THREE ESSAYS IN EMPIRICAL LABOUR ECONOMICS

Miroslav Kučera

A Thesis

In the Department

of

Economics

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

Concordia University

Montréal, Québec, Canada

November 7, 2010

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Miroslav Kucera

Entitled: Three Essays in Empirical Labour Economics

and submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (Economics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. S. Shaw	Chair
Dr. C. Belzil	External Examiner
Dr. S. Kairouz	External to Program
Dr. N. Gospodinov	Examiner
Dr. A. Prokhorov	Examiner
Dr. J. Hansen	Thesis Supervisor

Approved by

Dr. N. Gospodinov, Graduate Program Director

October 26, 2010

Dr. B. Lewis, Dean
Faculty of Arts and Science

ABSTRACT

Three Essays in Empirical Labour Economics

Miroslav Kucera, Ph.D.
Concordia University, 2011

The following thesis consists of three essays, each one being a study of issues of accumulation of and returns to human capital using real-world individual-level data. The first study examines what underlies differences in educational attainment between the children of immigrants to Canada and the children of the Canadian-born parents. It concludes that the children of immigrants have done better in terms of schooling, and that individual and family variables as well as unobserved characteristics such as ability cannot fully account for this difference.

The second study utilizes unique Canadian surveys to investigate the effects of overeducation on wages of post-secondary graduates. It confirms that jobs requiring a post-secondary degree pay substantially higher wages than jobs that do not require education beyond high-school, and also finds a large variation both in returns to required education as well as in overeducation premia across genders, degrees and fields of study.

The last essay proposes and estimates a structural dynamic model of optimal schooling and wages to explain differences between American whites and ethnic minorities of Afro-Americans and Hispanics. The study finds, among other things, that differences in educational attainment between the three ethnics can largely be explained by differences in individual endowments, while behavioural differences seem to be more important in explaining wage differences.

ACKNOWLEDGEMENTS

I would like to thank my supervisor and mentor Professor Jorgen Hansen for all his help and guidance throughout my studies, as well as to my family and friends whose support means a great deal to me. I would also like to thank the members of my PhD committee, Professors Sylvia Kairouz, Christian Belzil, Nikolay Gospodinov, Artem Prokhorov and Steven Shaw for their work and valuable comments and suggestions.

TABLE OF CONTENTS

INTRODUCTION

The accumulation of human capital, how it is rewarded in the labour market, and what explains variation in its stock and returns have been at the core of Labour Economics since the beginnings of the field. The work presented here addresses these issues using a variety of models and data sources. The three essays that constitute the thesis intersect as well as complement each other in a number of ways and dimensions. At the most general level, all three topics focus on education and labour market outcomes of individuals who operate in North America. Every one of these essays ultimately represents a scientific inquiry that builds on a set of relevant and up-to-date economic theories in order to develop empirically tractable models and estimation strategies to answer selected research questions within the constraints imposed by available data and computational resources. As a whole, the thesis is a series of investigations into both the individual and structural factors that describe and explain the dynamics of educational attainment and labour market outcomes for selected groups of the populations of Canada and the United States. The first essay focuses on second-generation immigrants in Canada; the analysis has been designed and carried out in a way that allowed for differences in education between males and females. The gender dimension remains important in the second study, in which the effects of factors that explain wage differentials among post-secondary graduates are clearly distinguished for men and women. Regardless of a country or region one chooses to investigate, specific groups, whether defined by gender or on the basis of a minority status in the general population

(specifically, the ethnic minorities of African-Americans and Hispanics, and the children of immigrants in this thesis) are invariably of interest not just to academic researchers but also to policy makers and the general public. The reason is obvious: women and minorities have often been identified as being disadvantaged – whether in terms of schooling, labour market outcomes or both – when contrasted with their respective 'mainstream' groups in the populations of interest. This aspect is further underscored in the third essay which studies the schooling and wage disparities between white males and males from two ethnic minorities – African-Americans and Hispanics. It is precisely this focus as well as the empirical nature of the study that make it an integral component of the thesis. The following paragraphs provide a more detailed overviews of the content, method and findings.

The first essay centers around the educational attainment of the children of immigrants to Canada. It uses cross-sectional data from the 2001 General Social Survey to estimate an ordered-choice model. Unlike a number of other studies, mostly of European provenience, the study concludes that the children of immigrants (second-generation immigrants) do better in terms of education than their contemporaries born to non-immigrant parents, and that the significant disparity in favour of second-generation immigrants remains even when differences in selected individual characteristics and unobserved abilities are controlled for.

The second essay uses unique data from two most recent cycles of the National Graduates Survey to explain variation in wages paid to graduates from Canadian post-secondary institutions, and to assess how wages are affected by the match between worker's educational attainment and education requirements of their jobs. Although the

results vary greatly across genders, degrees and fields of study, the essay concludes that workers with post-secondary degrees receive substantially higher wages relative to those who work in jobs that do not require education beyond high school. It also finds, with some exceptions and to a varying extent, that overeducated workers receive a wage premium over those who work the same jobs and are perfectly matched with these jobs in terms of education.

The third essay brings together the process and outcomes of both education and labour market, and focuses on explaining differences in educational attainment and wages between whites and ethnic minorities of blacks and Hispanics in the United States using the 1997 National Longitudinal Survey of Youth. It proposes and estimates a structural dynamic model in which individuals decide upon optimal years of schooling by maximizing their lifetime utility, and then uses the estimated model to analyze the sources of disparities in schooling and wages across the three ethnics. The analysis reveals, among other things, substantial differences in how the market rewards education and experience. Furthermore, the decompositions of the observed differentials between whites and minorities show that the differences in schooling attainment can largely be explained by differences in endowments, while behavioural differences play a more important role in explaining the ethnic wage gaps.

As mentioned above, the thesis consists of three essays that all follow a similar structure. Each essay starts with a brief overview (abstract), and then continues by introducing specific research questions, summarizing existing literature, conducting the empirical analysis, and ends by drawing conclusions from the findings. In order to ensure

consistency in the presentation, the tables and figures for each essay are placed in their respective appendices rather than mixed within the text.

ESSAY I

DIFFERENCES IN EDUCATIONAL ATTAINMENT BETWEEN CHILDREN OF IMMIGRANTS TO CANADA AND CHILDREN OF CANADIAN-BORN PARENTS

ABSTRACT

Using data from the 2001 General Social Survey, this study focused on differences in educational attainment between the children of immigrants to Canada and similarly-aged children of Canadian-born parents. Two definitions of second-generation immigrants were introduced. The first considered a Canadian resident with at least one immigrant parent to be a second-generation immigrant, while the second definition required that both parents were foreign-born. All first-generation immigrants were excluded from the sample, with the exception of those who had arrived in Canada at the age of 9 or younger; these young immigrants were then included among the second-generation immigrants. The results show that second-generation immigrants did better in terms of schooling attainment than their peers born to Canadian parents, and that a significant disparity in favour of second-generation immigrants remained even after controlling for differences in selected personal characteristics, family background and unobserved heterogeneity.

1 INTRODUCTION

The resurgence of immigration in many western countries has initiated an intensive debate over its effects. A major portion of current research has focused mainly on how first-generation immigrants integrate into the economic and social structure of the host country, despite the fact that the overall, long-term impact of immigration also depends on the adjustment process experienced by their children, commonly referred to as second-generation immigrants. In the context of current demographic trends and the role immigration plays in a number of developed countries, it is important to determine how both first and second-generation immigrants integrate and perform in host countries. Surprisingly, the research on the integration of the children of immigrants has been rather limited and many issues that are consequential to the immigration debate remain unaddressed.

One of the issues that has so far received only limited attention is the schooling attainment of second-generation immigrants. Considering that education is a strong determinant not only of subsequent labour market experiences but of successful social and economic integration in general, it is essential to study how children of immigrants differ in their educational attainments from the children of Canadian-born parents in order to evaluate past and current immigration policies. Although the existing literature is still rather sparse, a few recent studies looked at how children of immigrants fare in terms of education and labour market outcomes in comparison with their parents as well as with similarly-aged children of domestic-born parents.

For example, Borjas (1992, 1994), while focusing on intergenerational transfer of ethnic capital, summarized a number of observations about second-generation immigrants in the United States. He found substantial improvements across generations but also identified a large dispersion in educational attainment, as well as wages and occupational prestige scores, across different ethnic groups.

An explicit comparison between second-generation immigrants and similarly-aged children of German-born parents was the focus of Gang and Zimmermann (2000) who used data from the German Socio-Economic Panel to investigate the effects of parental education on a child's schooling, and to identify whether there were differences in educational attainment between second-generation immigrants, divided into five major ethnic groups, and native Germans of the same age. They found that the educational level of first-generation immigrants had no effect on the educational attainment of their children, while in the case of native-born Germans, parental education had an effect on the schooling of the progeny. They also detected a convergence in education between the children of immigrants and those born to Germans, but concluded that ethnic differences persisted within one's educational cohort even after controlling for parental human capital and other characteristics.

Another study by Riphahn (2003) analyzed the educational attainment of German-born children of immigrants using German Census data. She found that the educational outcomes of second-generation immigrants were significantly below that of natives and that even after controlling for various characteristics the overall educational gap between the children of immigrants and native Germans not only remained significant but actually widened over time.

Van Ours and Veenman (2003), using Dutch data, compared second-generation immigrants, divided into four major ethnic groups, both with first-generation immigrants and with natives of the same age group. They found that the differences in educational attainment which appeared in the data were largely driven by the differences in parental education rather than by ethnicity. In other words, the children of immigrants were worse off in terms of schooling because their parents had, on average, lower education than the parents of the natives. Van Ours and Veenman concluded that if these differences were taken into account, the gap between the native Dutch people and the second generation immigrants would to a large extent vanish.

More recently, Dustmann and Theodoropoulos (2006) used 1979 to 2005 data from the British Labour Force to investigate educational attainment and economic behaviour of ethnic minority immigrants and their children in Britain. They found that the second-generation ethnic minorities were on average more educated than their parents as well as than a comparable group of white natives. They, however, appeared to have lower employment probabilities than their British-born white peers. Dustmann and Theodoropoulos also report significant differences across immigrant/ethnic groups and genders.

Evidence on the performance of second-generation immigrants in comparison with similarly aged offsprings of non-immigrant parents differs substantially across countries, with U.S. and Canadian studies mostly presenting a more optimistic picture than European ones. For example, using the U.S. Census data, Card et al. (2000) found that children of immigrants had higher education and wages than children of non-immigrants even after controlling for parental background. Similarly for Canada,

Aydemir, Chen and Corak (2006) reported that second-generation immigrants had the educational attainments and labour market outcomes no worse, and in many ways better than those born to Canadian parents.

Aydemir and Sweetman (2006) examined differences in the characteristics and outcomes of first, second and third-generation immigrants to Canada and the United States. Using a sample from the 2001 Canadian Census and the data from 1998-2004 U.S. CPS, they found that immigrants to Canada had, on average, more years of schooling than the third generation while for the United States the opposite was true. Second-generation immigrants in both countries appeared to have accumulated more years of schooling than the third generation although the difference was reversed in Canada when ethnicity and geography were controlled for. While Aydemir and Sweetman considered educational differences across the three different immigrant groups (first-, second- and third-generation immigrants), a major limitation of their paper was their use of the census data. The Census does not contain information on various family and socio-economic characteristics (education of the parents, number siblings, etc.) that are known to significantly influence individuals' educational outcomes.

Worswick (2004), focused on differences in school outcomes using the data from three cycles of the National Longitudinal Survey of Children and Youth in Canada. He found that the children of immigrants had test scores in reading and mathematics comparable to the scores of the children of Canadian-born parents. He also found that children with a foreign mother tongue had low performance in vocabulary before age six but their performance in mathematics and reading was on a par with that of the children

of the Canadian-born at the age of 14, which may indicate a convergence in school outcomes as children moved through the Canadian educational system.

It has been recognized that acquiring a level of education equivalent to that of domestic population is one of the key elements that determines how immigrants and their children integrate into the economic and social structures of the host country. This paper addresses the issue by focusing on the schooling attainment of second-generation immigrants in Canada. Using detailed individual data from the 2001 General Social Survey (GSS), a sample of males and females between 16 and 65 years of age was analyzed to provide a comparison of the schooling attainment of children of immigrants to Canada with similarly-aged children of Canadian-born parents.

In the analysis, two definitions of what constitutes a second-generation immigrant were employed. The first definition considered a Canadian resident to be a second-generation immigrant if at least one of his/her parents was a foreign-born immigrant. This appears to be the definition most commonly used in the previous literature. The second definition was stricter as it required that both parents were foreign-born for an individual to be defined as a second-generation immigrant. The use of two alternative definitions, as well as accounting for other characteristics (mother tongue, parental education, family environment, etc.) and unobserved heterogeneity allowed to assess the importance of individual differences on schooling attainments.

For the purpose of this study, four levels of education were distinguished and a set of ordered-choice models was estimated separately for males and females. In contrast with most European studies, but in concordance with studies based on U.S. and Canadian data, the results suggest that the children of immigrants did better in terms of educational

attainment than their native Canadian counterparts even when the effects of selected individual characteristics were controlled for. Furthermore, the models estimated in this paper also allow for the possibility that there are unobserved individual differences that may be important in explaining the observed variation in educational attainments. To my knowledge, none of the existing studies of second-generation immigrants account for the effects of unobservables, despite the fact that such effect are unlikely to be negligible.

The paper is organized in the following way: Section 2 describes the data and the variables used in the analysis. Section 3 specifies the econometric model and provides a brief introduction to the methodology. Section 4 presents and discusses the results, and Section 5 concludes the paper.

2 THE DATA

The sample used in this study was extracted from the General Social Survey (GSS), a cross-sectional survey established in 1985 with main objectives being the collection of data on social trends, changes in living conditions and well-being of Canadians, and the supply of information on specific social policy issues or emerging interests. The GSS collects data over a 12-month period from the total population of 15 years of age and older, living in private households in the ten provinces of Canada. The survey excludes individuals living on the reserves and in the Territories, full-time members of the armed forces and institutionalized persons.

For this study, I used the 2001 GSS (cycle 15) public use microdata files. This particular cycle of GSS focused on issues of family history, and collected information

from approximately 25,000 respondents during the period from February to December 2001. After imposing age restrictions, excluding all first-generation immigrants except those who arrived in Canada at a young age (9 years old or younger) and after removing missing entries, the actual sample consisted of 12,018 individuals. This sample represents the population of over 12 million Canadian residents (51% males and 49% females), who were between 16 and 65 years of age in 2001.

2.1 Definitions of the Variables

Prior to any analysis, it is necessary to define what constitutes a second-generation immigrant. The definitions of a second-generation immigrant employed in existing literature vary substantially mainly due to legal and other differences across countries whose data were subjected to analyses. In this study, two definitions of second-generation immigrants have been used:

- Definition-1 second-generation immigrant is an individual born in Canada whose at least one parent was a foreign-born immigrant.
- Definition-2 second-generation immigrant is a Canadian-born individual whose both parents were foreign-born immigrants.

The rationale behind these definitions is that individuals raised by two immigrant parents may be substantially different in a number of ways from the domestic population. This, however, may not be the case if an individual has only one immigrant parent as this parent's influence may be diluted or even eliminated by the influence of the other, non-immigrant parent. The use of the definition 2 effectively excludes such mitigating effect. Furthermore, young immigrants – the first-generation immigrants who arrived in Canada

at the age of 9 or younger – were also included among the second-generation immigrants. Although there is no clear consensus over the cut-off age, it is often set to the age from which children start their elementary education. This reflects the hypothesis that if immigrant children enter the host country's schooling system early in life, differences and disadvantages they may have with respect to non-immigrants can be eliminated as they progress through the system.[1]

This study focuses on the highest level of education – a variable that takes on one of four mutually exclusive and ordered categories: less than high school, high school graduate, some post-secondary education (below university), and a university degree. Furthermore, as the sample also contained very young individuals, an indicator for censored observations had to be used to identify those who were still in school at the time of the survey, and for whom the final schooling attainment had not been observed.

Having a foreign mother tongue is often considered to have a negative effect on a child's schooling outcomes. In order to study the effect of mother tongue, an indicator was specified to distinguish individuals with a foreign mother tongue from those whose first language was English or French.

Besides the second-generation immigrant status and foreign mother tongue indicator, other variables that are commonly thought to influence an individual's schooling attainment were also included, namely variables for parental education and family size and structure (number of siblings, and an indicator for a complete two-parent

[1] Given the age-at-immigration categories in GSS, the choice of the cut-off was either 4 or 9 years of age. Neither the choice of the cut-off, nor complete exclusion of young immigrants from the sample changed the estimation results.

family). Two more variables, mother worked full-time and father worked full-time during respondent's childhood, were added to account for mother's and father's labour force status and work intensity, and to approximate household constraints in terms of time that the parents could devote to child-rearing as well as family income (with the father's work income presumably the largest fraction of it). Other individual differences, such as being raised in an urban area, the region of birth, and respondent's age were also controlled for.[2] No control for ethnic background could be included, as there was no variable in the 2001 GSS public use files that could be used to construct ethnicity indicators. It was only possible to determine whether the sampled individuals and their parents were born in Canada, Europe or some other unspecified place. Nevertheless, the information on the ethnic background was not central to the analysis. Given the dynamics of immigration to Canada, most second-generation immigrants in the 2001 GSS sample were still descendants of European immigrants. Although differences across individuals with ancestry from different parts of Europe may well have existed, their ethnic background was still more homogenous than not, and it did not matter in explaining differences in educational outcomes either.[3]

---

[2] Five birth regions were defined: Atlantic provinces, Western provinces, Ontario, Quebec and outside Canada, the last one to control for the birthplace of the young immigrants in the sample.

[3] In the 2001 GSS sample, over 70% of immigrant parents were from Europe. Differentiating between the second-generation immigrants with European background and those with non-European ancestry did not produce a significant impact on schooling attainment. Consequently, the corresponding indicator was dropped from the regression.

2.2 Sample Composition

As can be seen in table 1, the distributions of selected characteristics were quite similar for both men and women. The definition-1 and definition-2 second-generation immigrants made up over 23% and 13% of the two subpopulations, respectively. There were also over 4% of immigrants who arrived in Canada at the age of 9 or younger. More than 13% of women and almost 12% of men were still in school at the time of the survey. About 6.5% of individuals only spoke a language other than English or French in their childhood years, a majority of all respondents, more than 87%, were raised in a complete, two-parent family in which most fathers (around 97%) and almost 30% of mothers worked full-time. More than two thirds of the individuals grew up in an urban area.

In the 2001 GSS sample, how did the distributions of major characteristics differ between the second-generation immigrants and the children of the Canadian-born parents? First we look at the differences in schooling attainment of the sampled males as well as that of their parents as presented in tables 2 to 4 . Both definition-1 and definition-2 second-generation immigrant males appeared to be less represented in the two lowest schooling levels than the sons of the Canadian-born, but more concentrated in the highest category (university graduates). Their immigrant fathers and, to a lesser extent, also their mothers were less represented in the lowest schooling level (below high school), and more among the university graduates when compared to the Canadian-born parents.

Very similar differences in the parental and individual's own education could also be found between the daughters of immigrants and the daughters of the Canadian-born (tables 2 to 4). Furthermore, these differences, across both sexes as well as across the two

definitions of second-generation immigrants, were statistically significant. It appears that the children with immigrant background did somewhat better in terms of schooling attainment than the children of the Canadian-born, and had parents who were generally more educated than their domestic counterparts.

Finally, going back to table 1, we can see how selected observable characteristics were allocated across the two second-generation immigrant definitions. In all sub-samples, the children of immigrants and the children of Canadian-born parents were similarly represented among those who grew-up in complete families. The mothers of the second-generation immigrants were more likely to work full-time during their child's childhood than the Canadian-born mothers, while there was virtually no difference between the fathers of second-generation immigrants and the others as the majority of them worked full-time. Second-generation immigrants were somewhat more represented among the individuals who were still in school when the survey was conducted. Finally, a substantial fraction of the second-generation immigrants spoke foreign mother tongue during their childhood (about 24% of the definition-1 and over 38% of the definition-2 second-generation immigrants).

3 MODEL AND METHODOLOGY

In their 1998 paper in the Journal of Political Economy, Cameron and Heckman contrast the ordered-choice approach with the more traditional logistic model of grade transitions. They show that unlike the grade-transition model which implicitly assumes myopia on the part of agents, a simple ordered-choice model has is consistent with rational-agent

behaviour (agents observe their endowments at birth and choose the level of schooling that maximizes net returns to schooling). Following their reasoning, this section proposes an ordered-choice model that accounts for right-censored observations and unobserved individual heterogeneity. An individual's preference for education, y^*, is assumed to be determined by a number of personal and family characteristics, $\mathbf{x}$, an a random term, ε, in a linear fashion:

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon .$$

There are four (observed) levels of education, y, in the model: less than high school (0), high school (1), beyond high school but less than university (2), and a university degree (3). For those who had completed their schooling by the time of the survey, their 'taste' for education links to the attainment as follows:

$$y = \begin{cases} 0 & \text{if} \quad y^* \le \mu_1 \\ 1 & \text{if} \quad \mu_1 < y^* \le \mu_2 \\ 2 & \text{if} \quad \mu_2 < y^* \le \mu_3 \\ 3 & \text{if} \quad y^* > \mu_3 \end{cases}$$

For individuals who were still in school at the time of the survey and whose schooling spell had not yet been completed, there was no reason to presume that the level of education they had completed was also their final one. For these right-censored cases, we only know that their desired educational attainment y^* must have exceeded the cut-off point that defined the level of education they had completed by the year of the survey, that is

$$y = \begin{cases} 0 & \text{if} \quad y^* > -\infty \\ 1 & \text{if} \quad y^* > \mu_1 \\ 2 & \text{if} \quad y^* > \mu_2 \\ 3 & \text{if} \quad y^* > \mu_3 \end{cases}$$

Constructing the likelihood from the rules above is straightforward. Let c be an indicator such that $c = 1$ that if an individual was still in school at the time of the survey (censored observation) and 0 if otherwise, and let $F(\cdot)$ denote the cdf of ε. Then for the uncensored individuals, the attainment probabilities are

$$\begin{aligned} \Pr(y = 0 \mid \mathbf{x}, c = 0) &= F(\mu_1 - \mathbf{x}'\boldsymbol{\beta}) \\ \Pr(y = 1 \mid \mathbf{x}, c = 0) &= F(\mu_2 - \mathbf{x}'\boldsymbol{\beta}) - F(\mu_1 - \mathbf{x}'\boldsymbol{\beta}) \\ \Pr(y = 2 \mid \mathbf{x}, c = 0) &= F(\mu_3 - \mathbf{x}'\boldsymbol{\beta}) - F(\mu_2 - \mathbf{x}'\boldsymbol{\beta}) \\ \Pr(y = 3 \mid \mathbf{x}, c = 0) &= 1 - F(\mu_3 - \mathbf{x}'\boldsymbol{\beta}) \end{aligned}$$

and for the right-censored observations

$$\begin{aligned} \Pr(y = 0 \mid \mathbf{x}, c = 1) &= 1 \\ \Pr(y = 1 \mid \mathbf{x}, c = 1) &= 1 - F(\mu_1 - \mathbf{x}'\boldsymbol{\beta}) \\ \Pr(y = 2 \mid \mathbf{x}, c = 1) &= 1 - F(\mu_2 - \mathbf{x}'\boldsymbol{\beta}) \\ \Pr(y = 3 \mid \mathbf{x}, c = 1) &= 1 - F(\mu_3 - \mathbf{x}'\boldsymbol{\beta}) \end{aligned}$$

The model also takes into account unobserved heterogeneity through the inclusion of latent classes. Specifically, it is assumed that the distribution of the error term ε is a three-point mixture of standard normals (low, medium and high ability types).[4] The probability of belonging to type k is parameterized as a logistic transform

[4] On the basis of the Bayesian Information Criterion, the model was best fitted with three types of individuals.

$$p_k = \frac{\exp(q_k)}{\sum_{m=1}^{2}\exp(q_m)+1},\ k = 1,2,3,$$

where q_3 is set to zero for identification purposes. The contribution of an individual of type k to the likelihood function is

$$L_k = \prod_{j=0}^{3}\left[\Pr(y = j \mid \mathbf{x}, c = 0, \text{type} = k)\right]^{d_j(1-c)} \cdot \left[\Pr(y = j \mid \mathbf{x}, c = 1, \text{type} = k)\right]^{d_j c} \qquad (1)$$

where $d_j = 1$ if an individual had completed the *j*-th level of education and 0 if otherwise, and the unconditional (with respect to type) individual likelihood is

$$L = p_1 L_1 + p_2 L_2 + p_3 L_3 . \qquad (2)$$

The parameters of the model are obtained by maximizing the sum of individual likelihoods.

An important question regarding this model is whether it is identified. Cameron and Heckman (1997) showed that the structural parameters and the distribution of unobserved heterogeneity of the very same ordered-choice model as used in this study are identified even without invoking specific assumptions about the distribution of ε. Specifically, the μ's can be recovered up to an affine transformation, and the and β's can be recovered up to a scale transformation as in a standard binary choice model. Typically, standard ordered-choice models recover slopes and intercepts by assuming that $\mathrm{E}(\varepsilon) = 0$ (or $\text{Median}(\varepsilon) = 0$) to tie down the location of the distribution of the unobservables, and by normalizing its variance to a constant value to tie down the scale. This study follows Cameron and Heckman, and utilizes these assumptions as a matter of

computational convenience. Lastly, as the usefulness of parameter estimates in ordered-choice models is only limited, marginal effects (probability differences for the dichotomous ones) have to be calculated in order to assess the actual effects of selected variables.

4 ESTIMATION RESULTS

When summarizing estimation results from an ordered-choice model, the information value of the coefficient estimates is only limited. Generally, the signs of the coefficients can be directly related to changes in the conditional probabilities of the lowest and the highest schooling levels, but what happens to the probabilities of attaining the schooling levels in between is ambiguous. Thus, in order to fully assess the impact of the variables of interest on schooling, marginal effects have to be computed and analyzed. All analyses in this section begin with the presentation of the results from the ordered-probit regressions as specified in the methodology section. Each model was estimated both under the assumption of no unobserved heterogeneity, and then assuming a three-type finite mixture distribution. Such an approach allows to highlight biases in the parameter estimates when unobservable differences are unaccounted for. As for the choice of explanatory variables, it was based on existing theories as well as on empirical studies of educational attainment, and refined on statistical grounds. The chosen variables are typically employed to explain individual differences in schooling attainment. Using the regression estimates, the effects of all selected factors that affect schooling attainment are

discussed. Finally, detailed examinations of the marginal effects of second-generation immigrant status and foreign mother tongue conclude the analysis.

4.1 Parameter Estimates and Marginal Effects of Selected Characteristics: Sample of Males

The starting point for the discussion are the ordered-probit parameter estimates for the sub-sample of males as presented in table 5. The table shows estimated parameters and the associated z-statistics obtained under both definitions of second-generation immigrants, and for models with and without unobserved heterogeneity. First result to notice is that the effect of being a second-generation immigrant is positive and statistically significant across both definitions, and regardless whether unobservables are accounted for or not.[5] In table 7, we can see that this translates into second-generation immigrant males being more likely to attain higher levels of education (some post-secondary or a university degree), and less likely to end up in the lower ones (high-school or below) in comparison with sons of Canadian-born parents. The positive effect of being a second-generation immigrant on educational attainment contrasts with results reported in a number of studies, mainly of European provenience, in which children of immigrants appear to be disadvantaged in terms of schooling when compared to the children of non-immigrant populations. Even more surprising is perhaps the fact that the effect remained positive and statistically significant even when all relevant observable and unobservable

[5] Throughout this study, the level of significance is 5% unless stated otherwise.

characteristics of an individual were accounted for. In fact, including unobserved heterogeneity in the model actually resulted in an increase in the magnitude of the effect.

Another interesting result in table 5 is the positive coefficient corresponding to foreign mother tongue. The estimated marginal effects shown in table 8 confirm that individuals with foreign mother tongue had, on average, a higher educational attainment than those whose mother tongue was English or French. Furthermore, when unobserved individual heterogeneity was taken into account, the marginal effect of foreign mother tongue have actually increased. Without access to appropriate data, it is impossible to provide an explanation for such a result. Learning first a foreign language instead of the official language of the country in which one lives could conceivably cause a delay in the development of child's language skills, and be reflected in poorer schooling outcomes. On the other hand, speaking another language besides the official language of the country which a child would presumably "pick up" from friends, and through learning and interactions in a pre-school facility, could provide an additional stimulus, and enhance his/her cognitive abilities and capacity for learning. In Canada, Christopher Worswick (2004) provided some support for this hypothesis. Using data from the National Longitudinal Survey of Children and Youth, Worswick found that children of immigrants whose first language was neither English nor French were at a disadvantage in the early school years (had lower vocabulary test scores) compare to their English/French-speaking peers, but by the age of fourteen, their performance in reading and mathematics was at least as good as that of children of Canadian-born parents.

As for the effects of the remaining regressors in the fully-specified model, they all have expected signs, and are in line with current theories and empirical evidence.

Children from cities do better in terms of schooling attainment than children from rural areas. Family size (approximated by the number of siblings) correlates negatively with schooling attainment. The negative effect of the mother working full-time on child's schooling outcome can be explained as an effect of lower investment into the child's quality, as mothers with jobs cannot devote as much time to their children as mothers who stay at home. The positive effect of the father working full-time most likely serves as a proxy for higher household income compared to the families in which the father did not work. The higher income presumably translates into a greater investment into children which, in turn, would be reflected in a higher schooling attainment.

In order to analyze the effects of second-generation immigrant status and foreign mother tongue in a greater detail, four probability differences were estimated, each of them representing a different type of the marginal effect depending on the actual or otherwise specified values of the second-generation immigrant status and the foreign mother tongue indicator; the estimates for the sample of males are presented in tables 7 and 8. Let $s = 1$ if an individual is a second-generation immigrant, and 0 if otherwise, and let the foreign mother tongue variable, f also be defined as a binary indicator in a similar fashion, and $d = s, f$. Due to their binary nature, the marginal effects of these variables are calculated as the differences in probability of attaining the *j*-th level of schooling, $\Pr(y = j \mid \mathbf{x}, d = 1) - \Pr(y = j \mid \mathbf{x}, d = 0)$, averaged over the sample.

Continuing the analysis, the estimates in table 7 from the models that do not account unobserved heterogeneity show that being a second-generation immigrant male decreased the probabilities of attaining the two lower schooling levels, while increased the probabilities of having the two higher ones. The result is statistically significant, and

consistent across both definitions of second-generation immigrants. The general pattern is preserved even when unobserved heterogeneity is added into the regression model, although the actual estimates do change. The effect of being second-generation immigrant male now appears to lower the probability of having less than high school even more than it did previously. This is especially true for individuals from families with both parents immigrants (definition 2). For these males, the probability of education below high school was initially about 3.3 percentage points lower than for their counterparts from Canadian families; once unobservables were included, this difference more than doubled to -7.3 percentage points. Extending the model for unobserved heterogeneity somewhat lowered the probability of attaining the highest level of education (university), but greatly increased the probability of attaining the level right below (some post-secondary). The effect on the probability of having a high school degree become either insignificant (definition 1 males), or rather small in magnitude (definition 2).

The overall marginal effect of a foreign mother tongue in the sample of males (table 8) follows a pattern similar to the effect of the second-generation immigrant status. Individuals with a foreign mother tongue appear to have an advantage over those with French or English. They have lower probabilities of attaining high school or below, while they are more likely to obtain a post-secondary degree. Without accounting for unobservables, this effect is mostly statistically significant. However, when unobserved individual heterogeneity is included in the model, the marginal effect of foreign mother tongue on each schooling-level probability is not significantly different from zero.

4.2 Parameter Estimates and Marginal Effects of Selected Characteristics: Sample of Females

Parameter estimates for the sample of females in table 6 show a similar pattern as those for the sample of males (table 5). The coefficients corresponding to the second-generation immigrant status are positive and statistically significant across the two definitions of second-generation immigrants, and regardless whether unobserved heterogeneity is accounted for or not. One larger difference appears to be in the foreign-mother tongue effects which are of smaller magnitudes than the estimates for males and, in any case, statistically insignificant.

The probability differences in table 7 once again represent the marginal effects of the second-generation immigrant status. As was the case in the sample of males, second-generation immigrant females also have a lower probability of finishing with a below high-school education, and are more likely to have a post-secondary degree than similarly-aged daughters of Canadian-born parents. In fact, this positive effect of being a second-generation immigrant is even stronger for females than it was for males. Moreover, being from a family with two immigrant parents appeared more beneficial than having just one immigrant parent. This is indeed a surprising result that runs contrary to the findings of a number of mainly European studies. These studies have documented that children, and especially females, from immigrant families often fall behind children born to the native population in terms of education.

As for the marginal effect of foreign mother tongue, estimates in table 8 show no significant impact on any of the four educational attainment probabilities. In other words, having a foreign mother tongue provided neither advantage nor disadvantage over the

children whose first language was English/French. Once again, this is an important result by itself. It indicates that a foreign-language environment, and the lack of exposition to the language of the host country are not ultimately detrimental to the educational attainment of children of immigrants.

5 SUMMARY AND CONCLUSIONS

The focus of this paper has been on the comparison of educational attainment of second-generation immigrants with the attainment of similarly-aged offsprings of Canadian-born parents. Despite the policy relevance of questions regarding the integration of immigrant children into the economy and society of the host country, researchers have so far paid much more attention to the adjustment process experienced by their parents, the first-generation immigrants, rather than to their children

For the purpose of this paper, a sample of Canadian males and females between 16 and 65 years of age was extracted from the cycle 15 of the 2001 General Social Survey, and two distinct definitions of what constitutes a second-generation immigrant were used. One definition regarded an individual as a second-generation immigrant if at least one of his/her parents was a foreign-born immigrant to Canada; the other definition was more restrictive as it required both parents to be foreign-born immigrants. Both definitions also included first-generation immigrants, as long as they arrived in Canada at the age of 9 or younger.

In order to discern the impact of being a second-generation immigrant on schooling attainment, four levels of education were distinguished and a set of ordered-

choice models was estimated in which the children of immigrants and similarly-aged children of Canadian-born parents were differentiated. For both the sample of males and the sample of females, the same models were estimated for the two chosen definitions of second-generation immigrants. In contrast with many previous studies, the results confirmed that the second-generation immigrants did better in terms of educational attainment than their peers born to Canadian parents even when the effects of selected observable characteristics were controlled for. This result, despite some differences in the magnitudes of the effect, held not only across the two definitions of second-generation immigrants but also across both genders.

Initially, the analysis of the effect of foreign mother tongue also yielded an interesting result, as the effect on educational attainment appeared to be positive and significant, although only for the sample of males. This effect, however, became statistically insignificant once unobserved heterogeneity was accounted for. In the female sample, the effect of foreign mother tongue was not significantly different from zero regardless whether unobservables were included in the model or not. This result contrasts with those reported in a number of European studies, but coincides with the study of second-generation immigrants in Canada by Worswick (2004): children of immigrants with a foreign mother tongue may have had some disadvantage in early school years, but by their mid-teens their performance on language (and mathematics) tests was at least as good as that of their English/French-speaking peers.

The main findings of this study appear to be mostly in contrast with the current body of research on immigrant children. They suggest that the Canadian immigration system was able to select immigrants whose children did, in terms of educational

attainment, at least as well as children of Canadian-born parents. The tribute should perhaps also be paid to the Canadian educational system; even if second-generation immigrants may have had some disadvantage in their early years, this disadvantage appears to have been eliminated as the children advanced through the school system, and had no impact on the final educational attainment.

As for policy implications, our results suggest that second-generation immigrants were not, at least until recently, a group that would require special attention. In terms of schooling, they were doing very well even without any actions or policies targeted at them. This should not, however, lead to complacency. As mentioned before, ethnicity was excluded as an explanatory variable, since the second-generation immigrant sample is ethnically homogenous to a large degree, with a vast majority being from families who immigrated to Canada from Europe or the United States. Although not yet reflected in the data, this is no longer true. At present, the majority of new immigrants are arriving from non-European countries, with the immigration from Europe and the U.S. shrinking. The experience from other parts of the world suggests that such changes in ethnic composition of the immigrant flow could eventually be reflected in a greater variation in educational outcomes of immigrant children, perhaps with a substantial number of them being at risk of significantly lagging behind the general population. Thus, further examination of this issue will be needed once appropriate data are available. Furthermore, even though in this paper the second-generation immigrants appeared to have done better in terms of schooling attainment than their counterparts from non-immigrant families, the question stands whether the advantage of higher education also translated into an adequate success in the labour market. It remains to be seen, whether the Canadian

economy provided enough opportunities for these immigrant children to capitalize on their effort and achievement in school, and to fully utilize their potential.

REFERENCES

Aydemir, A., & Sweetman, A. (2006). *First and Second Generation Immigrant Educational Attainment and Labor Market Outcomes: A Comparison of the United States and Canada*. Discussion Paper no. 2298, IZA Bonn.

Aydemir, A., Chen W. H., & Corak, M. (2006). *Intergenerational Earnings and Mobility among the Children of Canadian Immigrants*. Discussion Paper no. 2085, IZA Bonn.

Borjas, G. J. (1992). Ethnic Capital and Intergenerational Mobility. *Quarterly Journal of Economics*, 107(1), 123-150.

Borjas, G. J. (1994). Long-Run Convergence of Ethnic Skills Differentials: The Children and Grandchildren of the Great Migration. *Industrial and Labor Relations Review*, 47(4), 553-573.

Borjas, G. J. (1994). Immigrant Skills and Ethnic Spillovers. *Journal of Population Economics*, 7(2), 99-118.

Cameron, S. V., & Heckman, J. J. (1998). Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy*, 106(2), 262-333.

Card, D., DiNardo, J., & Estes, E. (2000). *The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s,* in Issues in the Economics of Immigration, ed. G. Borjas, NBER.

Chiswick, D. (1988). Differences in Education and Earnings across Racial and Ethnic Groups: Tastes, Discrimination, and Investment in Child Quality. *Quarterly Journal of Economics,* 103(3), 571-597.

Dustmann, C., & Theodoropoulos, N. (2006). *Ethnic Minority Immigrants and their Children in Britain*. Unpublished manuscript.: University College London.

Gang, I. N., Zimmerman, K. F. (2000). Is Child Like a Parent? Educational Attainment and Ethnic Origin. *Journal of Human Resources,* 35(3), 550-569.

Greene, W. H. (2003). *Econometric Analysis* (5th ed.). Prentice Hall, New Jersey.

Haveman, R., & Wolfe, B. (1995). The Determinants of Children's Attainment: A Review of Methods and Findings. *Journal of Economic Literature*. 33(4), 1829-1878.

Heckman, J. J., & Singer, B. (1984). A Method of Minimizing the Impact of Distributional Assumptions for Duration Data. *Econometrica,* 52, 271–320.

McKelvey, R. D., & Zavoina, W. (1975). A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology*, 4, 103-120.

Riphahn, R. T. (2003). Cohort Effects in the Educational Attainment of Second Generation Immigrants in Germany: An Analysis of Census Data. *Journal of Population Economics*, 16, 711-737.

Van Ours, J. C., & Veenman, J. (2003). The Educational Attainment of Second Generation Immigrants in the Netherlands. *Journal of Population Economics*, 16, 739-753.

Veall, M. R., & Zimmermann, K. F. (1996). Pseudo-R2 Measures for Some Common Limited Dependent Variable Models. *Journal of Economic Surveys*, 10(3), 241-259.

Worswick, C. (2004). Adaptation and inequality: children of immigrants in Canadian schools. *Canadian Journal of Economics*, (37)1, 53-77.

APPENDIX: TABLES

TABLE 1: PERCENTAGE DISTRIBUTIONS OF SELECTED OBSERVABLE CHARACTERISTICS

		Males (6,129 obs.)		Females (5,889 obs.)	
		Children of the Canadian-born	Second-gen. immigrants	Children of the Canadian-born	Second-gen. immigrants
Sample based on definition 1 of second-gen. immigrants (At least one parent immigrant)	Still in school	11.1	14.2	13.1	16.6
	Foreign mother tongue	1.2	23.8	1.1	24.1
	Complete family	88.8	88.4	85.8	88.4
	Father worked F/T	97.4	98.5	96.8	97.3
	Mother worked F/T	25.7	34.5	28.2	35.0
	Born in urban area	66.4	84.7	65.1	81.7
Sample based on definition 2 of second-gen. immigrants (Both parents immigrants)	Still in school	11.3	15.2	13.1	18.7
	Foreign mother tongue	1.4	38.4	1.4	39.7
	Complete family	88.4	90.8	85.9	89.7
	Father worked F/T	97.5	98.6	96.8	97.6
	Mother worked F/T	26.2	37.9	28.5	38.4
	Born in urban area	67.8	89.2	66.3	86.6

TABLE 2: PERCENTAGE DISTRIBUTIONS OF EDUCATIONAL ATTAINMENT: GSS RESPONDENTS

		Males (6,129 obs.)		Females (5,889 obs.)	
	Level of schooling	Children of the Canadian-born	Second-gen. immigrants	Children of the Canadian-born	Second-gen. immigrants
Sample based on definition 1 of second-gen. immigrants (At least one parent immigrant)	l.t. high school	17.8	13.8	14.7	11.5
	high school	18.3	16.6	20.2	16.6
	some PSE	41.9	37.5	44.4	42.3
	university	21.9	32.1	20.8	29.6
Sample based on definition 2 of second-gen. immigrants (Both parents immigrants)	l.t. high school	17.8	11.1	14.3	11.5
	high school	18.2	16.7	20.3	13.1
	some PSE	41.7	42.6	43.9	43.9
	university	22.4	29.7	21.5	31.5

TABLE 3: PERCENTAGE DISTRIBUTIONS OF EDUCATIONAL ATTAINMENT: FATHERS OF GSS RESPONDENTS

		Males (6,129 obs.)		Females (5,889 obs.)	
	Level of schooling	Canadian-born fathers	Second-gen. immigrant fathers	Canadian-born fathers	Second-gen. immigrant fathers
Sample based on definition 1 of second-gen. immigrants (At least one parent immigrant)	l.t. high school	50.8	42.3	50.5	40.7
	high school	21.8	20.8	21.0	19.9
	some PSE	14.0	16.0	15.6	18.5
	university	13.5	20.9	12.9	21.0
Sample based on definition 2 of second-gen. immigrants (Both parents immigrants)	l.t. high school	48.8	39.0	49.7	38.0
	high school	21.7	22.2	21.0	18.9
	some PSE	15.5	16.6	15.5	21.6
	university	14.0	22.1	13.8	21.5

TABLE 4: PERCENTAGE DISTRIBUTIONS OF EDUCATIONAL ATTAINMENT: MOTHERS OF GSS RESPONDENTS

		Males (6,129 obs.)		Females (5,889 obs.)	
	Level of schooling	Canadian-born mothers	Second-gen. immigrant mothers	Canadian-born mothers	Second-gen. immigrant mothers
Sample based on definition 1 of second-gen. immigrants (At least one parent immigrant)	l.t. high school	41.4	36.4	43.8	37.5
	high school	31.4	32.2	27.1	26.1
	some PSE	16.0	17.3	19.1	21.1
	university	11.2	14.0	10.0	15.3
Sample based on definition 2 of second-gen. immigrants (Both parents immigrants)	l.t. high school	39.9	38.6	42.4	42.1
	high school	31.5	32.0	27.4	23.3
	some PSE	17.1	15.4	19.5	20.2
	university	11.5	14.1	10.7	14.4

TABLE 5: ORDERED-PROBIT PARAMETER ESTIMATES: MALES

	Defn. 1 (At least one parent immigrant)				Defn. 2 (Both parents immigrants)			
	No unobs. het.		3 types		No unobs. het.		3 types	
Second-gen immigrant	0.149	(3.129)	0.184	(2.937)	0.166	(2.314)	0.325	(2.488)
Father's education								
High school graduate	0.159	(3.396)	0.203	(3.292)	0.159	(3.389)	0.254	(2.634)
Some PSE	0.296	(5.214)	0.366	(4.763)	0.297	(5.221)	0.548	(2.922)
University graduate	0.713	(11.162)	0.948	(9.412)	0.711	(11.138)	1.349	(5.457)
Mother's education								
High school graduate	0.229	(5.341)	0.339	(5.935)	0.232	(5.400)	0.470	(3.887)
Some PSE	0.421	(7.278)	0.549	(7.142)	0.426	(7.346)	0.790	(3.824)
University graduate	0.586	(8.445)	0.748	(8.016)	0.591	(8.514)	1.044	(4.290)
Foreign mother tongue	0.224	(2.476)	0.195	(1.605)	0.213	(2.254)	0.170	(1.061)
Complete family	0.228	(3.957)	0.334	(4.484)	0.222	(3.860)	0.393	(4.005)
Father worked full time	0.366	(3.241)	0.397	(2.887)	0.368	(3.254)	0.448	(2.987)
Mother worked full time	-0.070	(-1.655)	-0.115	(-2.052)	-0.071	(-1.667)	-0.137	(-1.855)
No. of siblings	-0.067	(-5.827)	-0.090	(-5.934)	-0.068	(-5.877)	-0.109	(-4.627)
Lived in urban area	0.171	(4.577)	0.236	(4.914)	0.175	(4.677)	0.307	(3.931)
Log-likelihood	-6,163.0		-6,125.0		-6,165.3		-6,129.5	

Z-statistics in parentheses. The baseline category for father's and mother's education is less than high-school. Coefficients for age and province of residence, and the estimates of the cut-off points are not included in the table for space considerations.

TABLE 6: ORDERED-PROBIT PARAMETER ESTIMATES: FEMALES

	Defn. 1 (At least one parent immigrant)				Defn. 2 (Both parents immigrants)			
	No unobs. het.		3 types		No unobs. het.		3 types	
Second-gen immigrant	0.181	(4.013)	0.239	(3.834)	0.332	(4.649)	0.420	(4.159)
Father's education								
High school graduate	0.193	(4.353)	0.267	(4.295)	0.191	(4.309)	0.265	(4.267)
Some PSE	0.329	(6.421)	0.490	(6.031)	0.322	(6.259)	0.480	(5.941)
University graduate	0.765	(12.676)	1.132	(9.137)	0.767	(12.713)	1.163	(9.002)
Mother's education								
High school graduate	0.282	(6.791)	0.350	(5.917)	0.292	(7.022)	0.358	(6.046)
Some PSE	0.479	(9.736)	0.649	(7.942)	0.485	(9.859)	0.658	(8.045)
University graduate	0.668	(10.125)	0.878	(8.211)	0.680	(10.302)	0.894	(8.211)
Foreign mother tongue	0.074	(0.890)	0.025	(0.228)	-0.017	(-0.185)	-0.068	(-0.575)
Complete family	0.172	(3.444)	0.235	(3.613)	0.166	(3.321)	0.230	(3.511)
Father worked full time	0.222	(2.591)	0.200	(1.887)	0.219	(2.547)	0.202	(1.899)
Mother worked full time	-0.042	(-1.093)	-0.080	(-1.551)	-0.043	(-1.107)	-0.082	(-1.567)
No. of siblings	-0.089	(-8.360)	-0.124	(-7.574)	-0.090	(-8.457)	-0.126	(-7.639)
Lived in urban area	0.050	(1.476)	0.071	(1.593)	0.048	(1.406)	0.067	(1.506)
Log-likelihood	-7215.8		-7185.7		-7213.0		-7184.5	

Z-statistics in parentheses. The baseline category for father's and mother's education is less than high-school. Coefficients for age and province of residence, and the estimates of the cut-off points are not included in the table for space considerations.

TABLE 7: THE MARGINAL EFFECT OF BEING A SECOND-GENERATION IMMIGRANT ON THE PROBABILITY OF ATTAINING J-TH LEVEL OF SCHOOLING

Level of schooling (j)	Defn. 1 (At least one parent immig.)				Defn. 2 (Both parents immigrants)			
	Males		Females		Males		Females	
Unobserved individual heterogeneity unaccounted for								
Less than high school	-0.031	(-3.276)	-0.031	(-4.264)	-0.033	(-2.485)	-0.052	(-5.484)
High school graduate	-0.018	(-3.032)	-0.025	(-3.910)	-0.020	(-2.236)	-0.046	(-4.548)
Some post-secondary	0.003	(3.646)	0.002	(2.136)	0.003	(2.595)	-0.004	(-1.012)
University graduate	0.045	(3.049)	0.053	(3.895)	0.051	(2.232)	0.102	(4.385)
Unobserved heterogeneity follows 3-type mixture								
Less than high school	-0.035	(-2.740)	-0.073	(-3.910)	-0.073	(-3.910)	-0.124	(-4.378)
High school graduate	-0.001	(-0.468)	0.010	(3.080)	0.010	(3.080)	0.011	(2.956)
Some post-secondary	0.015	(1.966)	0.040	(3.963)	0.040	(3.963)	0.073	(4.464)
University graduate	0.021	(2.651)	0.023	(3.566)	0.023	(3.566)	0.040	(3.900)

Z-statistics in parentheses.

TABLE 8: THE MARGINAL EFFECT OF FOREIGN MOTHER TONGUE ON THE PROBABILITY OF ATTAINING J-TH LEVEL OF SCHOOLING

Level of schooling (j)	Defn. 1 (At least one parent immig.)				Defn. 2 (Both parents immigrants)			
	Males		Females		Males		Females	
Unobserved individual heterogeneity unaccounted for								
Less than high school	-0.043	(-2.760)	-0.013	(-0.922)	-0.042	(-2.498)	0.003	(0.184)
High school graduate	-0.027	(-2.393)	-0.010	(-0.880)	-0.026	(-2.180)	0.002	(0.186)
Some post-secondary	0.001	(0.472)	0.001	(1.682)	0.002	(0.610)	-0.001	(-0.173)
University graduate	0.069	(2.361)	0.022	(0.874)	0.066	(2.153)	-0.005	(-0.186)
Unobserved heterogeneity assumed to follow a 3-type mixture								
Less than high school	-0.037	(-1.688)	-0.008	(-0.228)	-0.032	(-1.117)	0.021	(0.572)
High school graduate	-0.002	(-0.486)	0.001	(0.231)	0.001	(0.264)	-0.002	(-0.557)
Some post-secondary	0.015	(1.510)	0.004	(0.228)	0.002	(0.203)	-0.012	(-0.573)
University graduate	0.023	(1.486)	0.002	(0.227)	0.030	(0.992)	-0.006	(-0.576)

Z-statistics in parentheses.

ESSAY II

WAGE RETURNS TO POSTSECONDARY DEGREES AND OVEREDUCATION PREMIA: A PANEL STUDY OF CANADIAN GRADUATES

ABSTRACT

This study contributes to the literature on the wage effects of mismatches between job-education requirements and individual schooling attainment. It employs unique information from the two most recent cohorts of the National Graduates Survey, and makes use of an empirical model that acknowledges the role of both the demand and supply sides of the labour market in the formation of wages. The panel structure of the data is utilized so as to mitigate potential biases in the estimated effects due to unobserved heterogeneity. The results show that jobs that required post-secondary education paid on average higher wages than jobs that did not, and that the gap increased over time. Regardless of gender, the highest returns were in jobs that required a master's degree. The analysis also reveals substantial differences across the fields of study; graduates from "soft" fields (such as arts and social sciences) had substantially lower wages than those who graduated from health fields, business and "hard" fields (such as physics, mathematics or engineering).

1 INTRODUCTION

One of the goals of educational policies in Canada has been to increase the level of schooling of the country's population. This goal corresponds to the recognition that investments in human capital are essential for innovations and economic growth, and it is also a response to the increasingly stronger orientation of the labour market towards skilled workers. As a consequence, Canada's workforce is among the most educated in the world, and spending on education now represents a considerable portion of government expenditures.

There is a general agreement among researchers and policy makers that education benefits both the individual and the society, and not only in economic terms. Indeed, the body of research, especially on the economic returns to education, has expanded substantially over the past three decades, and continues to grow. Increasingly more attention is also being paid to the synchronization between the educational system and the labour market, and especially to the issue of underutilization of education and skills.

Although overeducation is sometimes seen as a result of misallocation of private and public resources, it is not necessarily a wastage. In fact, existing research shows that having more education than is required by one's job often pays off. In other words, there is usually a positive premium for extra schooling, presumably because workers with more education are more productive in their jobs. Nonetheless, there are some concerns that overeducation can also result from a long-term excess supply of highly educated workers for whom there are not enough appropriate jobs. Virtually all developed countries have

experienced two trends: a substantial increase in the educational level of the work force, and an increase in the demand for skilled workers. Which of these two trends dominates has an impact on both the rate of overeducation in the workforce as well as for wages. In their meta-analysis, Groot and van den Brink (2000) concluded that the overall incidence of overeducation in the labour market had been around 26%, and that it had remained much the same since the 1970s. They also found that the return to one year of education required rose from 7.9% in the 1970s and 1980s to about 12% in the 1990s, and that the return to one year of overeducation was rather small, around 2.6%.

The body of research that focuses on mismatches between the actual and required levels of schooling, and on the resulting differences in wages and other economic outcomes is still rather sparse. Vahey (2000) pointed out that no studies prior to his had examined the relationship between job-educational mismatch and wages in Canada. In his analysis of the returns to educational mismatch, Vahey used the National Survey of Class Structure and Labour Process in Canada (NSCS), and found that there were positive returns to overeducation for men in jobs that required a university bachelor's degree, but no significant returns to overeducation at all other levels or for women. Needless to say that although interesting, Vahey's attempt to analyze links between educational mismatch and wages has to be taken with caution. The use of the NSCS data was somewhat unfortunate as the survey's sample size was only about 3,000 respondents (Vahey's working sample consisted of 993 individuals), and the data were cross-sectional and not very recent as they came from interviews conducted in 1982. Furthermore, the regression analysis ignored the likely bias in the returns to required education and overeducation due to endogeneity.

Another Canadian study, Frenette (2003), used a sample of young graduates from the 1982, 1986 and 1990 cycles of the National Graduates Survey, and a definition of overeducation that is similar to the one used in this paper (see Section 2 for details). Using the ordinary least squares and first-difference panel data estimators, Frenette found that before controlling for unobserved heterogeneity both college and bachelor's graduates incurred rather large earning penalties for overeducation (about 10% and 19%, respectively), overeducated master's graduates faced a penalty of only about 3% , and there was no evidence of a penalty at the PhD level. When the first-difference estimator was used, the negative effects of overeducation became much smaller. [6]

One more inquiry into the issue of job-education mismatches in Canada came from Boothby (2002) who used the 1994 International Adult Literacy Survey (IALS) to examine the relationship between schooling, literacy and wages. Although the IALS is only a cross-sectional survey, it provides direct measures of literacy skills and literacy use at work which can be used to approximate unobserved ability and presumably mitigate the endogeneity problem in wage regressions. In his analysis, Boothby derived years of overeducation (and undereducation) from the difference between individuals' years of schooling and the years of training required for their jobs. Having included these measures along with literacy scores, and the measure of literacy use at work, in his wage regressions, he found that (a) overeducated workers earned more than workers in jobs

[6] Frenette regressed annual earnings on individual's schooling attainment rather than schooling required by the job. In the regression of this type, the coefficient on overeducation is usually negative, hence the term "penalty". As Hartog (2000) points out, this simply brings out the fact that overeducated workers are in lower level jobs than those who are not overeducated. For more on Frenette's paper, see Section 4.

with comparable educational requirements whose educational attainment matched these requirements, and (b) overeducated workers earned less than workers with comparable schooling whose schooling matched the requirements of the job. Furthermore, he also found that when added to the regression, both the measures of literacy skills and the measure of literacy use at work reduced the magnitude of the estimated effects of overeducation, although only for men.

This paper contributes to the research on the labour market effects of overeducation by examining wage differences among Canadian post-secondary graduates through the use of the 1995 and 2000 cycles of the National Graduates Survey (NGS). The focus is on individuals 35 years of age or younger who represent the majority of post-secondary graduates in both cycles. The NGS provides rich information about a variety of background characteristics of the respondents, and details about their education, fields of study and labour market outcomes. The approach adopted here extends the existing literature in a number of ways. First, since the NGS is a longitudinal survey (respondents are interviewed twice, two and five years after graduation), a panel data estimator has been used to eliminate potential bias in the returns to required schooling and overschooling due to endogeneity. With some exceptions, this problem has mostly been overlooked in the previous studies. Second, a specification of the wage regression has been used that takes into account both the demand and supply sides of the labour market, and allows to identify both the wage returns and the overeducation premia for each job-education category. Third, by including fields of study into the wage model, and by estimating the regression separately for the two NGS cycles and for the genders, the paper explores dimensions of the issue that are important but have so far been

ignored. Lastly, the paper demonstrates that the chosen wage model that allows for job-educational mismatch is superior to the competing models based on the theories of human capital and job competition.

The rest of the paper is organized as follows: Section 2 introduces the NGS, describes in detail all variables, and provides summary statistics for the samples used in the analysis. Section 3 describes the econometric model, its theoretical underpinnings, and issues pertaining to the identification and estimation of the model's parameters. Section 4 discusses the results. Section 5 provides a summary of the most important outcomes, and concludes the paper.

2 DATA

2.1 The National Graduates Surveys

This study uses data from the National Graduates Survey (NGS), a series of surveys conducted regularly by Statistics Canada. Each NGS samples over 20,000 individuals from a particular class of graduates from post-secondary institutions in Canada, and interviews them twice: two years and five years after graduation. So far five cohorts of graduates have been fully followed in NGS: classes 1982, 1986, 1990, 1995 and 2000. The surveys were designed primarily for monitoring and analyzing relationships between individuals' post-secondary studies and labour market experiences and outcomes in short and medium terms after graduation. They are well suited to address a variety of important topics such as the extent to which graduates of post-secondary programs had been successful in obtaining employment since graduation, the relationship between programs

and fields of study and subsequent employment, the graduates' job and career satisfaction, as well as the incidence and persistence of mismatches between graduates' qualifications and the requirements of the jobs they hold.

There are a number of advantages to the NGS when it comes to issues pertaining to the labour market outcomes of post-secondary graduates in Canada. All NGS samples are representative of their respective cohorts of graduates, and provide a wealth of detailed information on the sampled individuals, such as their work and schooling history, the programs and fields of study from which they graduated, details about the educational requirements and other characteristics of the jobs they held after graduation, and other information. An important feature of the NGS is its two-interview design which allows to employ panel data techniques to control for unobserved heterogeneity in wage regressions. This combination of detailed variables and longitudinal design makes the NGS a unique source of data for studies of post-secondary graduates in Canada.

The obvious limitation of the NGS, especially with respect to wage studies, comes from the focus only on graduates from post-secondary institutions. Thus, the NGS cannot be used to estimate rates of return to schooling since there is no information on individuals who do not have a post-secondary degree.[7] Nonetheless, given that post-secondary graduates represent a large and growing portion of the Canadian workforce, a

[7] The returns to post-secondary degrees presented in this paper have to be interpreted as relative to PSE graduates who worked in jobs that did not require a post-secondary degree.

close examination of their outcomes is well warranted and of interest to researchers, policy makers, as well as the general public.[8]

2.2 Measures of Job-educational Requirement and Overeducation

Following Hartog (2000), there are three distinct ways in which required education and overeducation can be measured: (1) systemic job-analysis by professional job analysts; (2) method of realized matches, and (3) worker self-assessment. The job analysis approach relies on trained experts who specify the required level and type of education for the job titles in an occupational classification. Essential to the objectivity of the procedure are clear definitions and detailed measurement instructions. However, a diligent analysis is costly when carried out on a large scale; consequently, updates to the education/skill profiles of the jobs are infrequent and often only partial. Thus, while conceptually appealing, job analysis has in reality serious drawbacks that limit its usefulness and can put the validity and reliability of the resulting information in doubt.

The method of realized matches derives a measure of job-educational requirement from a statistical measure of what is the common education for the job, typically the mean or the median of the distribution of years of schooling. Since the method is easy to apply, and it is often the only one available due to the absence of job requirement information from occupational analysis or worker self-assessment, it has been widely

[8] According to Statistics Canada Census of Population (2006), 48% of the population between 25 and 64 years of age, and 56% of those aged 25 to 34 have a post-secondary degree. When trade diplomas and certificates are included among post-secondary degrees, the proportion of post-secondary graduates increases to 60% overall and to 66.6% in the youngest cohort.

used in overeducation/undereducation literature. There are, however caveats associated with the use of realized-matches measures. When the job requirement is defined in terms of years of schooling, which has mostly been the case, the outcomes of the analysis can be affected by researcher's choice of a central-tendency statistic that defines the "typical" schooling for each job category. Depending on the actual shape of the required schooling distributions, choosing either the mean or the median can produce very different incidences of educational mismatches. Proportions of overeducated/undereducated in the sample can also be affected by the choice of distance from the required years of schooling beyond which an educational mismatch occurs. There is no scientifically defined standard as to what the distance ought to be. Some studies define educational mismatch simply as schooling greater or less than required, while others require schooling attainment to be outside a specified interval, such as one standard deviation above or below the required years of schooling.

The third approach to deriving required schooling and educational mismatches – the one used in this paper – is based on worker self-assessment of the schooling requirements of their job. Measures of overeducation/undereducation are derived by comparing the job requirement with respondent's educational attainment. The method of worker self-assessment of job requirements can have considerable merits over the job-analysis and realized-matches methods. It deals explicitly with the respondent's actual job, and not with broader constructs such as occupational categories. It also provides information that is up to date. However, like the other two methods, worker self-assessment is not without potential shortcomings. As Hartog (2000) points out, respondents may overstate the requirements of their job to inflate the status of their

position, or simply reproduce current hiring standards. In the latter case, the incidence of overeducation may be understated if job-education requirements increase over time as employers raise hiring standards in response to increasing educational attainment of the workforce while the jobs themselves have not really changed.

The choice of measures for job-educational requirement and educational mismatch ultimately depends on data availability. To my knowledge, there are no appropriate sources based on expert job-analyses for Canada.[9] Generally, the self-assessment approach, despite its subjective nature, is considered to have stronger advantages and fewer drawbacks than the statistical approach based on realized-matches, especially when a large micro-data set with specific information about respondents' jobs and education, such as the NGS, is available. In the 1995 and 2000 NGS, the sampled individuals were explicitly asked what was the level of education required to get their job at the time of the survey.[10] Their responses were then recoded into standard educational categories and, for the purpose of this study, further collapsed into five major groups: no post-secondary degree required, a college diploma below bachelor's, bachelor's degree or higher but below master's, master's degree, and PhD. Overeducation status was then

[9] Although the JobFutures website (www.jobfutures.ca) maintained by Service Canada may be a potentially useful source as it contains descriptions of typical skill, experience and schooling requirements for 256 occupational categories, it appears to rely heavily on quantitative analyses of data provided by Statistics Canada rather than on detailed expert probes into actual job requirements. Moreover, occupation descriptions are much broader than the more straightforward job-educational requirements derived from respondents' assessments in the NGS. Neither is it clear how often job descriptions are reviewed and updated.

[10] The "job" refers to the main job held in the week prior to the interview.

derived by comparing the job requirement with respondents' educational attainment. Since the incidence of undereducation was very small in the samples that consist entirely of post-secondary graduates (below 2% in each of the two NGS cohorts), the undereducated individuals were excluded. Relevant descriptive statistics for required education, overeducation as well as other variables are presented and discussed later in this section.

2.3 Other Variables

Besides the required-schooling and overeducation variables, the analysis makes use of information about respondents' educational attainment, earnings, work patterns, experience, job satisfaction, parental education, current family situation, and geographical location. It also explicitly incorporates respondent's choice of field of study, which is unique in the literature on earnings implications of overeducation. A respondent's level of education is based on the degree obtained by graduation from the program in the reference year 1995 or 2000 ("reference program").[11] Educational attainment is categorized in the same way as required schooling, except that it does not include the "no PSE required" category as the NGS samples consist only of PSE graduates.

[11] As some respondents had already held a post-secondary degree before graduating from the reference programs, an alternative definition based on the highest degree held in 1995/2000 was also tested. Varying the definition of educational attainment had only a very small impact on the incidence of overeducation, and virtually none in the wage regressions. This is not surprising as most respondents with a previous PSE degree obtained either a higher or at least an equally high degree in the reference years.

The fields of study were assigned so that they would correspond to the reference programs. Since the 2000 NGS adopted a new coding system based on the Classification of Instructional Programs (CIP), the fields-of-study codes in the 1995 NGS that were based on the Community College Student Information System (CCSIS) and the University Student Information System (USIS) had to be converted into CIP codes to allow for a comparison across the two surveys. In order to ensure a reasonably large samples within each field, the CIP categories were further collapsed into five major groups according to the prevailing contents: (1) "soft" fields,[12] whose curricula have either no or only limited quantitative component, pool together Education, Arts, Humanities, Social and Behavioural Sciences, and Law; (2) "hard" fields such as Physical and Life Sciences and Technologies, Mathematics, Computer Sciences, Architecture, and Engineering have a strong focus on quantitative skills; (3) "business" is a shorthand for Business, Management and Public Administration; (4) "health" fields include Health as well as Recreation and Fitness; (5) Finally the "other" category pools all remaining CIP fields, such as Agriculture, Natural Resources and Conservation, Personal Improvement and Leisure, whose samples were not large enough for a separate analysis, and which could not be justifiably included within the previous four groups.

The hourly wage variable used in the regressions was generated from respondent's gross annual earnings (in constant dollars) for the job held during the week

[12] The terms "soft" and "hard" are used only as a shorthand for the lack of other suitable denotations. They refer to the amount of emphasis on quantitative and/or mathematical content in the curricula, but they are in no way indicative of the level of complexity of the fields or the amount of abilities, skills and effort required to master them.

prior to the survey, and from the information on the usual hours worked in that job. There is a substantial difference in the definition of the earnings variable used in the 1995 and 2000 NGS, and the definitions used in the earlier surveys. Both the 1995 and 2000 NGS report an estimate of annual earnings for the job held in the reference week which was derived from respondent's reported salary, how it was paid and the usual hours worked, while previously the NGS respondents were asked to estimate their annual gross pay themselves.[13] Although it is generally unclear how this change could affect the behaviour of earnings, some researchers have noted its likely importance. For example, Gunderson and Krashinsky (2008) report that at the first interview the average real earnings for the cohort that graduated in 1995 were only 84% of the real earnings of those who graduated in 1990, a drop of some 6,000 in constant dollars. Such a large decrease suggests that the 1990 NGS earnings may have been overstated. The differences in earnings definitions are, unfortunately, impossible to reconcile across the surveys which is the main reason why the NGS cohorts prior to 1995 are not included in this study.

Capturing an individual's work experience represents a special problem when working with the NGS 1995 cohort. While there is a variable for accumulated work experience in the NGS 2000 file, there is no counterpart to it in the 1995 one. I used the information in the NGS 2000 to regress an individual's years of work experience accumulated before starting the program in 2000 on his/her prior schooling and age at the

[13] The actual question was, "Working your usual hours, approximately what would be your annual earnings before taxes and deductions at the reference week job?".

time. I then used the estimated parameters to impute years of work experience to the respondents in the 1995 cohort.[14]

2.4 Sample Selection and Summary Statistics

After excluding missing observations, the 1995 NGS contains over 21,500, and 2000 NGS about 20,600 individuals. As mentioned before, the NGS does not provide much information about respondents' work lives prior to the survey. This implies difficulties in controlling for conceivably large differences in labour market outcomes between graduates with and without an extensive work history. To deal with this problem, the paper focuses on graduates who were 35 years old or younger at the time of graduation, many of whom had a limited or no prior work experience. These individuals represent the majority of graduates in both surveys, about 73% of the initial 1995 sample, and 79% of the 2000 one. Further selection conditions had to be imposed to ensure that the respondents were observed at both interviews and that panel data analysis would be feasible. All individuals who pursued additional education after graduating from their reference programs were excluded (30% in 1995, 39% in 2000), as well as those who were either unemployed or out of labour force in any of the two interviews (12% in 1995,

[14] The regression model had a 69% fit. Also, for the 2000 cohort, using the proxy in the place of the actual work experience made virtually no difference in the results of the wage regressions.

6% in 2000), and those who were self-employed (about 5% in each sample).[15] The final working samples contain 7,204 graduates from the NGS 1995, and 6,703 from the NGS 2000. The descriptive and regression analyses are presented separately for the two NGS classes of 1995 and 2000, as well as for both genders as tests for poolability and the regression results themselves strongly suggest that combining data across the graduation cohorts and/or genders would be erroneous.

The gender compositions of the two NGS samples were quite different: women represented about 47% of the 1995 sample while they accounted for over 54 % of the class 2000. This is no surprise as the trend of growing female participation in post-secondary education in Canada has been reported in a number of government and academic studies. Regarding the distribution of hourly wages, table 1 reveals some variation across the genders and levels of education required by the job, but it appears that overeducation status plays a more prominent role in determining wages than gender does. In every interview and at every job-education level, the mean wage was higher for those who were overeducated than those who were perfectly matched. No such unambiguous pattern can be detected from comparing wages for men and women.

Table 2 shows the distributions of the respondents by job-education requirement and overeducation status. The majority of men and women from the 1995 class were

[15] Although inferences based on subsamples may be subject to a selection bias, it could be argued (see, for example, Verbeek and Nijman, 1992) that as long as the selection mechanism for a given individual does not change over time, its effect is absorbed in the time-invariant heterogeneity component, and removed through fixed-effect differencing. In that case, the Hausman-Taylor estimator used in this paper can consistently estimate paramaters corresponding to the time-varying variables in the regression.

working in jobs that required either a college diploma or a bachelor's degree, and the same pattern carried over to the 2000 class. For both NGS samples, the proportions of individuals working in these jobs somewhat declined from the first to the second interview in favour of higher-level jobs which suggest that some of the respondents moved up into more demanding jobs. Over the same period, the proportions of those who worked in the lowest-level jobs stayed rather stable, with the exception of men, class 1995, for whom it slightly increased over time. Nonetheless, there did not seem to be any clear trend either in the upward or downward mobility.

As for job-education mismatches, for both NGS classes as well as genders, the highest proportions of overeducated were by far in jobs that required a bachelor's degree (ranging from about 24.3% to 30.3% depending on the gender and period of observation). Between the 1995 and 2000 NGS , there was a noticeable increase in the proportion of overeducated men in master's level jobs (i.e., mismatched doctoral graduates). With the data at hand, it cannot be distinguished whether the increase in the incidence of overeducation among male PhDs was a signal of a longer-term structural imbalance between the supply and demand in this particular segment of the labour market, or whether the observed increase was only a transient phenomenon caused, perhaps, by momentary economic conditions that were more favourable for the 1995 class whose members were entering the labour market during a period of high economic growth. Although the scope of this study does not allow for a detailed probe into this event, it is worth noticing that the regression analysis (tables 5 and 6) revealed a significant wage premium for these men. In other words, male PhD's from the 2000 class who were nominally overeducated were getting a significant and large wage premium over Master's

graduates in similar jobs. Thus, even though the incidence of job-education mismatches for these overeducated men with advanced post-graduate degrees rose over time, their wages were actually rather high.

The distributions in table 3 show great differences between men and women in terms of their choice of the field of study. Men in both NGS cohorts were heavily represented in hard fields (close to 60%), and under-represented in health fields. Women appeared more evenly distributed over the four major field-of-study groups. In both surveys, the majority of females graduated from soft fields (28.6 and 33.4%, respectively), with business and health following with somewhat smaller numbers. A comparison of females from the 1995 and 2000 samples reveals a noticeable shift from business fields (5 percentage points smaller than the corresponding group in 1995) in favour of health and, especially, soft fields. The field with the largest portion of overeducated was, somewhat surprisingly, business and management. This was true for both NGS cohorts, genders as well as for both survey interviews. Almost half of the students of business ended up overeducated (the highest figure was 51.5% for the 2000 NGS men). Nonetheless, the remaining fields of study also had rather large portions of people whose schooling was mismatched with the requirements of the job they held after graduation.

Table 4 is a set of matrices that show transitions into or out of being overeducated between the 2-year-after and 5-year-after interviews. The fractions of the overeducated are very similar for both NGS cohorts, and do not differ substantially even across genders. In all periods, about one third of women and men had more education than was required by their jobs. The data, however, reveal large differences in the numbers of state

transitions between the 1995 and 2000 classes. About 11% of men who graduated in 1995 and were overeducated in the first interview moved into jobs that matched their schooling by the second interview, while at the same time over 14% of those who were not overeducated at first became so by the second interview. Women were not very different in this respect – 12.6% experienced transition out of overeducation and about 12.3% into it. By contrast, transitions in and out of states were much less common for the 2000 graduates, both male and female. The actual number was around 4% in each direction separately which suggests a substantial state dependence, and may, perhaps, reflect a less flexible labour market in the early 2000s.

Although the proportions of overeducated in all survey years appear high, they are in fact consistent with the findings of previous studies, foreign and Canadian. Besides, the fact that some people have higher educational credentials than required for the jobs they hold, does not mean much by itself. There may be a number of reasons why they are mismatched. Some may be mismatched involuntarily simply because there are not enough jobs available that would require their level of education, while for others a mismatch may be a matter of choice determined by their personal preferences, life circumstances or expectations. Even though a detailed analysis of the causes and effects of overeducation is beyond the scope of this study, a close examination of how the market rewards post-secondary degrees, and whether overeducation matters in explaining variations in wages is important as it sheds light on the implications of job-educational mismatch for the welfare of individuals.

3 MODEL AND METHODOLOGY

3.1 Theoretical Background

The principal idea behind the specification of wage regression introduced here comes from a model of labour market proposed by Hartog (1986a, 1986b). Hartog's labour market consists of workers characterized by skills (the supply side) and jobs characterized by "job difficulty" (the demand side), and it takes a short run view by assuming that the skill level of individuals and the level of difficulty of jobs are fixed.[16] The market then has to solve the problem of matching workers to jobs, i.e. allocating worker's with different skills to jobs characterized by different levels of difficulty. Hartog argues that wages are instrumental in the process, and that the dependence of the wage rate on both worker's level of skills and level of job difficulty should be allowed. Although the lack of data makes it infeasible to model the actual allocation process, the wage regression specified below recognizes that the return to skills can, at least in the short run, depend on both the supply and demand sides of the labour market. Furthermore, the specification not only allows to study wage consequences of overeducation, but can also be tested against two competing specifications: one based on the human capital model, and another on the job-competition theory.

[16] Most empirical work, including this study, equate workers' skills with schooling attainment (years of schooling or highest level completed), and characterize job difficulty in terms of job-educational requirement (schooling required to obtain or hold the current job).

The long-run equilibrium specification of the human capital theory embodied in the traditional Mincerian wage equation dominates the literature on wage differentials among individuals. The theory attributes differences in wages solely to differences in individual characteristics, most prominently in education and work experience. The human capital model ignores variables determined by the demand side of the labour market, and does not allow for the possibility of a mismatch between worker's skills and the skill requirement of his job. However, individual characteristics on their own are unlikely to fully account for wage differences. As Hartog (1987b) argues, the structure of the demand side of the labour market, and the allocation of individuals within it, are important for individual earnings.

Contrary to the human capital theory, the job competition model proposed by Thurow (1975) regards worker's human capital as irrelevant for wage rates which are assumed to depend entirely on job characteristics, thus giving prominence to the demand side of the labour market. According to Thurow, individual characteristics may affect the distribution of workers across jobs, but they have no impact on wage rates.[17] Clearly, Hartog's "allocation" model lies between the two extremes of the human-capital and job-competition theories; it takes a short-run view of the labour market which assumes that both the supply and demand sides are important in explaining variation in wages. The next section proposes a way to test the validity of all three models empirically.

[17] The job-competition model can be seen as a subset of what is sometimes referred to as segmented labour market theories; for more details, see a survey by Cain (1976).

3.2 Wage Regression

Job-educational requirements are categorized into five mutually exclusive levels: (1) below post-secondary, (2) a college degree below bachelor's, (3) bachelor's degree, (4) master's degree, and (5) PhD. An individual's schooling attainment is categorized in the same way, but only in levels (2) through to (5) as the NGS samples only graduates from post-secondary programs. In terms of schooling, individuals and jobs can either be matched, when an individual's schooling attainment exactly corresponds to the schooling required by his job, or mismatched. In the context of this paper and the patterns observed in the NGS data, a job-educational mismatch is assumed to take only a specific form of overeducation by one schooling level.[18] Let $i = 1, \ldots, n$ be the individual (worker) subscript, and $t = 1, 2$ denote the time periods (i.e. interviews conducted for each of the NGS cohorts); the regression model can be written as

$$\ln(w_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\gamma} + \sum_{j=2}^{5} \theta_j \cdot r_{jit} + \sum_{k=2}^{4} \varphi_k \cdot (over_{it} \cdot r_{kit}) + \mu_i + \varepsilon_{it}, \qquad (3)$$

where w_{it} denotes hourly wage, $\mathbf{x}_{it}$ is a vector of time-varying individual background variables, and $\mathbf{z}_i$ contains all time-invariant observable characteristics. The job-educational requirement is captured by the dummy variable r_{jit} that equals 1 if in period t an individual i held a job that required *j*-th level of education. The variable $over_{it}$ takes

[18] The model can be easily extended to include undereducation as well as job-education mismatches by more than one level. However, in the NGS samples used in this paper only a very small number of respondents experienced these phenomena, and there was no reason to include them explicitly in the wage model in (3).

value of 1 if the individual's education attainment exceeded the schooling level required by his/her job, and 0 if they perfectly matched. Interacting the required levels with the overeducation dummy explicitly allows for the possibility of different premia for overeducation at each required level of education. The remaining two terms in the regression are the random wage shock, ε_{it}, and the unobserved time-invariant effect, μ_i, which is considered to reflect an individual's innate ability, motivation, attitudes, etc.

The baseline level of job-education requirement is "no PSE required". As the NGS samples consist entirely of PSE graduates, anyone who worked in a job of this category was overeducated. Also, by definition, the highest required level of education as well as the highest possible schooling attainment is Ph.D. It follows that those who worked in jobs that required a doctorate could not be overeducated. The following schematics summarizes which "returns" to education and overeducation that can be identified in equation (3):[19]

		Worker's schooling attainment			
		2 College	3 Bachelor's	4 Master's	5 PhD
Schooling required by the job	1 Below PSE	*Baseline*	×	×	×
	2 College	θ_2	$\theta_2 + \varphi_2$	×	×
	3 Bachelor's	×	θ_3	$\theta_3 + \varphi_3$	×
	4 Master's	×	×	θ_4	$\theta_4 + \varphi_4$
	5 PhD	×	×	×	θ_5

[19] The term "return" does not have the standard connotation of the "rate of return" as in economic theory. In this study, it is used to avoid wordiness in describing differences in the market rewards to different levels of post-secondary education relative to the chosen reference group.

3.3 Estimation Issues

The estimation of parameters in equation (3) poses a problem due to the unobserved individual-specific effect, μ_i, which cannot be directly controlled for, but is likely to be correlated with at least some of the explanatory variables. The schooling and field-of-study variables are likely to be correlated with the unobservables, and ignoring their endogeneity, as has often been the case in earlier studies, results in biased estimates of their effects on wage. The way to alleviate the problem in this study is to use the two-interview design of the National Graduates Survey (NGS). Although ordinary least-squares and random-effects, two popular estimators, are of no use as they both require the regressors to be uncorrelated with μ_i, a fixed-effects estimator is still a viable option. In fixed-effects, the unobserved individual heterogeneity is effectively removed by taking differences of the variables from their time-means. Unfortunately, this also differences out all other time-constant variables so that their effects ($\boldsymbol{\gamma}$) on wage are not identified. Given the focus of this paper, the (relative) returns to required schooling and the overeducation premia could still be estimated using the fixed-effects approach, but not the other effects of interest, such as that of the field of study.

An alternative estimator that can deal with the endogeneity problem, and allows to identify all regression coefficients has been proposed by Hausman and Taylor (1981). It is based on the assumption that some of the explanatory variables are uncorrelated with the unobserved individual heterogeneity and can, therefore, be used to construct suitable instruments for the endogenous regressors. Let $\mathbf{x}_{1it}$ and $\mathbf{z}_{1i}$ be the vectors of exogenous

regressors, and $\mathbf{x}_{2it}$ and $\mathbf{z}_{2i}$ contain the variables presumably correlated with μ_i, then the model in (3) can be recast as[20]

$$y_{it} = \mathbf{x}'_{1it}\boldsymbol{\beta}_1 + \mathbf{x}'_{2it}\boldsymbol{\beta}_2 + \mathbf{z}'_{1i}\boldsymbol{\gamma}_1 + \mathbf{z}'_{2i}\boldsymbol{\gamma}_2 + \mu_i + \varepsilon_{it}, \quad (4)$$

where the error term ε_{it} is assumed independently and identically distributed as $\mathrm{N}(0,\sigma_\varepsilon^2)$, and uncorrelated with all the right-hand side variables and the unobserved individual term μ_i. Hausman and Taylor showed that the deviations from individual means, $(\mathbf{x}_{1it} - \bar{\mathbf{x}}_{1i\cdot})$ and $(\mathbf{x}_{2it} - \bar{\mathbf{x}}_{2i\cdot})$, vector $\mathbf{z}_{1i}$ and that the individual time means $\bar{\mathbf{x}}_{1i\cdot}$ can serve as the remaining instruments. The condition for this instrumental variable estimator to be identified is $\dim(\mathbf{x}_{1it}) \geq \dim(\mathbf{z}_{1i})$.[21]

The obvious limitation of the Hausman-Taylor approach is that it requires specification of which explanatory variables are correlated with μ_i and which are not, and that the link between the identifying assumptions and job-educational variables may be somewhat tenuous. It is, nevertheless, possible to check whether the instrumentation of the selected endogenous variables is sufficient to remove any correlation between the

---

[20] To simplify the notation in regression (4), the vector of endogenous time-varying variables, $\mathbf{x}_{2it}$, now also contains the job-education requirement dummy, r_{jit}, and the overeducation status indicator, $over_{it}$.

[21] Amemiya and MaCurdy (1986) proposed a similar and generally more efficient IV estimator. that utilizes the entire history $\mathbf{x}_{1it},\ldots,\mathbf{x}_{1iT}$ and not just the time average $\bar{\mathbf{x}}_{1i\cdot}$. Other projections (see e.g., Breusch, Mizon and Schmidt, 1989) could also be used to generate suitable instruments. In this study, the Amemiya-McCurdy version of the IV estimator was also used and compared to the estimates from the original Hausman-Taylor estimator. The choice of the particular specification had almost no effect on the estimates, which may be due to the short time dimension (only two periods) of the panel used for this study.

person-specific heterogeneity and the remaining regressors. Given the assumptions stated above, the Hausman-Taylor estimator is both a consistent and efficient estimator of the regression parameters. Under the same conditions, the fixed-effects estimator is also consistent but it is not efficient. This information can be used to test whether the exogeneity assumptions $\mathrm{E}(\mathbf{x}'_{1it}\mu_i) = \mathbf{0}$ and $\mathrm{E}(\mathbf{z}'_{1i}\mu_i) = \mathbf{0}$ hold using a Hausman test. Ideally, one would like to expand the Hausman-Taylor framework to include exogenous variables from outside the sample data which would allow to test the sensitivity of the results towards the identifying assumptions.[22] However, to my knowledge, no such instruments are available for use with Canadian data.

4 RESULTS

The Hausman-Taylor estimates of the regression model in equation (3) are presented in full in table 5; table 6 then summarizes the relative wage differences and overeducation premia in order to simplify interpretation. This section focuses on two main aspects of the results: the change over time (i.e., differences between the classes 1995 and 2000), and differences between the genders. It is important to note that the wage differences as discussed below need to be interpreted as relative to the reference category which consists of PSE graduates who worked in jobs that did not require a post-secondary degree. As the NGS surveys only graduates from post-secondary institutions, all individuals in the baseline category were overeducated by design. The fact that

[22] Hansen and Wahlberg (2005), for example, included indicators for exogenous changes in the Swedish educational system, specifically in compulsory schooling, as additional instruments in their version of the Hausman-Taylor model.

individuals with schooling levels below post-secondary are absent from the samples is balanced by the extent and depth of information the NGS surveys have to offer. To my knowledge, such a degree of control for relevant observables has not yet been implemented in any study of the wage effects of job-education mismatches. Besides educational variables, such as the level of schooling required by respondent's job, his/her degree and field of study, a variety of individual-background variables was also utilized in the regressions (parental education, marital status and an indicator for dependant children), along with controls for the region of residence and local economic conditions (provincial unemployment rate).

As mentioned in the previous section, the choice of the Hausman-Taylor estimator dictates that assumptions are made as to which right-hand side variables in the wage regression may be endogenous. Given the focus of this study, the data constraints and the identification condition for the estimator, the variables representing the level of schooling required by the job, the overeducation status, and the field-of-study indicators are assumed endogenous. The Hausman specification tests indicate that the imposed exogeneity conditions hold for both NGS cohorts as well as for both genders.

4.1 Relative Returns for Post-Secondary Degrees

The first thing to notice in table 6 is that the magnitudes of the relative wage differences varied considerably across the job-educational requirement categories. As expected, the higher the degree required on the job, the higher the wage – with a notable exception of PhD-level jobs which, in terms of wage, never quite matched master's. The second

pattern to notice is that for all schooling levels, regardless of the gender, the relative market rewards increased between the two cycles of the NGS.

For men and women alike, the highest payoff was in jobs that required a master's degree. For men in master's level occupations, the wage gain over those in the reference group rose from 28.3% to 39.8% between the two NGS cycles. For women in master's level jobs, the relative "return" to schooling was already very high in the 1995 cycle, 43.6%, and it further increased to 52.1% in the 2000 cycle. Bachelor's and PhD-level jobs were also paying relatively higher wages, and the gap between them and the reference category of unskilled job increased over time as well. There were also wage premia in occupations that required only a college diploma below bachelor's (with the exception of the 1995 sample of men), but they were substantially smaller in their magnitudes that those offered by higher-skilled jobs. In fact, wages offered by college-diploma occupations were closer to the wages paid by unskilled jobs than to the wages paid by bachelor's or higher level ones.

Over time, for both men and women, the largest increase in the relative wage return occurred in PhD-level jobs. This was especially true for men. While men from the 1995 class in doctoral-level jobs were being paid a mere 12.7% over jobs that did not require any post-secondary degree, those who graduated in 2000 found the market considerably more rewarding as these jobs were now paying 30.6% over the baseline. However, despite the increase, PhD-level jobs still offered less than those that required only a master's degree, even though the gap between them had narrowed greatly.[23]

---

[23] Tables 5 and 6 show these changes in detail.

4.2 Overeducation Premia

Overall, across both cycles and gender, only those who were overeducated in jobs requiring college diploma were earning a large and significant bonus compared to those who worked the same jobs but were not mismatched. At this level, the premium was about 23% for the NGS 1995 men, and fell to 16.9% for those who graduated in 2000. For women, this premium showed a very slight increase between the two NGS cycles – from 16% to 17.5%, respectively. Women who were overeducated at bachelor's level also received a significant premium over those who were perfectly matched (24.4% for the class 1995, and 16.4% for the class 2000,), and so did men from the 1995 cohort (8.1%). However, there was virtually no premium at this level for the NGS 2000 men.

Most surprisingly, PhD graduates working in master's level jobs did not receive any extra benefit over those who exactly fit the job-education requirement – with one notable exception: men who graduated from a doctoral program in 2000 and went to a job that required master's degree were earning a premium of 24.2% over those with a master's degree in a similar job. In fact, the premium was so large that these mismatched PhDs enjoyed, on average, a higher wage than PhD graduates in PhD-level jobs. This was, however, the only exception, as the female PhD's and men from the 1995 class did not receive any significant overeducation premium.

For both men and women, the overeducation premium in bachelor's level occupations declined (by 8.9 and 8 percentage points, respectively) between the two NGS surveys. The premium for job-educational mismatch in PhD-level jobs rose for both genders, but the increase was significant only for men. There was also a decrease in the

overeducation premium in college-diploma occupations for men, while the premium for overeducated women in these jobs did not change significantly.

4.3 The Effects of Other Observable Characteristics

Besides the schooling and overschooling, a set of controls was used for other relevant differences among the sampled individuals. The choice was guided by the economic theory as well as empirical practice and statistical measures. According to the regression results in table 5, years of work experience had a significant impact on wages which was somewhat stronger for men than women, and declined for both between the surveys. The coefficient on experience-squared is significant and negative as expected. Having a permanent job, as opposed to a temporary one, appeared to boost the wage. The effect was again stronger for men, especially those who graduated in 2000, and it strengthened over time. Dissatisfaction with the job had a significant negative impact on wage across both cycles and genders. This is not surprising as job satisfaction is generally presumed to affect employee's performance and productivity, and be reflected in wages accordingly.

Mother's and father's schooling had an overall positive impact on wage, although not statistically significant across the board. The link between parental education and an individual's wage also appeared to have eroded over time. Family circumstances seemed to matter more for women: women who were either married or cohabiting with a common-law partner had a slightly higher wage (by about 2%) than those who did not. Having dependent children had a negative effect on wage that was generally small but somewhat more pronounced for the 1995 female sample. The effect of geographical

location was significant only for men from the 1995 class and women from the 2000 class. In these samples, all regions had lower wages on average in comparison with Ontario, the chosen baseline. Local unemployment rate had no effect on men's wages; it was positive and statistically significant for women, but virtually negligible in its magnitude.

The overeducation literature has so far ignored the potentially important role of the field of study in explaining variation in wages. Such an exclusion of an influential explanatory variable may be explained by the lack of appropriate data, but it is otherwise hardly defendable. A failure to control for the field of study may lead to substantial omitted-variable biases in other regression parameters, and also avoids answering questions of interest to policy makers. For example, there is a growing concern in Canada with regards to the applicability of the so-called "soft" fields, such as social sciences and arts, in the labour market. There are indications that graduates from programs with little or no technical and/or applied content may generally have more difficulties to find an appropriate employment, advance their careers, and reap the benefits of their investment into higher education than graduates from "hard" fields. The results in table 5 suggest that the concern about the labour market outcomes of graduates from "soft" fields may be to some extent warranted. It appears that across both NGS surveys as well as genders, the choice of "business", "hard" sciences and "health" paid off compared to the "soft" fields.[24] For men, the biggest gainers were "hard" fields in the 1995 cohort (17%

[24] The one and only exception being the 1995 female graduates from business fields, who had no gain over those who graduated from the "soft" fields.

premium over “soft” fields), and “hard” and “business” fields in the 2000 class (both about 18%). For women, the highest premium was, by far, in “health”: 17.2% for the class 1995, and 19.5% for the 2000 cohort. “Hard” sciences had also benefited women, but the premia were less than half of those offered to graduates from “health” fields. As can be seen in table 5, all the relative gains (over “soft” fields) also increased between the two NGS surveys, although to a varying extent.

4.4 Unobserved Individual Heterogeneity and the Endogeneity Bias Problem

The existing literature on the wage effects of overeducation has so far largely ignored the potential endogeneity in the schooling variables. Two exceptions, Bauer (2002) and Frenette (2003), addressed the problem by resorting to a fixed-effect model for the unobserved individual components. Despite differences in data and methods, both authors concluded that unobserved heterogeneity matters, and that a failure to account for it can severely bias estimates the return to education as well as of the effect of under/over-education on earnings. This study comes to a similar conclusion; as documented in tables 8 and 9, the difference between the least-squares and Hausman-Taylor coefficient estimates are in some cases substantial. In the 1995 NGS sample, OLS clearly overestimates the returns to education required by the job for both men and women. There does not appear to be a clear-cut pattern in the 2000 cohort. Similarly, no pattern seems to be underlying biases (some of which are very large) in the overeducation premia. Nevertheless, it is clear from both tables, that the differences between the OLS and Hausman-Taylor estimates are substantial, and that correcting for unobservables is rather important.

4.5 Regression with Overeducation vs the Human Capital and Job Competition Models

As pointed out in the theory section, the wage model in (3) not only allows us to study wage consequences of overeducation, but it is also well suited for testing two competing paradigms: the human capital embodied in Mincer's (1974) wage regression, and one derived from the job competition model of Thurow (1975). The two models are nested within (3), and their implications can be translated into parameter restrictions and tested using a t-test.

The human capital model suggests that individual earnings are entirely dependent on individual characteristics, most importantly on education and work experience. This implies that job characteristics (in this case job-education requirements) have no impact on wages; workers with the same level of schooling will receive the same return regardless of whether they are overeducated or properly match their job's educational requirement. With respect to the regression in (3), this translates into $\theta_j + \varphi_j = \theta_{j+1}$ for all $j = 1,2,3,4$, or in a more compact way $\theta_5 = \varphi_2 + \varphi_3 + \varphi_4$.

According to Thurow, only job characteristics determine wages while worker characteristics play no role. Thus, the job-competition model emerges from equation (3) if $\varphi_2 = \varphi_3 = \varphi_4 = 0$. Tests performed for both NGS cohorts and genders show that the wage model with overeducation in equation (3) is superior to the specifications based on the human-capital and job-competition models.

4.6 Comparison with Existing Literature

The existing studies of job-educational mismatch vary greatly in terms of regression models and data sources used to estimate them. Nevertheless, there are commonalities in terms of what they can tell us about the incidence and wage implications of such a mismatch. For example, Hartog (2000) evaluated a set of empirical studies from the United States, the Netherlands, Portugal, Spain and the United Kingdom, and concluded that in most cases a proper job-educational match was about 60%. Regarding the implications for wages, the return to overeducation was positive, but smaller than the return to required education (both measured in years).

How do these general results compare to the findings in this study? As tables 2, 3 and 4 show, the incidence of overeducation varies very much with the NGS cohort, gender, job-educational requirement and field of study. Nevertheless, the overall incidence rate stayed within 30-35% range at all times for both genders, which is similar to the numbers reported elsewhere. The literature on overeducation also suggests – and this study confirms – that a regression model with separate returns to required schooling and overschooling is preferable to regression equations derived from the human-capital and job-competition models.[25]

As for the implications of overeducation for wages, from the patterns described by Hartog, we would expect that while having more education than required by one's job is generally beneficial, an overeducated worker would get a higher return to his/her

[25] See e.g. Duncan and Hoffman (1981), Hartog and Oosterbek (1988), Hartog (2000) or Bauer (2002).

schooling if he moved into a job that would actually require his level of education. This means that we would expect the overeducation premia to be positive, i.e. $\varphi_j > 0$, and at the same time $\theta_j + \varphi_j < \theta_{j+1}$ for $j = 2,3,4$. As previously discussed, the results presented in this paper do not generally support these hypotheses. Depending on the gender, survey cohort, level of education required by the job and the field of study, some estimates of overeducation premia are not significantly different from zero, while others are very high (see table 6 for an overview).

It would indeed be interesting to see how the results in this paper compare to what has so far been found in Canadian data. The most natural point of comparison in this case would be the study by Frenette (2003). Frenette used three older cohorts of the NGS, namely 1982, 1986 and 1990, to estimate a first-difference wage regression model with individual schooling attainment (rather than schooling required by the job) and overschooling status. The differences in methodologies make a direct comparison of estimates impossible, but some general points of correspondence can be established. For one, there is an agreement over the role of unobserved heterogeneity; if unaccounted for, the correlations between the unobservables and the schooling and overschooling variables can lead to severely biased estimates of the effects of these variables.[26] Frenette also reports substantial variation in the effects of overeducation across the levels of education which can be confirmed in the more recent samples of the NGS used in this study.

[26] Evidence of the importance of controlling for unobserved individual heterogeneity also comes from foreign studies. For example, Bauer (2002) reports a substantial endogeneity bias in his study based on the German Socio-economic Panel.

Ideally, I would have liked to estimate the same model applied in this paper using the older NGS cohorts, and directly compare the results. Unfortunately, that was not possible due to the inaccessibility of the data. Reversing the problem, and estimating Frenette's version of the overeducation model is not a viable option either. A number of assumptions on which Frenette based his analysis would not hold with the more recent data used in this paper. For example, in his regression all three NGS cohorts and both genders were pooled together. Such aggregations, however, cannot be justified in the more recent data used here. Specification tests clearly prefer the separate-regressions approach. Yet another problem with Frenette's approach is his use of annual earnings rather than hourly wage as the dependent variable. Frenette claims that since his sample only contains full-time employees, the effect of differences in hours worked is small, which justifies the use of log earnings or the left-hand side of the regression. Although this may have been the case in the earlier NGS, it was not so in the cohorts used in this paper. There was still a substantial variation in hours worked even among those who reported working full-time. In summary, using Frenette's model with the 1995 and 2000 NGS would be incorrect. Needless to say, however, that a direct comparison of results based on the NGS classes prior to the 1995 with those from 1995 onwards would in any case had little meaning regardless of which model would be used. As mentioned in Section 2, the definition of individual earnings changed substantially between the 1990 and 1995 NGS which makes any comparison of earnings/wages across these periods virtually impossible.

5 SUMMARY AND CONCLUSIONS

This paper uses data from the 1995 and 2000 cohorts of the National Graduates Survey (NGS) to investigate the relative returns to post-secondary degrees and the wage implications of overeducation for young Canadian graduates. The NGS provides detailed information about respondents' higher education and labour market outcomes which allows to control for a variety of factors that are important for explaining variation in wages, but have mostly been ignored due to the lack of data. The wage regression applied here acknowledges the role of both the supply and demand side of the labour market. It allows to distinguish common market rewards for post-secondary degrees from premia for overeducation, and is shown to be superior to specifications based on the human-capital and job-competition models. Furthermore, the potential problem of a bias due to unobserved individual heterogeneity and sample selection is countered by the use of the Hausman-Taylor panel-data estimator. It is reasonable to expect that the selection mechanism for a given individual does not change over time and its effect is absorbed in the fixed-effect and removed through first differences. It is important, however, to keep in mind that the estimates of the coefficients that correspond to endogenous time-invariant variables may still be subject to a selection bias. Thus, the results presented in this study should be interpreted as pertaining to the population of young (at most 35 years of age) post-secondary graduates with a strong attachment to the labour market (those observed as having a job at both survey interviews). Although these individuals may not be fully representative of the population of post-secondary graduates as a whole, they still

represent a sizeable group, and their labour market outcomes warrant an interest on the side of researchers, policy makers and general public as well.

Given the estimates of market returns and overeducation premia across post-secondary degrees within each of the two NGS samples, as well as their comparisons across the genders, the most important results can be summarized as follows:

- Relative to jobs that did not require post-secondary education (the baseline), jobs that did require a PS degree paid more. Generally – with the exception of PhD-level jobs – the higher the degree required, the higher the wage. The highest wage return was in jobs that required a master's degree, ranging from 28.3% (men, class 1995) to 52.1% (women, class 2000) over the baseline.
- In all PSE categories, the relative returns were higher for the NGS 2000 class. This suggests that the gap between unskilled ("no PSE required") jobs and jobs that required higher education grew over time. The biggest "gainers" were PhD-level jobs which, in terms of wage return, appeared to be catching up with jobs that required only a master's degree.
- Overeducation premia for those who worked in jobs below their level of education differed substantially across the levels of schooling required by the jobs, between men and women, and over the two NGS cycles. In the NGS 1995, those who were overeducated in jobs that required either a college diploma or bachelor's degree received a positive and significant wage premium over individuals who worked the same jobs but were not overeducated. The same pattern appeared in the NGS 2000, but only for women; overeducated men in bachelor's-level jobs did not get any premium. With one exception – men, NGS class 2000 – graduates from doctoral

programs who worked in master's level jobs received no wage premium over those who worked in the same jobs and exactly matched the job-educational requirement.

- Compared to the "soft" fields of study, "business", "hard" fields and "health" all offered higher wages. The greatest relative return for men was in "hard" fields, for women it was in "health"; their wages were well in excess of 17% over the "soft" fields in both NGS cohorts.

The analysis in this paper suggests that the returns to post-secondary education were not, as some expected, diminishing over the period between the late 1990s and 2005. Wages in jobs that required a post-secondary degree appear to have increased both in dollar terms as well as relative to the wage paid by low-skilled jobs. Indeed, it is going to be interesting to see whether this conclusion still holds when the new NGS data are available.

The results also suggest that the quality of match between educational requirement of the job and individual's actual schooling had become increasingly more important. With some exceptions, and to a varying extent, having more education than required on the job did provide a wage premium over those who were perfectly matched in the same job, but the premium appears to have diminished over time.

Lastly, the analysis have confirmed that those who chose "soft" fields of study were worse off in terms of wages than graduates from any other field. Perhaps now is the time when both researchers and policy makers need to look deeper into the causes and implications of this phenomenon.

REFERENCES

Amemiya, T., & MaCurdy, T. (1986). Instrumental-Variable Estimation of an Error-Components model. *Econometrica,* 54(4), 869-890.

Bauer, T. K. (2002). Educational mismatch and wages: a panel analysis. *Economics of Education Review*, 2, 221-229.

Breusch, T. S., Mizon, G. E., & Schmidt, P. (1989). Efficient Estimation Using Panel Data. *Econometrica*, 57, 695-700.

Boothby, D. (2002). Literacy Skills, Occupational Assignment and the Returns to Over and Under-education. HRSDC and Statistics Canada Report.

Frenette, M. (2003). The Overqualified Canadian Graduate: The Role of Academic Program in the Incidence, Persistence, and Economic Returns to Overqualification. *Economics of Education Review,* 29-45.

Groot, W., & Van den Brink, H. M.. (2000). Overeducation in the Labour Market: A Meta-Analysis. *Economics of Education Review*.

Gunderson, M., & Krashinsky, H. (2008). Do Differences in Earnings and Other Labour Market Outcomes of Canadian Graduates Persist over Time? Human Resources and Skills Development Report.

Hansen, J. (2006). Returns to University Level Education: Variation within Disciplines, Occupations and Employment Sectors. Human Resources and Skills Development Canada Report (Ottawa).

Hansen, J., & Wahlberg, R. (2005). Endogenous Schooling and the Distribution of the Gender Wage Gap. *Empirical Economics*, 30, 1-22.

Hartog, J. (1986a). Earnings Functions: Beyond Human Capital. *Applied Economics*, 18, 1291-1309.

Hartog, J. (1986b). Allocation and Earnings Function. *Empirical Economics*, 11, 97-110.

Hartog, J. (2000). Over-Education and Earnings: Where Are We, Where Should We Go? *Economics of Education Review*.

Hausman, J. A., & Taylor, W. E. (1981). Panel Data and Unobservable Individual Effects. *Econometrica*, 49(6), 1377-1398.

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

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4), 281-302.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. NBER Press (New York).

Sicherman, N. (1991). Overeducation in the Labor Market. *Journal of Labor Economics*, 9(2), 101-122.

Statistics Canada. (2008). *Educational Portrait of Canada, 2006 Census*. Ottawa (2008), Statistics Canada Report.

Thurow, L. C. Generating Inequality. (1975). Basic Books (New York).

Thurow, L. C., & Lucas, R. E. B. (1972). *The American Distribution of Income: A Structural Problem*. US Government Printing Office (Washington).

Vahey, S. P. (2000). The Great Canadian Training Robbery: Evidence on the Returns to Educational Mismatch. *Economics of Education Review*, 19, 219-227.

Verbeek, M., & Nijman, T. (1992). Testing for Selectivity Bias in Panel Data Models. *International Economic Review*, 33(3), 681-703.

APPENDIX: TABLES

TABLE 9: AVERAGE HOURLY WAGE AND OVEREDUCATION PREMIUM BY JOB-EDUCATION REQUIREMENT, COHORT AND GENDER[a]

Education required by the job	Hourly wage 2 years after graduation		Hourly 5 years after graduation	
	Matched	Overeducated	Matched	Overeducated
Men, 1995 sample				
No PSE required	n.a.	14.24	n.a.	18.52
PSE below bachelor's	15.81	19.01	19.94	26.25
Bachelor's	18.55	25.07	25.08	30.96
Master's	23.23	26.63	29.03	32.02
PhD	20.20	n.a	26.51	n.a.
Women, 1995 sample				
No PSE required	n.a.	12.06	n.a.	14.66
PSE below bachelor's	14.26	18.52	16.76	21.11
Bachelor's	17.75	22.56	21.08	25.85
Master's	22.54	24.52	27.07	31.69
PhD	20.98	n.a.	25.99	n.a.
Men, 2000 sample				
No PSE required	n.a.	15.21	n.a.	17.76
PSE below bachelor's	17.73	23.04	20.43	25.84
Bachelor's	22.39	27.25	26.44	32.51
Master's	26.68	29.77	33.36	33.95
PhD	26.88	n.a.	30.80	n.a.
Women, 2000 sample				
No PSE required	n.a.	14.07	n.a.	15.26
PSE below bachelor's	15.22	20.88	16.54	22.55
Bachelor's	20.77	24.87	22.71	28.69
Master's	25.12	31.15	27.98	33.27
PhD	25.41	n.a.	30.85	n.a.

[a] Amounts in constant 2002 dollars.

TABLE 10: DISTRIBUTION OF JOB-EDUCATIONAL REQUIREMENTS AND OVEREDUCATION STATUS

	2 years after graduation		5 years after graduation	
	% of the sample (sum = 100)	% of overeducated within degree	% of the sample (sum = 100)	% of overeducated within degree
Men, 1995 sample				
No PSE required	19.3	100.0	22.4	100.0
PS below bachelor's	35.2	6.6	32.6	8.7
Bachelor's	31.6	30.3	30.4	29.0
Master's	7.5	15.8	8.6	17.4
PhD	6.4	0.0	6.0	0.0
Women, 1995 sample				
No PSE required	18.2	100.0	18.9	100.0
PS below bachelor's	36.1	18.9	34.0	10.2
Bachelor's	32.2	27.1	32.9	24.3
Master's	9.8	6.5	10.7	6.7
PhD	3.7	0.0	3.6	0.0
Men, 2000 sample				
No PSE required	15.1	100.0	15.5	100.0
PS below bachelor's	28.3	18.5	27.8	19.0
Bachelor's	44.6	28.5	43.8	26.4
Master's	7.0	23.5	7.6	22.7
PhD	4.9	0.0	5.4	0.0
Women, 2000 sample				
No PSE required	16.0	100.0	16.9	100.0
PS below bachelor's	32.9	19.6	32.4	21.1
Bachelor's	42.3	25.6	41.2	24.3
Master's	6.6	7.9	7.1	9.7
PhD	2.2	0.0	2.4	0.0

TABLE 11: DISTRIBUTION OF FIELDS OF STUDIES AND OVEREDUCATION STATUS

	% of the sample (sum = 100)	% of overeducated within field 2 years after graduation	5 years after graduation
Men, 1995 sample			
Soft fields	14.1	36.0	36.3
Business	12.7	44.9	39.6
Hard fields	59.4	29.4	34.0
Health	4.5	23.6	21.4
Other	9.3	33.0	46.0
Women, 1995 sample			
Soft fields	28.6	32.6	31.6
Business	25.6	38.9	38.9
Hard fields	19.0	29.2	29.6
Health	22.0	20.0	20.0
Other	4.8	41.2	41.2
Men, 2000 sample			
Soft fields	17.7	40.0	39.3
Business	11.6	50.5	51.5
Hard fields	57.3	28.7	27.9
Health	6.0	29.3	29.3
Other	7.4	49.2	46.1
Women, 2000 sample			
Soft fields	33.4	32.7	33.3
Business	20.7	40.7	40.6
Hard fields	18.2	32.0	32.2
Health	23.8	29.5	31.0
Other	4.0	39.7	42.2

Soft fields comprise of Education, Arts, Humanities and Social and Behavioural Sciences. Hard fields group Physical and Life Sciences, Mathematics, Computer and Information Sciences, Architecture and Engineering.

TABLE 12: TRANSITIONS INTO AND OUT OF OVEREDUCATION

Men, 1995 sample		Overeducated 5 years after graduation: No	Overeducated 5 years after graduation: Yes	Total
Overeducated 2 years after graduation	No	53.4 (79.0)	14.2 (21.0)	67.6 (100)
	Yes	11.0 (34.0)	21.4 (66.0)	32.4 (100)
Total		64.4	35.6	100

Women, 1995 sample		Overeducated 5 years after graduation: No	Overeducated 5 years after graduation: Yes	Total
Overeducated 2 years after graduation	No	56.4 (82.1)	12.3 (17.9)	68.7 (100)
	Yes	12.6 (40.2)	18.7 (59.8)	31.3 (100)
Total		69.0	31.0	100

Men, 2000 sample		Overeducated 5 years after graduation: No	Overeducated 5 years after graduation: Yes	Total
Overeducated 2 years after graduation	No	61.4 (94.1)	3.8 (5.9)	65.2 (100)
	Yes	4.5 (13.0)	30.3 (87.0)	34.8 (100)
Total		66.0	34.0	100

Women, 2000 sample		Overeducated 5 years after graduation: No	Overeducated 5 years after graduation: Yes	Total
Overeducated 2 years after graduation	No	62.1 (93.8)	4.1 (6.2)	66.2 (100)
	Yes	3.5 (10.2)	30.3 (89.8)	33.8 (100)
Total		65.6	34.3	100

Numbers in the parentheses are row percentages.

TABLE 13: HAUSMAN-TAYLOR REGRESSION ESTIMATES[a]

	Men, 1995 sample		Men, 2000 sample		Women, 1995 sample		Women, 2000 sample	
Job requirement[b]								
College diploma (below Bachelor's)	0.014	(0.011)	0.126*	(0.019)	0.059*	(0.012)	0.066*	(0.015)
Bachelor's and above (below Master's)	0.195*	(0.017)	0.355*	(0.022)	0.226*	(0.016)	0.340*	(0.017)
Master's and above (below PhD)	0.283*	(0.025)	0.398*	(0.034)	0.436*	(0.022)	0.521*	(0.024)
PhD	0.150*	(0.026)	0.367*	(0.031)	0.329*	(0.035)	0.484*	(0.033)
Overeducated at job requirement								
College diploma (below Bachelor's)	0.231*	(0.031)	0.169*	(0.049)	0.160*	(0.030)	0.175*	(0.040)
Bachelor's and above (below Master's)	0.081*	(0.032)	-0.008	(0.046)	0.244*	(0.036)	0.164*	(0.042)
Master's and above (below PhD)	0.001	(0.063)	0.242*	(0.102)	-0.015	(0.077)	0.039	(0.102)
Field of study[c]								
Business	0.134*	(0.022)	0.181*	(0.024)	0.014	(0.016)	0.070*	(0.015)
Hard sciences	0.170*	(0.017)	0.180*	(0.017)	0.071*	(0.018)	0.090*	(0.016)
Health	0.104*	(0.031)	0.145*	(0.031)	0.172*	(0.016)	0.195*	(0.014)
Other	0.026	(0.024)	0.085*	(0.027)	-0.029	(0.027)	-0.037	(0.026)
Experience	0.112*	(0.006)	0.054*	(0.006)	0.098*	(0.005)	0.040*	(0.005)
Experience^2	-0.002*	(0.000)	-0.001	(0.000)	-0.003*	(0.000)	-0.001*	(0.000)
Dissatisfied with the job	-0.085*	(0.010)	-0.090*	(0.011)	-0.087*	(0.010)	-0.094*	(0.009)
Permanent job	0.072*	(0.013)	0.172*	(0.016)	0.014	(0.011)	0.054*	(0.011)
Mother's education[d]								
Some PSE (below bachelor's)	0.065*	(0.017)	0.042	(0.022)	0.015	(0.017)	0.022	(0.016)
Bachelor's and above (below master's)	0.073*	(0.019)	0.047*	(0.023)	0.026	(0.018)	0.039*	(0.017)
Master's and above (below PhD)	0.095*	(0.024)	0.059*	(0.027)	0.044*	(0.024)	0.068*	(0.022)
PhD	0.137*	(0.039)	0.019	(0.039)	0.027	(0.038)	0.104*	(0.033)

Table 13 continues on the next page.

Table 13 continues from the previous page.

	Men, 1995 sample		Men, 2000 sample		Women, 1995 sample		Women, 2000 sample	
Father's education[d]								
Some PSE (below bachelor's)	0.037*	(0.017)	0.033	(0.021)	0.055*	(0.016)	0.030	(0.016)
Bachelor's and above (below master's)	0.036*	(0.018)	0.030	(0.021)	0.063*	(0.018)	0.024	(0.016)
Master's and above (below PhD)	0.054*	(0.022)	0.043	(0.024)	0.077*	(0.022)	0.043*	(0.019)
PhD	0.055*	(0.025)	0.091*	(0.028)	0.101*	(0.026)	0.063*	(0.023)
Married	-0.002	(0.010)	0.033	(0.011)	0.022*	(0.009)	0.023*	(0.008)
Have dependent children	-0.028*	(0.012)	0.030	(0.013)	-0.061*	(0.012)	-0.026*	(0.010)
Region of residence[e]								
West	-0.043*	(0.014)	-0.113	(0.018)	-0.020	(0.015)	-0.089*	(0.015)
East	-0.192*	(0.025)	-0.301	(0.028)	-0.175*	(0.024)	-0.262*	(0.022)
Quebec	-0.047*	(0.020)	-0.108	(0.021)	-0.029	(0.020)	-0.090*	(0.017)
Local unemployment rate	0.004	(0.003)	0.005	(0.004)	0.008*	(0.003)	0.008*	(0.003)

[a] Standard errors in parentheses, * denotes statistical significance at the 5% level.

[b] Reference category: jobs that did not require a post-secondary degree

[c] Reference category: "soft" fields (Education, Arts, Humanities and Social and Behavioural Sciences)

[d] Reference category: schooling below post-secondary

[e] Reference category: Ontario

TABLE 14: COMPARISON OF RELATIVE WAGE AND OVEREDUCATION PREMIA

	Men		Women	
	1995 sample	2000 sample	1995 sample	2000 sample
θ_2	0.014	0.126*	0.059*	0.066*
$\theta_2 + \varphi_2$	0.245	0.295*	0.219*	0.241*
θ_3	0.195*	0.355*	0.226*	0.340*
$\theta_3 + \varphi_3$	0.275*	0.347*	0.470*	0.503*
θ_4	0.283*	0.398*	0.436*	0.521*
$\theta_4 + \varphi_4$	0.283*	0.640*	0.421*	0.559*
θ_5	0.150*	0.367*	0.329*	0.484*
$\theta_3 - (\theta_2 + \varphi_2)$	-0.050*	0.060*	0.007*	0.098*
$\theta_4 - (\theta_3 + \varphi_3)$	0.007*	0.052*	-0.034*	0.017*
$\theta_5 - (\theta_4 + \varphi_4)$	-0.134*	-0.273*	-0.093*	-0.075*

The table was generated using estimates in Table 13. Asterisk * distinguishes estimates that are statistically significant at the 5% level; θ_j is the (estimated) return to schooling in a job that required level j education, and φ_j is the corresponding overeducation premium. The levels of education are: college (2), bachelor's (3), master's (4), and PhD (5). The differences $\theta_j - (\theta_{j-1} + \varphi_{j-1})$ demonstrate how much higher/lower were the wages of those who were perfectly matched in terms of schooling with their job compared to those who had the same education but worked in jobs that required less schooling.

TABLE 15: CHANGES OVER NGS CYCLES IN RELATIVE WAGE AND OVEREDUCATION PREMIA

	1995 sample	2000 sample	Change 2000-1995
Men			
Job requirement			
College diploma (below Bachelor's)	0.014	0.126*	0.112*
Bachelor's and above (below Master's)	0.195*	0.355*	0.160*
Master's and above (below PhD)	0.283*	0.398*	0.116*
PhD	0.150*	0.367*	0.218*
Overeducated at job requirement			
College diploma (below Bachelor's)	0.231*	0.169*	-0.062*
Bachelor's and above (below Master's)	0.081*	-0.008	-0.089*
Master's and above (below PhD)	0.001	0.242*	0.241*
Women			
Job requirement			
College diploma (below Bachelor's)	0.059*	0.066*	0.007*
Bachelor's and above (below Master's)	0.226*	0.340*	0.113*
Master's and above (below PhD)	0.436*	0.521*	0.084*
PhD	0.329*	0.484*	0.156*
Overeducated at job requirement			
College diploma (below Bachelor's)	0.160*	0.175*	0.015*
Bachelor's and above (below Master's)	0.244*	0.164*	-0.080*
Master's and above (below PhD)	-0.015	0.039	0.054

TABLE 16 COMPARISON OF SELECTED OLS AND HAUSMAN-TAYLOR REGRESSION ESTIMATES: NGS 1995[a]

	Men 1995				Women 1995			
	OLS		HT		OLS		HT	
Job requirement[b]								
College diploma (below Bachelor's)	0.049*	(0.011)	0.014	(0.011)	0.068*	(0.011)	0.059*	(0.012)
Bachelor's and above (below Master's)	0.209*	(0.014)	0.195*	(0.017)	0.294*	(0.014)	0.226*	(0.016)
Master's and above (below PhD)	0.372*	(0.021)	0.283*	(0.025)	0.459*	(0.020)	0.436*	(0.022)
PhD	0.214*	(0.024)	0.150*	(0.026)	0.379*	(0.033)	0.329*	(0.035)
Overeducated at job requirement								
College diploma (below Bachelor's)	0.194*	(0.027)	0.231*	(0.031)	0.233*	(0.023)	0.160*	(0.030)
Bachelor's and above (below Master's)	0.193*	(0.019)	0.081*	(0.032)	0.142*	(0.021)	0.244*	(0.036)
Master's and above (below PhD)	0.005	(0.044)	0.001	(0.063)	0.168*	(0.059)	-0.015	(0.077)
Field of study[c]								
Business	0.136*	(0.020)	0.134*	(0.022)	0.031*	(0.015)	0.014	(0.016)
Hard sciences	1.152*	(0.015)	0.170*	(0.017)	0.088*	(0.016)	0.071*	(0.018)
Health	0.102*	(0.028)	0.104*	(0.031)	0.182*	(0.015)	0.172*	(0.016)
Other	0.020	(0.021)	0.026	(0.024)	-0.020	(0.024)	-0.029	(0.027)
Experience	0.072*	(0.005)	0.112*	(0.006)	0.083*	(0.005)	0.098*	(0.005)
Experience^2	-0.003*	(0.000)	-0.002*	(0.000)	-0.003*	(0.000)	-0.003*	(0.000)

[a] Standard errors in parentheses, * denotes statistical significance at the 5% level. Note that the regressions in this table have exactly the same regressors as the regressions reported in table 5; only selected parameters are reported for spatial concerns (the complete set can be obtained upon request).

[b] Reference category: jobs that did not require a post-secondary degree

[c] Reference category: soft fields (Education, Arts, Humanities and Social and Behavioural Sciences)

TABLE 17 COMPARISON OF SELECTED OLS AND HAUSMAN-TAYLOR REGRESSION ESTIMATES: NGS 2000[a]

	Men 2000				Women 2000			
	OLS		HT		OLS		HT	
Job requirementb								
College diploma (below Bachelor's)	0.140*	(0.016)	0.126*	(0.019)	0.034*	(0.013)	0.066*	(0.015)
Bachelor's and above (below Master's)	0.270*	(0.016)	0.355*	(0.022)	0.361*	(0.013)	0.340*	(0.017)
Master's and above (below PhD)	0.567*	(0.026)	0.398*	(0.034)	0.500*	(0.021)	0.521*	(0.024)
PhD	0.433*	(0.029)	0.367*	(0.031)	0.512*	(0.031)	0.484*	(0.033)
Overeducated at job requirement								
College diploma (below Bachelor's)	0.153*	(0.027)	0.169*	(0.049)	0.355*	(0.020)	0.175*	(0.040)
Bachelor's and above (below Master's)	0.348*	(0.020)	-0.008	(0.046)	0.166*	(0.017)	0.164*	(0.042)
Master's and above (below PhD)	-0.146*	(0.049)	0.242*	(0.102)	0.388*	(0.058)	0.039	(0.102)
Field of studyc								
Business	0.140*	(0.022)	0.181*	(0.024)	0.082*	(0.014)	0.070*	(0.015)
Hard sciences	0.164*	(0.016)	0.180*	(0.017)	0.093*	(0.014)	0.090*	(0.016)
Health	0.157*	(0.028)	0.145*	(0.031)	0.194*	(0.013)	0.195*	(0.014)
Other	0.051*	(0.024)	0.085*	(0.027)	-0.021	(0.024)	-0.037	(0.026)
Experience	0.056*	(0.006)	0.054*	(0.006)	0.048*	(0.004)	0.040*	(0.005)
Experience^2	-0.003*	(0.000)	-0.001	(0.000)	-0.001*	(0.000)	-0.001*	(0.000)

[a] Standard errors in parentheses, * denotes statistical significance at the 5% level. Note that the regressions in this table have exactly the same regressors as the regressions reported in table 5; only selected parameters are reported for spatial concerns (the complete set can be obtained upon request).

[b] Reference category: jobs that did not require a post-secondary degree

[c] Reference category: soft fields (Education, Arts, Humanities and Social and Behavioural Sciences)

ESSAY III

DISPARITIES IN SCHOOLING CHOICES AND WAGES BETWEEN ETHNIC MINORITIES AND WHITES: EVIDENCE FROM THE NLSY97

This essay has been produced in collaboration with Jorgen Hansen and Xingfei Liu.

ABSTRACT

This study uses data for a sample of males from the 1997 cohort of the National Longitudinal Survey of Youth to estimate a structural dynamic model of schooling attainment and wages, and to understand sources of disparities between the white majority, and the ethnic minorities of blacks and Hispanics in the United States. We find that whites have a higher return to schooling than blacks but lower than Hispanics. Of all three ethnic groups, whites have the lowest return to work experience. By simulating schooling and wages under various scenarios, we are able to assess the relative effects of behavioural and endowment differences in explaining ethnic gaps in outcomes. Our decompositions show that ethnic differentials in schooling attainments can be largely explained by differences in endowments across the ethnics. Behavioural differences (differences in parameters) appear to play a larger role in explaining ethnic differences in wages.

1 INTRODUCTION

Disparities in educational and labour market outcomes between various ethnic groups in the United States have been an active area of research, and the existence of such gaps especially between whites, blacks and Hispanics has been extensively documented.[27] Ethnic minorities have been shown to be, on average, less educated and to earn less than their white counterparts. The literature suggests that differences in schooling between ethnic groups can mainly be explained by parental education, family environment and individual abilities rather than by credit constraints. With regard to wages, the central question has been and remains whether the observed variation across the ethnic groups are due to unequal market prices of skills and experience (wage discrimination), or due to differences in the distributions of education and abilities.

The social sciences literature abounds with studies of ethnic gaps in earnings, education and many other dimensions. Unfortunately, many of them suffer from various data and methodological problems and often offer very limited insight at best. In this paper, we recognize the need to respect the structure and dynamic nature of the process that is in the base of individuals' schooling decisions and their labour market success. Our point of departure in the literature is Cameron and Heckman (2001). In their paper,

---

[27] According to the Longman Dictionary of Contemporary English, an *ethnic* is someone who comes from a group of people who are of a different race or religion, or who have a different background from most other people in the country. As the term *ethnic* is more general and clearly encompasses *race*, we choose to refer to both blacks and Hispanics as ethnic rather than racial groups.

Cameron and Heckman used a dynamic model of schooling attainment to investigate racial and ethnic disparity in schooling (focusing primarily on college attendance). Contrary to then conventional thinking, they found that parental background and family environment were more important in explaining ethnic differences in schooling then family credit constraints. Besides its importance for policy making, the paper is also an important contribution to empirical economics for its recognition of the dynamics of the schooling attainment process, and attention it paid to important issues such as the effects of unobserved abilities.

Racial differences in schooling and earnings were also subject of a paper by Keane and Wolpin (2000) who estimated a structural dynamic model of school attendance, work and occupational choice, and tested implications of two policy proposals that were expected to have a differential racial impacts: a high-school graduation bonus for students from low-income families, and a wage subsidy to low-wage workers. Although they recognized the potential effect these schemes may have on the size of the schooling and earnings gaps, they emphasized that equalizing endowments that individuals have when they begin making independent decisions about their future (age 16 in Keane and Wolpin's model) would by itself go a long way toward eliminating ethnic differences in labour market success.

Recent contributions to the literature on ethnic differences, namely Carneiro, Heckman and Masterov (2005), bring into focus the role of cognitive skills and discrimination in explaining ethnic wage gaps. Urzua (2008) extends this focus further by adding non-cognitive skills, and the distinction between measured and unobserved skills,

to his study of black-white differences in schooling choices and labour market outcomes. He finds that even after controlling for differences in abilities, significant labour market gaps still exist.

While a great deal of existing research focused on individuals who were completing their schooling and entering the labour market in the late 1970s and early 1980s, much less is known about their successors. Although the younger 1997 cohort of the National Longitudinal Survey of Youth (NLSY97) has been available for some time, its use by researchers has so far been rather sporadic. To our knowledge, only Altonji, Bharadwaj and Lange (2008) have made a more extensive use of the NLSY97 when comparing it to the older 1979 cohort. To our knowledge, no study has yet used the NLSY97 to analyze the existence and sources of ethnic disparities in schooling and earnings.

Our paper addresses this issue by analyzing a sample of men from the 1997 cohort of the NLSY. We propose a structural dynamic model of schooling choice and wages, and use it to estimate various parameters of interest, and to analyze sources of ethnic gaps. We find, among other things, that whites, blacks and Hispanics all face different returns to schooling (lowest for blacks) and work experience (lowest for whites). Furthermore, we find that ethnic differentials in schooling attainments can largely be explained by differences in endowments across the ethnics, and that behavioural differences (differences in parameters) play a prominent role in explaining ethnic differences in wages.

The paper is organized as follows: Section 2 provides description and summary statistics for our sample. Section 3 introduces the structural dynamic model and outlines our estimation strategy. Section 4 presents the key estimation results and decompositions of the ethnic gaps in schooling and wages. Finally, Section 5 concludes the paper. Note that all tables referred to in the text can be found in Appendix A and figures in Appendix B.

2 DATA AND SAMPLES

In this paper we utilize data from the 1997 National Longitudinal Survey of Youth (NLSY97). Unlike the 1979 NLSY which has long been a major source of information about the transition of young Americans into the labour market, the use of the 1997 cohort has until recently been limited by the young age of the respondents and insufficient observations of their labour market experiences and outcomes. As the NLSY97 consists of youths who were 12 to 16 years old at the end of 1996, a meaningful analysis of school to work transitions and labour market outcomes is only now becoming feasible. By 2007 – the latest data release available to us – a majority of the surveyed individuals had aged enough to have completed their schooling (about 85% of our sample) and entered the workforce. Nevertheless, it is important to keep in mind that we are still observing only early stages of these individuals' work histories.

In this study, we use data for 3,578 males from 1997 to 2007 cycles of the NLSY97. Whites represent 52.7 percent of our sample, blacks 27.1 and Hispanics 20.2 percent. For each of the three ethnic groups, we have a sufficiently large number of

observations to warrant a separate analysis and avoid pooling the three ethnics together. Table 1, and figures 1 to 6 summarize all the major characteristics of our samples, as well as the differences in schooling attainments and wages across the ethnic groups.

First thing to notice in table 1 are differences in family backgrounds between the white majority and the minorities. Parents of whites are on average more educated than those of blacks, and Hispanic parents have substantially lower education than both white and black parents. Parental income is comparable for blacks and Hispanics, but substantially lower than income of white parents. Furthermore, black and Hispanic families have more children than white families. A very large difference between whites and Hispanics on the one side and blacks on the other is in the family conditions in which they were raised in their formative years. While over 60 percent of whites and 55 percent of Hispanics lived in complete families (with both biological parents) until their mid-teens, a full 73 percent of blacks grew up with only one biological parent.

There are also differences in schooling between whites and the minorities. Although at age 16 (the starting point in our model) all three ethnics have, on average, about 10 years of schooling, the average final educational attainment of whites is more than one year higher than those of blacks and Hispanics. Close to 28 percent of whites complete 16 or more years of schooling, while only 9.5 percent of blacks and 12.3 percent of Hispanics do so (figure 2). In fact, the schooling distribution for whites appears almost bimodal with spikes at grades 12 and 16, while the schooling distributions of blacks and Hispanics are unimodal, peaking at grade 12. Figure 3 also confirms that the schooling patterns of the two ethnic minorities are similar to each other but very

different from that of the white majority. For example, whites have almost 60 percent probability of completing and continuing past grade 12. In contrast, blacks and Hispanics are only about 40 percent likely to do so.

In order to assess ethnic disparities in abilities, we created a composite index as an average of six specific-ability test scores from the Armed Services Vocational Aptitude Battery (ASVAB).[28] This set of tests was administered from the summer of 1997 through the spring of 1998 to the NLSY97 respondents who were of varying ages and schooling. To eliminate the effect of these differences on the test results, we use residuals from the regression of our composite ability score on the highest grade completed at the time when the tests were taken. The non-parametric estimates of the distributions of the composite ability score in figure 4 suggest that blacks and Hispanics have similar bell-shaped ability distributions that are centered close to the zero mark. With respect to the minority distributions, the white ability distribution appears to be shifted to the right. Both the mean and the median scores for whites are about one point greater than the corresponding statistics for blacks and Hispanics. The white ability distribution also exhibits slightly larger variability than those of the minorities.

Figures 5 and 6 show that differences in wages between whites and the minorities are substantial, although more pronounced for blacks than Hispanics. Average hourly

---

[28] Similar constructs are used in Heckman, Stixrud and Urzua (2006), and Urzua (2008). The six scores averaged in our ability index are for arithmetic reasoning, mathematics knowledge, paragraph comprehension, word knowledge, coding speed, and numerical operations. Details about the ASVAB tests and their administration can be found on the NLSY97 website (http://www.bls.gov/nls/nlsy97.htm).

wage of blacks starts to diverge from the wages earned by whites rather early, and by their mid-twenties blacks earn on average about 20 percent less than whites. By the same age, Hispanics also earn less than whites, about 15 percent, but their age-wage profile is similar to that of whites for longer than in the case of blacks. In fact, only the last two averages, corresponding to ages 25 and 26, are substantially lower than those of whites. Indeed, given the young age of the NLSY97 respondents at the time of our last observation in 2007, it would be premature to draw any conclusions as to whether the observed divergence of minority wages from the wages of whites will continue, stabilize or diminish.

3 STRUCTURAL MODEL OF SCHOOLING AND WAGES

This section introduces our empirical model which is similar to those in Belzil and Hansen (2002, 2007). Individuals in our model decide sequentially whether it is optimal to enter the labour market or continue to accumulate schooling. They are assumed to be rational, forward-looking, and to maximize their discounted expected lifetime utility over a finite horizon which is set to the age of 65 (retirement age). For each period an individual decides whether to invest into an additional year of schooling, or terminate his education and enter the labour market permanently. That is, once an individual makes his decision to leave school and start working, he continues to work

until retirement.[29] Note that the model as described below applies to all three ethnic groups – whites, blacks and Hispanics – and is estimated separately for every one of them.

3.1 Utility of Schooling

We assume that when in school in period t, an individual i receives an instantaneous monetary return at the amount of ξ_{it}.[30] We further assume that the utility of being in school is logarithmic in this income and depends on selected time-invariant background characteristics in vector $\mathbf{x}_i^{\xi}$, individual's initial schooling grade at age 16, s_0, his abilities both observed, a_i (ability test score), and unobserved, ν_i^{ξ}, and on a random shock, ε_{it}^{ξ}.

$$\mathrm{U}_{it}^{\xi} = \log(\xi_{it}) = \mathbf{x}_i^{\xi\prime}\boldsymbol{\theta}^{\xi} + \gamma_0^{\xi} \cdot s_{0i} + \sum_{j=s_{0i}+1}^{20} \delta_j^{\xi} \cdot \mathrm{I}(s_{it} = j) + \alpha^{\xi} \cdot a_i + \nu_i^{\xi} + \varepsilon_{it}^{\xi}. \qquad (5)$$

This specification is flexible enough to control for differences in a range of individual endowments, family environment variables and initial schooling. It also allows for the

---

[29] A simplifying assumption such as this is dictated by the data availability. Even at the latest survey available to us, the NLSY97 respondents were still quite young. Once data permit, the model could be extended, for example, to incorporate unemployment or re-entry to schooling.

[30] Expressing the utility of schooling in monetary terms makes it comparable to the utility of working (wage) later on. Although we cannot explicitly incorporate parental transfers or costs of schooling into our model due to the lack of data, it is possible to think of schooling utility as an equivalent of income support received from familial and/or other sources to cover the cost associated with continuing schooling in a given period.

possibility that the costs of schooling vary over schooling grades by including the grade-indicator function, $\mathrm{I}(s_{it} = j)$, which takes on value of 1 if an individual completes grade j, and 0 otherwise. This makes our model more realistic and better aligned with empirical facts.

3.2 Utility of Working

We assume that hourly wage, w_{it}, depends on an individual's initial schooling grade, s_{i0}, his educational attainment, s_{it}, years of work experience, z_{it}, and his abilities both observed, a_i, and unobserved, ν_i^w, and on a random shock, ε_{it}^w. The instantaneous utility of working in period t is logarithmic in wage:[31]

$$\mathrm{U}_{it}^w = \log(w_{it}) = \gamma_0^w \cdot s_{i0} + \alpha^w \cdot a_i + \varphi^w \cdot (s_{it} - s_{i0}) + \kappa_1 \cdot z_{it} + \nu_i^w + \varepsilon_{it}^w . \qquad (6)$$

More developed specifications of the wage equation are indeed possible. One extension, for example, would be to allow the returns to schooling to vary across individuals and schooling grades. Although we have forgone this distinction to keep the computational costs of estimating the model manageable, our more parsimonious design is sufficiently realistic for the purpose of this study and fits the data rather well.

[31] Our specification omits the usual quadratic work experience term. The age-wage profiles of young individuals in the NLSY97 are still very much linear and do not yet exhibit the familiar concave shape.

3.3 Schooling Interruptions

In order to conform more closely to the empirical facts, we also allow for schooling interruptions. For simplicity, we incorporate them as a state which occurs with an exogenous probability π_t, and is captured by a binary indicator variable I_{it}. If an interruption happens in a given period ($I_{it}=1$), the decision problem is frozen and the stock of schooling remains constant over that period until the beginning of the next one. Due to the lack of data on parental transfers in the NLSY97, we do not distinguish monetary payoff the individual receives when in school from payoff when school is interrupted.[32]

3.4 Initial Schooling Model

It is plausible that the permanent personal endowments that help explain schooling decisions are also instrumental in determining how much schooling one acquires by age 16 when individuals start to decide whether to continue in school or enter the labour market. A failure to account for this possibility could seriously bias the estimates of the structural parameters. Consequently, we choose to model initial schooling as an ordered-choice, and let the initial-schooling grade probabilities depend on a vector of observable individual characteristics as well as on unobserved abilities.

---

[32] In the absence of information about the reasons for and the activities during schooling interruptions, the interruption state in our model can be thought of as encompassing a variety of events such as illness or injury, travel, temporary work or academic failure.

3.5 Unobserved Abilities

The intercept terms of the utility of attending school, ν_i^{ξ}, the log-wage function ν_i^{w}, and in the initial-schooling latent regression, ν_i^{0}, are individual specific. As is customary in this type of models, we model unobserved heterogeneity as a finite mixture. We assume that there are K types of individuals, and express the probability of belonging to type k as a logistic transform

$$p_k = \frac{e^{q_k}}{\sum_{j=1}^{K} e^{q_j}},\ k = 1, 2, \ldots, K,$$

with the restriction that $q_K = 0$. For the purpose of this study, we consider $K = 5$ distinct types of individuals.

3.6 Value Functions

To simplify the notation, let $\mathbf{s}_{it}$ denote the vector of all (predetermined and random) state variables, and d_{it} be the control variable. The choice in this case is simple: an individual either invests into additional year of education ($d_{it} = 1$), or terminates his schooling and enters the labour market ($d_{it} = 0$). The choice of entering the labour market is assumed to be permanent. That is $d_{it} = 0$ implies $d_{ij} = 0$ for all periods $j = t+1, ..., T$. The terminal period T is set to age 65 at which the individuals retire.

The net present value of the decision to remain in school, given state variables, can be expressed by the Bellman equation

$$\begin{aligned} \mathrm{V}_{it}^{\xi}(\mathbf{s}_{it}) = \mathrm{U}_{it}^{\xi} + \beta \cdot \Big\{ & \pi_{t+1} \cdot \mathrm{E}\Big[\mathrm{V}_{it+1}^{I}(\mathbf{s}_{it+1}) \Big] \\ & + (1-\pi_{t+1}) \cdot \mathrm{EMax}\Big[\mathrm{V}_{it+1}^{\xi}(\mathbf{s}_{it+1}), \mathrm{V}_{it+1}^{w}(\mathbf{s}_{it+1}) \Big] \end{aligned} \tag{7}$$

where β is the discount rate, π_{t+1} is the probability that a schooling interruption will occur in the next period, and $\mathrm{V}_{it+1}^{I}(\mathbf{s}_{it+1})$ denotes the value an invidual receives when he is in the state of interruption in period $t+1$. Our data do not allow us to distinguish between income the individual receives while in school and income when school is interrupted, instead we assume that the value of schooling interruption is identical to the value of attending school.

The value of terminating schooling and entering the labour market is given by

$$\mathrm{V}_{it}^{w}(\mathbf{s}_{it}) = \mathrm{U}_{it}^{w} + \beta \cdot \mathrm{E}\big[\mathrm{V}_{it}(\mathbf{s}_{it+1}) \,|\, d_{it} = 0 \big]. \tag{8}$$

in which the second term on the right-hand side is simply the discounted expected value of working from $t+1$ until retirement

$$\begin{aligned} & \mathrm{E}\big[\mathrm{V}_{it+1}(\mathbf{s}_{it+1}) \,|\, d_{it} = 0 \big] \\ & \qquad = \sum_{j=t+1}^{T} \beta^{j-(t+1)} \Big[\Pr(w_{ij} > 0) \cdot \mathrm{E}(\mathrm{U}_{ij}^{w} \,|\, w_{ij} > 0) + \Pr(w_{ij} \le 0) \cdot 0 \Big] \end{aligned}$$

Finally, an individual chooses to terminate schooling and enter the labour market permanently if

$$\mathrm{V}_{it}^{w}(\mathbf{s}_{it}) \ge \mathrm{V}_{it}^{\xi}(\mathbf{s}_{it}). \tag{9}$$

3.7 Likelihood Function

The dynamic programming problem is solved using backward recursion, and the parameters of the model are estimated by maximum likelihood. For the estimation, additional assumptions are needed about the distributions of the random shocks in the schooling utility and wage equations: $\varepsilon_{it}^{\xi} \sim \text{iid}\,\text{N}(0,\sigma_{\xi}^2)$ and $\varepsilon_{it}^{w} \sim \text{iid}\,\text{N}(0,\sigma_{w}^2)$. If no interruption occurs in period t (that is, $I_{it}=0$), the probability of leaving school in that period is

$$\begin{aligned}\Pr(d_{it}=0) &= \Pr\left[\text{V}_{it}^{w}(\mathbf{s}_{it}) \geq \text{V}_{it}^{\xi}(\mathbf{s}_{it})\right] \\ &= \Pr(\varepsilon_{it}^{w} - \varepsilon_{it}^{\xi} \geq R_{it}^{*})\end{aligned} \tag{10}$$

where R_{it}^{*} is an element in the sequence of reservation values that can be derived by combining equations (7), (8) and (9) into

$$\begin{aligned}R_{it}^{*} = (\text{U}_{it}^{\xi} - \varepsilon_{it}^{\xi}) + \beta \cdot \Big\{ & \pi_{t+1} \cdot \text{E}\left[\text{V}_{it+1}^{I}(\mathbf{s}_{it+1})\right] \\ & + (1-\pi_{t+1}) \cdot \text{EMax}\left[\text{V}_{it+1}^{\xi}(\mathbf{s}_{it+1}), \text{V}_{it+1}^{w}(\mathbf{s}_{it+1})\right]\Big\} \\ & - \Big\{(\text{U}_{it}^{w} - \varepsilon_{it}^{w}) + \beta \cdot \text{E}\left[\text{V}_{it+1}(\mathbf{s}_{it+1}) \mid d_{it}=0\right]\Big\}\end{aligned} \tag{11}$$

The probability of remaining in school for another year is given by

$$\Pr(d_{it}=1) = \Pr(\varepsilon_{it}^{w} - \varepsilon_{it}^{\xi} < R_{it}^{*}). \tag{12}$$

Given these inputs, the likelihood function, conditional on unobserved components, can be constructed as a composite of:

- The probability of observing a particular sequence of schooling/interruption histories:

$$L_{1i}(k) = \Pr\left\{\left[d_{i0}(k), I_{i0}(k)\right], \left[d_{i1}(k), I_{i1}(k)\right], ..., \left[d_{i\tau}(k), I_{i\tau}(k)\right]\right\}.$$

- The probability of entering the labour market in period $\tau + 1$, at observed wage $w_{i\tau+1}$ which can be expressed as a product of normal conditional probability and a marginal wage density $f(\cdot)$:

$$\begin{aligned} L_{2i}(k) &= \Pr[d_{i\tau+1}(k) = 0, w_{i\tau+1}(k)] \\ &= \Pr[d_{i\tau+1}(k) = 0 \mid w_{i\tau+1}(k)] \cdot f(w_{i\tau+1}(k)) \end{aligned}$$

- The joint denstities of observed wages from $\tau + 2$ until the last observed period T_i:

$$L_{3i}(k) = f\left[w_{i\tau+2}(k), ..., w_{iT_i}(k)\right],$$

which can be expressed as a product of marginal densities, conditional on the unobserved heterogeity components.

The complete individual contribution to the likelihood is

$$L_i = \sum_{k=1}^{4} p_k \cdot L_{1i}(k) \cdot L_{2i}(k) \cdot L_{3i}(k), \tag{13}$$

where p_k is the probability of type k. The parameters of the model are estimated by maximizing the sum of the log of the individual likelihoods over all i.

4 KEY ESTIMATION RESULTS

We estimated the structural model separately for whites, blacks and Hispanics. This section summarizes the estimates as presented in tables 2, 3 and 4. Throughout it, we will point out differences in the parameter estimates across the ethnics, and provide a more detailed investigation into the sources of ethnic disparities in schooling and wages.

4.1 Family Background and Individual Abilities in the Utility of Schooling

As can be seen in table 2, leaving aside for the moment the type-specific intercepts and effects of initial schooling, only two covariates appear to be uniformly significant for all three ethnics. One is the effect of observed scholastic ability as measured by the composite ability test score; it is positive and of similar magnitude for whites and blacks (0.034 and 0.029, respectively) while stronger for Hispanics (0.055). This is not unexpected. Presumably, individuals who exhibit higher scholastic ability have lower psychical costs of schooling which would be reflected in higher utility of attending school. The other effect that is significant across all three ethnics is the effect of being raised by both biological parents (variable *nuclear*). It is positive and substantial, especially for Hispanics (0.209) and whites (0.168), and somewhat weaker for blacks (0.066). Furthermore, for all three ethnics, growing up in a complete family appears to be the most important of all family-environment characteristics considered in the utility of schooling equation. In comparison, family income is virtually inconsequential. It is insignificant for whites and blacks, and positive but small for Hispanics.

Family size (number of siblings) appears to have a negative effect on blacks, but no significant impact on the schooling utility of whites or Hispanics. With regards to intergenerational transfer of education, we observe a positive and significant correlation between the education of parents and that of their offsprings, although the relationship is not uniformly significant. In our results, mother's education has a positive effect on the schooling utility of whites and blacks, while father's education is positive and significant for whites and Hispanics.

The effects of unobserved abilities are difficult to gauge. They not only work through the type-specific intercepts, but also interact with initial schooling (schooling at age 16). An important fact is that unobserved heterogeneity, represented by the type-specific intercepts in table 2, is significant in determining the schooling utility of whites, but that only selected types are of importance for Hispanics and blacks. We provide a closer look at the impact of unobserved heterogeneity on explaining ethnic differentials in schooling and wages later in this section.

4.2 The Effects of Schooling, Work Experience and Abilities on Wage

The wage returns to schooling and work experience are presented in table 3. The return to one year of schooling is the largest for Hispanics (7.9%) followed by the return for Whites (6.1%) and blacks (3.5%). Similarly, labor market experience also has a positive effect on wage, although of smaller magnitude. For whites the return to one year of work experience is about 2.9%, less than half of the return to one year of schooling. It is somewhat bigger for blacks and Hispanics, about 3.5% and 3.9% respectively.

Individual ability as measured by the composite ability test score does increase wage for whites and blacks (insignificant in the case of Hispanics), but the effect is rather small, especially when compared to the magnitudes of the effects of unobserved heterogeneity. The estimates of the type-specific intercepts in the wage equation suggest that type 1 is a dominant high-ability type for blacks, while for Hispanics type 3 is high-ability. In the case of whites, type 3 is low-ability, and there is no clearly dominating high-ability type, as the intercepts for types 1, 2 and 3 have similarly high magnitudes.

Admitedly, our current specification of the wage regression is somewhat limited. It could be improved, for example, by making returns to schooling vary across heterogeneity types, or by relaxing the assumption of local return to schooling being constant across schooling levels. Nevertheless, we believe that our model is an improvement over the standard approaches used in the returns to schooling literature.[33] Despite some variation across the three ethnic groups, our estimates of the wage returns to schooling are lower than those normally found in the traditional ordinary least squares (OLS) literature. The choice of OLS is justified only if realized schooling and unobserved market ability are uncorrelated, a central assumption that is hard to justify. Unlike traditional approaches, we maintain that individuals are heterogenous with respect to ability in school as well as in the labour market. Our model allows us to estimate the returns to schooling without any need to assume orthogonality between labour market

[33] See, for example, Belzil and Hansen (2007) for a more flexible specification of the wage regression within a similar structural model, and Belzil and Hansen (2002) for a discussion and comparison of structural dynamic models against traditional OLS and instrumental-variable (IV) approaches.

ability and schooling attainment, and without the estimates suffering from the otherwise ubiquitos ability bias.

4.3 Unobserved Heterogeneity, Schooling Attainments and Wages

In table 5, we present the predicted schooling attainments and wages by the five heterogeneity types along with the estimates of type probabilities (population proportions). We incorporate a rich specification of unobserved heterogeneity which enters all essential parts of our model. Unobserved abilities and tastes determine initial schooling levels, and directly enter the utility of attending school as well as the wage equation. Consequently, the effects of heterogeneity on individual's optimal schooling decisions and wage income are non-trivial. Furthermore, there are differences in how heterogeneity is distributed and how it operates across the three ethnic groups we consider. Majority of whites (39.7%) are of type 4, and so are blacks (38.5%). The predominant type for Hispanics is 3 (39%). Predictions in table 5 show great deal of variation in schooling across ethnics and ability types. For whites, type 3 individuals appear to be those most successful in scholastic terms. Similarly dominating are type 5 individuals in the case of Hispanics. For blacks, type 5 appears to have the highest attainment, but the predicted 12.6 years of schooling is not much higher than the 12 years predicted for types 3 and 4.

4.4 In-Sample Predictions and the Fit of the Model

In this section we examine the performance of our model in terms of how well it replicates the actual data that were used to estimate it. Figures 7, 8 and 9 show that the model predicts the schooling attainment for all three ethnics well. For blacks and Hispanics in particular, the model reproduces the actual schooling fairly closely. The fit is somewhat looser for whites, perhaps because of the more complicated bimodal shape of their schooling distribution, but it is still quite accurate.

Regarding the wage predictions (figures 10, 11 and 12), the model also shows a very satisfactory performance. As can be seen in the graphs, the predictions are close to the actual mean wage for ages 18 and over. The predictions for ages 16 and 17 are imprecise, but they are not of much interest as only few individuals would work at such a young age. We can conclude that, overall, the model fits the actual schooling and wage observations well, especially considering the limited amount of data available to estimate it.

4.5 Sources of Ethnic Gaps in Schooling and Wages

In the descriptive part of this paper, we pointed out the differences in characteristics and outcomes that exist between the white majority, and the minority groups of blacks and Hispanics. In this section, we investigate these differences more closely using our dynamic structural model. In particular, we focus on the relative importance of differences in endowments, resources and prices in explaining ethnic gaps in educational attainments and wages.

The first step in our assessment was to estimate the model separately for each of the three ethnic groups, thus imposing neither parameter equality nor equality of the distributions of unobserved heterogeneity. In this section, we proceed to summarize the overall importance of behavioural differences and endowments in explaining ethnic differences in the outcomes of interest.[34] That is, our goal is to decompose the mean difference in an outcome, y, between the majority ethnic group, W (whites) and the minority group, M (blacks or Hispanics):

$$\Delta^{WM} = \mathrm{E}_{\boldsymbol{\theta}^W}(y \mid \mathbf{x}^W) - \mathrm{E}_{\boldsymbol{\theta}^M}(y \mid \mathbf{x}^M), \tag{14}$$

where $\mathrm{E}_{\boldsymbol{\theta}^W}(y \mid \mathbf{x}^W)$ denotes the expectation of y conditional on the covariates of group W and evaluated at the parameter vector of group W, and $\mathrm{E}_{\boldsymbol{\theta}^M}(y \mid \mathbf{x}^M)$ is interpreted in the same fashion. Depending on the choice of reference group, there are two alternative ways of decomposing the difference in (14)

$$\Delta^{WM} = [\mathrm{E}_{\boldsymbol{\theta}^W}(y \mid \mathbf{x}^W) - \mathrm{E}_{\boldsymbol{\theta}^M}(y \mid \mathbf{x}^W)] + [\mathrm{E}_{\boldsymbol{\theta}^M}(y \mid \mathbf{x}^W) - \mathrm{E}_{\boldsymbol{\theta}^M}(y \mid \mathbf{x}^M)], \tag{15}$$

and

$$\Delta^{WM} = [\mathrm{E}_{\boldsymbol{\theta}^W}(y \mid \mathbf{x}^M) - \mathrm{E}_{\boldsymbol{\theta}^M}(y \mid \mathbf{x}^M)] + [\mathrm{E}_{\boldsymbol{\theta}^W}(y \mid \mathbf{x}^W) - \mathrm{E}_{\boldsymbol{\theta}^W}(y \mid \mathbf{x}^M)], \tag{16}$$

[34] We follow the same terminology as Cameron and Heckman (2001). Thus, differences due to parameters are 'behavioral differences', and covariates are 'endowments'. Furthermore, in the decompositions, ethnic differences in heterogeneity distributions are contained in the "behavioral difference".

In both equations, the first difference on the right-hand side represents the gap due to behavior, and the second one is the gap due to endowments. Note that this type of decomposition can potentially be sensitive to the choice of reference group and, in principle, one can get two very different estimates of the relative importance of endowments and behavior in explaining ethnic differences in education and wages. Therefore, in tables 6 and 7 we report both alternatives for a comparison.

In table 6, whites are predicted to attain on average 1.43 more years of schooling than blacks. Regardless which decomposition we employ, only a small portion of the schooling gap can be explained by behavioral differences (21% or 8.4%). That is, the white-black differences in educational attainment appear to be primarily determined by differences in endowments. However, this is not the case with wages. Whites are predicted to earn about 26% more per hour than blacks, and this gap seems to be mostly determined by differences in parameters which explain more than two thirds of the predicted wage gap regardless of the decomposition approach.

Differences between whites and Hispanics (table 7), both in schooling and wages, are not as pronounced as between whites and blacks. On average, whites are predicted to have higher educational attainment, about 0.89 years more, than Hispanics, and earn about 1.16 dollars more per hour. Similarly to blacks, the schooling gap between whites and Hispanics can largely be explained by differences in endowments. Depending on the decomposition approach, the behavioral differences can only explain 6.7% or 3.3% of the educational attainment differential. As for wages, differences in parameters and

differences in endowments both seem to explain about half of the white-Hispanic wage gap.

5 SUMMARY AND CONCLUSIONS

In this paper, we propose a structural dynamic programming model of schooling and wages, and estimate it separately for white, black and Hispanic males using the data from the 1997 to 2007 cycles of the NLSY97. The model respects the dynamic nature of schooling decisions made by rational, forward-looking agents, and employs a rich set of observables as well as a model for unobserved heterogeneity to isolate the effects of various individual characteristics on schooling attainment and wages. We find that certain components of family environment have a substantial impact on individual's schooling. Namely, growing up in a complete family (with both biological parents) appears to have a positive and significant effect on educational attainment across all three ethnics. Family income and, somewhat surprisingly, also parental education either have no impact on the utility of attending school, or their effect is not uniform across the three ethnics. The insignificance of family income suggests that policies based on providing monetary incentives to individuals from low-income families to continue schooling may not have the desired outcome.

Our structural estimates of the returns to schooling and work experience reveal some differences in how the market rewards the three ethnics. The return to one year of schooling is the highest for Hispanics at 7.9%, followed by 6.1% for whites and 3.5% for blacks. Whites have the lowest return to work experience (2.9%) of the three ethnic

groups (Hispanics 3.9% , blacks 3.5%). Our estimates, especially of the returns to schooling, are smaller than generally found through traditional least-squares analysis.

Having estimated the model parameters, we then simulate schooling and wages for all three ethnics under different assumptions, and decompose the observed differences in outcomes into the part explained by behavioral differences (parameters) and the one explained by differences in endowments (covariates). We find that differences in educational attainments can be to a large extent explained by differences in endowments between whites and the minority groups. While behavioral differences explain only a small part of the differences in schooling, they seem to play an important role in explaining differences in wages. This is especially true when comparing whites and blacks as more than two thirds of the black-white wage gap is explained by differences in parameters. Parameter differences explain about half of the white-Hispanic wage differential.

REFERENCES

Altonji, J. G., Bharadwaj, P. & Lange, F. (2008). *Changes in the Characteristics of American Youth: Implications for Adult Outcomes*. NBER Working Paper, no. 13883.

Belzil, C., & Hansen, J. (2002). Unobserved Ability and the Return to Schooling. *Econometrica*, 70(5), 2075-2091.

Belzil, C., & Hansen, J. (2003). Structural Estimates of the Intergenerational Education Correlation. *Journal of Applied Econometrics*, 18, 679-696.

Belzil, C., & Hansen, J. (2007). A Structural Analysis of the Correlated Random Coefficient Wage Regression Model. *Journal of Econometrics*, 140, 827–848.

Cameron, S. V., & Heckman, J. J. (2001). The Dynamics of Educational Attainment for Black, Hispanic and White Males. *Journal of Political Economy*, 109(3), 455-499.

Carneiro, P., Heckman J. J., & Masterov, D. V. (2005). Labor Market Discrimination and Racial Differences in Premarket Factors. *Journal of Law and Economics*, 58, 1-39.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3), 411-482.

Keane, M. P., & Wolpin, K. I. (2000). Eliminating Race Differences in School Attainment and Labor Market Success. *Journal of Labor Economics*, 18(4), 614-652.

Urzua, S. (2008). Racial Labor Market Gaps: The Role of Abilities and Schooling Choices. *Journal of Human Resources*, 43(4), 919-971.

APPENDIX A: TABLES

TABLE 18: SAMPLE MEANS/PROPORTIONS OF SELECTED VARIABLES

	White		Black		Hispanic	
Father's education	13.57	(2.80)	12.10	(2.26)	10.37	(3.90)
Mother's education	13.40	(2.44)	12.32	(2.05)	10.36	(3.56)
Parental income	35.68	(25.78)	23.80	(17.35)	24.45	(17.94)
Num. of siblings	2.28	(1.05)	2.65	(1.43)	2.74	(1.31)
Ability test score	0.48	(1.76)	-0.69	(1.67)	-0.46	(1.62)
Initial education	10.00	(0.76)	9.75	(0.98)	9.88	(0.86)
Final education	13.36	(2.54)	12.04	(2.36)	12.35	(2.30)
Nuclear	0.61		0.27		0.55	
Number of obs.	1,884		971		723	

Standard deviations in parentheses.
Education measured in completed years off schooling.
Parental income in thousands of 1997 dollars.
Nuclear equals 1 if the resp. lived with both biological parents until age 14.

TABLE 19: PARAMETER ESTIMATES: UTILITY OF SCHOOLING

	White		Blacks		Hispanics	
Intercept type 1	**-0.617**	**(0.286)**	0.435	(0.643)	**-1.233**	**(0.432)**
Intercept type 2	**-1.089**	**(0.216)**	**-0.956**	**(0.348)**	-0.545	(0.457)
Intercept type 3	**0.920**	**(0.001)**	**2.090**	**(0.296)**	**-1.198**	**(0.451)**
Intercept type 4	**-0.454**	**(0.169)**	-0.284	(0.169)	-1.075	(0.566)
Intercept type 5	**-0.970**	**(0.282)**	-0.184	(0.169)	**0.025**	**(0.002)**
Initial ys of educ. type 1	**0.107**	**(0.029)**	-0.138	(0.076)	**0.139**	**(0.043)**
Initial ys of educ. type 2	**0.114**	**(0.023)**	**0.102**	**(0.040)**	**0.087**	**(0.034)**
Initial ys of educ. type 3	**0.167**	**(0.005)**	**-0.244**	**(0.040)**	**0.155**	**(0.040)**
Initial ys of educ. type 4	**0.056**	**(0.017)**	0.002	(0.022)	**0.157**	**(0.048)**
Initial ys of educ. type 5	**0.148**	**(0.035)**	**-0.050**	**(0.026)**	**0.353**	**(0.011)**
Father's education	**0.030**	**(0.004)**	0.010	(0.006)	**0.011**	**(0.005)**
Mother's education	**0.020**	**(0.003)**	**0.027**	**(0.007)**	0.008	(0.004)
Family income	0.001	(0.001)	0.002	(0.002)	**0.004**	**(0.002)**
Number of siblings	-0.006	(0.009)	**-0.033**	**(0.010)**	-0.012	(0.013)
Nuclear	**0.168**	**(0.020)**	**0.066**	**(0.034)**	**0.209**	**(0.036)**
Ability score	**0.034**	**(0.007)**	**0.029**	**(0.012)**	**0.055**	**(0.014)**

Standard errors in parentheses. Estimates statistically significant at 5% or lower printed in bold.

TABLE 20: PARAMETER ESTIMATES: UTILITY OF WORKING

	White	Blacks	Hispanics
Intercept type 1	**1.753 (0.148)**	**2.353 (0.342)**	**1.314 (0.125)**
Intercept type 2	**1.735 (0.097)**	**1.420 (0.102)**	**0.948 (0.214)**
Intercept type 3	**0.789 (0.001)**	**1.222 (0.184)**	**1.500 (0.090)**
Intercept type 4	**1.116 (0.117)**	**1.008 (0.125)**	**1.127 (0.178)**
Intercept type 5	**1.751 (0.267)**	0.206 (0.304)	**-0.007 (0.002)**
Initial ys of educ. type 1	**0.093 (0.017)**	**0.096 (0.041)**	**0.052 (0.016)**
Initial ys of educ. type 2	**0.024 (0.010)**	**0.043 (0.012)**	**0.131 (0.025)**
Initial ys of educ. type 3	**-0.111 (0.004)**	**0.130 (0.022)**	**0.053 (0.020)**
Initial ys of educ. type 4	**0.093 (0.015)**	**0.086 (0.019)**	**0.174 (0.021)**
Initial ys of educ. type 5	**0.239 (0.030)**	**0.212 (0.040)**	**-0.107 (0.004)**
Years of schooling	**0.061 (0.005)**	**0.035 (0.008)**	**0.079 (0.011)**
Years of work exper.	**0.029 (0.006)**	**0.035 (0.008)**	**0.039 (0.014)**
Ability score	**0.028 (0.005)**	**0.016 (0.005)**	0.011 (0.007)

Standard errors in parentheses. Estimates statistically significant at 5% or lower printed in bold.

TABLE 21: PARAMETER ESTIMATES: INITIAL-SCHOOLING MODEL

	White	Black	Hispanic
Father's education	**0.033 (0.007)**	**0.070 (0.013)**	**0.033 (0.010)**
Mother's education	**0.048 (0.007)**	**0.145 (0.015)**	**0.047 (0.010)**
Family income	0.001 (0.001)	0.002 (0.004)	0.006 (0.003)
Number of siblings	-0.020 (0.022)	**-0.069 (0.023)**	-0.018 (0.029)
Nuclear	**0.136 (0.055)**	0.087 (0.077)	**0.329 (0.078)**
Ability score	**0.056 (0.016)**	0.017 (0.023)	0.020 (0.030)

Standard errors in parentheses. Estimates statistically significant at 5% or lower printed in bold.
Estimates of the cut-off points not reported for space considerations (can be provided on request).

TABLE 22: MEAN YEARS OF SCHOOLING BY ETHNICITY AND TYPE

	Type	Proportion of population	Schooling
Whites	1	0.189	14.3
	2	0.265	11.5
	3	0.080	19.6
	4	0.397	14.0
	5	0.069	15.1
Blacks	1	0.034	11.1
	2	0.283	10.9
	3	0.071	12.0
	4	0.385	12.0
	5	0.227	12.6
Hispanics	1	0.260	11.6
	2	0.213	11.0
	3	0.390	13.5
	4	0.093	12.7
	5	0.043	19.2

TABLE 23: DECOMPOSITIONS OF WHITE-BLACK GAPS IN SCHOOLING AND WAGES

	Schooling (years)	Wage (dollars)
$E_{\theta^W}(y \mid \mathbf{x}^W) - E_{\theta^B}(y \mid \mathbf{x}^B)$	1.43 (whites 11.8% more)	2.87 (whites 26.0% more)
$E_{\theta^W}(y \mid \mathbf{x}^W) - E_{\theta^B}(y \mid \mathbf{x}^W)$	0.30 (explains 21% of the gap)	1.95 (explains 67.9% of the gap)
$E_{\theta^B}(y \mid \mathbf{x}^W) - E_{\theta^B}(y \mid \mathbf{x}^B)$	1.13	0.92
$E_{\theta^W}(y \mid \mathbf{x}^B) - E_{\theta^B}(y \mid \mathbf{x}^B)$	0.12 (explains 8.4% of the gap)	1.98 (explains 69% of the gap)
$E_{\theta^W}(y \mid \mathbf{x}^W) - E_{\theta^W}(y \mid \mathbf{x}^B)$	1.31	0.89

Decompositions of the wage differential based on the predicted wage at age 26.

TABLE 24: DECOMPOSITIONS OF WHITE-HISPANIC GAPS IN SCHOOLING AND WAGES

	Schooling (years)	Wage (dollars)
$E_{\theta^W}(y \mid \mathbf{x}^W) - E_{\theta^H}(y \mid \mathbf{x}^H)$	0.89 (whites 7% more)	1.16 (whites 9.1% more)
$E_{\theta^W}(y \mid \mathbf{x}^W) - E_{\theta^H}(y \mid \mathbf{x}^W)$	0.06 (explains 6.7% of the gap)	0.56 (explains 48.3% of the gap)
$E_{\theta^H}(y \mid \mathbf{x}^W) - E_{\theta^H}(y \mid \mathbf{x}^H)$	0.83	0.60
$E_{\theta^W}(y \mid \mathbf{x}^H) - E_{\theta^H}(y \mid \mathbf{x}^H)$	0.03 (explains 3.3% of the gap)	0.59 (explains 50.9% of the gap)
$E_{\theta^W}(y \mid \mathbf{x}^W) - E_{\theta^W}(y \mid \mathbf{x}^H)$	0.86	0.57

Decompositions of the wage differential based on the predicted wage at age 26.

APPENDIX B: FIGURES

FIGURE 1: INITIAL SCHOOLING AT AGE 16

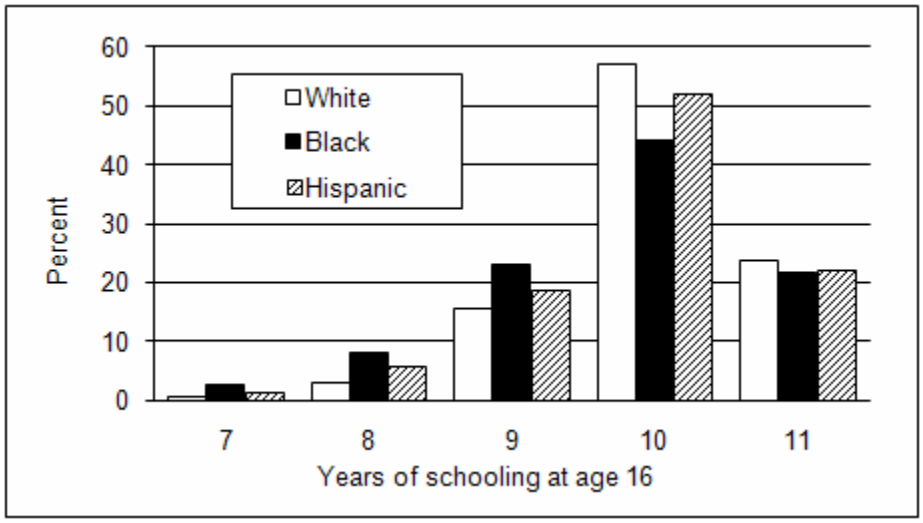

FIGURE 2: FINAL SCHOOLING ATTAINMENT

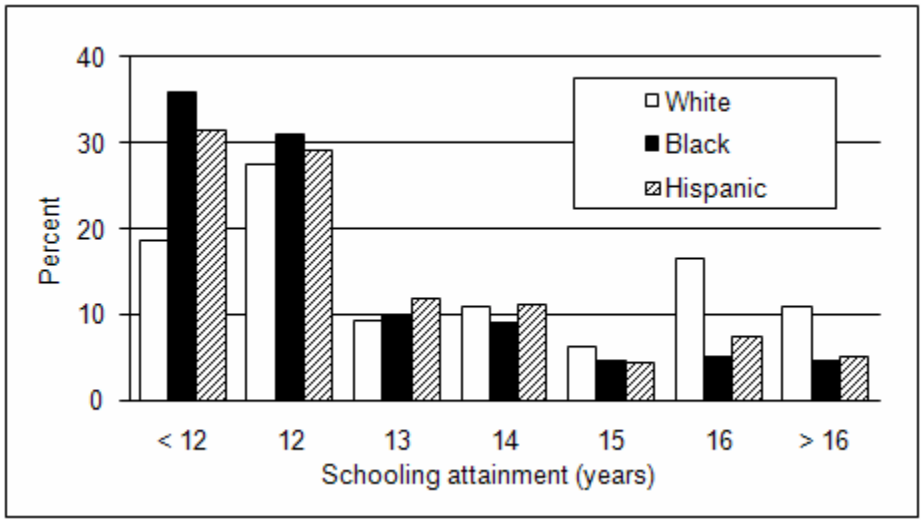

FIGURE 3: ESTIMATED PROBABILITIES OF COMPLETING AND CONTINUING PAST A GIVEN SCHOOL GRADE

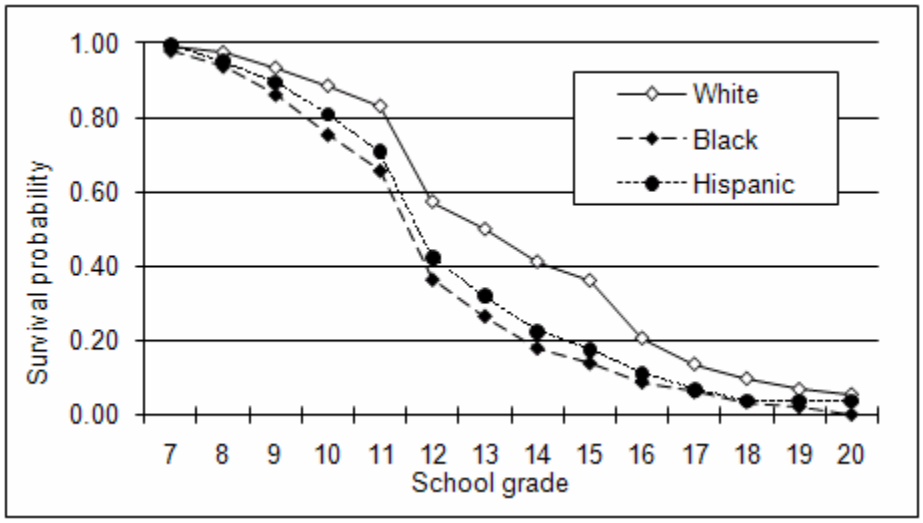

FIGURE 4: ESTIMATED PROBABILITY DENSITIES OF THE ABILITY TEST SCORES

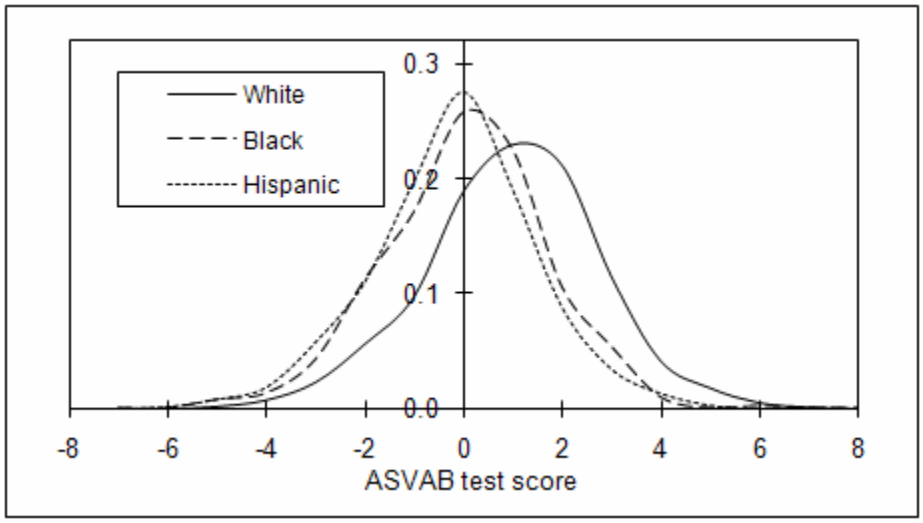

FIGURE 5: AGE-WAGES PROFILES

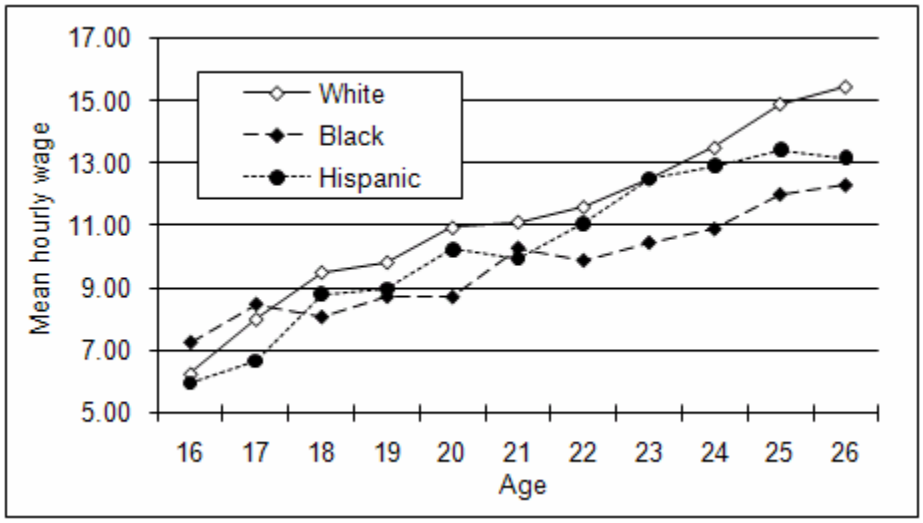

FIGURE 6: MINORITY WAGE AS A PERCENTAGE OF WHITE WAGE

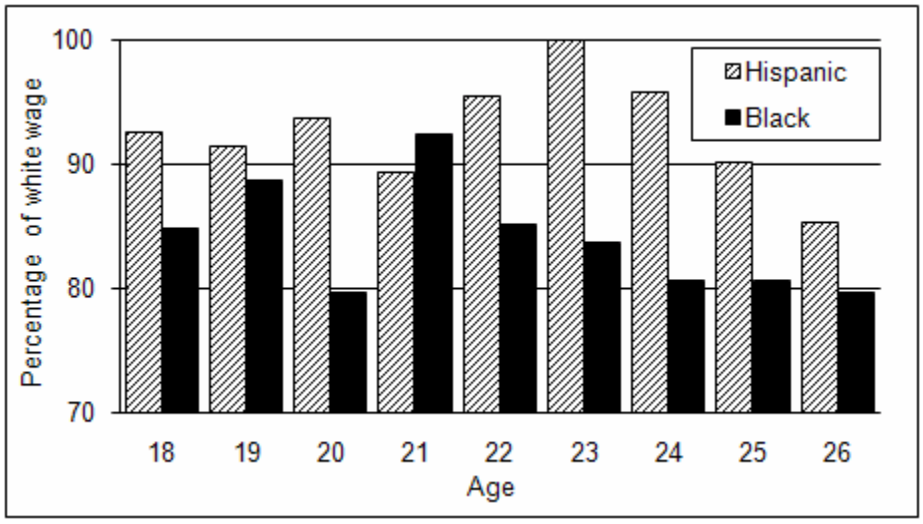

FIGURE 7: ACTUAL AND PREDICTED SCHOOLING: WHITES

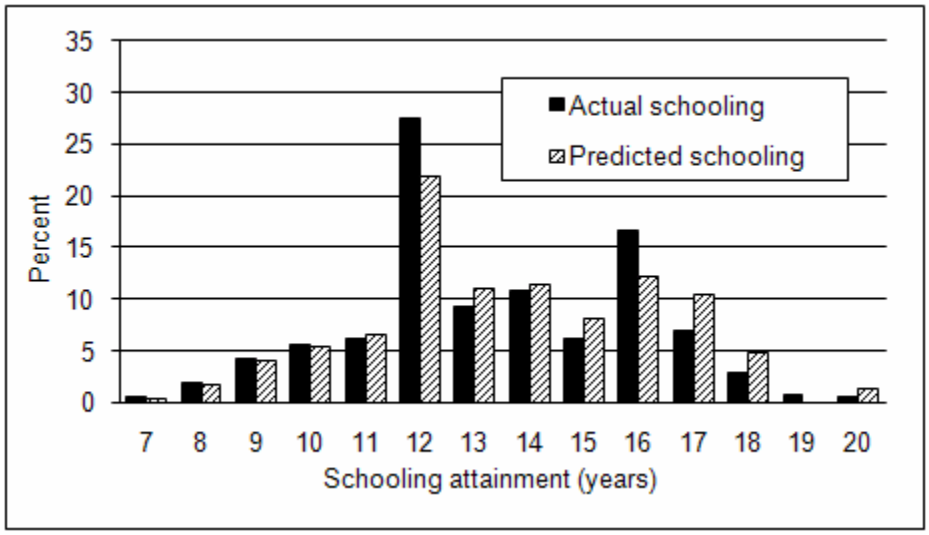

FIGURE 8: ACTUAL AND PREDICTED SCHOOLING: BLACKS

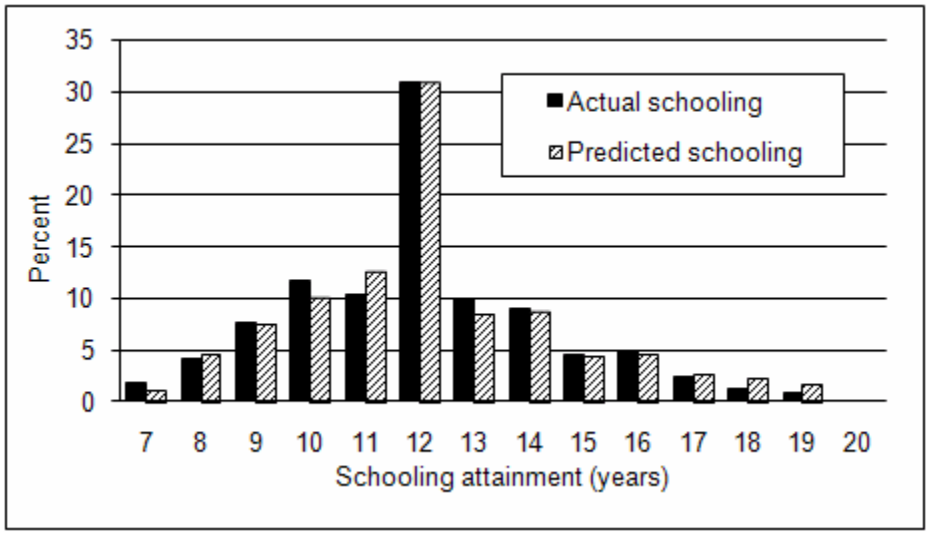

FIGURE 9: ACTUAL AND PREDICTED SCHOOLING: HISPANICS

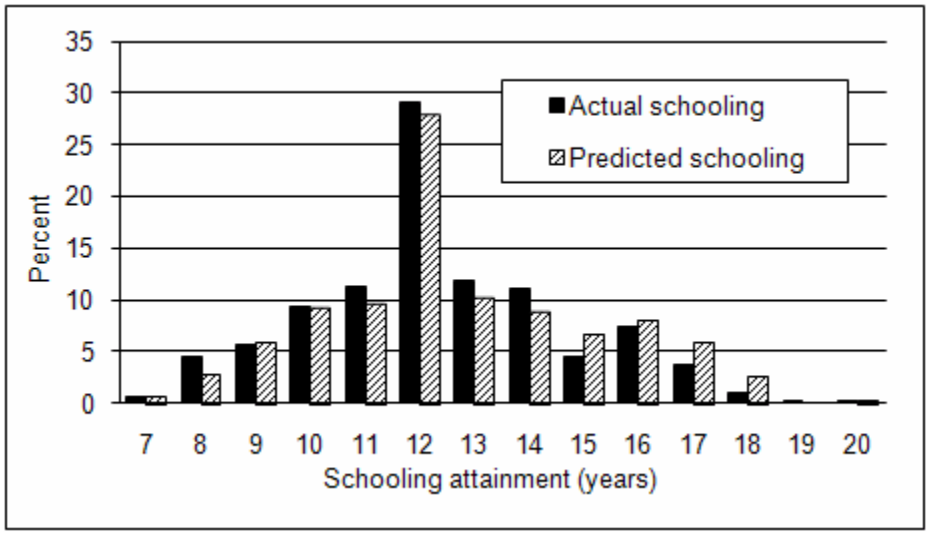

FIGURE 10: ACTUAL AND PREDICTED WAGES: WHITES

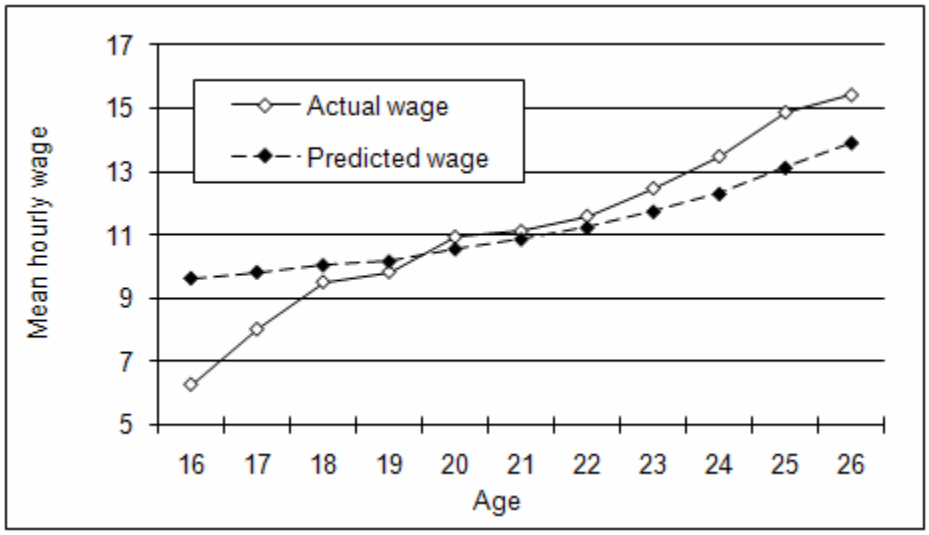

FIGURE 11: ACTUAL AND PREDICTED WAGES: BLACKS

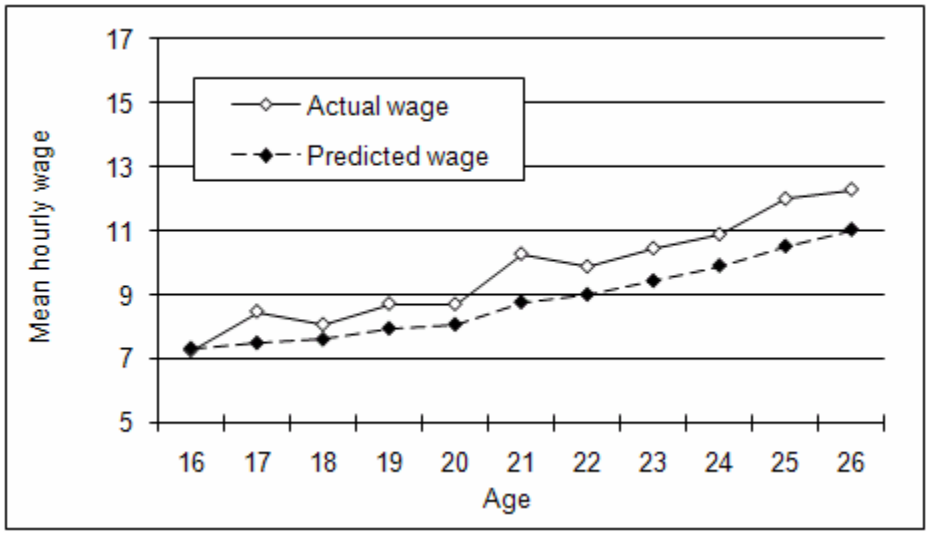

FIGURE 12: ACTUAL AND PREDICTED WAGES: HISPANICS