APPENDIX: A NOTE ON SELECTION BIAS

In economic and historical research involving topics such as migration, social mobility, and savings behavior, selection bias is an abiding concern.[1] Were the Irish immigrants who opened accounts in the Emigrant Savings Bank in the 1850s typical of Irish immigrants generally in that period? Were those who could not be traced or tracked down in the census or elsewhere differ from those who could? Did those account holders who migrated within the U.S. differ from those who stayed put? The answer to all these questions is 'yes', but it is also 'yes, but not to the extent that the inferences drawn in the paper are invalid'.

First, the occupational profile of male account holders did not differ much from that of Irish males generally in the 1855 census of New York (Table A1). As might be expected, unskilled laborers were underrepresented and professionals and white-collar workers over-represented but, perhaps surprisingly, the differences were modest. Second, Table A2 shows that we showed that while unskilled workers were proportionately fewer among those we managed to link across censuses than those we could not link, again the differences were too small to be a cause for worry. Moreover, the average values of first deposits—$115 for tracked, $117 for untracked—were very similar.

Third, Table A3 and A4 compare some characteristics of tracked and non-tracked accounts. Table A3 focuses on accounts open until 1860, while Table A4 includes all those matched either in 1860-70 or 1870-80. In Table A3 the average savings of untracked unskilled workers were lower than those of tracked, but the opposite was the case for skilled workers. Similarly, untracked unskilled workers were less likely to keep their accounts open, while untracked skilled workers were more so. Table A4 focuses instead on median values of certain characteristics. The message is broadly the same: yes, there is some selection bias across most dimensions; and, therefore, inferences drawing on the

[1] Tyler Anbinder, Dylan Connor, Cormac Ó Gráda, and Simone Wegge', The Problem of False Positives in Automated Census Linking: Evidence from Nineteenth-Century New York's Irish Immigrants', School of Economics, University College Dublin, Working Paper 21/14.

bank account holders tell a story of social progress that is somewhat rosier than the truth. That is to be expected, but the bias seems to be modest in size.

TABLE A1
IRISH-BORN MALES BY OCCUPATIONAL CATEGORY: ESB'S NEW YORK CITY CUSTOMERS (1850-1858) AND ALL NEW YORKERS IN 1855

	ESB		**All New York Irish**	
Occupational Category	*N*	*%*	*N*	*%*
Professionals	20	0.4	16	0.3
Business Owners	397	8.0	310	5.8
Lower-Status White-Collar	370	7.4	260	4.9
Skilled Workers	1,738	34.8	2,022	37.9
Petty Entrepreneurs	255	5.1	90	1.7
Unskilled Workers	2,117	42.4	2,567	48.1
Others	91	1.8	73	1.4
Total (N)	4,988	100	5,338	100

Sources: Emigrant Savings Bank Depositor Database; 1855 New York State Census [10 percent sample of employed Irish-born adults]. Both are available at https://dataverse.harvard.edu/dataverse/anbinder. Note that the 1855 NY Census columns do not include one of the city's twenty-two wards because the returns for Ward Seventeen are not extant. The occupational categories are explained in the text.

TABLE A2
IRISH-BORN MALES BY OCCUPATIONAL CATEGORY:
THE TRACKED AND THE NON-TRACKED

	Tracked 10+ Years in U.S.		**Tracked Less Than 10 Years in U.S.**	
Occupational Category	*N*	*%*	*N*	*%*
Professionals	10	0.8	0	0
Business Owners	103	9	78	7
Lower-Status White-Collar	102	9	96	8
Skilled Workers	444	38	397	33
Peddlers	63	5	74	6
Unskilled Workers	449	38	496	42
Others	9	0.8	45	4
Total	1,180	100	1,186	100

Sources: Emigrant Savings Bank database.

TABLE A3
SOME CHARACTERISTICS OF TRACKED AND NON-TRACKED ACCOUNTS OPEN TILL 1860

Category	Tracked	Untracked
Average Savings, Unskilled Workers	$509	$611
% Accounts Still Open, Unskilled Workers	38	54
Average Savings, Skilled Workers	$617	$544
% Accounts Still Open, Skilled Workers	42	33

TABLE A4
SOME CHARACTERISTIS OF LINKED AND NON-LINKED DEPOSITORS: MEDIAN VALUES

Characteristic	*Linked*	*Not linked*
Age in 1860	35	38
Peak savings [$]	411	171
Peak savings [$], males only	412	200
Peak savings [$], arrived pre-1846	524	320
Peak savings [$], arrived post-1845	331	150
Peak savings if has other account	498	302
Opening deposit [$]	100	60
Opening deposit [$], males only	100	70
Opening deposit [$], pre-1846	120	100
Opening deposit [$], post-1845	80	50
Years to highest deposit	2.83	1.23
Year of arrival	1849	1850
Joint account? (%)	31.4	22.4
Other account? (%)	58.0	31.7
Median duration of account (years)	5.33	2.56
Number of transactions	15	9
Female (%)	30.2	42.5
Number of accounts	c. 1,260	c. 9,830
Note: “Linked” includes all those matched in either 1860-70 or 1870-80		