

Measuring Socioeconomic Status in Historical Censuses of Canada

Shari Eli, Nicholas Li, Janki Shah*

January 4, 2023

Abstract

We use the new complete count Canadian Census records spanning from 1871 to 1901 to explore measures of socioeconomic status including income, occupation, literacy, and prestige, which scholars commonly use to investigate economic mobility and inequality within society. We first explore the availability and comparability of measures across census years. Then, using individuals in the 1901 Census which contains wages, we document selection into the wage sample as well the characteristics of individuals found in different segments of the wage distribution. Lastly, we offer strategies to mitigate sample selection bias in future research.

JEL codes: C80, I32, J31, J62, N31, N32

* Shari Eli: University of Toronto and NBER, shari.eli@utoronto.ca; Nicholas Li: Toronto Metropolitan University, nicholas.li@torontomu.ca; Janki Shah: Bank of Canada, janki.shah@mail.utoronto.ca. We thank George Cunfer, Lisa Dillon, Alain Gagnon, Rob Gillezeau, Kris Inwood and Marc St-Hilaire as well as participants at The Canadian Peoples project meeting at the University of Saskatchewan for their comments. Census data used in this research were made available by The Canadian Peoples project (PI: Kris Inwood). The data collection and digitization were funded by the Canadian Foundation for Innovation, provincial research funds in Quebec, Ontario and Alberta as well as Ancestry.ca. Funding was also provided by the following academic institutions: Laval University, Queen's University, University of Alberta, University of Guelph, University of Montreal, University of Ottawa, University of New Brunswick, University of Saskatchewan, and the University of Toronto.

1. Introduction

There is a large and growing literature on economic mobility and inequality in Canada and the United States that exploits measures of status drawn from population censuses (Antoine et al 2022; Atkinson; Piketty and Saez 2011; Chetty, Hendren and Katz 2016; Corak 2020; and Long and Ferrie 2013). This body of research often requires scholars to link individuals over their entire lifetimes, as well as multiple generations, to answer topical questions concerning, for instance, the effects of government policy on socioeconomic status or the intergenerational transmission of social status (Abramitzky, Boustan and Eriksson 2014; Aizer et al 2016; Eli 2015; and Olivetti and Paserman 2015). Therefore, accurate measures of status are crucial, and in fact foundational, to the validity of these types of seminal and pioneering studies.

Measuring status in the post-World War II era is straightforward as there are consistent measures of status such as income, educational attainment, tax payments and government transfers, for nearly all individuals in the population of both Canada and the U.S.¹ However, for the years prior to 1901 in Canada, and prior to 1940 in the United States, population census records do not contain any income measures, which is problematic since income is often seen as one of the most important measure of socioeconomic status.² Thus, scholars have explored many proxy measures for income and overall socioeconomic status such as occupational income scores and other indices that rank individuals enumerated prior to 1950 (Duncan et al, 1961; Featherman and Stevens 1981; and Saavedra and Twinam 2020).

Recent scholarship on pre-1950 socioeconomic status measures has focused on the U.S., as opposed to Canada, in part because the full count censuses from 1850-1940 have been available to researchers for more than a decade through IPUMS USA. The complete count Censuses of Canada, spanning from 1871 to 1921, will become newly available to researchers starting in 2023 through The Canadian Peoples (TCP) project and will mark a new opportunity for enhanced research on economic mobility and inequality that is specific to Canadian populations.³ Therefore, our study provides a framework for empirical analyses that use socioeconomic status measures with these new data.

Our first contribution is to show the extent to which measures are available and comparable over census years. Second, we document selection into the wage sample and explore the associated characteristics of outliers in the wage distribution from 1901. Thus, we show how future analyses using wages may be biased and offer strategies to mitigate these biases. Third, we rank 1901 occupations using the median income in each occupation. Then, we impose our occupation status ranking from 1901 onto occupations in the prior census years. In this way, we capture changes in the occupational status distribution over time. We therefore explore changes

¹ These measures are arguably crude in that they do not capture measures of social capital that could be considered part of one's social status, but they are consistently available across years and so have been the focus of recent scholarship in economics and related disciplines. For a list of research on the US post-1950 that employs census data, see for instance: <https://opportunityinsights.org/paper/>. For research on Canada post-1950, see for instance: <https://mileskorak.com/research/income-mobility/canadian-geography-of-intergenerational-mobility/>.

² For a concise review of the history of grading socioeconomic status by income as well as other measures in the field of sociology, see Boyd (2008) as well as Goldthorpe and Hope (1974).

³ For details on data availability for researchers, see <https://thecanadianpeoples.com/>.

in status for Canadian populations in the late nineteenth century but also discuss the shortcomings of using occupational income scores in this period.

Finally, to address the fact that individuals of high socioeconomic status are unlikely to report any earnings, even in years when census enumerators collected earnings information, we turn to society registers or blue books to learn more about these individuals. Focusing on the cities of Toronto and Hamilton, we newly digitize the society registers from the era.⁴ These registers provide information on exact address of primary residence – which, unlike in the US, is not available in Canadian censuses – as well as prestigious club memberships, aristocratic and political titles, and locations of additional residences. We then hand match these individuals to our census records to characterize social status within the elite. Understanding the dispersion of status *within* high-status groups allows us to better understand just how profound socioeconomic inequality was in an era when there was pervasive underreporting of income amongst elites, as was the case in Canada a century ago.

2. Background

2.1 Occupational Income Score

A commonly used proxy for income in the US census is *occupational income score* (OCCSCORE), which is a variable created by IPUMS USA that assigns a score to each occupation based on where the median income within an occupation is in the 1950 income distribution.⁵ The concern with using the OCCSCORE variable for the years prior to 1950 is that the rank of median income for each occupation changes, in some cases dramatically, from 1850 to 1950 (Goldin and Margo 1992; Goldin and Katz 2008; Katz and Margo 2013). For instance, a physician in 1950 would likely be in a higher part of the income distribution relative to a physician in 1850. Conversely, a farmer in 1950 would likely be in a lower part of the income distribution relative to a farmer in 1850.

Scholars have explored a variety of strategies to create socioeconomic rankings or scores that are an improvement on the occupational income score. Researchers have used auxiliary datasets to assign measures of socioeconomic status to individuals found in the federal census by assuming that individuals in auxiliary data are representative, or can be reweighted to be made representative, of the general population. For instance, the Iowa State Census of 1915 contains income for all individuals and thus allows researchers to create measures of *predicted* income for non-Iowan individuals in the federal population census. To do this, scholars have used variables that are comparable in both data sources such as age and household size to predict income (Aizer et al 2016). Saavedra and Twinam (2020) use the 1915 Iowa State Census to outline the biases created when using occupational income scores and then provide a new machine learning approach to adjust the occupational score. The approach taken by Saavedra and Twinam – using an auxiliary dataset to adjust the occupational income score – is not possible in the Canadian context as there were no provincial censuses containing income information prior to 1901.⁶

⁴ This digitization process is ongoing and will be featured in upcoming work.

⁵ The use of occupation as a measure of socioeconomic status originates from the field of sociology. See Duncan (1966) and Erikson and Goldthorpe (1992). For more information on the construction of occupational score, see here: <https://usa.ipums.org/usa/chapter4/chapter4.shtml#occscore>.

⁶ There were provincial censuses such as the Census of the Northwest Provinces in 1906, however they are not useful to exploit as they all took place after 1901, which is the first year in which the federal census enumerates income for each individual. For more on provincial censuses, see: <https://www.bac-lac.gc.ca/eng/census/Pages/census.aspx>. See also: <https://publications.gc.ca/site/eng/9.830420/publication.html>.

However, Inwood, Minns and Summerfield (2019) exploit the fact that the 1911, 1921 and 1931 Canadian Censuses contain income data and thus create occupational income scores for each year in order to compare the scores with actual annual real earnings. Thus, they evaluate the use of occupational income scores and show that, in fact, they provide a poor approximation for actual earnings in instances when groups face declining earnings without changes in occupational profile.

Even during census years – such as 1911 and 1921 – when enumerators did collect information on income, the variable is still found to be missing for many groups. For instance, groups that are self-employed, such as farm owners, do not have earnings in the 1940 U.S. Census. Similarly in Canada, earnings coverage is not complete for the self-employed as well as for most farmers. In the Canadian Census of 1901, over half of households are still engaged in farming, yet fewer than five percent of these households report earnings.⁷

2.2 Literacy and Education

Given that income data can often be missing even in years when enumerators recorded wage and salary information, scholars have sought out other measures of socioeconomic status. These include literacy and years of education, which is especially useful as a proxy for educational attainment and the ability to assimilate amongst immigrant groups (Borjas 1994; Abramitzky, Boustan, Eriksson 2020; Lleras-Muney and Shertzer 2015).

2.3 Societal Measures of Status

Enrollment in elite universities, or even memberships in clubs within these educational institutions, has also been used as a measure of status in research on intergenerational mobility (Clark 2014; Clark, Andrew and Pottenger 2020; Michelman, Price and Zimmerman 2022). These societal measures of status are particularly useful in instances when the wage distribution within an occupation is narrow or in instances when elites do not report income. Conversely for those at the bottom of the socioeconomic status distribution, the collection of social assistance has been used as status measures (Aizer et al 2016). While these measures can be subject to bias, they provide a more nuanced measure of status in countries such as Canada that still experienced an entrenched class system during the late nineteenth and early 20th century. In the case of Canada, class boundaries were the result of a long colonial history and membership in the Commonwealth, which led many (elite) Canadians to retain distinctions, such as aristocratic titles, political appointments, membership in the clergy or military honors, all of which would not necessarily be reflected in income, education or occupation measures.⁸

3. Status Measures in the Census: Occupation, Income and Literacy

3.1 Availability and Comparability of Measures Across Years

⁷ Authors' calculations.

⁸ For instance, French nobles living in Lower Canada had lost much of their influence after the French Revolution but still experienced the benefits of high social status in Canada. Similarly, members of the clergy in both Ontario and Quebec also retained prominence and prestige. See Belshaw (2015) for a pre-Confederation history of Social Classes in Canada.

Table 1 shows the availability and comparability of status measures within the census across the years 1871 to 1921. These measures are divided into the following three categories: 1) occupation; 2) income; and 3) education and literacy. The only status measure that is available for all the census years is the first one: occupation. In this field, census enumerators supplied the occupation of the individual as well as whether the individual was retired but not both (e.g., the data does not contain information on the occupation an individual had before retiring). In 1911 and 1921, enumerators took information on an individual's employment in work other than in his primary field. From 1891 onward, information on place of work becomes available, though the exact question changes from a binary question ("working at trade in factory or in home") to a more specific question about where work takes place (e.g., "on farm" or "foundry").

Table 1 also shows the presence, or lack thereof, of questions on income and earnings in the censuses. From 1901 onwards, census enumerators recorded information on earnings from occupations or trades. For comparison, the U.K. first asks questions about income in 1921 and the U.S. does not ask the question until 1940. Therefore, the complete count 1901 Census of Canada provides the first opportunity to study income mobility and income inequality for the entire population of a country in North America during the early 20th century.⁹ In addition, in 1921, there are three questions concerning unemployment: 1) whether out of work; 2) weeks unemployed in the past 12 months; and 3) weeks unemployed in past 12 months due to illness.¹⁰

Finally, Table 1 also contains information on education and literacy. Unlike the questions in the U.S. census, Canadian enumerators do not record any information on years of schooling. Instead, they record information on the number of months at school during the year. This is a useful measure when considering the status of parents with certain occupations that benefit from child labor such as farm work during the harvest season. From 1891 onward, enumerators also record whether the individual can read and write, which is a crude measure of literacy and education. Finally, in 1911, enumerators ask about the costs of education at a college, convent, or university during the year for those in the household who are over 16. Taken together, the three panels of Table 1 illustrate the need for proxy measures of status especially for the years prior to 1891.

3.2 Response rates across Status Measures

3.2.1 Occupation

Table 2 shows that 75% percent of men aged 18-65 in the 1901 Census reported an occupation of some kind and that reporting an occupation was positively correlated with the amount of annual wage reported.¹¹ Individuals who reported themselves as an "employer" comprised 12% of the population of men aged 18-65 and were less likely to report a wage amount. Conversely, 42% of individuals reported themselves as employees and these individuals were highly likely to report their wage (the correlation between being an employee and having a non-missing wage is 0.68).

3.2.2 Income

⁹ Sager (2000); Inwood, Minns and Summerfield (2019); and Antoine et al (2022) use subsamples of these data.

¹⁰ Note: The 1901 Census contains data on months of work.

¹¹ Note: Individuals may report a wage without reporting an occupation.

When reporting wages, census enumerators were instructed to provide the income of each individual “who is employed in any industrial or other occupation and is paid salary, wages or other money allowance for his or her service...”¹² Therefore, those who were self-employed, such as proprietors as well as farmers, were not required to report their earnings. Only 17% of individuals enumerated in the 1901 census reported their wages. Restricting the sample of individuals (men and women) to those aged 18 to 65, we find that 27% reported a wage. When we further restrict the sample to include only *men* aged 18 to 65, we find that only 42% reported their wages (see Table 2). Finally, only 40% of those who report an occupation were also found to report their wages, and 25% of individuals with wages did not report an occupation.¹³

For individuals who report an occupation (regardless of whether or not they report a wage), we can compute the median wage within their occupation for individuals that report wages for the same occupation and then assign this median wage to them based on their occupation. Table 2 shows that those who report occupations with a higher median wage are more likely to report a wage. (Also see Figure 6, which is further discussed in Section 6).

In Column 1 of Table 3, we provide the rate at which individuals in each occupation category reported wages. For instance, we find that less than 5% of farmers reported a wage. Conversely, more than 83% of machinists reported their wages. It also appears that the rate of wage reporting tends to increase from the bottom until the middle of the income distribution, and then wage reporting falls again when moving from the middle to the top of the distribution. Therefore, selection into the wage sample is a concern for researchers as median occupation wages may not be representative of the entire population within an occupation group. In addition, particular attention should be paid to any analysis that investigates farmers in the late nineteenth and early twentieth century as farmers were not low-income and instead typically earned the median wage.¹⁴ We further explore selection into the wage sample in Section 4.

3.2.3. Education and Literacy

Restricting our sample to men aged 18 to 65 in the 1901 Census, Table 2 shows that 86% of these men could read and write. Expectedly, those who reported that they could read and write were also more likely to report having a wage as well as an occupation. They were also more likely to report higher wage amounts relative to their illiterate counterparts and to be employed in occupations with higher median wages.

Next, we consider school attendance patterns of children. Of men aged 18 to 65 with children aged 7 to 13, 73% of their children attended school for at least some part of the year. On average, their children attended school for 6.45 months.¹⁵

3.3 The Wage Distribution: Right Tails

¹² See page 16, section 62 of *Fourth Census of Canada 1901, Instructions to Chief Officers, Commissioners and Enumerators*, published in 1901 by the Government Printing Bureau.

¹³ Of those who report a non-farmer occupation, 66% report wages.

¹⁴ See Costa and Kahn (2008) for instance.

¹⁵ We explore how sensitive the results are to different child age restrictions and show them in Table 2. For instance, if we restrict child ages from 6 to 14 years, we find that 58% of them attended school. On average, these children attended school for 5 months.

In Figure 1, we plot the distribution of wages in 1901, which contains income information for 17% of the entire sample. First, Figure 1 shows that there is heaping of wage reporting around round numbers such as annual wages of \$500 or \$1000. Second, the figure shows that there are very few individuals who report an income of zero (only 129 individuals in this category). Census enumerators leave this field blank for individuals who are unemployed. Third, and most importantly, the figure shows a long right tail even though we exclude the top 1% of income earners (e.g., those who earn more than \$2,290).

The top income earners in 1901 report annual wages of \$35,000 but most of these wage reports are implausible. For instance, one of these individuals is a man named George Randall born on December 23, 1859, in Nova Scotia. He reports that he is a miner in Yarmouth, Nova Scotia. Upon looking at the image of his census entry, as opposed to the Ancestry database, the census enumerator wrote that George Randall earned \$350.00, not \$35000 per year. Indeed, the median wage for a miner in 1901 is \$450 per year (see Table 3). Similarly, H Georg French, who was born on September 16, 1868, born in Ontario, and worked as a Carpenter also appears in the top 1% wage group in the 1901 Census database on Ancestry. However, upon closer inspection of the image, it is again clear that he earns only \$350.00 per year as the last two zeroes are superscript and underlined by the enumerator to signify cents as opposed to dollars. Therefore, studies using wage data from this census should consider the median wage in the occupation reported by the top earner and, if feasible given the number of observations, consult the original images from the 1901 census. In addition, the best practice would be to drop (or check) observations in which wages widely deviate from the occupation median or to drop the top 1% of wage earners within each occupation.

In Column 5 of Table 3, we provide the average wages listed in *Bulletin 1, Wage-Earners by Occupation*, which was published in 1907, and which contains wage tabulations using the 1901 Census that were made by the government of Canada. In nearly all cases, the medians that we find in our calculations using the complete count census of 1901 (see column 2) coincide relatively well with the averages reported in the 1907 bulletin. The exceptions are for hotel keepers, insurance agents, dentists, barristers, and civil engineers. All these occupations for which the median wage and average do not match well are occupations in the top tier of the income distribution (see Figure 1), which further emphasizes the need for additional proxy measures of socioeconomic status for groups employed in high status occupations. Said differently, there is a wider wage distribution within high status occupations than within lower status ones.

4. Selection into the Wage Sample

4.1 Correlation of Status Measures for Individuals

In Figure 2, we show the relationship between the head of household's annual wage and the average number of months in school for the children in the household in 1901. As expected, low-wage households have lower school attendance, as measured in months, relative to high-wage households. However, those who attend school for the least number of months, or who do not attend school at all, do not come from the lowest earning household but instead come from households earning approximately \$450 annually, which was the median income earned by a machinist or miner in the era. The pattern shown in Figure 2 suggests that child labor varied

significantly for those in the bottom half of the income distribution and shows that the trend was not linear.

Given that children could contribute meaningfully to farm work in this era, a natural question is whether children in farm households were less likely to attend school relative to children living in non-farm households. Figure 3 shows the distribution of months spent in school for farm versus non-farm households. Children in non-farm headed households accumulated more months of school than their counterparts in non-farm headed households.

Next, we consider wage distributions by province in Figure 4. We show that the nominal wage distribution in British Columbia skews to the right of the distributions from all other provinces. At first glance, this geographic wage pattern – in which western regions have higher nominal wages relative to eastern ones – seems similar to patterns shown in the United States in the same era. Indeed, Rosenbloom (1990) showed that nominal as well as real wages converged across regions in the US as the national labor market became more integrated.

A benefit of Rosenbloom’s work is his use of cost-of-living measures, such as food and housing prices, to convert nominal wages into real ones. Using this approach, he argues that there was indeed real wage convergence during the second half of the nineteenth century. Similarly, Emery and Levitt (2003) show that in Canada there were large cost differentials across regions prior to 1914. However, they show that Canada did not experience real wage convergence to the same extent as the United States. A shortcoming of Rosenbloom’s work, however, is that data are drawn from only 12 cities, and the West region is represented only by the city of San Francisco. Therefore, using the complete count 1901 Canadian census provides researchers with a more representative picture of nominal wages, especially in rural regions.

Figure 5a shows wage distributions for those born within Canada, those born in the United Kingdom and all other groups born abroad. The graph shows that the foreign born earn more on average relative to the Canadian born. In particular, those born in the United Kingdom earn the most relative to all other groups including the Canadian born. Figure 5b shows wage distributions by race. In the 1901 census, there were four main races: 1) “white” to denote Caucasian; 2) “yellow” to denote Asian; 3) “black” to denote African descent; and 4) “red” to denote Indigenous people.¹⁶ The distribution of wages for whites is clearly to the right of the distribution for all other races. While racial interactions and dynamics are beyond the scope of this paper, the race differentials in wages that we find warrant two important caveats. First, enumerators were often tasked with deciding the race of an individual as they filled out census registers. Given that race is viewed as a social construct or political category outside the discipline of economics and given that many individuals are not clearly identifiable as one race or another, there is considerable measurement error in the race variable (Roberts 2012; Rose 2022; Schaub 2019). Second, Liebler et al (2017) has shown that individuals self-report their own race differently from census to census, and Cornwell, Rivera and Schmutte (2017) show that self-reports of race change as an individual’s wage fluctuates. Thus, it is important to consider issues of context and subjectivity when using race variables, especially when intersected with wage data.

¹⁶ In the 1901 Census data as cleaned and provided by *The Canadian Peoples* project, race variables appear in the following way: 1) “w” for white; 2) “b” for black; and 3) “r” for red, etc. In Quebec, however, “b” is denoted for *blanc* in French, which translates to *white* in English. Most individuals in the French province of Quebec have a “b” designation for race in the 1901 Census of Canada.

4.2 Representativeness of Individuals in the Wage Sample

In order to assess the extent to which individuals in the wage sample are representative of men aged 18 to 65 in the 1901 census, we present (OLS) wage regressions in Table 4. First, we regress the annual wage earned by men aged 18 to 65 on three variables: 1) reads and writes; 2) whether an employer or not; and 3) whether a servant is employed in the home. Table 4, Column 1 shows the coefficient estimates on each of the three variables. The coefficients are positive, large and statistically significant for each of the three variables suggesting that each are large contributors to wage levels. For instance, the ability to read and write is associated with more than \$100 extra dollars of wages per year, which is a quarter of the mean wage \$408.27 for men aged 18 to 65.

Table 4, Column 2 shows coefficient estimates when the dependent variable is not wage amount but instead whether the individual reported any wage at all or, said differently, whether the wage was “non-missing” in the data. Column 2 show that the wage sample is slightly positively selected on literacy, slightly negatively selected on being an employer, and firmly negatively selected on having a servant in the home.

In Table 4, Columns 3 and 4, we restrict the sample to men who are 18 to 65 years old *and* in households with children aged 7 to 13 years old. In those households, reporting any wage is slightly negatively associated with literacy, and literacy contributes *less* to annual income (\$83.07) than it does for men in the overall sample (\$100.69). For this group of men with children in the home, we also include the variable *mean months in school*, which is the mean number of months in school for children aged 7-13. This age range was chosen because at least 67% of children in each age within this range attended school. (Our results are similar when we choose wider age range such as 6 years to 16 years old where at least 40% of children attend some school at every age in the range.) If children in the household attend an average of an extra month of school, the associated annual wage increase is \$3. This also suggests that the wage sample is likely to contain men from households with children that spend more time in school.

In Table 4, Columns 5 to 12, we show results when the sample is split by the median wage earned in a given occupation. In particular, we choose the annual wage cutoffs of \$200, \$500 and \$720 as these cutoffs represent the 50th, 90th and 99th percentiles of the distribution of the median wage per occupation category. For all groups except those in occupations having a median wage of less than \$200 per year, literacy is associated with wage reporting. To the extent that literacy is associated with numeracy, illiterate individuals may face difficulty reporting their exact earnings to census enumerators. Indeed, Figure 1 shows heaping around round numbers for individuals earning under \$300 per year, which may indicate that illiterate individuals have difficulty recalling wage amounts that are not round numbers. Conversely, we find that those in occupations that have median wages in the 99th percentile are twice as likely to report their wages relative to individuals with occupations in the 90th or 50th percentile.

Overall, results are consistent for those engaged in occupations in which the median wage is between \$200 and \$500 annually (Columns 7 to 10). For these groups, we observe positive selection on literacy and the average number of months children in the household spend in school. We also observe negative selection on being an employer and having a servant in the home.

5. Occupational Rankings by Median Income, 1871-1891

In census years prior to 1901, individuals were not asked to provide information on their annual wages. However, since we have information on their occupations, we are able to construct an occupational ranking by median income in 1901 in order to rank occupations in prior census years. However, an occupational ranking of this kind imposes the occupational-wage hierarchy from 1901 onto earlier census years, which may not be suitable given the historical context as noted in the discussion of occupational income scores in Section 2.1.

In order to construct the occupational ranking, we first compute the 1901 median wage for each occupation group (using the 3-digit occupation variable). We then assign the occupation median wage to each individual within the occupation group in all census years prior to 1901. Finally, we compute the mean of all of the median wages by each occupation for individuals that report an occupation during the census year and present the results in Table 5.

First, Table 5 shows that the percent of the population that reports any occupation increases from 26% in 1871 to 38% in 1901. When we consider all men and women who report an occupation in 1871, we find that the mean of the median wages across occupations is \$260 per year. In 1901, the mean for this group rises to \$274 per year, which suggests that between 1871 and 1901, individuals are on average experiencing a rise in occupational ranking (given the 1901 distribution of median wages per occupation). When we restrict the sample to men only, we find that the rise is even more profound (from \$265 to \$288 annually). Finally, we see the largest rise when we restrict the sample to men aged 16 to 60. The results of this exercise are consistent with literature on the nineteenth century North American experience which highlights the rise in income and living standards for individuals in the era.¹⁷

6. Status and Prestige Measures from Social Registers

The final measures of status that we explore are those which we call “prestige” measures and which we find in historical society registers. In particular, we turn to the *Toronto and Hamilton Society Blue Book* published in 1902 by Wm. Tyrell & Co. with the stated purpose of providing, “a reliable Directory to over 2,500 of the Elite Families of Toronto and Hamilton, arranged alphabetically, with much additional regarding Families, Club Membership, Summer Residences, Maiden Names, Receiving Days, and other items of social interest” (p. 7).

An entry in the register appeared in the following way:

Gooderham, Mr. and Mrs. George
“Wavney,” 135 St. George Street
Mrs., nee Dean.
Receive Friday
Miss Violet Gooderham.
Mr., 1-3-17-31-42-52-55-61.

This entry is for George Gooderham, who was born in 1830 and died in 1905. He was the son of William Gooderham Sr., who was a founder of Gooderham and Worts distillery as well as a banker and elite business leader. His home at 135 St. George Street is the location of one of the most prestigious social clubs in Toronto today, The York Club, which was incorporated in 1909 and thus featured in the *1911 Toronto Society Blue Book*. Finally, the club memberships for Mr. Gooderham are listed on the last line and pertain to the following clubs: Albany Club, Argonaut

¹⁷ For a review, see Fogel (2004) and Cutler, Deaton, Lleras-Muney (2006).

Rowing Club, Granite Club, Ontario Jockey Club, Royal Canadian Yacht Club, Toronto Club, Toronto Hunt and the Victoria Club.

The society registers are particularly useful because they contain full names of individuals which aids in the process of matching them to the 1901 census. In this way, we can compile information from the registers that is not present in the census but that can be used for matching individuals across censuses such as the maiden names of wives as well as the names of children living in the household. In addition, match rates are likely to be high for these individuals as they are more likely to be literate as well as interested in having a relatively public profile, as they are found in a society blue book to which they themselves contribute information.

Our efforts to digitize the entirety of the Toronto Society Blue Books from the first half of the twentieth century are ongoing, and the results of our empirical analysis will be included in Eli et al (2023). However, in this work we are interested in understanding selection into the wage sample amongst elites and so we ask, what are the blue book characteristics of elites who are in occupations in which the median wage is in the 99th percentile (above \$720 annually) but who do not themselves report a wage?

Considering the Gooderham descendants provides an interesting case study. There are three Gooderham descendants who are found in the sample of individuals who are men between the ages of 18 and 65. Their names are Ross, Edward, and Henry Gooderham and their ages are 24, 41 and 24, respectively.¹⁸ Of the three men, Edward is the only one with children between the ages of 7 and 13 living in the household, and so his information would be contained in columns 11 and 12 of Table 4. Edward's occupation is "Manager," and he does not report a wage. Upon searching through the 1902 blue book, we observe that Edward's primary residence is located at 40 Madison Avenue in Toronto and that he is a member of two clubs, The Albany Club and the Ontario Jockey Club.

Conversely, we search for Ross and Henry Gooderham – both of whom are listed as barristers in the 1901 census and both of whom do not report a wage – in the 1902 blue book. Ross reports that he lives at 204 St. George Street and is part of just one club, the Royal Canadian Yacht Club. Henry is reported by his father Charles H. Gooderham as living in his father's residence located at 592 Sherbourne Street and is not reported to be a member of any clubs.

For comparison, we search for Arthur T. Lowe, who lives in the same census sub-division as Edward Gooderham and who is a 45-year-old civil engineer. He does not report his earnings, but he does report his *extra* earnings as \$800 per year. Arthur Lowe is not found in the 1902 blue book. Though these are only a few anecdotal examples, our preliminary hand matching efforts have so far shown that wage reporting declines as wage increases when restricting to those who report themselves as being in occupations with median wages above approximately \$250 per year. Indeed, Figure 6 shows that wage reporting amongst those who report an occupation first increases and then decreases by median wage within that occupation.

Taken together, our matching efforts show the importance of considering context as well as auxiliary datasets for elites. Important measures of status that can be found in the blue books include the following: 1) exact street addresses of primary and other residences, which can be matched to land records to determine land values; 2) the number of clubs joined, which can be a proxy for social proclivities; and 3) aristocratic, military, and other titles. Taken together, these measures provide a more complete picture of social status for members of the elite.

¹⁸ The given name variable for Henry Gooderham contains the entry "Nancy" as opposed to "Henry." Upon inspection of the census image files, the entry of "Nancy" appears to be a digitization error.

7. Conclusion

We use the newly digitized complete-count 1901 Census of Canada to explore measures of socioeconomic status that scholars often use in research on economic mobility and inequality. We show the difficulty in using common status measures such as earnings, occupation, literacy school attendance, and prestige measures when ranking individuals. In particular, we focus on selection into the sample of individuals who report their wages. We show that individuals who report their wages are most likely to be from the middle of the median wage (by occupation) distribution and that wage reporting is reduced amongst individuals of low and high occupational status. We also show the ways in which researchers must use caution when interpreting status measures in empirical research using the historical Canadian census records.

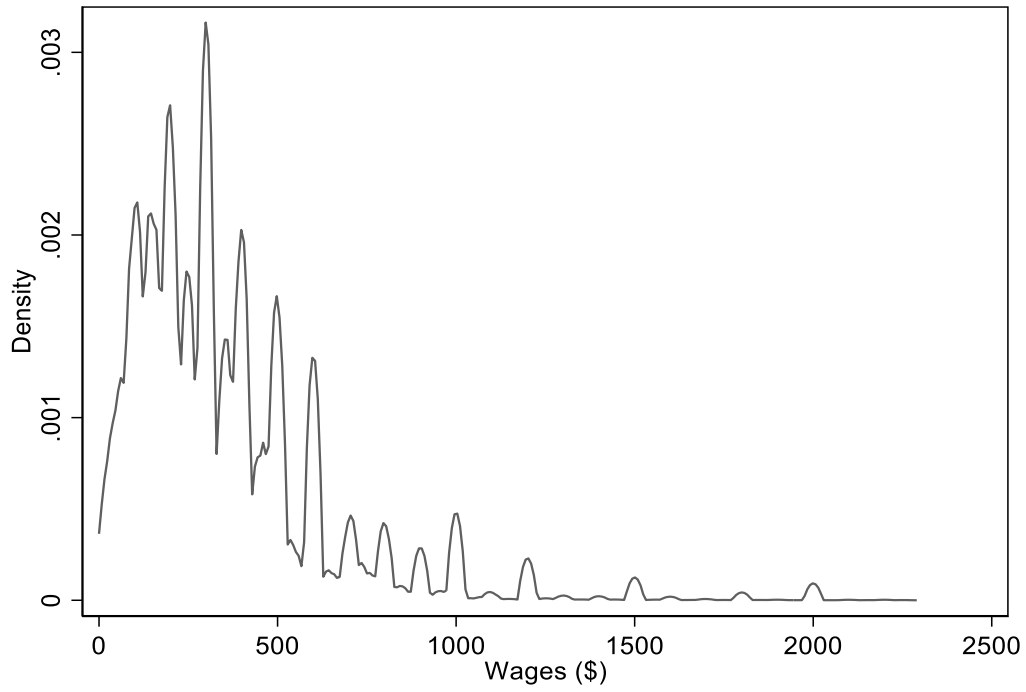
References

- Abramitzky, R., L. Boustan, and K. Eriksson. Do Immigrants Assimilate More Slowly Today Than in the Past? *American Economic Review: Insights* 2(1): 2020. doi: 10.1257/aeri.20190079
- Abramitzky, R., L. Boustan, and K. Eriksson. A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy*, 122(3): 2014. doi: 10.1086/675805.
- Aizer, A., S. Eli, J. Ferrie and A. Lleras-Muney. The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review* 106(4): 2016. doi: 10.1257/aer.20140529.
- Antoine L., K. Inwood, C. Minns, and F. Summerfield. Intergenerational Mobility in a Mid-Atlantic Economy: Canada 1871-1901. *The Journal of Economic History* 82(4): 2022. doi: 10.1017/S0022050722000353.
- Atkinson, A. B., Piketty, T., and E. Saez. Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1): 2011. doi: 10.1257/jel.49.1.3.
- Belshaw, J. D. 2015. Canadian History: Pre-Confederation. BCcampus.
- Borjas, G. Long Run Convergence of Ethnic Skill Differentials: The Children and Grandchildren of the Great Migration. *Industrial and Labor Relations Review* 47(4): 1994. doi: 10.1177/00197939940470040
- Boyd, M. A Socioeconomic Scale for Canada: Measuring Occupational Status from the Census. *Canadian Review of Sociology* 45(1): 2008. doi: 10.1111/j.1755-618X.2008.00003.x.
- Census and Statistics. 1907. *Bulletin 1, Wage-Earners by Occupation*. Ottawa: S.E. Dawson.
- Census of Canada. 1901. *Fourth Census of Canada 1901: Instructions to Chief Officers Commissioners and Enumerators*, Ottawa: Government Printing Bureau.
- Chetty, R. N. Hendren, and L. Katz. The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, 106(4): 2016. doi: 10.1257/aer.20150572
- Clark, G. L., A. Leigh, and M. Pottenger. Frontiers of Mobility: Was Australia 1870-2017 a More Socially Mobile Society than England? *Explorations in Economic History* 76(5): 2020. doi: 10.1016/j.eeh.2020.101327.
- Clark, G. 2014. *The Son Also Rises*. Princeton, NJ: Princeton University Press.

- Cornwell, C., J. Rivera, and I. M. Schmutte. Wage Discrimination When Identity is Subjective Evidence from Changes in Employer-Reported Race. *Journal of Human Resources*, 52(3): 2017. doi: 10.3368/jhr.52.3.0815-7340R1.
- Corak, M. The Canadian Geography of Intergenerational Income Mobility. *The Economic Journal* 130(631): 2020. doi: 10.1093/ej/uez019.
- Costa, D., and M. Kahn. 2008. *Heroes and Cowards: The Social Faces of War*. Princeton, NJ: Princeton University Press.
- Cutler, D., A. Deaton, and A. Lleras-Muney. The Determinants of Mortality. *The Journal of Economic Perspectives* 20(3): 2006. doi: 10.1257/jep.20.3.97.
- Dau's Blue Books. 1902. *The Society Blue Book, Toronto: A Social Directory 1902*. Toronto, Canada: Wm. Tyrell & Co.
- Duncan, O. D., and A.J. Reiss. A Socioeconomic Index for All Occupations. *Class: Critical Concepts* 1(1): 1961.
- Duncan, O. D. 1966. Methodological Issues in the Analysis of Social Mobility. In *Social Structure and Mobility in Economic Development*, ed. N. J. Smelser and S. M. Lipset, 51–97. Chicago: Aldine.
- Eli, S. Income Effects on Health: Evidence from Union Army Pensions. *The Journal of Economic History* 75(2): 2015. doi: 10.1017/S0022050715000674.
- Emery, J. C. H, and C. Levitt. Cost of living, real wages and real incomes in thirteen Canadian cities, 1900–1950. *Canadian Journal of Economics*, 35(1): 2002. doi: 10.1111/1540-5982.00123.
- Erikson, R. and J. H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
- Featherman, D. L., and G. Stevens. A Revised Socioeconomic Index of Occupational Status. *Social Science Research*, 10(4): 1981. doi: 10.1016/0049-089X(81)90011-9.
- Fogel, R. 2004. *The Escape from Hunger and Premature Death, 1700-2100*. Cambridge: Cambridge University Press.
- Liebler, C. A., S. R. Porter, L. E. Fernandez, J. M. Noon, and S. R. Ennis. 2017. America's Churning Races: Race and Ethnicity Response Changes Between Census 2000 and the 2010 Census. *Demography*, 54(1): 2017. doi: 10.1007/s13524-016-0544-0.
- Long, J., and J. Ferrie. Intergenerational Occupational Mobility in Great Britain and the United States since 1850. *American Economic Review*, 103(4): 2013. doi: 10.1257/aer.103.4.1109.
- Goldin, C., R. A. Margo. The Great Compression: The Wage Structure in the United States at Mid-Century. *The Quarterly Journal of Economics*, 107(1): 1992. doi:10.2307/2118322.
- Goldin, C., L.F Katz. *The Race Between Education and Technology*. Cambridge, MA: Harvard University Press.
- Goldthorpe, J. H., and K. Hope. 1974. *The Social Grading of Occupations: An New Approach and Scale*. Oxford, England: Clarendon Press.
- Inwood, K., C. Minns, and F. Summerfield. Occupational Income Scores and Immigrant Assimilation. Evidence from the Canadian Census. *Explorations in Economic History* 72(4): 2019. doi: 10.1016/j.eeh.2019.02.001.
- Lleras-Muney, A., and A. Shertzer. Did the Americanization Movement Succeed? An Evaluation of the Effect of English-Only and Compulsory Schooling Laws on Immigrants. *American Economic Journal: Economic Policy* 7(3): 2015. doi: 10.1257/pol.20120219.

- Michelman, V., J. Price, and S. D. Zimmerman. Old Boys' Clubs and Upward Mobility Among the Educational Elite. *The Quarterly Journal of Economics* 137(2): 2022. doi: 10.1093/qje/qjab047
- Olivetti, C., M.D. Paserman. In The Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940. *American Economic Review* 105(8): 2015. doi: 10.1257/aer.20130821.
- Roberts, D. 2012. *Fatal Intervention: How Science, Politics, and Big Business Re-create Race in the Twenty-first Century*. New York, NY: The New Press.
- Rose, E. A Constructivist Perspective on Empirical Discrimination Research. *Journal of Economic Literature*, forthcoming.
- Rosenbloom, J. L. One Market or Many? Labor Market Integration in the Late Nineteenth-Century United States. *The Journal of Economic History* 50(1): 1990. doi: 10.1017/S0022050700035737.
- Saavedra, M., T. Twinam. A Machine Learning Approach to Improving Occupation Income Scores. *Explorations in Economic History* 75: 2020. doi: 10.1016/j.eeh.2019.101304.
- Sager, E. W. National Data on Working-Class Earnings: The 1901 Census of Canada. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 33(4): 2000. doi: 10.1080/01615440009598966.
- Schaub, J. F. 2019. *Race is About Politics: Lessons from History*. Princeton: Princeton University Press.

Figure 1: Distribution of Annual Wages in 1901



Note: N=914,903, 17.01% of the total sample in the 1901 Census of Canada. Top 1% of wage earners, who earn more than \$2,229 per year in 1901, are excluded from this figure.

Figure 2 - Months Spent in School by Household Wage, 1901

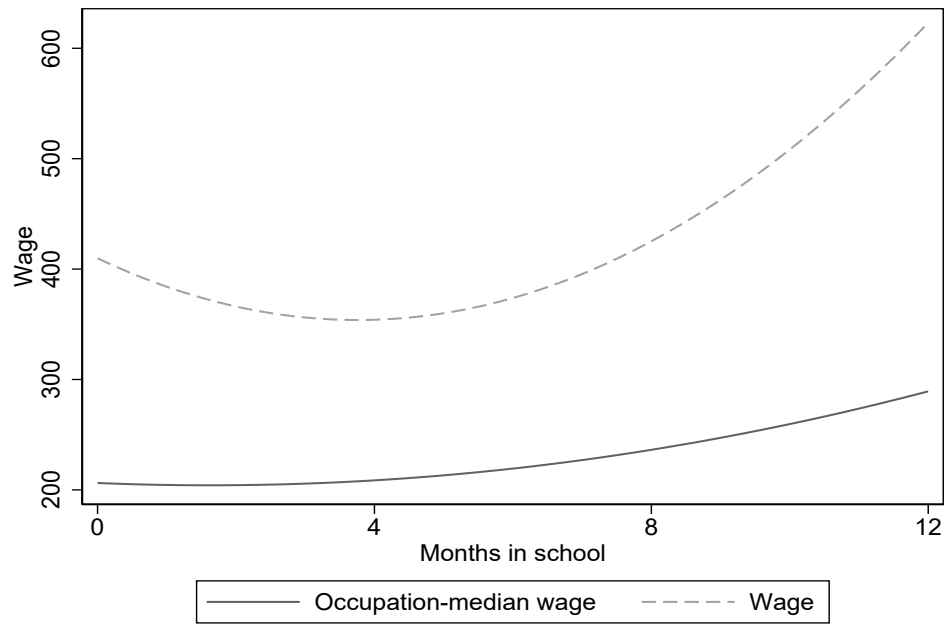


Figure 3 - Distribution of Child's Months Spent in School, Farm vs. Non-farm households

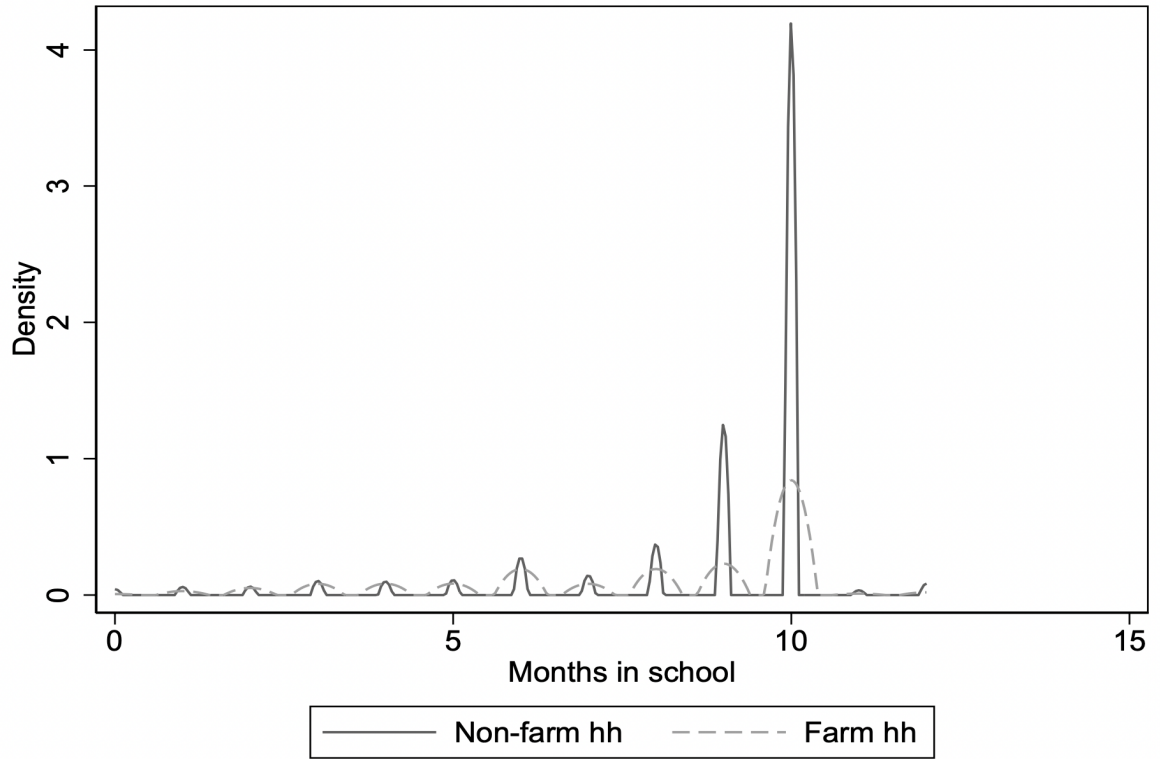


Figure 4 - Wage Distributions by Province in 1901

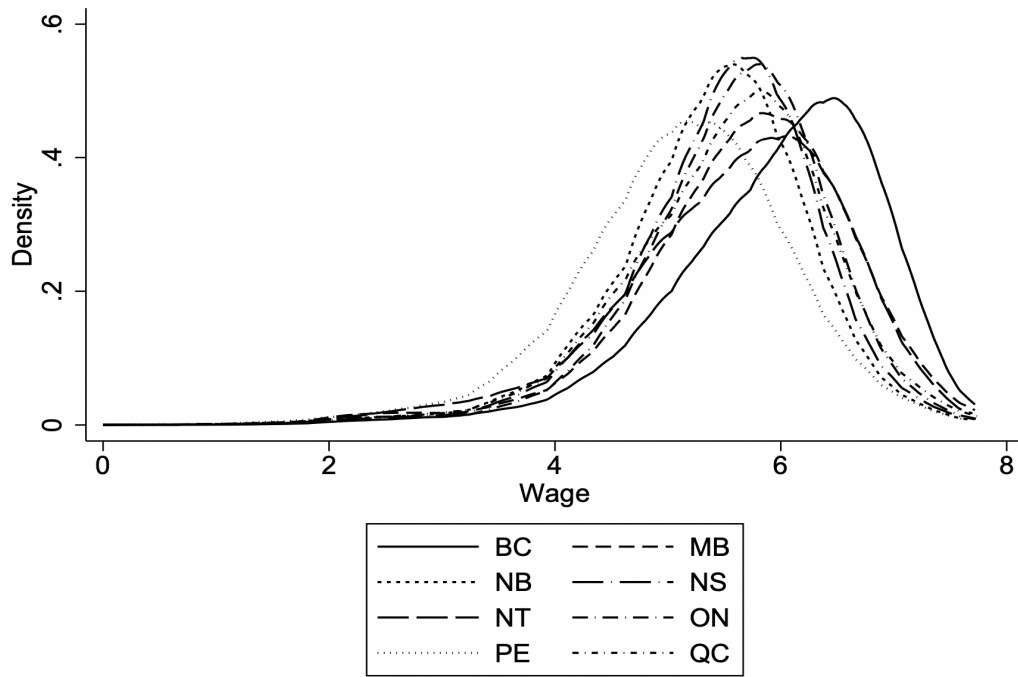


Figure 5a - Wage Distributions for Canadian Born versus New Immigrants

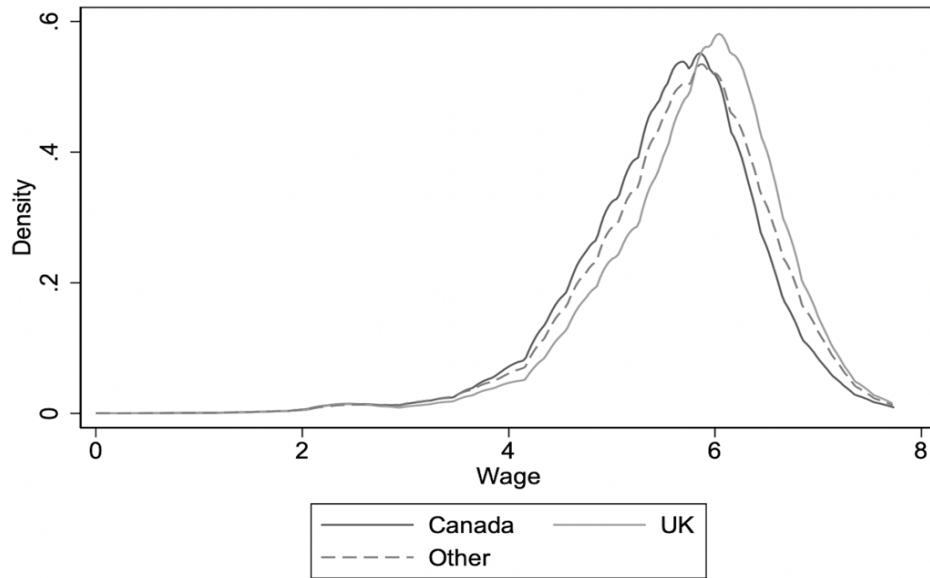


Figure 5b - Wage Distributions for each Race ("Colour")

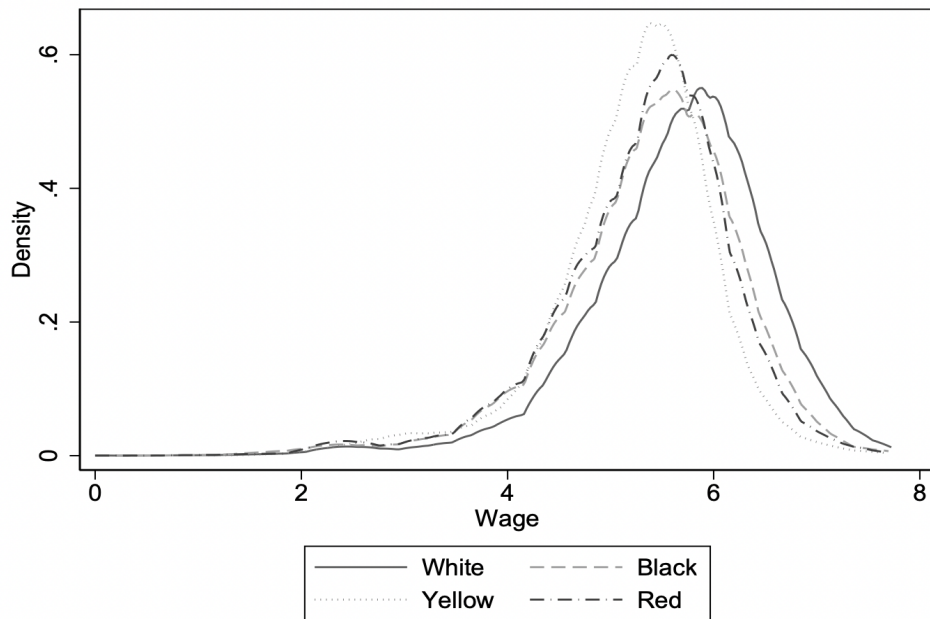
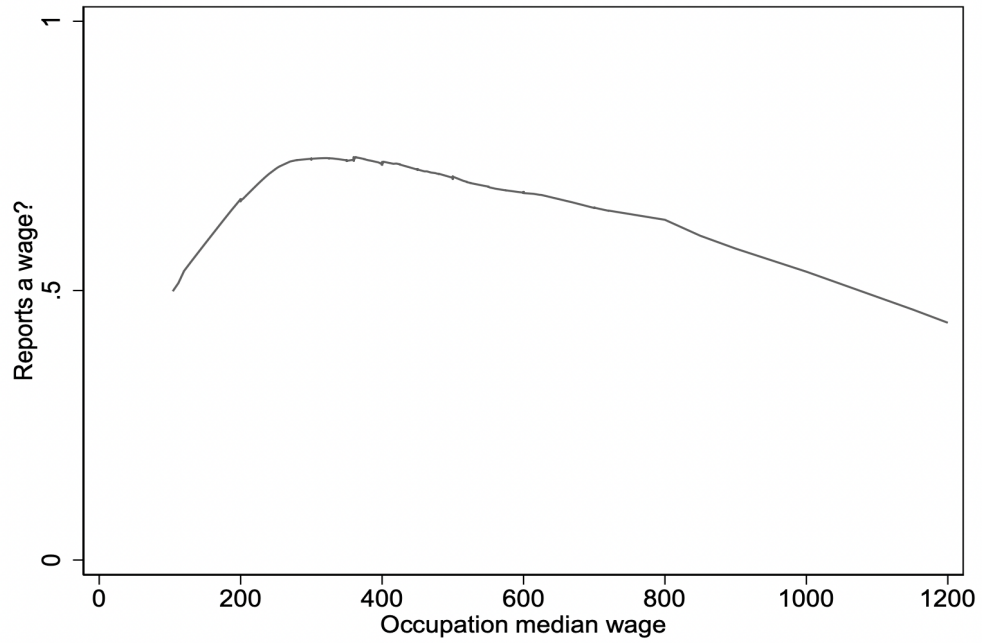


Figure 6 - Wage Reporting by Median Wage within Occupation



Note: We use the Stata lowess command to plot the distribution of median wages within an occupation. In this way, each row in the data represents an occupation.

Table 1. Measures of Status in Canadian Census data, 1871-1921

Status Measure	Census Year					
	1871	1881	1891	1901	1911	1921
<i>Occupation</i>						
Profession or Occupation (& if retired)	x*	x*	x	x	x	x
Employment other than at chief occ. or trade, if any					x	x
Employer or Employee			x**	x	x	x
Working on Own Account				x	x	x
Working at Trade in Factory or in Home			x***	x		
Place of employment (e.g. "on farm" "foundry")					x	x
Mo. Employed Occ. other than Trade/Factory/Home				x	x****	
Wage Earner			x			
Hours of working time per week					x	
<i>Income</i>						
Living on Own Means (i.e. annuities, pensions)				x		
Earnings from Occupation or Trades				x	x	x
Extra Earnings				x	x	
Out of work						x
Weeks unemployed in past 12 mos.						x
Weeks unemployed in past 12 mos. due to illness						x
<i>Education and Literacy</i>						
Months at School in Year				x	x	x
Can Read/Write			x	x	x	x
Costs for those over 16 at College, Convent or Univ.					x	

Notes: *In 1871 and 1881, asked if in trade as well as profession or occupation. **In 1891, census enumerators asked whether the person was an employer (not whether an employee) and how many "hand" or employees the individual had. ***In 1891, enumerators asked if the individuals was working in a trade in a factory (but not whether in a home, as in 1901). ****1911 census as for weeks, not months, employed. 1871 census includes Nova Scotia, New Brunswick, Quebec and Ontario only. 1881 Census includes British Columbia, Manitoba, New Brunswick, Nova Scotia, Ontario, PEI, Quebec and Northwest Territories. 1891 Census includes British Columbia, Manitoba, New Brunswick, Nova Scotia, Ontario, Prince Edward Island, Quebec, and the Northwest Territories (Alberta, Assiniboia, and Saskatchewan). 1901 Census includes British Columbia, Manitoba, New Brunswick, Nova Scotia, Ontario, Prince Edward Island, Quebec, and the North-West Territories (which included the districts of Alberta, Assiniboia, Athabasca, Franklin, MacKenzie, Saskatchewan, and Ungava). 1911 census includes Alberta, British Columbia, Manitoba, New Brunswick, Nova Scotia, Ontario, Prince Edward Island, Quebec, and Saskatchewan, and two territories - the Yukon Territory and the Northwest Territories. 1921 census includes same regions as the 1911 census as well as the Royal Canadian Navy.

Table 2: Summary Correlations for Men age 18-65 in 1901 Census

	N	Mean	SD	Correlations			
				Wage Amount	Occ-Wage	Non-Missing Wage	Non-missing occ.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<u>Occupation</u>							
Non-missing Occupation	1,336,978	0.75	0.43	0.14	.	-0.01	1
Median wage within an occupation*	1,007,211	302.35	133.85	0.42	1	0.38	.
Individual is an Employer	1,336,978	0.12	0.33	0.11	0.04	-0.02	0.09
Individual is an Employee	1,336,978	0.42	0.49	-0.06	0.36	0.68	-0.05
Own account worker (self-employed)	1,336,978	0.32	0.47	0.04	-0.08	-0.12	0.22
Own means	1,336,978	0.34	0.47	0.03	0.05	0.15	0.04
<u>Income</u>							
Wage amount	565,891	408.27	288.33	1	0.42	.	0.14
Non-missing Wage	1,336,978	0.42	0.49	.	0.38	1	-0.01
<u>Education & Literacy</u>							
Read and write	1,336,978	0.86	0.35	0.17	0.12	0.05	0.1
Share in school (children age 7-13)	551,196	0.73	0.42	0.07	0.1	0.09	0.05
Mean months of school (age 7-13)	551,196	6.45	4.05	0.12	0.17	0.13	0.04
Share in school (age 6-16)	714,317	0.58	0.43	0.09	0.1	0.08	0.06
Mean months of school (age 6-16)	714,317	4.99	4.03	0.13	0.17	0.12	0.05

Note: We restrict to households with 10 or fewer individual. This table restricts the sample to those who report an occupation of any sort. Wage amount are amounts earned annually. *For the median wage within an occupation, we assign the median wage within a particular occupation to all individuals who report that occupation, even if the individual does not report a wage.

Table 3. Earnings by (Select) Occupations, 1901 Census

<i><u>Occupation</u></i>	Percent Reporting Wages (1)	Median Wage from Full Count Census (2)	Wages: 10th Percentile (3)	Wages: 90th Percentile (4)	Avg. Wages, <i>Census Bulletin</i> , for males* (5)
Servant	63.50	120	48	280	181.90
Farmer's Son	26.59	200	75	336	199.83
Fisherman	54.56	200	80	400	203.30
Laundryman	69.82	200	75	480	193.23
Farmer	4.54	210	60	750	207.55
Teacher	66.43	250	90	600	245.76**
Lumberman	68.28	300	120	600	299.40
Mason	71.35	400	150	700	427.96
Rancher	21.96	400	120	1000	357.50
Blacksmith	59.24	400	150	800	434.06
Painter	73.71	400	150	600	409.36
Carpenter	70.32	400	150	600	411.86
Printer	80.85	400	124	750	452.08
Butcher	58.56	400	150	800	385.22
Machinist	83.41	450	156	720	485.17
Miner	57.75	450	200	900	513.77
Electrician	81.61	480	160	840	518.12
Hotel Keeper	32.28	520	200	1500	735.34
Policeman	70.46	528	274	840	568.83
Clergyman	26.37	600	180	1100	712.02
Insurance Agent	48.30	600	240	1500	799.49
Dentist	34.44	700	150	2200	397.44
Architect	48.29	800	150	2000	921.08
Barrister	28.97	1000	200	2500	762.5
Civil Engineer	63.11	1000	300	2300	1316.31

Notes: *Column (5) contains averages from Table II of the Census of 1901 report in *Bulletin 1, Wage-Earners by Occupation*; **female wages listed for teachers instead of male ones.

Table 4: Representativeness of Men Aged 18 to 65 in the Wage Sample

	All Men Aged 18 to 65		With Children aged 7-13		In occupations with med wage<200 (50pct)		In occupations with median wage>200 (50pct)		In occupations with median wage>500 (90pct)		In occupations with med wage>720 (99pct)	
	<i>Wage Amount</i> (1)	<i>Non-missing wage</i> (2)	<i>Wage Amount</i> (3)	<i>Non-missing wage</i> (4)	<i>Wage Amount</i> (5)	<i>Non-missing wage</i> (6)	<i>Wage Amount</i> (7)	<i>Non-missing wage</i> (8)	<i>Wage Amount</i> (9)	<i>Non-missing wage</i> (10)	<i>Wage Amount</i> (11)	<i>Non-missing wage</i> (12)
<i>Reads and Writes</i>	100.693*** (1.390)	0.018*** (0.001)	83.073*** (2.098)	-0.015*** (0.002)	77.190*** (2.687)	-0.028*** (0.002)	85.749*** (2.527)	0.119*** (0.003)	165.860*** (16.337)	0.107*** (0.014)	360.429*** (127.155)	0.253*** (0.054)
<i>Employer</i>	48.828*** (1.242)	-0.016*** (0.001)	56.810*** (1.989)	-0.022*** (0.002)	40.785*** (3.477)	0.014*** (0.002)	55.718*** (2.102)	-0.040*** (0.003)	25.838*** (8.695)	-0.016** (0.008)	-4.490 (41.135)	-0.007 (0.024)
<i>Servant in Home</i>	107.553*** (1.724)	-0.038*** (0.002)	126.617*** (2.796)	-0.044*** (0.003)	32.906*** (3.981)	0.031*** (0.003)	112.281*** (2.370)	-0.043*** (0.003)	236.299*** (7.862)	-0.081*** (0.007)	280.505*** (30.612)	-0.158*** (0.017)
<i>Mean months school</i>			3.024*** (0.164)	0.003*** (0.000)	3.703*** (0.264)	0.001*** (0.000)	2.454*** (0.179)	0.007*** (0.000)	3.685*** (0.751)	0.008*** (0.001)	1.655 (3.675)	0.004* (0.002)
Observations	422,443	1,007,216	169,506	417,448	43,011	266,502	158,485	230,905	18,862	28,947	1,393	3,075
R-squared	0.275	0.421	0.297	0.441	0.136	0.252	0.252	0.187	0.172	0.177	0.122	0.188
Adj R-squared	0.274	0.421	0.296	0.441	0.135	0.252	0.251	0.186	0.170	0.175	0.110	0.182

Notes : Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10. In Column 3-4, *mean months school* is the mean number of months in school for children aged 7-13. This age group was chosen because it is the group with at least 67% school attendance. In column 5 and 6, we restrict to men aged 18-65 in occupations with median wage equal or below the overall median of the medians for each occupation in the entire sample, which is \$200 per year. In columns 7 and 8, we restrict to men aged 18-65 in occupations with median wage equal or above \$200 per year. In columns 9 and 10, we restrict to men above the 90th percentile of the medians wages for each occupation, which is \$500 per year. In columns 11 and 12, we restrict to men above the 99th percentile, which \$720 per year.

Table 5: Occupational Status Rankings by Median Income, 1871-1891

Census Year	Percent of Population with an occupation	Mean of occupation median wage – men and women	Mean of occupation median wage - only men	Mean of occupation median wage - men aged 16-60
1871	26	260	265	269
1881	30	266	272	277
1891	34	272	282	288
1901	38	274	288	297

Notes: We compute the 1901 median wage for each occupation (using 3-digit occupation variable). Then, we assign the 1901 occupation's median wage to each individual with said occupation in every prior census year. Finally, we compute the mean for the population with a particular occupation in each year.