Online Appendix for

US Immigrants' Secondary Migration and Geographic Assimilation during the Age of Mass Migration

Ariell Zimran
Vanderbilt University and NBER

February 23, 2022

C Additional Tables and Figures

Table C.1: Linkage rates for the foreign born

	Immigrants		Natives	
	(1)	(2)	(3)	(4)
Link	Start	Linked	Start	Linked
1850–1860	722,412	47,329	2,980,784	384,687
		(0.066)		(0.129)
1850–1870	722,412	45,080	2,980,784	347,741
		(0.062)		(0.117)
1860–1870	1,260,619	71,745	3,840,785	383,199
		(0.057)		(0.100)
1860–1880	1,260,619	69,337	3,840,785	374,981
		(0.055)		(0.098)
1870–1880	1,446,944	102,422	4,488,154	575,268
		(0.071)		(0.128)
1880–1900	1,403,439	118,378	6,892,494	777,946
		(0.084)		(0.113)
1900–1910	2,603,941	228,521	12,669,268	1,546,951
		(0.088)		(0.122)
1900–1920	2,171,777	220,391	11,303,818	1,450,252
		(0.101)		(0.128)
1910–1920	3,523,597	309,799	15,268,612	2,201,059
		(0.088)		(0.144)
1910–1930	3,096,356	271,036	14,175,313	2,065,500
		(0.088)		(0.146)

Numbers in parentheses are the fraction of individuals who were successfully linked.

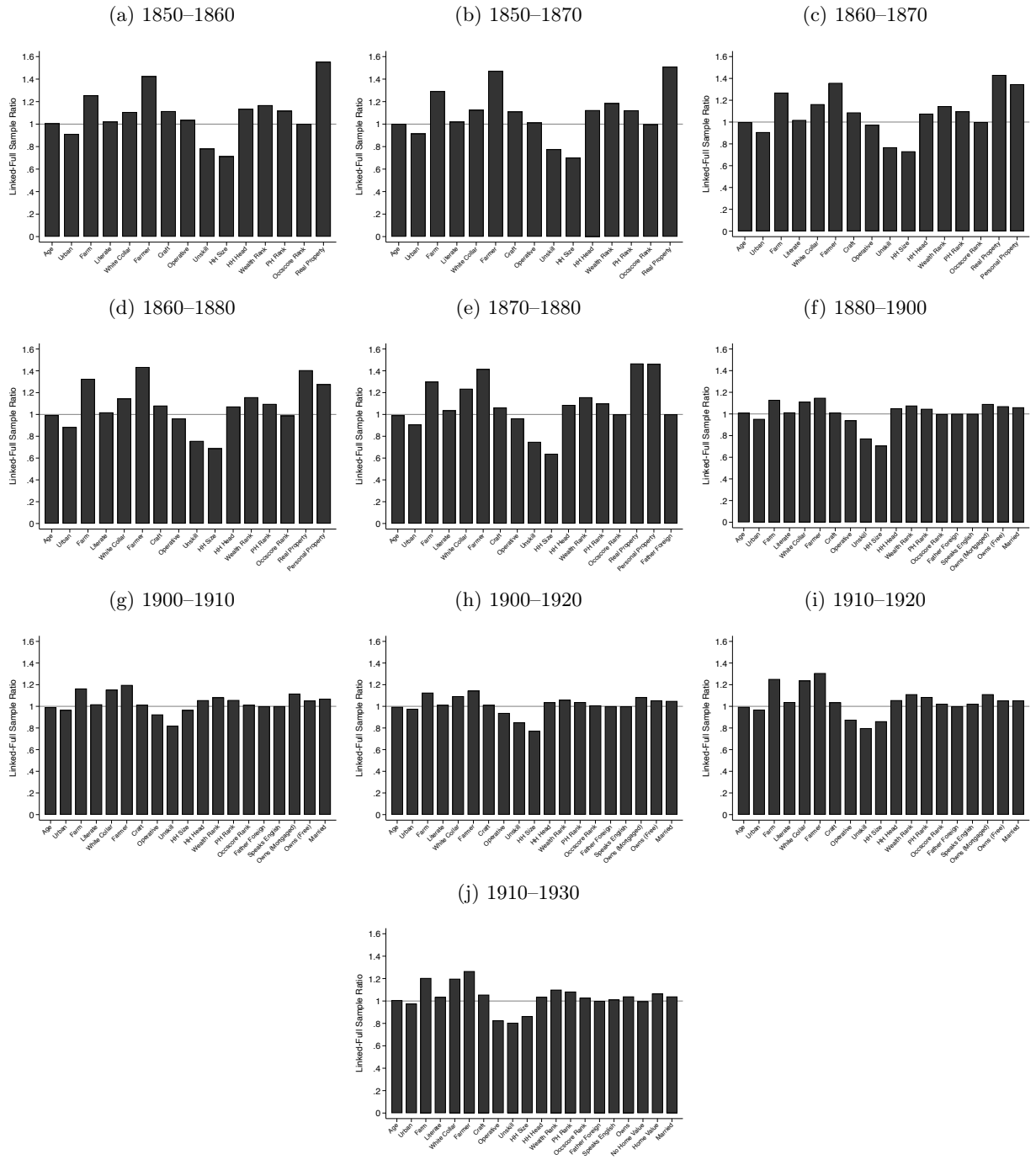


Figure C.1: Representativeness of the linked sample, foreign-born individuals

Note: Each bar presents the ratio of the mean of each variable in the linked sample relative to the full sample at risk for linkage. For panels (a)–(e), the sample at risk is from the initial census year. For panels (f)–(j), the sample at risk is from the final census year, limiting to individuals whose ages and years of arrival put them at risk for linkage.

Table C.2: Observables available in each initial census

Variable	1850	1860	1870	1880	1900	1910
Age	X	X	X	X	X	X
Urban	X	X	X	X	X	X
City Population[a]	X	X	X	X	X	X
Farm	X	X	X	X	X	X
Literacy	X	X	X	X	X	X
Marital status				X	X	X
Occupation[b]	X[c]	X	X	X	X	X
Household size	X	X	X	X	X	X
Household head	X	X	X	X	X	X
Birthplace	X	X	X	X	X	X
Place of residence	X	X	X	X	X	X
Real property	X	X	X			
Personal property		X	X			
Father's nativity[d]				X	X	X
Years in US[e]					X	X
Speaks English[e]					X	X
Owns home (mortgaged)					X	X
Owns home (free and clear)					X	X

[a]: used only for alternate definitions of urban residence in Online Appendix G
[b]: used to create occupational categories, the three occupational rank measures, and the average occupational rank measure, which are available for every census year
[c]: farmers are probabilistically recategorized following Collins and Zimran (2021)
[d]: native born only
[e]: foreign born only

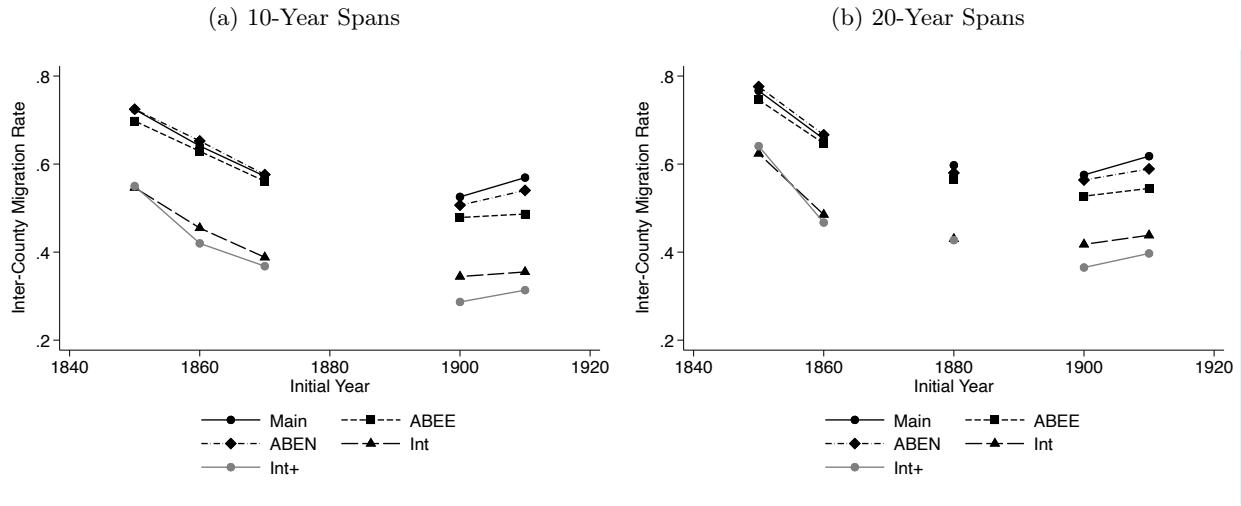


Figure C.2: Uncorrected inter-county migration rates for the foreign born

Note: Each figure shows the probability that a foreign-born individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, according to each linkage method. All observations are weighted by inverse linkage probability.

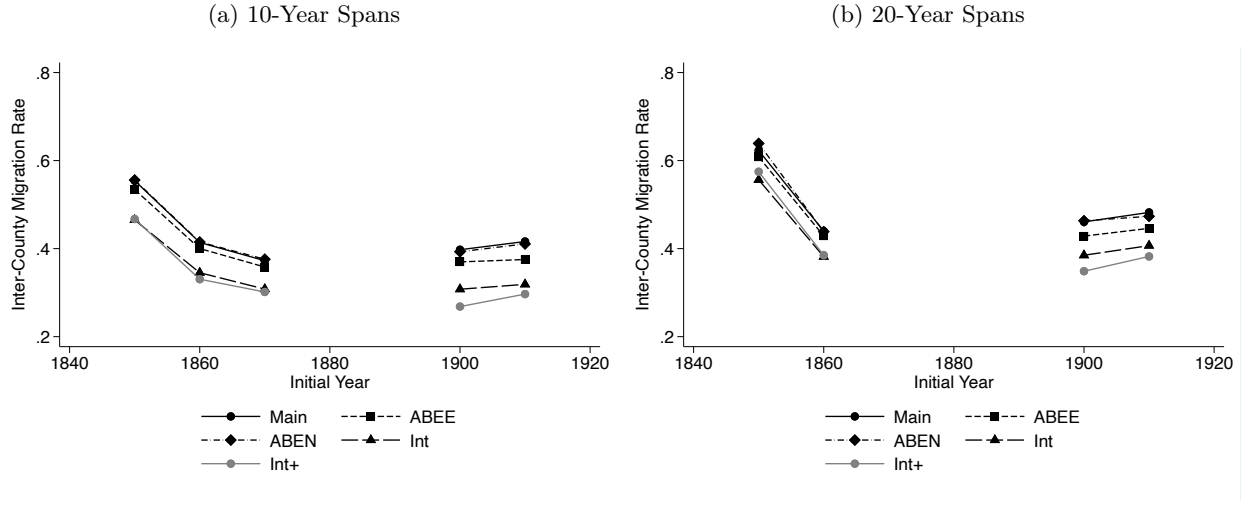


Figure C.3: Corrected inter-county migration rates for the foreign born

Note: Each figure shows the probability that a foreign-born individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, according to each linkage method, corrected for false matches according to the method described in Appendix A. All observations are weighted by inverse linkage probability.

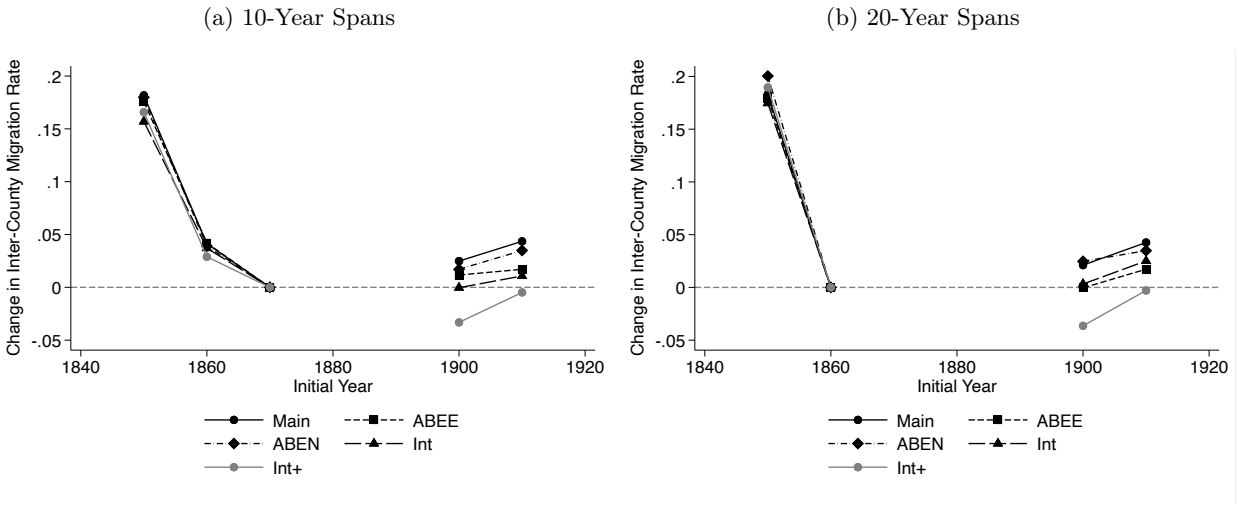


Figure C.4: Changes in corrected inter-county migration rates for the foreign born

Note: Each figure shows the change in the estimated inter-county migration rate corrected for false matches (Figure C.3 relative to that of the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

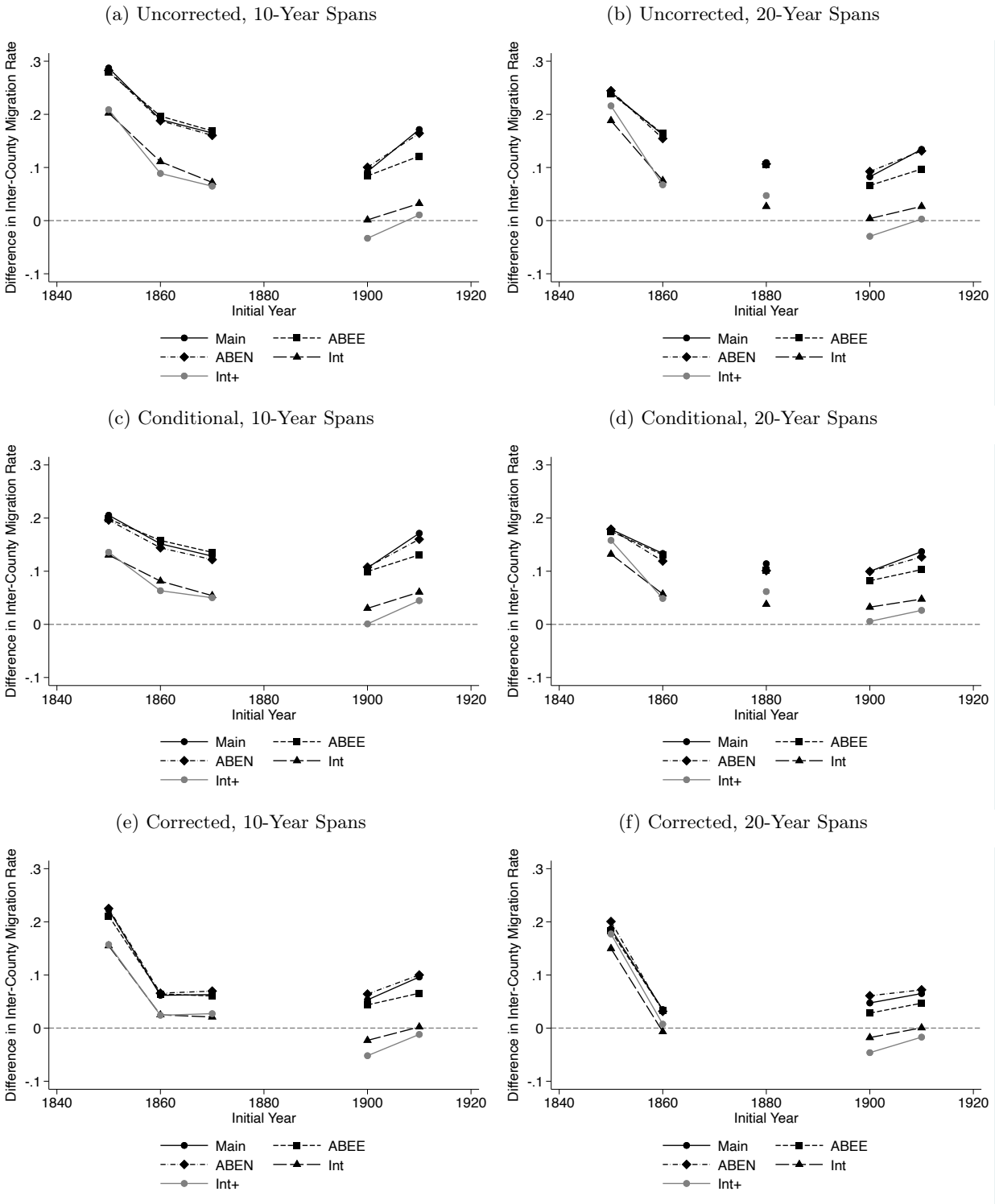


Figure C.5: Differences in inter-county migration rates

Note: These figures present estimated differences in inter-county migration rates by nativity. Panels (a) and (b) present unconditional differences. Panels (c) and (d) are unconditional differences after correcting for false matches as described in text and in Appendix A. Panels (e) and (f) are from regressions including all controls available for both immigrants and natives in the census of the initial year of the span. All observations are weighted by inverse linkage probability.

D Defining Inter-County Migration

In principle, it is straightforward to determine whether an individual made an inter-county move between two censuses—every census reports an individual's county of residence, and an individual is a migrant if he does not live in the same county in each census. But changing county boundaries across censuses complicates this definition. The example in Figure D.1 illustrates how I address this issue. This figure depicts several counties in northeastern Illinois in 1850 and in 1870 that experienced boundary changes in this span. In 1850, points A and B on the one hand and C and D on the other are each in the same county and cannot be distinguished from one another (because only the county of residence is known). Similarly, in 1870, points B and C are in the same county and cannot be distinguished from one another. A move from point A to C, which crosses county boundaries as defined in either year, thus cannot be considered an example of internal migration because it cannot be differentiated from a move from point A to B, from B to C, or from continuous residence at point B. This issue arises whenever there is any overlap between an individual's residence county in two census years. Thus, I only consider as inter-county migration moves between counties with no geographic overlap—in the example of Figure D.1, moves from A or B to D or from C or D to A.[64]

Figure D.1: Definition of inter-county migration

Note: These figures show a set of counties in northeastern Illinois that experienced changing boundaries between 1850 and 1870. In 1850, it is not possible to distinguish between points A and B or between points C and D. In 1870, it is not possible to distinguish between points B and C. As a result, only moves from A or B to D or from C or D to A will be considered an inter-county move.

[64]Due to slight variations in the drawing of county boundaries by Manson et al. (2019) in different years, a more nuanced definition is required to avoid seeing an overlap in two non-overlapping counties simply because of a drawing imprecision. I consider an overlap to occur only if the overlapping area is at least 0.5 percent of at least one county's area.

E Results with Alternative Linkage Methods

This appendix repeats the main results with alternate linkage methods. Appendices E.1 and E.2 use the ABE-Exact and ABE-NYSIIS conservative methods from Abramitzky et al. (2020), respectively. Appendix E.3 uses the intersection-of-matches method—matches that are made by both of the ABE methods and the one used in the main text. Appendix E.4 uses the intersection-of-matches plus corroboration method, which uses the subset of the intersection-of-matches links that can be confirmed by information not used in the linking, such as the parents' birthplaces.

E.1 ABE-Exact

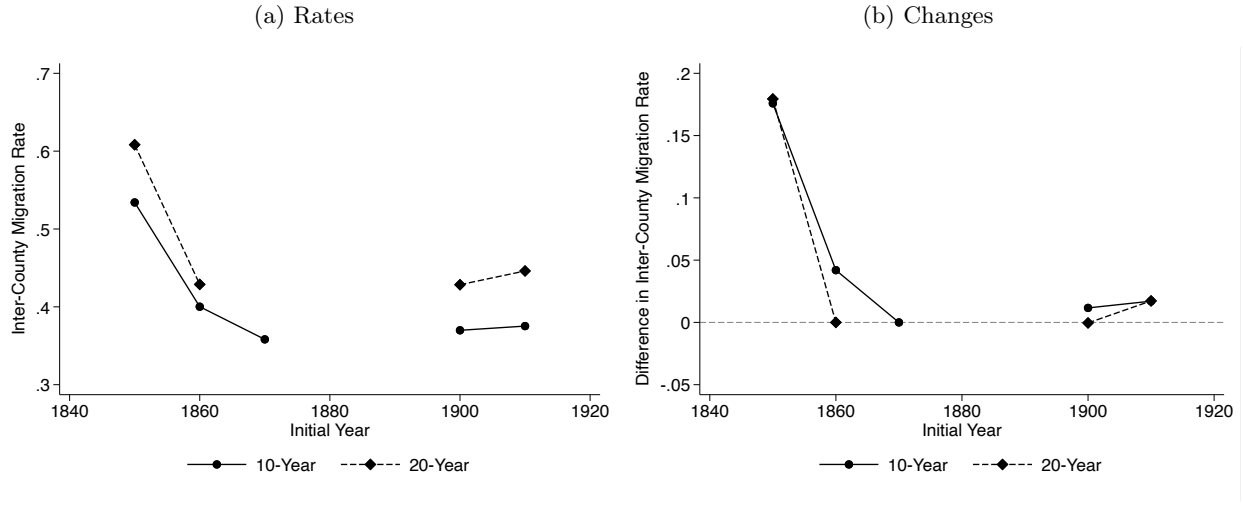


Figure E.1.1: Inter-county migration rates and changes, corrected for false matches

Note: Panel (a) shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

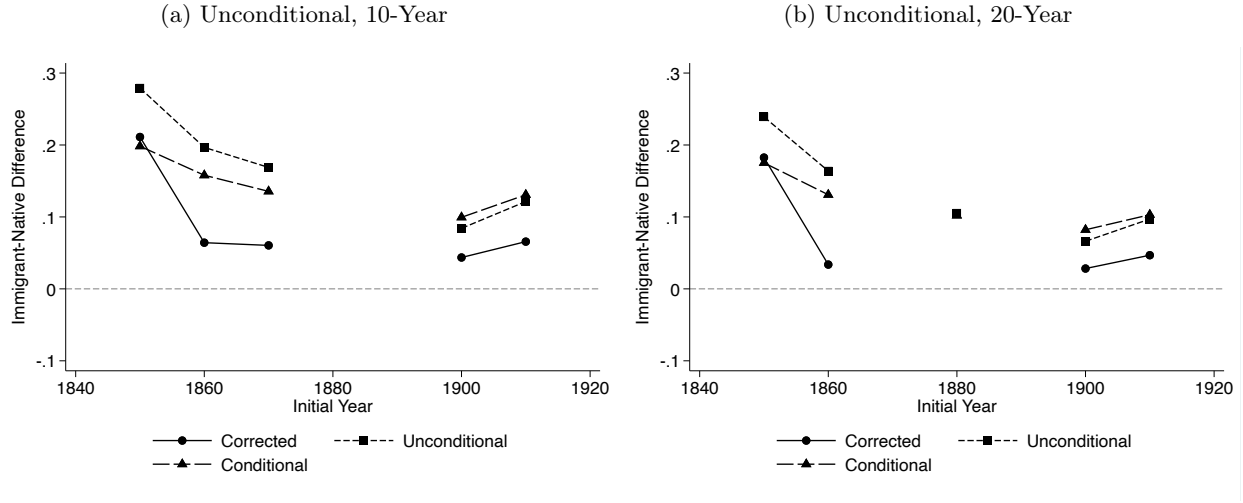
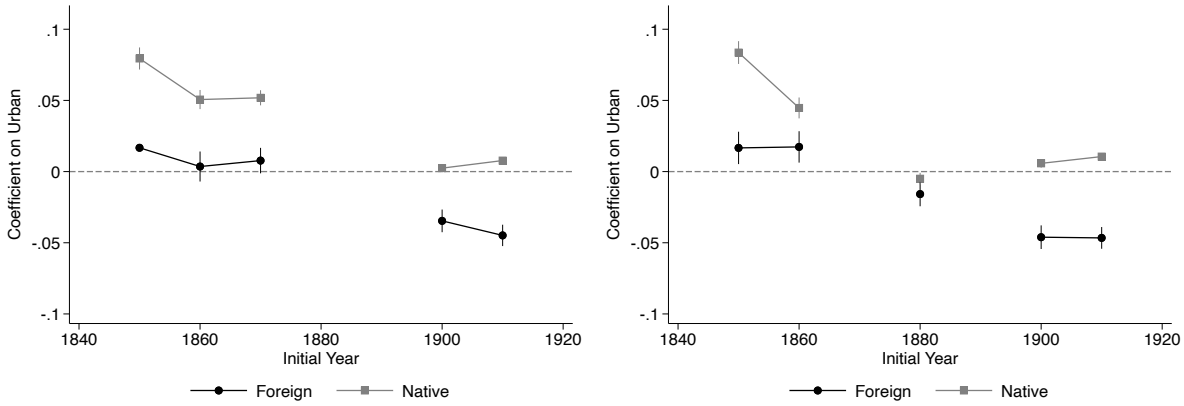


Figure E.1.2: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

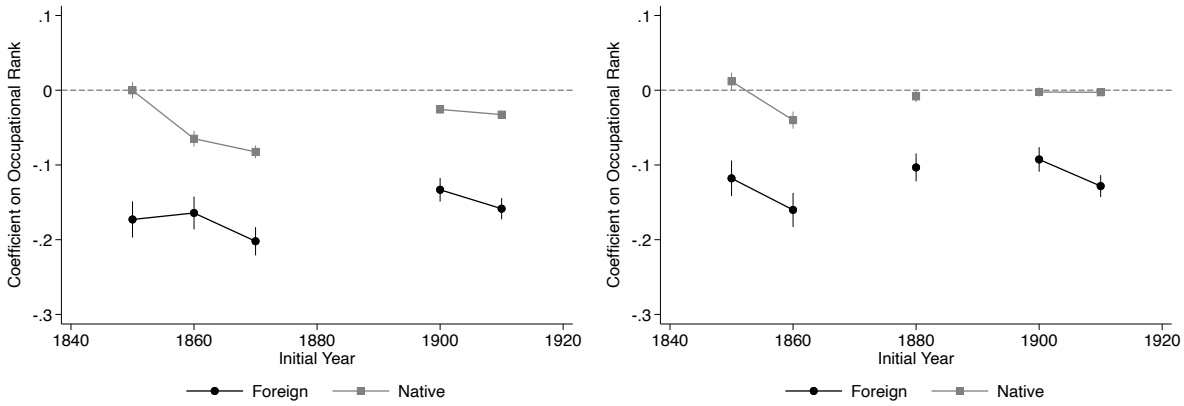
(a) Urban, 10-Year

(b) Urban, 20-Year



(c) Occupational Rank, 10-Year

(d) Occupational Rank, 20-Year



(e) Literacy, 10-Year

(f) Literacy, 20-Year

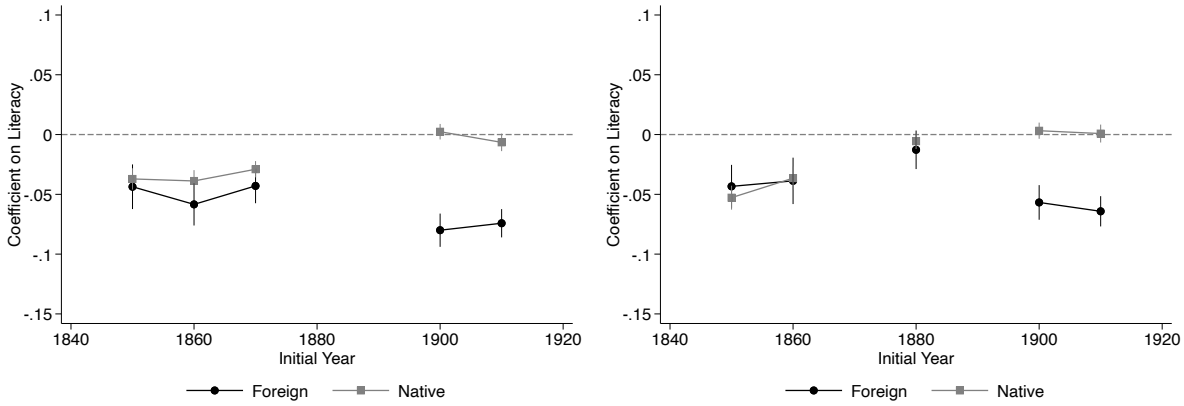


Figure E.1.3: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

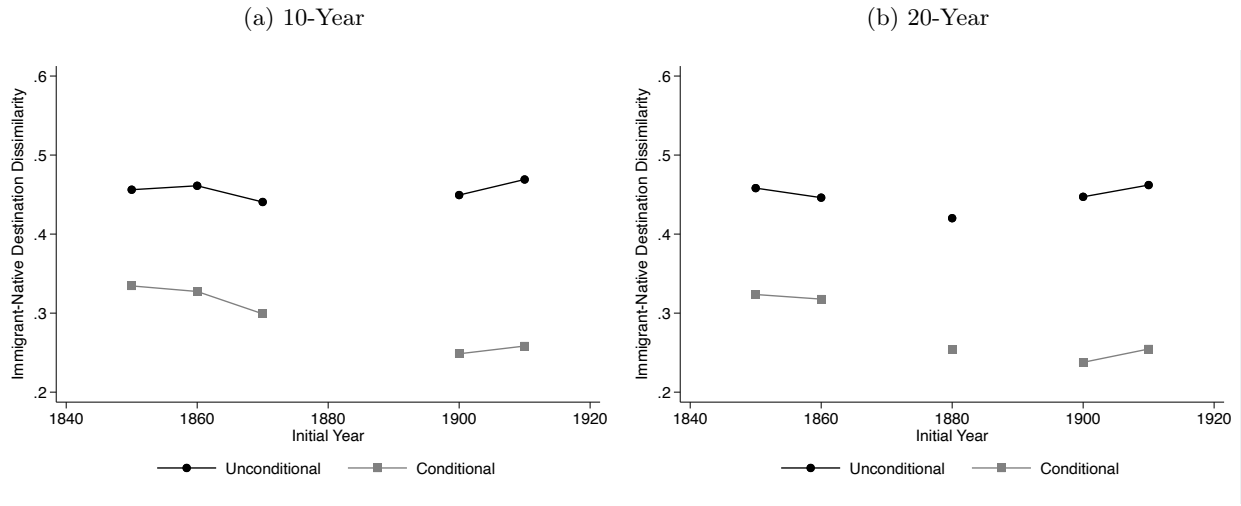
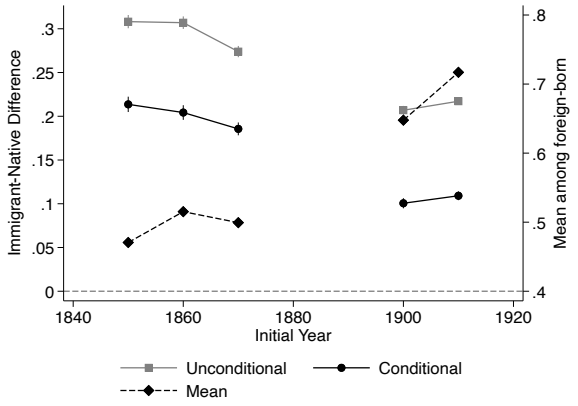


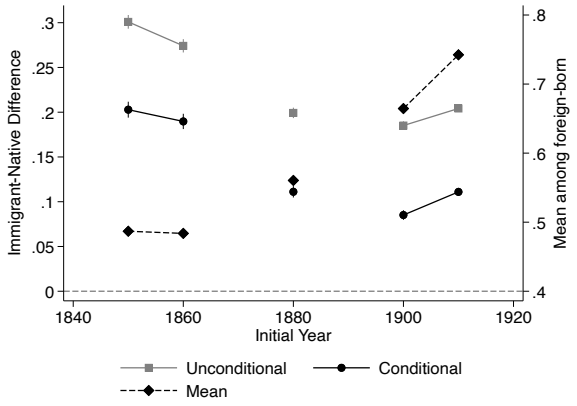
Figure E.1.4: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 45.

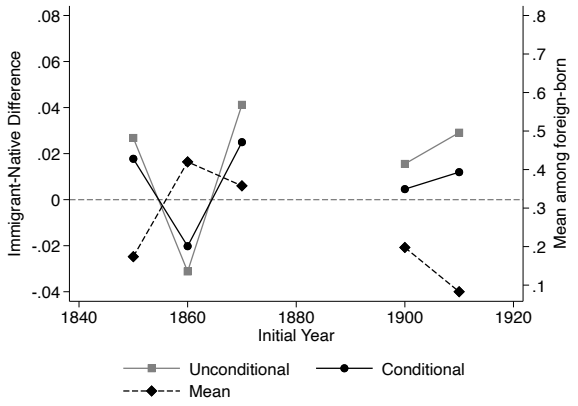
(a) Urban, 10-Year



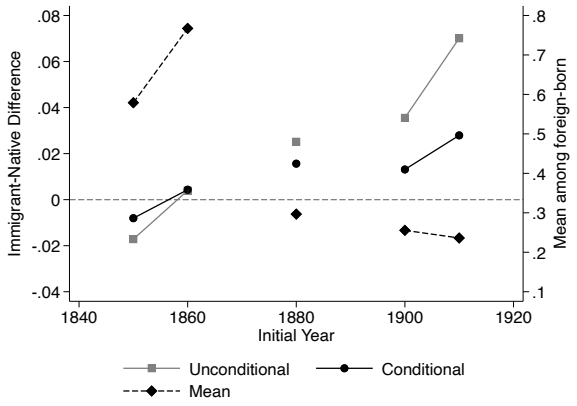
(b) Urban, 20-Year



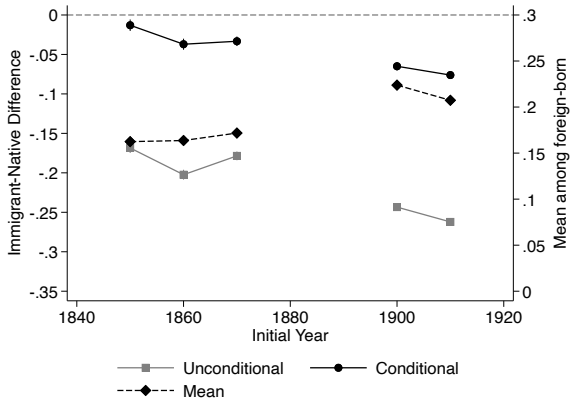
(c) Labor Demand Growth, 10-Year



(d) Labor Demand Growth, 20-Year



(e) South or West, 10-Year



(f) South or West, 20-Year

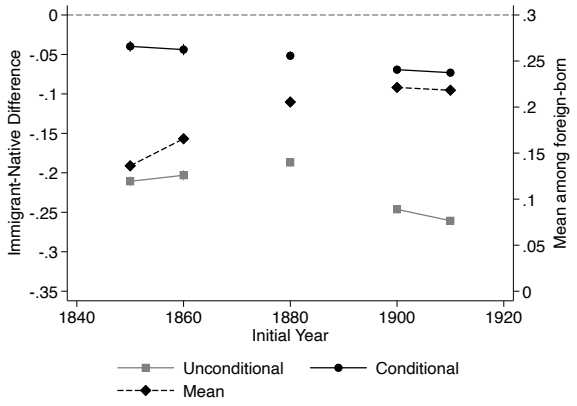


Figure E.1.5: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

(a) 10-Year

(b) 20-Year

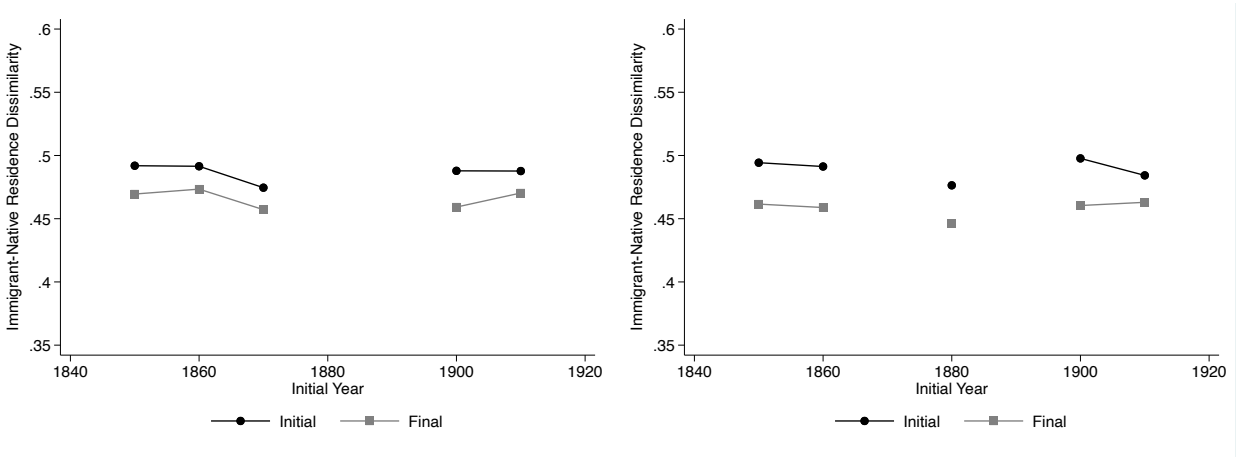


Figure E.1.6: Dissimilarity in counties of residence

Note: These figures present the dissimilarity index between the counties of residence of the foreign and native born in my linked samples in the initial and final year of the span beginning in the year on the x-axis.

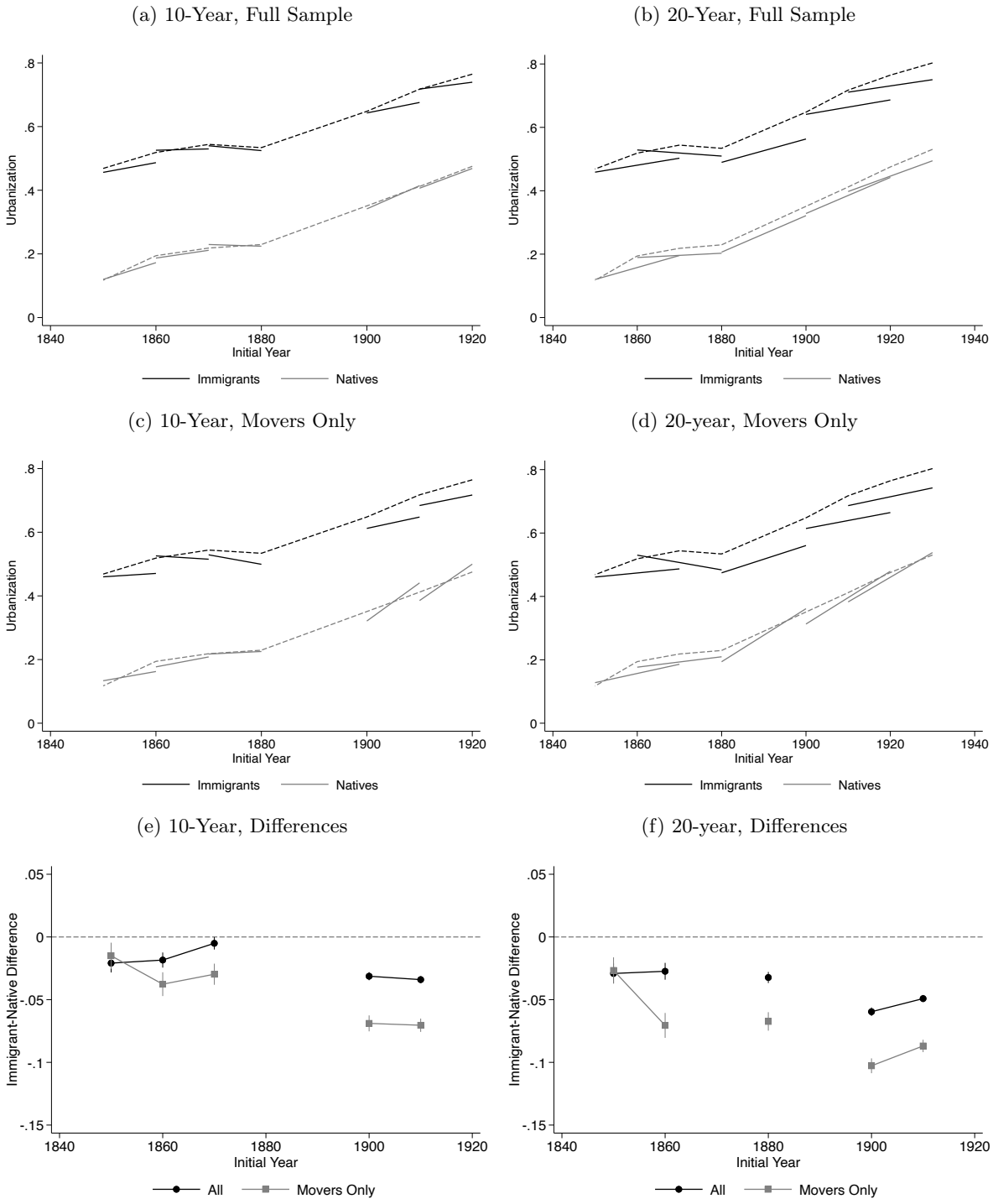
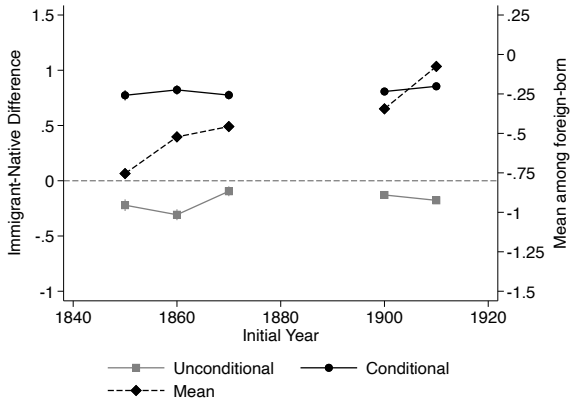


Figure E.1.7: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

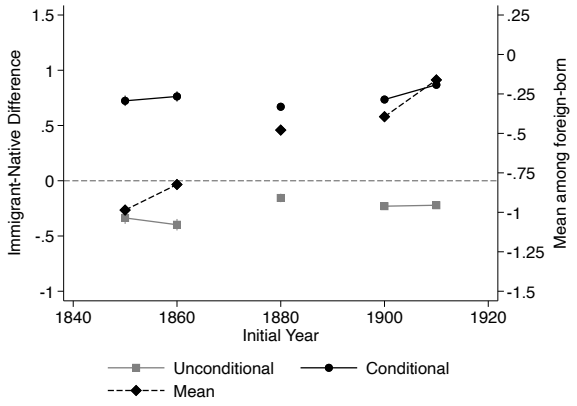


Figure E.1.8: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

E.2 ABE-NYSIIS

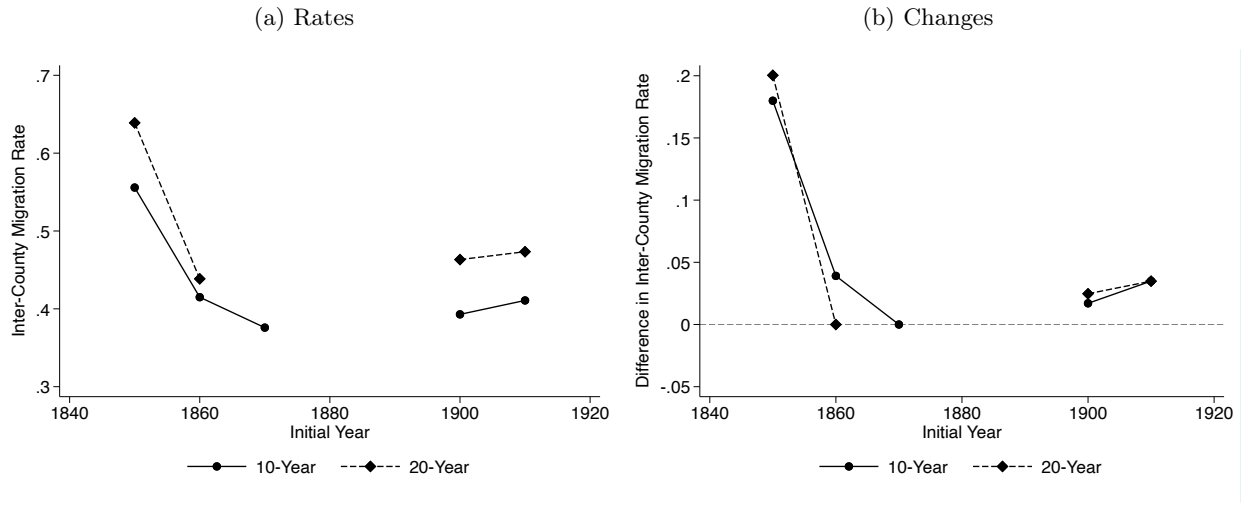


Figure E.2.1: Inter-county migration rates and changes, corrected for false matches

Note: Panel (a) shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

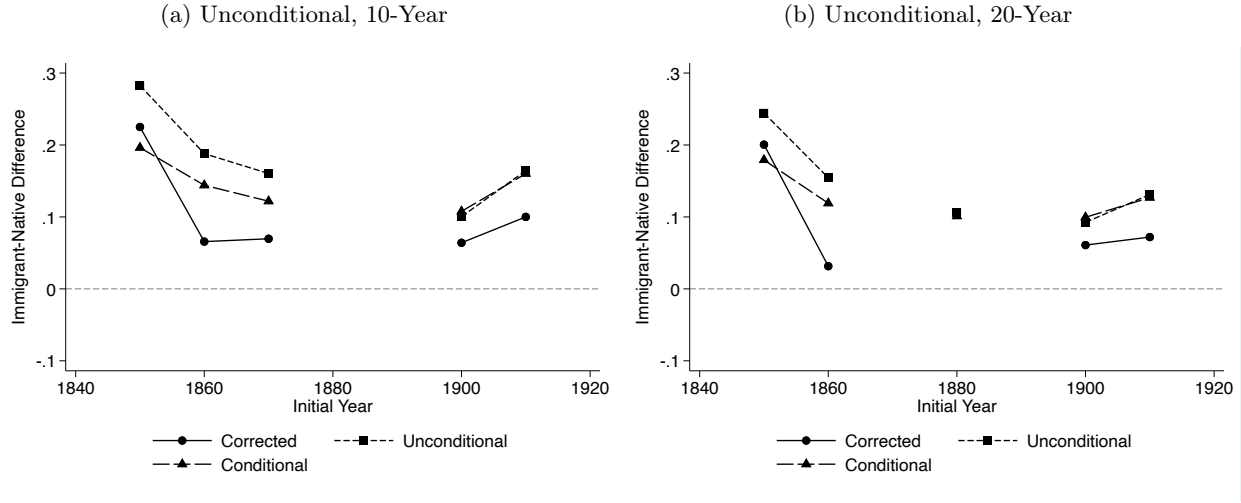
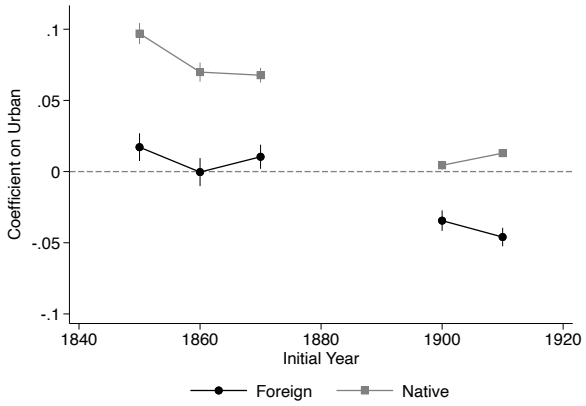


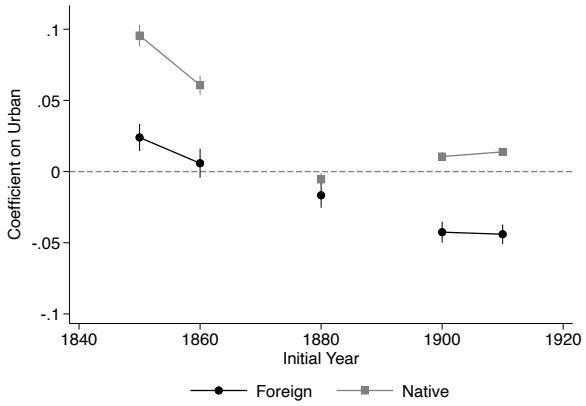
Figure E.2.2: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

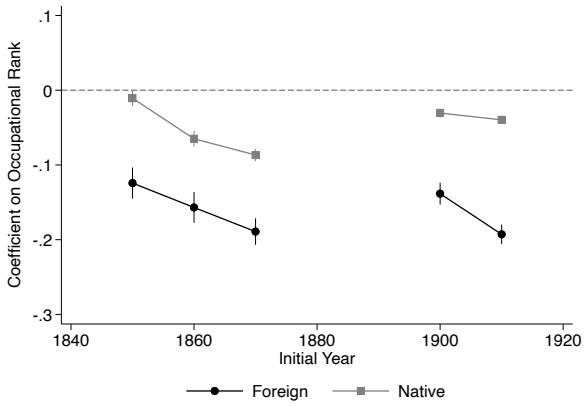
(a) Urban, 10-Year



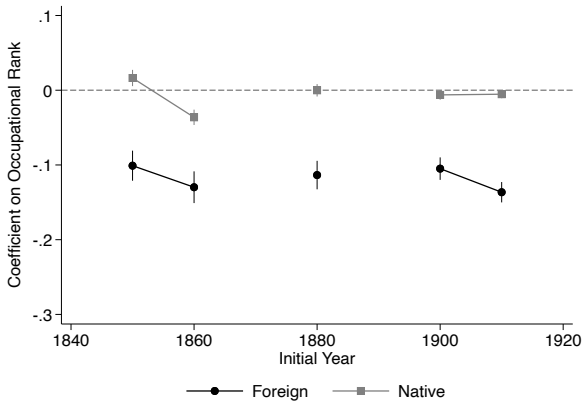
(b) Urban, 20-Year



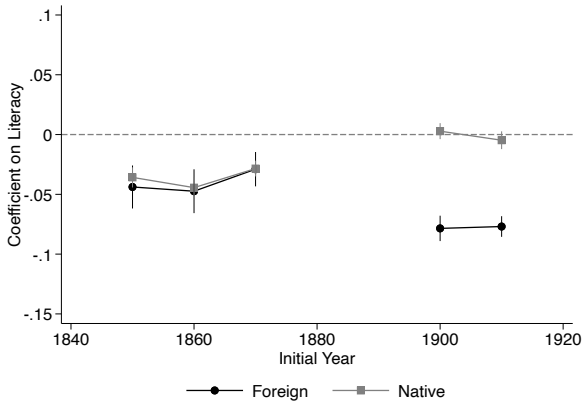
(c) Occupational Rank, 10-Year



(d) Occupational Rank, 20-Year



(e) Literacy, 10-Year



(f) Literacy, 20-Year

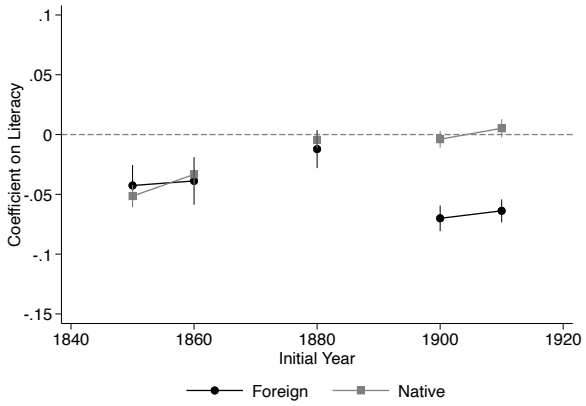


Figure E.2.3: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

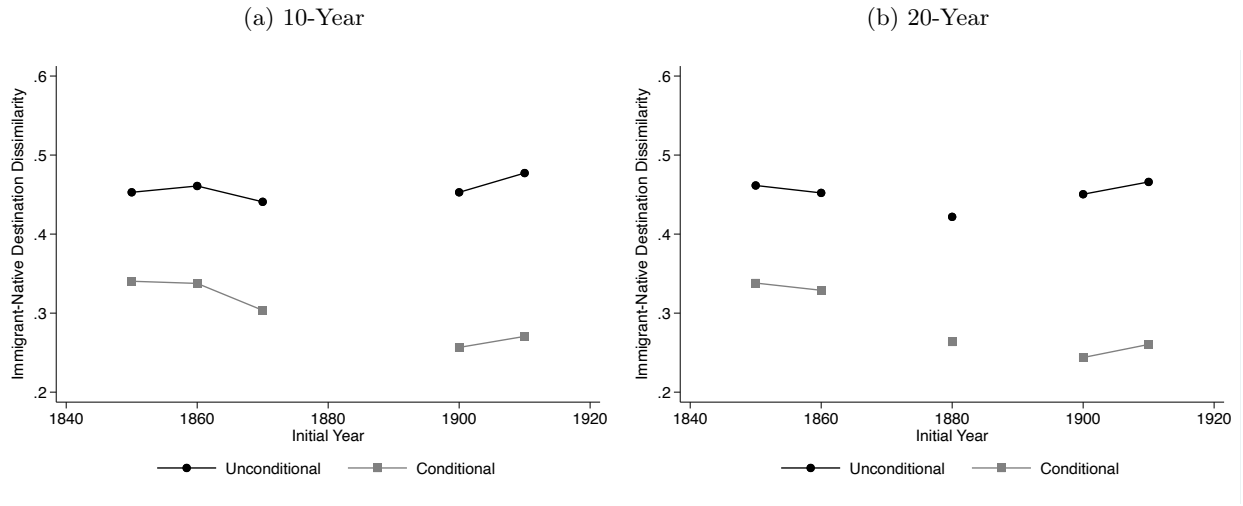
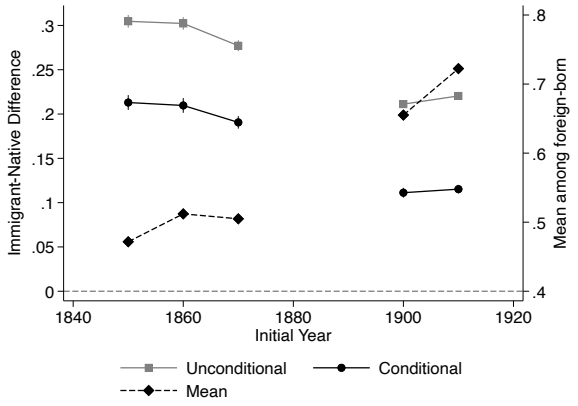


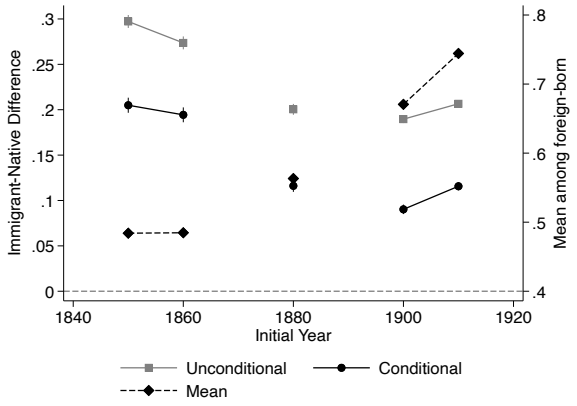
Figure E.2.4: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 45.

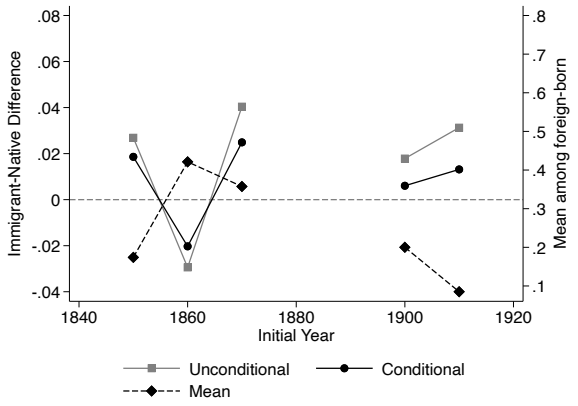
(a) Urban, 10-Year



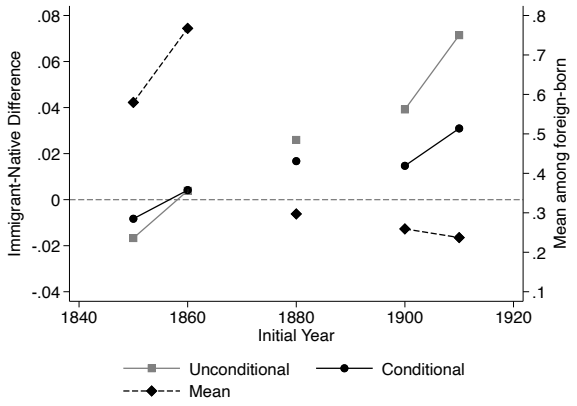
(b) Urban, 20-Year



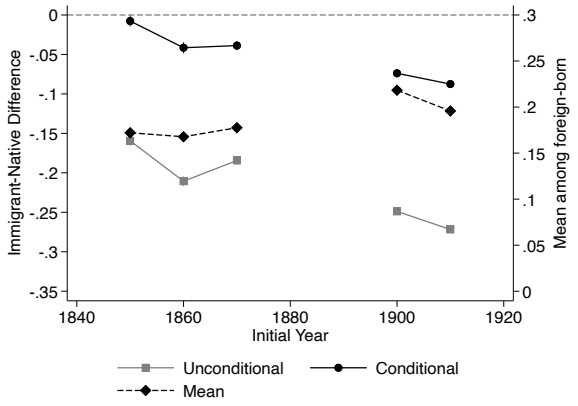
(c) Labor Demand Growth, 10-Year



(d) Labor Demand Growth, 20-Year



(e) South or West, 10-Year



(f) South or West, 20-Year

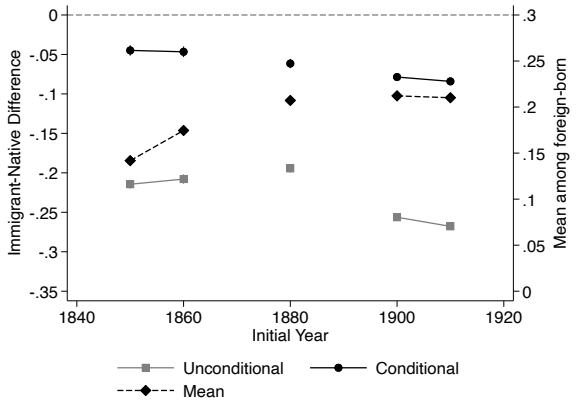


Figure E.2.5: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

(a) 10-Year

(b) 20-Year

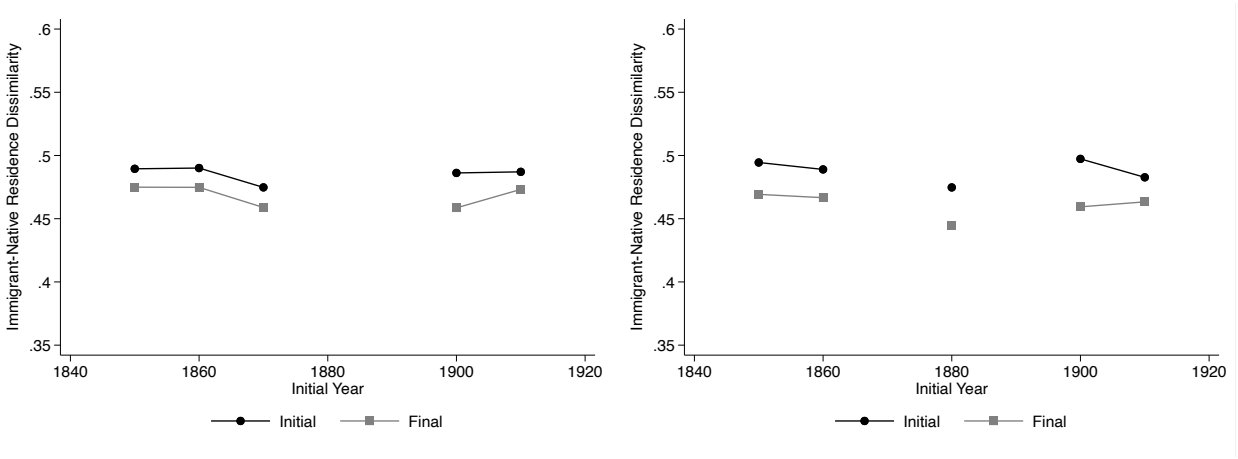


Figure E.2.6: Dissimilarity in counties of residence

Note: These figures present the dissimilarity index between the counties of residence of the foreign and native born in my linked samples in the initial and final year of the span beginning in the year on the x-axis.

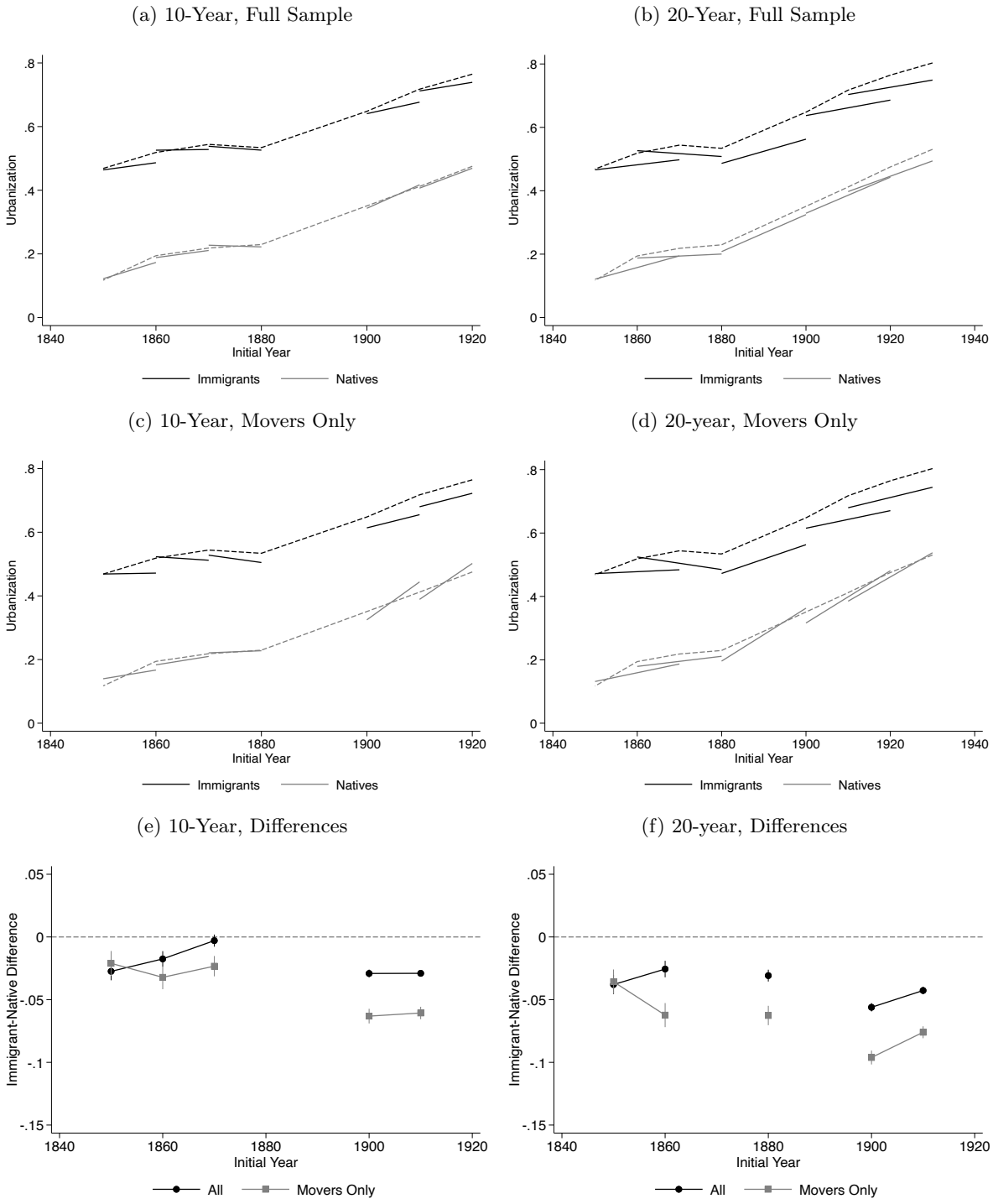
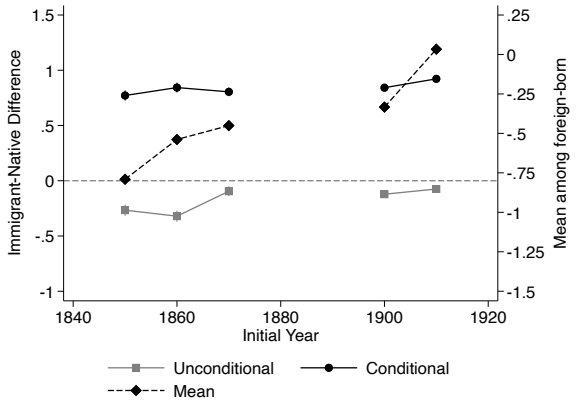


Figure E.2.7: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

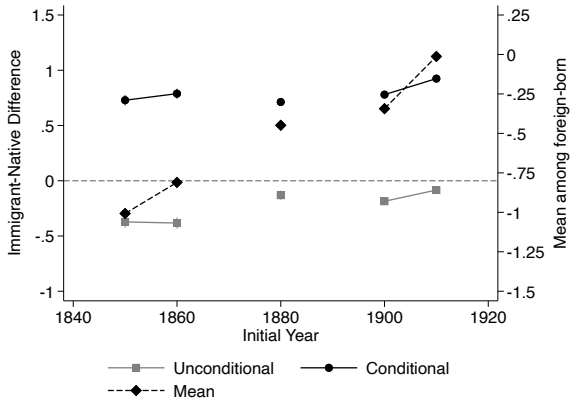


Figure E.2.8: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

E.3 Intersection of Matches

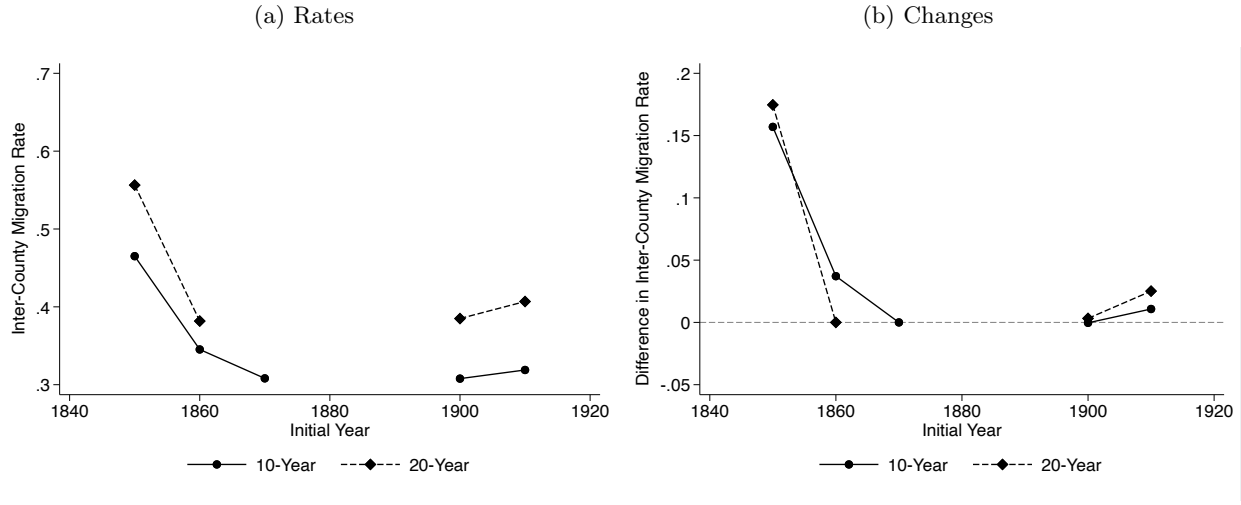


Figure E.3.1: Inter-county migration rates and changes, corrected for false matches

Note: Panel (a) shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

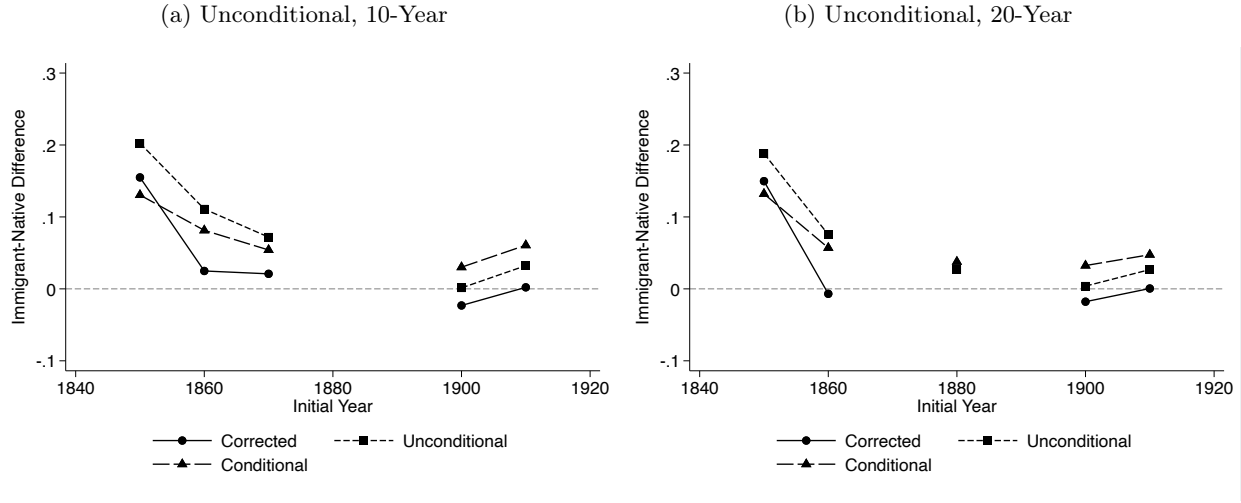


Figure E.3.2: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

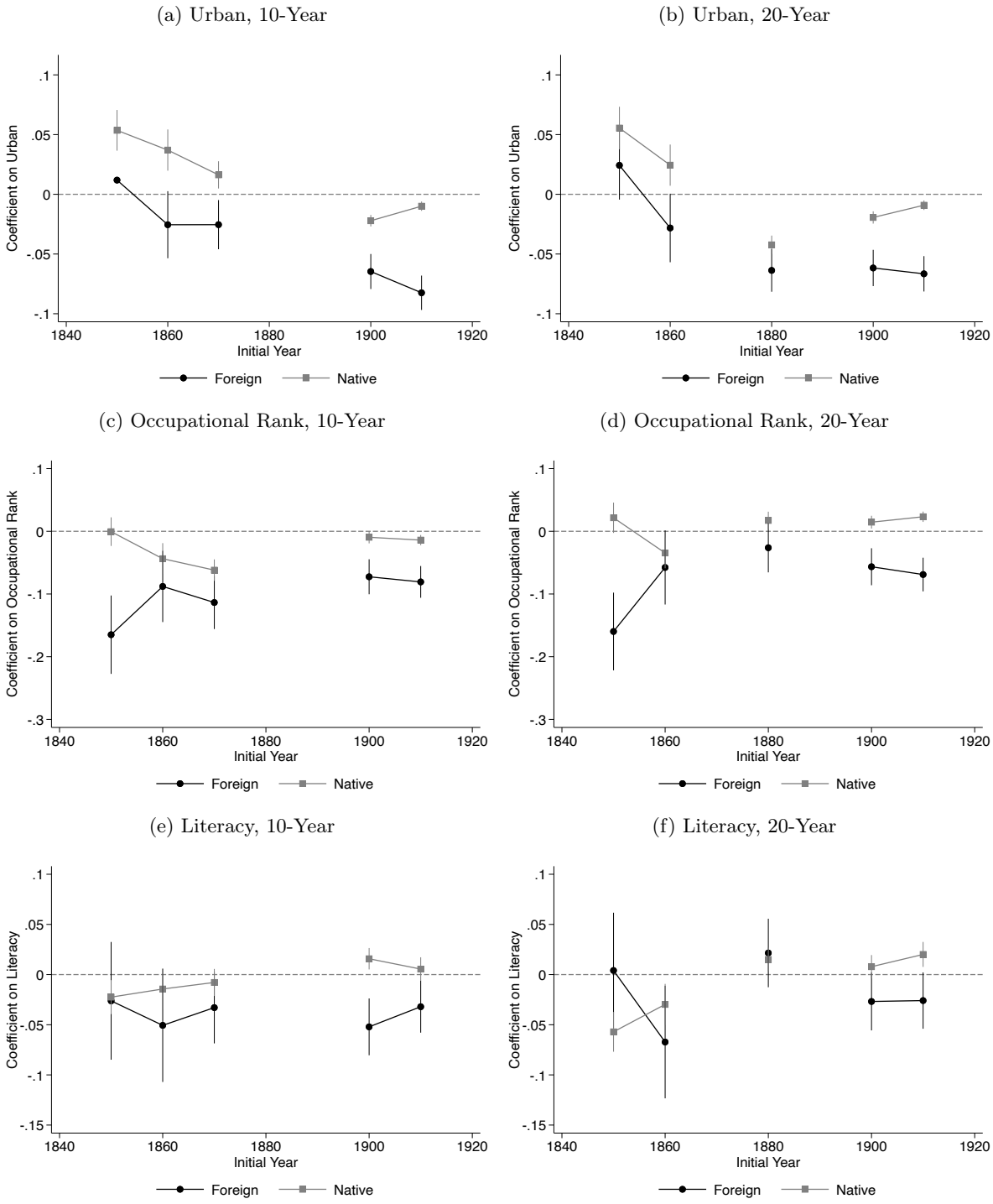


Figure E.3.3: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

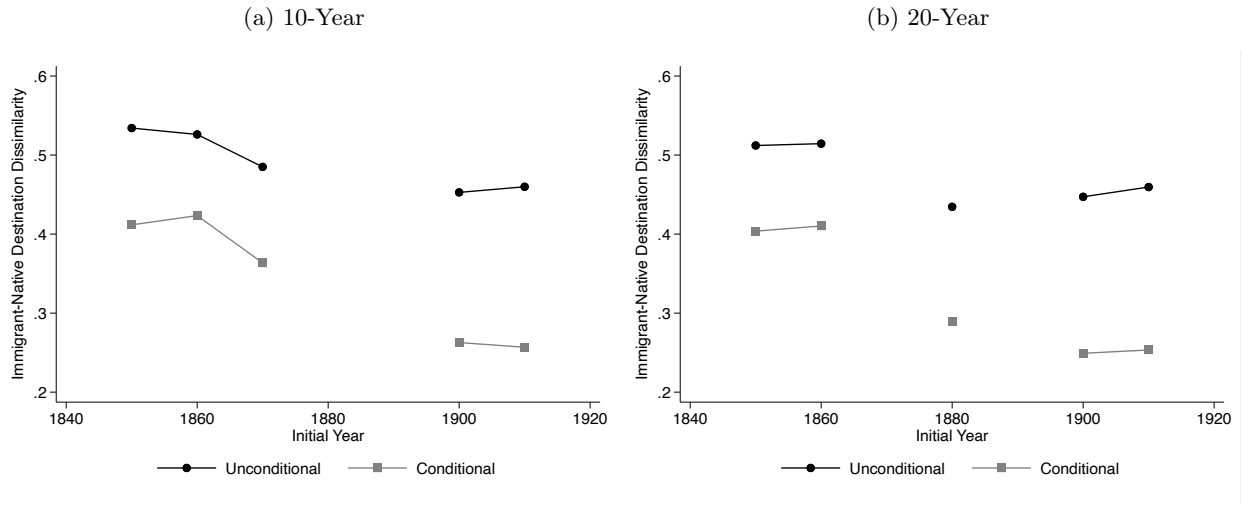
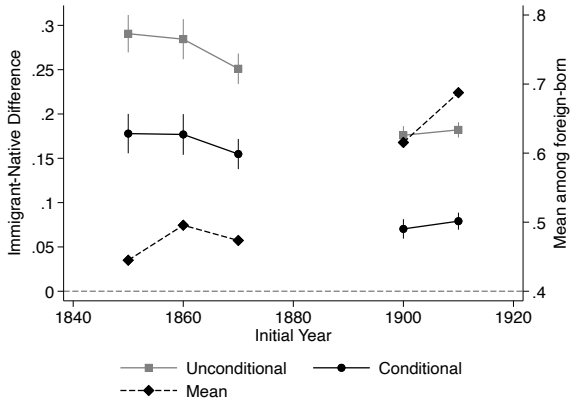


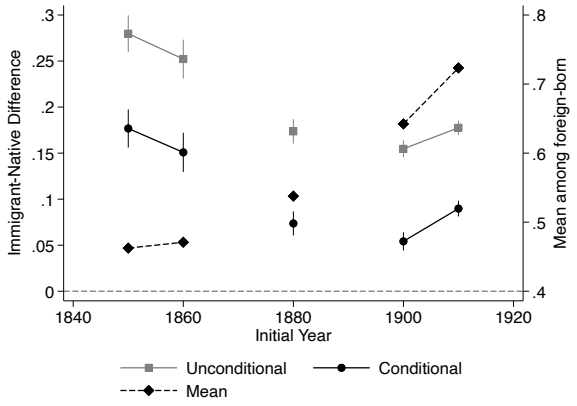
Figure E.3.4: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 45.

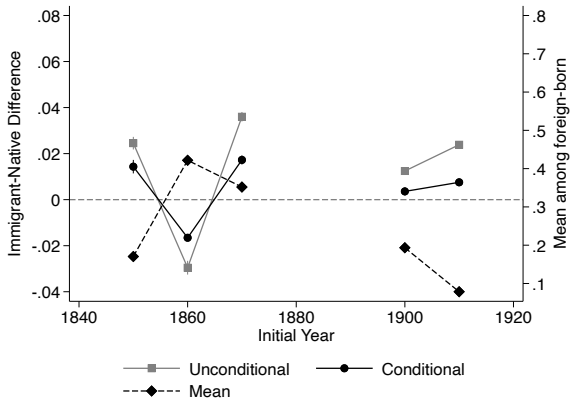
(a) Urban, 10-Year



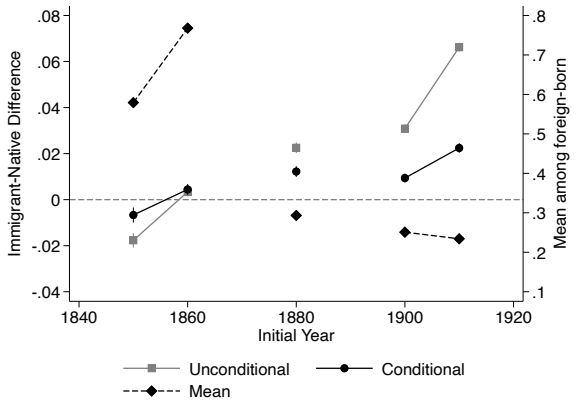
(b) Urban, 20-Year



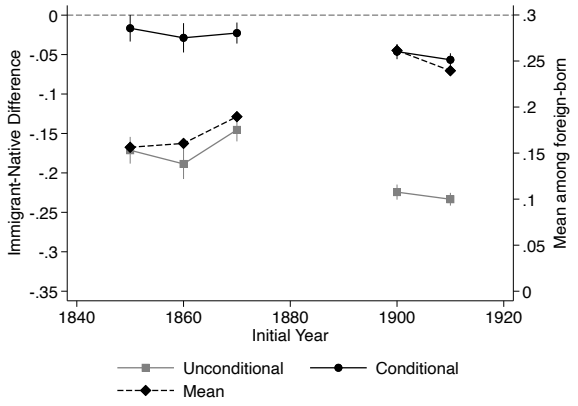
(c) Labor Demand Growth, 10-Year



(d) Labor Demand Growth, 20-Year



(e) South or West, 10-Year



(f) South or West, 20-Year

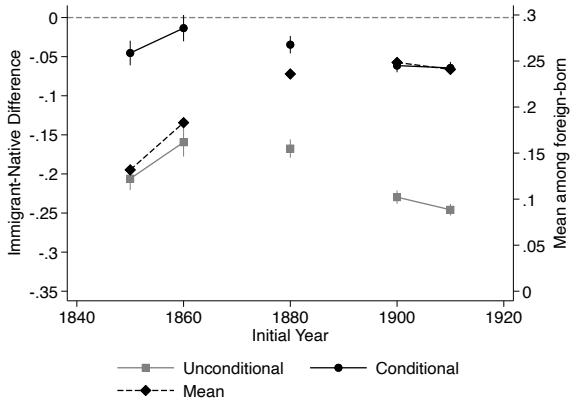


Figure E.3.5: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

(a) 10-Year

(b) 20-Year

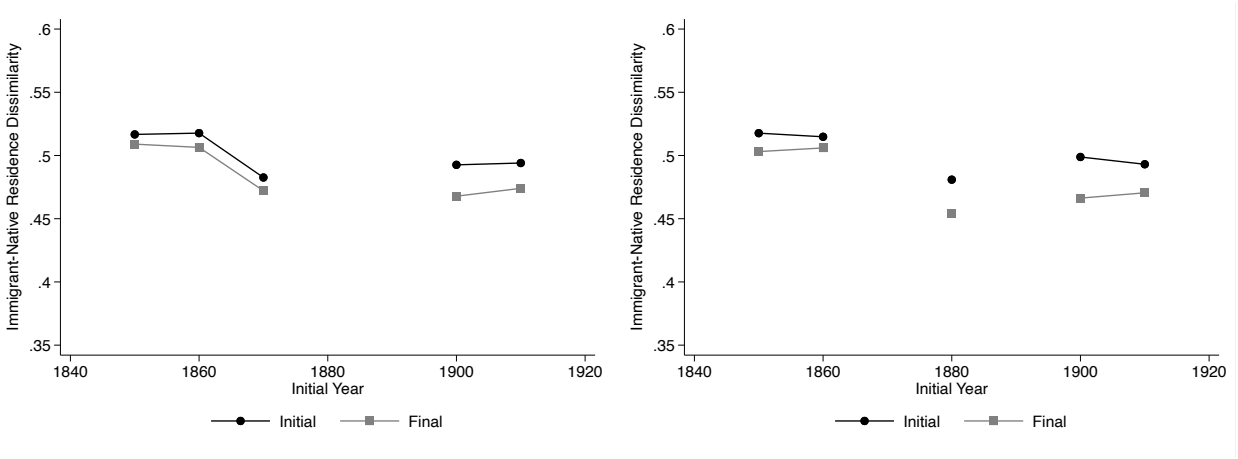


Figure E.3.6: Dissimilarity in counties of residence

Note: These figures present the dissimilarity index between the counties of residence of the foreign and native born in my linked samples in the initial and final year of the span beginning in the year on the x-axis.

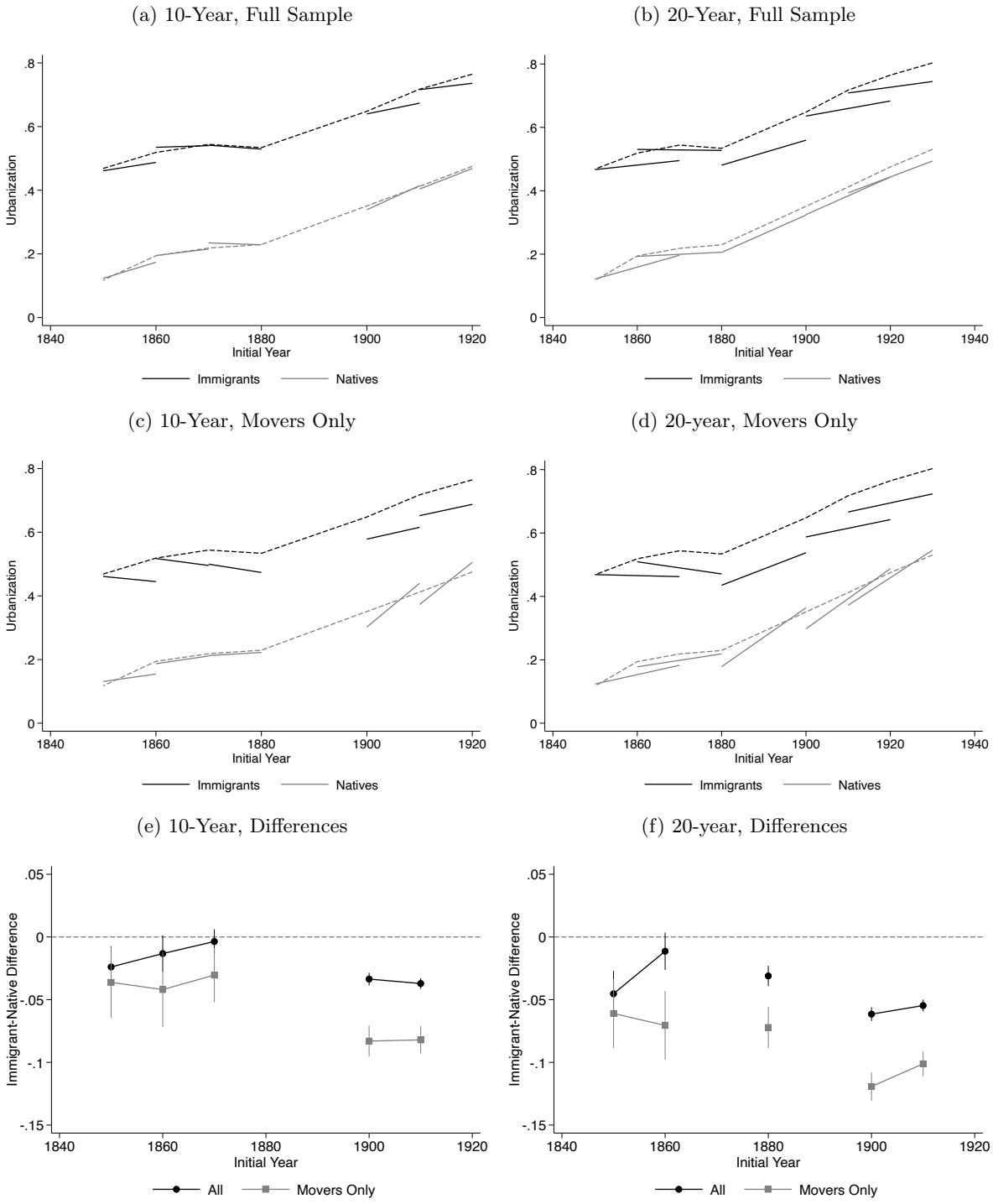
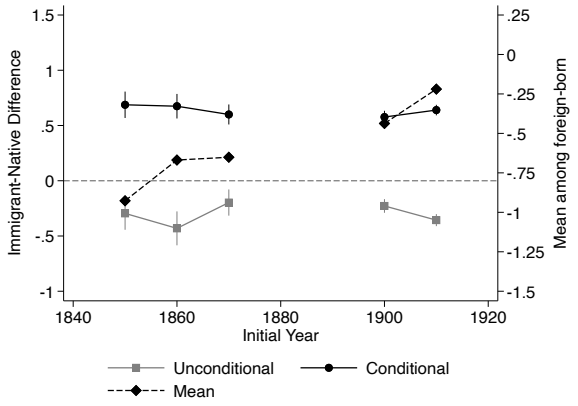


Figure E.3.7: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

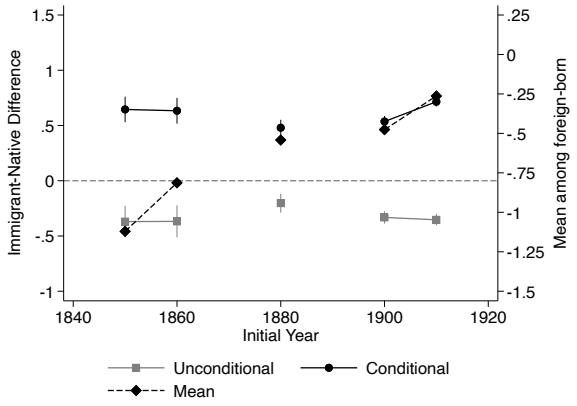


Figure E.3.8: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

E.4 Intersection Plus Corroboration

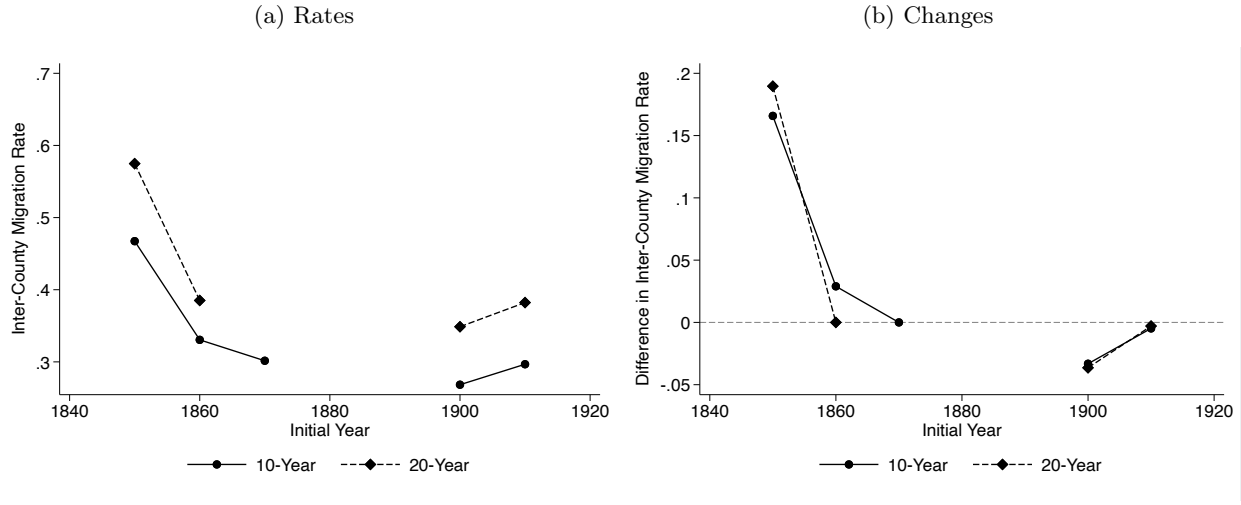


Figure E.4.1: Inter-county migration rates and changes, corrected for false matches

Note: Panel (a) shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

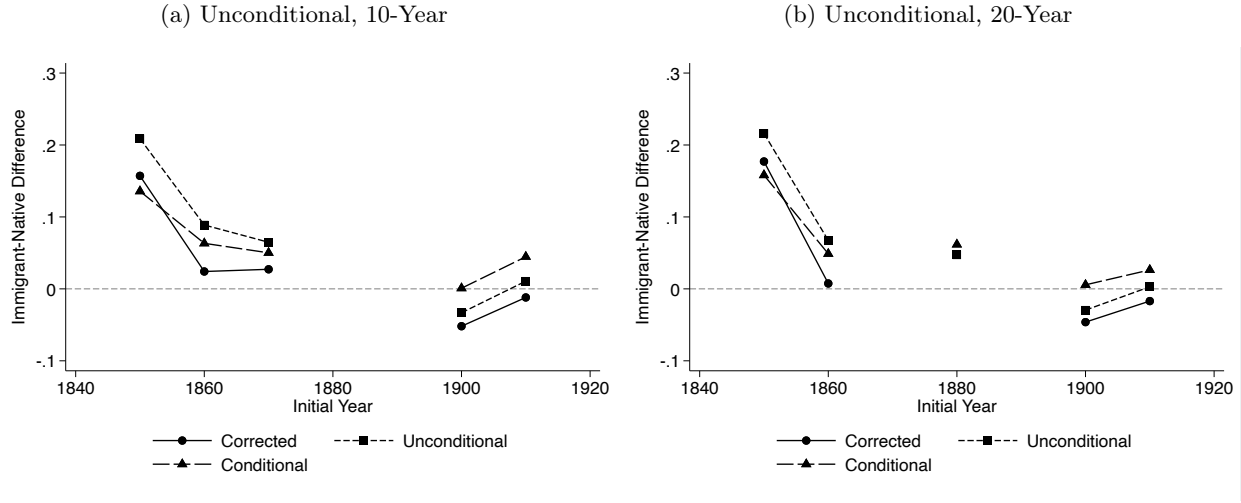
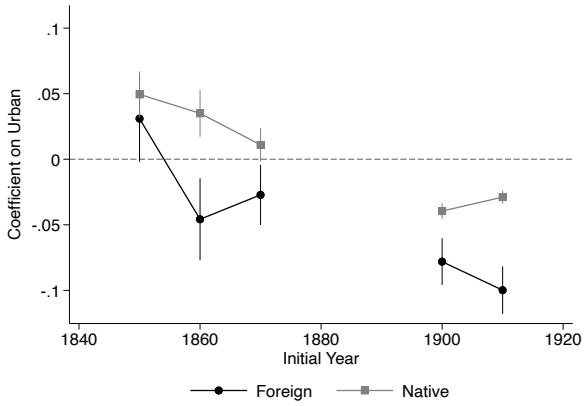


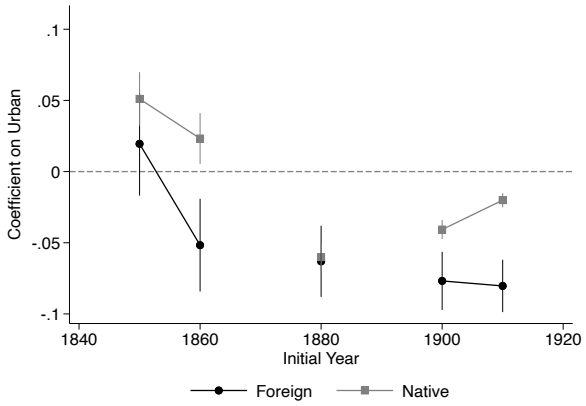
Figure E.4.2: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

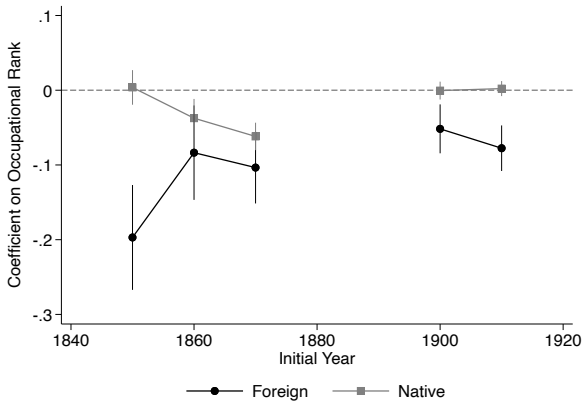
(a) Urban, 10-Year



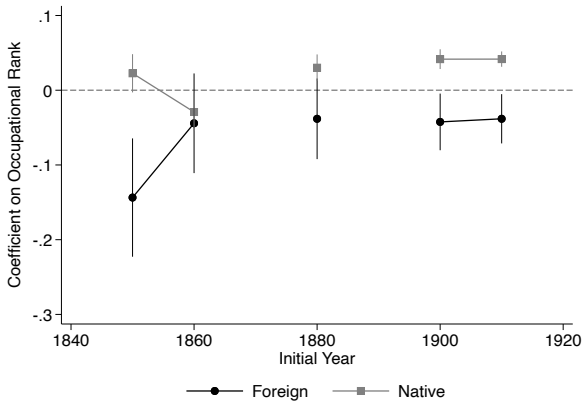
(b) Urban, 20-Year



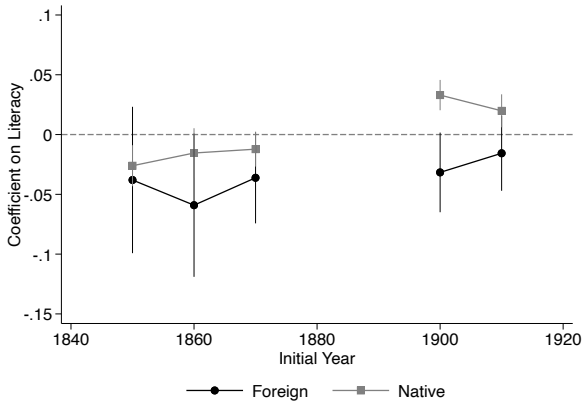
(c) Occupational Rank, 10-Year



(d) Occupational Rank, 20-Year



(e) Literacy, 10-Year



(f) Literacy, 20-Year

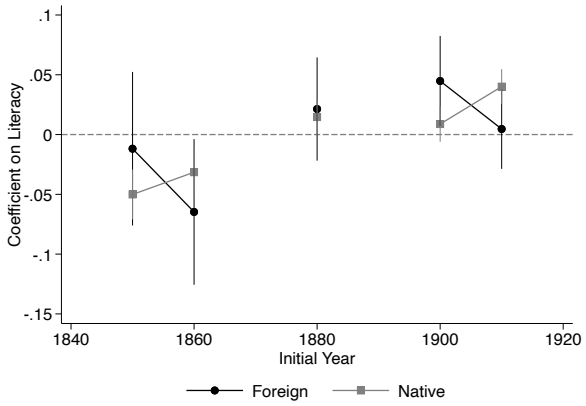


Figure E.4.3: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

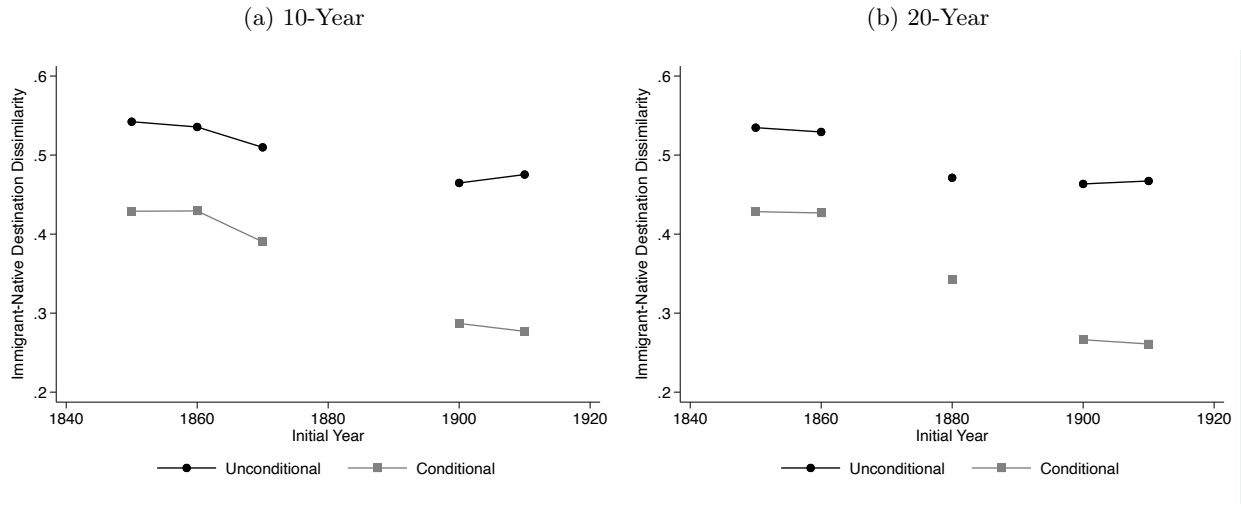
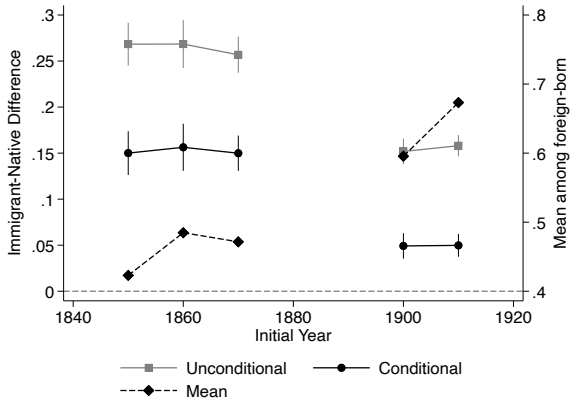


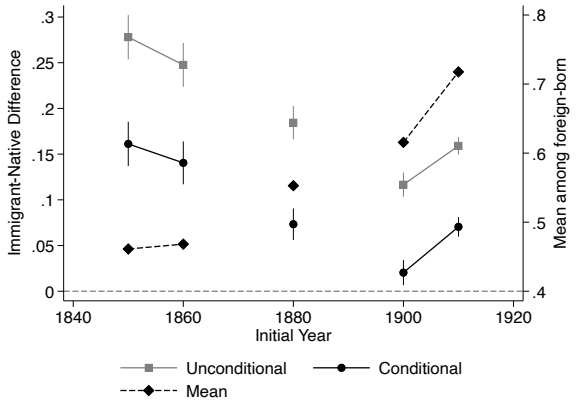
Figure E.4.4: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 45.

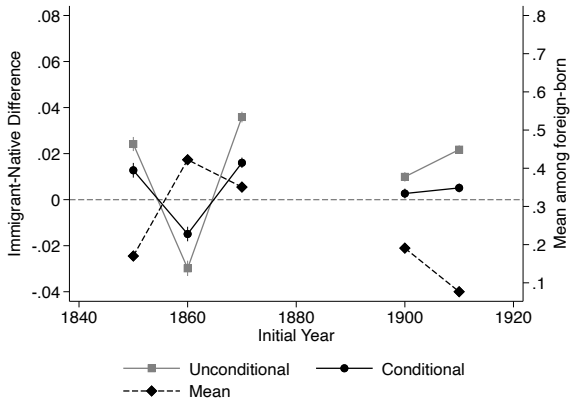
(a) Urban, 10-Year



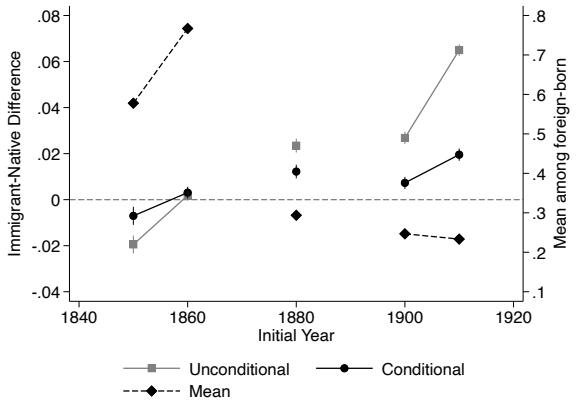
(b) Urban, 20-Year



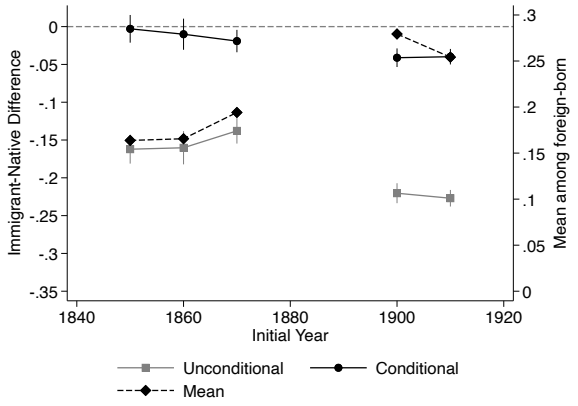
(c) Labor Demand Growth, 10-Year



(d) Labor Demand Growth, 20-Year



(e) South or West, 10-Year



(f) South or West, 20-Year

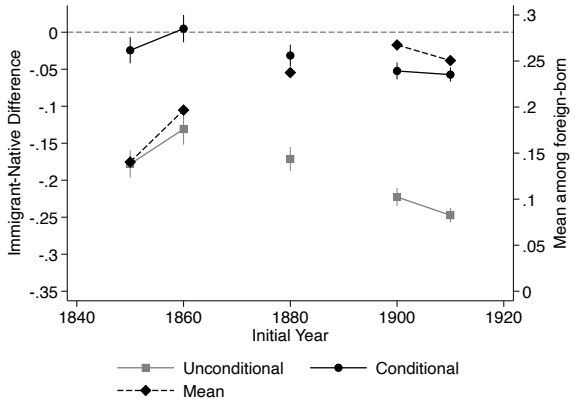


Figure E.4.5: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

(a) 10-Year

(b) 20-Year

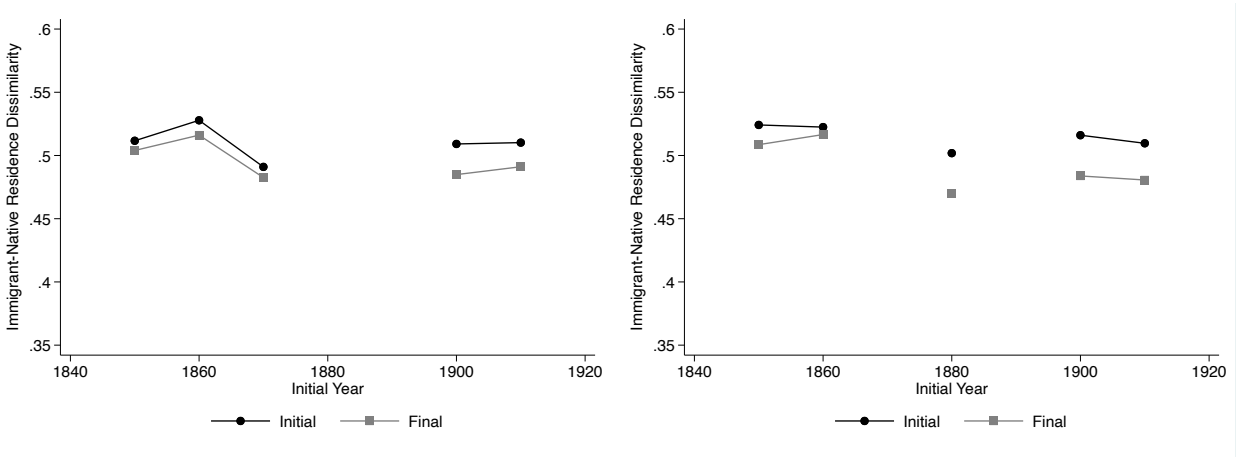


Figure E.4.6: Dissimilarity in counties of residence

Note: These figures present the dissimilarity index between the counties of residence of the foreign and native born in my linked samples in the initial and final year of the span beginning in the year on the x-axis.

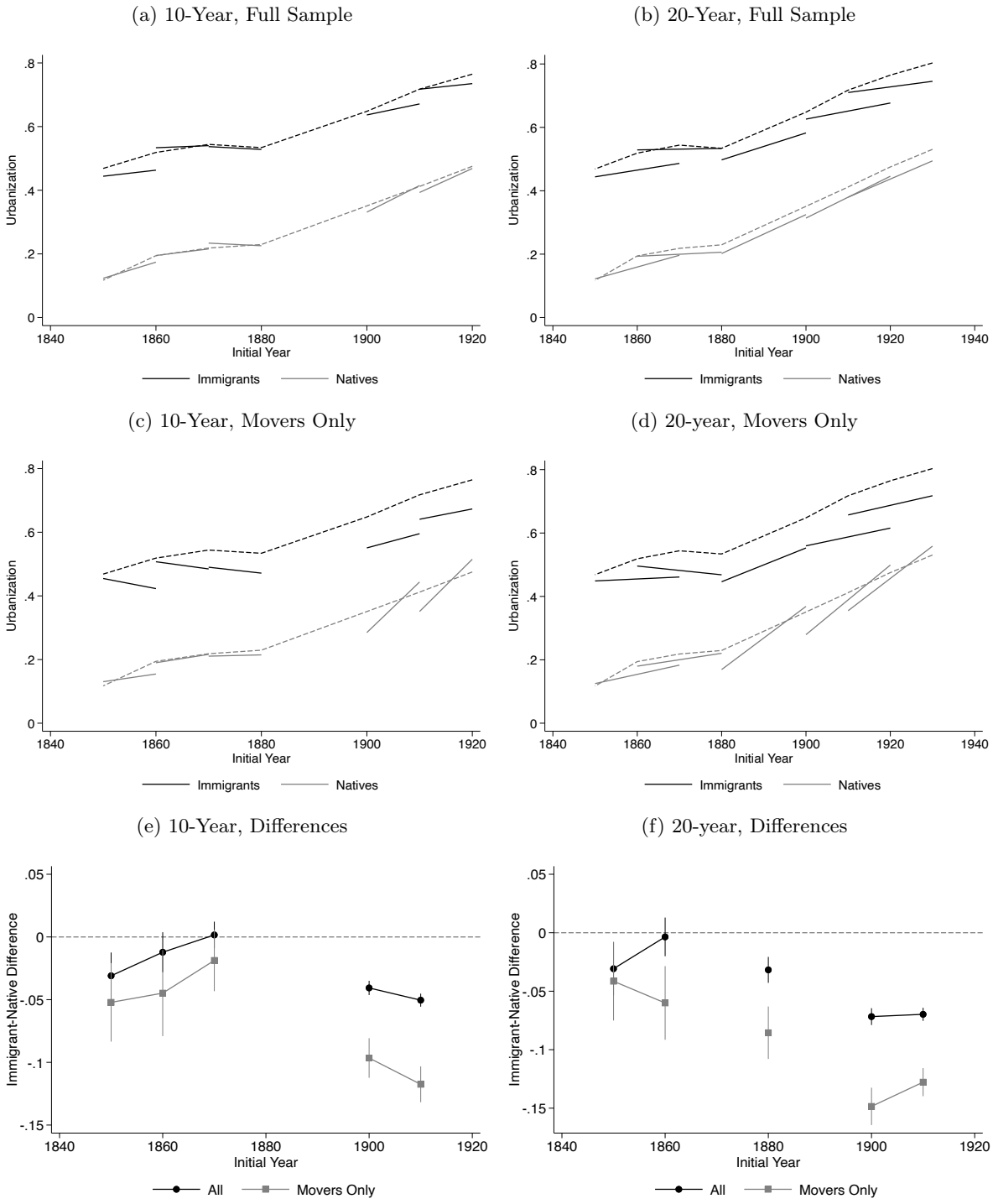
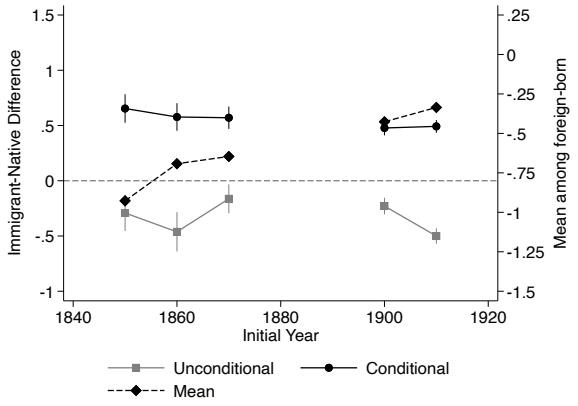


Figure E.4.7: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

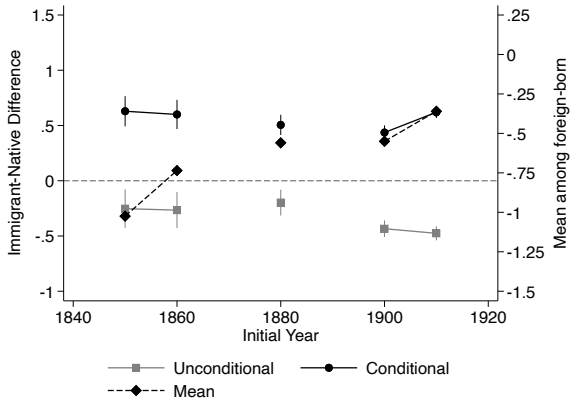


Figure E.4.8: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

F Results with Imputed Occupational Codes

The data for the censuses of 1900–1930 released by Ruggles et al. (2021) are preliminary, and in many cases individual's occupations have not yet been standardized. Instead, these observations have been given a code of 979 ("Not Yet Classified"). Because I use occupational information to create weights to correct for selection into linkage, such individuals must be omitted from the analysis. To verify that my results are not affected by this omission, this appendix aims to repeat the main analysis incorporating these individuals. In particular, following Collins and Zimran (2021), I assign to each individual the modal occupational code assigned to individuals whose occupational strings match that of the individual in question on the basis of the NYSIIS score. Occupations that were assigned a code in the original data retain that code.

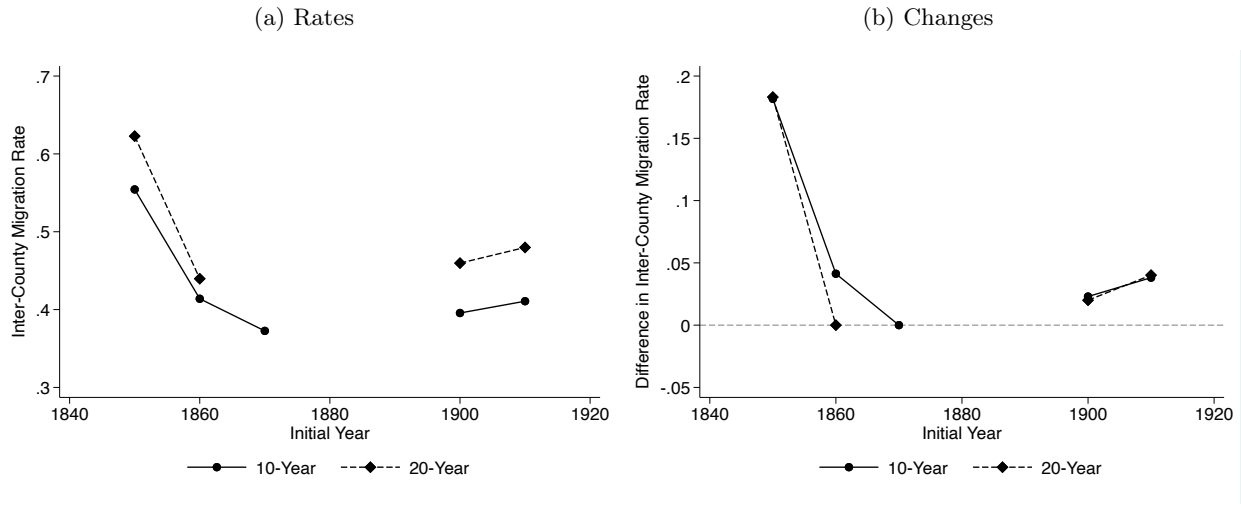


Figure F.1: Inter-county migration rates and changes, corrected for false matches

Note: Panel (a) shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

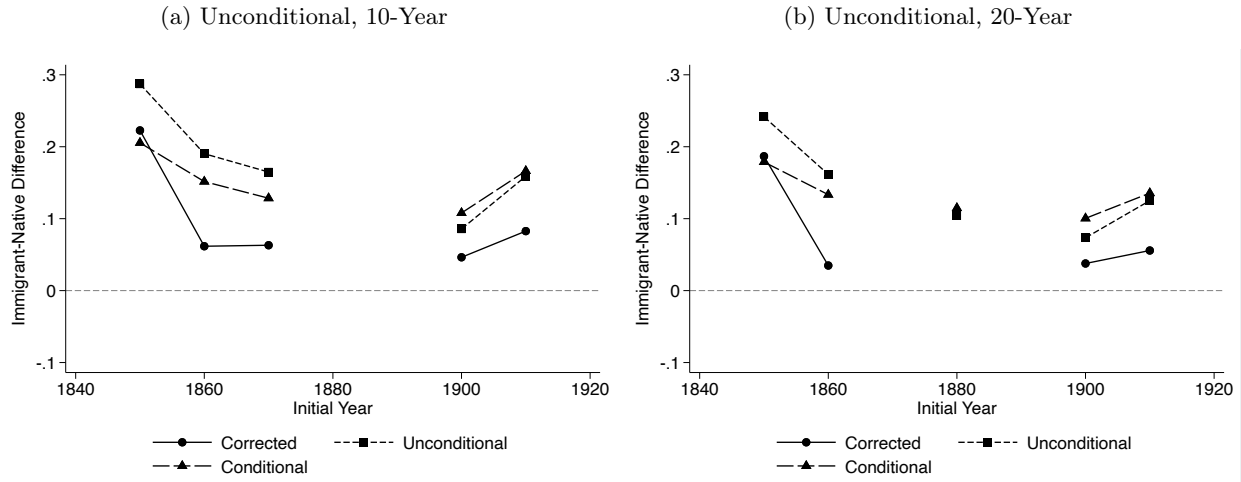


Figure F.2: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

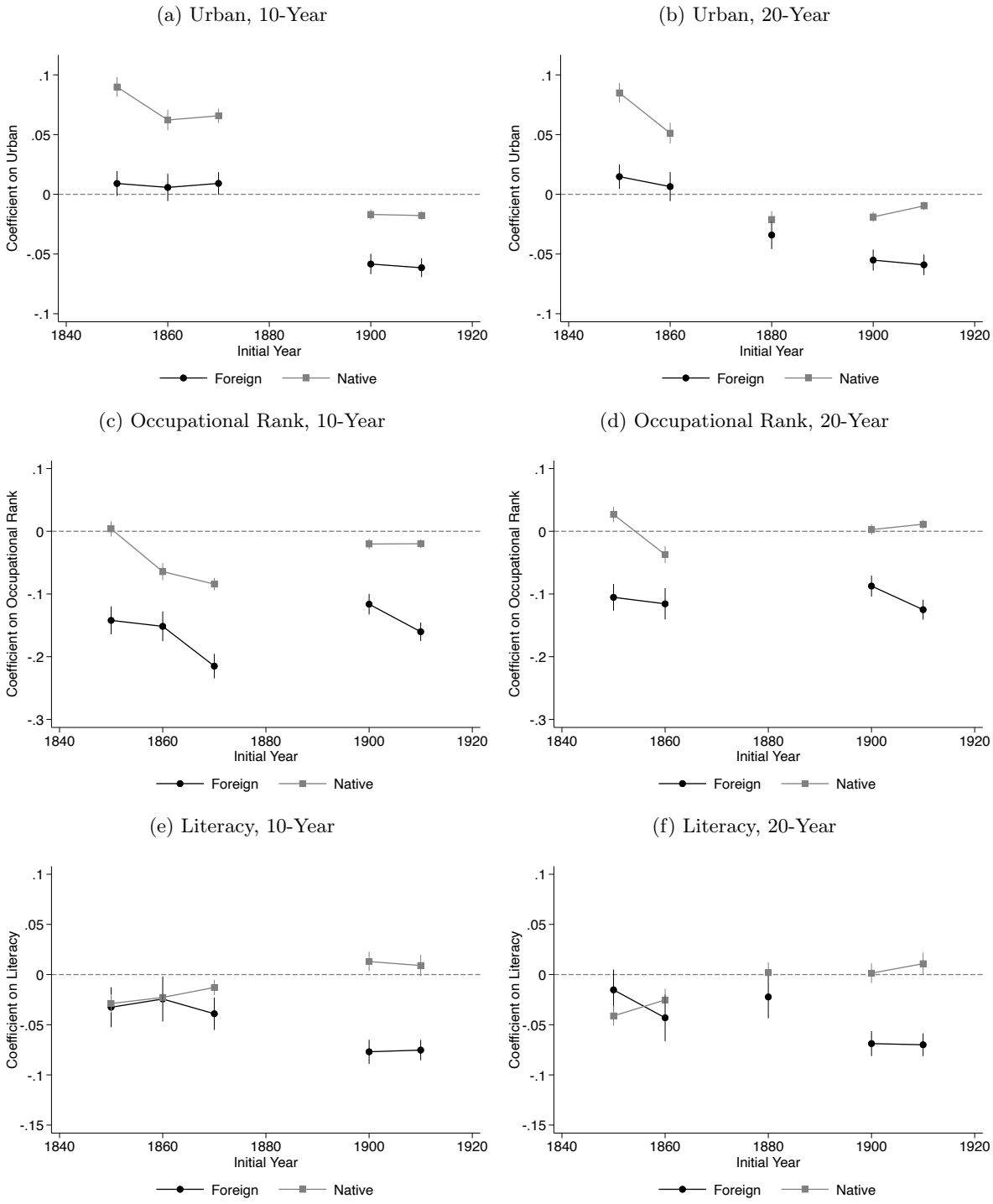


Figure F.3: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

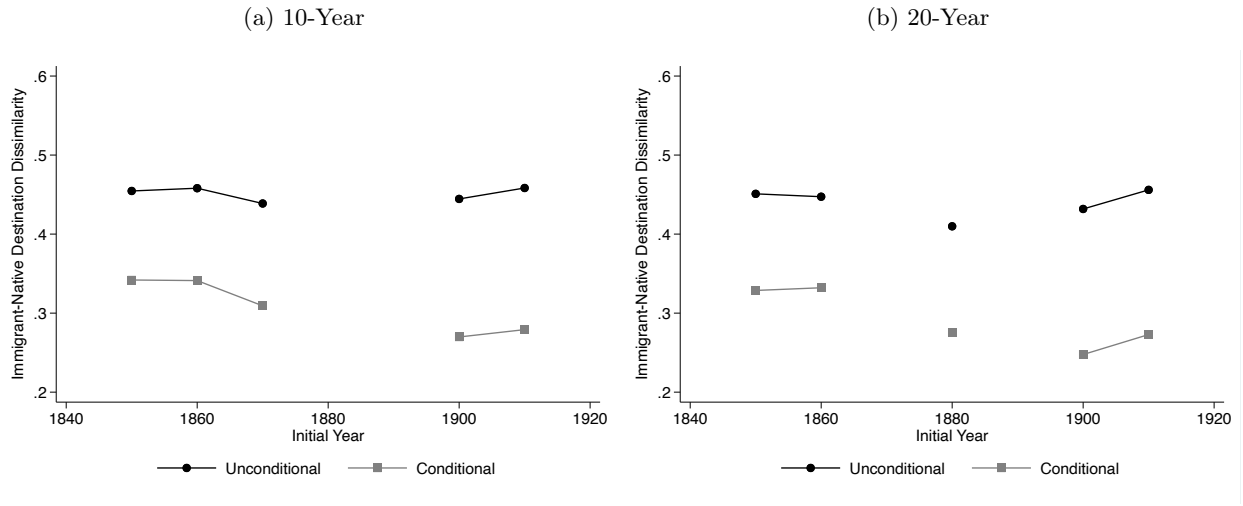
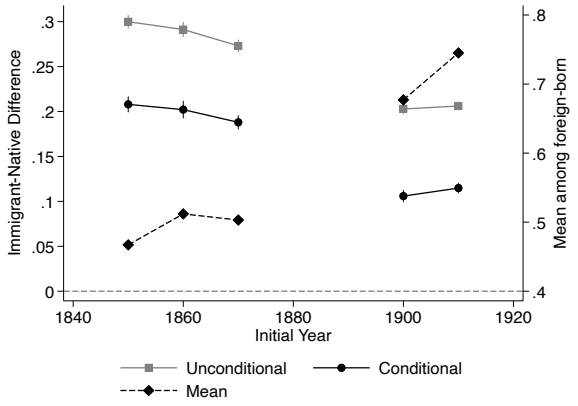


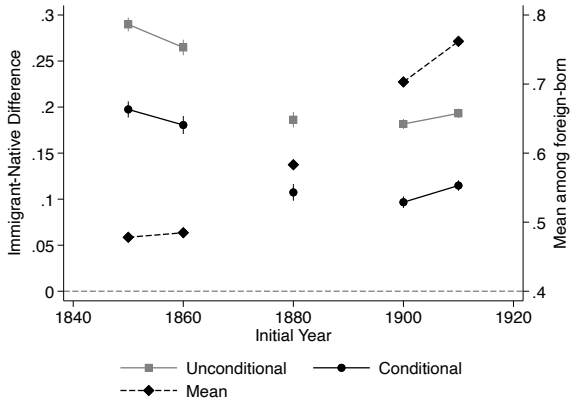
Figure F.4: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 45.

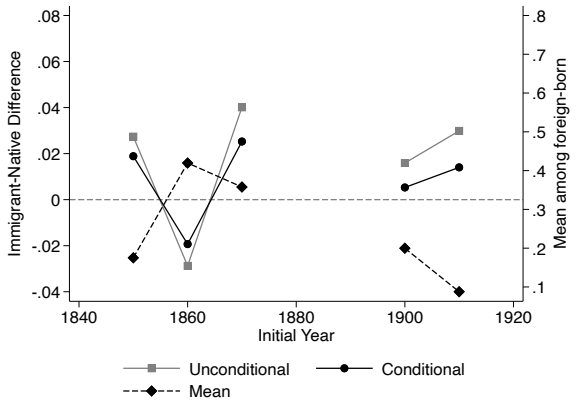
(a) Urban, 10-Year



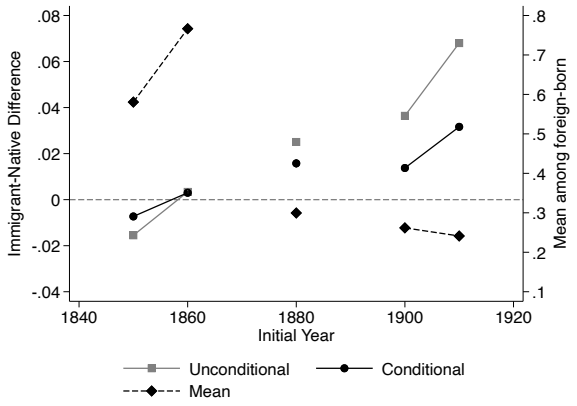
(b) Urban, 20-Year



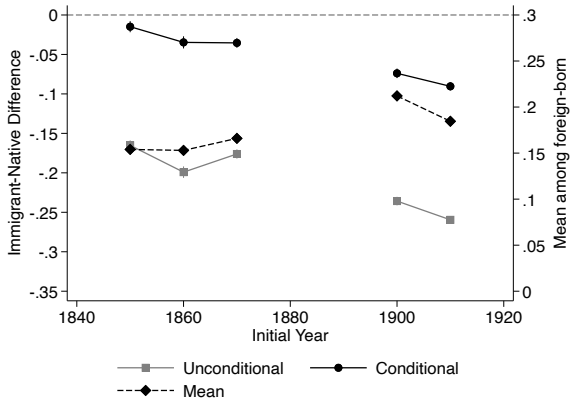
(c) Labor Demand Growth, 10-Year



(d) Labor Demand Growth, 20-Year



(e) South or West, 10-Year



(f) South or West, 20-Year

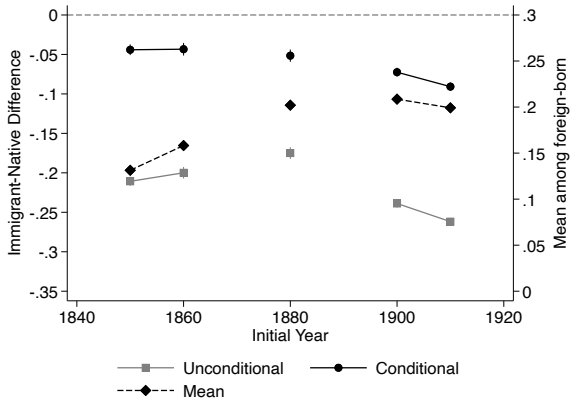


Figure F.5: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

(a) 10-Year

(b) 20-Year

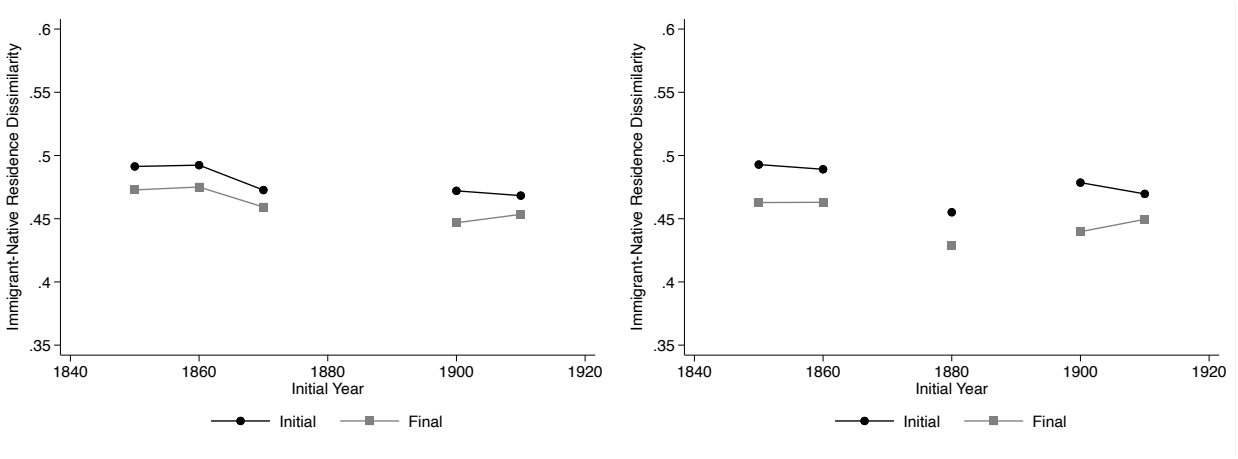


Figure F.6: Dissimilarity in counties of residence

Note: These figures present the dissimilarity index between the counties of residence of the foreign and native born in my linked samples in the initial and final year of the span beginning in the year on the x-axis.

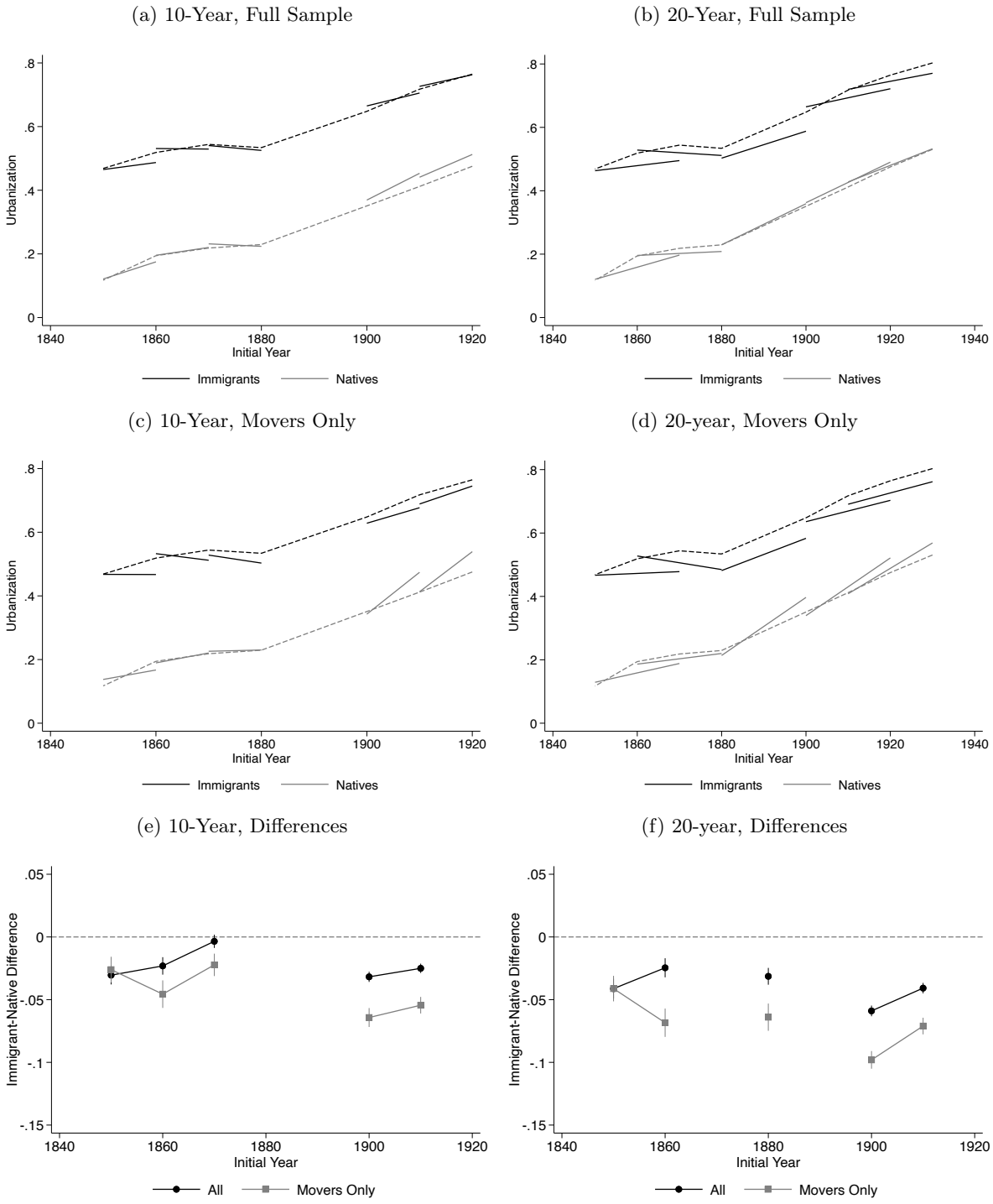
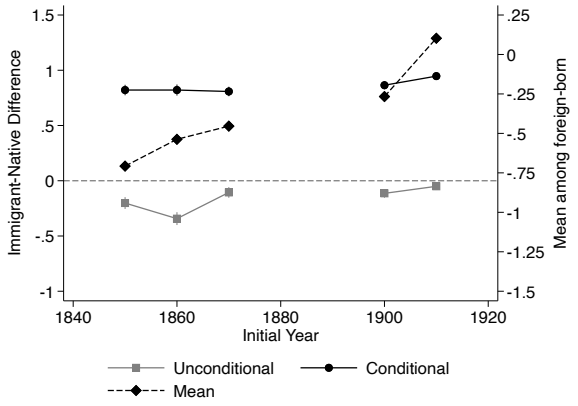


Figure F.7: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

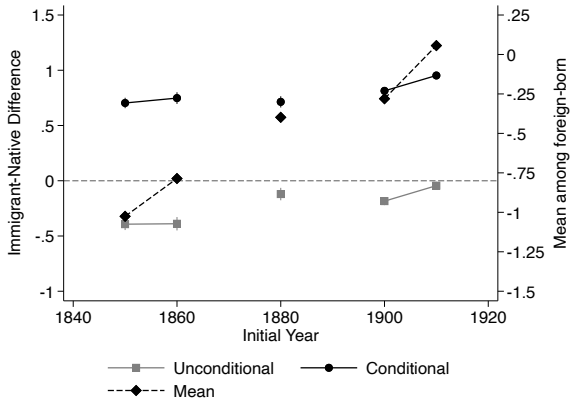


Figure F.8: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

G Alternate Urban Definitions

G.1 Over 25,000

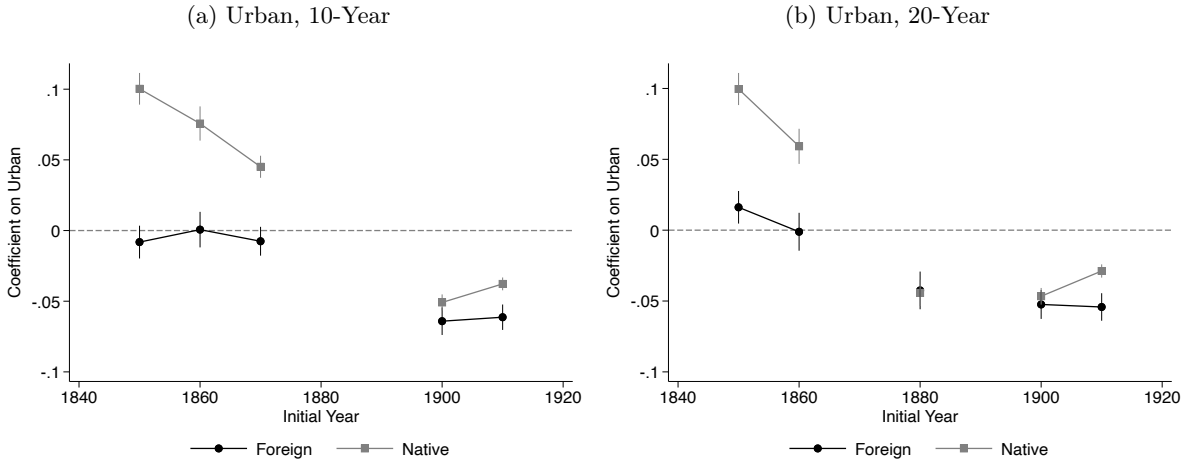


Figure G.1.1: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

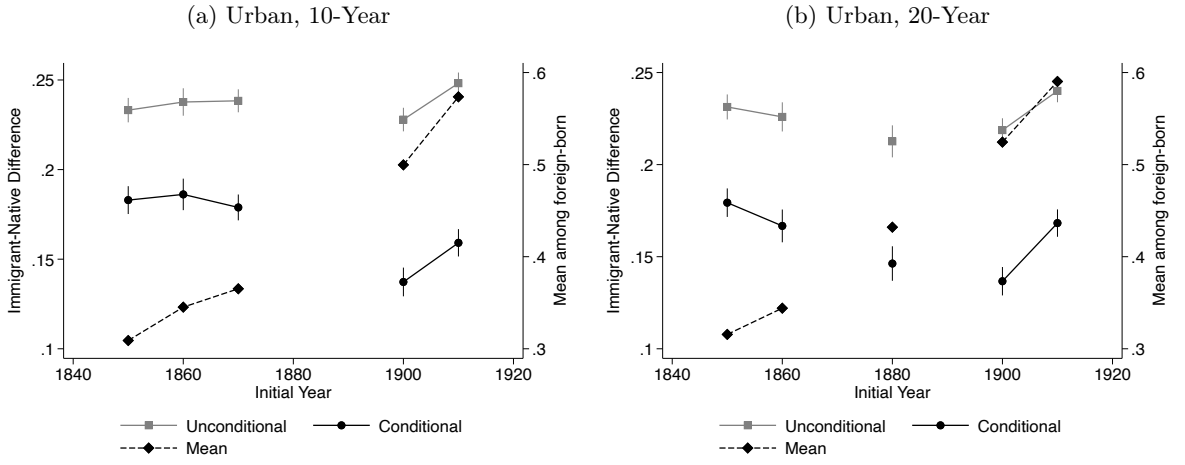


Figure G.1.2: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

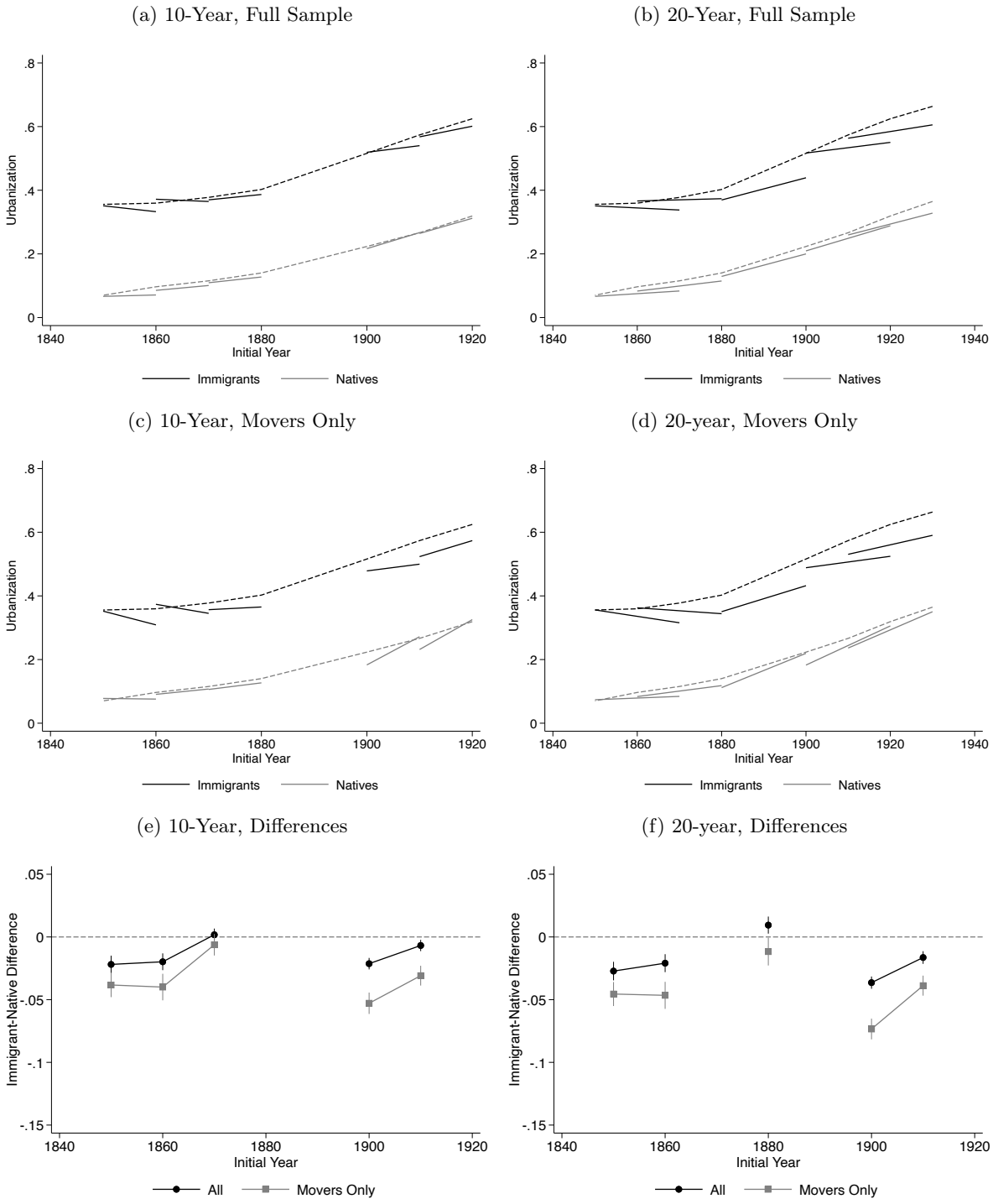


Figure G.1.3: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

G.2 20 Largest Cities

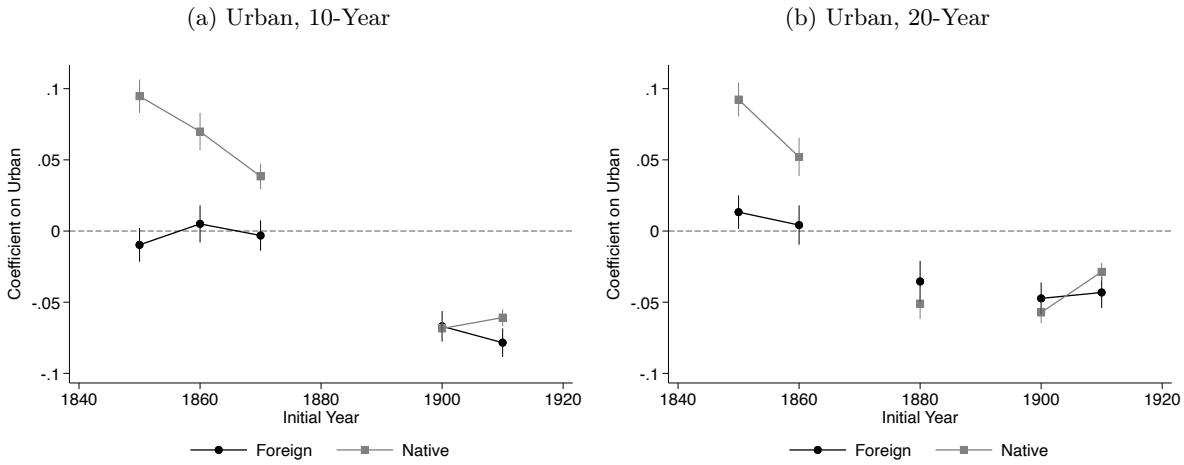


Figure G.2.1: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

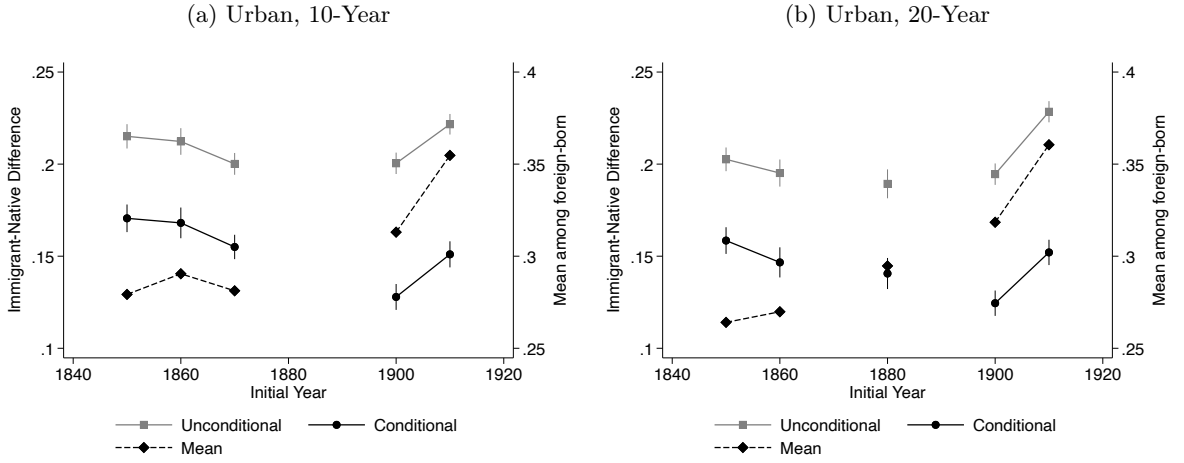


Figure G.2.2: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

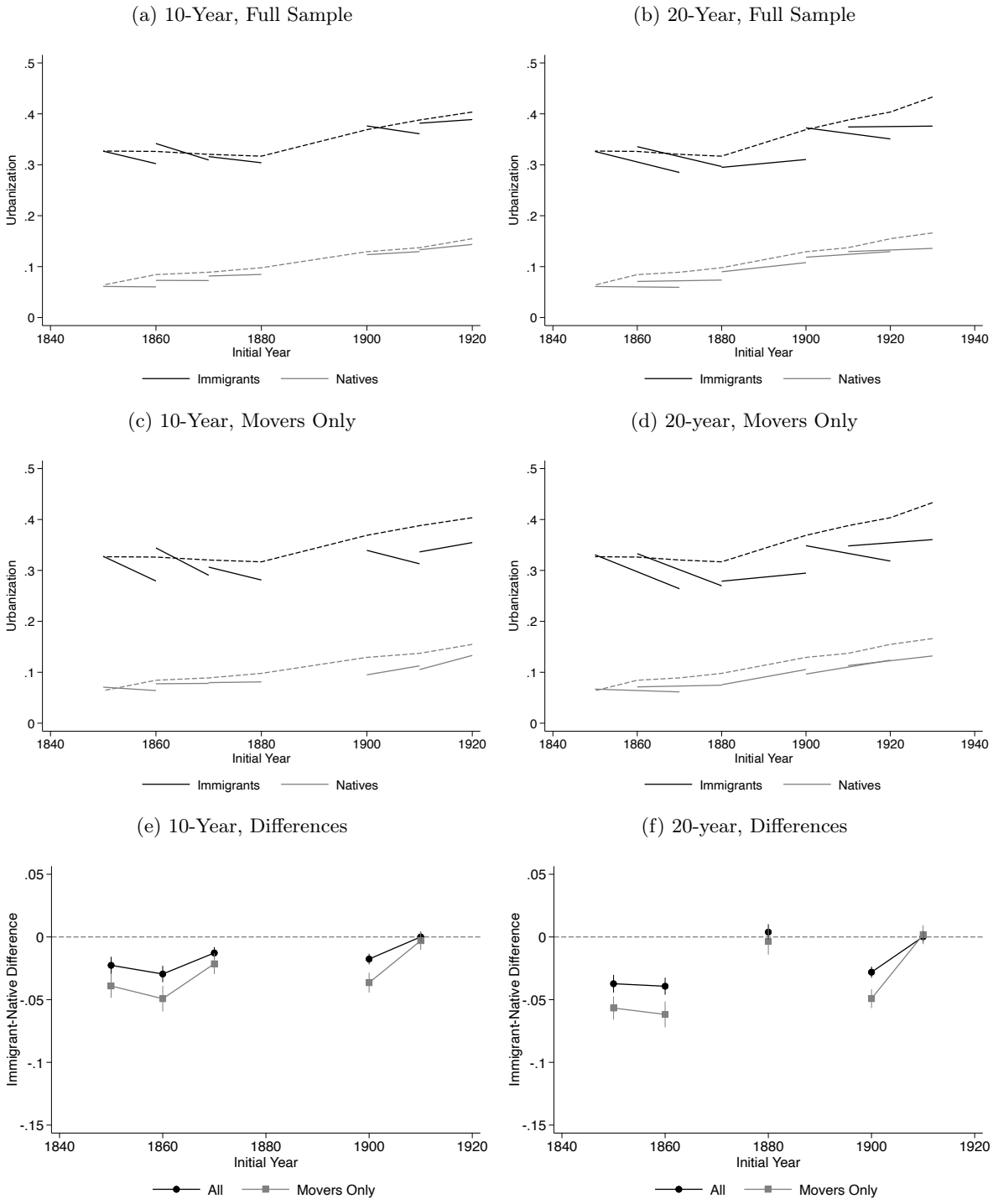


Figure G.2.3: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

G.3 Port Cities

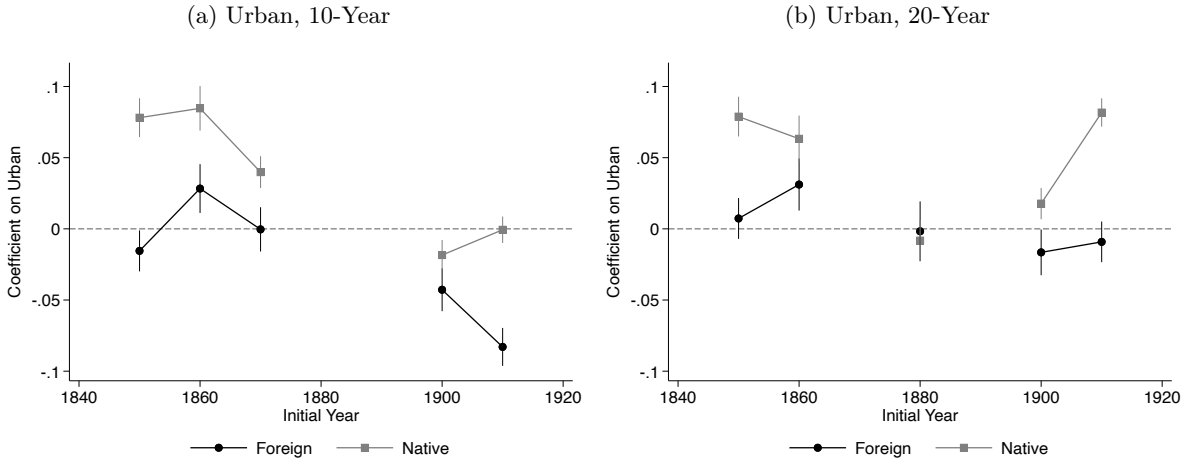


Figure G.3.1: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

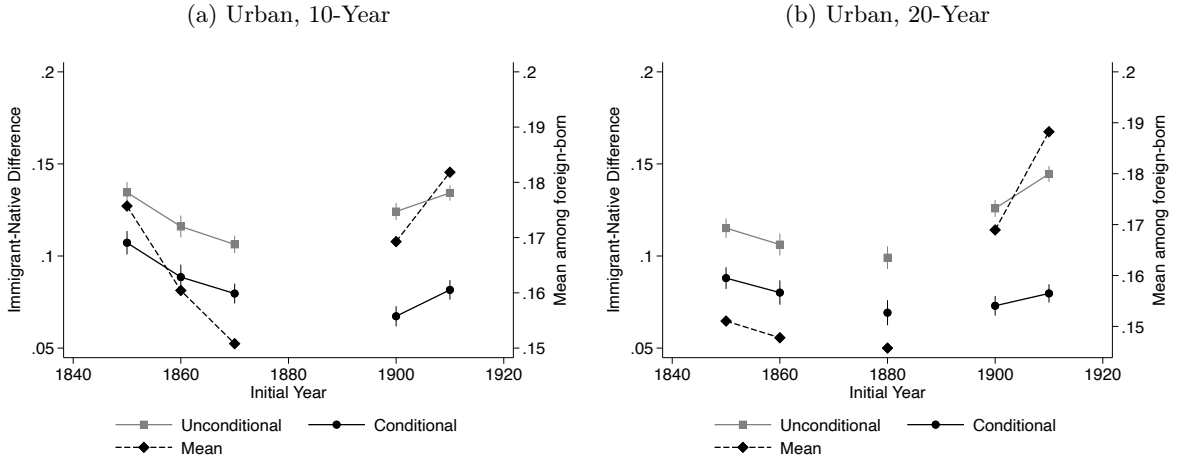


Figure G.3.2: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

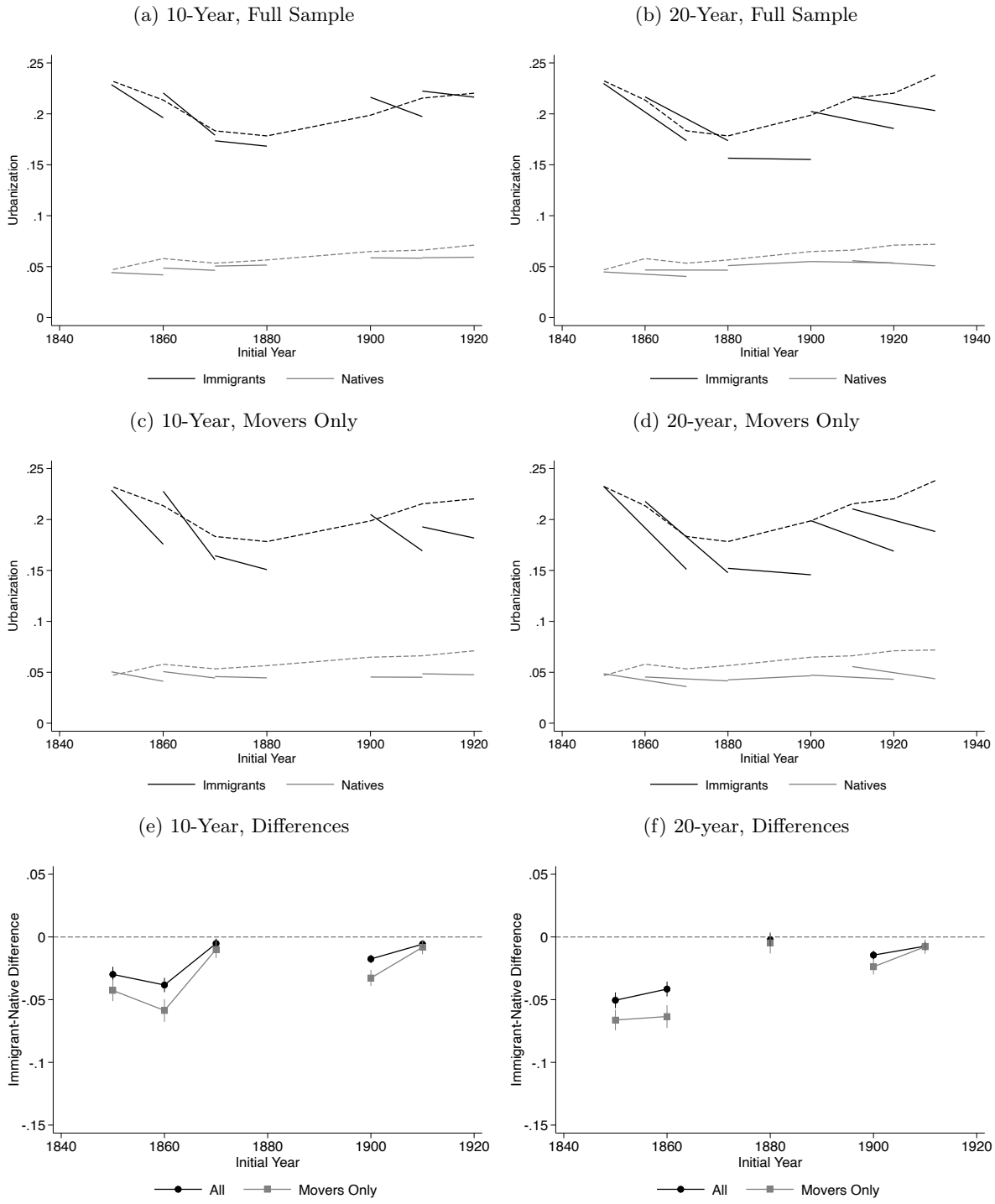


Figure G.3.3: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

H Results for Natives

This appendix presents analogous results for selection into linkage and migration rates to those of the main text, but for natives rather than immigrants. The results in this appendix are similar but not identical to those of Zimran (2021). The chief difference is that the selection into linkage (and resulting computation of weights) in Zimran (2021) is always based on the initial census of the span, whereas here the selection is based on the latter census in cases where that census provides information on immigrants' year of arrival. This is done to keep these results consistent with those for immigrants.

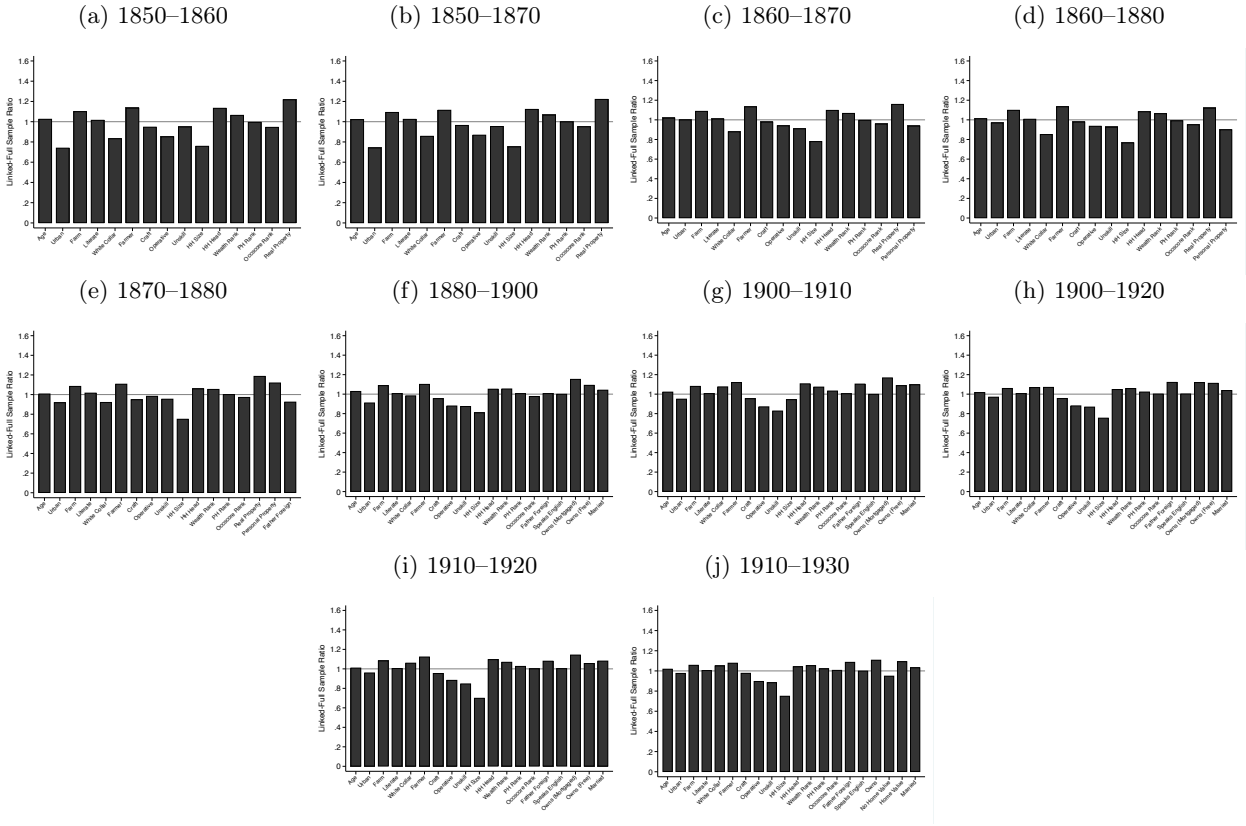


Figure H.1: Representativeness of the linked sample

Note: Each bar presents the ratio of the mean of each variable in the linked sample relative to the full sample at risk for linkage. For panels (a)–(e), the sample at risk is from the initial census year. For panels (f)–(j), the sample at risk is from the final census year, limiting to individuals whose ages put them at risk for linkage.

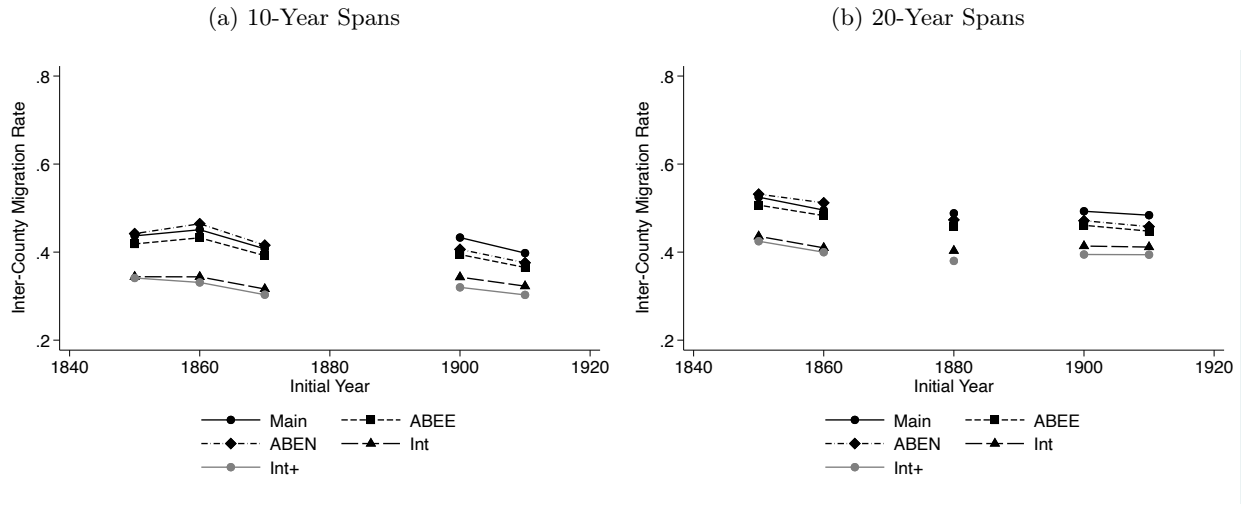


Figure H.2: Inter-county migration rates for the native born

Note: Each figure shows the probability that a foreign-born individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, according to each linkage method. All observations are weighted by inverse linkage probability.

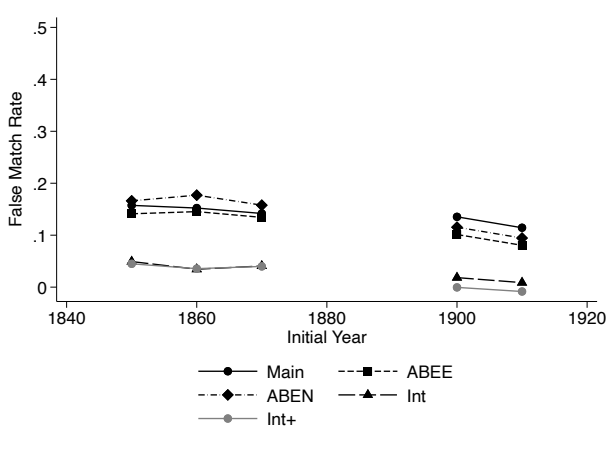


Figure H.3: Estimated rates of false linkage, native-born individuals

Note: This figure shows the estimated rate of false matching for each linkage method and 10-year span according to the comparison of inter-state migration estimates by the linkage method and by the method based on the birth places and ages of children. The rate is computed according to Bayes's Theorem, as explained in text and in Appendix A.

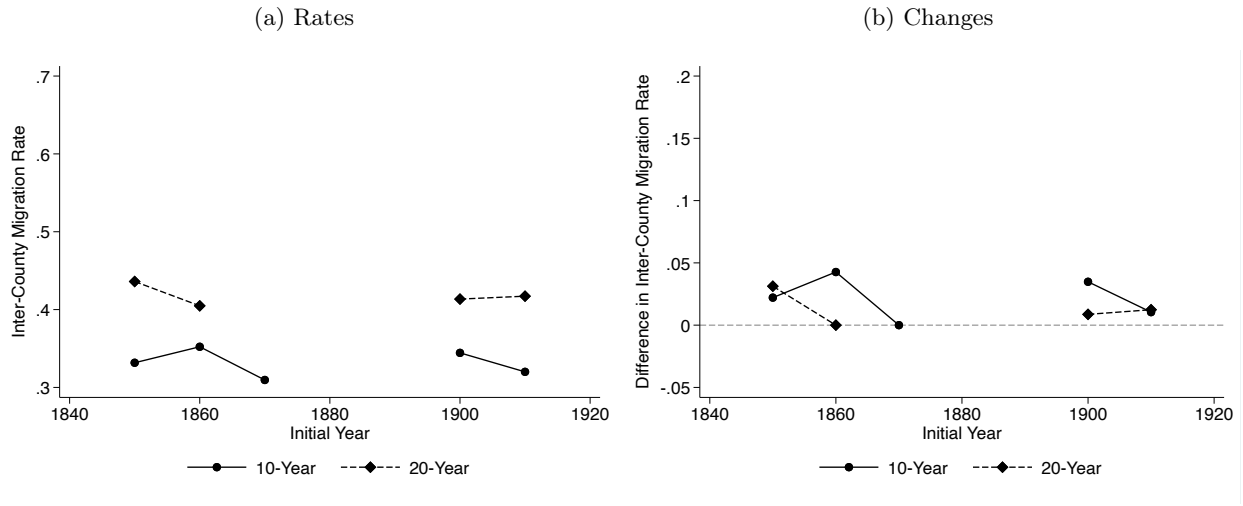


Figure H.4: Inter-county migration rates and changes, corrected for false matches, natives

Note: Panel (a) shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

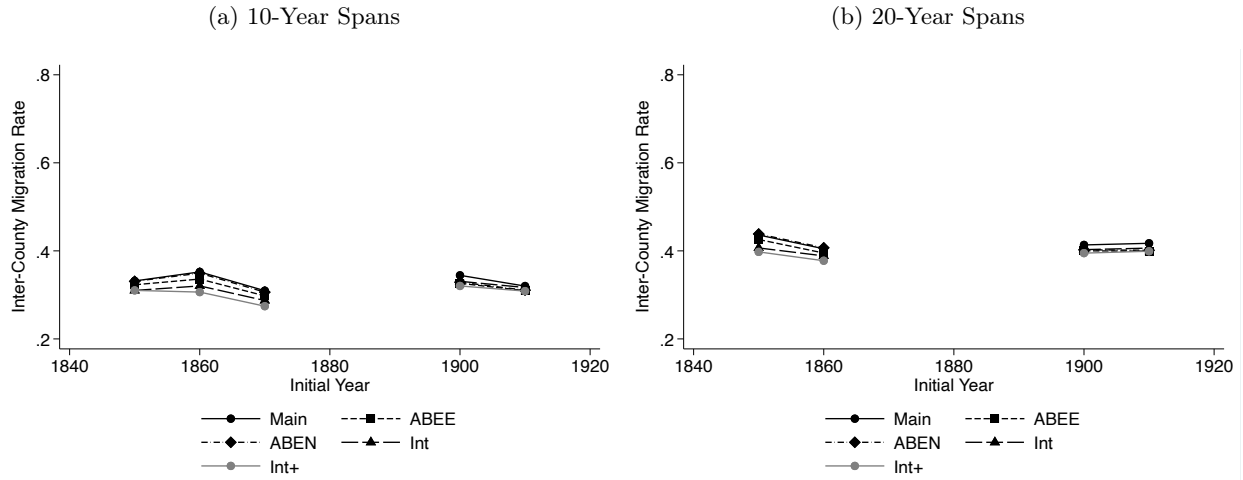


Figure H.5: Corrected migration rates for natives by all linkage methods

Note: Each figure shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, according to each linkage method. All observations are weighted by inverse linkage probability. The estimates are then adjusted for false matches according to the method presented in text.

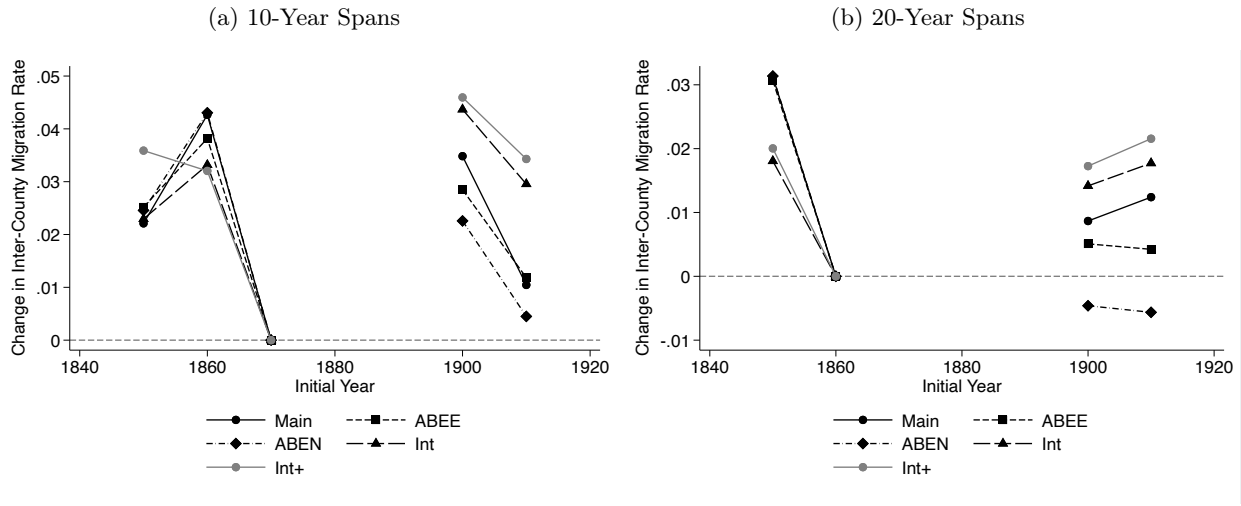


Figure H.6: Changes in corrected inter-county migration rates for the native born

Note: Each figure shows the change in the estimated inter-county migration rate corrected for false matches (Figure H.5 relative to that of the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

I Results for Longer-Distance Moves

In this appendix, I repeat the analysis of the main text, but change the definition of migration to require that a move not only cross county lines, but that the distance between the initial and final counties' centroids be at least 150 miles. The analysis is carried out in the same way in all respects as that of the main text, except for the correction of the estimated migration rates. In the standard analysis, I can simplify equation (A.3) by setting $P(m|f)$ to 1, since that is essentially the probability of linking an individual to someone in a different county when the link is made in error. But the probability that a false link is made to someone at least 150 miles away is smaller, since much of the false matching is determined by the distribution of birthplaces. To this end, I compute, for each individual in the initial census, the probability that an individual in the final census of a span to whom he could be linked (i.e., had a matching birthplace and sufficiently similar age-implied birthyear) lived in a county at least 150 miles away. I then compute $P(m|t)$ without the simplification used in the rest of the analysis. Because the sample is unchanged relative to the main analysis, I use the same estimates of $P(f)$ as in the main analysis

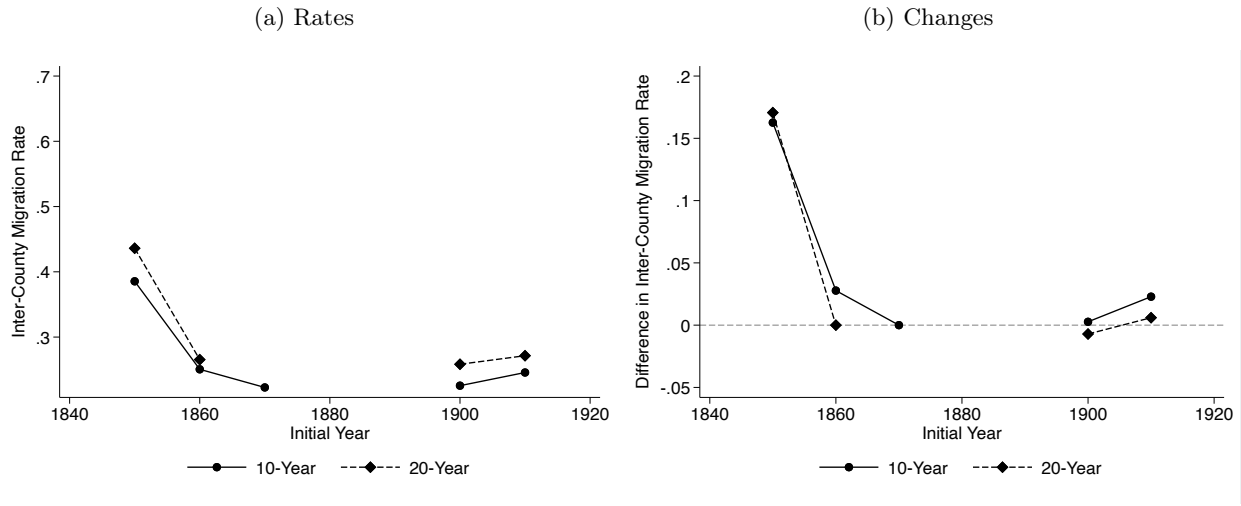


Figure I.1: Inter-county migration rates and changes, corrected for false matches

Note: Panel (a) shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county at least 150 miles from their original county 10 or 20 years later, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

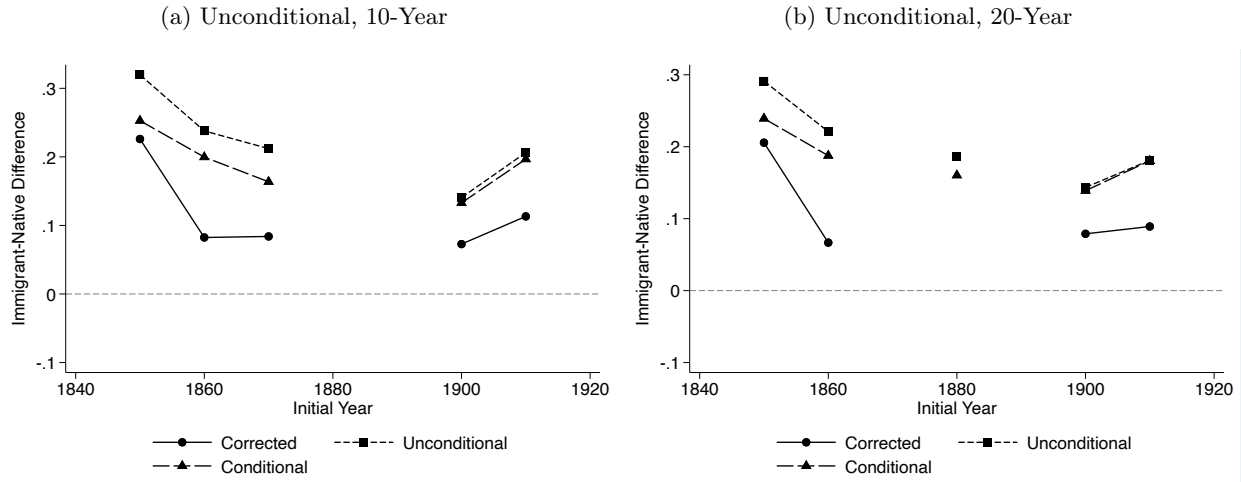


Figure I.2: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

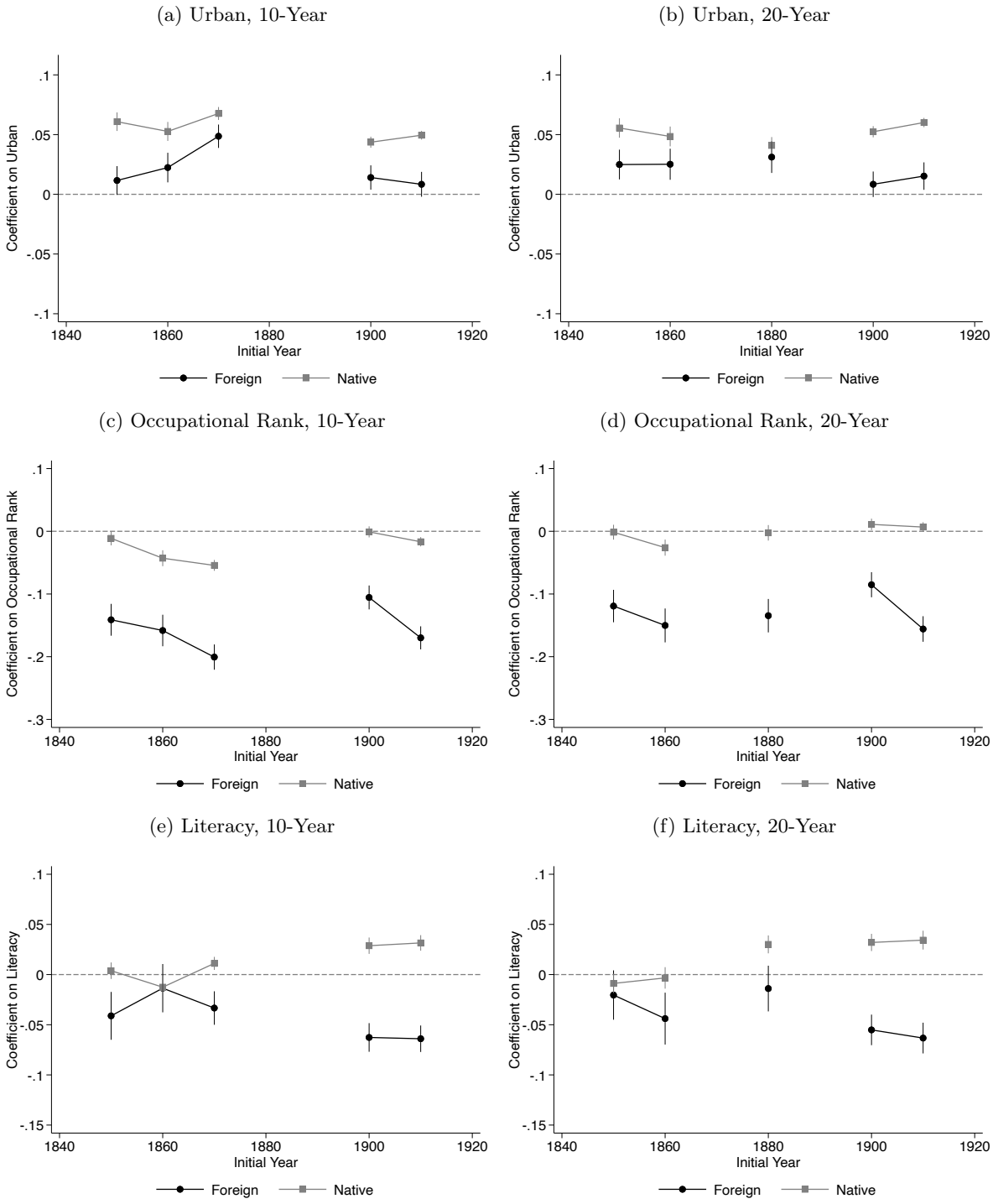


Figure I.3: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

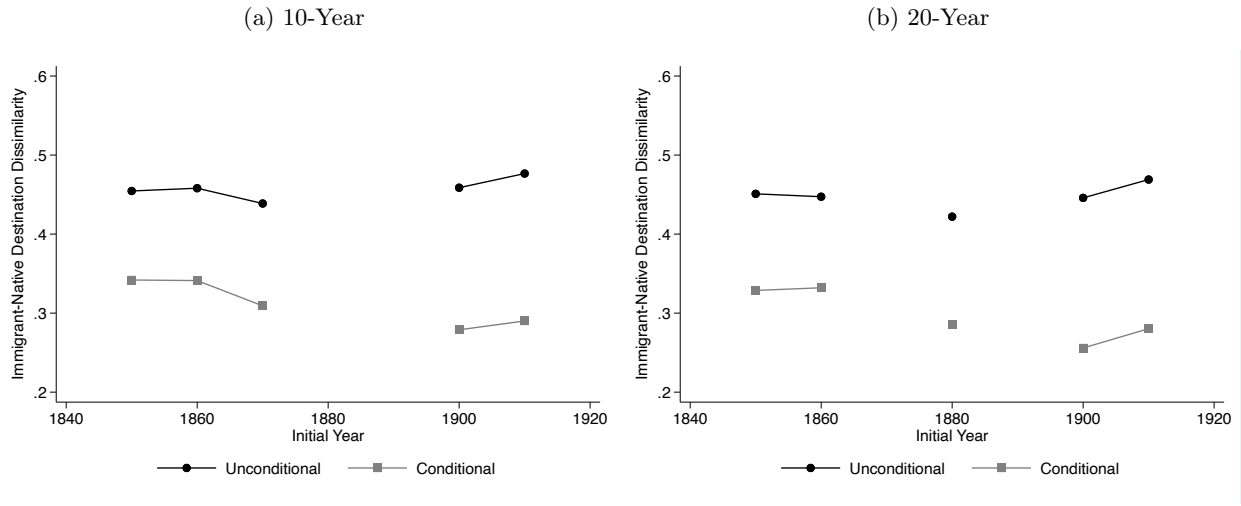


Figure I.4: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 45.

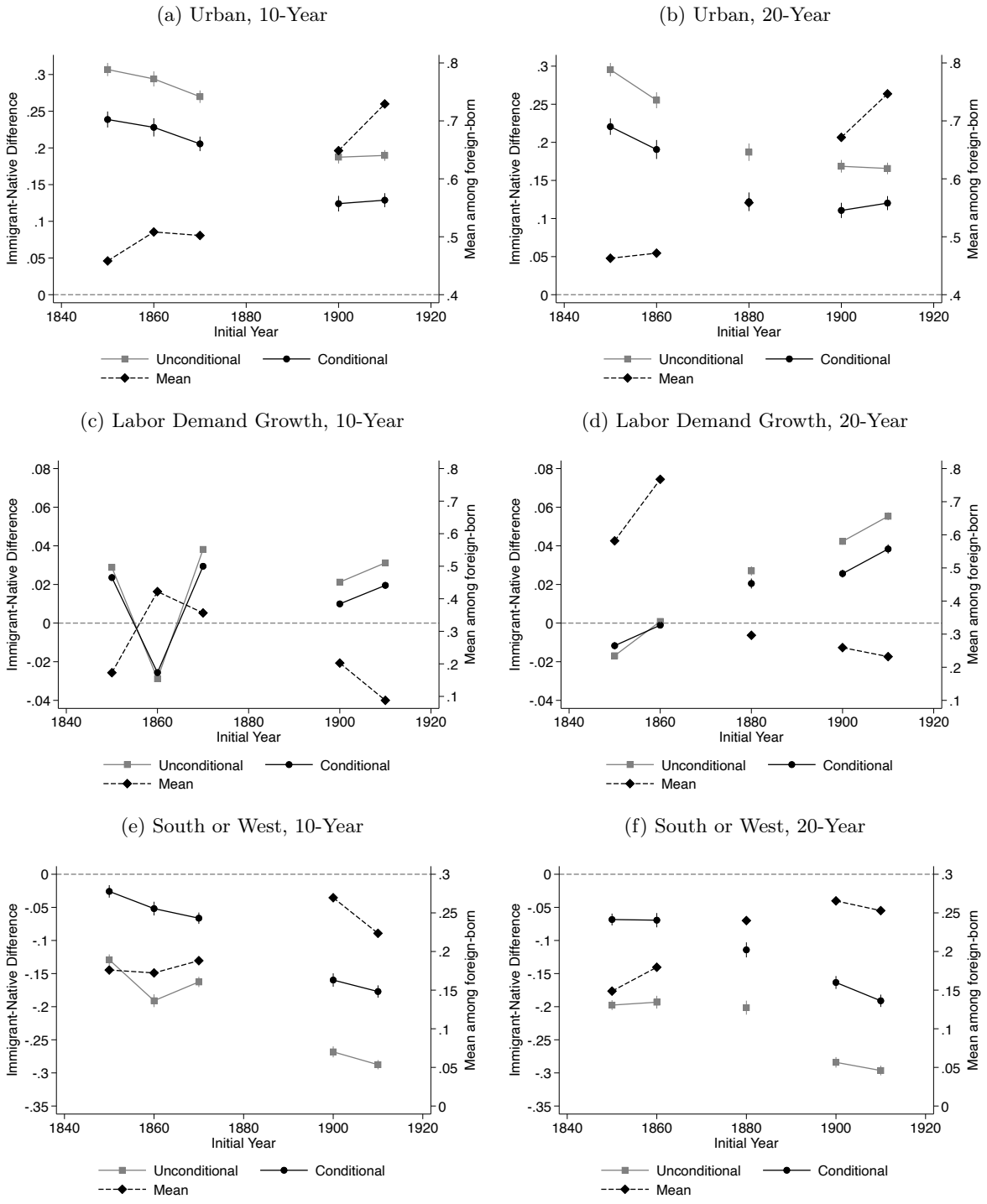


Figure I.5: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

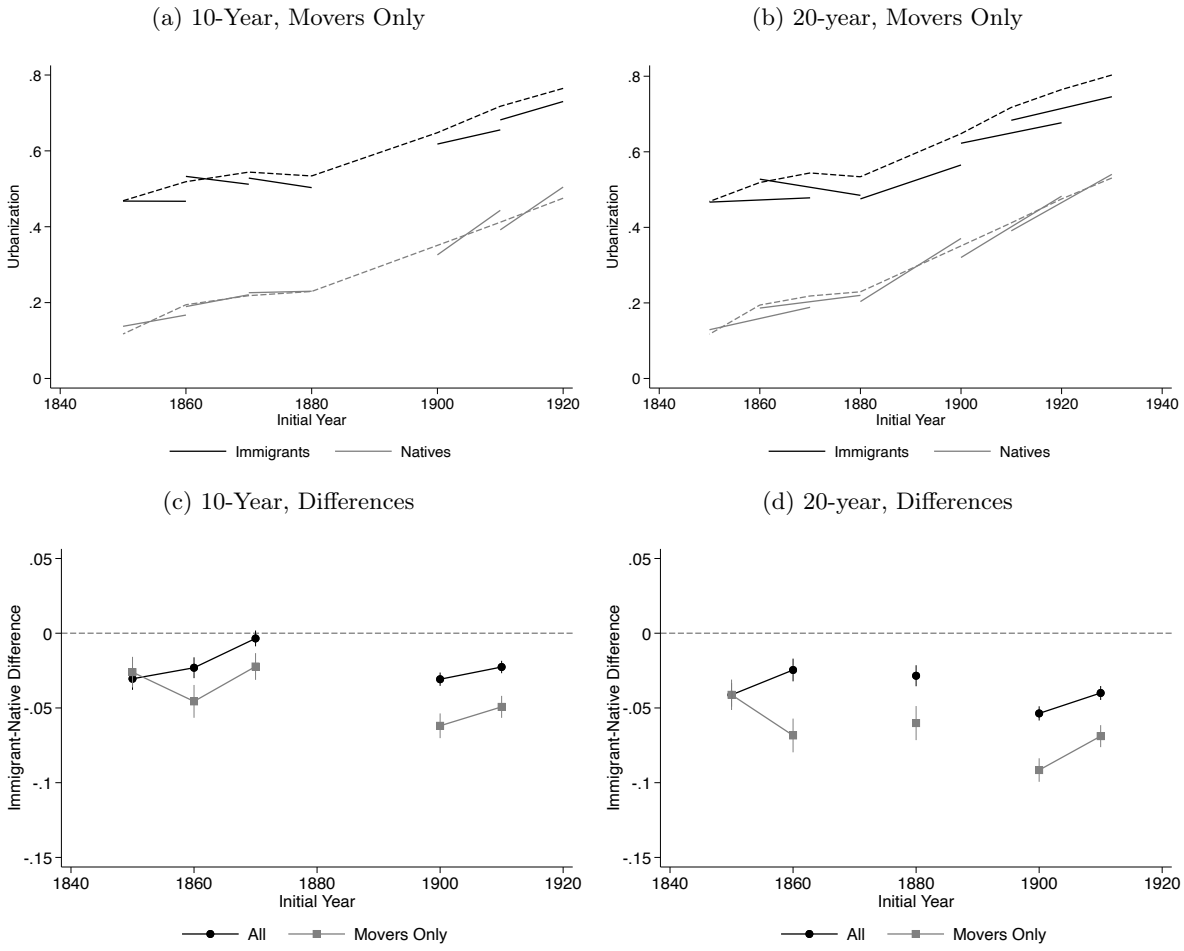
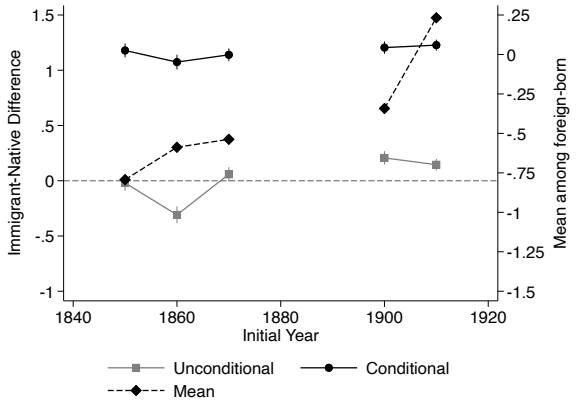


Figure I.6: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Samples are limited to movers only.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

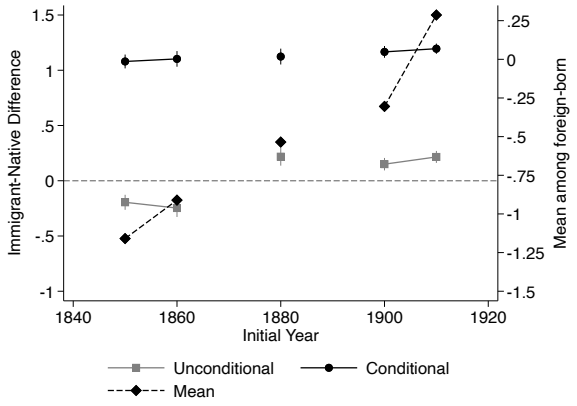


Figure I.7: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

J Results for Recent Immigrants

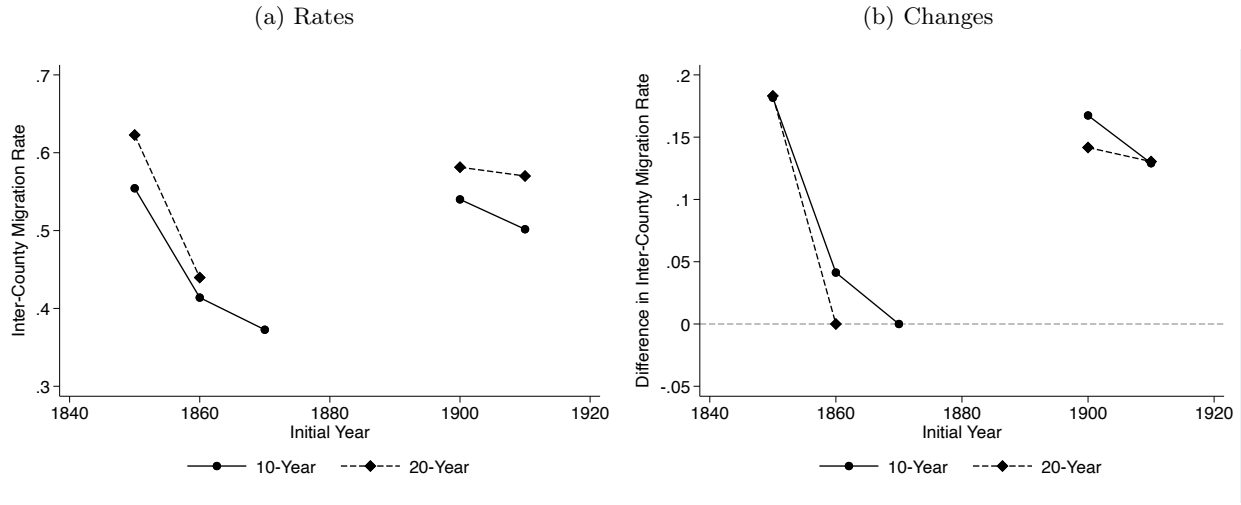


Figure J.1: Inter-county migration rates and changes, corrected for false matches

Note: Panel (a) shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

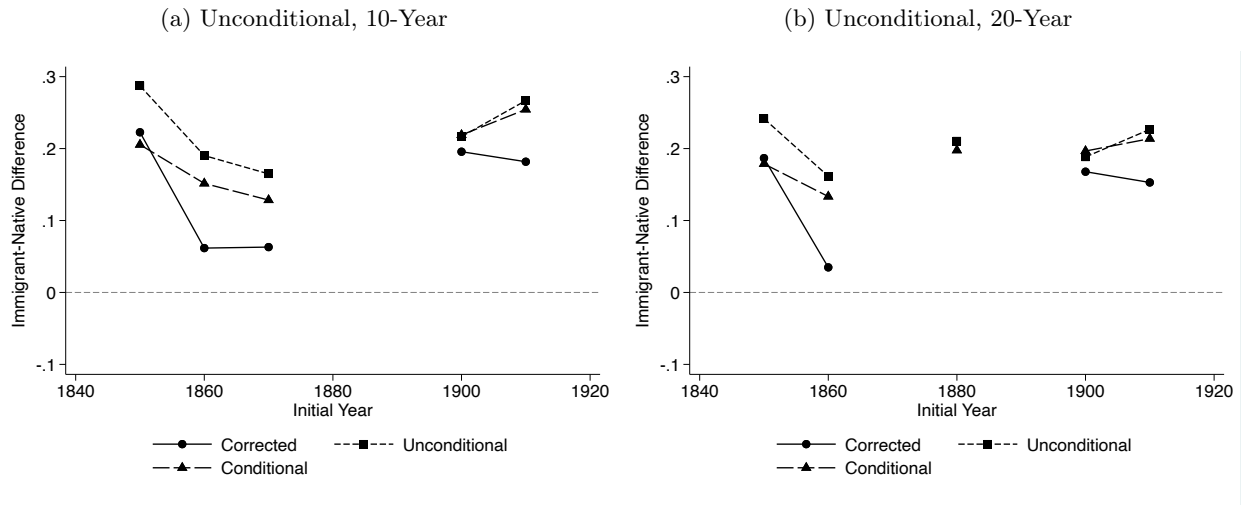


Figure J.2: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

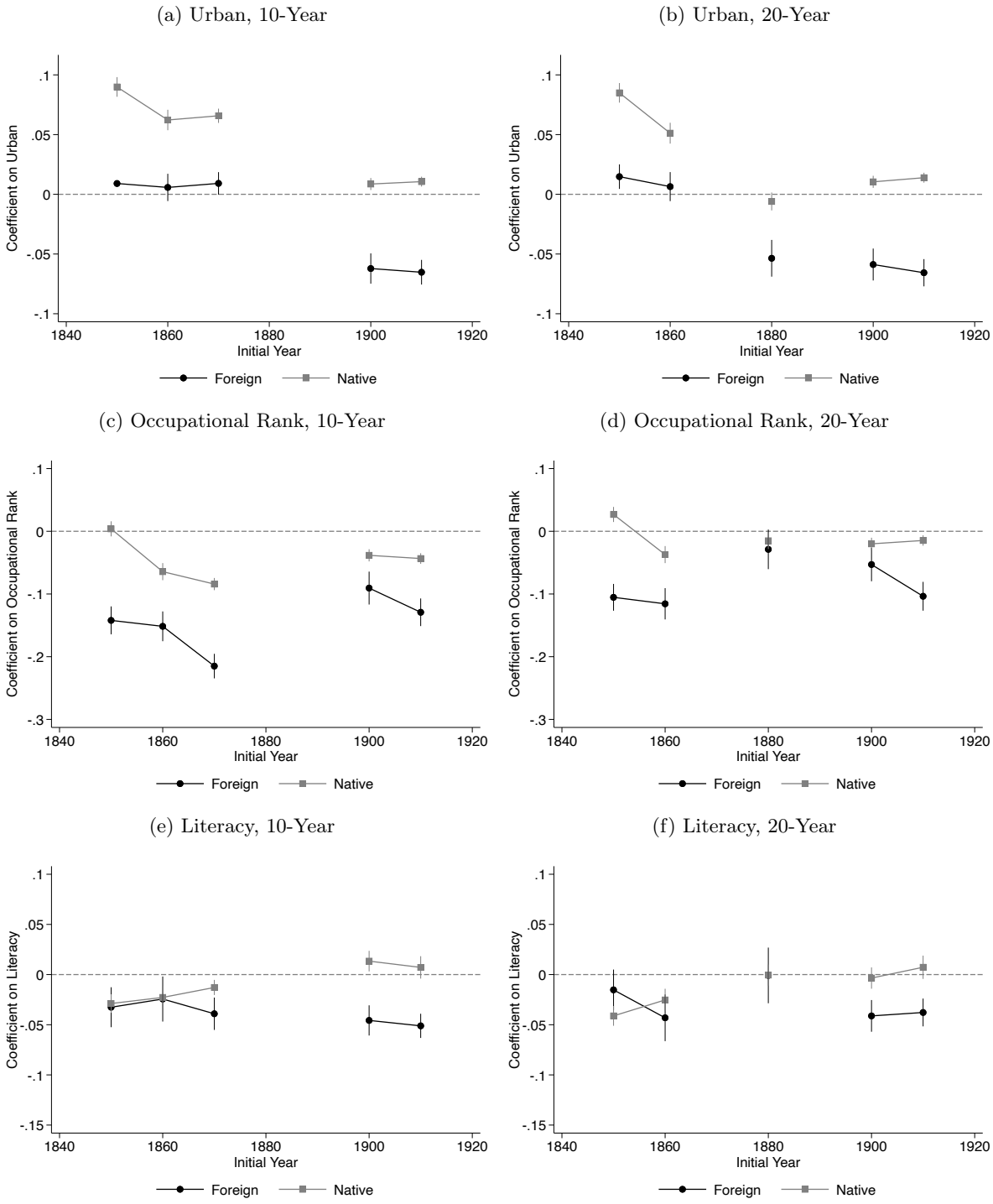


Figure J.3: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

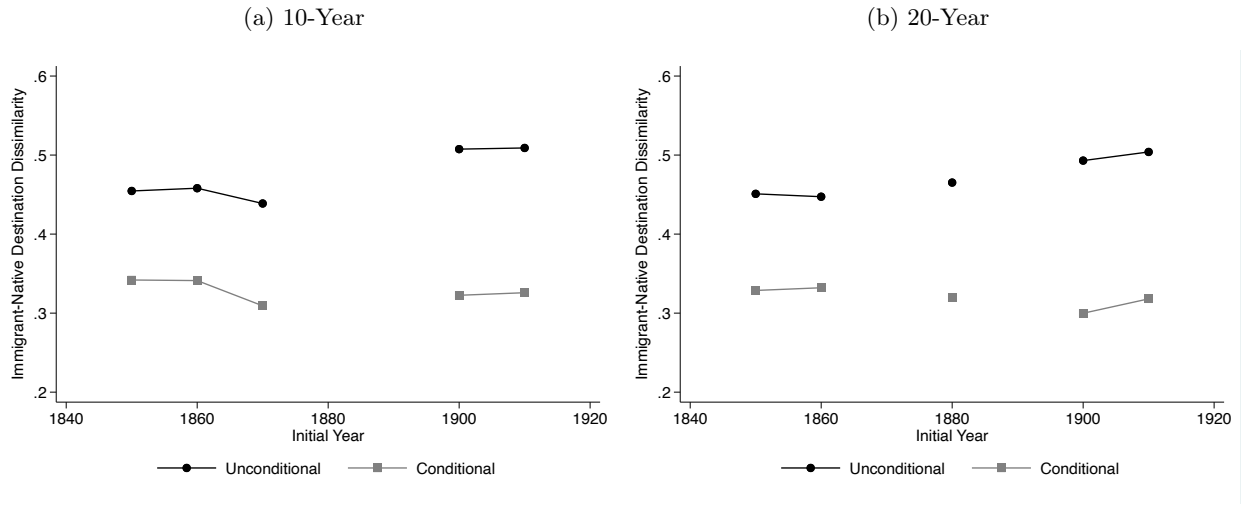
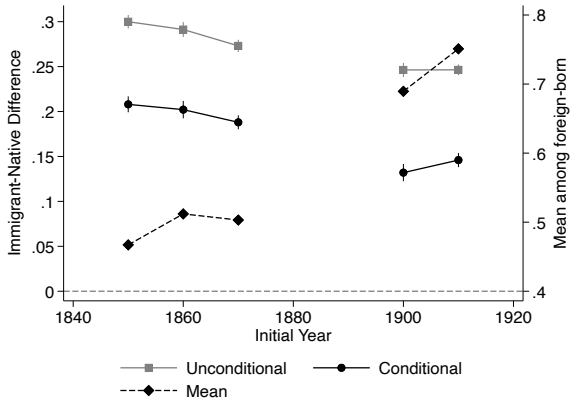


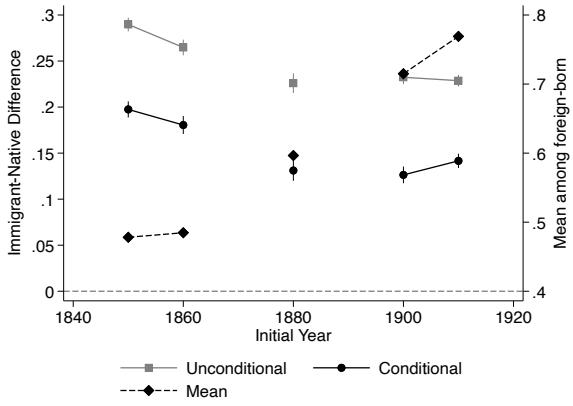
Figure J.4: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 45.

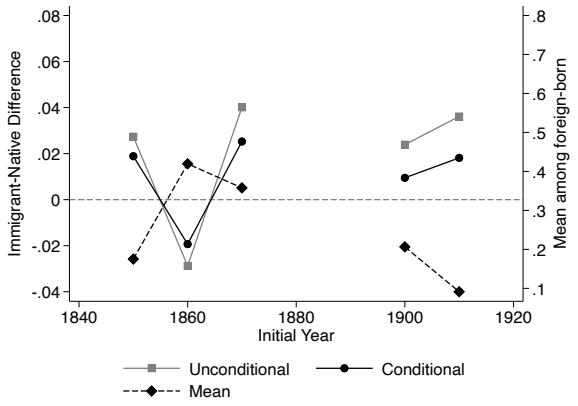
(a) Urban, 10-Year



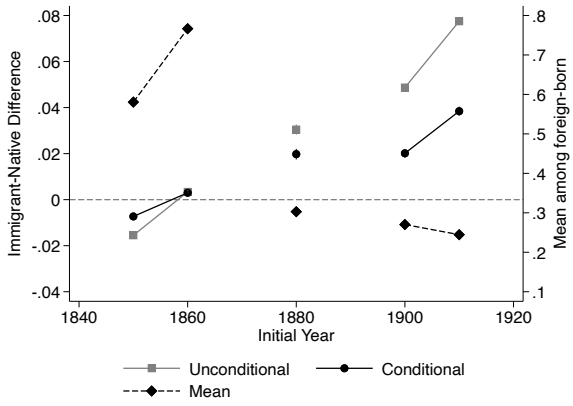
(b) Urban, 20-Year



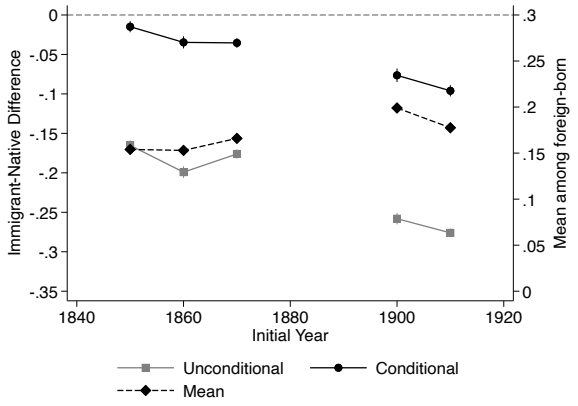
(c) Labor Demand Growth, 10-Year



(d) Labor Demand Growth, 20-Year



(e) South or West, 10-Year



(f) South or West, 20-Year

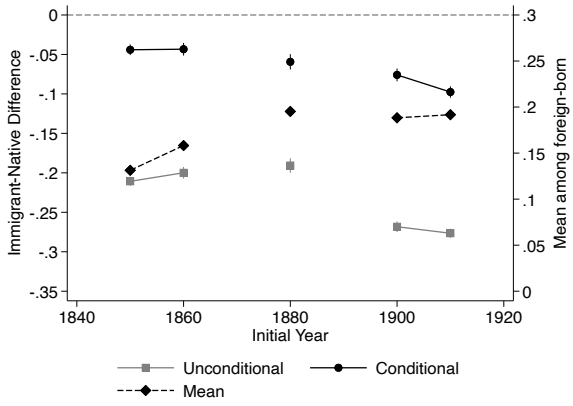
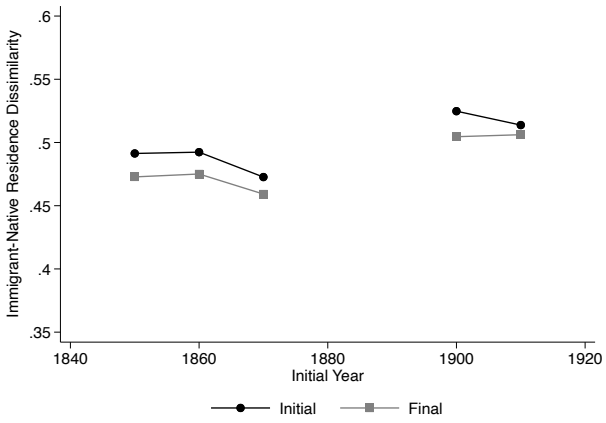


Figure J.5: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

(a) 10-Year



(b) 20-Year

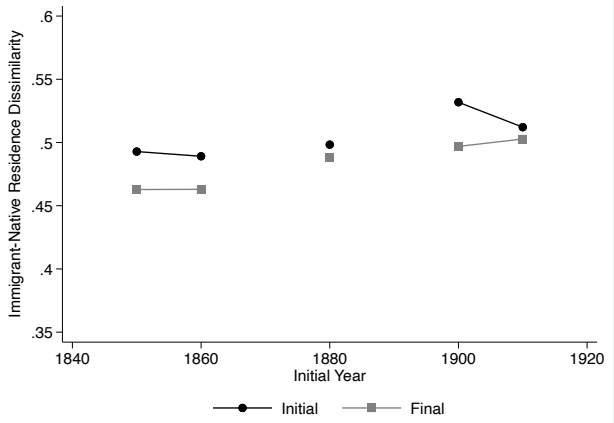


Figure J.6: Dissimilarity in counties of residence

Note: These figures present the dissimilarity index between the counties of residence of the foreign and native born in my linked samples in the initial and final year of the span beginning in the year on the x-axis.

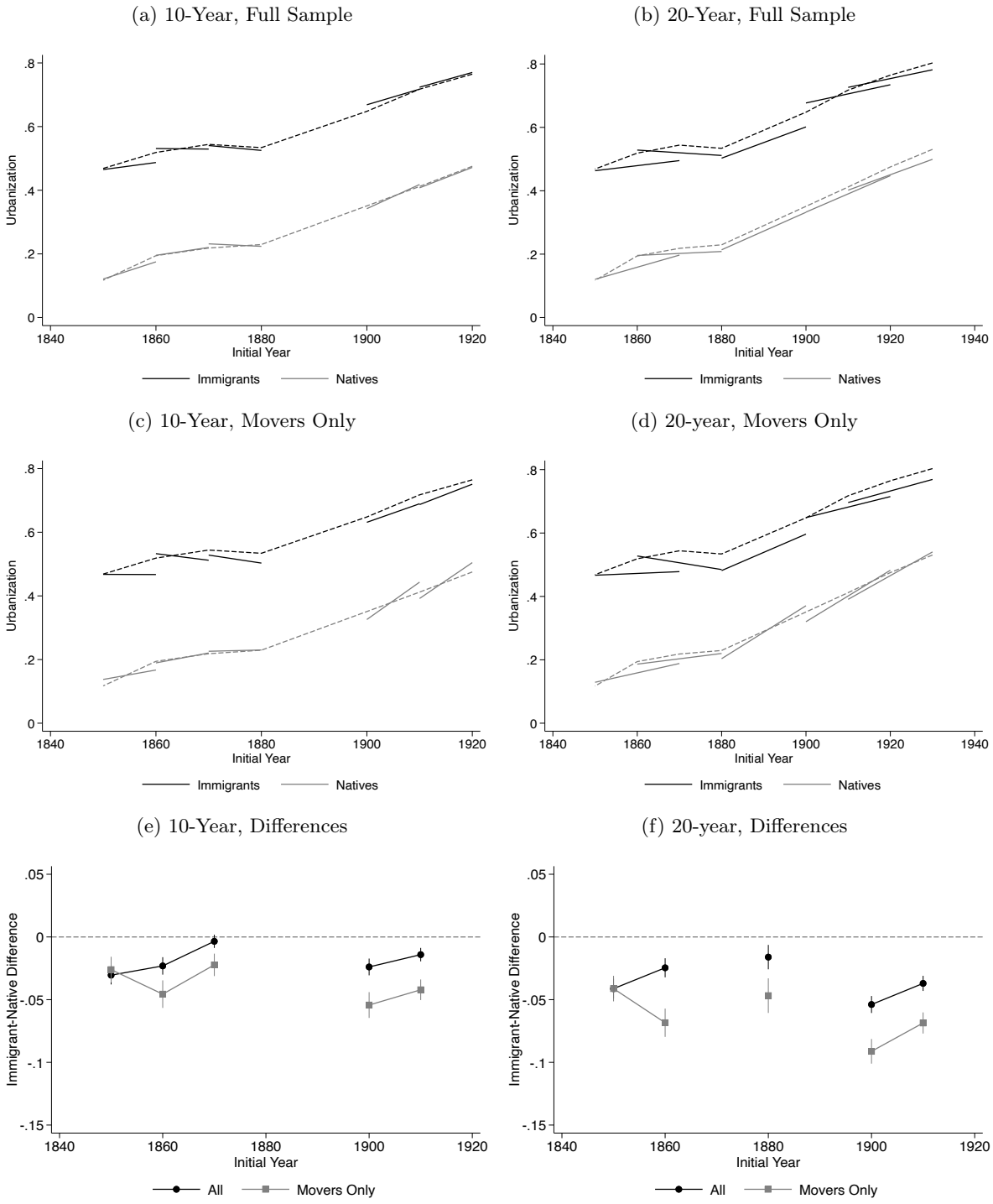
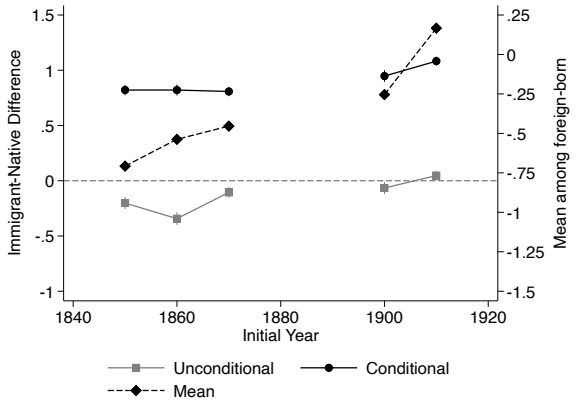


Figure J.7: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

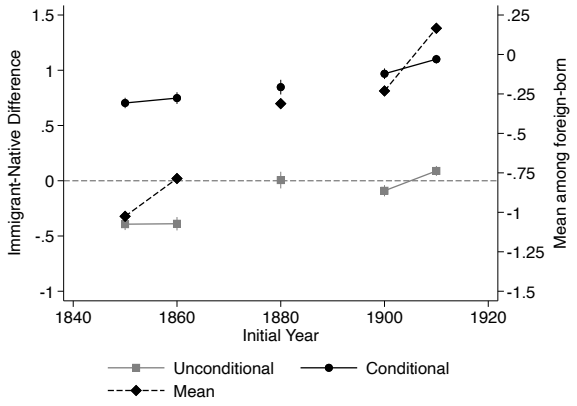


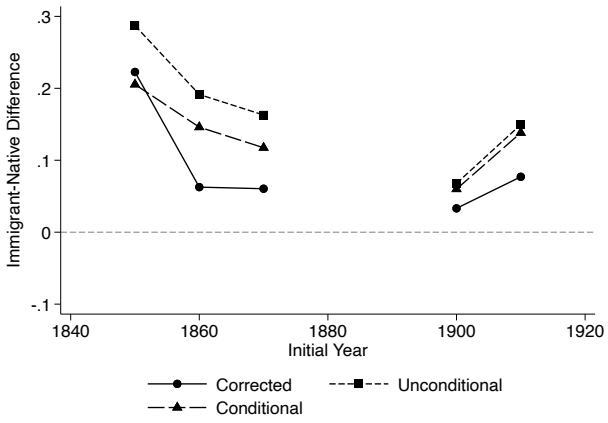
Figure J.8: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

K Results Dropping Second-Generation Immigrants

In this appendix, I drop natives whose father was an immigrant. Since this affects only the comparisons of immigrants and natives, I only replicate results that make such a comparison.

(a) Unconditional, 10-Year



(b) Unconditional, 20-Year

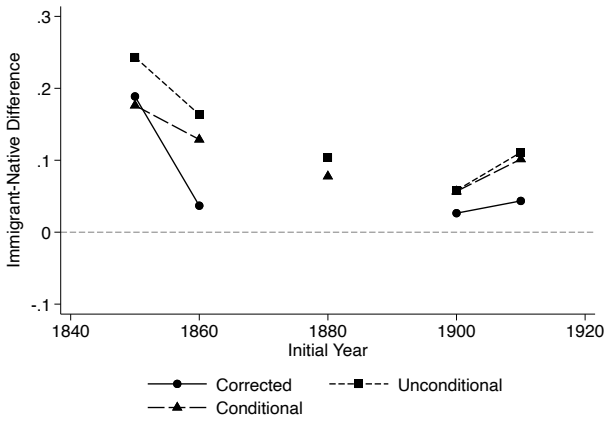


Figure K.1: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

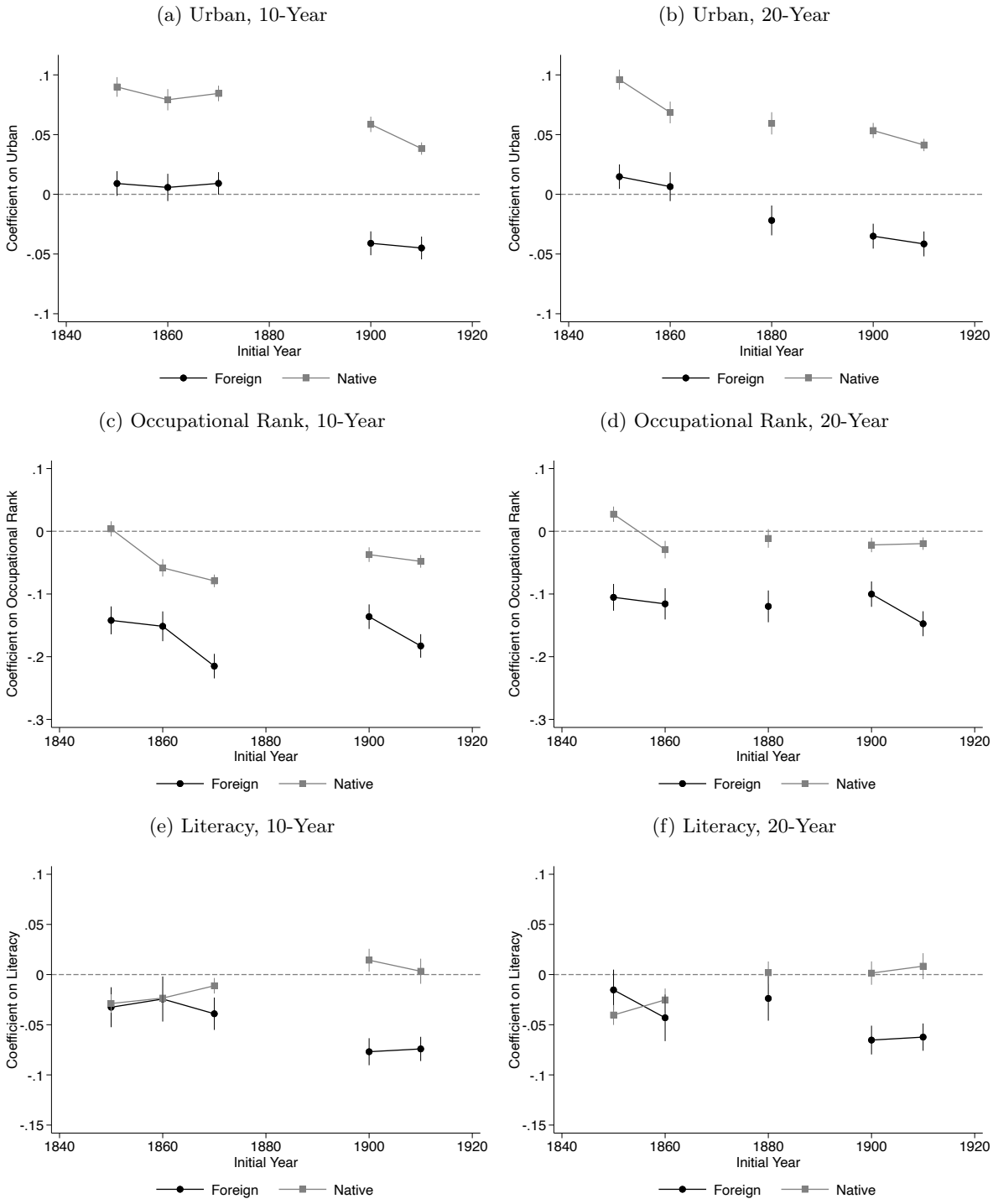


Figure K.2: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

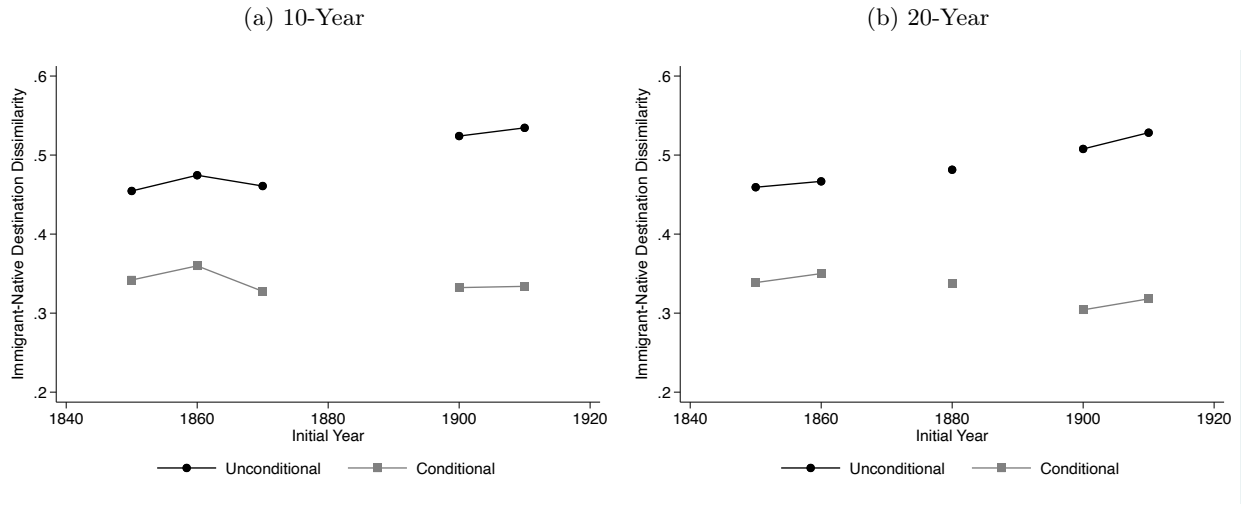
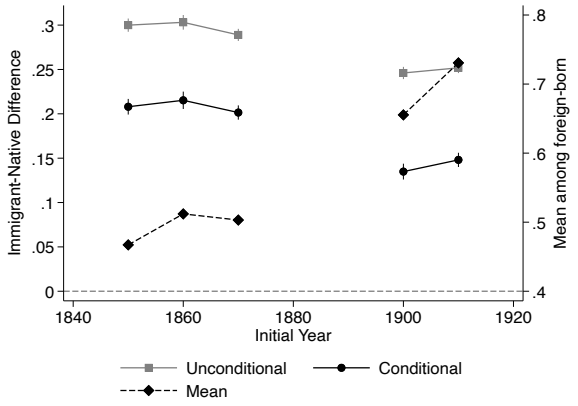


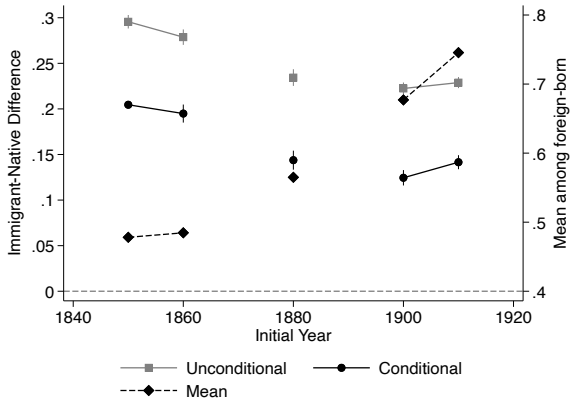
Figure K.3: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 45.

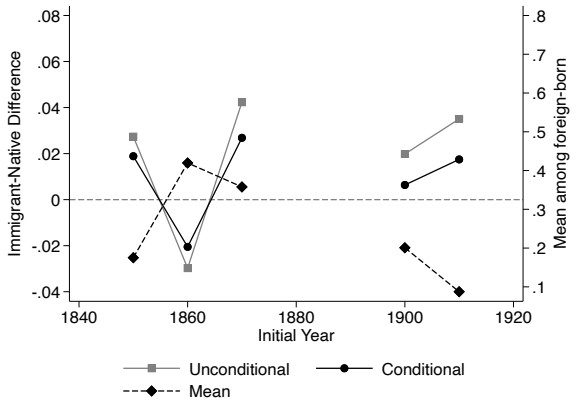
(a) Urban, 10-Year



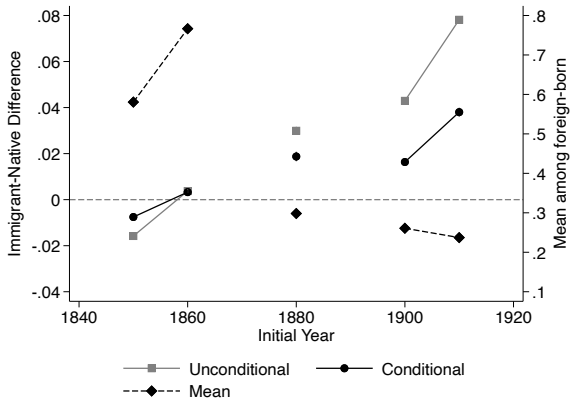
(b) Urban, 20-Year



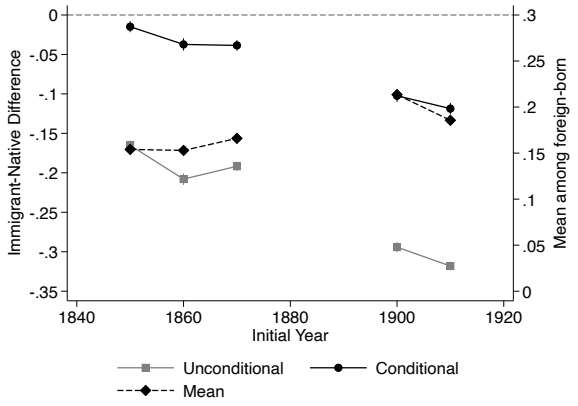
(c) Labor Demand Growth, 10-Year



(d) Labor Demand Growth, 20-Year



(e) South or West, 10-Year



(f) South or West, 20-Year

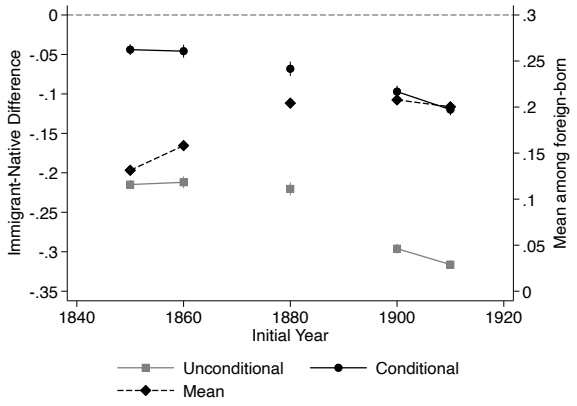


Figure K.4: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

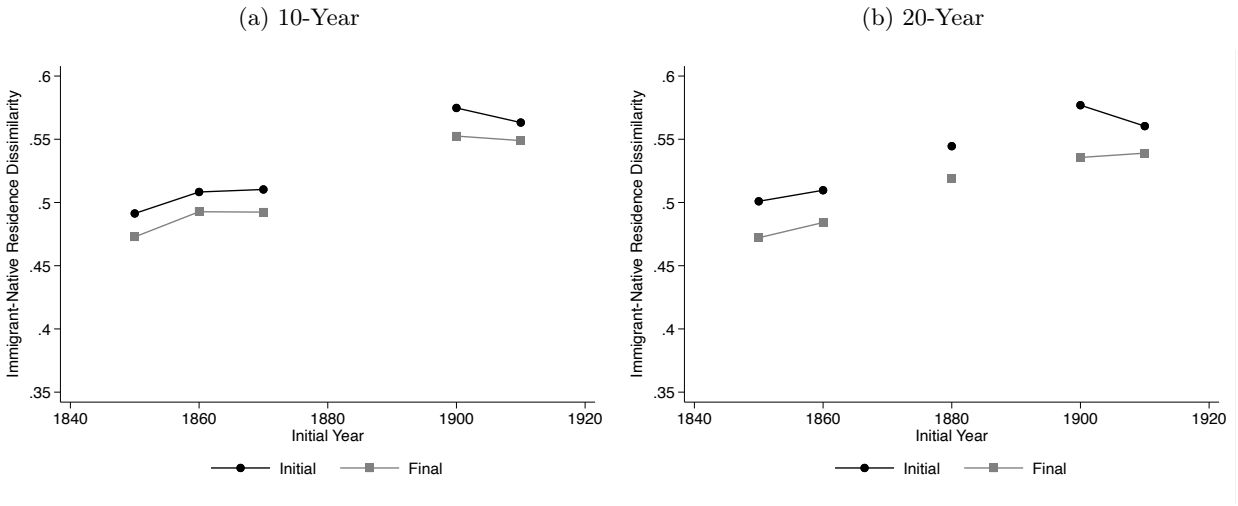


Figure K.5: Dissimilarity in counties of residence

Note: These figures present the dissimilarity index between the counties of residence of the foreign and native born in my linked samples in the initial and final year of the span beginning in the year on the x-axis.

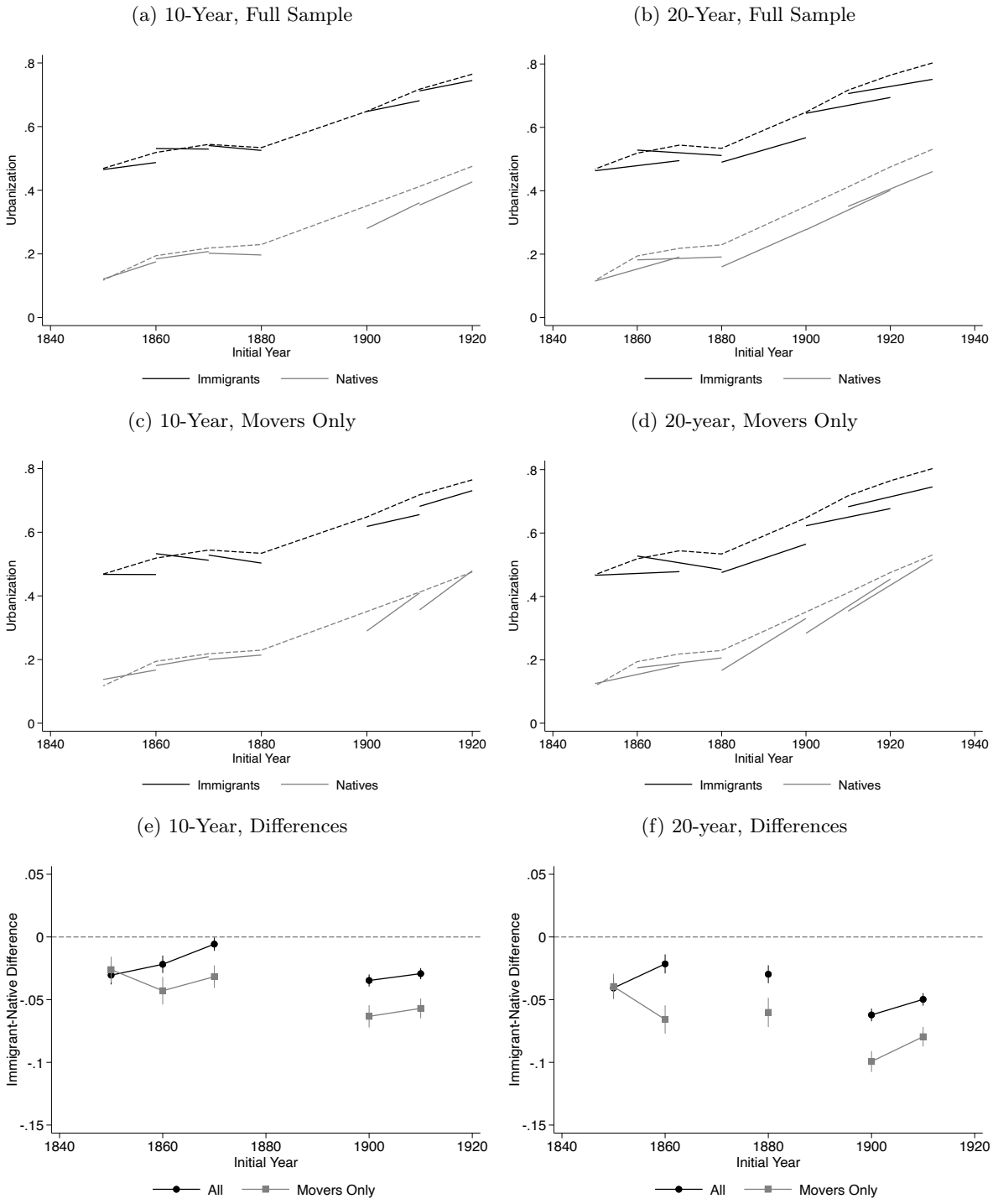
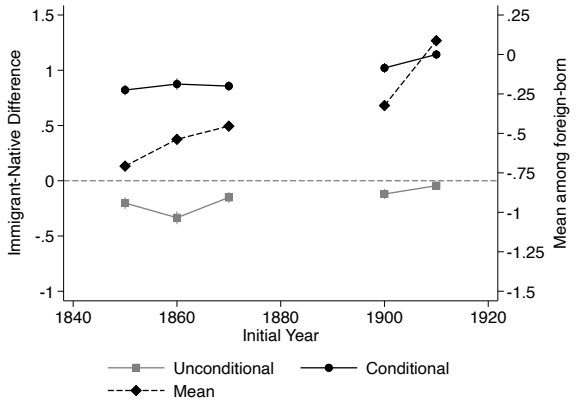


Figure K.6: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

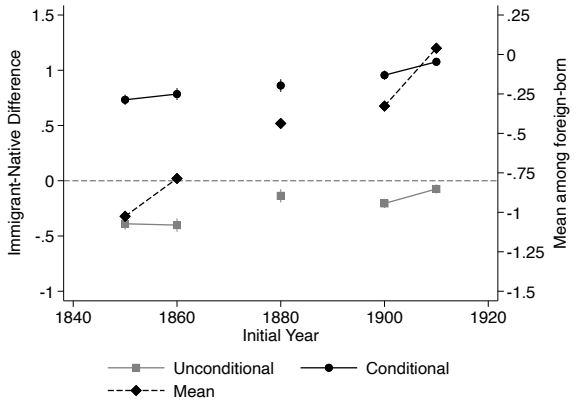


Figure K.7: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

L Results Dropping Southerners

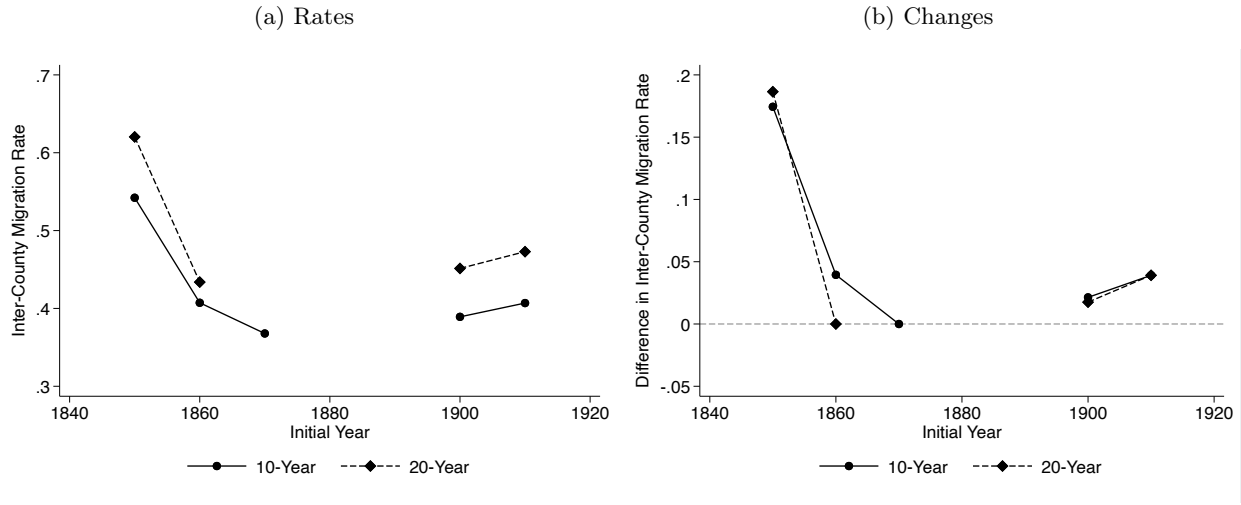


Figure L.1: Inter-county migration rates and changes, corrected for false matches

Note: Panel (a) shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

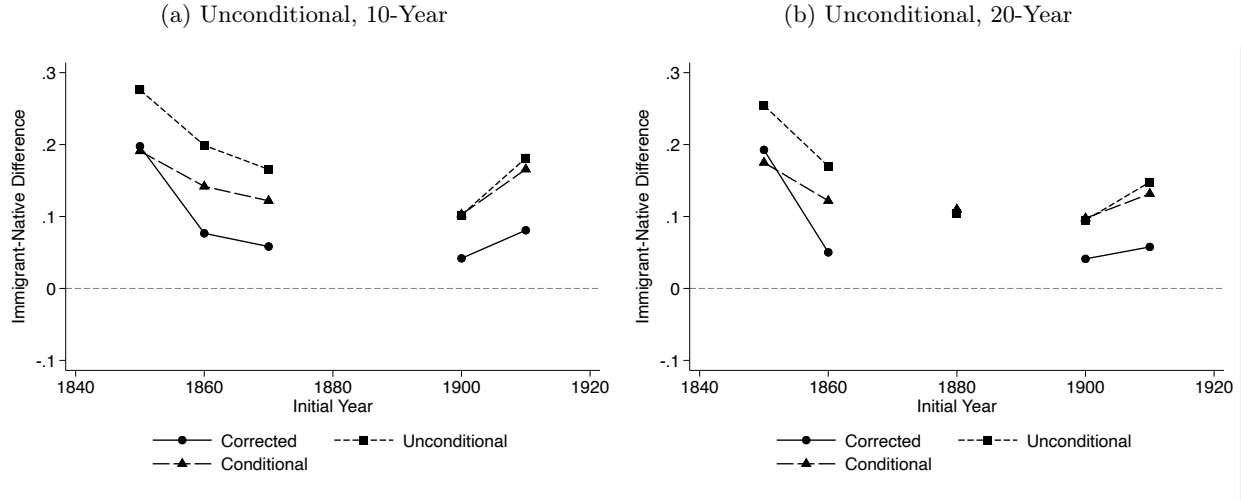
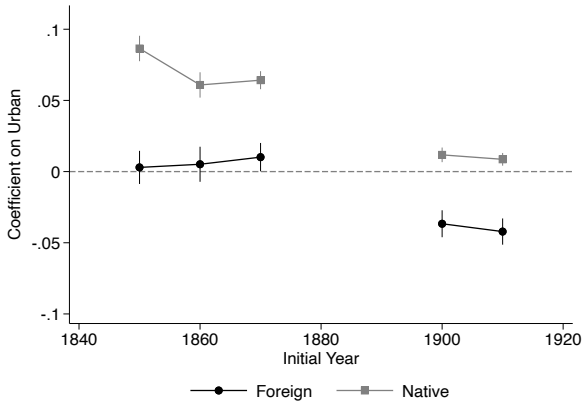


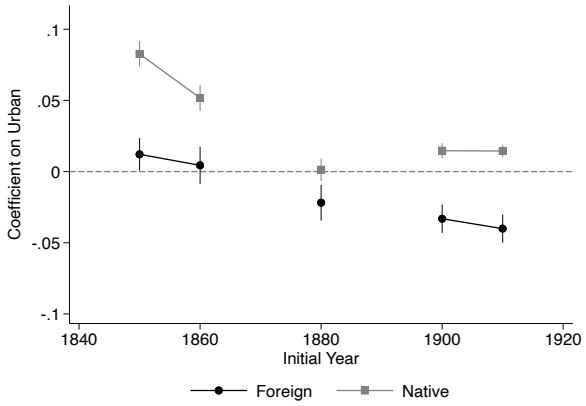
Figure L.2: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

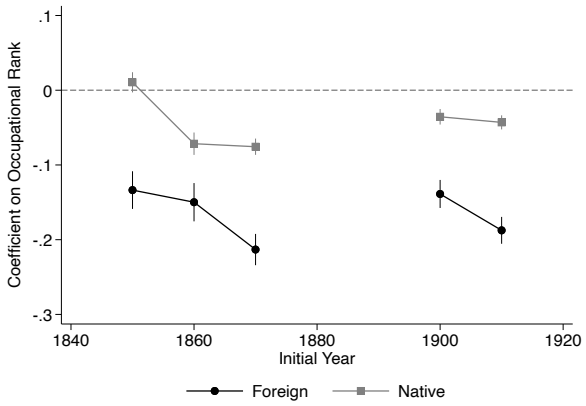
(a) Urban, 10-Year



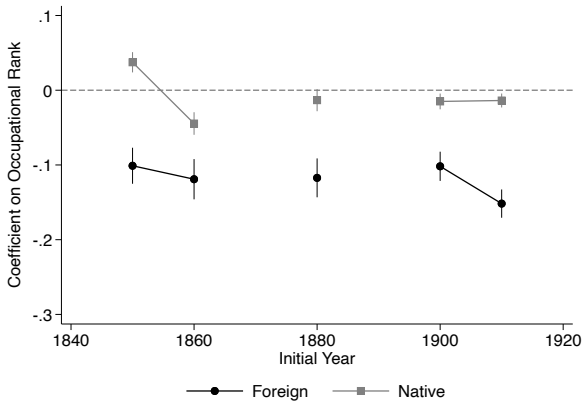
(b) Urban, 20-Year



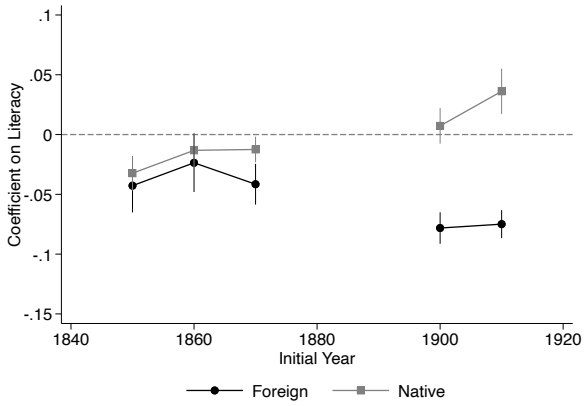
(c) Occupational Rank, 10-Year



(d) Occupational Rank, 20-Year



(e) Literacy, 10-Year



(f) Literacy, 20-Year

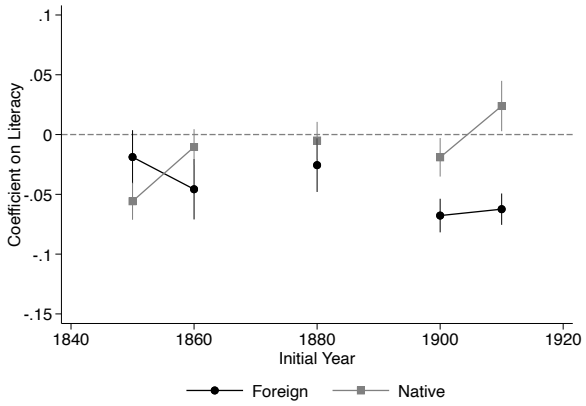


Figure L.3: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

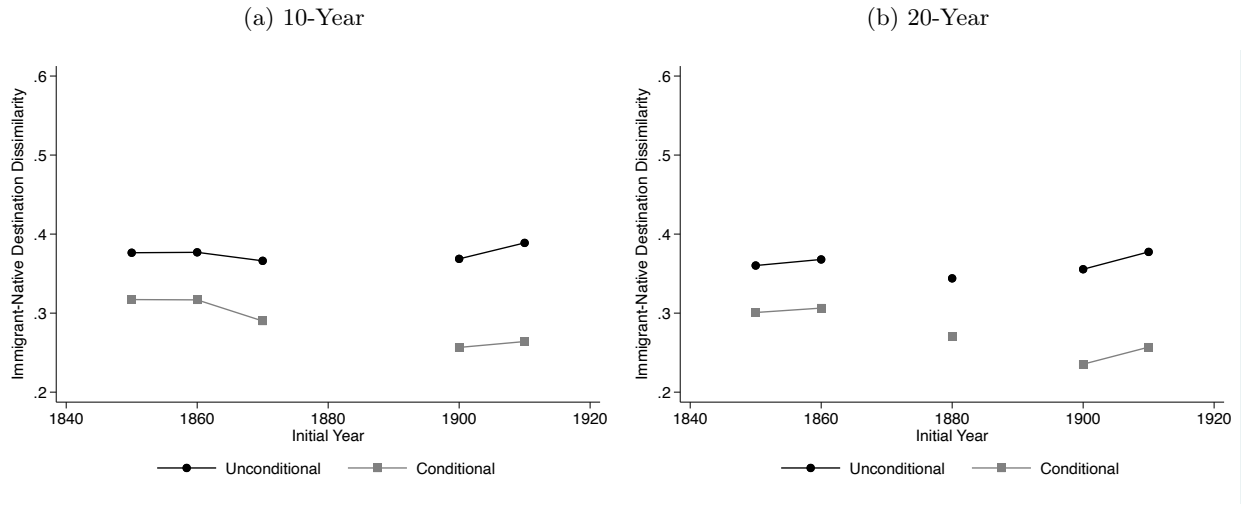
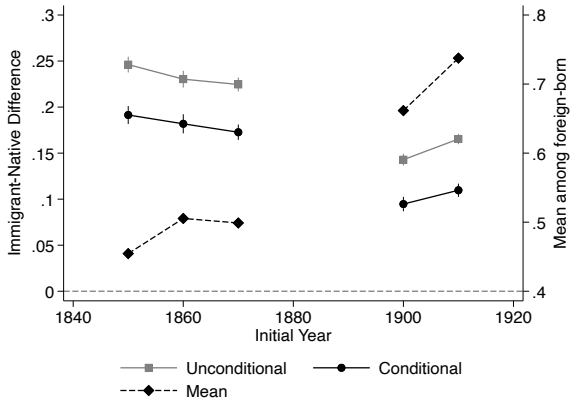


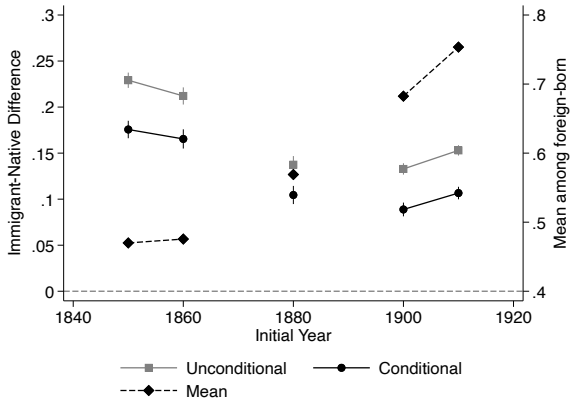
Figure L.4: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 45.

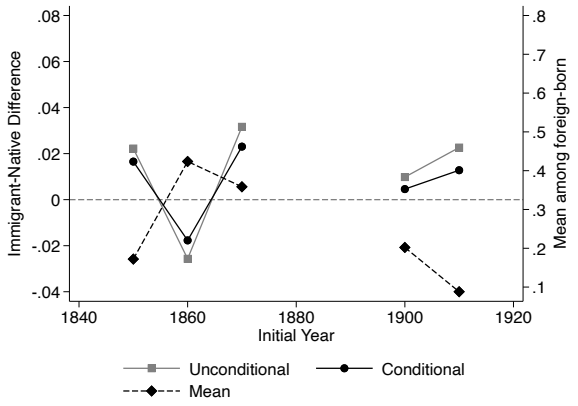
(a) Urban, 10-Year



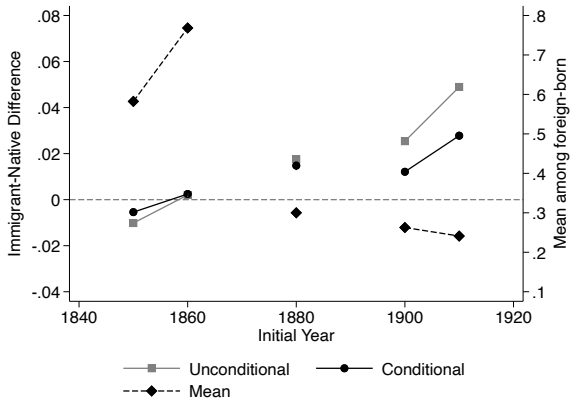
(b) Urban, 20-Year



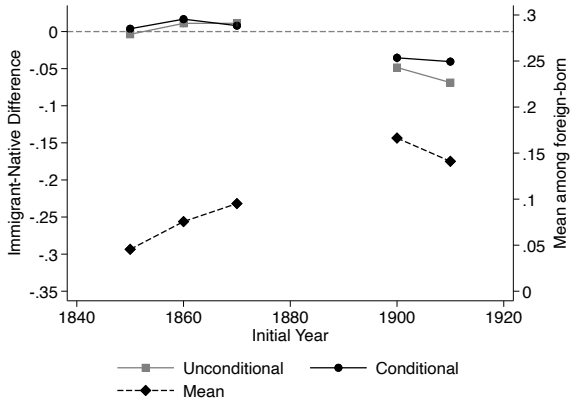
(c) Labor Demand Growth, 10-Year



(d) Labor Demand Growth, 20-Year



(e) South or West, 10-Year



(f) South or West, 20-Year

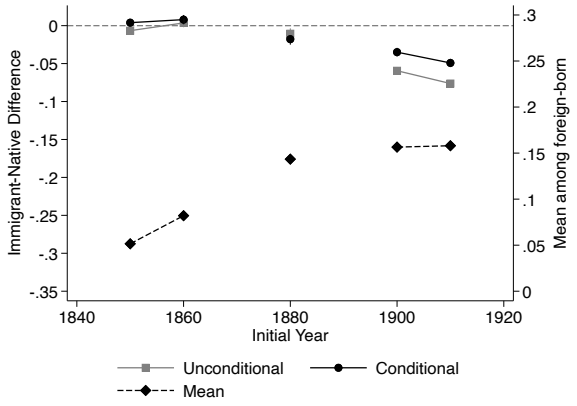


Figure L.5: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the destination characteristic for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are robust 95-percent confidence intervals.

(a) 10-Year

(b) 20-Year

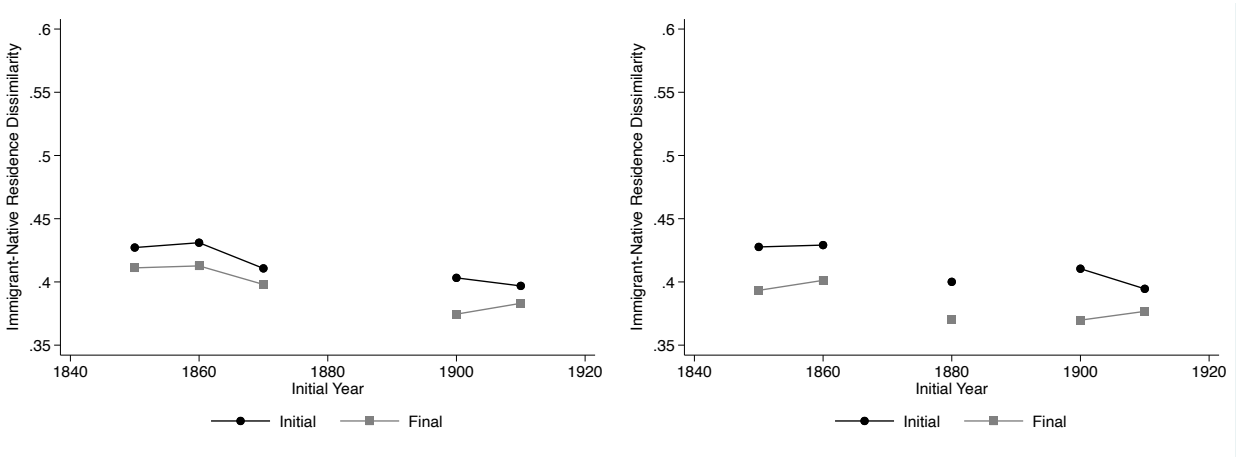


Figure L.6: Dissimilarity in counties of residence

Note: These figures present the dissimilarity index between the counties of residence of the foreign and native born in my linked samples in the initial and final year of the span beginning in the year on the x-axis.

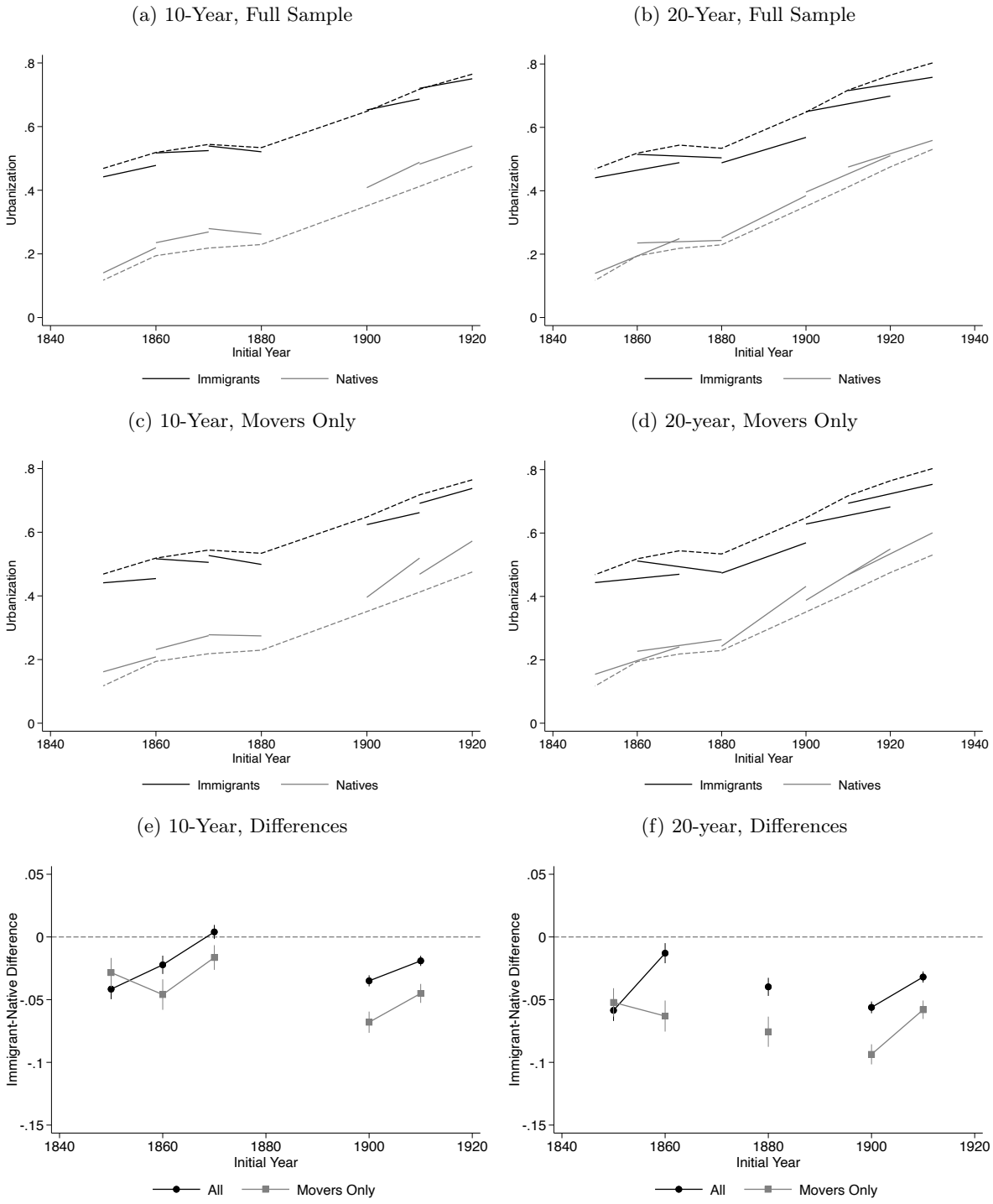
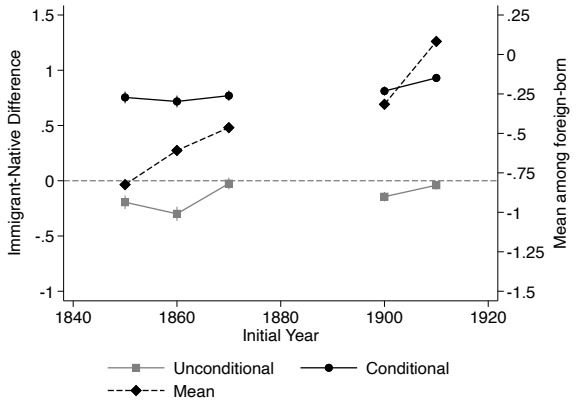


Figure L.7: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. Each line represents a cohort, and links its initial urbanization to its final. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

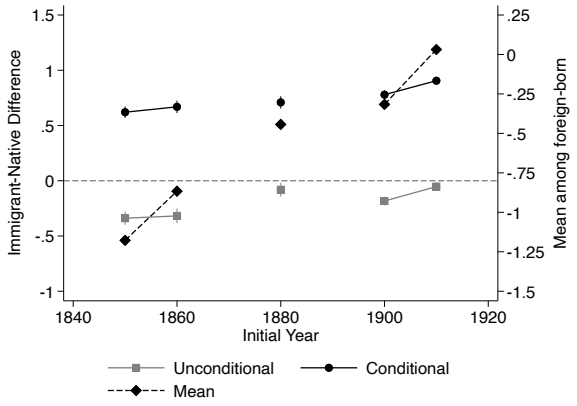


Figure L.8: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

References

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, Santiago Pérez, and Myera Rashid (2020). *Census Linking Project: Version 2.0 [dataset]*. `https://censuslinkingproject.org`.

Collins, William J. and Ariell Zimran (2021). "Working Their Way Up? US Immigrants' Changing Labor Market Assimilation in the Age of Mass Migration." NBER Working Paper 26414.

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles (2019). *IPUMS National Historical Geographic Information System: Version 14.0* [database]. Minneapolis: IPUMS.

Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek (2021). *Integrated Public Use Microdata Series: Version 11.0* [dataset]. Minneapolis: University of Minnesota.

Zimran, Ariell (2021). "Internal Migration in the United States: Rates, Selection, and Sorting, 1850–1940." Mimeo., Vanderbilt University.