THE ABORIGINAL-WHITE EARNINGS DIFFERENTIAL IN CANADA

BY

JOANNA ROBERTS

AN ESSAY SUBMITTED TO THE DEPARTMENT OF ECONOMICS

AS PART OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

QUEEN'S UNIVERSITY, KINGSTON, CANADA

DECEMBER 2011

COPYRIGHT © JOANNA ROBERTS

1. Introduction

The economic disparity of aboriginal people is a major problem in Canada. Government policy perennially acknowledges that the improvement of economic outcomes of aboriginals is a priority; a simply-stated goal, but large in scale. Those reporting aboriginal ancestry on the 2006 Census make up 5.4 percent of the population, approximately 1.7 million people, and programs targeting this group receive approximately $10 billion per year in funding, according to the Ministry of Indian and Northern Affairs (2010). One reason that the issue is so central to Canadian public policy is the huge cost of foregone output from the unrealized economic potential of aboriginal Canadians. The Report of the Royal Commission on Aboriginal Peoples (1996) comprehensively estimates the cost of the economic marginalization of aboriginal Canadians. It calculates the potential income that could be generated if the aboriginal population had similar levels of employment, education, and investment to the rest of the country. The total value of production foregone is estimated to be $5.8 billion for the year 1996 alone. (Although this study is somewhat outdated, the study is uniquely broad in scope and helpful to grasp the magnitude of the economic gap.) It amounts to approximately seven-tenths of one percent of Canada's gross domestic product for that year. The economic success of the aboriginal population is therefore important to that of the entire country.

A major source of the disparity is the labour market performance of aboriginal Canadians. This can be measured in many ways, and this paper focuses on the earnings differential between aboriginal and non-aboriginal Canadians. Under the theories of human capital investment and discrimination in the labour market, the earnings differential between two groups can be thought of as a combination of two factors: first, a difference in productivity and second, a difference in the value assigned by the market to the labour of each group, controlling for human capital (see, for example, Cotton 1988). This value assigned by group status alone may reflect discrimination. A technique to decompose the earnings differential into these two parts, a human capital component and an 'unexplained,' or potential discrimination component, was proposed by Blinder (1973) and

Oaxaca (1973), and is now standard in the literature. An earnings function is estimated for each group by regressing the logarithm of wages or earnings on a vector of characteristics, personal and locational. The personal characteristics attempt to capture the level of human capital, and include factors such as level of education, work experience and marital status (Mincer 1974, 83). The differential that is not attributable to these observable characteristics is the part of (most) interest. It may be due to a difference in inherent, unmeasurable productivity, or it may also be due to discrimination. Concluding that discrimination is present requires the strong assumption that all the factors that determine the wage are contained in the vector of observed characteristics, and is obviously subject to dispute. For example, the intensity of work per unit time might differ across the two groups, and thus the model is misspecified. However, with a reasonably complete and accurately measured set of personal characteristics, this estimate serves as a starting point and empirical norm for discussion.

This paper will estimate the earnings functions of aboriginal and non-aboriginal male Canadians and analyze the determinants of the earnings differential, based on a 1994 study by George and Kuhn. Their paper was one of the first in Canada to use regression analysis to study the aboriginal-white wage differential, and many similar papers have been published since. This study will use their methods on a newer and larger set of Canadian data, the master file of the 2006 Census. The earnings differential will be decomposed according to the Blinder-Oaxaca method, allowing us to measure the fraction of the differential that may be attributed to wage discrimination. In addition, a sample selection model based on Heckman (1979) will correct for the non-random participation in the work force, and allow the structure of our earnings function to better represent the whole population of working age.

2. Review of the Literature

Before reviewing the literature, it is worthwhile to clearly define the terms used in this subject area. This paper will use the terminology of the 2006 Canadian Census dictionary. The term 'aboriginal' refers to the ethnic groups North American Indian, Métis, and Inuit. It replaces the terms 'native' and 'Indian' that are found in previous studies. The comparison group will be referred to as 'white,' those who are neither a visible minority nor report any aboriginal ethnicity. The 2006 Census identifies aboriginal ethnicity according to several different concepts, and the wording of the relevant questions is given in Appendix A. The one used in this study is *aboriginal ancestry*. This refers to anyone who reported at least one North American Indian, Métis, or Inuit ancestry to question 17 on the 2006 Census. Ancestry is the most similar concept to that of the 1986 Census, and will be used for comparability with George and Kuhn. The 1986 question asked, "To which ethnic or cultural group(s) do you or did your ancestors belong?" followed by a series of choices to check off (Profile of Canada's Aboriginal Population, 1995). The 2006 question asked "What were the ethnic or cultural origins of this person's ancestors?" with spaces to write up to eight responses (2006 Census Dictionary, Statistics Canada, 2010). This modification to the question and changes in the social environment may affect responses, but it is essentially the same concept.

Another important distinction is that between those who report *solely* aboriginal ancestry and those who report *any* aboriginal ancestry, possibly in combination with non-aboriginal ancestry. This paper will refer to the former group as 'aboriginal origins only' and the latter, which is broader and includes the previous group as a subset, as 'any aboriginal origins.' A note to make about this approach is that the Métis group descends from aboriginal and non-aboriginal ancestors. Therefore, in this study, those who report solely Métis ancestry will be strictly in the 'any aboriginal origins' group.

Aboriginal identity is another potentially useful concept, but it will not be used in this study. It is a relatively new concept, only included in the Canadian Census since 1996. It denotes those who report identification with an aboriginal group to question 18 on the 2006 Census. As an indication of the relative size of the groups, in the 2006 Census, there were 1 138 285 who reported aboriginal identification and 1 678 200 who reported aboriginal ancestry (Aboriginal Peoples Technical Report, 2006 Census). Since it was not reported on the 1986 Census, and a primary goal of this paper is comparability with the study by George and Kuhn, it will not be used in this study. As well, ancestry is perhaps more useful than identity to study discrimination against the aboriginal population, since it uses the entire group that is recognized by employers, rather than only those who self-identify as being a member.

The number of empirical studies conducted on aboriginal-white earnings differentials in Canada is less than one would expect for an issue that receives so much attention in the popular media and in government policy. Part of the reason is a lack of data. The Census is the only source of data in Canada with a large enough sample size to examine small aboriginal populations, particularly because many communities are small and confidentiality is a problem. In addition, many reserves do not participate in the Census, and this causes potentially significant representativeness issues. However, a Canadian literature exists nonetheless, and has important findings.

Using the public-use file of the 1986 Census, George and Kuhn (1994) find that the raw earnings differential between aboriginals and whites is significant, but is smaller than that of other disadvantaged groups, at approximately 14 percent of log earnings for males working full-time and full-year. Their findings are summarized in Table 1. The earnings gap is between whites and the specified aboriginal group, using the white earnings as a reference. The 'explained' quantity is the amount of the total differential that is attributable to human capital and locational characteristics. This decomposition uses the earnings structure of the white sample as the basis of comparison,

although, as will be discussed, there are other methods with other bases of comparison that can be used. The authors also account for self-selection into the full-time, full-year work force with a Heckman selectivity model. These latter results are of low statistical significance and the decomposition of the wage gap is apparently affected little by this adjustment.

TABLE 1: EARNINGS DIFFERENTIAL BETWEEN WHITE AND ABORIGINAL MALES, FULL-TIME AND FULL-YEAR WORKERS, WITH AND WITHOUT CORRECTION FOR SELECTION INTO FULL-TIME, FULL-YEAR WORK, 1986 CANADIAN CENSUS. (GEORGE AND KUHN, 1994.)

	Total gap, as a percent of log earnings of whites	Percent explained, uncorrected for selection into full-time, full-year work	Percent explained, corrected for selection into full-time, full-year work
Any aboriginal origins	14.0	51	50
Aboriginal origins only	24.2	59	58

The earnings gap is smaller for the more inclusive group with any aboriginal origins (and therefore mixed ancestry) than for those with solely aboriginal origins. George and Kuhn are limited to examining only aboriginals living off-reserve, because of the small sample size of the on-reserve population. They find that observable characteristics account for about one-half of the differential for males with any aboriginal origins and three-fifths for those with aboriginal origins only. George and Kuhn note that, from the earnings regressions, the structure of wages appears to be the same for those of mixed origins and solely aboriginal origins, differing only in the intercept. Therefore, all those with any aboriginal origins are pooled together and mixed ancestry is distinguished from solely aboriginal origins with a dummy variable. When examining the determinants of participation in the labour force for aboriginals, George and Kuhn find a negative selection coefficient into full-time, full-year work. This suggests that the subsample of workers included in the earnings regression is in fact less 'able' in terms of unobserved characteristics than the average of the population. As such, it is unlikely that the earnings gap is underestimated. The

details of the model, based on Heckman (1979) and Greene (2003), will be discussed in more detail in the methodology section below.

De Silva (1999) uses much of the same methodology of the previous paper, with data from the 1991 Canadian Census. He finds a raw earnings gap of 15 percent, between white and those with any aboriginal origins, for full-time and full-year male workers. Potential discrimination accounts for less than half of the earnings differential, 43 percent. One flaw of this study, however, is that it fails to distinguish between aboriginals living on- and off-reserve. Economic conditions are far worse on reserves, and it is inaccurate to assume that workers face the same labour market prospects. The present study will avoid this problem by using the larger master file of the Census, rather than the public-use file, so as to identify those living on reserves. De Silva finds no significant selectivity effect into full-time, full-year work; however, his Heckman selection model is likely misspecified because of the missing reserve data.

Pendakur and Pendakur (2011) explore the earnings disparity according to the various concepts of 'aboriginality' defined by the more recent Census cycles. They divide those with some aboriginal origin into seven groups: ancestry (aboriginal origins only or mixed aboriginal origins), identification (North American Indian, Métis, or Inuit), and legal status (registered Indians on-reserve and registered Indians off-reserve). Using Census data from 1996, 2001, and 2006, they find consistency in the male earnings gap across the years. Each year, legal-status aboriginals on-reserve had the highest earnings gap, approximately 48 to 53 percent, those reporting aboriginal identification, 18 to 37 percent, and those with multiple-origin aboriginal ancestry the lowest, between 6 and 9 percent. They also find that earnings disparity in general declined for aboriginals over the period 1995-2005. The earnings gap decreased as much as one-half for the Métis identity group and one-third for the North American Indian identity group. By exploiting the richer data available from the recently-added Census questions on identity, Pendakur and Pendakur shed light

on the complex nature of aboriginal ethnicity and how the varying definitions play a role in labour market outcomes.

Kuhn and Sweetman (2002) hypothesize that assimilation is the driving force behind the labour market success of the aboriginal population in Canada. They compare aboriginals to immigrants, whose economic outcomes improve with the acquisition of the traits of the culture and labour market into which they enter. Although aboriginals do not immigrate per se, they are a minority in their own country, and the economic culture is very different from their traditional one. By examining phenomena such as the improved outcomes of the descendants of intermarriage of aboriginals and non-aboriginals, Kuhn and Sweetman argue that aboriginals have some characteristics of 'unwilling immigrants' to Canada.

Hurst (1997) examines the aboriginal earnings differential in the United States, using the 1990 US Census. The majority group in this study is American-born non-aboriginals, however, instead of non-visible minorities as is generally used in the Canadian literature. Similar to Pendakur and Pendakur, he uses two different concepts from the Census, ancestry and 'race' (comparable to identity on the Canadian Census), to define several categories of aboriginal ethnicity. Contrary to Pendakur and Pendakur (2011), however, Hurst finds that there is less of a negative earnings differential for single-ancestry aboriginals than for mixed-ancestry aboriginals, once human capital characteristics are controlled for. The unexplained differential for the mixed-ancestry group is 41 percent of the total, compared to 13 percent for single-ancestry aboriginals. This suggests that this type of discrimination in the American labour market is relatively minor and increasing the human capital of the aboriginal population would eliminate a large amount of earnings disparity.

Australia also has a substantial literature on the economic outcomes of its aboriginal population. Daly (1994) finds that the wage differential is very highly attributable to differences in human capital, using the 1991 Australian Census. In fact, for females, it is wholly due to the

difference in endowments, and only 10 percent is unexplained by human capital for males. The study has some significant flaws, however, and is not entirely comparable to the Canadian literature. The concept of ethnicity is somewhat less nuanced, with the population simply divided into aboriginal and non-aboriginal. Additionally, the study includes only those working full-time and full-year, but ignores the possible selection bias that this likely generates. It is highly doubtful, given that the Australian and Canadian aboriginal populations display similarly poor economic outcomes, that there is not some sort of selectivity effect introduced by including only this relatively better-paid group of workers.

The literature on South American indigenous wage differentials is small but growing. For example, Wood and Patrinos (1994) find, in Bolivia, that 72 percent of the wage differential between urban indigenous and non-indigenous males is explained by human capital endowments. However, this study, along with others from the region, are perhaps less relevant to the present one since the economic conditions of the country as a whole are very different from those in Canada. Income inequality is higher overall, with the top-earning quintile receiving 57.5 percent of the country's total earned income. Half of the population is below the poverty line (Psacharopoulos et al, 1992). Most importantly, the majority of indigenous people in Bolivia live in rural areas, and there is insufficient data on the rural population's incomes to include them in the study of wage differentials and so Wood and Patrinos could only examine the urban population. As data collection improves in South America, however, this body of research may become more useful for comparison to studies such as the present one.

3. Data Overview and Analysis Sample Used in the Study

The data for the present study come from the population universe of the 2006 Canadian Census Master File, which contains data on 20 percent of Canada's population. We exclude institutional residents and those under 21 or over 64 years of age. To provide cleaner estimates of the earnings differential specifically for aboriginals, we do not include visible minorities in our sample. In addition, for clarity in the results, the northern territories will be excluded from this study. George and Kuhn (1994) find a very different wage structure in the northern territories, for both aboriginals and whites. The earnings differential between the two groups is considerably higher there, especially for women. The mean wages for both aboriginals and whites in the North are higher relative to the rest of Canada because of the relatively high cost of living there, but it is the difference in this 'northern premium' on wage that affects the earnings differential. The higher earnings gap can be attributed to a smaller northern wage premium for aboriginal males, approximately 5 percent, than for white males, at 10 percent (George and Kuhn, 1994). This phenomenon merits its own study, and the factors that cause it may cloud the results of the present one.

Focusing exclusively on males in the present study also removes another potential complication in its findings. The analysis of females' earnings is much more affected by differences in labour supply within a household, relative to males. This issue will be avoided in the present study for this reason. On-reserve aboriginals will also be excluded from the main earnings regressions in the present study, for several reasons. Under the *Indian Act,* income earned on a reserve is not subject to federal or provincial taxation, and this can affect both wages paid and the supply of labour. According to the 2006 Census, the mean unemployment rate on reserves was 23 percent, another factor that can have an effect. Finally, a small number of reserves remain incompletely enumerated, approximately 2.5 percent of all reserves in the 2006 Census. Caution must be used with these data, and it is therefore not included in the study. However, the subsample

of the aboriginal population that lives off-reserve is likely not randomly selected, and examining them alone may potentially introduce bias. This will be considered later in this study.

The hourly wage rate would be the best measure to use in a study of this type; however, it is not reported in the Census. The annual, individual measure 'Wages and Salaries', defined as the pre-tax income earned in the year 2005 from wages, salaries, commissions, and tips, will be used instead. For this reason, only full-time, full-year (FTFY) workers are used in the earnings regressions, defined as those working at least 30 hours a week for at least 49 weeks in 2005. This necessitates a correction for selectivity bias, as selection into the FTFY group is almost certainly non-random. Persons who are self-employed and those who report zero wage and salary earnings will also be excluded, as this measure of wages is unreliable for these workers. Finally, those who report a family income that is less than their own wage and salary earnings (perhaps due to negative investment or self-employment income) are excluded. The log of family income net of the individual's wage and salary earnings is used in the selectivity model, and those with a negative value must be dropped.

The final sample of working-age, off-reserve residents of the provinces, to be used in the selection model, is as follows. It consists of 1 011 646 white males and 57 105 males with any aboriginal origins, of which 12 753 males have solely aboriginal origins and 44 352 males have mixed aboriginal and non-aboriginal origins. The subsample of FTFY workers used in the earnings regressions consists of 657 714 white males and 31 279 males with any aboriginal origins, 5 926 males with solely aboriginal origins and 25 353 with mixed aboriginal and non-aboriginal origins.

4. Methodology Followed in the Study

4.1 Earnings Regressions

The earnings function is estimated with a standard human capital semi-log equation.

$$w_i = x_i\beta + \varepsilon_i \quad (1)$$

The dependent variable w_i is the logarithm of the individual's annual wage and salary income, described in the previous section. The vector of human capital and locational characteristics, x_i, contains the following: level of education (represented by a set of six education dummy variables, which includes a category for non-university training), age and its square, province or region of residence, an urban/rural indicator, marital status (represented by three dummy variables), and language (represented by four dummies). The Census does not provide a measure of work experience, so age is used as a proxy for potential work experience. Its squared value is also included to capture a possible non-linear effect, as experience is expected to have positive but diminishing returns (Mincer, 1974). The sample means of these explanatory variables are given in Table 2. The 'mixed aboriginal origins' category, in the rightmost column, consists of those with mixed aboriginal and non-aboriginal ancestry, therefore is the 'any aboriginal origins' group less the members with solely aboriginal origins. This group is not used by itself in the regressions, but the sample means are included to illustrate the differences within the larger group with any aboriginal origins. Also, note that the 'aboriginal language' identifier is to be used in the selection model to come, and not in the basic earnings regressions.

TABLE 2: SAMPLE MEANS, CANADIAN MALES AGED 25-64, WORKING FULL-TIME AND FULL-YEAR, LIVING IN THE PROVINCES AND OFF-RESERVE (2006 CANADIAN CENSUS)

	White	Any Aboriginal Origins	Aboriginal Origins Only	Mixed Aboriginal Origins
PEI/ Newfoundland	0.0178	0.0204	0.0324	0.0176
Nova Scotia	0.0316	0.0302	0.0225	0.0320
New Brunswick	0.0254	0.0193	0.0179	0.0196
Quebec	0.2622	0.1678	0.2302	0.1532
Ontario	0.3769	0.3039	0.2676	0.3124
Manitoba	0.0351	0.0986	0.0828	0.1022
Saskatchewan	0.0270	0.0591	0.0657	0.0576
Alberta	0.1171	0.1642	0.1314	0.1719
British Columbia	0.1069	0.1365	0.1495	0.1335
Age	42.160	39.406	39.470	39.391
Rural	0.1876	0.2232	0.2413	0.2190
Never married	0.3054	0.3825	0.4781	0.3602
Married	0.5861	0.4978	0.3909	0.5228
Widowed/separated/ divorced	0.1085	0.1197	0.1310	0.1171
Less than high school	0.1237	0.1978	0.2809	0.1783
High school diploma	0.2479	0.2716	0.2678	0.2725
Some post-secondary	0.2476	0.2265	0.1857	0.2360
Bachelor's degree	0.1376	0.0761	0.0501	0.0822
Post-graduate degree	0.0729	0.0349	0.0218	0.0380
Non-university training	0.1703	0.1931	0.1937	0.1930
English spoken only	0.6601	0.7139	0.7364	0.7096

French spoken only	0.1199	0.0667	0.1025	0.0583
Bilingual	0.2189	0.2154	0.1587	0.2287
Neither official language spoken	0.0011	0.0040	0.0024	0.0034
Aboriginal language	-	0.0278	0.1186	0.006576
Mixed aboriginal origin	-	0.8378	-	-
Wage and salary income ($)	60326.55	48893.81	42298.55	50435.38
N	657 714	31 279	5 926	25 353

The aboriginal and white populations show differences in important characteristics. Aboriginal Canadians are more likely to live in rural areas, even with on-reserve populations omitted. They are younger on average, with a mean age almost three years younger than that of the white sample. Whites are more likely to be married, and aboriginals are more likely to be separated, divorced, widowed, or never married. A higher proportion of whites have a university degree (both bachelor's and postgraduate), and a higher proportion of aboriginals have non-university training. Whites are more likely to speak French only than any of the aboriginal groups, and aboriginals are more likely to speak English only.

The mean annual wage and salary earnings for a male, white full-time, full-year worker is $60 327, $48 894 for a male FTFY worker with any aboriginal origins, and $42 299 for a male FTFY worker with aboriginal origins only.

The earnings regressions will be estimated with ordinary least squares for three groups, those with any aboriginal origins, those with aboriginal origins only (which is a subset of the previous group), and whites. The group with any aboriginal origins will have included in their variable of characteristics a dummy variable that indicates mixed aboriginal origins. As mentioned

previously, this is the strategy of George and Kuhn (1994), who found that it sufficiently captured the difference in earnings between those with solely aboriginal origins and those with mixed aboriginal origins. They found that the level of earnings differed across the two groups, but through the intercept rather than the wage structure, and so a dummy variable suffices.

4.2 Blinder-Oaxaca Decomposition

The technique used to decompose the earnings differential is based on the methodology of Blinder (1973) and Oaxaca (1973), with modifications as detailed below. The observed difference in earnings can be divided as follows, denoting aboriginals with *a* and non-aboriginals with *n.* Here $\bar{w}$ signifies the mean of the log of earnings:

$$\bar{w}^n - \bar{w}^a = (\bar{x}^n - \bar{x}^a)'\hat{\beta}^n + \bar{x}^{a\prime}(\hat{\beta}^n - \hat{\beta}^a) \quad (2)$$

The left-hand side of the equation is the observed difference in mean log earnings between the two groups as reported in the data. The first term on the right-hand side is the vector of differences in mean human capital (including locational) characteristics, multiplied by the observed coefficients (the $\hat{\beta}_j$) from the earnings regression for non-aboriginals. This is the 'explained' component, that is, the part attributable to differences in human capital and other productivity-related endowments. The wage structure of the majority (non-aboriginal) group is used here as the base of comparison for illustration purposes, but this choice will be discussed in more detail in the following paragraph. The rightmost term is the vector of differences in observed coefficients, multiplied by the mean human capital characteristics for an aboriginal. This is the 'unexplained' component, and may be an indicator of possible labour market discrimination. It measures the difference in the white and aboriginal wage structures as evaluated by the labour market, that is,

the difference in the $\hat{\beta}_j$, and multiplies that by the vector of characteristics of a representative aboriginal person. The structure of the decomposition relies on the fact that the mean residual is zero in an OLS regression with an intercept term. This follows directly from the fact that the regression line passes through the point of each sample mean for each regressor. It can be useful to examine the relative importance of a certain regressor or group of them. In this study, determining the amount of the explained component attributed to, for example, education, provides additional insight into the structure of the earnings gap.

When reporting a differential as a fraction or a percentage, the white log earnings will always be used as the denominator. The differential as a fraction is calculated by

$$1 - e^{-d} \tag{3}$$

where d is the raw differential expressed in log points. Since, for this study, d is always the difference between white log earnings and those of the aboriginal group in question, the white log earnings are the reference.

The Blinder-Oaxaca decomposition is appealing in its intuitive explanation. The explained term $(\bar{x}^n - \bar{x}^a)'\hat{\beta}^n$ estimates the difference in earnings that a white male would see, if his productivity characteristics were changed to those of a representative aboriginal male. The unexplained term $\bar{x}^{a\prime}(\hat{\beta}^n - \hat{\beta}^a)$ predicts the amount that a representative aboriginal male's earnings would change if he were subject to the same treatment as a white male in the labour market. If there were no difference between the treatment of whites and aboriginals in the labour market, then $\hat{\beta}^n$ would be equal to $\hat{\beta}^a$ and the unexplained term would disappear.

The positions of the two groups can be reversed in equation (2), of course. One could use the coefficients/wage structure for aboriginals in the explained component, and the characteristics of a representative white person in the unexplained. The unexplained component now reflects the white 'premium' on wage, rather than the aboriginal 'penalty.' In addition, the explained component

uses the wage structure of aboriginals as the 'norm.' The most representative wage structure as the norm has been the topic of several papers since the original methodology was proposed. A hypothetical, neutral wage structure is sought, one that would prevail in the absence of discrimination. For example, Neumark (1988) suggests that a non-discriminatory norm for the wage structure can be found in a pooled regression. This study, however, will use the white earnings structure as the norm. This relies on the reasonable assumption that white males are the least likely to face labour market discrimination of this sort. Therefore, this earnings structure will serve as the basis of comparison for the analysis of discrimination.

One potential flaw in the Blinder-Oaxaca method arises from the focus on the residual. The measure of productivity-related factors in x_i cannot realistically be exhaustive and completely capture all the determinants of productivity. Missing factors might include inherent motivation, ability, past work experience, and so on. The Canadian Census does not contain enough data to control for all these productivity factors, leaving a possible omitted-variables bias. Since the residual is the quantity of interest, it must be stressed that this is, unfortunately, a somewhat imperfect measure.

4.3 Heckman Selection Model for Participation in Full-time, Full-year Work

Since earnings are observed only for those employed, and of those who are employed, only full-time and full-year workers are included in our regressions, our estimates from the previous regressions are likely nonrepresentative of the entire population of working-age Canadians. Selectivity bias may occur in a situation such as this, because selection into the subsample is likely non-random. To correct for this, Heckman's (1979) selectivity-adjustment method will be used, as follows.

Let the logarithm of the reservation wage of worker i be denoted ${w_i}^s$. It is determined by a vector of observed characteristics q_i according to:

$$ {w_i}^s = {q_i}'\varphi + \varepsilon_i \tag{3} $$

The logarithm of market wage offered to worker i is ${w_i}^m$ and determined according to equation (1) by the observed characteristics x_i. However, ${w_i}^m$ is observed only if ${w_i}^m > {w_i}^s$. Let this difference be ${w_i}^*$. It is a function of the vector z_i, which contains all the determinants of the shadow and market wages, as shown in the reduced-form equation (4). The observed variable p_i equals one when there is participation in FTFY work and zero otherwise.

$$ {w_i}^* = {z_i}'\gamma + \delta_i \tag{4} $$

$$ p_i = 1({w_i}^* > 0) \tag{5} $$

For the model to be identified, the vector z_i must contain at least one element not in x_i. Again, using George and Kuhn's methodology, we include the level of household income net of the individual's. It may reasonably be thought to affect the individual's reservation wage, but not the offered market wage.

Observing only the market wage, using equation (1), and that x_i is a subset of z_i (Wooldridge 2009, 609):

$$ E({w_i}^m | x_i, p_i = 1) = E({w_i}^m | x_i, z_i, \delta_i) \tag{6} $$

$$ = E({x_i}'\beta + \varepsilon_i | z_i, \delta_i) \tag{7} $$

$$ = {x_i}'\beta + E(\varepsilon_i | z_i, \delta_i) \tag{8} $$

However, since this is conditional on participation in FTFY work, we know that $w_i^* > 0$, and therefore $\delta_i > -z_i'\gamma$. So using the observed market wage to estimate the population wage structure, we have:

$$E(w_i^m | x_i, p_i = 1) = x_i'\beta + E(\varepsilon_i | \delta_i > -z_i'\gamma) \quad (9)$$

The rightmost term introduces bias, as it cannot be assumed to equal zero. The correction for this requires several assumptions. Let us assume the error terms ε_i, δ_i are independent of the explanatory variables and have a joint normal distribution, with means of zero and correlation ρ according to (10).

$$(\varepsilon_i, \delta_i) \sim N(0,0,\sigma_\varepsilon, 1, \rho) \quad (10)$$

The variance of δ_i is normalized to equal one as is standard for the probit selection equation (Greene 2003, 783). Under these assumptions, we have

$$E(\varepsilon_i | \delta_i > -z_i\gamma) = \rho\sigma_\varepsilon E(\delta_i | \delta_i > -z_i\gamma) \quad (11)$$

and, moreover, the assumption of the normal distribution gives us:

$$E(\delta_i | \delta_i > -z_i'\gamma) = \left[\frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)}\right] = \lambda(-z_i'\gamma) \quad (12)$$

The term in square brackets is the inverse of Mills ratio, that is, the ratio of the probability density to the cumulative normal distribution. It is denoted $\lambda(z_i'\gamma)$. Therefore, an unbiased estimate of the observed market wage, given only the nonrandom sample (Greene 2003, 784) is:

$$E(w_i^m | z_i, w_i^m \ observed) = x_i'\beta + \rho\sigma_\varepsilon \left[\frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)}\right] \quad (13)$$

$$= x_i'\beta + \rho\sigma_\varepsilon \lambda(-z_i'\gamma) \quad (14)$$

An estimate of γ is obtained from the probit specified by equation (4), run separately for the aboriginal group and the white group, using the entire sample of working age members for each. With $\hat{\gamma}$, we calculate for each observation i:

$$\hat{\lambda}_i = \lambda(-z_i'\hat{\gamma}) = \left[\frac{\phi(z_i'\hat{\gamma})}{\Phi(z_i'\hat{\gamma})}\right] \tag{15}$$

The coefficient of this term is expressed as a single parameter, $\theta = \rho\sigma_\varepsilon$, referred to as the selection coefficient. The resulting wage equation for FTFY workers is then:

$$w_i = x_i'\beta + \theta\lambda(-z_i'\gamma) + e_i \tag{16}$$

with $E(e_i) = 0$ even using only wage observations for FTFY workers. With the additional estimated term $\lambda(-z_i'\hat{\gamma})$, and the assumptions listed previously, the measured differentials and decomposition are now considered to more accurately represent the entire population.

Note that if ρ=0, the rightmost term disappears, and the market wage regression does not need the selection correction to consistently estimate β. So when the unobserved factors in the market and shadow wage equations are uncorrelated, the selectivity adjustment is unnecessary.

4.4 Decomposition of the Selection Term

Because the term correcting for possible sample selection bias, λ, is included as a regressor in the earnings regressions, it becomes part of the Blinder-Oaxaca decomposition. There exist several methods in the literature to accommodate it, none currently accepted as standard, but each suited to different assumptions of the model. Group differences in the selection term $\theta\lambda$ are, as shown previously, a combination of the difference in the inverse Mills ratio, λ, and the difference in the selection coefficient, θ. The decomposition now appears as follows:

$$\overline{w}^n - \overline{w}^a = (\bar{x}^n - \bar{x}^a)'\hat{\beta}^n + \bar{x}^{a\prime}\left(\hat{\beta}^n - \hat{\beta}^a\right) + \left(\hat{\theta}^n\bar{\lambda}^n - \hat{\theta}^a\bar{\lambda}^a\right) \quad (17)$$

For the other regression terms, it is straightforward to decompose each into a human capital component and a structural component. Using the previous terminology, these are the explained and unexplained terms in the decomposition equation (2). In the case of the group difference in the selection term $\hat{\theta}^n\bar{\lambda}^n - \hat{\theta}^a\bar{\lambda}^a$, however, it is not exactly clear how it breaks into those two parts. Neuman and Oaxaca (2004) review several methods and discuss their implications for the assessment of potential discrimination.

One method, which is noncommittal as to the endowment/potential discrimination components of the term $\hat{\theta}^n\bar{\lambda}^n - \hat{\theta}^a\bar{\lambda}^a$, simply defines it as a third component of the decomposition (Neuman and Oaxaca, 2004). Thus, the differential is split into three: a part explained by the difference in human capital endowments, a part due to the different coefficients of the earnings structure, and a part due to selectivity. It makes few assumptions about the nature of the selection term, and whether it can be attributed to observable characteristics or potential discrimination.

Another method, which essentially avoids decomposing the selectivity-adjustment term, subtracts the difference in correction terms from the wage differential. This leaves only the previously defined explained and unexplained terms as shown in (18).

$$(\overline{w}^n - \overline{w}^a) - (\hat{\theta}^n\bar{\lambda}^n - \hat{\theta}^a\bar{\lambda}^a) = (\bar{x}^n - \bar{x}^a)'\hat{\beta}^n + \bar{x}^{a\prime}(\hat{\beta}^n - \hat{\beta}^a) \quad (18)$$

This provides consistency with the model that was uncorrected for selectivity, in terms of the components of the decomposition. It allows a direct comparison of the contribution of each regressor in the corrected versus the uncorrected earnings differential. It has one obvious drawback, however: it does not decompose the observed wage differential (Neuman and Oaxaca, 2004). For this reason, this study will not use this method.

The method that this study does use, out of comparability to George and Kuhn (1994), assumes that differences in the inverse Mills ratio, as a function of FTFY-determining variables, are endowment effects, and the difference in the selectivity term is a difference in the earnings structure. The decomposition now appears as follows:

$$\bar{w}^n - \bar{w}^a = [(\bar{x}^n - \bar{x}^a)'\hat{\beta}^n + (\bar{\lambda}^n - \bar{\lambda}^a)\hat{\theta}^n] + [\bar{x}^{a\prime}(\hat{\beta}^n - \hat{\beta}^a) + \bar{\lambda}^a(\hat{\theta}^n - \hat{\theta}^a)] \quad (19)$$

Since the determinants of FTFY employment are the same human capital variables as those used in the earnings regressions, with the family income net of the individual's added, it seems reasonable to include that term in the part of the differential attributed to human capital endowments. Differences in the wage effects of selectivity, included in the unexplained portion of the differential, contribute to the differential attributed to possible discrimination (Neuman and Oaxaca, 2004). As before, these assumptions are subject to qualification: the group differences may be attributable to unmeasured productivity characteristics. Thus, any declaration of discrimination, especially when it is not agreed-upon how to accommodate the selectivity term, should be subject to further examination before any conclusion is reached.

4.5 Selection Model for Off-reserve Living

For aboriginals, there is a second potential form of sample selection bias. It comes from the exclusion of on-reserve aboriginals in the earnings regressions. As George and Kuhn (1994) show, wage structures are quite different on reserves, in particular the very low returns to education for on-reserve workers as compared to off-reserve workers. From an earnings regression for on-reserve, FTFY workers run on this sample with the 2006 Census Individual File data (the results are shown in Table B2 in the Appendix), the same pattern appears. Although there is not much difference in the returns to a high school diploma, post-secondary education shows less impact on wages. As well, the locational variables have different effects. For those with any aboriginal origins

residing off-reserve, living in a rural area has a positive effect on wages, but on-reserve, there is a negative effect. As well, the coefficients of the provincial controls are quite different from both off-reserve aboriginals and whites. These two observations might indicate a difference in the industries and occupations in which workers are employed on- and off-reserve, which would cloud the analysis of the wage differential. Therefore, this study includes only off-reserve aboriginals in the final sample.

The decision to live on or off-reserve is dependent on a fairly complex set of factors, and is likely subject to considerations beyond simply economic ones. It is beyond the scope of this study to examine the choice in its entirety, and instead, the decision to live off-reserve will be accounted for in another application of the Heckman method. It will use the same assumptions and framework used for the FTFY selection correction, with a few differences. The observed variable in the probit equation is an indicator of off-reserve living. The probability of living off-reserve is a function of the regressors of the earnings regression, plus the necessary identifying variable. The variable in this case should be a determinant of the decision to live on a reserve, but not affecting the reservation or market wages. As used by George and Kuhn (1994), this paper uses a variable to indicate if an individual's mother tongue was an aboriginal language. The definition of mother tongue as used here, according to the Census Dictionary, is "the first language learned at home in childhood and still understood by the individual at the time of the Census," (Statistics Canada, 2010, p. 83), and can be an indicator of a preference for living on-reserve.

The estimates from this probit will, again, be used to construct an inverse Mills ratio, λ_R, as in the FTFY selection model. It is included as a regressor for the aboriginal earnings regression. While the earnings function for the white subsample remains the same as equation (1), for aboriginals it now appears as follows:

$$w_i = x_i'\beta + \theta_R\lambda_R(-r_i'\vartheta) + e_i \tag{20}$$

In equation (20), r_i is the vector x_i plus the indicator of an aboriginal mother tongue, and ϑ the vector of coefficients estimated in the probit equation for off-reserve living. This additional regressor, again, permits us to estimate the $\hat{\beta}_j$ of the earnings regression consistently, and can now more accurately represent the entire aboriginal population working full-time and full-year. The correction for FTFY employment is dropped in this model, as it is not possible to correct for two separate selection effects using the Heckman method as shown here.

4.6 Decomposition of the Selection Term for Off-reserve Living

The selection term will be treated similarly in this implementation of the Blinder-Oaxaca decomposition to the selection term correcting for FTFY employment. The coefficient will be treated as part of the earnings structure, and the value of the inverse Mills ratio will be treated as an endowment. The difference, however, between this model and the previous one, is that the selection term appears only in the aboriginal wage regressions. So the decomposition appears as follows:

$$\bar{w}^n - \bar{w}^a = [(\bar{x}^n - \bar{x}^a)'\hat{\beta}^n] + [\bar{x}^{a\prime}(\hat{\beta}^n - \hat{\beta}^a) - \hat{\theta}_R\bar{\lambda}_R] \quad (21)$$

The part explained by human capital characteristics is in the first set of square brackets, and the part unexplained, therefore potentially attributed to discrimination, is in the second set. Notice that the explained component is *unchanged* from that in the uncorrected decomposition. It is because the coefficients of the white wage structure are used as the norm, those that would exist in the absence of discrimination. This model, then, assumes that the selection effect for living off-reserve affects the relative contribution of each regressor to the unexplained component of the differential, but it does not affect the make-up of the explained component. Therefore, this selection term provides an explanation for how the wage structure differs for aboriginal and non-aboriginal FTFY workers, since only the aboriginal workers face the decision to live on or off a reserve.

5. OLS Earnings Regressions

The structure of earnings without any selectivity adjustment, determined by the ordinary-least-squares regressions described in the previous section, is displayed in Table 3. The sample here is male Canadians, aged 25-64, living off-reserve and employed full-time and full-year. The three categories are white (non-aboriginal and non-visible minority), those with any aboriginal origins, and those with aboriginal origins only. The regressors are the same for all three groups, except that in the regressions for the 'any aboriginal origins' group, an indicator variable was used to designate those with mixed aboriginal origins. The omitted categories for the dummy variables are as follows: residing in Ontario and an urban area, never married, having less than a high school education, and speaking English only.

TABLE 3: OLS EARNINGS REGRESSIONS, FULL-TIME, FULL-YEAR WORKERS. (DEPENDENT VARIABLE: LOG OF ANNUAL WAGES)

	White N=657 714	Any Aboriginal Origins N=31 279	Aboriginal Origins Only N=5 926
Prince Edward Island	-0.2269*** (29.29)	-0.1532*** (4.89)	-0.1471*** (2.99)
Newfoundland	-0.3242*** (28.91)	-0.3281*** (4.30)	-0.2455 (0.75)
Nova Scotia	-0.2331*** (44.67)	-0.2254*** (8.63)	-0.1731*** (2.95)
New Brunswick	-0.2429*** (45.65)	-0.2549 (7.28)	-0.2643*** (3.16)
Quebec	-0.1281*** (30.07)	-0.0664*** (3.57)	0.0845** (2.34)
Manitoba	-0.1551*** (32.45)	-0.1278*** (8.10)	-0.2245*** (4.52)
Saskatchewan	-0.1281*** (23.42)	-0.1289*** (7.29)	-0.2643*** (5.22)
Alberta	0.0765*** (24.08)	0.1292*** (9.13)	0.0781** (2.13)
British Columbia	-0.0340*** (11.06)	-0.0009 (0.64)	-0.6262* (1.71)
Age	0.0860*** (123.05)	0.08334*** (23.02)	0.0759*** (9.24)
Age^2/100	-0.0912*** (109.12)	-0.0910*** (20.23)	-0.0809*** (8.01)
Rural	-0.0315*** (13.15)	0.0261** (2.54)	-0.0012 (0.05)
Married	0.2509*** (106.60)	0.2817*** (25.11)	0.2674*** (9.85)
Widowed/separated/ Divorced	0.0981*** (26.77)	0.0893*** (5.15)	0.0915** (2.52)
High school diploma	0.1331*** (41.63)	0.1323*** (9.50)	0.1234*** (4.07)
Non-university Training	0.2001*** (59.93)	0.2041*** (14.21)	0.2253*** (7.53)
Some post-secondary	0.3002*** (93.81)	0.2818*** (25.11)	0.2484*** (7.42)
Bachelor's degree	0.5267*** (140.70)	0.4842*** (24.68)	0.4749*** (10.77)
Post-graduate degree	0.6024*** (129.20)	0.5547*** (20.14)	0.5404*** (5.09)
French spoken only	-0.0654*** (14.03)	-0.0688*** (2.83)	-0.0242 (0.54)
Bilingual	0.0204*** (6.06)	0.0367** (2.54)	0.0613* (1.78)

Neither official language spoken	-0.2859*** (8.56)	0.2897* (1.75)	0.2318 (1.36)
Mixed aboriginal ancestry	-	0.0939*** (7.78)	-
Intercept	8.5251*** (619.67)	8.4246*** (120.78)	8.5887*** (54.19)
	R^2=0.1988 F(22,N)=5868.68	R^2=0.1813 F(23,31 255)=229.00	R^2=0.1519 F(22,5903)=36.26

Note: Figures in parentheses are absolute values of coefficient t-ratios.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

The results are, in general, consistent with the previous literature on earnings structure based on human capital. Age, representing work experience, has a positive effect on wages, but decreasingly so. The turning point is approximately 46 to 47 years old for all the groups. Marriage is associated with higher earnings for all groups. Being widowed, separated or divorced has a lesser effect but is still linked to higher earnings than the never-married. These findings generally uphold those of George and Kuhn (1994), whose OLS estimates are reported in Table B1 in the appendix. The returns to education steadily increase with each higher level attained. The return to a high school diploma is almost identical across groups. A post-secondary certificate below the bachelor's level, a bachelor's degree, and a post-graduate degree (including professional programs) have slightly higher returns for the white group than the two aboriginal groups. Non-university training has the highest returns for the aboriginal origins only group. This is different from the structure reported by George and Kuhn (1994). The aboriginal group in their study had lower returns to each of these levels of education.

Geographical differences appear, and are common across groups. Wages are highest in Alberta and Ontario, with Newfoundland and New Brunswick the lowest. Living in a rural area has a negative effect on whites' earnings, but a positive effect for those with any aboriginal origins. Living in a rural area is found to have no significant effect on wage earnings for those with only aboriginal origins. This is perhaps due to the small sample size for this group, as this regression is restricted to

the population living off-reserve. These regional effects are similar to George and Kuhn (1994), and it is notable (although inconsequential to this paper) that the highest-earning provinces have stayed somewhat stable over the past twenty years.

Language effects found are similar to those of George and Kuhn. For all three groups, bilingualism has a small but statistically significant positive effect on wages. Speaking only French has a negative effect on the wages of the white and any aboriginal origins groups. Possibly because of small cell sizes, this effect is not statistically significant for the aboriginal origins only group. Speaking neither official language has a statistically significant and negative effect on wages for the white group, while the effect is, unexpectedly, positive and significant at the 10 percent level for the any aboriginal origins group. For the group with aboriginal origins only, the effect is also positive although not statistically significant. This may indicate the demand for those who speak aboriginal languages in the workplace, although it is surprising that English or French would not be required as well.

Table 4 lists the observed annual log earnings for each group, along with the unadjusted differential between it and the predicted majority (white) log earnings and, for comparison, the findings of George and Kuhn (1994). The earnings differentials presented in the third row of Table 4 are calculated according to equation (3), using the raw differentials in the second row. The log earnings gap is 16.4 percent for those with any aboriginal origins, and 28.0 percent for those with only aboriginal origins. These are slightly higher figures than those found by George and Kuhn, who find a gap of 14.0 percent for the former group and 24.2 percent for the latter. These numbers are just the starting point, however, in the analysis of the earnings differential, and will be subjected to further decomposition and examination in the next section.

TABLE 4: OBSERVED LOG EARNINGS AND WAGE GAPS, BY GROUP

	Whites	Any Aboriginal Origins	Aboriginal Origins Only
Observed ln(earnings)	10.776	10.597	10.447

Differential (in ln(earnings))	-	0.179	0.329
Differential, as a percentage of white earnings	-	16.4	28.0
Differential found by George and Kuhn (1994)	-	14.0	24.2

6. Decomposition of the Earnings Differential

As described previously, the gap in earnings can be broken down into a component that is explained by the differences in endowments between the two groups, and a part that is due to the wage structure or how those endowments are remunerated by the Canadian labour market. This technique is first applied to the gap between whites and those with any aboriginal origins. The set of coefficients must be the same across the two groups, otherwise the amount unexplained can be inflated by the omission of a regressor. Therefore, a new structure for the 'any aboriginal origins' group must be estimated with the dummy variable indicating mixed aboriginal and non-aboriginal ancestry omitted. The regression results for this new earnings structure are very similar to the previous one, and are displayed in Table B3 in the Appendix.

TABLE 5: DECOMPOSITION OF THE EARNINGS DIFFERENTIAL

	Earnings differential (percent)	**Fraction explained (reference earnings structure: white)**	**Earnings differential after adjustment for observable characteristics**
Any aboriginal origins	16.4	0.56	7.2
Aboriginal origins only	28.0	0.53	13.2

Note: The differentials in Columns 1 and 3 are expressed as a percentage of white log earnings.

When the white earnings structure is used as the norm, 56 percent of the differential can be attributed to differences in characteristics (explained), while 44 percent is due to differences in the earnings structure (unexplained). The explained and unexplained components are the first and second terms, respectively, on the right-hand side of equation (2), divided by the term on the left-hand side. The counterfactual idea implied by this method suggests that a white person given the characteristics of the average person with any aboriginal origins in the sample would see approximately a 9.2 percent reduction in log earnings (ie., 56 percent of the gross earnings differential of 16.4 percent), eliminating 56 percent of the wage gap. After adjustment for all observable productivity characteristics, the any aboriginal-white earnings differential remaining is 7.2 percent (ie. the gross differential 16.4 percent minus 9.2 percent).

When the aboriginal sample is restricted to those with only aboriginal origins, the proportion of the gap explained by endowments is quite similar. When the white wage structure is used, 53 percent is due to differences in characteristics, and 47 percent to differences in coefficients. So, a male with solely aboriginal origins and the mean human capital endowments of his group would be expected to earn 13.2 percent less than an identically-endowed white worker.

The earnings differential, according to this analysis, is considerably larger for the more strictly defined aboriginal sample, whether adjusted or unadjusted for observable characteristics. The raw differential is 16.4 percent for those with any aboriginal origins versus 28.0 percent for those with aboriginal origins only. After adjusting for differences in productivity-related characteristics, the remaining wage gaps are 7.2 percent and 13.2 percent, respectively. This indicates the fraction of each total differential that can be attributed to possible discrimination is close to one-half. Although it is possible that the unexplained share of the earnings differential is due to unmeasured productivity characteristics, it is conceivable that at least some could be due to labour market discrimination.

George and Kuhn (1994) find a similar amount explained by observable characteristics. For the group with any aboriginal origins, the amount explained is 51 percent, compared to 56 percent observed here. The adjusted gap they find for this group is 6.8 percent, very close to the value of 7.2 percent found in this study. They find that the differential is 59 percent explained for those with aboriginal origins only, compared to 53 percent in this study. The earnings gap remaining after adjustment for observable characteristics is 10.1 percent in their study, which is somewhat less than the adjusted gap of 13.2 percent in this present study. This suggests that there has been little change in the proportion of the differential due to productivity-related characteristics, and the earnings gap attributed to discrimination has perhaps increased.

TABLE 6: DECOMPOSITION OF THE EXPLAINED COMPONENT OF THE EARNINGS DIFFERENTIAL

	Fraction attributed to education	Fraction attributed to experience	Contribution to total differential	Earnings differential after adjustment (percent)
Any aboriginal origins	0.302	0.158	0.255	12.2
Aboriginal origins only	0.268	0.089	0.189	22.8

Note: The differential in Column 4 is expressed as a percentage of white log earnings.

The explained component of the differential can be broken down further, as mentioned before. Of particular interest are the amounts attributed to education and work experience, as they are human capital factors that receive a significant amount of policy attention and (in the case of education) funding. For the any aboriginal origins group, education accounts for 30.2 percent of the explained component and potential work experience (using age as a proxy, and its square to capture a possible non-linear effect) accounts for 15.8 percent (shown in Table 6). This indicates that 46.0 percent of the explained gap, and therefore 25.5 percent of the total gap, could potentially be erased by increasing the education and work experience levels of the aboriginal population to those of the white population. This counterfactual conclusion is clearly made for illustrative purposes only, but it offers some insight. For the aboriginal origins only group, the same

decomposition is shown in Table 6. For both groups, the earnings differential decreases a significant amount with this hypothetical change in these characteristics, to 12.2 percent and 22.8 percent for the any aboriginal origins group and the solely aboriginal origins group, respectively.

7. Selection into Full-time, Full-year Work

The first selection effect that will be examined in this study is that into full-time and full-year work. The sample from which the FTFY workers are selected consists of the working-age males in each of the three previously-used categories, still subject to the exclusion of those on-reserve, visible minorities, and the self-employed. The sample means of the regressors of the probit equation are presented in Table 7. Again, the "mixed aboriginal origins" category (the remainder of the 'any aboriginal origins' group with the 'aboriginal origins only' group removed) has been included for information, although it is not used by itself in the earnings regressions.

Many of the differences from the FTFY subsample are consistent across all groups. For instance, there is a higher proportion with less than a high school diploma. As well, the proportion who live in a rural area is higher. The mean age is also lower for this more inclusive group. The fraction of the group participating in FTFY employment varies from almost two-thirds, for the white category, to just under a half, for those with solely aboriginal origins. While these numbers have a certain amount of descriptive power for the differing characteristics of the FTFY-employed and the entire working age samples, the selection model will use them to assess the determinants of the decision to participate FTFY, and what effect this might possibly have on observed wages.

TABLE 7: SAMPLE MEANS, CANADIAN MALES AGED 25-64, LIVING OFF-RESERVE. (2006 CANADIAN CENSUS)

	White	Any Aboriginal Origins	Aboriginal Origins Only	Mixed Aboriginal Origins
PEI/ Newfoundland	0.0239	0.0295	0.0453	0.0250
Nova Scotia	0.0335	0.0308	0.0187	0.0343
New Brunswick	0.0289	0.0228	0.02	0.0236
Quebec	0.2775	0.1761	0.2243	0.1622
Ontario	0.3565	0.2757	0.2225	0.2910
Manitoba	0.0323	0.0951	0.0867	0.0975
Saskatchewan	0.0261	0.0639	0.0762	0.0604
Alberta	0.1116	0.1616	0.139	0.1681
British Columbia	0.1097	0.1445	0.1673	0.1379
Age	41.449	38.35	38.251	38.3785
Rural	0.2019	0.2474	0.2635	0.2428
Never married	0.3687	0.4641	0.5485	0.4398
Married	0.5207	0.4156	0.3227	0.4423
Widowed/Separated/Divorced	0.1106	0.1203	0.1288	0.1179
Less than high school	0.1424	0.2354	0.3356	0.2066
High school diploma	0.2554	0.2675	0.2474	0.2733
Non-university training	0.1733	0.1928	0.191	0.1933
Some post-secondary	0.2349	0.2081	0.1717	0.2186
Bachelor's degree	0.1274	0.0665	0.0387	0.0745
Post-graduate degree	0.0666	0.0297	0.0156	0.0338
English spoken only	0.6424	0.7123	0.7521	0.7009
French spoken only	0.1322	0.0727	0.1024	0.0642
Bilingual	0.224	0.2107	0.1429	0.2302

Neither official language spoken	0.0014	0.0043	0.0026	0.0048
Native language	-	0.0378	0.1417	0.0079
Mixed aboriginal origin	-	0.8122	-	-
In FTFY employment	0.6579	0.5569	0.4747	0.5805
N	1 011 646	57 105	12 753	44 352

Probit results from the first step of the Heckman two-step correction are presented in Table 8. The regressors are the same as the previous earnings structure regressions, with an identifying variable added (log of family income of the net of the individual's). The sample for this regression is all workers, both FTFY and non-FTFY, hence the larger sample size. Age displays the same effect on participation as it does on wage: it increases the likelihood of being employed FTFY, however, the squared age term is negative, so this effect decreases after a certain point. This point is approximately 42 to 43 years for the three groups. Living outside of a metropolitan area decreases the probability of FTFY employment. Across the three groups, education has a positive effect on working FTFY. Non-university training has the least impact on such employment. The impact of education generally increases with each higher level of schooling attained, with the exception of the post-graduate level. Also important to note is that the effects of education, while displaying this same pattern for the three groups, are highest for the only aboriginal origins group, followed by the any aboriginal origins group, and lowest for the non-aboriginal group. This finding is also consistent with Frenette (2011), who, using the same cycle of the Canadian Census, finds that "additional schooling is generally associated with a larger decline in the probability of being unemployed for aboriginal people than non-aboriginal people" (p. 3). Overall, it appears that the determinants of full-time, full-year participation in the labour force are similar for aboriginal and non-aboriginal males, as evidenced by the similarities in the probit results.

TABLE 8: PROBIT RESULTS FOR FULL-TIME, FULL-YEAR WORK

	White N=1 011 646	Any Aboriginal Origins N=57 105	Aboriginal Origins Only N=12 753
Prince Edward Island	-0.6010*** (62.09)	-0.5971*** (19.59)	-0.5328 (9.81)
Newfoundland	-0.3280*** (17.67)	-0.5780*** (4.72)	-0.5019* (1.78)
Nova Scotia	-0.1797*** (23.77)	-0.1405*** (4.08)	0.0224 (0.23)
New Brunswick	-0.2598*** (31.97)	-0.2892*** (7.92)	-0.3477*** (3.63)
Quebec	-0.0478*** (8.42)	-0.1340*** (6.32)	-0.1554*** (3.62)
Manitoba	0.0609*** (7.78)	-0.0636*** (3.05)	-0.2749*** (5.47)
Saskatchewan	-0.0249*** (2.92)	-0.2300*** (10.02)	-0.3159*** (6.62)
Alberta	-0.0075 (1.63)	-0.0652*** (3.77)	-0.1682*** (4.15)
British Columbia	-0.1559*** (34.14)	-0.1898*** (10.30)	-0.2874*** (7.13)
Age	0.1620*** (190.28)	0.1050*** (29.49)	0.0923*** (12.35)
Age^2/100	-0.1934*** (194.60)	-0.1254*** (28.95)	-0.1062*** (11.60)
Rural	-0.1272*** (37.89)	-0.2259*** (18.41)	-0.1748*** (6.64)
Married	0.4100*** (115.01)	0.4667*** (33.97)	0.3552*** (12.47)
Widowed/Separated/Divorced	0.1177*** (23.64)	0.1200*** (6.23)	0.0582 (1.46)
High school diploma	0.1820*** (43.01)	0.2977*** (19.40)	0.3550*** (11.38)
Some post-secondary	0.2639*** (59.20)	0.3143*** (19.12)	0.2761*** (7.88)
Bachelor's degree	0.2882*** (55.09)	0.4547*** (17.81)	0.5662*** (8.36)
Post-graduate degree	0.2537*** (39.90)	0.4208*** (11.86)	0.6224*** (6.05)
Non-university training	0.1411*** (30.42)	0.1976*** (12.12)	0.2116*** (6.58)
French spoken only	-0.1140*** (17.14)	0.0151 (0.53)	0.0011 (0.02)
Bilingual	-0.0602*** (12.44)	0.0261 (1.56)	0.0824** (1.98)
Neither official	-0.4306***	-0.0270	0.0077

language spoken	(13.08)	(0.22)	(0.06)
Log of family income net of individual's	-0.0516*** (58.34)	-0.0455*** (12.55)	-0.0423*** (5.58)
Intercept	-2.4880*** (123.31)	-1.6783*** (20.65)	-1.5671*** (9.25)

Note: Figures in parentheses are absolute values of coefficient t-ratios.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

The identifying variable, log of family income net of the individual's income, is highly significant in predicting the probability of FTFY participation for all three subsamples. It has a negative effect on the likelihood of FTFY employment, as expected. George and Kuhn (1994) find it significant for all groups, even with a much smaller sample size.

The sample mean of the inverse Mills ratio, $\boldsymbol{\lambda}$, and the estimated selection coefficient, $\widehat{\boldsymbol{\theta}}$, for each group is presented in Table 9. The estimated selection coefficient is negative but is not statistically significant for any of the groups. As stated before, since the selection coefficient is the product of the variance of the error term of the shadow wage equation, which is always positive, and the correlation coefficient of the errors, it indicates the direction of correlation. A negative value indicates that the unobserved factors that cause an increase in the likelihood of participation will cause a decrease in the market wage received. This implies that the restriction of the sample for the earnings regression to only the FTFY workers, without any sort of correction, might underestimate the earnings of the entire population of working-age males.

TABLE 9: CORRECTION FOR SELECTION INTO FULL-TIME, FULL-YEAR WORK

	White	**Any Aboriginal Origins**	**Aboriginal Origins Only**
Mean of inverse Mills ratio, λ	0.5245	0.6553	0.7693
Selection coefficient, θ	-0.0255 (1.28)	-0.0152 (0.15)	-0.1363 (0.61)

Note: Figures in parentheses are absolute values of coefficient t-ratios.

George and Kuhn (1994) find a similar result to this study: the selection coefficients are statistically insignificant and estimated to be negative. This runs counter to the intuitive argument that they postulate, and instead implies that "the sample of workers employed full-year, full-time is thus estimated to be on average somewhat *less* able on dimensions that are unobserved by the econometrician that the rest of the population" (p. 33). The coefficients of the earnings structure, implementing the Heckman selectivity adjustment with respect to FTFY work, are shown in Table 10. Compared to the unadjusted results shown in Table 3, there are very few changes.

TABLE 10: EARNINGS REGRESSION RESULTS WITH CORRECTION FOR SELECTION INTO FULL-TIME, FULL-YEAR WORK

	White N=657 714	Any Aboriginal Origins N=31 279	Aboriginal Origins Only N=5 926
Prince Edward Island	-0.2186*** (22.23)	-0.1477*** (3.11)	-0.1003 (1.14)
Newfoundland	-0.3198*** (27.31)	-0.3231*** (3.87)	-0.2062 (0.62)
Nova Scotia	-0.2309*** (42.31)	-0.2243*** (8.46)	-0.1751*** (2.96)
New Brunswick	-0.2396*** (40.75)	-0.2524*** (6.77)	-0.2342** (2.55)
Quebec	-0.1196*** (29.82)	-0.0653*** (3.45)	-0.0714* (1.78)
Manitoba	-0.1560*** (32.26)	-0.1273*** (8.06)	-0.2026*** (3.48)
Saskatchewan	-0.1280*** (23.40)	-0.1269*** (6.02)	-0.2381*** (3.74)
Alberta	0.0765*** (24.09)	0.1298*** (9.12)	0.0918** (2.22)
British Columbia	-0.0321 (9.46)	-0.0077 (0.46)	-0.0393 (0.81)
Age	0.0836*** (43.17)	0.0823*** (10.72)	0.0675*** (4.19)
Age^2/100	-0.0883*** (38.59)	-0.0898*** (9.79)	-0.0713*** (3.80)
Rural	-0.0298*** (11.00)	0.0282* (1.74)	0.0143 (0.40)
Married	0.2460*** (55.28)	0.2780*** (9.80)	0.2406*** (4.72)
Widowed/Separated/Divorced	0.0964*** (24.55)	0.0882*** (4.50)	0.0862** (2.31)
High school diploma	0.1306*** (35.23)	0.1295*** (5.85)	0.0928 (1.54)
Non-university training	0.1981*** (54.02)	0.2022*** (10.79)	0.2062*** (4.93)
Some post-secondary	0.2967*** (70.66)	0.2606*** (10.95)	0.2242*** (4.12)
Bachelor's degree	0.5229*** (108.85)	0.4803*** (14.24)	0.4281*** (4.85)
Post-graduate degree	0.5990*** (111.98)	0.5510*** (15.28)	0.4913*** (4.06)
French spoken only	-0.0639*** (13.27)	-0.0689*** (2.84)	-0.2499 (-0.56)
Bilingual	0.0211*** (6.19)	0.0365** (2.54)	0.0544 (1.47)
Neither official	-0.2796***	0.2902*	0.2334

language spoken	(8.29)	(1.75)	(1.37)
Intercept	8.5896*** (167.07)	8.4573*** (36.39)	8.8789*** (17.46)
	R^2=0.1988 F(23,657690)=5616.88	R^2=0.1813 F(24,31254)=219.55	R^2=0.1520 F(23,5902)=34.82

Note: Figures in parentheses are absolute values of coefficient t-ratios.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

The statistical insignificance of the selection coefficient could indicate that the probit equation that examines the determinants of FTFY work is misspecified, and this model does not capture the selection effect into FTFY work. It could, however, support the notion that there is no significant unobserved difference between the FTFY workers and the rest of the working-age population, and thus examining the FTFY group alone captures the desired results.

TABLE 11: DECOMPOSITION OF THE EARNINGS DIFFERENTIAL, CORRECTED FOR SELECTION INTO FULL-TIME, FULL-YEAR EMPLOYMENT

	Fraction explained (using white earnings structure)	**Earnings differential after adjustment for observable characteristics, selectivity-corrected**	**From Table 5: Earnings differential after adjustment for observable characteristics, uncorrected**
Any aboriginal origins	0.57	7.0	7.2
Aboriginal origins only	0.54	12.8	13.2

Note: The differentials in Column 2 and 3 are expressed as a percentage of white log earnings.

The addition of the new regressor λ for the selectivity correction slightly increases the fraction of the earnings differential that is explained by the productivity-related characteristics, as shown in the first column of Table 11. Using the white earnings structure as reference, the unexplained component of the Blinder-Oaxaca decomposition in equation (2) decreases slightly to 43 percent, for those with any aboriginal origins. As such, a male with any aboriginal origins working full-time, full-year with human capital characteristics at the mean levels for his group,

could possibly earn 7.0 percent less than a white male with those same characteristics, that is, of the representative FTFY worker with any aboriginal origins. This gap is the difference in the coefficients of the wage structure for whites and aboriginals, multiplied by the means of the characteristics of a representative FTFY worker with any aboriginal origins. The wage structures used here are corrected, by including the inverse Mills ratio (λ) as a regressor, to represent part-time and part-year employees in addition to the FTFY group used previously.

The adjusted earnings gap for the group with solely aboriginal origins is also reduced slightly by the selection correction, to 12.8 percent of log earnings. However, as stated before, the error in the estimation of the selection coefficient θ was too high to consider it statistically significant, for each group and therefore, this may not be a reliable measure of the effects of selection into FTFY work for this group. This is the same problem that George and Kuhn (1994) faced.

TABLE 12: DECOMPOSITION OF THE EXPLAINED COMPONENT OF THE EARNINGS DIFFERENTIAL, CORRECTED FOR SELECTION INTO FULL-TIME, FULL-YEAR EMPLOYMENT

	Amount attributed to education	Amount attributed to experience	Contribution to total differential	Earnings differential after adjustment
Any aboriginal origins	0.527	0.276	0.457	8.9
Aboriginal origins only	0.490	0.163	0.354	18.1

Note: The differential in Column 4 is expressed as a percentage of white log earnings.

Table 12 shows the contributions of education and work experience to the explained component of the differential. The fractions are similar to the uncorrected results shown in Table 6, with approximately half explained by education for both groups, approximately a quarter explained by potential work experience for the group with any aboriginal origins and around 16 percent for the group with aboriginal origins only. This suggests that, although the selectivity correction has changed the structure of earnings to account for the self-selection into FTFY work, it has little impact on the relative contribution of these particular factors. The same conclusions can be drawn

regarding the possible impact of increasing the levels of these variables. Education and potential work experience account for a large proportion of the differential, and a great reduction in the wage gap appears to be possible with an increase in these variables in the aboriginal population to the level of the non-aboriginal population.

If we are to assume the model is correctly specified, the statistical insignificance of the selectivity coefficient for any of the three groups analyzed in this study indicates several things. It can be said that the unobserved productivity characteristics (eg., ability) are neutral in the decision to work FTFY, that is, they have the same effect on the market wage and shadow wage. Therefore, the subsample of FTFY workers is not found to have significantly different unobserved productivity characteristics from the entire working-age population. If this correlation were found in the data, an individual's decision to work FTFY indicates that he would be more likely to earn a higher wage, even part-time or part-year. Since there is no evidence that there is a correlation in the unobserved productivity characteristics and unobserved factors in the decision to work FTFY, this indicates that the earnings of the FTFY workers are not inherently likely to have higher earnings than those of the working-age population. This reflects positively on the representativeness of our sample. This study analyzes the wage structure for FTFY workers only, because of the availability of earnings data. However, the results here suggest that unobserved productivity characteristics have a neutral effect on the selection into that group and so the estimated wage structure is unbiased in that respect. This also suggests that the size and structure of the differential is representative of the entire working-age population, since correcting for this selectivity effect has only a negligible effect on the earnings structure, and therefore a small effect on the decomposition of the earnings differential.

8. Selection into Off-reserve Living

This section examines the possible bias introduced by examining only the earnings of the members of the aboriginal groups who live off-reserve. Selection off reserves is likely to be nonrandom and this would render the wage structure and decompositions unrepresentative of the general aboriginal population.

OLS estimates of a basic earnings function for the on-reserve population working full-time and full-year are given in Table B2 in Appendix B. The mean log earnings on-reserve are lower than those off-reserve by 36.1 percent.[1] The coefficients are also somewhat different from the off-reserve regression for the same group (presented previously in Table 3). The returns to a high school diploma are almost the same, the returns to a bachelor's degree are slightly higher on reserves, and the returns to non-university training and a post-graduate degree are lower on-reserve.

When correcting for the possible bias arising from omitting the population on reserves, we drop the correction for selection into FTFY work. This is in order to prevent the two effects from possibly confounding one another. As a result, we also drop the correction for the comparison group (ie. white FTFY workers) when decomposing the wage gap. The estimates from the probit model for living off-reserve (for FTFY workers) are presented in Table 13.

[1] However, it must be stressed that income tax is not paid on these earnings, and this will likely affect the wages offered. In addition, it may be more instructive to compare with the off-reserve earnings of the aboriginal origins only group, as the on reserve population is almost entirely in that category. The earnings gap in that case is 24.5 percent.

TABLE 13: PROBIT RESULTS FOR THE PROBABILITY OF LIVING OFF-RESERVE, FOR FULL-TIME, FULL-YEAR WORKERS WITH ANY ABORIGINAL ORIGINS

	Any Aboriginal Origins N=47 283	Aboriginal Origins Only N=20 250
Prince Edward Island	-1.761***	2.163***
	(20.32)	(21.94)
Newfoundland	-1.630***	0.0004
	(17.56)	(0.00)
Nova Scotia	-1.992***	-0.2967***
	(20.86)	(3.82)
New Brunswick	-1.474***	-0.3563***
	(17.17)	(4.30)
Quebec	-1.663***	0.7670***
	(19.97)	(17.96)
Manitoba	-1.774***	-0.1704***
	(20.92)	(3.92)
Saskatchewan	-1.700***	0.0056
	(19.90)	(0.13)
Alberta	-1.371***	0.2980***
	(16.29)	(7.81)
British Columbia	-2.264***	-0.4698***
	(26.90)	(12.79)
Age	-0.0194***	-0.0024
	(3.53)	(0.30)
Age^2/100	0.0206***	-0.0014
	(3.13)	(0.16)
Rural	-1.542***	-1.452***
	(91.05)	(60.07)
Married	0.2625***	-0.0228
	(14.17)	(0.88)
Widowed/Separated/ Divorced	0.2125***	0.1254***
	(7.50)	(3.16)
High school diploma	0.2144***	0.1444***
	(9.52)	(4.59)
Non-university training	0.0987***	0.0207
	(4.41)	(0.504)
Some post-secondary	0.0044	-0.1298***
	(0.20)	(4.01)
Bachelor's degree	0.2064***	0.1209**
	(5.21)	(1.99)
Post-graduate degree	0.2708***	0.1269
	(4.59)	(1.39)
French spoken only	0.0952**	-0.3962***
	(2.17)	(7.13)
Bilingual	0.6261***	0.0926**
	(20.61)	(1.99)

Neither official language spoken	0.5991*** (4.61)	0.5501*** (4.03)
Aboriginal language	-1.463*** (70.89)	-0.8520*** (31.80)
Intercept	1.7995*** (16.42)	0.7564*** (4.87)
	χ^2(23)=26821.85	χ^2(23)=7939.38

Note: Figures in parentheses are absolute values of coefficient t-ratios.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

The variable included for identification, having an aboriginal language as a mother tongue, significantly increases the individual's likelihood of residing on-reserve. The majority of reserves in Canada are outside of urban areas; therefore it is not surprising that living in a rural location has a strong negative effect on living off-reserve. For the group with any aboriginal origins, age displays a negative effect on the probability of living off-reserve up to age 47, then a positive effect. The effects of age are not statistically significant for the aboriginal origins only group. Every category of education attained above the omitted category (less than a high school diploma) except for 'Some post-secondary' displays a positive and significant effect on living off-reserve for the any aboriginal origins group. For those with solely aboriginal origins, a high school diploma and a bachelor's degree increase the probability of living off-reserve, whereas having 'Some post-secondary' increases the likelihood of residing on-reserve.

TABLE 14: CORRECTION FOR SELECTION OFF-RESERVE

	Any Aboriginal Origins	**Aboriginal Origins Only**
Mean of inverse Mills ratio, λ	0.2477	0.6561
Selection coefficient, θ	-0.1076*** (3.63)	-0.1186* (1.79)

*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.
Note: Figures in parentheses are absolute values of coefficient t-ratios.

Shown in Table 14, the selection coefficient in the off-reserve selectivity-adjusted earnings equation is negative and statistically significant for both groups (for the aboriginal origins only group, it is significant only at the 10% level). Therefore, there is a negative correlation between the unobserved factors in the decision to live off-reserve and the unobserved factors in the earnings structure. This suggests that a FTFY worker with a stronger unobserved preference for living on-reserve would have unobserved characteristics that would lead to higher earnings, whether on or off-reserve. Therefore, excluding the population on reserves from the earnings regressions is not likely to overstate the earnings of the representative aboriginal worker. It is more likely to understate the earnings. However, this is remedied by including the selection correction term. George and Kuhn find that, for males, the selection coefficient is not statistically significant, although their identifying variable (the same as used here) has a strong and statistically significant effect. They use, however, a smaller sample size, consisting of 136 living on-reserve and 932 off-reserve.

The estimates from the earnings regressions with the Heckman selectivity correction, shown in Table 15, are very similar to those without (seen in Table 3). The coefficients have the same sign for every variable, with the direction of the effects consistent with the human capital theory of wages. Some variables, for example speaking both or neither official languages, lose their statistical significance when the inverse Mills ratio is included, indicating its likely collinearity with the other regressors.

TABLE 15: EARNINGS REGRESSION RESULTS WITH CORRECTION FOR SELECTION OFF-RESERVE

	Any Aboriginal Origins N=31 279	Aboriginal Origins Only N=5 926
Prince Edward Island	-0.1981***	-0.2453***
	(6.11)	(3.30)
Newfoundland	-0.3282***	-0.2431
	(4.35)	(0.75)
Nova Scotia	-0.2278***	-0.1545***
	(8.72)	(2.59)
New Brunswick	-0.2448***	-0.2299***
	(6.91)	(2.66)
Quebec	-0.0734***	-0.1085***
	(3.93)	(2.89)
Manitoba	-0.1206***	-0.2020***
	(7.68)	(4.18)
Saskatchewan	-0.1225***	-0.2531***
	(7.08)	(5.06)
Alberta	0.1258***	0.0661*
	(8.84)	(1.77)
British Columbia	0.0093	-0.0328
	(0.60)	(0.82)
Age	0.0835***	0.0761***
	(23.04)	(9.29)
Age^2/100	-0.0912***	-0.0806***
	(20.27)	(7.96)
Rural	0.0896***	0.1096*
	(4.36)	(1.66)
Married	0.2809***	0.2695***
	(24.70)	(9.94)
Widowed/Separated/Divorced	0.0862***	0.0838**
	(4.94)	(2.29)
High school diploma	0.1298***	0.1116***
	(9.13)	(3.55)
Non-university training	0.2054***	0.2227***
	(14.23)	(7.41)
Some post-secondary	0.2697***	0.2566***
	(18.89)	(7.65)
Bachelor's degree	0.4882***	0.4721***
	(24.78)	(10.67)
Post-graduate degree	0.5578***	0.5364***
	(20.23)	(5.07)
French spoken only	-0.0794***	-0.0261
	(3.18)	(0.58)
Bilingual	0.0254	0.0442
	(1.61)	(1.23)
Neither official	0.2491	0.2066

language spoken	(1.53)	(1.21)
Intercept	8.5137*** (122.49)	8.6352*** (52.59)
	R^2=0.1796 F(23,31255)=34.82	R^2=0.1527 F(23,5902)=225.38

Note: Figures in parentheses are absolute values of coefficient t-ratios.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

The size of the differential and the proportions explained and unexplained by human capital characteristics using the Blinder-Oaxaca decomposition are exactly the same with or without the correction for the selection effect into living off-reserve. Therefore, the results for this model are identical to those in Table 5. As stated in the description of the methodology, using the white earnings structure as the norm, which does not have the Heckman correction for off-reserve selection, there is no change in the component that multiplies the vector of differences in the regressors by the white earnings coefficients. It is the other part of the differential that changes – the part consisting of the difference between the white and aboriginal values of each coefficient of the earnings regression, multiplied by the mean of the regressor for the aboriginal group. The relative contribution of each category of regressors (education, potential work experience, etc.) changes with the inclusion of the inverse Mills ratio as a regressor. Since the inverse Mills ratio is positive and the selection coefficient is negative, this term has a negative contribution to the differential – that is, it would have the effect of decreasing the size of the unexplained component. However, because the size of the unexplained component remains the same, the other regressors must have a greater effect. Note that all regressors are positive, so a negative contribution to the differential implies that the coefficient is greater in the white earnings structure. Therefore, the selection term partially accounts for the earnings structure of aboriginal FTFY workers living off-reserve rewarding them less than an identically-endowed white worker. As the sample in the earnings regressions is restricted to only those living off-reserve, and without including the inverse Mills ratio to correct for it, the results of the first set of earnings regressions for the aboriginal

groups may overstate the unexplained component and, therefore, the amount attributed to potential discrimination.

These results, while plausible, must be analyzed with caution, however. This is a very restricted view of the decision to live off-reserve, since it only examines the subsample of aboriginal FTFY workers. As well, the method depends on the mother tongue as being a reliable predictor of the preference for living on-reserve. There are certainly more factors that influence this decision, for example the unemployment rate relative to a possible place to live off-reserve. As well, there is the usual caveat when using Canadian Census data for on-reserve populations: there are a number of unenumerated reserves, and these data may misrepresent the full population. Therefore, with these considerations in mind, these results are perhaps not as reliable as would be ideal.

9. Summary and Conclusions

This study has examined the earnings differential between aboriginal and non-aboriginal, non-visible minority Canadian males working full-time and full-year in the labour market. The average differential was decomposed into two parts according to the Blinder-Oaxaca decomposition. The first part is that attributed to average human capital and locational characteristics, and the second part to differences in the earnings structure paid to the workers and therefore possibly discrimination. This paper uses the Master File of the 2006 Canadian Census and a 20 percent sample size to update a study by George and Kuhn (1994) that used the Public Use Sample Tape, with a 2 percent sample, of the 1986 Census.

The raw gap between off-reserve aboriginal and white log annual earnings is 16.4 percent, or 28.0 percent when the aboriginal group includes only those with strictly aboriginal ancestry. These

differentials are slightly larger than those found by George and Kuhn, who found gaps of 14.0 percent and 24.2 percent, respectively. Thus, this study finds that the aboriginal-white earnings gap has likely widened over this period.

The proportion of the differential that is explained by human capital and locational characteristics is similar for both definitions of aboriginality. Using the white earnings structure as the basis of comparison, 56 percent is attributed to differences in productivity-related characteristics for those with any aboriginal origins, and 53 percent for those with solely aboriginal origins. Therefore, there remains a gap of 7.2 percent (for those with any aboriginal origins) or 13.2 percent (for those with solely aboriginal origins) of log earnings that arises from differences in the earnings structure. This 'unexplained' residual captures unmeasured differences in productivity, but also may reflect some degree of discrimination in the wages and jobs offered.

When selection into the group of full-time and full-year labour force participants is accounted for, the amount explained by human capital and locational characteristics remains approximately the same. It rises to 57 percent for those with any aboriginal origins and 54 percent for those with only aboriginal origins. George and Kuhn (1994) find the amount explained stays very similar to their uncorrected results, at 50 percent for the group with any aboriginal origins and 58 percent for the group with aboriginal origins only. Their selectivity coefficient in their selectivity-adjusted earnings equations is found to be statistically insignificant, as the present study also finds, indicating that the full-time and full-year subsample does not differ in unobserved characteristics from the rest of the working-age population. Although the correction method is obviously an approximate way of accounting for individuals left out of the sample, it provides some insight. It indicates that the full-time, full-year participants in the labour force do not appear to have different unobserved characteristics in this respect than the rest of the working-age population with our larger and more recent data set. Adjusting the earnings structure to accommodate for these unmeasured differences between the full-time, full-year workers and the rest of the working-age

population causes no significant change in the amount that productivity-related characteristics account for the earnings gap between white and aboriginal workers. When the earnings of FTFY aboriginal males are examined independently of the decision to participate FTFY, their inherent, unmeasurable productivity characteristics may make it appear that they experience less discrimination. When these unobserved factors leading them to have higher earnings are corrected for, their productivity-related characteristics explain less of the differential. This indicates that further study of labour market discrimination against aboriginal Canadians would benefit from the analysis of a wider scale of employment participation.

The fraction of each group who participate in full-time, full-year work indicates the large number of workers who are excluded from this study. While almost two-thirds of the non-aboriginal, non-visible minority, male working-age population are employed full-time and full-year, less than half (48 percent) of the equivalent population with aboriginal origins only is employed FTFY, and 56 percent of the population with any aboriginal origins is so employed, shown in Table 6. This indicates the possibility of another form of labour market discrimination, separate from wage discrimination. Previous studies (Pendakur, 2010) postulate that job characteristics are at least as susceptible to ethnic discrimination as the wages paid to workers. It could be that aboriginal workers are channeled into seasonal work, or lower-paying industries.

Selection effects off-reserve were found to have a small but statistically significant effect. The selection coefficient is negative, suggesting a negative correlation between the unobserved factors in the decision to live off-reserve and in the market wage equation. Therefore, those with an inherent preference for living on a reserve are more likely to have a higher wage. The correction does not change the relative proportions of the differential explained and unexplained by observable human capital characteristics, however. This is because the white wage structure, which does not include the selection term, is used as the reference. George and Kuhn (1994) find that this selection effect is insignificant for males, with a smaller sample size. This selection model is only

applicable to FTFY aboriginal workers, and is therefore somewhat limited. Because of the high unemployment rates on reserves, there is likely some interplay between the decision to work FTFY and to live on-reserve, and this is another area of further research opportunity.

The results of this study suggest that there has not been a great deal of progress in eliminating the aboriginal-white earnings gap over the past twenty years. The raw earnings gap has, in fact, increased very slightly for the two definitions of aboriginal FTFY worker analyzed, those with any aboriginal origins and those with solely aboriginal origins. The amount of that gap unexplained by characteristics has decreased for the former group, from 49 to 44 percent, but increased for the latter, from 41 to 47 percent. As this is the component that can potentially be attributed to discrimination, these numbers do not provide evidence that wage discrimination is being eliminated in Canada.

The sizeable proportion of the differential that is attributed to human capital-related characteristics, especially to education, indicates a possible way to reduce the differential. The educational attainment of the aboriginal population has increased greatly since the previous study, which shows that perhaps some of the government's objectives in improving outcomes have worked. The proportion of aboriginal Canadians participating full-time and full-year in the labour force (as defined in this paper and that of George and Kuhn) has increased as well, from 36 percent in 1986 to 56 percent in 2006, another improvement. As various indicators show, progress has been made in some outcomes for aboriginal workers. However, there remains a large gap to be closed. It will require translating increases in the level of education and labour force participation into earnings, as well as, perhaps, eliminating wage and job discrimination.

References

Blinder, A. (1973). "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8, 436-455.

Chiswick, B.R. and J. Mincer (1972). "Time Series Changes in Personal Income Inequality in the United States From 1939, With Projections to 1985." *Journal of Political Economy* 88, 35-38.

Cotton, J. (1988). "On the Decomposition of Wage Differentials." *Review of Economics and Statistics* 70, 236-243.

Daly, A. (1994). "The Determinants of Employment Income for Indigenous Australians." CAEPR Discussion Paper no. 68, CAEPR, Canberra.

De Silva, A. (1999). "Wage Discrimination against Natives." *Canadian Public Policy* 25, 65-85.

Frenette, M. (2011). "Are the Labour Market Benefits to Schooling Different for Aboriginal and Non-Aboriginal People?" Canadian Labour Market and Skills Researcher Network Discussion Paper no. 79, CLSRN, University of British Columbia.

George, P. and P. Kuhn (1994). "The size and structure of native-white wage differentials in Canada." *Canadian Journal of Economics* 27, 20-43.

Greene, W. H. (2003). *Econometric Analysis*, 5th edition, (Upper Saddle River, NJ: Prentice-Hall Inc.).

Heckman, J. (1974). "Shadow prices, market wages and labor supply." *Econometrica* 42, 679-94.

Heckman, J. (1979). "Sample selection bias as a specification error." *Econometrica* 47, 153-161.

Hurst, M. (1997). "The Determinants of Earnings Differentials for Indigenous Americans: Human Capital, Location, or Discrimination?" *Quarterly Review of Economics and Finance* 37, 787-807.

Kuhn, P. and A. Sweetman (2002). "Aboriginals as unwilling immigrants: Contact, assimilation and labour market outcomes." *Journal of Population Economics* 15, 331–55.

Maddala, G.S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*, (Cambridge: Cambridge University Press).

Mincer, J. A. (1974). *Schooling, Experience, and Earnings*, (New York: Columbia University Press).

Minister of Indian Affairs and Northern Development and Federal Interlocutor for Métis and Non-Status Indians (2010). *A Progress Report on Aboriginal Initiatives from the Government of Canada (2009-2010).* (Ottawa: Indian and Northern Affairs Canada).

Minister of Supply and Services (1996). *Report of the Royal Commission on Aboriginal Peoples.* (Ottawa: Indian and Northern Affairs Canada).

Neuman, S. and R. Oaxaca (2004). "Wage decompositions with selectivity-corrected wage equations: A methodological note." *Journal of Economic Inequality* 2, 3-10.

Neumark, D. (1988). "Employers' Discriminatory Behaviour and the Estimation of Wage Discrimination." *Journal of Human Resources* 23, 279-95.

Oaxaca, R. (1973). "Male-female wage differentials in urban labor markets." *International Economic Review* 14, 693-709.

Pendakur, K. (2010). "Visible Minorities and Aboriginal Peoples in Vancouver's Labour Market." (Ottawa: Human Resources and Skills Development Canada).

Pendakur, K. and R. Pendakur (2011). "Aboriginal Income Disparity in Canada." *Canadian Public Policy* 37, 62-84.

Psacharopoulos, G., A. Fiszbein, H. Lee, S. Morley, and B. Wood. (1992). "Poverty and Income Distribution in Latin America: The Story of the 1980s." Report No. 27. World Bank, Latin America and the Caribbean Technical Department, Regional Studies Program, Washington, D.C.

Statistics Canada (1995). *Profile of Canada's Aboriginal Population*, Catalogue No. 94-325 (Ottawa: Statistics Canada).

Statistics Canada (2010). *2006 Census Dictionary* (Ottawa: Ministry of Industry).

Statistics Canada (2010). *Aboriginal Peoples Technical Report, 2006 Census, Second Edition* (Ottawa: Ministry of Industry).

Wood, B. and H.A. Patrinos. (1994). "Urban Bolivia" from *Indigenous People and Poverty in Latin America: An Empirical Analysis.* Psacharopoulos, G. and H. A. Patrinos (eds). World Bank, The International Bank for Reconstruction and Development , Washington, D.C.

Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach*, 7th ed. Mason, (Ohio: South-Western College Publishing).

Appendix A

Ethnic Origin Questions from the 2006 Canadian Census of Population (Statistics Canada, 2010)

17 What were the ethnic or cultural origins of this person's **ancestors**?

An ancestor is usually more distant than a grandparent.

For example, Canadian, English, French, Chinese, Italian, German, Scottish, East Indian, Irish, Cree, Mi'kmaq (Micmac), Métis, Inuit (Eskimo), Ukrainian, Dutch, Filipino, Polish, Portuguese, Jewish, Greek, Jamaican, Vietnamese, Lebanese, Chilean, Salvadorean, Somali, etc.

Following were eight write-in boxes.

18 Is this person an Aboriginal person, that is, North American Indian, Métis or Inuit (Eskimo)?

If " Yes", mark the circle(s) that best describe(s) this person now.

Possible Responses:

No

Yes, North American Indian

Yes, Métis

Yes, Inuit (Eskimo)

20 Is this person a member of an Indian Band / First Nation?

Possible Responses:

No

Yes, member of an Indian Band/First Nation.

21 Is this person a Treaty Indian or a Registered Indian as defined by the Indian Act of Canada?

Possible Responses:

No

Yes, Treaty Indian or Registered Indian

Appendix B

TABLE B1: GEORGE AND KUHN (1994) , USING 1986 CANADIAN CENSUS DATA

OLS EARNINGS REGRESSIONS, FULL-TIME, FULL-YEAR WORKERS. (DEPENDENT VARIABLE: LOG OF ANNUAL WAGES)

	White	**Any Aboriginal Origins**
Newfoundland	-0.1189**	-0.0724
	(0.0175)	(0.1554)
Nova Scotia	-0.1318**	-0.0022
	(0.0138)	(0.1705)
New Brunswick, Prince Edward Island	-0.1356**	-0.2850**
	(0.0127)	(0.1354)
Quebec	0.0779**	0.0018
	(0.0088)	(0.0860)
Manitoba	0.0727**	-0.0842
	(0.0114)	(0.0739)
Saskatchewan	0.0582**	-0.1960**
	(0.0131)	(0.0907)
Alberta	0.0500**	0.0776
	(0.0079)	(0.0648)
British Columbia	0.0212**	0.1335**
	(0.0076)	(0.0656)
Age	0.0841**	0.0523**
	(0.0015)	(0.0147)
$Age^2/100$	-0.0877**	-0.0547**
	(0.0018)	(0.0188)
Urban	0.0427**	0.0195
	(0.0046)	(0.0436)
Married	0.2646**	0.3696**
	(0.0066)	(0.0594)
Widowed/Separated/Divorced	0.1378**	0.2959**
	(0.0108)	(0.0917)
Grade 5-8	0.2085**	0.3209
	(0.0207)	(0.2064)
Grade 9-10	0.3366**	0.3111
	(0.0205)	(0.2034)
Grade 11-13	0.4668**	0.4368**
	(0.0203)	(0.2023)
University 1-4	0.6323**	0.6078**
	(0.0207)	(0.2096)
University 5+	0.7832**	0.6579**
	(0.0215)	(0.2242)
Training	0.0236**	0.0119
	(0.0018)	(0.0187)
French spoken only	-0.0422**	-0.0433
	(0.0110)	(0.1297)

Bilingual	0.0260** (0.0076)	-0.0602 (0.0666)
Neither official language spoken	-0.1999** (0.0403)	-0.2067 (0.4808)
Mixed aboriginal ancestry	-	0.0785** (0.0482)
N	65 705	932
R^2	0.2275	0.1747

Note: robust standard error in parentheses.
** indicates significance at 5% level.

TABLE B2: OLS EARNINGS REGRESSION FOR ABORIGINAL MALES WORKING FULL-TIME AND FULL-YEAR AND LIVING ON-RESERVE

Dependent variable: log earnings	Any aboriginal origins N=16 004	Aboriginal origins only N=14 324
Prince Edward Island	0.3116***	0.2958***
	(3.73)	(3.37)
Newfoundland	0.0587	0.1538
	(0.29)	(0.91)
Nova Scotia	-0.0497	-0.0403
	(1.22)	(0.94)
New Brunswick	-0.0891*	-0.0662
	(1.87)	(1.30)
Quebec	0.2082***	0.2249***
	(6.54)	(6.65)
Manitoba	-0.1310***	-0.1324***
	(5.92)	(5.67)
Saskatchewan	-0.0830***	-0.0829***
	(3.43)	(3.25)
Alberta	0.0621***	0.0583**
	(2.60)	(2.33)
British Columbia	0.0751***	0.0601**
	(3.26)	(0.015)
Age	0.0677***	0.0666***
	(12.41)	(11.29)
Age^2/100	-0.0657***	-0.0634***
	(10.44)	(9.37)
Rural	-0.0887***	-0.0762***
	(3.68)	(2.90)
Married	0.1894***	0.1842***
	(12.35)	(11.24)
Widowed/Separated/ Divorced	0.1080***	0.1062***
	(4.61)	(4.28)
High school diploma	0.1337***	0.1247***
	(6.40)	(5.56)
Non-university training	0.1456***	0.1455***
	(7.26)	(6.81)
Some post-secondary	0.2709***	0.2622***
	(15.37)	(14.01)
Bachelor's degree	0.5553***	0.5563***
	(16.81)	(17.30)
Post-graduate degree	0.4810***	0.4415***
	(8.78)	(8.27)
French spoken only	-0.1661***	-0.1895***
	(3.79)	(4.00)
Bilingual	0.0631*	0.0435
	(1.65)	(0.97)

Neither official language spoken	-0.1271 (1.22)	-0.1235 (1.16)
Intercept	8.3558*** (70.31)	8.3422*** (64.40)
Predicted log earnings	9.407	9.388
	R^2=0.1015 **F(22, 15981)=75.06**	**R^2=0.0998** **F(22, 14301)=71.71**

Note: Figures in parentheses are absolute values of coefficient t-ratios.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

TABLE B3: OLS EARNINGS REGRESSION FOR ABORIGINAL MALES WORKING FULL-TIME AND FULL-YEAR, OMITTING IDENTIFIER FOR MIXED ABORIGINAL/NON-ABORIGINAL ANCESTRY

Dependent variable: log earnings	**N=31 279**
Prince Edward Island	-0.1638***
	(5.24)
Newfoundland	-0.3287***
	(4.36)
Nova Scotia	-0.2235***
	(8.56)
New Brunswick	-0.2575
	(7.35)
Quebec	-0.0777***
	(4.19)
Manitoba	-0.1257***
	(7.96)
Saskatchewan	-0.1306
	(7.33)
Alberta	0.1323***
	(9.35)
British Columbia	-0.0107
	(0.75)
Age	0.0828***
	(22.82)
Age^2/100	-0.0905***
	(20.09)
Rural	0.0251**
	(2.44)
Married	0.2895***
	(25.87)
Widowed/Separated/ Divorced	0.0933***
	(5.37)
High school diploma	0.1380***
	(9.93)
Non-university training	0.2103***
	(14.67)
Some post-secondary	0.2713***
	(19.04)
Bachelor's degree	0.4943***
	(25.24)
Post-graduate degree	0.5650***
	(20.57)
French spoken only	-0.0633***
	(2.61)
Bilingual	0.0456***
	(3.16)
Neither official	0.2287

language spoken	(1.38)
Intercept	8.5065***
	(122.56)
R^2=0.1792	
F(22, 31 256)=235.12	

Note: Figures in parentheses are absolute values of coefficient t-ratios.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.