**Occupational income scores and immigrant assimilation. Evidence from the Canadian Census[[1]](#footnote-1)**

Kris Inwood (Guelph), Chris Minns (LSE), and Fraser Summerfield (St. Francis Xavier)

January 2019

Abstract: Little evidence is available to assess the effect of substituting occupation-based income scores for individual incomes before 1940. The example of immigrant assimilation in Canada 1911-1931 reveals differences in the extent and even the direction of assimilation depending on whether income scores are used and how the occupational income score is constructed. Given the increasingly wide use of income scores, we summarize a number of procedures to address the limitations associated with the absence of individual level income variation. An adjustment of conventional income scores for either group earnings differences and/or intertemporal change using summary information for broad groups of occupations reduces the deviation between scores and actual incomes.

**Introduction**

From the 1850s to the 1930s the integration of immigrants was a defining issue for in North American society, spawning a large public debate about the desirability of immigration (Abramitzky and Boustan 2017; Ferrie and Hatton 2015). And yet, because the census did not collect information about individual earnings until 1901 in Canada and 1940 in the United States, our understanding of immigrant assimilation in the labour market relies almost entirely on occupations and occupation-based income scores. This practice, adapted from an earlier literature in sociology (Duncan 1961; Sobek 1995, 1996; Hauser and Warren 1997), imputes fixed earnings by occupation based on averages in a later census for which both income and occupation are available, or from ancillary sources of pay by occupation. In the last three decades there has been considerable use of income scores for the economic analysis of assimilation (Chiswick 1991; Borjas 1992, 1994; Hanes 1996; Minns 2000; Abramitzky, Boustan, and Erikkson 2012, 2014) and ethnic inequality (Darity et al 1997; Horton et al 2000; Collins and Wannamaker 2014, 2015). Occupation-based income scores are also used to proxy for individual incomes on other topics including intergenerational economic mobility (Olivetti and Paserman 2015), inequality over time (Lindert and Williamson 2016; Modalsli 2015), fertility decline (Aaronson, Lange and Mazumder 2014), policy change (Chen 2015; Fagernäs 2014), schooling (Stephens and Yang 2014; Lleras-Muney and Shertzer 2015) and the early life origins of health and human capital (Bleakley 2007; Saavedra forthcoming).[[2]](#footnote-2)

While the use of occupational income scores has become common, it comes with several limitations which are important when estimating labour market convergence between immigrants and the native-born. The presence of these limitations make the immigrant assimilation debate a good arena in which to test the performance of occupational income scores, and, if possible, come up with improvements. First, because the income measure is fixed for all observations with the same occupation, the variance of any estimation is reduced. The compression of variance, by itself, will reduce the estimated earnings gap between immigrants and the foreign-born. Abramitzky, Boustan, and Eriksson (2014, p. 12, footnote 21) observe that differences in occupation account for a third of total earnings differences in 1970. Moreover, immigrants and the native-born may have large differences in age-earnings patterns within occupations that are unobserved by the researcher. Second, the usefulness of income scores constructed in one year for use in another year depends on changes in labour market skills and the accompanying income hierarchy of occupations (Goldin and Margo 1992; Goldin and Katz 2008; Katz and Margo 2013). To the extent that occupational titles carry an element of social status or prestige that changes more slowly than income (Erikson and Goldthorpe, 1992), the modest rates of earnings convergence implied by immigrant occupational change may understate assimilation by not taking account of changes within occupations. However, even the direction of immigrant assimilation may be estimated incorrectly if movement between occupations over time does not match trends in earnings between occupations.

All this is reasonably well known, but there is little evidence on the empirical consequences of using occupational income scores to estimate immigrant labour market assimilation.[[3]](#footnote-3) In this paper we use Canadian census data from 1911, 1921, and 1931, which include individual earnings, to estimate immigrant earnings assimilation. This is based on a synthetic cohort approach, as materials are not yet available to construct a panel of immigrants with Canadian sources. Our approach follows Feigenbaum in that we use occupational income scores when individual earnings are available in order to compare estimates of immigrant assimilation with alternate measures. We find differences in the extent and even the direction of assimilation depending on which measure is used.[[4]](#footnote-4) Our results also show, however, that some variants of occupational income scores perform better than others in terms of capturing differences in levels in earnings and the patterns of assimilation in individual earnings over time. We use these better performing scores to devise adjustments to standard occupational income scores that can be applied using ancillary evidence in the US context, and indeed elsewhere.

**A formal representation of occupational income scores**

Occupational income scores measure individual earnings with error that may result in a misrepresentation of the direction of assimilation and an understatement of differences between immigrant and native-born earnings. We represent the earnings of individual *i* in occupation *o, wio* with a surrogate value that varies only across occupation,

 (1).

If immigrant location in the within-occupation earnings distribution is not random, the error term *eio* varies systematically by immigrant status. This becomes a straightforward example of measurement error in the dependent variable that is correlated with an independent variable k rather than random (Woodridge 2009, chapter 9). Suppose the true data generating process for earnings is

 (2),

such that the vector captures various characteristics related to earnings. The error term for estimation of an equivalent specification with occupation-scores

 (3),

is given by.

 (4).

Estimates of will be biased if . Since many of the characteristics that affect cross-occupation earnings are also likely to affect within-occupation earnings, positive bias may occur for groups with earnings below . In other words, the returns to characteristics such as experience or time since migration may be overestimated for immigrant populations, with implications for the rate of assimilation since it is estimated by comparing such returns to the native-born (below). In our empirical model that follows, the potential bias if any would be found in the cohort dummy indicators that capture to returns to time spent in Canada among the foreign-born.[[5]](#footnote-5)

**Data**

Our assessment of the implications of occupation-fixed earnings uses three recently released Census samples that capture 5, 4, and 3 percent of the records from the original 1911, 1921 and 1931 Census enumerations, respectively. The data are nationally representative and report a wide range of relevant personal characteristics. In addition to the usual Census variables available in other countries, the Canadian authorities were among the first to ask respondents to report their earnings. Earnings coverage is reasonably complete for adult men in urban areas; most manual workers and white-collar employees reported annual pay. Responses were less consistent for professionals, and few farmers or self-employed reported their pay.[[6]](#footnote-6) As a result, our analysis excludes farm owners and operators, who make up a significant share of the working population.[[7]](#footnote-7) We further limit our attention to adult men between the age of 16 and 65 in each Census sample, separating the native-born from the foreign-born in most of what follows.[[8]](#footnote-8)

**Earnings Distributions**

A brief examination of earnings and occupations for the three Canadian censuses shows how life-cycle patterns differ between individual earnings and fixed occupational income scores. Panel (i) of Figure 1 presents unconditional age-earnings profiles for native-born adult men from pooled Census data from 1911, 1921, and 1931. The comparison of individual earnings by age with the analogue for occupational incomes using 1911-based income scores shows that differences are largest for workers below the age of 35, and to a lesser extent those above the age of 55. An income score, therefore, may do an acceptable job of predicting peak earnings potential, and yet fail to capture the growth of earnings with experience among young workers. Panel (ii) of Figure 1 shows that in comparison to the native-born, occupation scores overstate the earnings of immigrants over a wider age range, and by a much larger amount among younger immigrants. This is unsurprising given the age at which foreign-born workers often arrive, and the subsequent accumulation of host country-specific human capital. It means, however, that a common age adjustment to occupational scores which may be fairly effective in other circumstances (Saavedra and Twinam, 2018) is unlikely to work well for studies of immigrant assimilation.

How well do occupational income scores perform over time compared to individual data from the same occupations? Panel (i) of Table 1 shows unconditional mean earnings of major (1-digit) occupation groups relative to that of labourers. We compare individual earnings with own-year occupation earnings and 1950 IPUMS-USA earnings for 1911, 1921, and 1931. Unsurprisingly, the own-year-based occupational score (OCC-Y) does not stray far from actual earnings in most comparisons. The 1950-based score understates the mean earnings of higher skilled groups relative to labourers in 1931, and it understates the earnings of service workers relative to labourers in 1911 and 1921. At the level of individual occupations (panel ii), the 1950-based score markedly understates income relative to labourers for all categories in 1931 and for most although not all occupations in 1911 and 1921. The greater the earnings premium over labourers, the more 1950-based occupational earnings understate the earnings gap. This is consistent with the compression of earnings that took place in the US and Canada between 1930 and 1950 (Goldin and Margo, 1992; Green and Green 2016). The most important finding here is that 1950-based occupational earnings miss entirely the considerable increase in inequality experienced in 1931.

Table 1 ignores variance in earnings within occupation categories. The absence of such information may be problematic in the context of theories explaining immigrant earnings assimilation. Within-occupation gaps in pay might be associated with labour market discrimination against immigrants (Buffum and Whaples, 1995). Acquisition of host-country human capital, in the form of language and other capabilities, should raise productivity within occupations, leading to changing within-occupation earnings gaps with age and experience. Figure 3 shows earnings gaps between the native-born and immigrants for the six 1-digit occupation categories in the 1931 sample.[[9]](#footnote-9) The figure shows that immigrant earnings were lower within all broad groups.[[10]](#footnote-10) The figure also makes clear that the variation in earnings within 1-digit groups was wide, particularly among immigrants in lower skill groups. The implication is that there was plenty of scope for substantial life-cycle changes in pay within occupations.

**Immigrant Earnings Assimilation**

We use a simple Mincer-type earnings function to model relative immigrant earnings in each cross-section. The regression takes the following form:

 (5)

On the right-hand side of equation (5), age and the square of age trace out the usual profile of the returns to experience.[[11]](#footnote-11) *Xi* is a vector of additional personal and geographic characteristics – in our models, these include whether the respondent speaks English, and the province or region of residence.[[12]](#footnote-12) Finally, *Ci* is a vector of cohort identifiers for immigrants that that serve capture inter-cohort differences in relative earnings in each cross-section. The data include both native-born and immigrant men, so the pattern of *δ* coefficients across the three cross-sections provides an indicator of longitudinal earnings progression relative to the native-born among synthetic cohorts of immigrants, as well as differences in entry effects and the relative earnings of cohorts at different points in time.

The dependent variable *Yi* is either real individual earnings, or one of five occupational income scores.[[13]](#footnote-13) The differences in how these occupation scores are constructed reflect how information is potentially lost under different aggregation strategies. Four of these measures are constructed directly from the Census data. We use the 1950 US Census occupation codes as a common base for mean earnings measures at the 3-digit level. In cases where an occupation score would be based on fewer than 20 observations, that earnings measure is represented instead with the corresponding 1-digit mean.

 The first occupation code, OCC, assigns individuals across all census years the average real earnings of employed adult men in their occupational category in the 1911 Census sample. The second measure, OCC-Y, differs from the first by assigning occupation-mean earnings specific to each census year.[[14]](#footnote-14) Relative to the first measure, OCC-Y allows for shifts in the return to occupational skill over time. The third measure, OCC-G, employs separate occupational scores for three groups in the 1911 census: the native-born, “free” immigrants from the United States or Britain (including Ireland) and “other” immigrants from continental European countries[[15]](#footnote-15). A fourth specification, OCC-A5, estimates occupational earnings for 5-year age bands. The final measure, OCC 50, is the IPUMS-USA occupational income score, which has been widely used in the recent literature. This score assigns median occupation wages in the 1950 US census to individuals in a relevant occupation.[[16]](#footnote-16)

While our empirical approach follows well-established lines in the literature, two shortcomings should be acknowledged. We analyse synthetic cohorts rather than a true panel of individuals observed at more than one point in time. Our results are therefore sensitive to issues of selection in terms of who remained in Canada in successive census years, among both the foreign-born and native-born Canadians. Recent research drawing upon panel data of individual immigrants for the United States has shown that the presence of selective return migration can lead to an overstatement of the rate of immigrant assimilation in repeated cross-sections (Abramitzky, Boustan and Eriksson 2014; 2016). Second, because of under-reporting by proprietors and the self-employed and previous evidence of greater immigrant mobility into proprietor-type activity over time (Minns, 2000), we may not fully capture the dynamics of immigrant labour markets in urban areas.

 The full set of regression coefficients from our model are presented in a companion on-line appendix.[[17]](#footnote-17) In each case we restrict our attention to the cohort born between 1865 and 1895, who are working age in 1911 and remain thus throughout the three periods. To illustrate how estimates of assimilation are affected by the choice or outcome measure, Table 2 presents the coefficients on cohort indicators that capture the relative earnings of each group against the native-born in each Census after controlling for other differences relative to the native-born for the alternative income measures used. Table 2 shows that restricting earnings to a fixed occupation measure has a substantial effect on estimates of labour market assimilation, and on level differences in earnings between immigrants and the native-born. When using individual earnings, we document modest assimilation between 1911 and 1921 for cohorts arriving 1901 to 1910 (about 5 log points). In the next decade, the 1901-10 cohorts diverge from the native born slightly and those arriving between 1911 and 1920 maintain their gap to the native-born. This pattern confirms earlier findings from Inwood, Minns, and Summerfield (2016), which shows sharper earnings convergence between 1911 and 1921 when the immigrant population is split between “free” migrants from the United States and United Kingdom (including Ireland) and other European migrants, followed by a striking pattern of divergence between European migrants and the native-born between 1921 and 1931. Occupational income scores tell a different story, however. When we look at cohort trends over time through fixed 1911 occupational earnings (OCC) or 1950-based US scores (OCC 50), there is less assimilation for the 1901-10 cohorts to 1921, and occupation-based earnings measures converge further to 1931 despite a widening gap in actual earnings.

 How do alternative income scores perform? Table 2 also lists cohort dummies from models that use OCC-G, OCC-Y, and OCC-A5 as the dependent variable. Two of these alternatives appear to have some merit in our application. In 1911 and 1931, group-specific scores (OCC-G) comes closer to approximating the earnings gaps at the individual level than standard fixed occupational scores, as well as the other alternative scores presented in Table 2. While group income scores do best in terms of estimates of the size of earnings gaps in two of the three samples, it is less effective than year-specific scores in approximating the changes in the position of immigrant relative to the native born over time. To illustrate, consider results for the 1901-1905 arrival cohort. OCC-G predicts 0.5 percentage point earnings convergence between 1911 and 1921 versus 4.6 percentage points in individual earnings; it misses entirely the divergence in earning between 1921 and 1931. OCC-Y predicts 3.2 percentage point convergence in earnings between 1911 and 1921 (70 percent of the actual change), and 2.3 percentage points of divergence between 1921 and 1931 (46 percent of the actual change). Measures that differentiate income scores by age (OCC-A5) yield coefficients on the age terms that are much more in line within individual income, but do not perform better than conventional occupational income scores for the study of immigrant earnings assimilation.[[18]](#footnote-18)

This exercise yields two takeaways for researchers. First, cross-sectional studies of earnings differences by demographic groups, including but not limited to race, ethnicity, language or immigrant status, would benefit from the use of occupation scores generated from sources that allow for group-specific earnings measures. Second, in studies which trace out life-cycle earnings, sourcing year-specific occupational rankings may be more important than group- or age-specific rankings.

**Improving occupational income scores**

The analysis above demonstrates that own-year scores and scores disaggregated for groups of analytical interest improve estimates of immigrant assimilation. Of course, own-year scores in the absence of individual-level earnings are rare. In some contexts, however, ancillary information on changes in returns to occupations and differences in earnings between immigrants and the native-born may be available and, if so, offer a useful adjustment to occupational income scores. To illustrate how this process might work, we construct group and year adjusted occupational income scores, starting from our fixed 1911 occupational income score OCC. We do this by a) rescaling occupational income scores to reflect differences in occupational incomes scores at the 1-digit occupational cluster for immigrants and the native born, and b) rescaling occupational income scores in 1921 and 1931 to reflect changes in the returns to occupations (again, at the 1-digit level) over time. Under the first adjustment, immigrant earnings for carpenters, for example, would be adjusted to match overall immigrant/native born differences in income for skilled craft workers. Under the second adjustment, occupational earnings for stenographers in 1921 would be rescaled to reflect trends in earnings for clerical workers between 1911 and 1921. In generating new occupational income scores that incorporate these adjustments (both separately and combined), we are able to provide a rough simulation of the implications of using imperfect contemporaneous data to deal with group differences in earnings and changes in returns to skill over time.

 Earnings regressions using adjusted scores are detailed in our companion appendix on-line, with cohort indicator variables summarized in Table 3.[[19]](#footnote-19) Adjusted scores perform well in terms of replicating the advantageous features of group-specific and year-specific income scores that are unavailable to most researchers. These results show that our method more closely approximates the movement of actual earnings. For the 1901-1905 and 1906-1910 arrival cohorts, this is most visible for series where occupational earnings are adjusted for changes in the structure of earnings over time (year-adjusted), or for a combined adjustment for year and differences between immigrant and native-born groups (year/group adjusted). For later arrivals, the year-adjusted income scores do not track the direction of change in relative earnings as well as group-adjusted scores do, but both approaches (individually and combined) more closely match level differences in actual earnings than standard occupational income scores.

**Conclusions**

 How well do occupational income scores substitute for individual incomes? Feigenbaum (2015) shows that the absence of individual earnings variation does not appear to have major implications for studies of intergenerational mobility. This reflects the fact that peak-to-peak earnings comparisons of native-born generations are relatively immune to unobserved variation in earnings over the life-cycle.[[20]](#footnote-20) Our findings from immigrant earnings assimilation in Canada between 1911 and 1931 shows that occupational income scores are less helpful when level differences between groups and changes in earnings over time are an important consideration. Our assimilation estimates based on fixed, aggregate occupation scores typically used in the literature capture a limited share of actual earnings convergence, and underestimate the gap in earnings between the native-born and foreign-born. Individual earnings and occupational scores move in opposite directions in several of our models between 1921 and 1931, when immigrants experienced a decline in earnings without a corresponding deterioration in occupational profile.

Although occupation scores remain the best available surrogate for income in many research purposes, to our knowledge there has been no public discussion about best practice in their use. Our analysis yields a number of observations about how best to deal with the limitations imposed by occupational income scores. Some of the strategies that conform to best practice cohere with the approaches taken by previous researchers who were mindful of the challenged posed in using occupational income scores. First, a partial accommodation for the reduction of variance introduced by occupational scores is to cluster standard errors by occupational group (Cameron and Miller, 2015). Second, our findings support the disaggregation of income scores by occupation as much as available data will allow, since the severity of any distortion is affected by the number of occupational categories and the distribution of individuals across categories. The arrival of complete count US census data for 1940 offers greater potential for disaggregation. It is especially helpful to disaggregate by population sub-groups whose outcomes are being compared (Collins and Wannamaker 2014; 2015). Third, comparisons at the same point in their life cycle will be relatively unaffected by age-varying differences between individual earnings and averages for an occupational group (Feigenbaum 2015). Fourth, we find that decade- or year- specific income scores do a better job of capturing some of the dynamics of Canadian immigrant earnings between 1921 and 1931. More generally, in periods of rapid change in relative wages, such as the early to mid-twentieth century, it may be preferable to create occupational scores from contemporary but less representative sources rather than a census several decades later. Several early contributions to the occupational score literature take this route (Preston and Haines 1990; Borjas 1994), as do some recent studies (Abramitzky, Boustan, and Eriksson 2014; Salisbury 2014; Long and Ferrie 2015).

Researchers may not be able to implement the strategies above in all data sources. Our results suggest an important further possibility for researchers wishing to adopt a best practice approach: the adjustment of conventional income scores for either group earnings differences (immigrant or native-born) and/or intertemporal differences (changes in returns to types of skill over time) using ancillary information for broad groups of occupations. In our Canadian example adjusted scores perform much better than conventional income scores in terms of matching level and trends in earnings over time. This approach points to a low cost way to improve income scores for researchers who do not have individual earnings information in their main data source, but do have contemporaneous evidence from alternative sources on broad changes in the return to occupational skill or immigrant-native born earnings differences.

What form might this evidence take? For those interested in adjusting for changes in pay by occupation in American data, evidence on skill premia in Goldin and Katz (2001) offers a starting point. Another possibility would be to take advantage of more detailed evidence on changes in pay by occupation groups from Canadian Census data, as summarized in Green and Green (2016).[[21]](#footnote-21) One caution here would be that skill premia in Canada fell less than in the US between 1910 and 1940. To adjust group income scores to reflect the typical earnings of immigrants relative to the native-born, wage evidence from the Reports of the Immigration Commission (1911) is a source that could be used to good effect.

**References**

Daniel Aaronson, Fabian Lange, and Bashkar Mazumder (2014). “Fertility transitions along the extensive and intensive margins.” *American Economic Review*, 104, pp. 3701-3724.

Ran Abramitzky and Leah Boustan (2017). “Immigration in American Economic History”. *Journal of Economic Literature*, 55(4): 1311–1345.

Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson (2012). “Europe’s Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration.” *American Economic Review*, 102, pp. 1832–1856

Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson (2014). “A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration.” *Journal of Political Economy,* 122, pp. 467-506.

Hoyt Bleakley (2007). “Disease and Development: Evidence from Hookworm Eradication in the American South.” *Quarterly Journal of Economics*, 122, pp. 73-117

George J. Borjas (1985). “Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants.” *Journal of Labor Economics*, 3, pp. 463-489.

George J. Borjas (1992). “Ethnic Capital and Intergenerational Mobility.” *Quarterly Journal of Economics*, 107, pp. 123-150.

George J. Borjas (1994). “Long Run Convergence of Ethnic Skill Differentials: the Children and Grandchildren of the Great Migration.” *Industrial and Labour Relations Review*, 47, pp. 553-573.

George J. Borjas and Bernt Bratsberg (1996). “Who Leaves? The Outmigration of the Foreign-Born.” *Review of Economics and Statistics*, 78, pp. 165-176.

David Buffum and Robert Whaples (1995). “Fear and Lathing in the Michigan Furniture industry: Employee-based Discrimination a Century Ago.” *Economic Inquiry*, 33, pp. 234-252.

A. Colin Cameron and Douglas L. Miller (2015). “A Practitioner’s Guide to Cluster-Robust Inference.” *Journal of Human Resources*, 50, pp. 317-372.

Joyce J. Chen (2015). “The Impact of Skill-Based Immigration Restrictions: The Chinese Exclusion Act of 1882.” Journal of Human Capital, 9, pp. 298-328.

Barry R. Chiswick (1978). “The Effect of Americanization on the Earnings of Foreign-born Men.” *Journal of Political Economy*, 86, pp. 897-921.

Barry R. Chiswick (1986). “Is the New Immigration Less Skilled than the Old?” *Journal of Labour Economics*, 4, pp. 168-192.

Barry R. Chiswick (1991). “Jewish Immigrant Skill and Occupational Attainment at the

Turn of the Century.” *Explorations in Economic History*, 28, pp. 64-86

William J. Collins and Marianne H. Wanamaker (2014). "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data." *American Economic Journal: Applied Economics*, 6, pp. 220-52.

William J. Collins and Marianne H. Wanamaker (2015). “The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants.” *Journal of Economic History*, 75, pp. 947-992.

William Darity Jr., Jason Dietrich, David K. Guilkey (1997). “Racial and Ethnic Inequality in the United States: A Secular Perspective.” Am*erican Economic Review*, 87, pp. 301-305.

Otis D. Duncan (1961). “A Socioeconomic Index for All Occupations.” In *Occupations and Social Status*, edited by Albert J. Reiss, Jr., pp. 109–38. New York: Free Press.

R. Erikson, R. and J. H. Goldthorpe, J. H. (1992). *The constant flux: A study of class mobility in industrial societies*. Oxford University Press.

Sonja Fagernäs (2014). “Papers, please! The effect of birth registration on child labor and education in early 20th century USA.” *Explorations in Economic History*, 52, pp. 63-92.

James J. Feigenbaum, (2015) “Intergenerational Mobility During the Great Depression.” Unpublished working paper.

Joseph P. Ferrie and Timothy J. Hatton (2015). “Two Centuries of International Migration”. In Barry R. Chiswick and Paul W. Miller, ed. *Handbook of the Economics of International Migration,* volume 1: 53-88.

Claudia Goldin and Larry F. Katz (2008). *The Race Between Education and Technology*. Cambridge, MA: Harvard University Press.

Claudia Goldin and Robert A. Margo (1992). “The Great Compression: The Wage Structure In the United States at Mid-Century.” *Quarterly Journal of Economics* 107, pp. 1-34.

Alan G. Green and Mary E. MacKinnon (2001). “The Slow Assimilation of British Immigrants in Canada: Evidence from Montreal and Toronto, 1901.” *Explorations in Economic History* 38, pp. 315-338.

Alan G. Green and David A. Green (2016). “Immigration and the Canadian Earnings Distribution in the First Half of the Twentieth Century.” *Journal of Economic History*, 76, pp. 387-426.

Christopher Hanes (1996). “Immigrants’ Relative Rate of Wage Growth in the Late 19th Century.” *Explorations in Economic History*, 33, pp. 35-64.

R. M. Hauser and J.R. Warren (1997). 4. Socioeconomic Indexes for Occupations: A Review, Update, and Critique. *Sociological Methodology*, *27*, pp. 177-298.

Hayward Derrick Horton, Beverlyn Lundy Allen, Cedric Herring and Melvin E. Thomas (2000). “Lost in the Storm: The Sociology of the Black Working Class, 1850 to 1990.” *American Sociological Review*, 65, pp. 128-137.

Kris E. Inwood, Chris Minns, and Fraser Summerfield (2014). “Reverse Assimilation? Immigrants in the Canadian Labour Market During the Great Depression.” *European Review of Economic History*, 20, pp. 299-321.

Kris E. Inwood, Chris Minns, and Fraser Summerfield (2018). “Occupation scores in Canada 1911-1931” Open-ICPSR Repository <http://doi.org/10.3886/E107288V1>.

Larry F. Katz and Robert E. Margo (2014). “Technical Change and the Relative Demand for Skilled Labor: the United States in Historical Perspective.” In L. Boustan, C. Frydman, and R.A. Margo, eds., *Human Capital in History*. University of Chicago Press and NBER. pp. 15-57.

Adriana Lleras-Muney and Allison Shertzer (2015). “Did the Americanization Movement Succeed? An Evaluation of the Effect of English-Only and Compulsory Schooling Laws on Immigrants,” *American Economics Journal: Economic Policy*, 7, pp. 258-90.

Peter H. Lindert and Jeffrey G. Williamson (2016). *Unequal Gains: American Growth and Inequality since 1700*. Princeton University Press.

Chris Minns (2000). “Income, Cohort Effects, and Occupational Mobility: a New Look at Immigration to the United States at the Turn of the 20th Century.” *Explorations in Economic History*, 37, pp. 326-350.

J. Modalsli (2015). “Inequality in the very long run: inferring inequality from data on social groups.”  *Journal of Economic Inequality*, 13, pp. 225-247.

Claudia Olivetti and M. Daniele Paserman (2015). “In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940.” *American Economic Review* 105, pp. 2695-2724

Martin Saavedra (2017), “Early-Life Disease Exposure and Occupational Status: the Impact of Yellow Fever during the 19th Century.” *Explorations in Economic History*, forthcoming <http://dx.doi.org/10.1016/j.eeh.2017.01.003>.

Martin Saaverda and Tate Twinam (2918). “A Machine Learning Approach to Improving Occupational Income Scores.” SSRN Working Paper, [http://dx.doi.org/10.2139/ssrn.2944870](https://dx.doi.org/10.2139/ssrn.2944870)

Laura Salisbury (2014). Selective migration, wages, and occupational mobility in nineteenth century America.” *Explorations in Economic History* 53, PP. 40-63.

Matt Sobek (1995). “The comparability of occupations and the generation of income scores.” *Historical Methods,* 28, pp. 47-71.

Matt Sobek (1996). “Work, status and income: Men in the American occupational structure since the late nineteenth century.” *Social Science History,* 20, pp. 169-207.

Gary Solon (1992). “Intergenerational Income Mobility in the United States.” *American Economic Review,* 82, pp. 393-408.

Melvin Stevens, Jr. and Dou-Yan Yang (2014). “Compulsory Education and the Benefits of Schooling.” *American Economic Review,* 104, pp. 1777-1792.

Jeffrey M. Wooldridge (2009). *Introductory Econometrics, Fourth Edition.* Mason OH: South-Western,Cengage Learning.

Table 1: Earnings Measures for Occupation Groups and Major Occupations, Relative to Labourers

|  |  |  |  |
| --- | --- | --- | --- |
|  | **1911** | **1921** | **1931** |
|  | **Earnings** | **OCC-Y** | **OCC 50** | **Earnings** | **OCC-Y** | **OCC 50** | **Earnings** | **OCC-Y** | **OCC 50** |
| 1. **Occupation groups**
 |
| Prof/Prop | 2.24 | 2.39 | 2.00 | 1.87 | 2.02 | 1.94 | 3.50 | 3.68 | 1.86 |
| Clerical | 1.39 | 1.46 | 1.32 | 1.28 | 1.32 | 1.32 | 2.10 | 2.19 | 1.33 |
| Craftsmen | 1.48 | 1.47 | 1.51 | 1.48 | 1.46 | 1.55 | 2.03 | 2.00 | 1.52 |
| Operative | 1.11 | 1.10 | 1.17 | 1.11 | 1.08 | 1.19 | 1.71 | 1.68 | 1.22 |
| Service | 0.80 | 0.80 | 0.61 | 0.85 | 0.83 | 0.66 | 1.08 | 1.52 | 0.49 |
| ***Labour*** | **1.00** | **1.00** | **1.00** | **1.01** | **1.00** | **1.00** | **1.02** | **1.00** | **1.00** |
| 1. ***Occupations***
 |
| Mngrs, officials | 3.06 | 3.09 | 2.19 | 2.57 | 2.64 | 2.18 | 5.08 | 5.33 | 2.17 |
| Bookkeepers | 1.33 | 1.34 | 1.15 | 1.23 | 1.24 | 1.14 | 1.98 | 2.01 | 1.14 |
| Stenographers | 1.09 | 1.11 | 1.15 | 1.10 | 1.12 | 1.14 | 1.73 | 1.82 | 1.14 |
| Clerical and kin | 1.31 | 1.32 | 1.30 | 1.25 | 1.26 | 1.30 | 2.09 | 2.15 | 1.29 |
| Salesmen | 1.37 | 1.40 | 1.25 | 1.34 | 1.38 | 1.25 | 2.07 | 2.14 | 1.24 |
| Carpenters | 1.42 | 1.42 | 1.25 | 1.34 | 1.33 | 1.25 | 1.60 | 1.60 | 1.24 |
| Operative | 1.16 | 1.16 | 1.20 | 1.07 | 1.07 | 1.20 | 1.48 | 1.46 | 1.19 |
| ***Laborers (nec)*** | **0.98** | **0.97** | **1.04** | **0.98** | **0.97** | **1.04** | **0.99** | **0.98** | **1.03** |

Notes: Earnings measure is actual annual real earnings. OCC-Y is the own-year average occupational earnings. OCC 50 is the IPUMS occupational earnings measure based on the 1950 US Census. Sample weights applied. 1-Digit average for labourers used as base category in Panel (i). 3-digit average for “Labourers (nec)” uses as a base category in Panel (ii).

Table 2: Relative immigrant earnings by arrival cohort, year, and earnings measure

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Arrival Cohort | Outcome | 1911 | 1921 | 1931 |
| 1901-1905 | **ln(earnings)** | **-.144** | **-.098** | **-.144** |
| ln(OCC) | -.108 | -.084 | -.056 |
| ln (OCC-G) | -.148 | -.143 | -.124 |
| ln(OCC-Y) | -.108 | -.076 | -.097 |
| ln(OCC-A5) | -.097 | -.068 | -.063  |
| 1906-1910 | **ln(earnings)** | **-.181** | **-.149** | **-.162** |
| ln(OCC) | -.125 | -.103 | -.101 |
| ln (OCC-G) | -.162 | -.154 | -.158 |
| ln(OCC-Y) | -.125 | -.096 | -.138 |
| ln(OCC-A5) | -.114  | -.121  | -.111  |
| 1911-1915 | **ln(earnings)** |  | **-.177** | **-.175** |
| ln(OCC) |  | -.132 | -.098 |
| ln (OCC-G) |  | -.182 | -.154 |
| ln(OCC-Y) |  | -.121 | -.140 |
|  ln(OCC-A5) |  | -.147  | -.105  |
| 1916-1920 | **ln(earnings)** |  | **-.221** | **-.190** |
| ln(OCC) |  | -.105 | -.102 |
| ln (OCC-G) |  | -.155 | -.157 |
| ln(OCC-Y) |  | -.104 | -.140 |
|  ln(OCC-A5) |  | -.120  | -.113  |
| 1921-1925 | **ln(earnings)** |  |  | **-.384** |
| ln(OCC) |  |  | -.175 |
| ln (OCC-G) |  |  | -.218 |
| ln (OCC-Y) |  |  | -.268 |
| ln(OCC-A5) |  |  | -.192 |
| 1926-1930 | **ln(earnings)** |  |  | **-.524** |
| ln(OCC) |  |  | -.162 |
| ln (OCC-G) |  |  | -.203 |
| ln (OCC-Y) |  |  | -.268 |
| ln(OCC-A5) |  |  | -.197 |

Notes: Summary of immigrant arrival cohort regression coefficients from Appendix Tables A2-A4. Source data are urban males ages 16-65 reporting occupation and earnings from the Canadian Census 1911, 1921 and 1931. The native born are the reference group. Rows compare estimates of cohort effects on earnings as dependent variable changes from ln(earnings), real annual earnings, and four different occupational income scores constructed by aggregating Canadian Census earning measures in the three ways described in the text. Columns compare across census years.

Table 3: Relative immigrant earnings by arrival cohort, with adjusted income scores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Arrival Cohort | Outcome | 1911 | 1921 | 1931 |
| 1901-1905 | **ln(earnings)** | **-.144** | **-.098** | **-.144** |
| ln(OCC) group adjusted | -.187 | -.163 | -.141 |
| ln (OCC) year adjusted | -.129 | -.098 | -.119 |
| ln (OCC) group/year adjusted | -.187 | -.168 | -.187 |
|  |  |  |  |
| 1906-1910 | **ln(earnings)** | **-.181** | **-.149** | **-.162** |
| ln(OCC) group adjusted | -.213 | -.196 | -.188 |
| ln (OCC) year adjusted | -.159 | -.145 | -.193 |
| ln (OCC) group/year adjusted | -.213 | -.205 | -.255 |
|  |  |  |  |
| 1911-1915 | **ln(earnings)** |  | **-.177** | **-.175** |
| ln(OCC) group adjusted |  | -.223 | -.188 |
| ln (OCC) year adjusted |  | -.173 | -.192 |
| ln (OCC) group/year adjusted |  | -.233 | -.253 |
|  |  |  |  |
| 1916-1920 | **ln(earnings)** |  | **-.221** | **-.190** |
| ln(OCC) group adjusted |  | -.199 | -.199 |
| ln (OCC) year adjusted |  | -.150 | -.211 |
| ln (OCC) group/year adjusted |  | -.205 | -.265 |
|  |  |  |  |
| 1921-1925 | **ln(earnings)** |  |  | **-.384** |
| ln(OCC) group adjusted |  |  | -.270 |
| ln (OCC) year adjusted |  |  | -.333 |
| ln (OCC) group/year adjusted |  |  | -.396 |
|  |  |  |  |
| 1926-1930 | **ln(earnings)** |  |  | **-.524** |
| ln(OCC) group adjusted |  |  | -.277 |
| ln (OCC) year adjusted |  |  | -.349 |
| ln (OCC) group/year adjusted |  |  | -.415 |
|  |  |  |  |

Notes: Summary of immigrant arrival cohort regression coefficients estimates from Appendix Tables A5 to A7. Source data are urban males ages 16-65 reporting occupation and earnings from the Canadian Census 1911, 1921 and 1931. The native born are the reference group. Rows compare cohort effects as the dependent variable changes from the log of real annual earnings to three alternative income measures described in the text. Columns compare across census years.

Figure 1: Unconditional Age-earnings Profiles, Canada 1911-31

(i) Native Born (ii) Foreign Born

  

Notes: Means of log earnings measures by individual age from pooled Canadian Census data, 1911, 1921 and 1931. Data restricted to urban males with birth years 1866-1985. ln(Earnings) is real annual earnings reported in the data. ln(OCC) is the 3-digit 1911 occupational average earnings, imposed on the full pooled sample. Shading indicates 95% confidence interval.

Figure 3: Earnings by occupation and nativity group, 1931



Notes: Bars plot the log of actual real annual earnings of immigrants and the native-born in (1-digit) major occupation groups in the 1931 Canadian Census. Data restricted to urban males born 1886-1895. Whiskers indicate 2 standard deviations in earnings.

1. We thank Ian Keay, Ariel Zimran, seminar participants at LSE and Lakehead University, and conference participants at the 2015 Canadian Economics Association, the 2015 European Historical Economics Society meetings, and the 2016 Social Science History Association for helpful comments. We have also benefitted from advice from the editor of this journal and two anonymous referees. This research was supported by funds to the Canadian Research Data Centre Network (CRDCN) from the Social Sciences and Humanities Research Council (SSHRC), the Canadian Institute for Health Research (CIHR), the Canadian Foundation for Innovation (CFI), and Statistics Canada. Although the research and analysis are based on data from Statistics Canada, the opinions expressed do not represent the views of Statistics Canada. A full Appendix is available on-line at <http://www.lse.ac.uk/Economic-History/Assets/Documents/WorkingPapers/Economic-History/2018/WP292.pdf>. Stata codes and instructions on how to access CCRI data are posted in Open-ICPSR at <http://doi.org/10.3886/E107288V1>. [↑](#footnote-ref-1)
2. A comprehensive list of studies using occupational income scores is listed in an accompanying on-line appendix (http://www.lse.ac.uk/Economic-History/Assets/Documents/WorkingPapers/Economic-History/2018/WP292.pdf). [↑](#footnote-ref-2)
3. The performance of income scores has been examined in more detail when estimating intergenerational economic mobility. Feigenbaum (2015) argues that income scores perform reasonably well as compared to individual earnings, while Saavedra and Twinam (2018) propose characteristics-adjusted income scores that yield plausible results. In a different literature, Modalsli (2915) attempts to correct for missing within-group variation in calculating historical income inequality. [↑](#footnote-ref-3)
4. More detail on immigrant assimilation in early 20th century Canada is available in Inwood, Minns, and Summerfield (2016). [↑](#footnote-ref-4)
5. Even if panel data are available, fixed-effects estimates will not account for any time-varying components of *eio*. Because fixed-effect is a within-estimator, the variation used in estimation will come only from individuals for whom the dependent variable changes over time, in this case occupation switchers. Thus $\hat{β}\_{k}$ will be a good estimate for the select group of individuals who change occupation while remaining at the same relative location within the earnings distribution of their new occupation, so as to make *eio* time invariant. [↑](#footnote-ref-5)
6. Over 80 percent of workers we would normally think of as employees reported earnings in all three censuses, but less than 20 percent of farmers, and only 50 to 60 percent of professionals and proprietors. [↑](#footnote-ref-6)
7. “Farmer” is probably the most difficult occupation to represent with an occupation score (see Minns, 2000). [↑](#footnote-ref-7)
8. More detailed results that break the foreign-born population into “free” migrants arriving from English-speaking countries with few policy restrictions (United States, United Kingdom, and Ireland) and “other” migrants arriving mostly from continental Europe. These results are available from the authors on request. [↑](#footnote-ref-8)
9. Similar figures for 1911 and 1921 are available on request. [↑](#footnote-ref-9)
10. Difference in log earnings are much larger between the native-born and immigrants from continental Europe. [↑](#footnote-ref-10)
11. Experimentation with higher order polynomials in previous research did not change the overall pattern of results in our sample. [↑](#footnote-ref-11)
12. We combine Nova Scotia, New Brunswick, and Prince Edward Island into one Maritime region, and present-day Saskatchewan and Alberta into one Northwest Region. [↑](#footnote-ref-12)
13. Real earnings measures derived using a price index from the appendix of Inwood, Minns, and Summerfield (2016) [↑](#footnote-ref-13)
14. In 1911, OCC is therefore the same as OCC-Y. [↑](#footnote-ref-14)
15. We do not estimate a model with separate coefficients for the two immigrant groups because the main consequences for occupational income scores are evident in simpler specifications. Alternative models that illustrate separate wage penalties for the two immigrant groups are available on request. [↑](#footnote-ref-15)
16. IPUMS variable name “OCCSCORE”. See documentation at the Minnesota Population Center for the construction of 1950 occupation scores: <https://usa.ipums.org/usa-action/variables/OCCSCORE#description_section>. An additional score, specific to province and census year, was also generated but did not provide any novel results [↑](#footnote-ref-16)
17. Please see tables A2 to A4 at http://www.lse.ac.uk/Economic-History/Assets/Documents/WorkingPapers/Economic-History/2018/WP292.pdf. [↑](#footnote-ref-17)
18. The full set of coefficients for these models are presented in Tables A2 to A4 in the on-line appendix at http://www.lse.ac.uk/Economic-History/Assets/Documents/WorkingPapers/Economic-History/2018/WP292.pdf. [↑](#footnote-ref-18)
19. See appendix tables A5 to A7 at http://www.lse.ac.uk/Economic-History/Assets/Documents/WorkingPapers/Economic-History/2018/WP292.pdf. [↑](#footnote-ref-19)
20. Solon (1992) argues that occupational earnings are less likely to be tainted by measurement error than individual earnings that reflect transitory shocks. [↑](#footnote-ref-20)
21. Table 3 of Green and Green (2016) lists differences in average (log) wages by occupation and age. [↑](#footnote-ref-21)