ONLINE APPENDICES FOR:

Intergenerational Mobility in a Mid-Atlantic Economy: Canada,1871-1901

LUIZA ANTONIE, KRIS INWOOD, CHRIS MINNS, AND FRASER SUMMERFIELD

The Journal of Economic History, Vol. 82, No. 4 (December 2022).

APPENDIX A: OVERVIEW OF RECORD LINKING METHODOLOGY

Identification of the same individuals in successive censuses relies on standard computing science methodology (Felligi and Sunter 1969; Winkler 2006; Christen 2012). Below is a summary of the two-stage method we have employed with Canadian census records; further details are reported elsewhere (Richards 2013; Antonie et al 2014a, 2014b, 2015, 2020).

In the first stage we compare each record at the beginning of a decade with every record at the end of the decade after excluding or blocking some pairs that are unlikely to be matched. Matching is based on a small set of time-invariant characteristics (Ruggles 2006). We require the same sex, consistent marital status, same birthplace (country, or province if born in Canada), same first letter of surname and edit distance between surnames less than 0.15. Edit distance is the minimum number of edit operations needed to transform one string into the other relative to the total number of characters. We also block on first name groups that encompass nicknames, abbreviations, misspellings and, importantly, French and English forenames (eg, Guillaume is combined with William and Bill; Beth is combined with Liz and Isabelle). Within each block, similarity is established using reported age (accepting +/- 2 years as identical), and first and last name similarities. Name similarity is assessed using edit distance, Jaro-Winkler and double metaphone metrics.

We use these features to characterize a hyperplane that classifies each record pair as a match or a non-match. The classification process relies on 11,700 record pairs known to be correct after detailed examination in studies that were conducted independently and had access to additional information. These 'true links' are selected to represent Canadian diversities – Anglophone families in Toronto, Francophones in Quebec City, the entire population of a farming township and a heterogenous group of industrial proprietors (two-thirds of whom lived on farms).[1] We randomly divide the true links into five parts; four parts are used for training and 1 part for testing. This is done five times, each using a different 1/5 for testing. This assessment process guides development of our model and confirms that fewer than 5% of links generated by the model are incorrect.

This first stage successfully, ie uniquely, links only 15% of the 1871 records (1871 records matching with one and only one 1881 record, and the 1881 record not matched with any other 1871 record). Many more records are multiply linked. By this, we mean the 1871 record is matched to more than one 1881 record or it is one of several 1871 records matched to a single 1881 record.

In a second stage we resolve ambiguity for a subset of the multiply-linked records using a Jaccard similarity measure that identifies the total number of items in the intersection

[1] Baskerville (2015); Hinton (2010); Inwood and Reid (2001). We confirm true links by 1) finding in both censuses at least one other household member with matching vital information, 2) ensuring consistency with church records where available (Toronto and Quebec City), 3) ensuring that significant contradictory information makes a link improbable (for example, when one family member matches, but three others do not) and 4) determining there is no other likely match in the 1881 Canadian census or the 1880 U.S. census.

of the two households (in different years) and divides it by the total number of items in the union of the two households (Richards 2013). Disambiguation roughly doubles the size of linked sample with no deterioration of the false positive rate. A careful comparison of first stage and second stage linked data shows that addition of a second stage does not, on balance, increase the selectivity of linking (Antonie et al 2020). These results are similar to those of Helgertz et al (2022) linking U.S. data over a single decade. Nevertheless, our linked records at both stages are not representative of the population, as Ferrie (1996) would have predicted. Consequently, we follow Bailey at al (2020) in reweighting observations for our inference about patterns in the broader population.

We link, independently, over each of the three decades and then use only those people we can follow through each decade. We do this, rather than linking 1871-1901 directly, for several reasons. (i) Unlike some previous researchers we have access to data for all decadal enumerations. (ii) One of our research objectives is to examine, in a separate paper, short-term life trajectories within the 30-year span. (iii) Obtaining 'true links' over a 30-year span would be difficult. (iv) Any bias at the second stage arising from the reliance on continued co-residence is minimized by limiting the time to ten years. (v) Most importantly, linking over a shorter span generates fewer false positive links or errors. The disappearance of someone due to death, migration or mis-reporting, in the presence of considerable name-age-birthplace duplication, is a common source of incorrect links. If the correct person is not available to be linked, or is reported imprecisely, in many cases an incorrect link with near identical characteristics will be accepted. Extending the link span to three decades exacerbates the problem because, on average, memory and perhaps also the commitment to precise reporting deteriorates with time. Loss of precision due to incorrect links is especially damaging because the Canadian population was small.

More than one-half of the linked boys are linked uniquely in each decade; the remainder rely on disambiguation for one or more decades. One-quarter of the records rely on disambiguation for all decades. The share of disambiguated records for each decade is reported below. A decline in the effectiveness of disambiguation in the 1880s and 1890s reflects the departure of many boys, who were younger than 15 years in 1871, from their parents' home.

Distribution of type of link made in each decade

	First stage	Second stage (disambiguation)
1871-1881	45%	55%
1881-1891	66%	34%
1891-1901	70%	30%

REFERENCES

Antonie, Luiza, Kris Inwood, Dan Lizotte, and J. Andrew Ross (2014a) "Tracking People over Time in 19th Century Canada." *Machine Learning* 96 (S1): 129-146.

Antonie, Luiza, Shawki Areibi, Gary Grewal, Kris Inwood, Laura Richards, and J. Andrew Ross (2014b) "Comparing Classifiers in Historical Census Linkage", In

Proceedings of the 2014 IEEE Data Mining Workshop (ICDMW), pp 1086-1094, DOI: 10.1109/ICDMW.2014.160.

Antonie, Luiza, Kris Inwood, and J. Andrew Ross (2015) "Dancing with Dirty Data: Problems in the Extraction of Life-course Evidence from Historical Censuses," in Gerrit Bloothooft, Peter Christen, Kees Mandemakers and Marijn Schraagen (eds.) Population Reconstruction. Cham, Switzerland: Springer International Publishing AG: 217-241.

Antonie, Luiza, Kris Inwood, Chris Minns and Fraser Summerfield, "Selection bias encountered in the systematic linking of historical records", *Social Science History* 44 (2020): 555-570

Bailey, Martha, Connor Cole, and Catherine Massey (2020), "Simple strategies for improving inference with linked data: a case study of the 1850-1930 IPUMS linked representative historical samples." *Historical Methods*, 53(2), pp. 80-93.

Baskerville, Peter (2015) "Wilson Benson Revisited: Movement and Persistence in Rural Perth County, Ontario, 1871-1881." In Peter Baskerville and Kris Inwood, eds., *Lives in Transition: Longitudinal Research from Historical Sources*. McGill–Queen's University Press, pp. 141-164.

Christen, Peter (2012) *Data Matching - Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Cham, Switzerland: Springer International Publishing AG.

Fellegi, Ivan P. and A. B. Sunter (1969) "A theory for record linkage." *Journal of the American Statistical Association* 64: 1183-1210.

Ferrie, Joseph P. (1996) "A new sample of males linked from the Public Use Micro Sample of the 1850 U.S. federal census of population to the 1860 U.S. federal census manuscript schedules." *Historical Methods* (29): 141–56.

Helgertz, Jonas, Joseph Price, Jacob Wellington, Kelly J Thompson, Steven Ruggles and Catherine A. Fitch (2021), "A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel." *Historical Methods*, 55, pp. 11-28

Hinson, Andrew (2010) *Migrant Scots in a British City: Toronto's Scottish Community, 1881-1911*. PhD diss. University of Guelph.

Inwood, Kris and Richard Reid, Richard (2001) "Gender and Occupational Identity in a Canadian Census." *Historical Methods* 32, pp. 57-70.

Richards, Laura (2013) *Disambiguating Multiple Links*. University of Guelph MSc thesis.

Ruggles, Steven (2006) "Linking historical censuses: A new approach." *History and Computing* 14 (1–2): 213–24.

Winkler, William E. (2006) *Overview of record linkage and current research directions*. United States Census Bureau Research Report Series: Statistics #2006-2.

APPENDIX B: SUPPLEMENTAL TABLES AND FIGURES

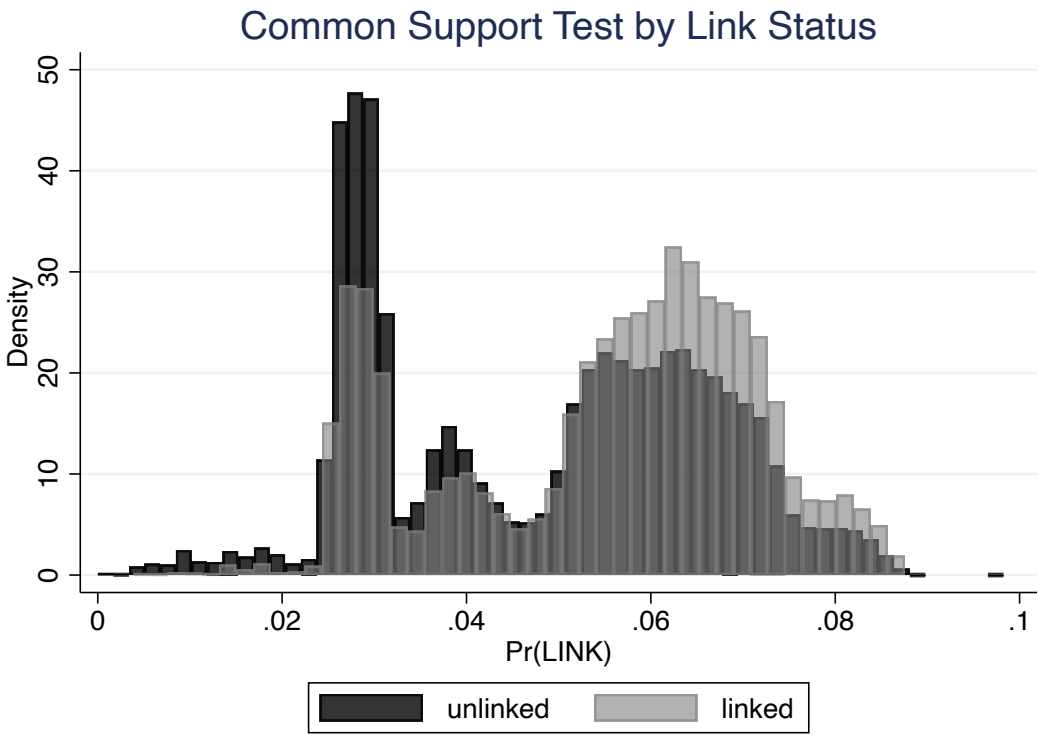

FIGURE B.1
DIAGNOSTICS FOR INVERSE PROPENSITY SCORE LINKING WEIGHTS

Note: Pr(LINK) are fitted values from probit regression for successful linkage of observations from 1871 full count. Relative densities shown over support [0,1].

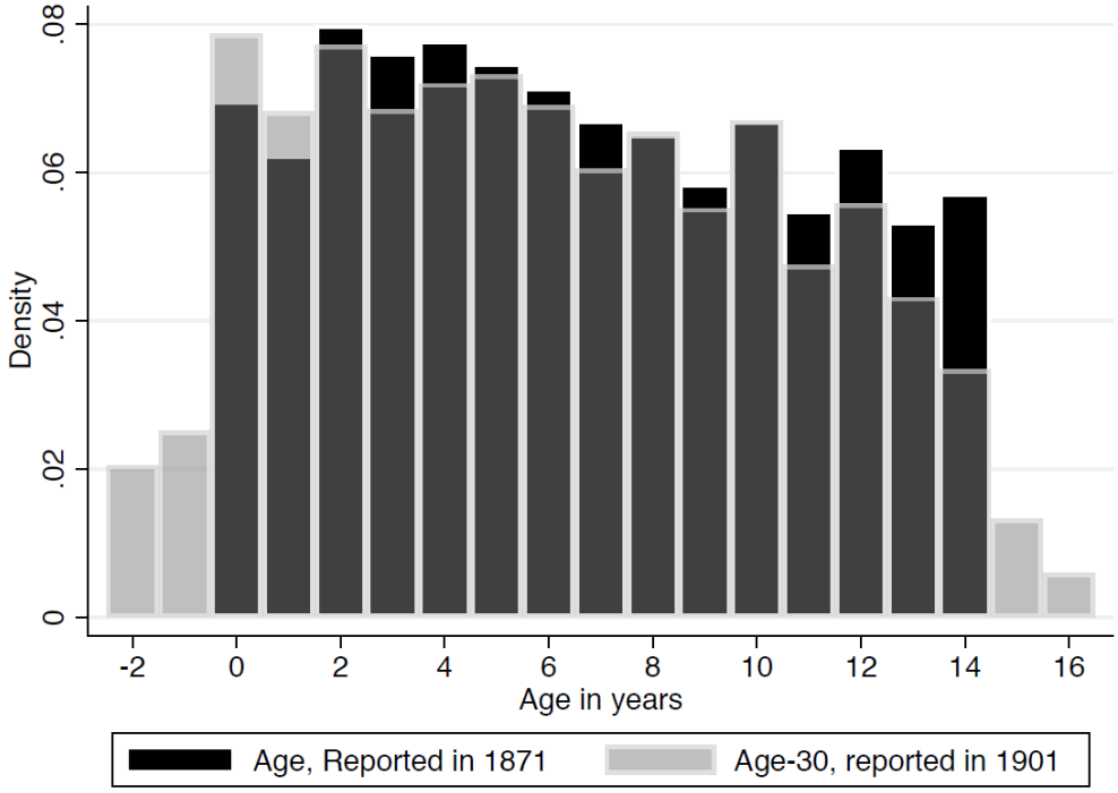

FIGURE B.2
HISTOGRAM OF AGE DISTRIBUTIONS 1871 AND 1901 FOR LINKED OBSERVATIONS

Note: Our linkage procedure allows for 2-year age discrepancy between 1871 and 1901 to account for differences in survey collection timing over the 4 census rounds that we link.

APPENDIX C: RESULTS USING OCCHISCO GROUPINGS

This section contains a full set of results computed using alternative occupation classifications. We group OCCHISCO codes into 4 categories similar to Long and Ferrie (2013). White collar includes proprietors and professionals, clerical works and sales workers (Codes 0 – 30000). Unskilled comprises service workers, labourers including agricultural labourers, primary industry workers and farmer's sons (codes 50000 – 60000; 62710 – 69999; 99120 – 99439). Farm includes farmers and farm owners (codes 6000 – 62709). Skilled/semi-skilled includes craftsmen and operatives (all remaining codes ranging from 70000 – 98900).

Table C.1 below shows that the main difference between the two groupings is the assignment of skilled/semi-skilled and workers white collar workers. There are also a few observations that are lost because we are unable to assign them a HISCLASS coding. However, the main results of the paper are largely unchanged regardless of which occupational grouping we use.

TABLE C.1
CONCORDANCE BETWEEN 4-GROUP OCCUPATIONAL ASSIGNMENT METHODS

L&F Group	HISCLASS Groupings: Sons 1901					
Sons 1901	Unclass.	White	Skill/Semi	Unskilled	Farm	Total
Unclassified	2,357	0	0	0	0	2,357
White	4	4,232	246	11	0	44,93
Skill/Semi	1	242	5,731	684	0	6,658
Unskilled	43	310	259	7,864	9	8,485
Farm	10	9	0	0	13,207	13,226
Total	2,415	4,793	6,236	8,559	13,216	35,219

L&F Group	HISCLASS Groupings: Fathers 1871					
Fathers 1871	Unclass.	White	Skill/Semi	Unskilled	Farm	Total
Unclassified	421	4	0	0	0	425
White	6	2,331	95	35	0	2,467
Skill/Semi	0	114	5,198	425	0	5,737
Unskilled	6	288	105	4,001	4	4,404
Farm	66	0	0	0	22,120	22,186
Total	499	2,737	5,398	4,461	22,124	35,219

Note: See main text for details on occupational coding. L&F Grouping matches closely to groups from Long and Ferrie (2013).

TABLE C2

LINKED AND UNLINKED SAMPLE CHARACTERISTICS, 1871

	(1) 1871 full count	(2) 1871-1901 linked	(3) 1871-1901 linked, weighted	(4) Unique links
1871 Age	6.8 (4.3)***	6.6 (4.2)	6.9 (4.3)	6.9 (4.2)***
% hhlds with 5+ children	0.55***	0.53	0.56	0.55***
Born NS	0.10***	0.14	0.11	0.15***
Born NB	0.08***	0.10	0.08	0.12***
Born QC	0.34***	0.22	0.33	0.21
Born ON	0.44***	0.53	0.45	0.49***
Born UK & Ireland	0.02**	0.01	0.01	0.01**
Born Elsewhere	0.04***	0.01	0.03	0.03
Reside NS	0.11	0.14	0.11	0.16***
Reside NB	0.08	0.10	0.08	0.12***
Reside QC	0.34	0.22	0.33	0.21
Reside ON	0.48	0.54	0.48	0.52***
Head white collar	*0.08*	0.07	0.07	0.07*
Head skilled/semi skilled	*0.19*	0.17	0.16	0.17**
Head unskilled	*0.19*	0.13	0.14	0.13
Head farm	*0.54*	0.64	0.64	0.62***
French Eth.	0.32***	0.19	0.31	0.18
Anglo Eth.	0.60***	0.71	0.60	0.68***
No Female >22 in hhld	0.03***	0.03	0.03	0.02***
N	733,355	32,484	32,484	17,309

Note: See text for sample descriptions. *, **, and *** denote significant differences between each unweighted sample and the linked sample in column (2) at 90, 95, and 99 percent confidence intervals. Unique links refers to people linked uniquely in all three decadal spans: 71-81, 81-91 and 91-01. Head occupation "unclassified" are omitted. Column (1) is identical the same column in main text Table 1. "Full count" sample limited to males aged 0 to 14 in 1871. Children defined as individuals enumerated with the same household id age 0-17, inclusive. Column 1 numbers in italics (father occupations) generated from the 7% 1871 census sample file. Anglophone includes ethnicities reported as English, Welsh, Scottish, Irish and North American.

TABLE C3
FULL CANADIAN TRANSITION MATRIX, 1871-1901

	Father				
Son	White Collar	Skilled/Semi-Skilled	Unskilled	Farm	Total
a) Unweighted					
White collar	1,134	1,103	488	1,654	4,379
	(.51)	(.21)	(.12)	(.08)	
Skilled/semi-skilled	495	2,217	1,098	2,737	6,547
	(.22)	(.41)	(.27)	(.13)	
Unskilled	292	1,085	1,592	54,38	8,407
	(.13)	(.20)	(.39)	(.26)	
Farm	323	949	931	10,948	13,151
	(.14)	(.18)	(.23)	(.53)	
Total	2,244	5,354	4,109	20,777	32,484
b) Weighted					
White collar	22,505	22,362	10,850	32,442	88,158
	(.49)	(.21)	(.12)	(.08)	
Skilled/semi-skilled	10,858	45,121	25,173	56,237	137,388
	(.24)	(.42)	(.28)	(.13)	
Unskilled	5,802	21,402	33,313	108,028	168,544
	(.13)	(.20)	(.37)	(.25)	
Farm	6,705	18,162	21,189	228,871	274,926
	(.15)	(.17)	(.23)	(.54)	
Total	45,868	107,046	90,524	425,578	669,016

Note: See main text for details of occupational coding. Column shares in parentheses. Weighted values rounded to nearest integer. Linkage weights similar to Bailey (2020) are described in the main text.

TABLE C4
CANADIAN ALTHAM STATISTICS

	(1) $d(\boldsymbol{P},\boldsymbol{J})$	(2) G^2	(3) $d(\boldsymbol{Q},\boldsymbol{J})$	(4) G^2	(5) $d(\boldsymbol{P},\boldsymbol{Q})$	(6) G^2
CAN 1871-1901	*16.2*	*6,570***				
CAN 1871-1901, weighted	*16.3*	*5,187***				
ONT 1871-1901	*15.4*	*3,358***				
QUE 1871-1901			*17.8*	*1,688***	*5.1*	*87.9****
MAR 1871-1901			*17.6*	*1,701***	*5.9*	*169****
Franco 1871-1901	17.7	1,255***				
Anglo 1871-1901			16.0	4,788***	4.1	41.7***
Franco in Quebec	17.4	1,053***				
Anglo in Quebec			18.2	602***	6.0	27.5***
Franco outside Quebec	19.6	204***				
Anglo outside Quebec			15.8	4,213***	6.2	17.5**

Note: G^2 for weighted sample calculated using normalized inverse propensity score weights that sum to N.

TABLE C5
TWO-WAY ODDS RATIOS OF RELATIVE REPRESENTATION OF SONS BY FATHER OCCUPATION

(a) Canada

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	8.5 (0.046)	1.9 (0.039)	0.8 (0.051)	0.3 (0.034)
Skilled/Semi-skilled	1.1 (0.053)	3.7 (0.032)	1.5 (0.038)	0.3 (0.028)
Unskilled	0.4 (0.064)	0.7 (0.037)	2.0 (0.035)	1.0 (0.026)
Farm	0.2 (0.061)	0.3 (0.038)	0.4 (0.039)	4.8 (0.027)

(b) Ontario

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	7.5 (0.063)	1.9 (0.051)	0.9 (0.070)	0.3 (0.044)
Skilled/Semi-skilled	1.1 (0.075)	4.1 (0.044)	1.9 (0.054)	0.3 (0.040)
Unskilled	0.4 (0.090)	0.6 (0.053)	1.3 (0.053)	1.4 (0.038)
Farm	0.2 (0.082)	0.3 (0.050)	0.5 (0.053)	4.2 (0.037)

(c) Quebec

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	9.3 (0.094)	2.4 (0.086)	0.9 (0.107)	0.2 (0.077)
Skilled/Semi-skilled	1.3 (0.102)	3.6 (0.072)	1.4 (0.080)	0.4 (0.060)
Unskilled	0.3 (0.146)	0.8 (0.086)	2.3 (0.075)	0.9 (0.059)
Farm	0.2 (0.120)	0.2 (0.093)	0.4 (0.080)	5.9 (0.060)

(d) Maritimes

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	10.6 (0.097)	1.6 (0.083)	0.7 (0.106)	0.3 (0.074)
Skilled/Semi-skilled	1.0 (0.111)	3.2 (0.063)	1.1 (0.074)	0.4 (0.056)
Unskilled	0.4 (0.119)	0.7 (0.065)	3.0 (0.063)	0.8 (0.049)
Farm	0.2 (0.148)	0.3 (0.075)	0.2 (0.088)	5.4 (0.058)

(e) Anglophone

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	8.3 (0.052)	1.8 (0.044)	0.9 (0.059)	0.3 (0.039)
Skilled/Semi-skilled	1.1 (0.061)	3.7 (0.038)	1.6 (0.046)	0.3 (0.034)
Unskilled	0.4 (0.073)	0.7 (0.043)	1.8 (0.043)	1.2 (0.032)
Farm	0.2 (0.070)	0.3 (0.044)	0.4 (0.048)	4.7 (0.032)

(f) Francophone

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	9.0 (0.118)	2.5 (0.100)	0.7 (0.132)	0.3 (0.089)
Skilled/Semi-skilled	1.3 (0.127)	4.0 (0.079)	1.3 (0.086)	0.3 (0.066)
Unskilled	0.4 (0.166)	0.8 (0.090)	2.5 (0.076)	0.8 (0.062)
Farm	0.3 (0.142)	0.2 (0.101)	0.4 (0.082)	5.2 (0.064)

(g) Quebec Francophones

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	8.2 (0.127)	2.4 (0.110)	0.8 (0.146)	0.3 (0.096)
Skilled/Semi-skilled	1.4 (0.134)	4.3 (0.087)	1.4 (0.095)	0.3 (0.071)
Unskilled	0.4 (0.179)	0.7 (0.103)	2.4 (0.087)	0.8 (0.070)
Farm	0.3 (0.151)	0.2 (0.114)	0.4 (0.093)	5.4 (0.071)

(h) Quebec Anglophones

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	9.6 (0.147)	2.3 (0.140)	1.2 (0.164)	0.1 (0.135)
Skilled/Semi-skilled	1.1 (0.166)	2.2 (0.139)	1.6 (0.152)	0.5 (0.113)
Unskilled	0.3 (0.258)	0.9 (0.158)	2.0 (0.148)	1.1 (0.114)
Farm	0.2 (0.201)	0.2 (0.168)	0.3 (0.166)	7.1 (0.114)

(i) Ontario/Maritime Francophones

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	16.4 (0.324)	3.0 (0.239)	0.6 (0.309)	0.2 (0.232)
Skilled/Semi-skilled	0.9 (0.420)	3.0 (0.196)	1.1 (0.204)	0.5 (0.167)
Unskilled	0.3 (0.443)	0.8 (0.188)	2.4 (0.158)	0.7 (0.133)
Farm	0.3 (0.416)	0.3 (0.216)	0.4 (0.173)	4.1 (0.143)

(j) Ontario/Maritime Anglophones

Sons	Fathers			
	White Collar	Skilled/Semi-skilled	Unskilled	Farm
White Collar	8.0 (0.056)	1.8 (0.046)	0.9 (0.063)	0.3 (0.041)
Skilled/Semi-skilled	1.1 (0.066)	3.8 (0.039)	1.6 (0.049)	0.3 (0.036)
Unskilled	0.4 (0.076)	0.6 (0.045)	1.8 (0.045)	1.2 (0.033)
Farm	0.2 (0.075)	0.3 (0.045)	0.4 (0.05)	4.5 (0.034)

Note: Authors' calculations $e^{\Theta_{i,j}}$. Odds ratios $\Theta_{i,j}$ are calculated following equation (2). Standard errors in parentheses calculated from equation 3.1 in Agresti (2002)

TABLE C6
PARTIAL ALTHAM STATISTICS FOR CANADA 1871-1901

	White Collar		Skilled / Semi-Skilled		Unkilled		Farm	
	d_W	d_{NW}	d_S	d_{NS}	d_U	d_{NU}	d_F	d_{NF}
CAN 1871-1901	12.0	10.9	6.7	14.8	6.9	14.7	10.7	12.2

Notes: See main text for calculation details.

APPENDIX D: RESULTS WITH 5 OCCUPATION GROUPS

The main results of this paper suggest that intergenerational immobility among white collar occupations is a defining feature of the Canadian labour market from 1871-1910. However, this pattern may be partly due to implicit thresholds in occupational group assignments.

To further test sensitivity among white collar workers, we present results here that are based on two different 5-group OCCHISCO systems, similar to the robustness exercises presented in Perez (2019). In section D.1, white collar workers are split into high-white collar (professional workers with HISCLASS codes 1-3) and low-white collar (clerical and sales workers with HISCLASS codes 4 and 5). In section D.2, we split unskilled workers into unskilled farm workers (HISCLASS 10 and 12) and unskilled non-farm workers (HISCLASS 11).

D.1 High and Low White Collar distinction

TABLE D.11
FULL CANADIAN TRANSITION MATRIX, 1871-1901

Son	Father					
	High White Collar	Low White Collar	Skilled/semi-skilled	Unskilled	Farm	Total
a) Unweighted						
High White collar	233	218	384	163	702	1,700
	(0.24)	(0.14)	(0.08)	(0.04)	(0.03)	
Low White Collar	215	549	757	336	1,108	2,966
	(0.22)	(0.36)	(0.15)	(0.08)	(0.05)	
Skilled/semi-skilled	203	342	2,049	980	2,548	6,122
	(0.21)	(0.23)	(0.41)	(0.23)	(0.12)	
Unskilled	156	196	925	1,768	5,420	8,465
	(0.16)	(0.13)	(0.18)	(0.42)	(0.26)	
Farm	152	215	907	927	10,911	13,112
	(0.16)	(0.14)	(0.18)	(0.22)	(0.53)	
Total	959	1,520	5,023	4,174	20,689	32,365
b) Weighted						
High White collar	4,528	4,318	8,103	3,593	13,728	34,270
	(0.23)	(0.14)	(0.08)	(0.04)	(0.03)	
Low White collar	4,616	11,129	15,133	7024	21,994	59,896
	(0.24)	(0.35)	(0.15)	(0.08)	(0.05)	
Skilled/semi-skilled	4,452	7,668	42,479	22,540	52,885	130,024
	(0.23)	(0.24)	(0.42)	(0.25)	(0.13)	
Unskilled	2844	3,903	18,100	36,436	106,767	168,050
	(0.15)	(0.12)	(0.18)	(0.40)	(0.25)	
Farm	3098	4,514	17,510	20,858	227,546	273,526
	(0.16)	(0.14)	(0.17)	(0.23)	(0.54)	
Total	19,538	31,532	101,325	90,451	422,920	665,766

Note: See main text for details on occupational coding. Column shares in parentheses. Weighted values rounded to nearest integer. Linkage weights similar to Bailey (2020) are described in the main text.

TABLE D12
ALTHAM STATISTICS FOR CANADA 1871-1901 AND COMPARISON COUNTRIES

	(1) $d(\boldsymbol{P},\boldsymbol{J})$	(2) G^2	(3) $d(\boldsymbol{Q},\boldsymbol{J})$	(4) G^2
CAN 1871-1901	*26.8*	*6,733****		
CAN 1871-1901, weighted	*27.0*	*5,434****		
US 1850-1880			*28.0*	– ***
UK 1851-1881			*32.6*	– ***
ARG 1869-1895			*23.3*	– ***
NOR 1865-1900			*44.7*	– ***
SWE 1880-1910			*31.3*	– –

Note: G^2 for weighted sample calculated using normalized inverse propensity score weights that sum to N. Values for $d(\boldsymbol{Q},\boldsymbol{J})$ taken from Perez (2019) and Berger et al. (2020), where values for G^2 are not provided. $d(\boldsymbol{P},\boldsymbol{Q})$ cannot be calculated without microdata or all of the 4-way odds ratios for all populations.

TABLE D13
TWO-WAY ODDS RATIOS OF RELATIVE REPRESENTATION OF SONS BY FATHER OCCUPATION

(a) Canada

Sons	Fathers: High White Collar	Low White Collar	Skilled/ Semi-skilled	Unskilled	Farm
High White Collar	6.5 (0.080)	3.3 (0.078)	1.6 (0.060)	0.7 (0.084)	0.4 (0.051)
Low White Collar	3.0 (0.079)	6.7 (0.057)	2.0 (0.045)	0.9 (0.060)	0.3 (0.040)
Skilled/Semi	1.2 (0.080)	1.3 (0.063)	3.9 (0.033)	1.4 (0.040)	0.3 (0.029)
Unskilled	0.5 (0.088)	0.4 (0.078)	0.6 (0.039)	2.4 (0.034)	1.0 (0.026)
Farm	0.3 (0.089)	0.2 (0.075)	0.3 (0.039)	0.4 (0.039)	4.8 (0.027)

D.2 Farm and Non-Farm Unskilled distinction

TABLE D.21
FULL CANADIAN TRANSITION MATRIX, 1871-1901

Son	Father: White Collar	Skilled/semi-skilled	Unskilled Nonfarm	Unskilled Farm	Farm	Total
a) Unweighted						
High White collar	1,215	1,142	383	116	1,810	4,666
	(0.49)	(0.23)	(0.13)	(0.09)	(0.09)	
Skilled/semi-skilled	545	2,049	807	173	2,548	61,22
	(0.22)	(0.41)	(0.28)	(0.14)	(0.12)	
Unskilled Nonfarm	146	422	522	113	1,396	2,619
	(0.06)	(0.08)	(0.18)	(0.10)	(0.07)	
Unskilled Farm	206	503	485	628	40,24	5,846
	(0.08)	(0.10)	(0.17)	(0.49)	(0.19)	
Farm	367	907	698	229	10,911	13,112
	(0.15)	(0.18)	(0.24)	(0.18)	(0.53)	
Total	2,479	5,023	2,895	1,279	20,689	32,365
b) Weighted						
High White collar	24,592	23,236	8,356	2,262	35,722	94,168
	(0.48)	(0.23)	(0.13)	(0.09)	(0.08)	
Skilled/semi-skilled	12,120	42,479	19,304	3,237	52,885	130,025
	(0.24)	(0.42)	(0.29)	(0.13)	(0.13)	
Unskilled Nonfarm	3,119	9,045	12,686	2,585	29,902	57,337
	(0.06)	(0.09)	(0.19)	(0.11)	(0.07)	
Unskilled Farm	3,623	9,055	9,859	11,305	76,866	110,708
	(0.07)	(0.09)	(0.15)	(0.47)	(0.18)	
Farm	7,613	17,510	16,015	4,842	227,546	273,526
	(0.15)	(0.17)	(0.24)	(0.20)	(0.54)	
Total	51,067	101,325	66,220	24,231	422,920	665,764

Notes: Column shares in parentheses. Weighted values rounded to nearest integer. Linkage weights similar to Bailey (2020) are described in the main text.

TABLE D22
ALTHAM STATISTICS FOR CANADA 1871-1901 AND COMPARISON COUNTRIES

	(1) $d(\boldsymbol{P},\boldsymbol{J})$	(2) G^2	(3) $d(\boldsymbol{Q},\boldsymbol{J})$	(4) G^2
CAN 1871-1901	*25.7*	*7,320****		
CAN 1871-1901, weighted	*26.3*	*6,024****		
US 1850-1880			*30.5*	– ***
UK 1851-1881			*32.4*	– ***
ARG 1869-1895			*22.3*	– ***
NOR 1865-1900			*37.1*	– ***
SWE 1880-1910			*26.4*	– –

Note: G^2 for weighted sample calculated using normalized inverse propensity score weights that sum to N. Values for $d(\boldsymbol{Q},\boldsymbol{J})$ taken from Perez (2019) and Berger et al. (2020), where values for G^2 are not provided. $d(\boldsymbol{P},\boldsymbol{Q})$ cannot be calculated without microdata or all of the 4-way odds ratios for all populations.

TABLE D23
TWO-WAY ODDS RATIOS OF RELATIVE REPRESENTATION OF SONS BY FATHER OCCUPATION

(a) Canada

Sons	Fathers White Collar	Skilled/ Semiskilled	Unskilled Nonfarm	Unskilled Farm	Farm
White Collar	7.4 (0.044)	2.0 (0.038)	0.9 (0.057)	0.6 (0.099)	0.3 (0.033)
Skilled/Semi	1.2 (0.050)	3.9 (0.033)	1.8 (0.044)	0.7 (0.083)	0.3 (0.029)
Unskilled Nonfarm	0.7 (0.087)	1.0 (0.056)	2.9 (0.053)	1.3 (0.094)	0.6 (0.041)
Unskilled Farm	0.4 (0.073)	0.5 (0.049)	0.9 (0.052)	4.8 (0.058)	1.3 (0.031)
Farm	0.2 (0.058)	0.3 (0.039)	0.4 (0.045)	0.3 (0.074)	4.8 (0.027)

Note: See main text for calculation details.