β_{FHO}	0.011**	0.007**	0.009**	0.008***	0.006***	0.007***	0.109**
	(0.005)	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)	(0.043)
Female Doctor (164,310 patients with 1,048 GPs)							
β_{FHO}	0.015	0.014**	0.013	0.009	0.007	0.009*	0.087
	(0.01)	(0.006)	(0.009)	(0.006)	(0.004)	(0.005)	(0.084)
Doctor with age_march2006<40 (88,485 with 490 GPs)							
β_{FHO}	0.012	0.010	0.013	0.007	0.006	0.008	-0.078
	(0.013)	(0.008)	(0.010)	(0.008)	(0.005)	(0.007)	(0.110)
Doctor with 40≤age_march2006<60 (585,785 with 2,080 GPs)							
β_{FHO}	0.011**	0.009***	0.011**	0.007**	0.008***	0.008***	0.142***
	(0.005)	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)	(0.044)
Doctor with age_march2006≥60 (123,085 with 429 GPs)							
β_{FHO}	0.014	0.002	0.004	0.015**	0.003	0.007	0.062
	(0.012)	(0.007)	(0.010)	(0.007)	(0.005)	(0.006)	(0.111)

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficients are defined in equation (2).

Table 3.8: Sensitivity Tests on the Dynamics of the Impact

	Intensive Margin			Extensive Margin			
	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Length of Stay in Hospital
Specification 3							
All							
β_{FHO_0}	0.004	0.004**	0.005*	0.003*	0.003**	0.003*	0.056**
	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.025)
β_{FHO_1}	0.000	-0.001	-0.001	0.001	-0.001	0.001	0.001
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.024)
Senior female patients							
β_{FHO_0}	0.014***	0.010***	0.008***	0.012***	0.009***	0.008***	0.117**
	(0.005)	(0.003)	(0.003)	(0.005)	(0.003)	(0.002)	(0.046)
β_{FHO_1}	-0.005	-0.004	-0.001	-0.004	-0.002	-0.003	-0.022
	(0.005)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)	(0.046)
Specification 4							
Add switch date in 2007 (172 GPs switched)							
All							
β_{FHO_0}	0.005*	0.004**	0.005**	0.004**	0.003**	0.004**	0.058**
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.022)
$\beta_{FHO_{2007}}$	-0.003	0.000	-0.003	-0.003	0.000	-0.002	-0.013
	(0.007)	(0.003)	(0.006)	(0.004)	(0.002)	(0.003)	(0.053)
Senior female patients							
β_{FHO_0}	0.012**	0.008***	0.010**	0.008***	0.007***	0.008***	0.102**
	(0.005)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.041)
$\beta_{FHO_{2007}}$	-0.002	-0.005	-0.002	-0.003	-0.001	-0.003	0.009
	(0.011)	(0.007)	(0.010)	(0.007)	(0.004)	(0.006)	(0.090)
Add switch date in 2008 (552 GPs switched)							
All							
β_{FHO_0}	0.004	0.003	0.004	0.002	0.002	0.002	0.046
	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.028)
$\beta_{FHO_{2008}}$	0.000	0.001	0.001	0.002	0.001	0.002	0.018
	(0.004)	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)	(0.035)
Senior female patients							
β_{FHO_0}	0.014**	0.008**	0.013**	0.007*	0.007***	0.008**	0.149***
	(0.006)	(0.004)	(0.005)	(0.004)	(0.003)	(0.003)	(0.053)
$\beta_{FHO_{2008}}$	-0.004	-0.001	-0.005	0.002	-0.001	-0.001	-0.079
	(0.007)	(0.004)	(0.006)	(0.004)	(0.003)	(0.004)	(0.064)
Add switch date in 2009 (331 GPs switched)							
All							
β_{FHO_0}	0.004	0.004**	0.004*	0.004**	0.003**	0.003**	0.061***

	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.024)
$\beta_{FHO_{2009}}$	0.001	0.000	0.001	-0.002	0.000	0.000	-0.017
	(0.005)	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)	(0.039)
Senior female patients							
β_{FHO_0}	0.011**	0.007***	0.009**	0.008***	0.006***	0.007***	0.093**
	(0.005)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.042)
$\beta_{FHO_{2009}}$	0.000	0.002	0.003	-0.002	0.003	0.002	0.039
	(0.008)	(0.005)	(0.007)	(0.005)	(0.004)	(0.005)	(0.073)
Add switch date in 2010 (299 GPs switched)							
All							
β_{FHO_0}	0.004*	0.004***	0.004**	0.004**	0.003***	0.004***	0.056***
	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.021)
$\beta_{FHO_{2010}}$	0.002	-0.004	-0.002	0.000	-0.005	-0.005	0.003
	(0.008)	(0.005)	(0.008)	(0.005)	(0.003)	(0.004)	(0.074)
Senior female patients							
β_{FHO_0}	0.010**	0.008***	0.009**	0.008***	0.007***	0.008***	0.094**
	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.039)
$\beta_{FHO_{2010}}$	0.022	0.004	0.016	0.005	-0.003	0.002	0.197
	(0.016)	(0.010)	(0.014)	(0.010)	(0.007)	(0.009)	(0.156)

Note: * p<0.10, ** p<0.05, *** p<0.01. Coefficients are defined in equation (3) and (4).

Table 3.9: Falsification Tests

	Intensive Margin			Extensive Margin			
	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Hospital Admission	Hospital Admission by Ambulance	Emergency Hospital Admission	Length of Stay in Hospital
Falsification 1 (random switch date)							
δ_0	0.002	0.001	0.004	0.000	0.001	0.001	0.015
	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.029)
Senior female patients							
δ_0	0.000	0.002	0.003	-0.004	-0.001	-0.001	0.037
	(0.006)	(0.004)	(0.005)	(0.003)	(0.002)	(0.003)	(0.054)
Falsification 2 (April 1, 2009)							
δ_0	0.000	0.001	0.002	-0.002	0.000	0.000	0.030
	(0.004)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.030)
Senior female patients							
δ_0	0.003	0.005	0.005	-0.001	0.002	0.001	0.027
	(0.006)	(0.003)	(0.005)	(0.003)	(0.002)	(0.003)	(0.053)
Falsification 2 (Mid-point of the sampling period)							
δ_0	0.003	0.002	0.003	0.001	0.001	0.001	0.041
	(0.004)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.033)
Senior female patients							
δ_0	-0.003	0.002	0.002	-0.004	0.001	0.000	0.011
	(0.006)	(0.004)	(0.005)	(0.004)	(0.002)	(0.003)	(0.056)
The δ_0 estimates in equation (2) for senior patients							
δ_0	0.011**	0.008***	0.010**	0.008***	0.006***	0.008***	0.103***
	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.038)

Note: * p<0.10, ** p<0.05, *** p<0.01. δ_0 estimates in equations (2) for senior patients are added in the last panel as a comparison for the previous panels.

Conclusion

This thesis consists of three chapters that addresses policy-relevant questions with a focus on the immigrants' labor market and physician payment models. Immigration policy has been debated politically for decades. The first chapter empirically investigates the relationship between quality of immigrants' source country educational outcomes and labor market outcomes in a receiving country. Many surveys undertaken as part of health services and health economics research usually suffer from nonresponse and attrition, so parameter estimates derived from these survey data without proper weights would result in bias and imprecision. The second chapter explores a technique that combines complementary population and survey data using a GMM technique that combines information from the two data sources to estimate micro-econometric models. The effectiveness of physician payment schemes in improving patient health has been debated over the past two decades. The third chapter explores the relationship between a diabetic patient's hospitalization and his family doctor's payment scheme.

In the first chapter, the results show that differences in the source country average quality of pre-immigration educational outcomes have substantial impacts on the Canadian labour market earnings of immigrants. The observed impact flows through the return to education, with those from source countries with higher test scores having much higher returns to education, so that the gap widens as years of schooling increases. Adding country-level controls, especially source country GDP per capita, does not appreciably alter the relationship so it is not a wealthy-country effect. Further, the return to education observed for an immigrants who arrived before age 10 is not a function of his/her source country quality of educational outcome. This reinforces the idea that it is the quality of educational outcomes, and not source country wealth effects *per se*, that is correlated with the return to education. Notably, the findings for the sample of all women differs somewhat from that for men, especially conditional on source country characteristics. However, in line with the literature on immigrant gender roles, when the sample is restricted to women who are unmarried or without children living in the household, the results are quite similar to those for men.

The second chapter is an exploration on a technique developed by Hellerstein and Imbens (1999). The results from Monte Carlo simulations show that the efficiency gain from using this approach can be substantial. It reduces the bias of the estimates in an endogenously stratified sample and the variance of the estimates in a simple random sample or an exogenously stratified sample. The results also show that the advantages of this technique are maintained in the face of variation in the sample sizes, explanatory power of independent variables, and the amount of heteroskedasticity in error terms. Different from an estimator identification using two individual datasets, this approach only requires marginal information at aggregate level from a population. Particularly, it estimates micro-econometric models combining survey data with marginal information from a complementary population using a GMM technique that matches the moments of the two data sources and recovers estimated parameters of population. A practical concern is the amount of overlap in the values of the variables between the population and sample. We show an example looking at midwife training and another one looking at an optometrist's location of work, to illustrate its use in a health human resource context.

In the third chapter, we identify Ontario patients with diabetes and empirically investigate their hospitalizations as a function of their family doctors' remuneration models. Using comprehensive administrative datasets maintained by the Ontario Ministry of Health and Long-

Term Care, we construct a longitudinal data for diabetic patients' inpatient hospitalization admissions as well as their GPs' payment models. We employ a difference-in-differences approach to control of both selection on observables and selection on unobserved time-invariant fixed effects to avoid the estimation bias in the identification. The treatment group is defined as those diabetic patients whose GPs switched from FHG to FHO and the comparison group is defined as those whose GPs stayed in FHG over the sampling period. The "prior to" and "after" periods depend on the intervention time that is the physicians' switch dates in this case. The results show that on both intensive and extensive margin, the hospital admissions for senior female patients economically and statistically significantly increased after their GP's remuneration model changed from FHG to FHO. In contrast, the impact on male patients and younger female patients were small and not statistically significant. The results provide a cautionary message regarding to the differences in practice patterns towards senior diabetic patients between GPs in the two models.