Online Appendix for

Internal Migration in the United States
Rates, Selection, and Destination Choice, 1850–1940

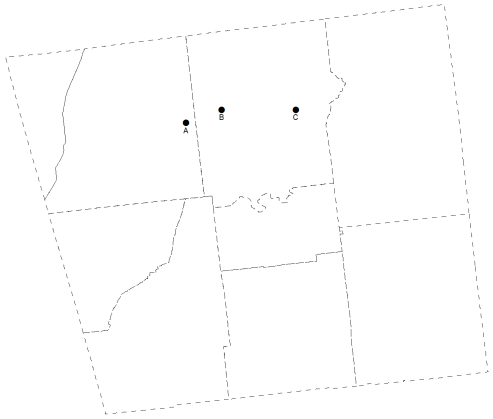
Ariell Zimran
Vanderbilt University and NBER

August 15, 2022

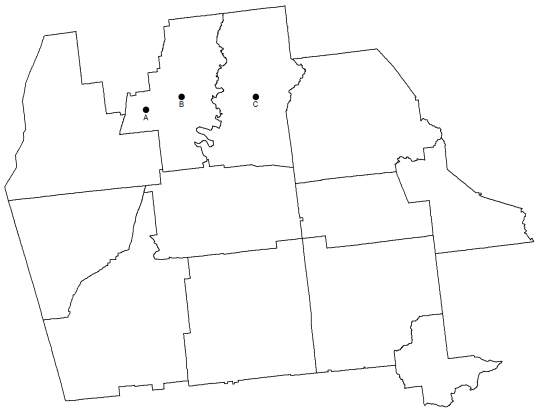
A. Additional tables and figures . . . 42
B. Results with imputed occupational codes . . . 52
C. Results for longer-distance moves . . . 58
D. Selection and sorting results with alternative linkage methods . . . 64
  D.1. ABE-Exact . . . 64
  D.2. ABE-NYSIIS . . . 69
  D.3. Intersection of matches . . . 74
  D.4. Intersection of matches plus corroboration . . . 79
E. Results dropping second-generation immigrants . . . 84

A Additional tables and figures

(a) 1850



(b) 1860



(c) Changing boundaries

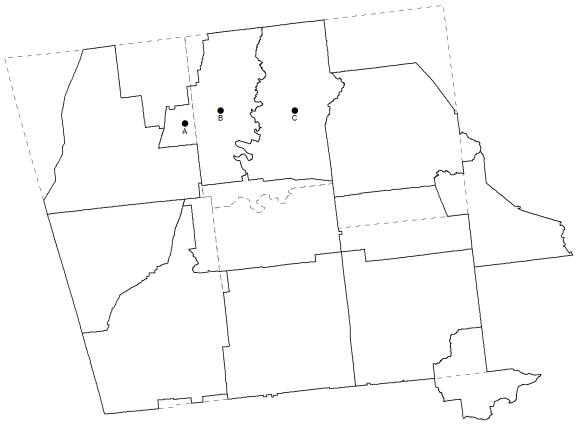


Figure A.1: Definition of inter-county migration

Note: Each map shows a group of counties in northwestern Georgia that experienced boundary changes 1850–1860. Points A and B are in the same county in 1860 but not in 1850. Points B and C are in the same county in 1860 but not 1850. Points A and C are never in the same county. Only moves between points A and C are classified as an inter-county move.

Table A.1: Linkage rates

	Natives	
	(1)	(2)
Link	Start	Linked
1850–1860	2,980,784	384,687
		(0.129)
1850–1870	2,980,784	347,741
		(0.117)
1860–1870	3,840,785	383,199
		(0.100)
1860–1880	3,840,785	374,981
		(0.098)
1870–1880	4,488,154	575,268
		(0.128)
1880–1900	6,527,283	776,295
		(0.119)
1900–1910	10,198,622	1,549,020
		(0.152)
1900–1920	10,198,622	1,450,093
		(0.142)
1910–1920	12,552,719	2,196,851
		(0.175)
1910–1930	12,552,719	2,061,328
		(0.164)
1920–1930	14,338,600	2,958,097
		(0.206)
1920–1940	14,338,600	2,646,060
		(0.185)
1930–1940	17,366,105	3,843,088
		(0.221)

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

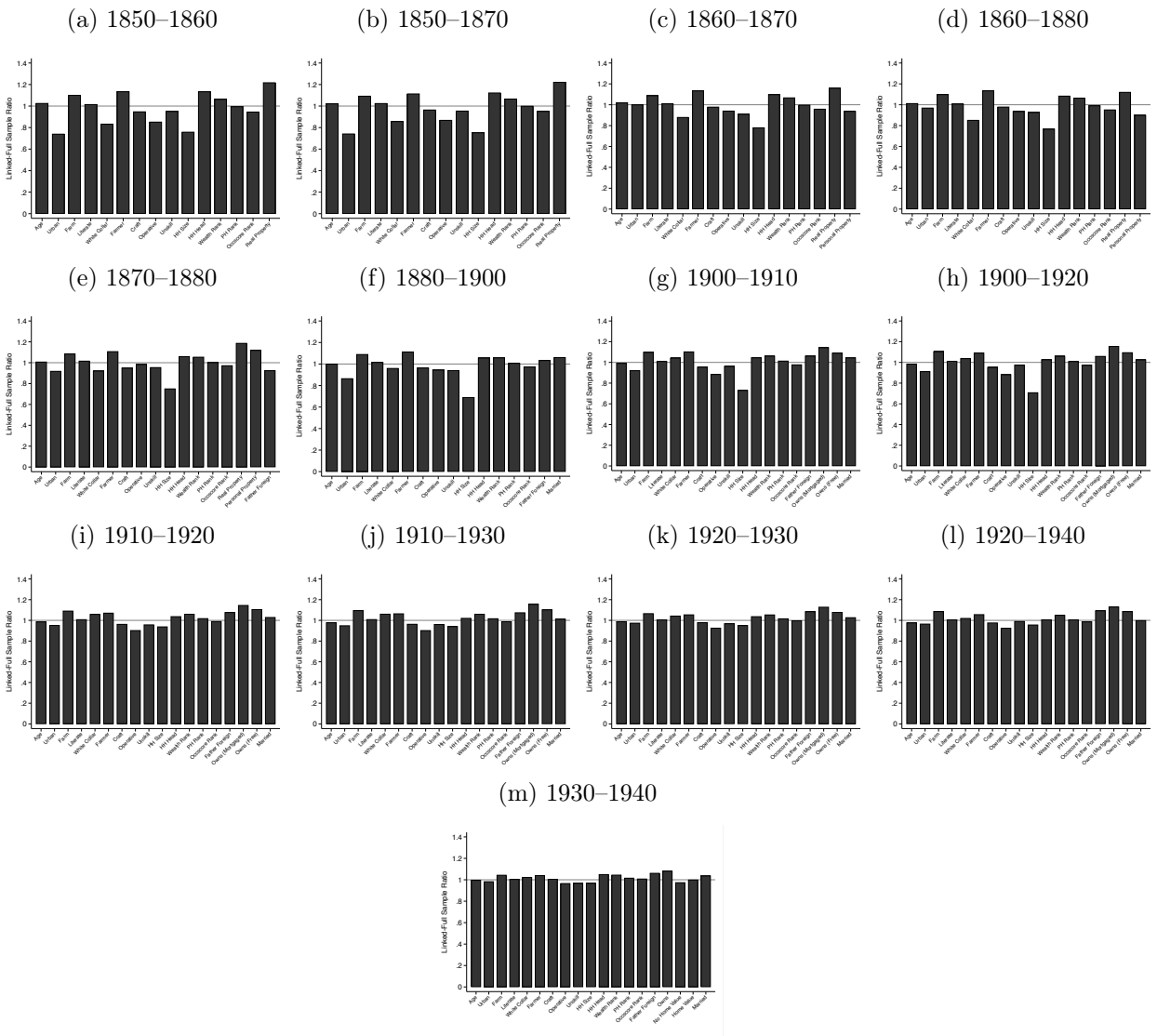


Figure A.2: Representativeness of the linked sample

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.

Table A.2: Covariates available in each initial census

Variable	1850	1860	1870	1880	1900	1910	1920	1930
Age	X	X	X	X	X	X	X	X
Urban	X	X	X	X	X	X	X	X
City Population†	X	X	X	X	X	X	X	X
Farm	X	X	X	X	X	X	X	X
Literacy	X	X	X	X	X	X	X	X
Marital status				X	X	X	X	X
Occupation*	X	X	X	X	X	X	X	X
Household size	X	X	X	X	X	X	X	X
Household head	X	X	X	X	X	X	X	X
Birthplace	X	X	X	X	X	X	X	X
Place of residence	X	X	X	X	X	X	X	X
Real property	X	X	X					
Personal property		X	X					
Father's nativity				X	X	X	X	X
Owns home (mortgaged)					X	X	X	
Owns home (free and clear)					X	X	X	
Owns home								X
Value of home								X

†: used only to create measures of urbanization
*: used to create occupational categories, occupational rank measures, and the average occupational rank measure, all of which are available for all initial years.

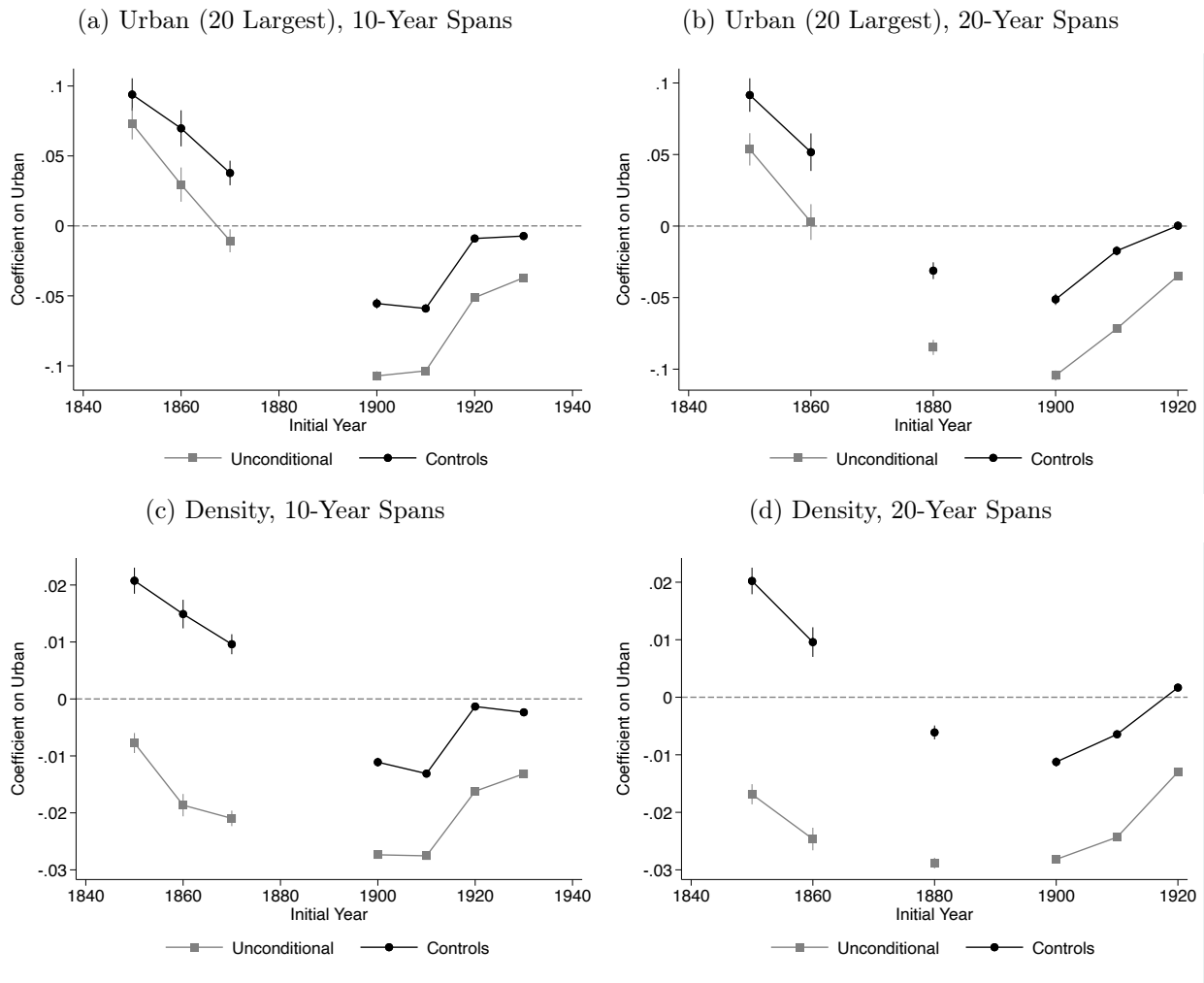
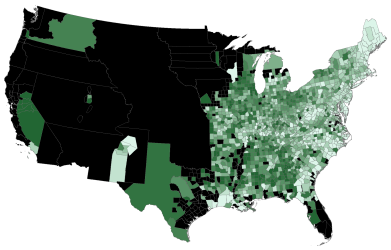


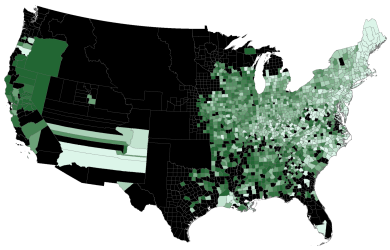
Figure A.3: Migrant selection by measure and span, alternate urban definitions

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as state fixed effects.

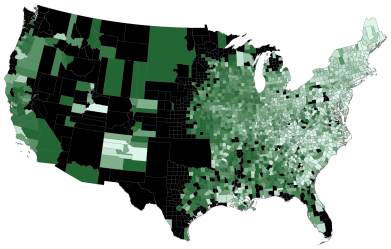
(a) 1850–1860



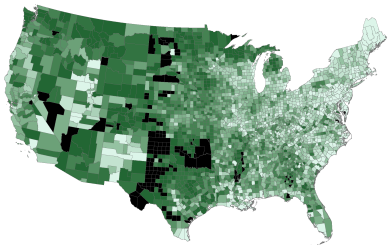
(b) 1860–1870



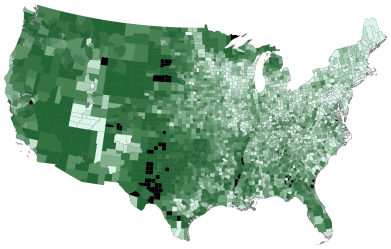
(c) 1870–1880



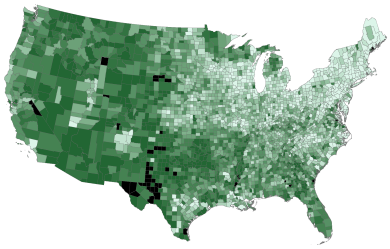
(d) 1900–1910



(e) 1910–1920



(f) 1920–1930



(g) 1930–1940

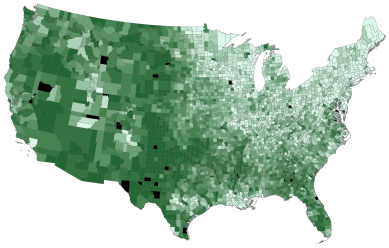
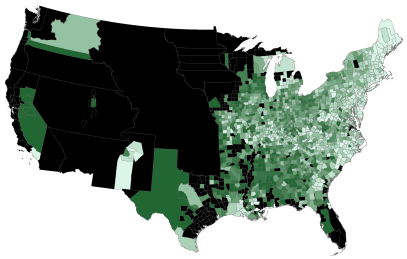


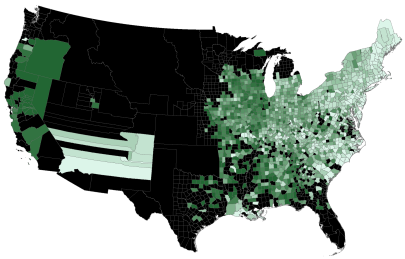
Figure A.4: County-level emigration rates, 10-year spans

Note: These are county-level emigration rates (i.e., the probability that an individual initially observed in that county would be observed in another county 10 years later) in the linked sample, weighted to correct for selection into linkage. Counties with fewer than 30 observations in the linked sample are dropped, and shown only in black. The scale is different in each year. It is based on deciles of migration rates across counties, with darker colors indicating higher migration rates.

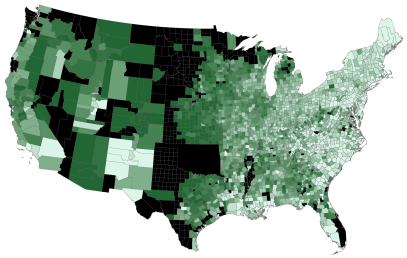
(a) 1850–1870



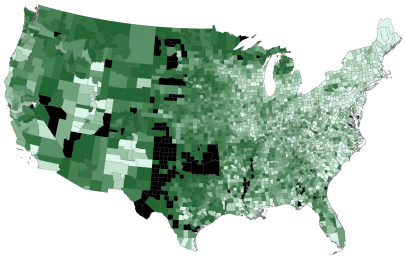
(b) 1860–1880



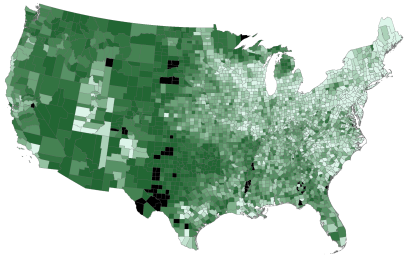
(c) 1880–1900



(d) 1900–1920



(e) 1910–1930



(f) 1920–1940

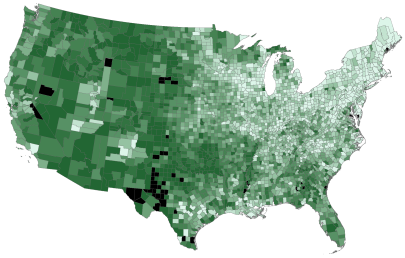
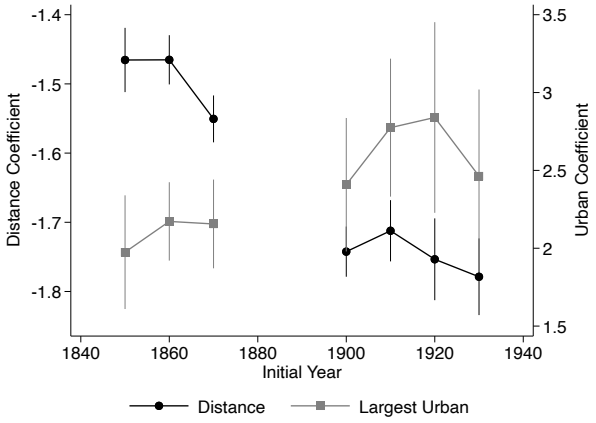


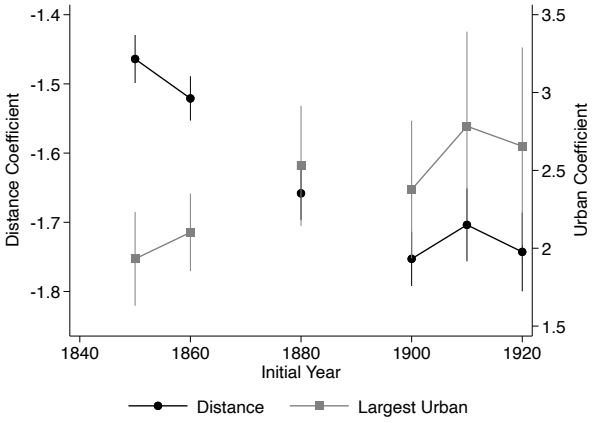
Figure A.5: County-level emigration rates, 20-year spans

Note: These are county-level emigration rates (i.e., the probability that an individual initially observed in that county would be observed in another county 20 years later) in the linked sample, weighted to correct for selection into linkage. Counties with fewer than 30 observations in the linked sample are dropped, and shown only in black. The scale is different in each year. It is based on deciles of migration rates across counties, with darker colors indicating higher migration rates.

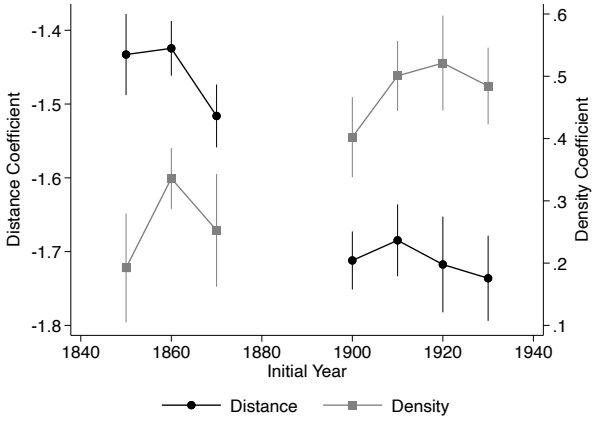
(a) Urban (20 Largest), 10-Year



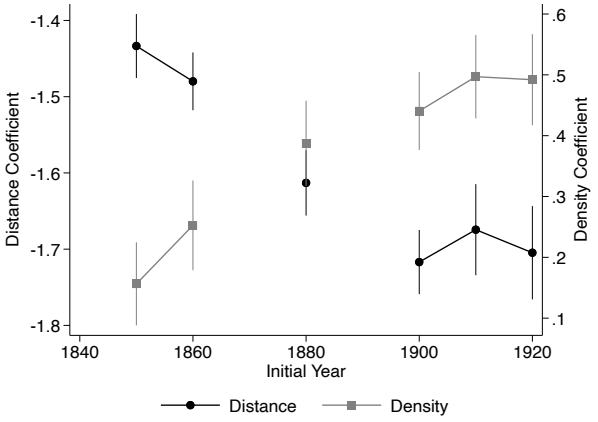
(b) Urban (20 Largest), 20-Year



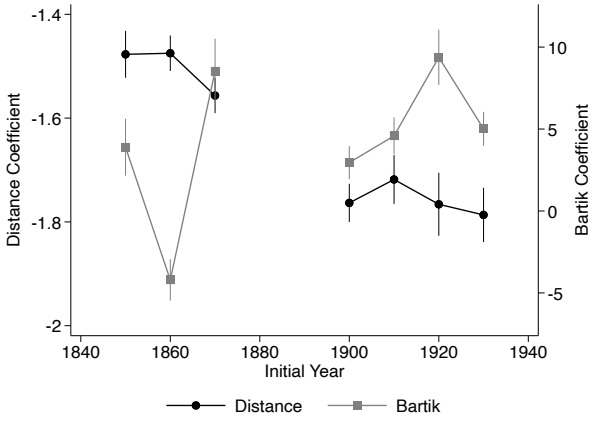
(c) Density, 10-Year



(d) Density, 20-Year



(e) Bartik, 10-Year



(f) Bartik, 20-Year

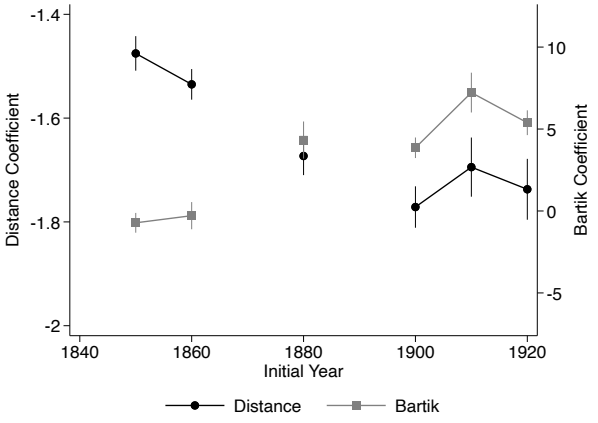


Figure A.6: Conditional logit results

Note: This figure presents coefficients on distance and the measure of urban residence or labor demand growth from the conditional logit estimation.

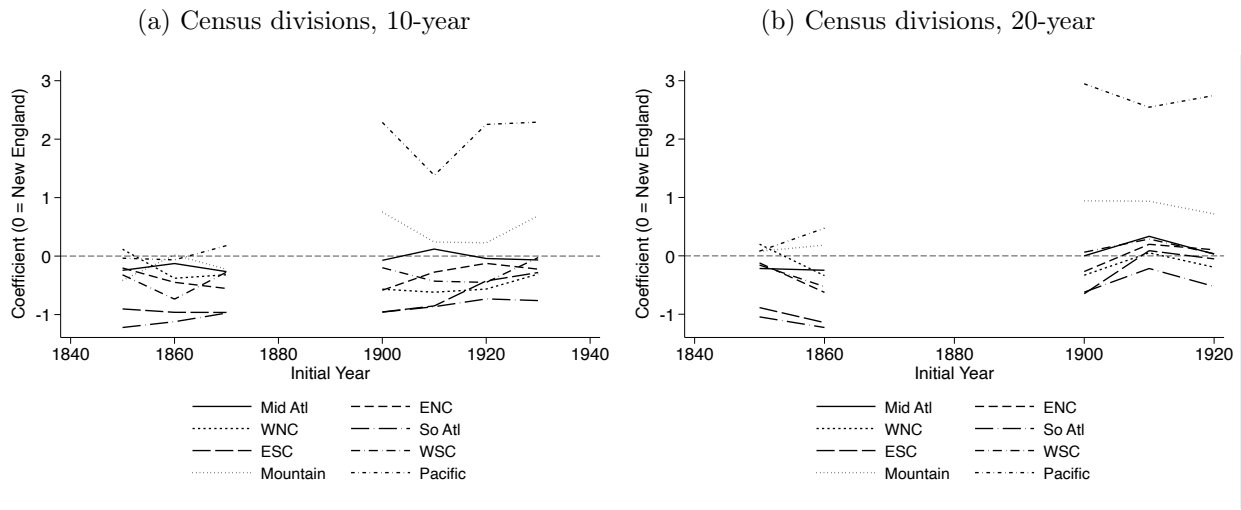


Figure A.7: Region indicators from conditional logit with labor demand growth

Note: These panels present the coefficients on the census division indicators, with New England excluded, controlling for labor demand growth.

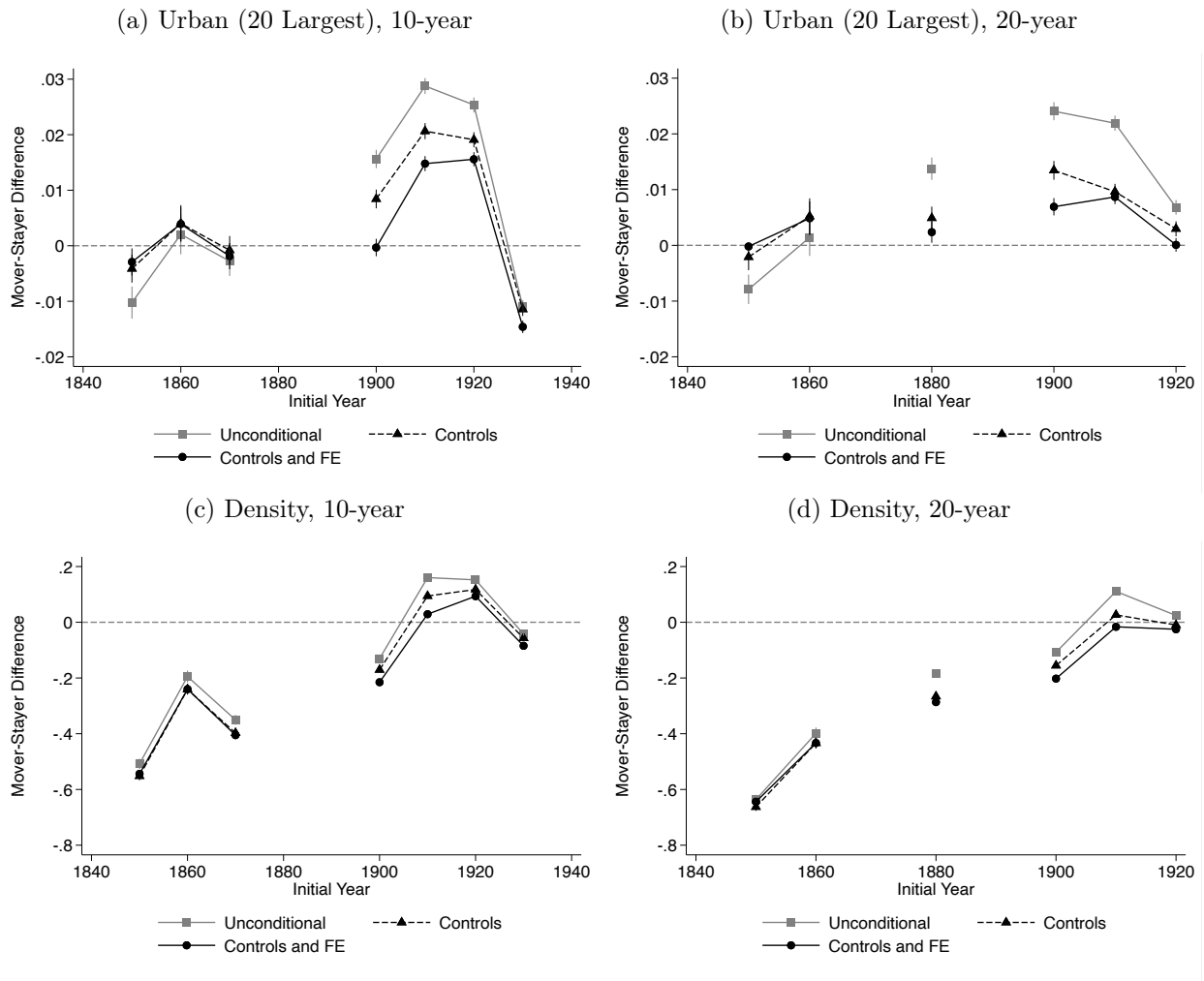


Figure A.8: Movers' change in urbanization and Bartik, alternate measures of urban residence

Note: This figure compares the growth in urbanization or density demand experienced by movers to that experienced by stayers, with either no controls, controlling for all observables, and controlling for all observables and county fixed effects.

B Results with imputed occupational codes

The census files provided by Ruggles et al. (2021) for census years 1900–1930 are preliminary. For the purposes of this paper, the main limitation is that in a significant number of cases, the occupational strings have not yet been assigned the three-digit codes that I use to assign them occupational ranks. Instead, they are listed as "Not Yet Classified" (occ1950=979). To bring these individuals back into analysis, I use Zimran's (2022) crosswalks to assign each such individual an imputed occupational code based on the original occupational string. The results in this appendix repeat the analysis of the main text, but incorporate these individuals.

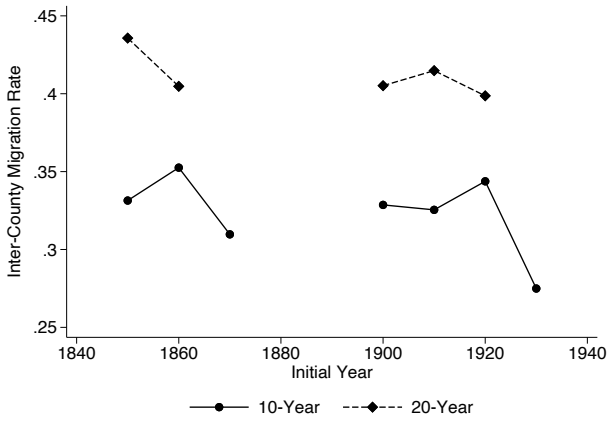


Figure B.1: Inter-county migration rates by linkage method and span, corrected for false matches

Note: This 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 the main 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.

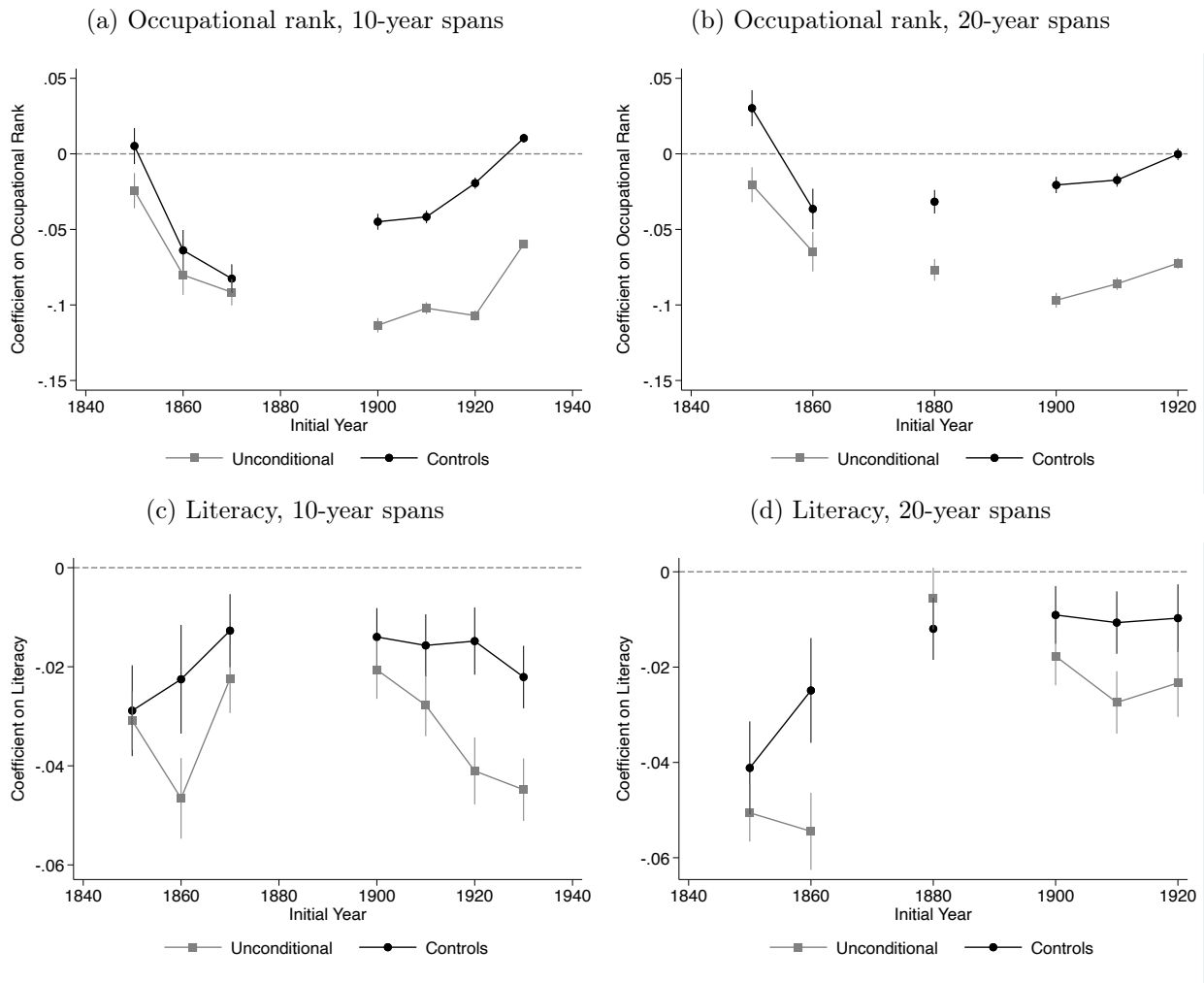


Figure B.2: Migrant selection by measure and span, individual characteristics

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as county fixed effects.

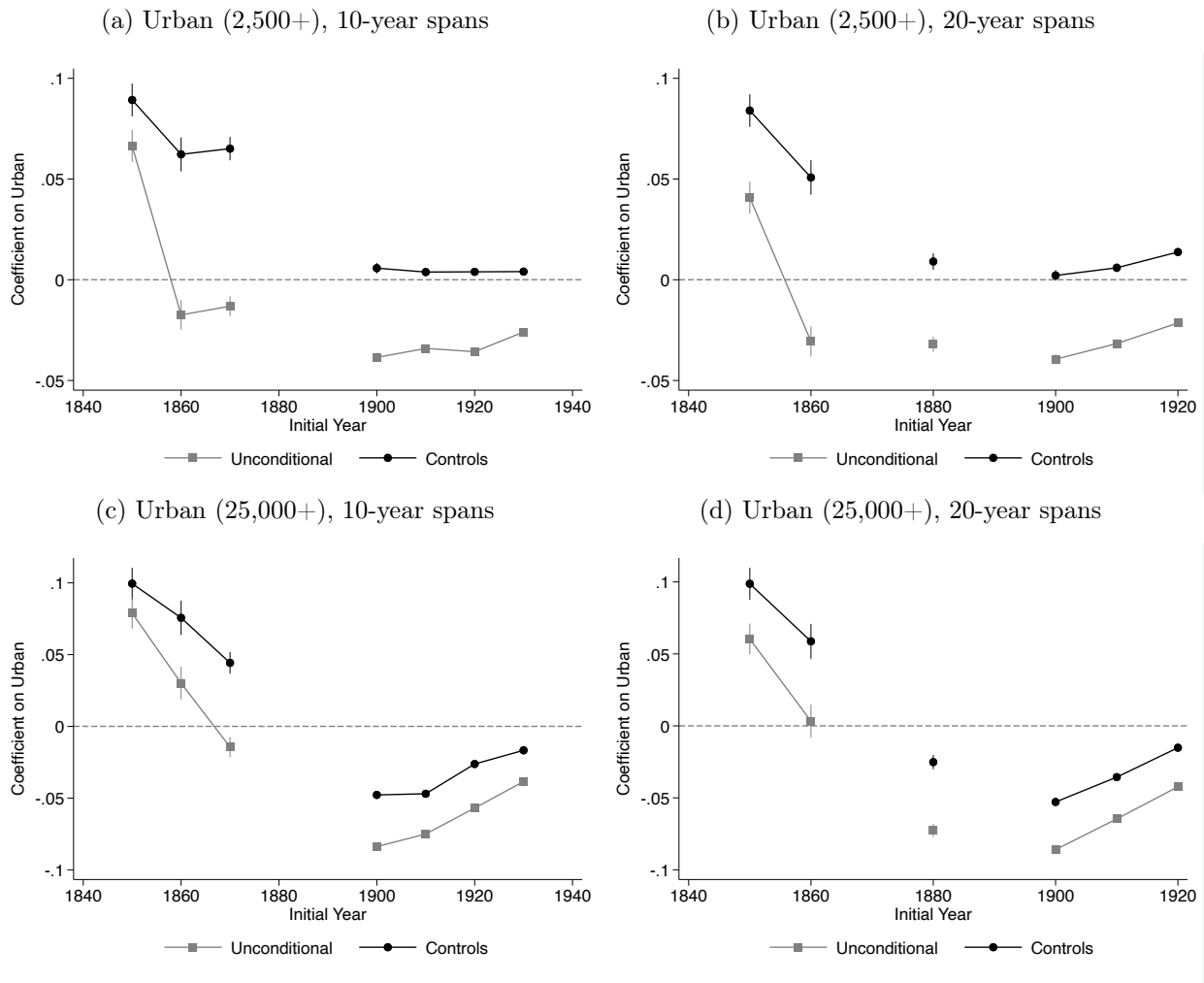


Figure B.3: Migrant selection by measure and span, urban residence

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as state fixed effects.

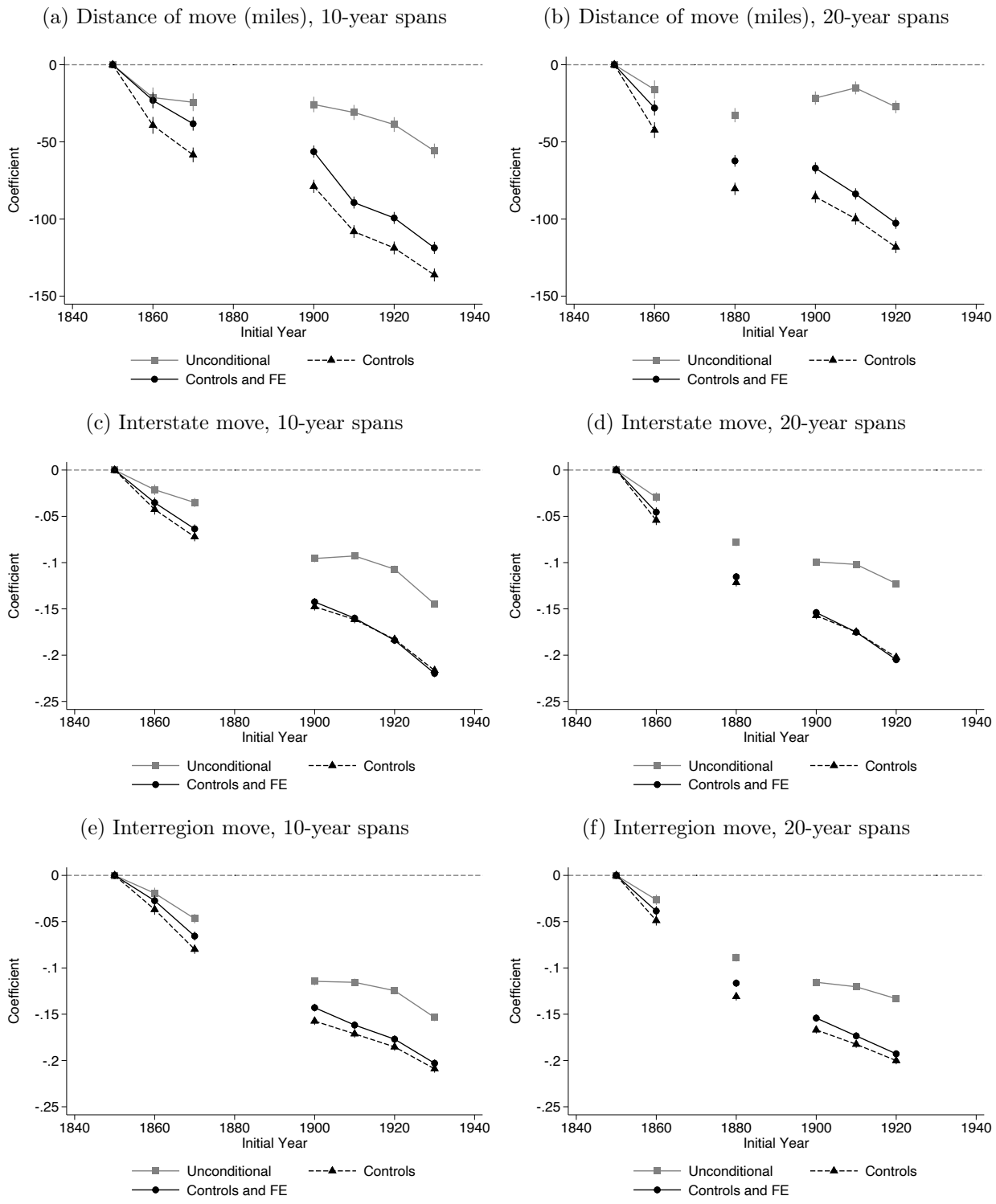


Figure B.4: Changes in distance of move

Note: These figures show the coefficients on initial-year fixed effects in regressions of destination characteristics with and without controls, with 1850 as the excluded year. "Controls" includes all controls available in all censuses, including state fixed effects. "Controls with FE' includes also county fixed effects. Vertical lines are 95-percent confidence intervals. Observations weighted to correct for selection into linkage.

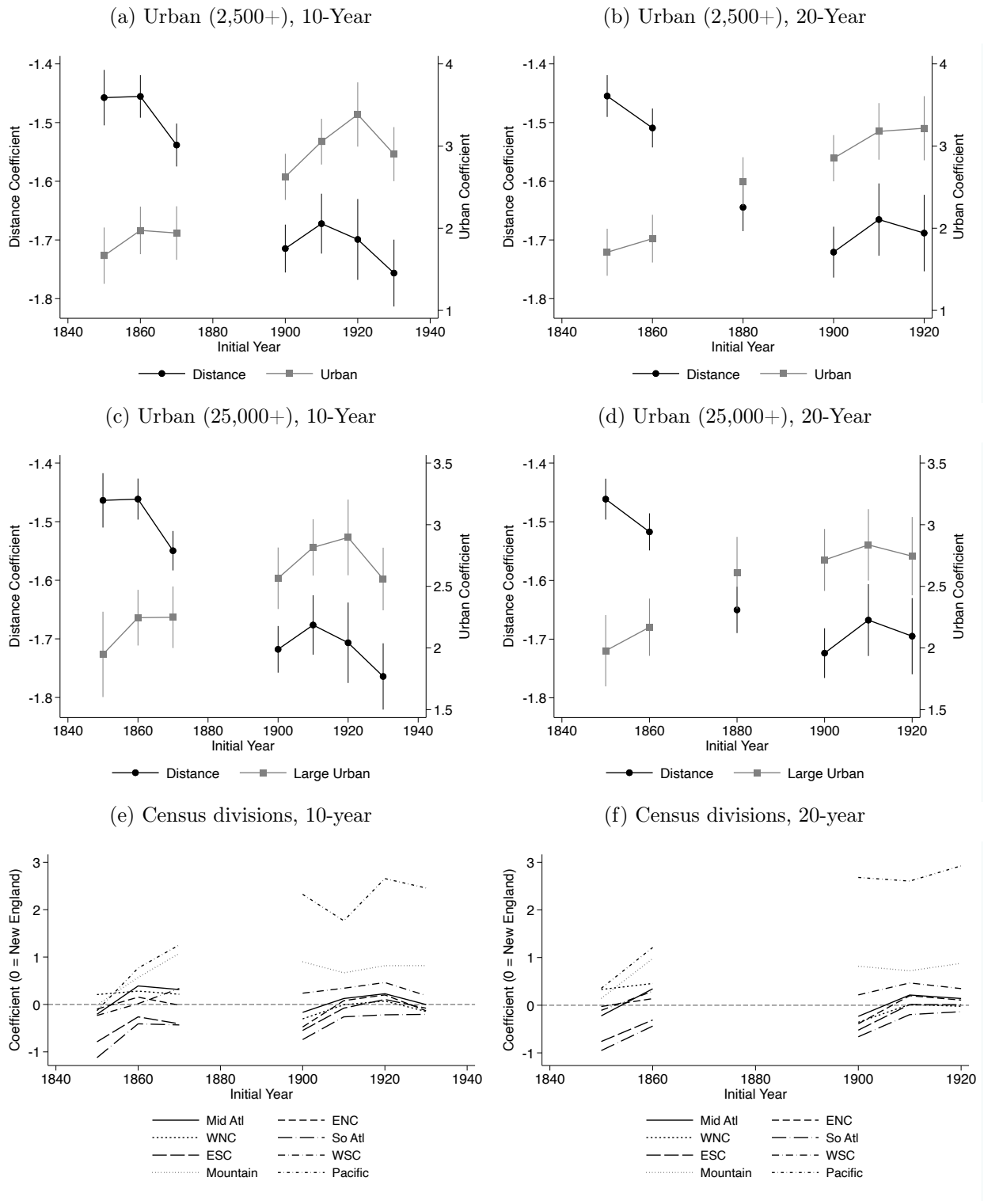


Figure B.5: Conditional Logit Results

Note: Panels (a)–(d) present coefficients on distance and the measure of urban residence from the conditional logit estimation. Panels (e) and (f) present the coefficients from the regressions of panels (a) and (b), respectively, on the census division indicators, with New England excluded.

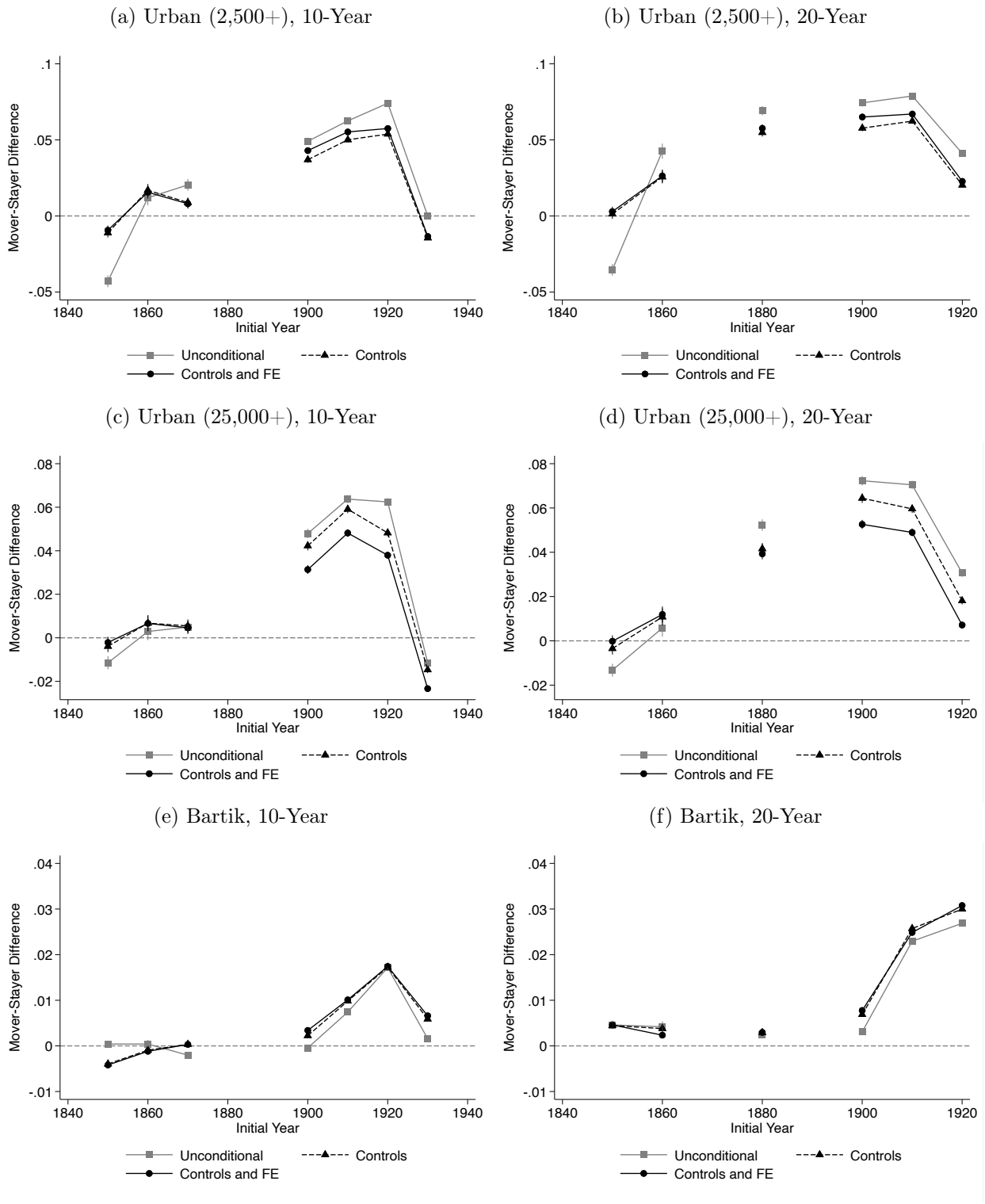


Figure B.6: Movers' change in urbanization and Bartik

Note: This figure compares the growth in urbanization or labor demand experienced by movers to that experienced by stayers, with either no controls, controlling for all observables, and controlling for all observables and county fixed effects.

C Results for longer-distance moves

In this appendix, I repeat the analysis of the main text, but redefine migration so that moves must be of at least 150 kilometers. The analysis is the same in all but two respects. First, I present migration rates only using the main linkage method, since the main text showed that estimates were insensitive to the method after correction. Second, for the correction of the estimated rates of migration for false linkage, I do not simplify equation (3) to equation (4) by setting $P(\text{county move}|\text{false match})$ to one. This is an accurate enough approximation in dealing with any inter-county move, but it is not in studying moves of at least 150 kilometers. The restriction of requiring a match to be to a person with the same birth state limits the probability that a false match is made to a distant person because individuals generally live in or near their state of birth. Thus, in this appendix, the correction is made using the unsimplified equation (3) in which $P(\text{county move}|\text{false match})$ is computed by determining, for each individual in the sample, the fraction of individuals to whom he could be linked (defined as those with the same birth place and an age-implied birthyear sufficiently similar that a match would be permitted) who in the second year of the span lived at least 150 miles away from the individual's county of residence in the first year of the span. The false match rate used in this calculation is the same as in the main text, since the same sample is being used.

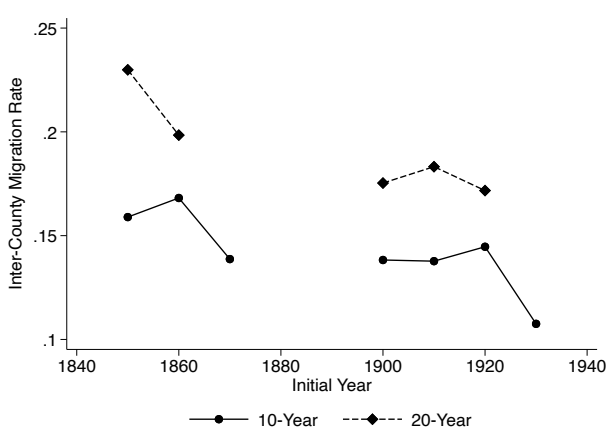


Figure C.1: Inter-county migration rates by span, corrected for false matches

Note: The 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 the main 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.

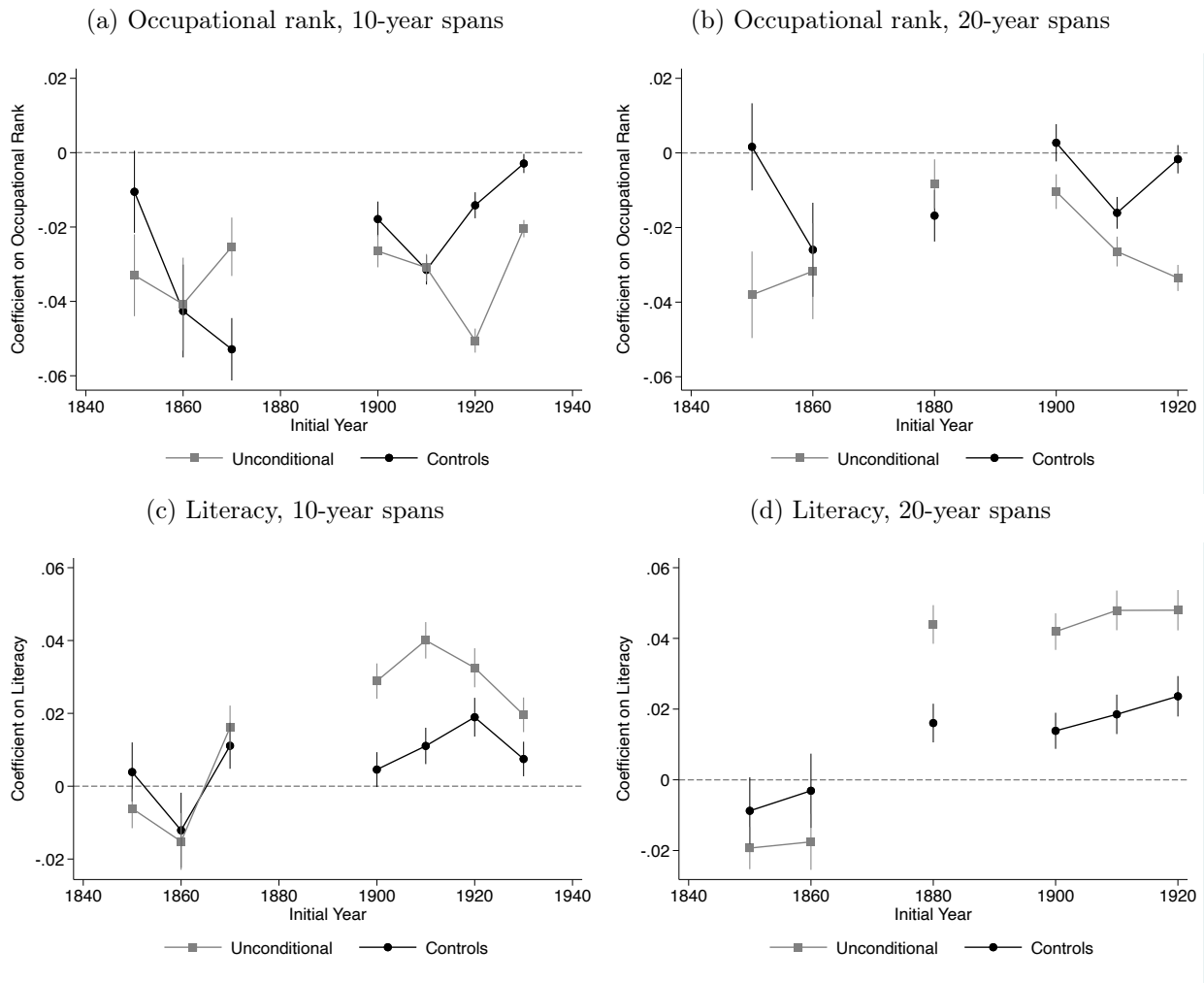


Figure C.2: Migrant selection by measure and span, individual characteristics

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as county fixed effects.

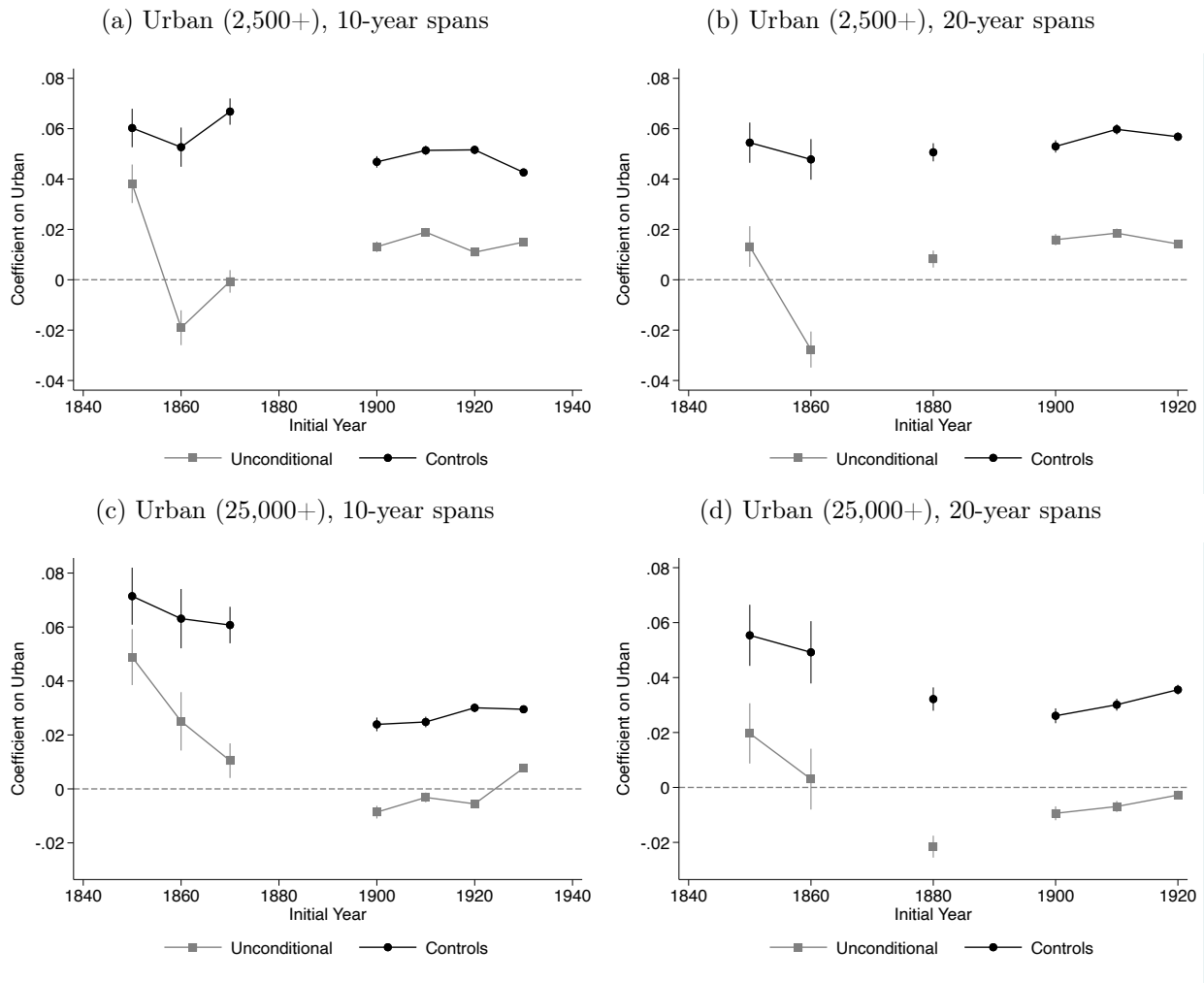


Figure C.3: Migrant selection by measure and span, urban residence

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as state fixed effects.

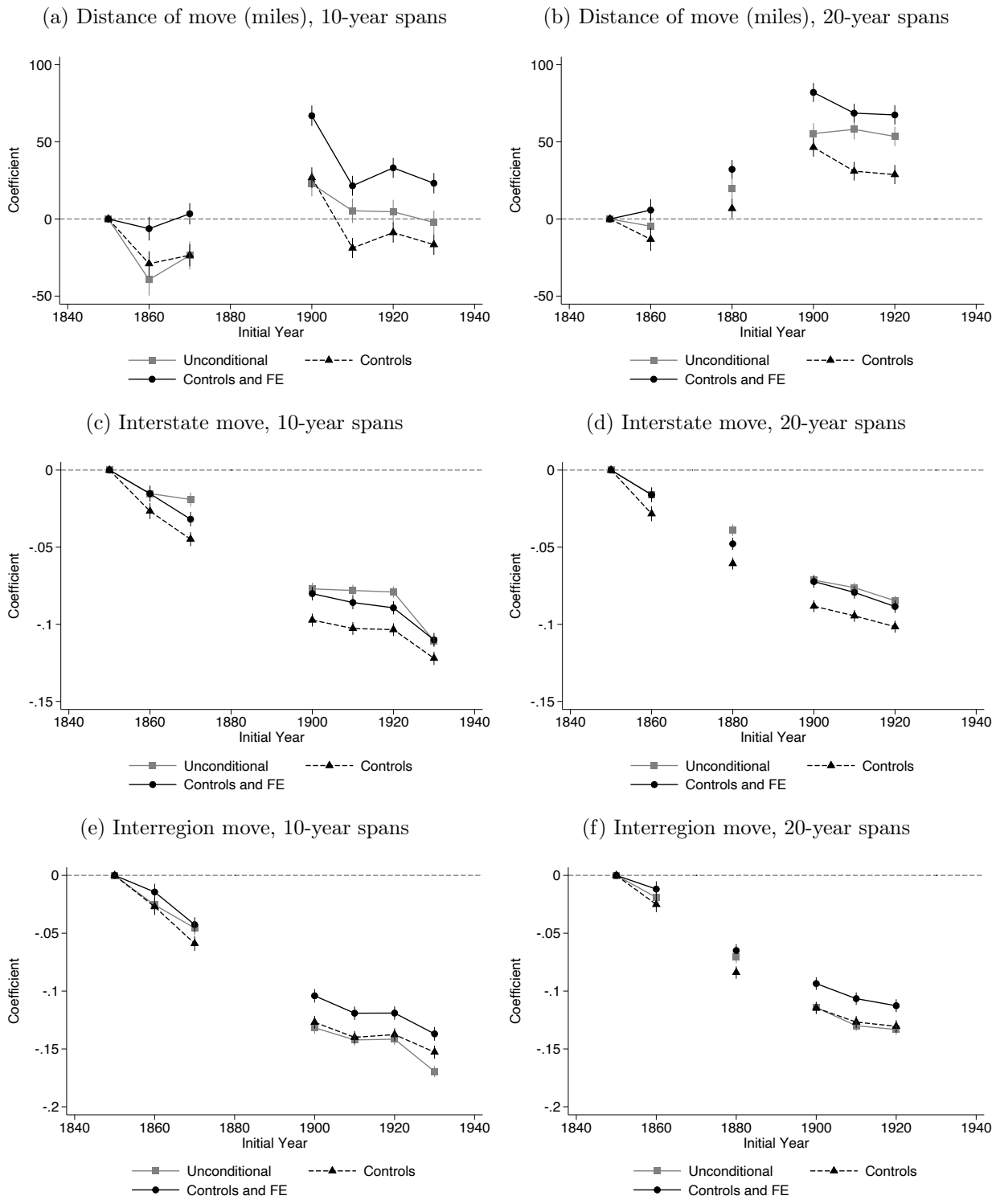


Figure C.4: Changes in distance of move

Note: These figures show the coefficients on initial-year fixed effects in regressions of destination characteristics with and without controls, with 1850 as the excluded year. "Controls" includes all controls available in all censuses, including state fixed effects. "Controls with FE' includes also county fixed effects. Vertical lines are 95-percent confidence intervals. Observations weighted to correct for selection into linkage.

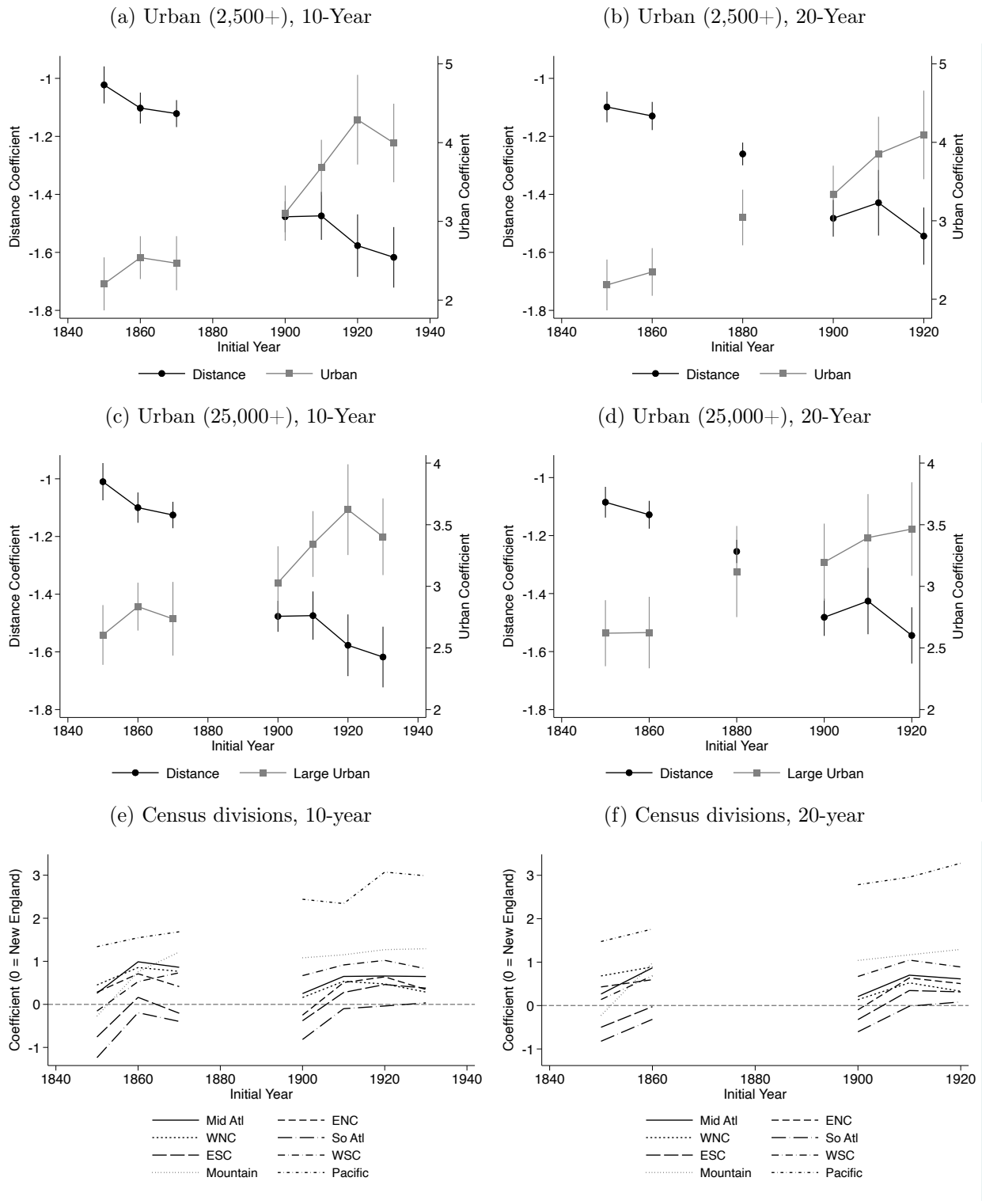


Figure C.5: Conditional Logit Results

Note: Panels (a)–(d) present coefficients on distance and the measure of urban residence from the conditional logit estimation. Panels (e) and (f) present the coefficients from the regressions of panels (a) and (b), respectively, on the census division indicators, with New England excluded.

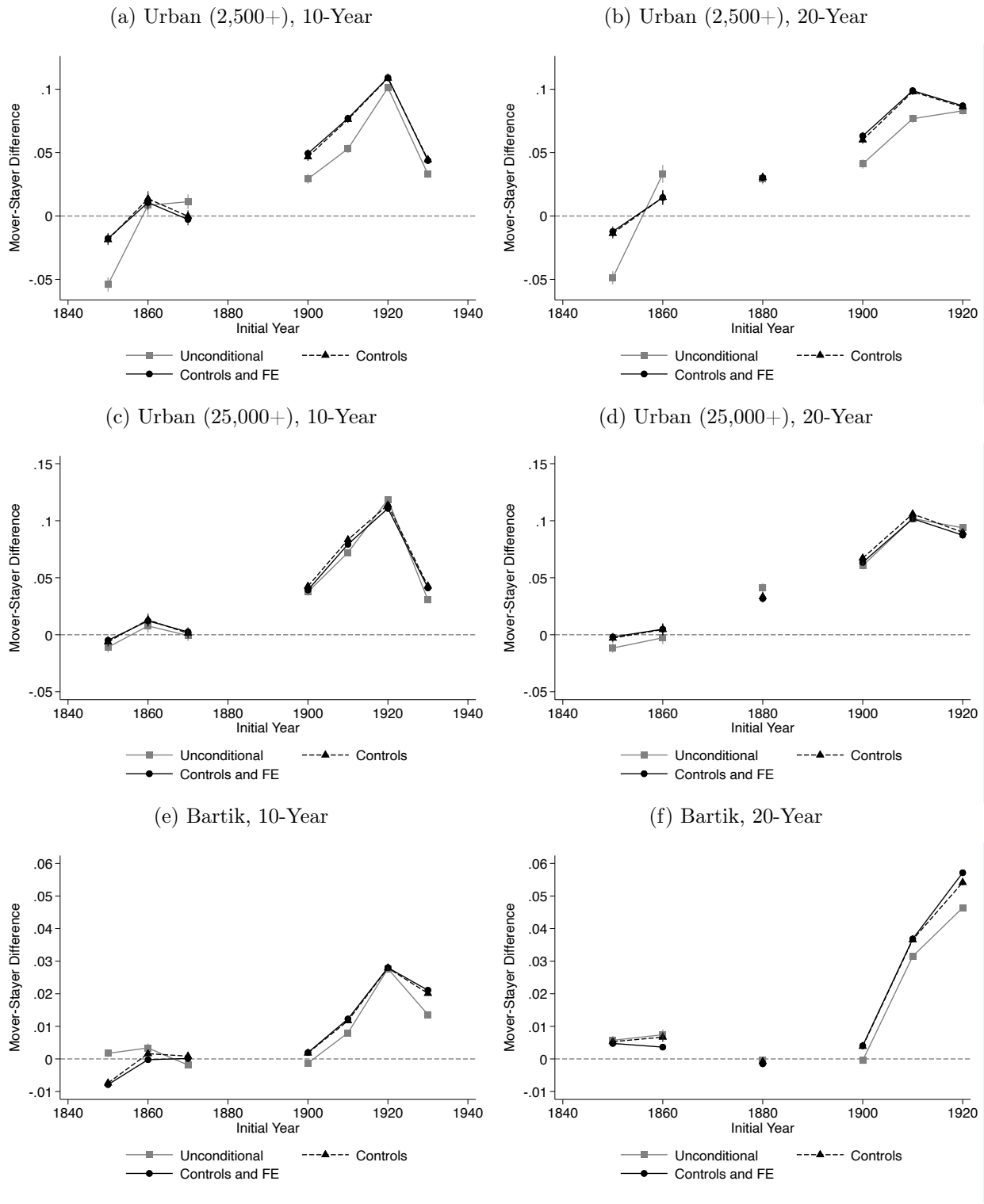


Figure C.6: Movers' change in urbanization and Bartik

Note: This figure compares the growth in urbanization or labor demand experienced by movers to that experienced by stayers, with either no controls, controlling for all observables, and controlling for all observables and county fixed effects.

D Selection and sorting results with alternative linkage methods

D.1 ABE-Exact

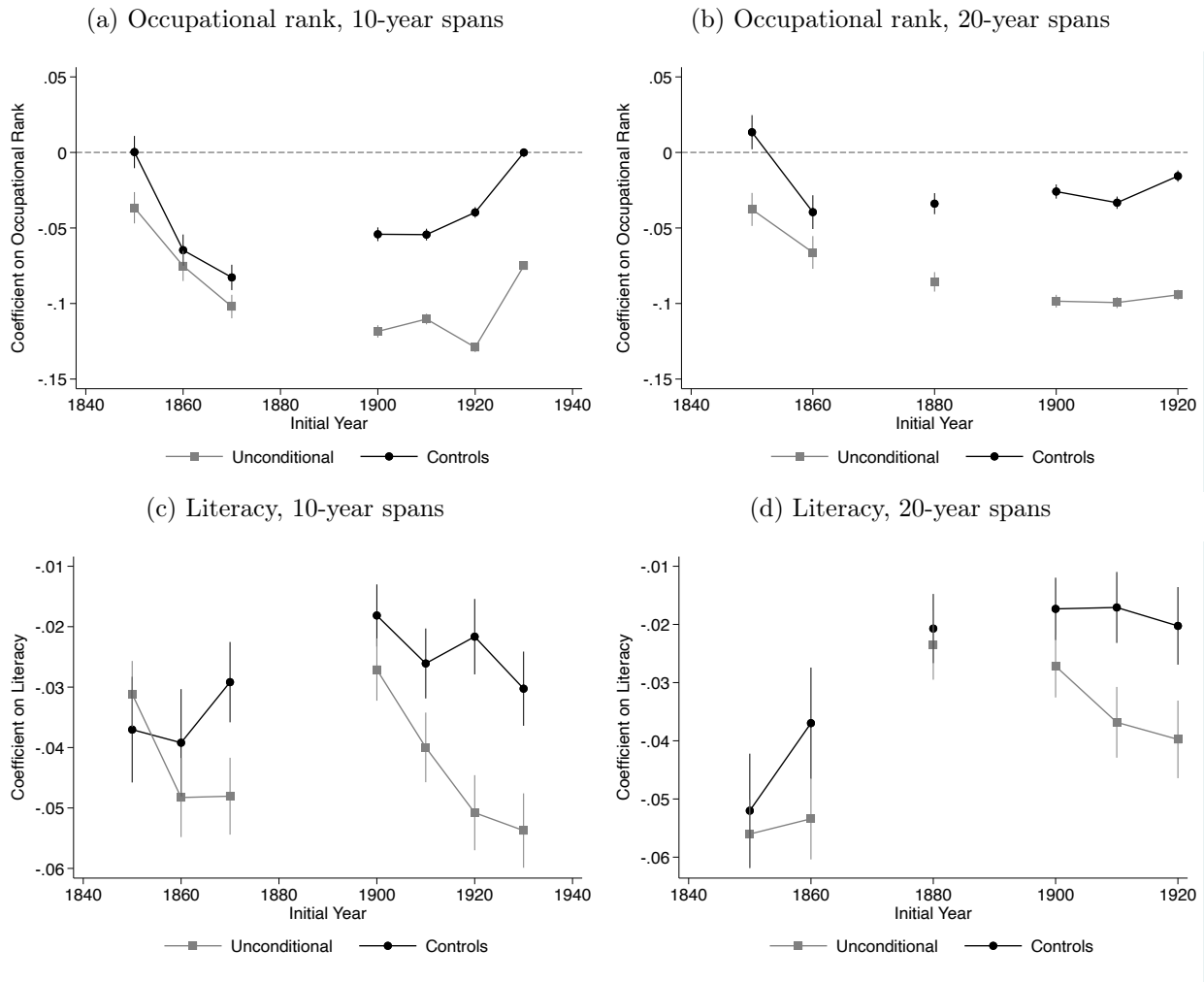


Figure D.1.1: Migrant selection by measure and span, individual characteristics

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as county fixed effects.

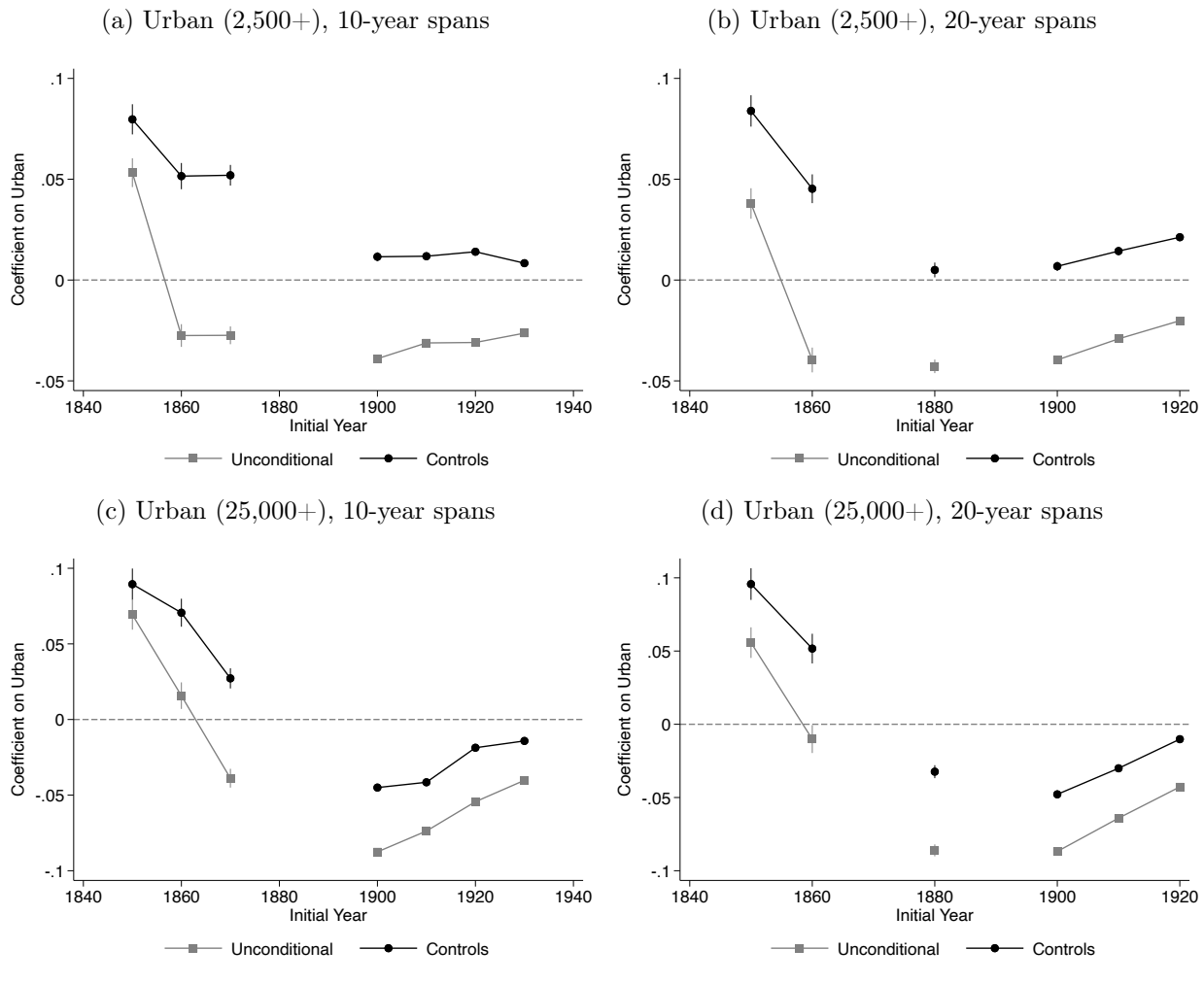


Figure D.1.2: Migrant selection by measure and span, urban residence

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as state fixed effects.

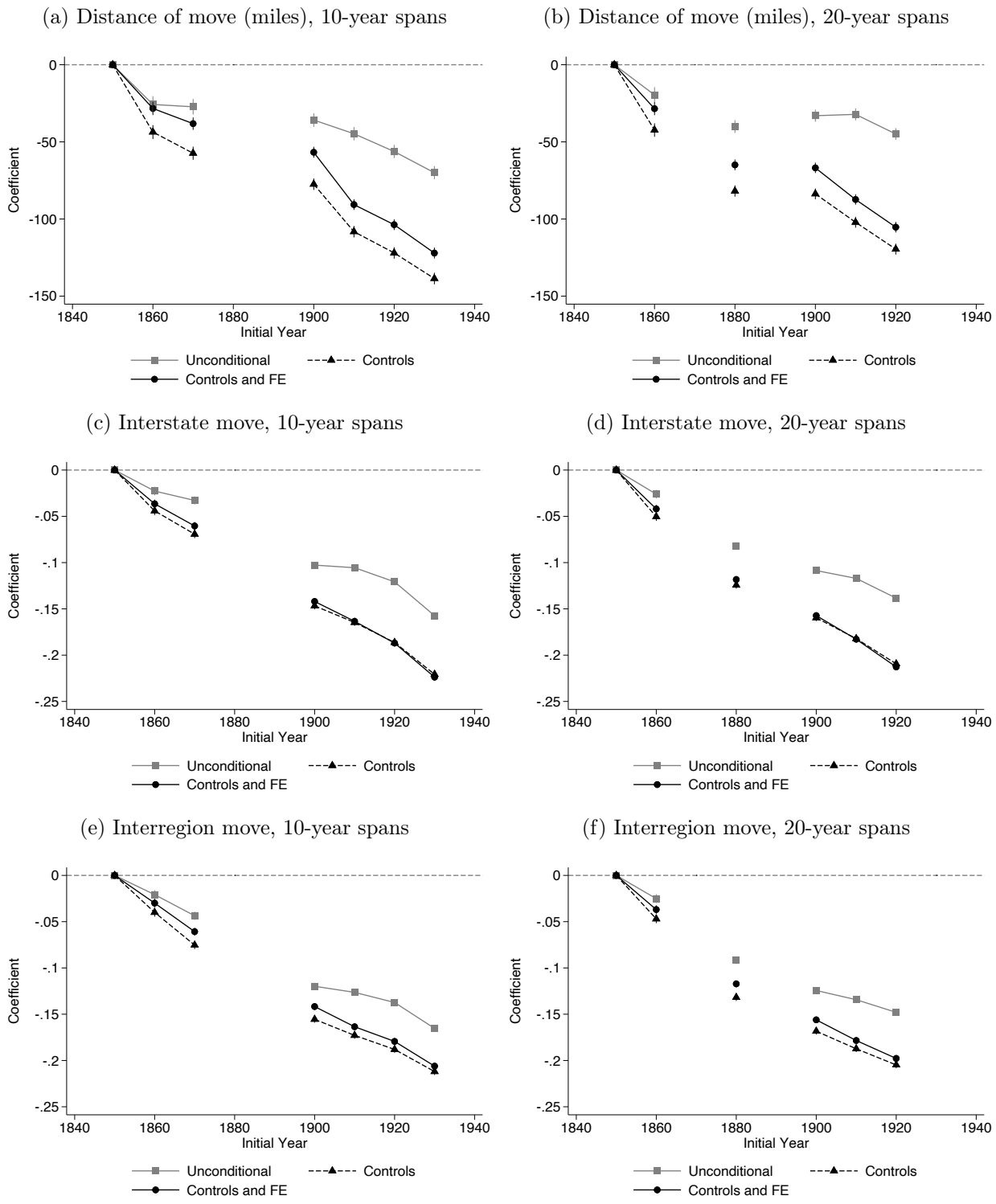


Figure D.1.3: Changes in distance of move

Note: These figures show the coefficients on initial-year fixed effects in regressions of destination characteristics with and without controls, with 1850 as the excluded year. "Controls" includes all controls available in all censuses, including state fixed effects. "Controls with FE' includes also county fixed effects. Vertical lines are 95-percent confidence intervals. Observations weighted to correct for selection into linkage.

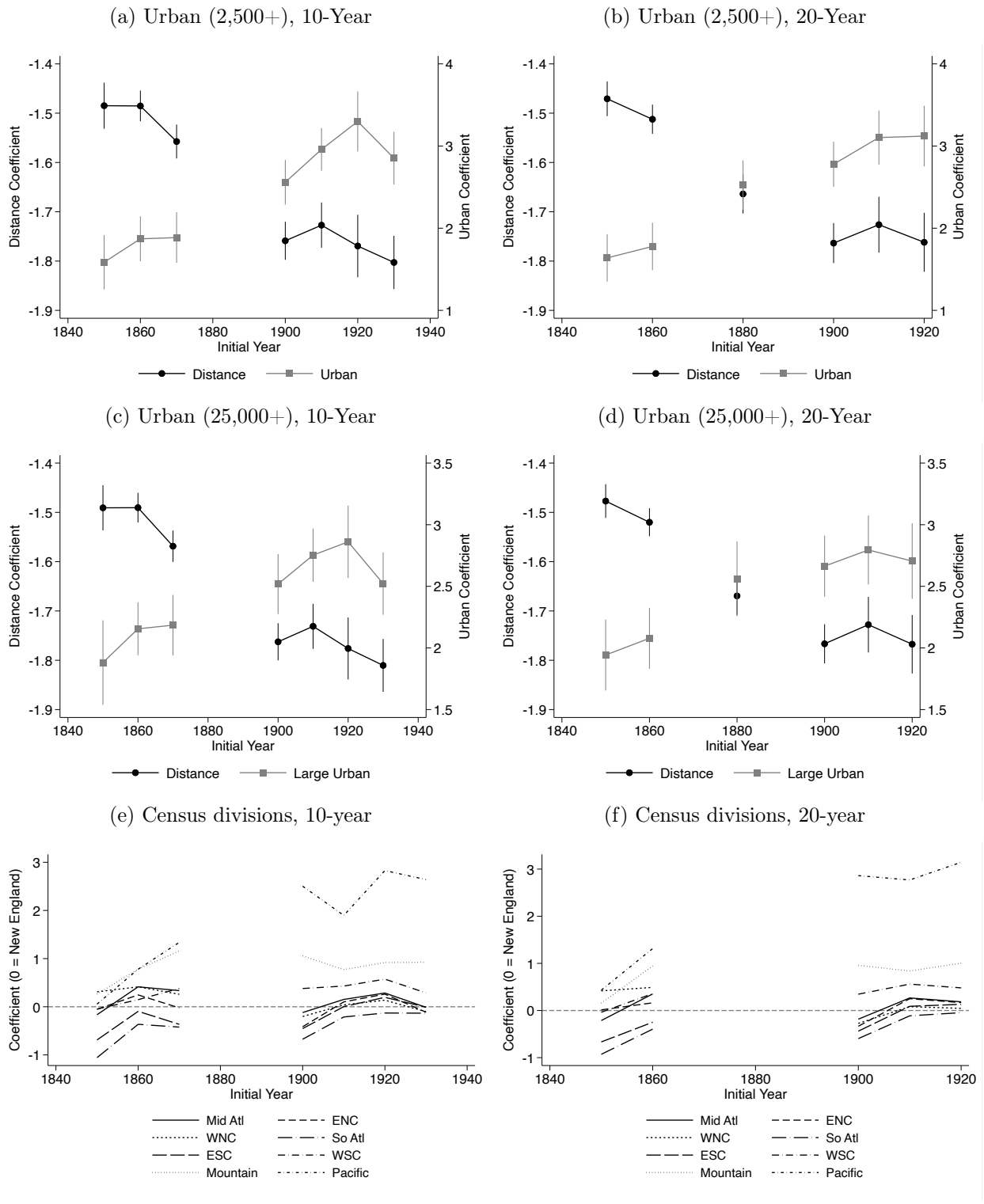


Figure D.1.4: Conditional Logit Results

Note: Panels (a)–(d) present coefficients on distance and the measure of urban residence from the conditional logit estimation. Panels (e) and (f) present the coefficients from the regressions of panels (a) and (b), respectively, on the census division indicators, with New England excluded.

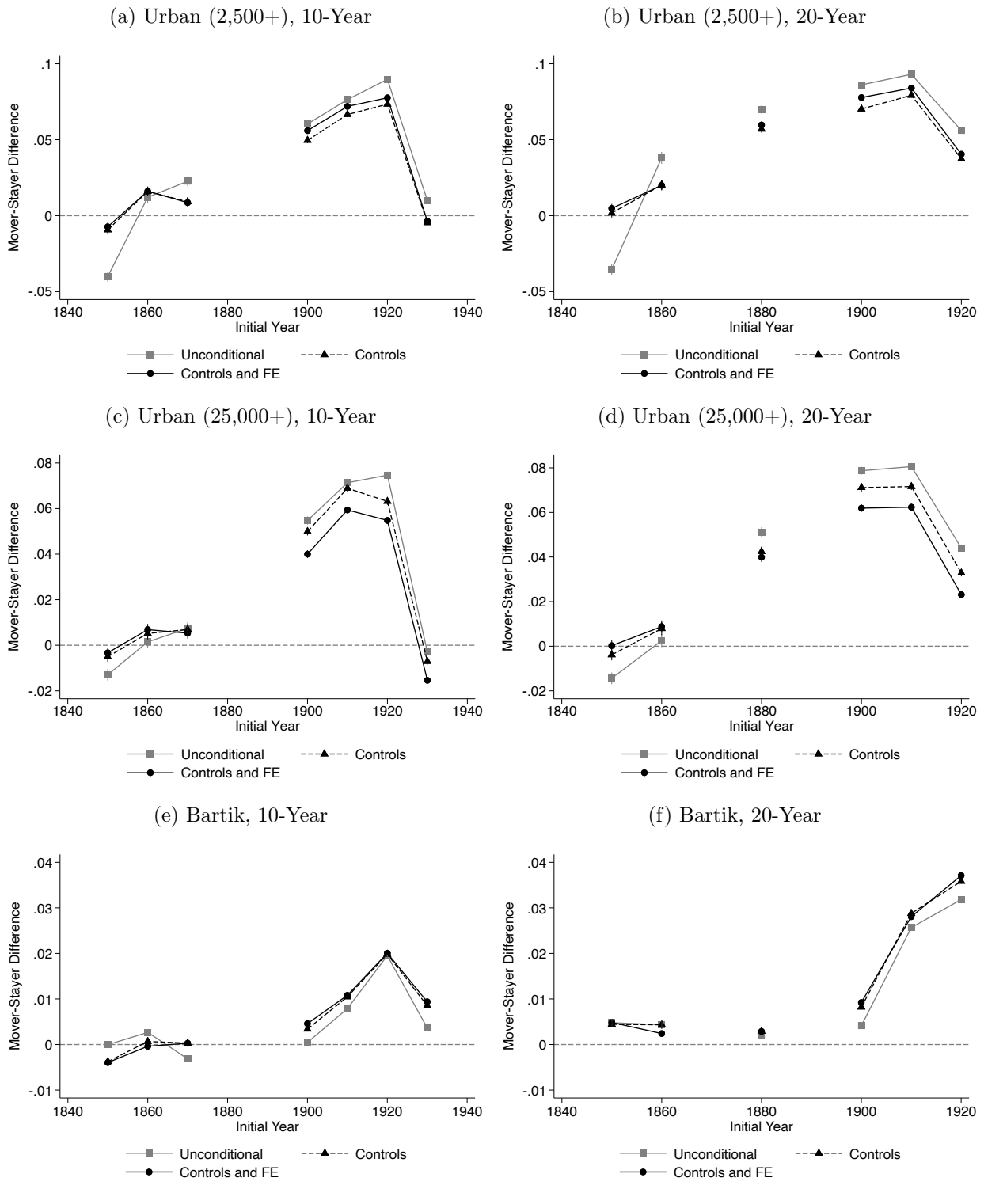


Figure D.1.5: Movers' change in urbanization and Bartik

Note: This figure compares the growth in urbanization or labor demand experienced by movers to that experienced by stayers, with either no controls, controlling for all observables, and controlling for all observables and county fixed effects.

D.2 ABE-NYSIIS

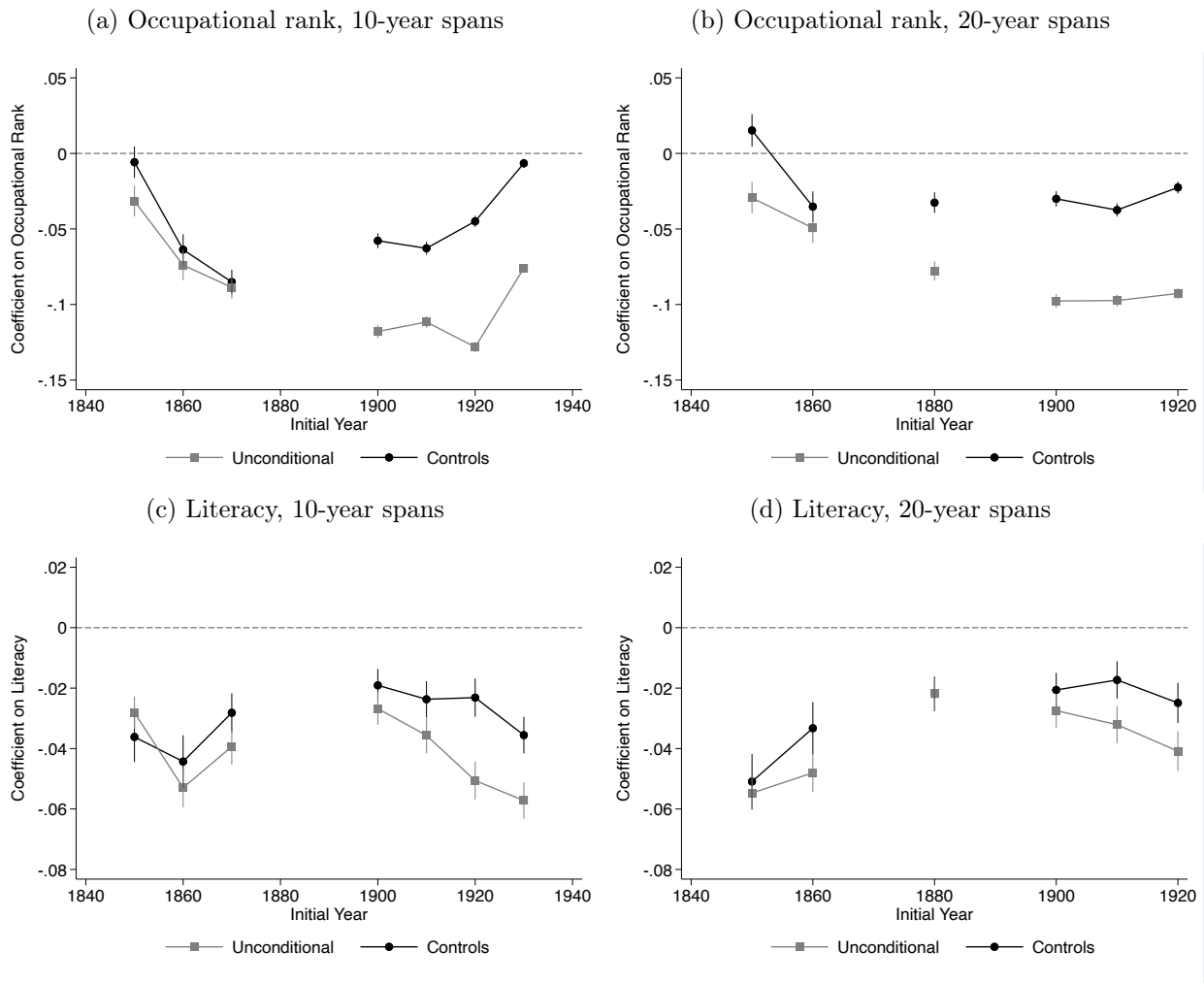


Figure D.2.1: Migrant selection by measure and span, individual characteristics

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as county fixed effects.

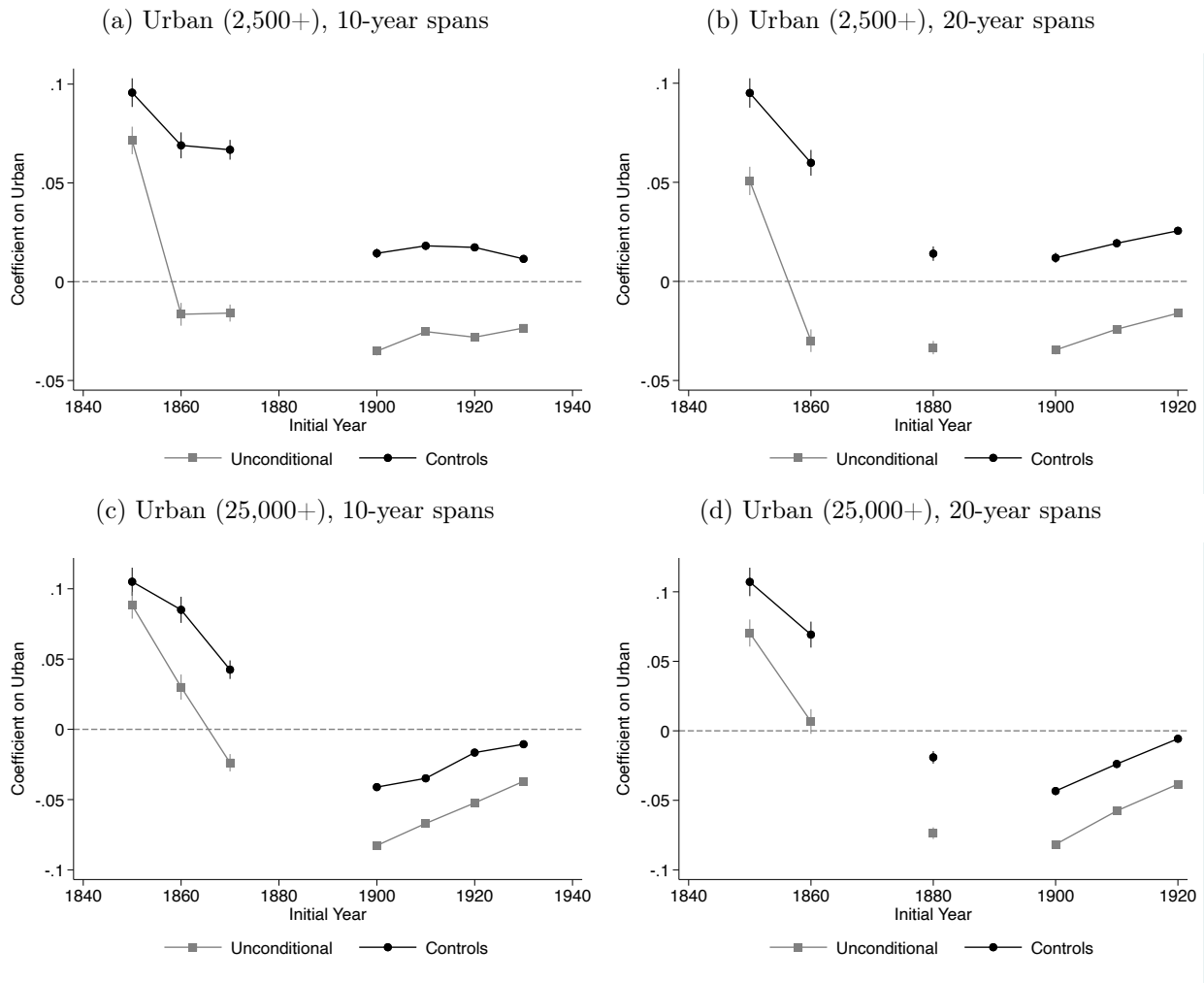


Figure D.2.2: Migrant selection by measure and span, urban residence

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as state fixed effects.

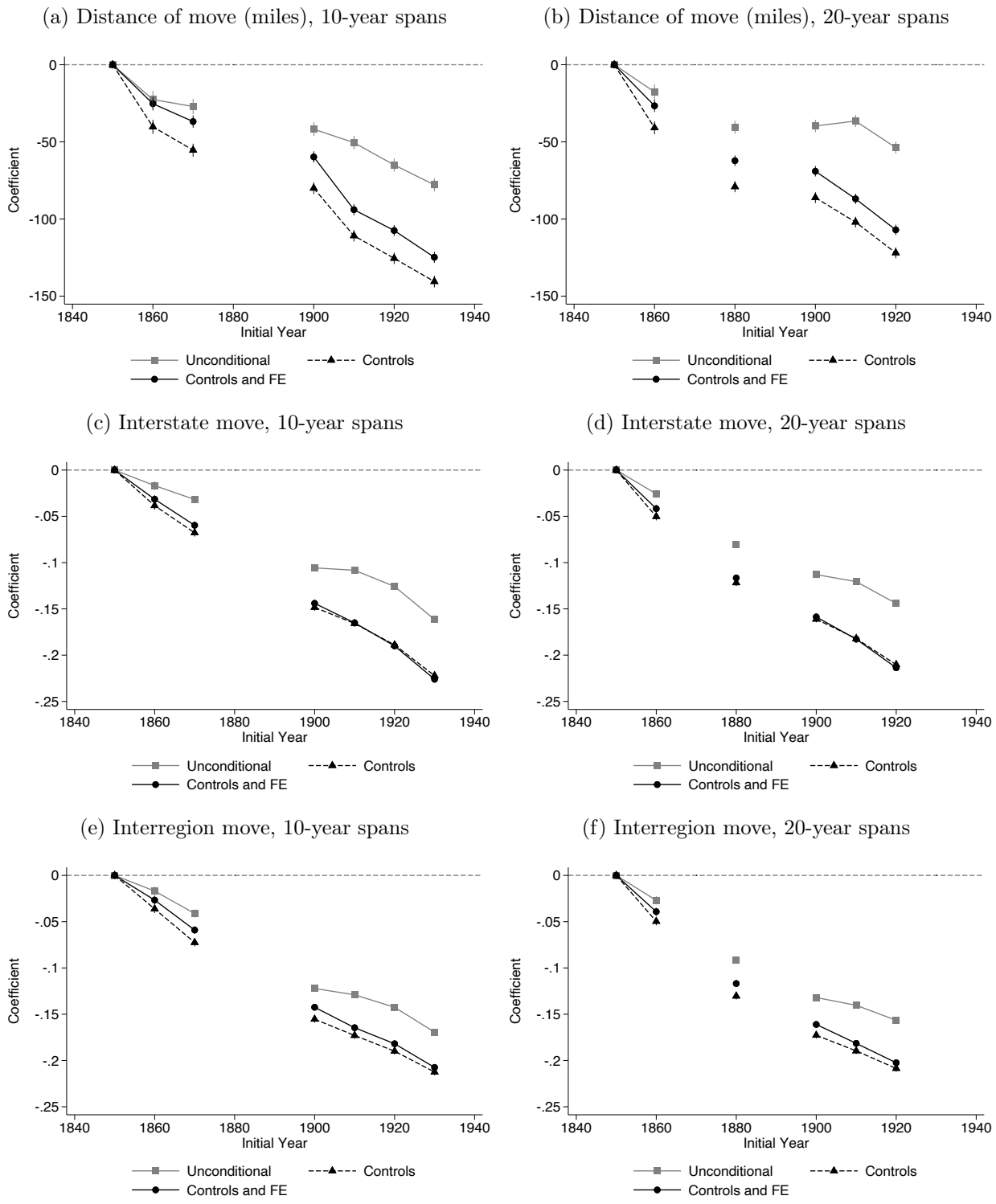
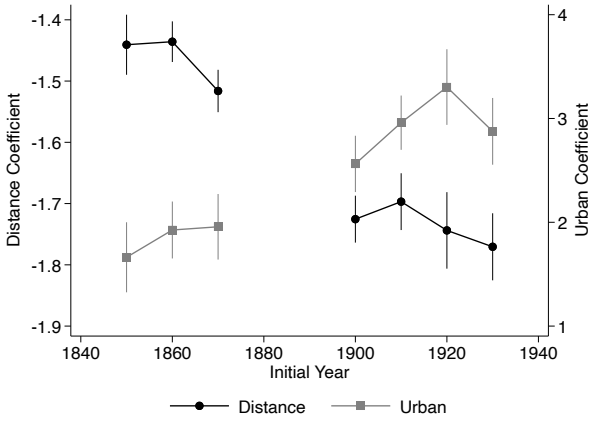


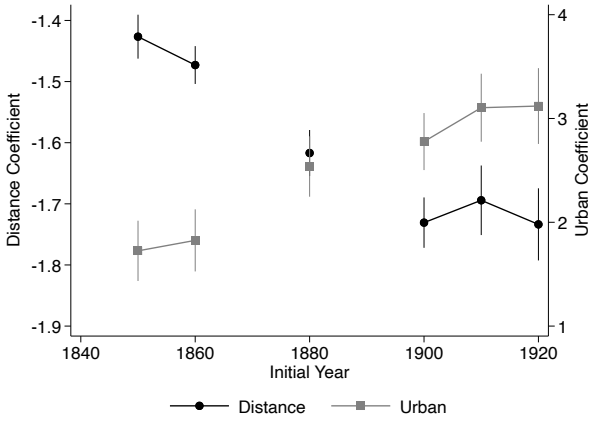
Figure D.2.3: Changes in distance of move

Note: These figures show the coefficients on initial-year fixed effects in regressions of destination characteristics with and without controls, with 1850 as the excluded year. "Controls" includes all controls available in all censuses, including state fixed effects. "Controls with FE' includes also county fixed effects. Vertical lines are 95-percent confidence intervals. Observations weighted to correct for selection into linkage.

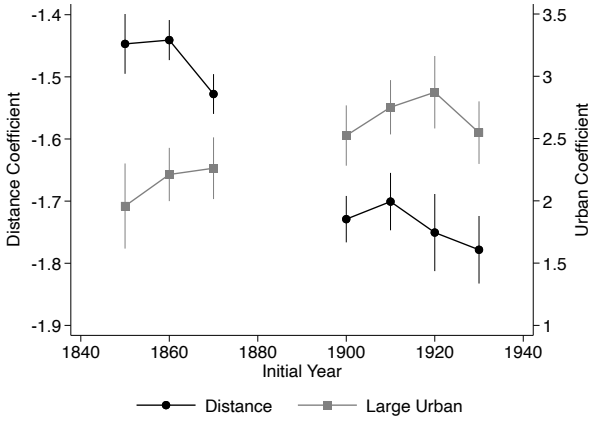
(a) Urban (2,500+), 10-Year



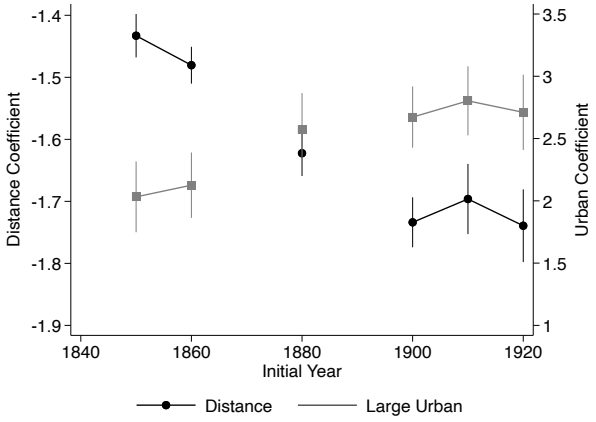
(b) Urban (2,500+), 20-Year



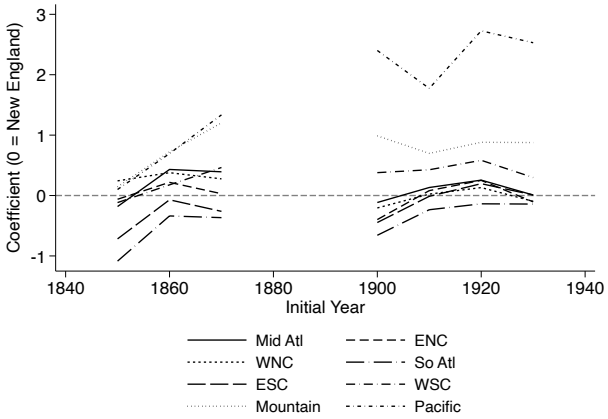
(c) Urban (25,000+), 10-Year



(d) Urban (25,000+), 20-Year



(e) Census divisions, 10-year



(f) Census divisions, 20-year

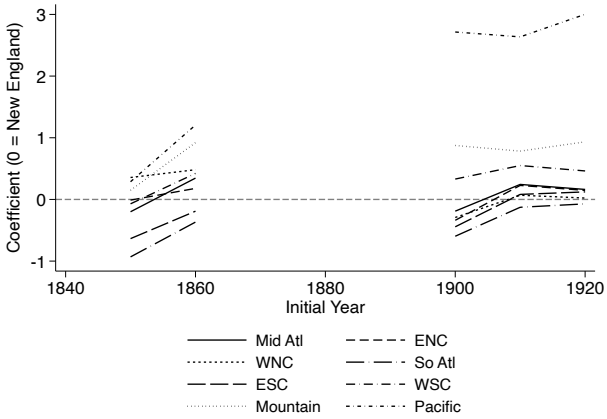


Figure D.2.4: Conditional Logit Results

Note: Panels (a)–(d) present coefficients on distance and the measure of urban residence from the conditional logit estimation. Panels (e) and (f) present the coefficients from the regressions of panels (a) and (b), respectively, on the census division indicators, with New England excluded.

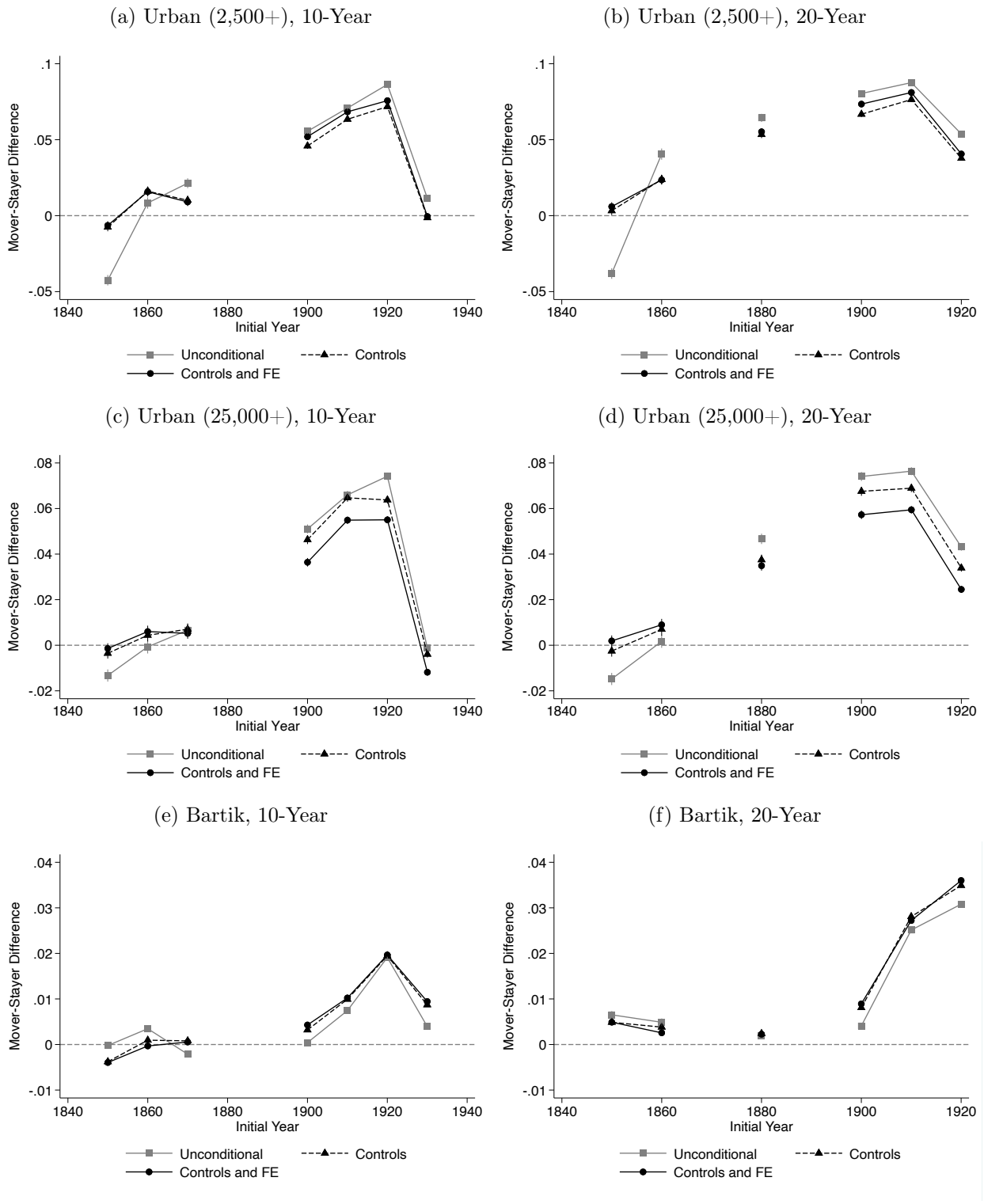


Figure D.2.5: Movers' change in urbanization and Bartik

Note: This figure compares the growth in urbanization or labor demand experienced by movers to that experienced by stayers, with either no controls, controlling for all observables, and controlling for all observables and county fixed effects.

D.3 Intersection of matches

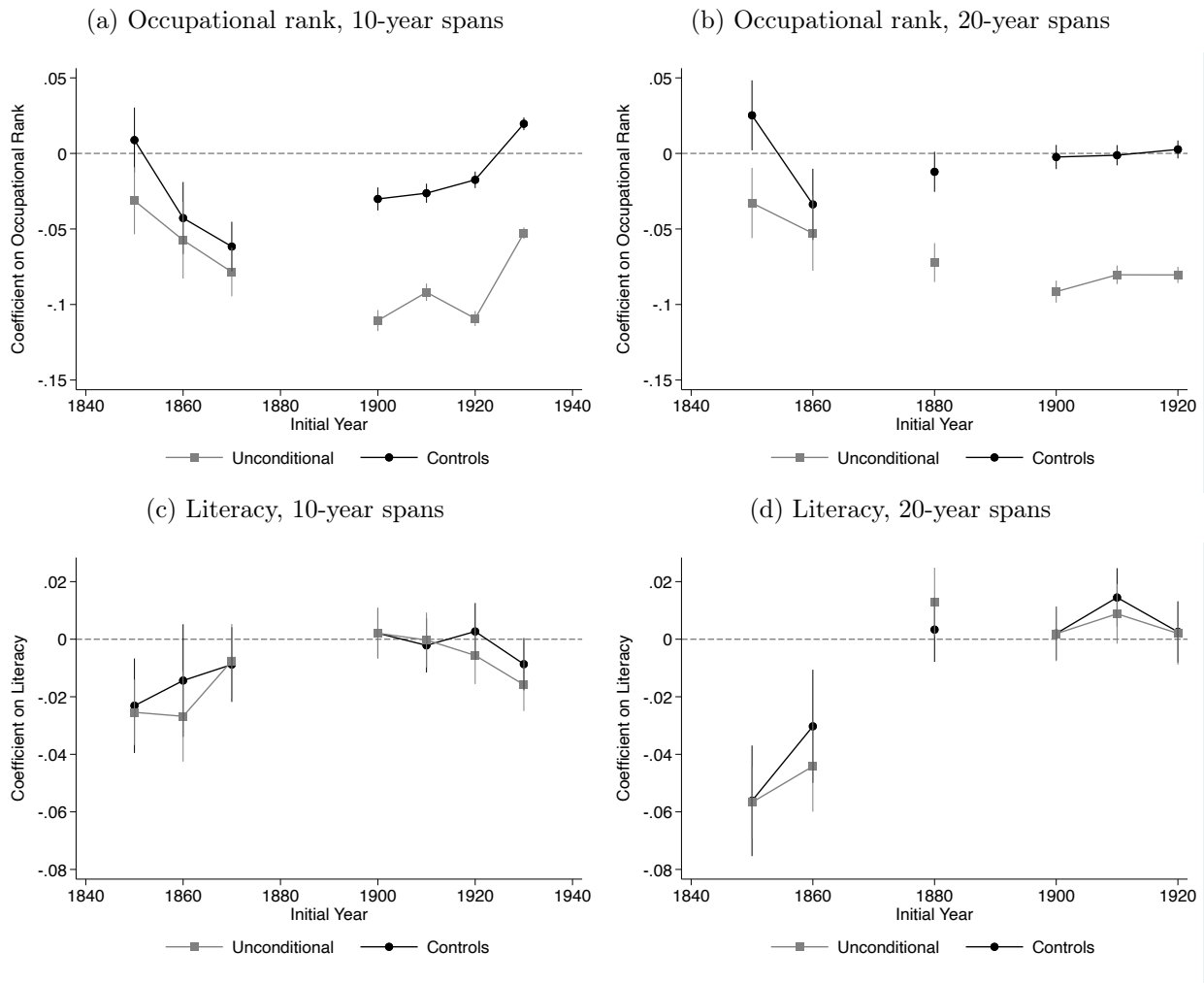


Figure D.3.1: Migrant selection by measure and span individual characteristics

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as county fixed effects.

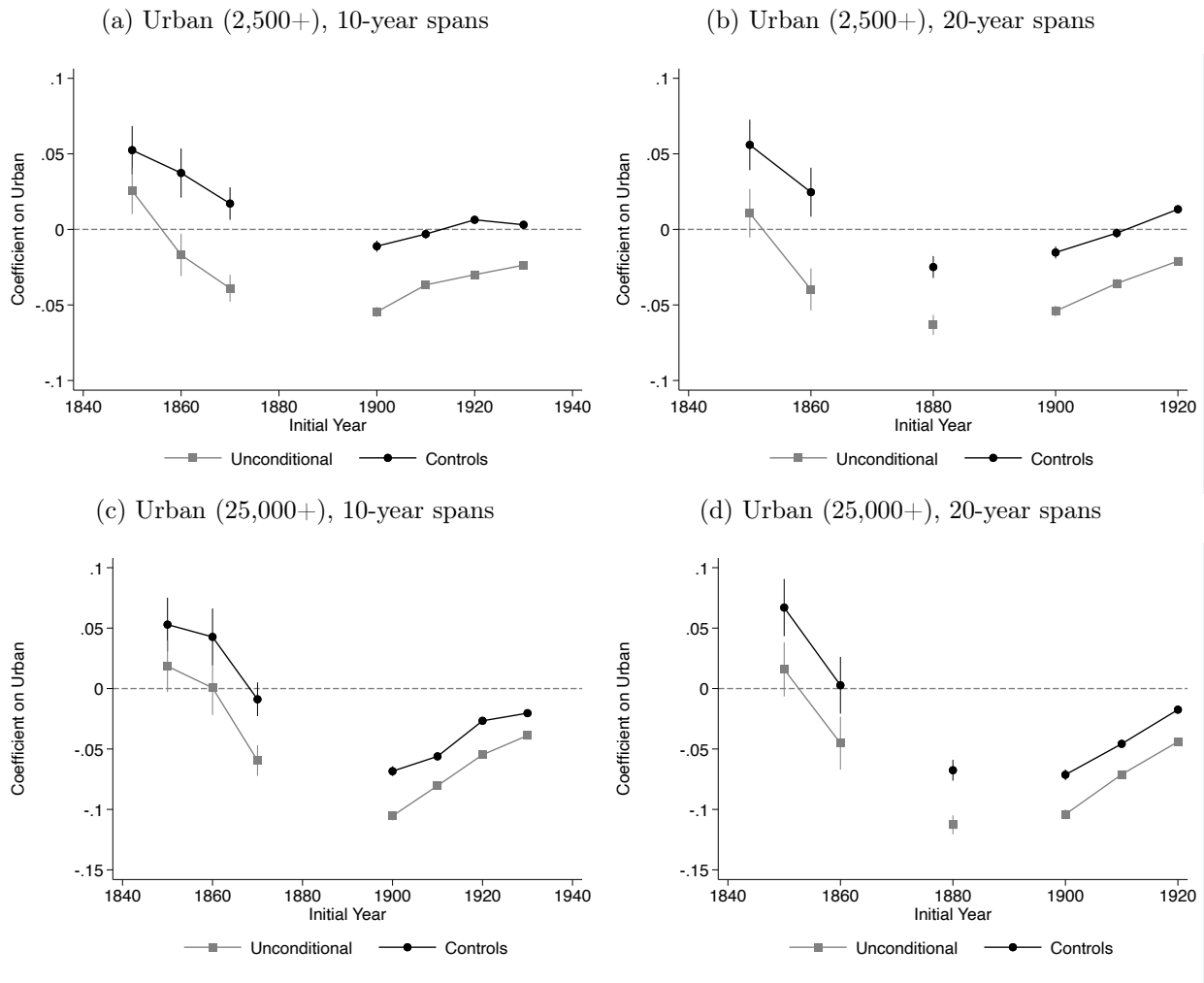


Figure D.3.2: Migrant selection by measure and span, urban residence

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as state fixed effects.

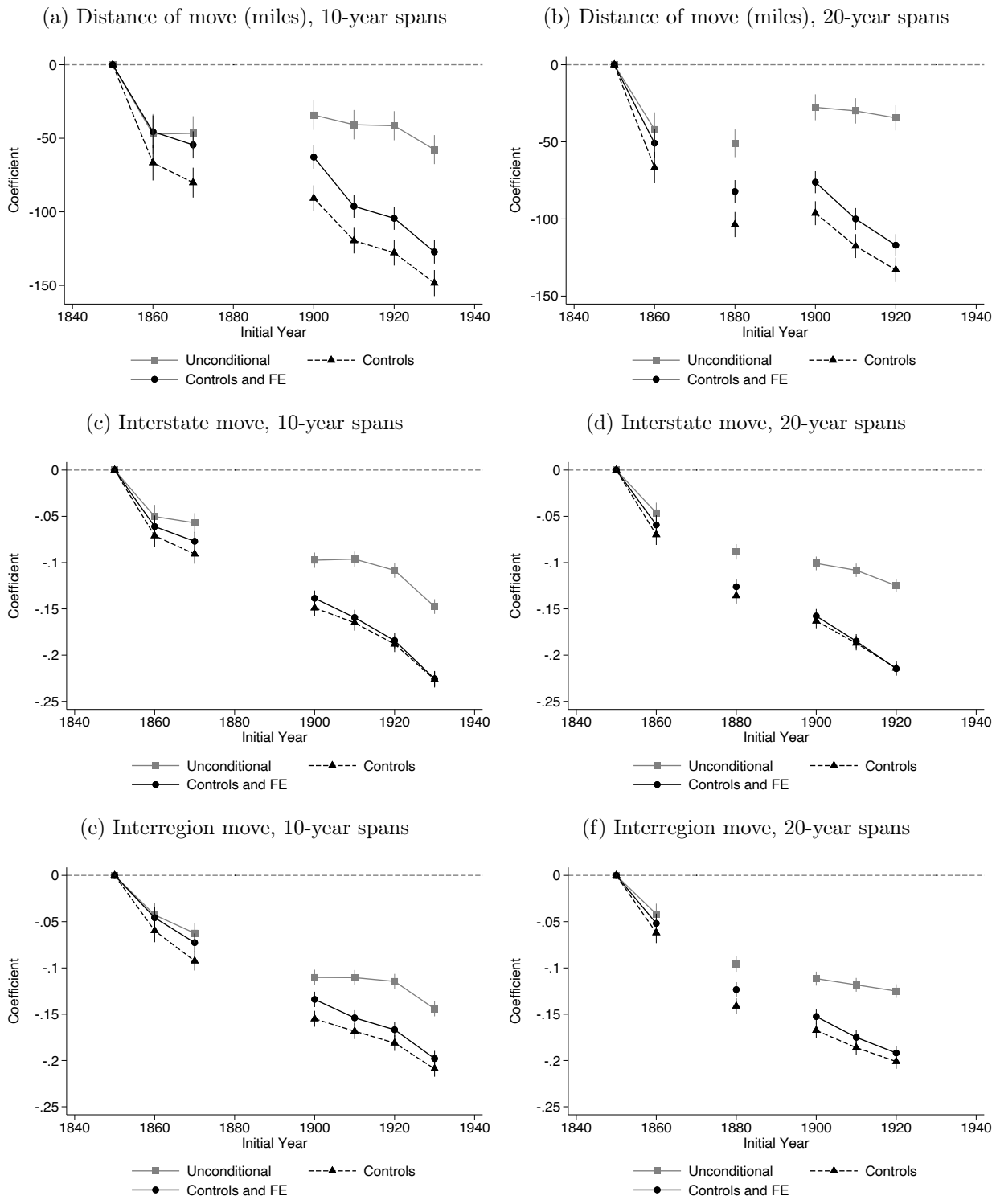


Figure D.3.3: Changes in distance of move

Note: These figures show the coefficients on initial-year fixed effects in regressions of destination characteristics with and without controls, with 1850 as the excluded year. "Controls" includes all controls available in all censuses, including state fixed effects. "Controls with FE' includes also county fixed effects. Vertical lines are 95-percent confidence intervals. Observations weighted to correct for selection into linkage.

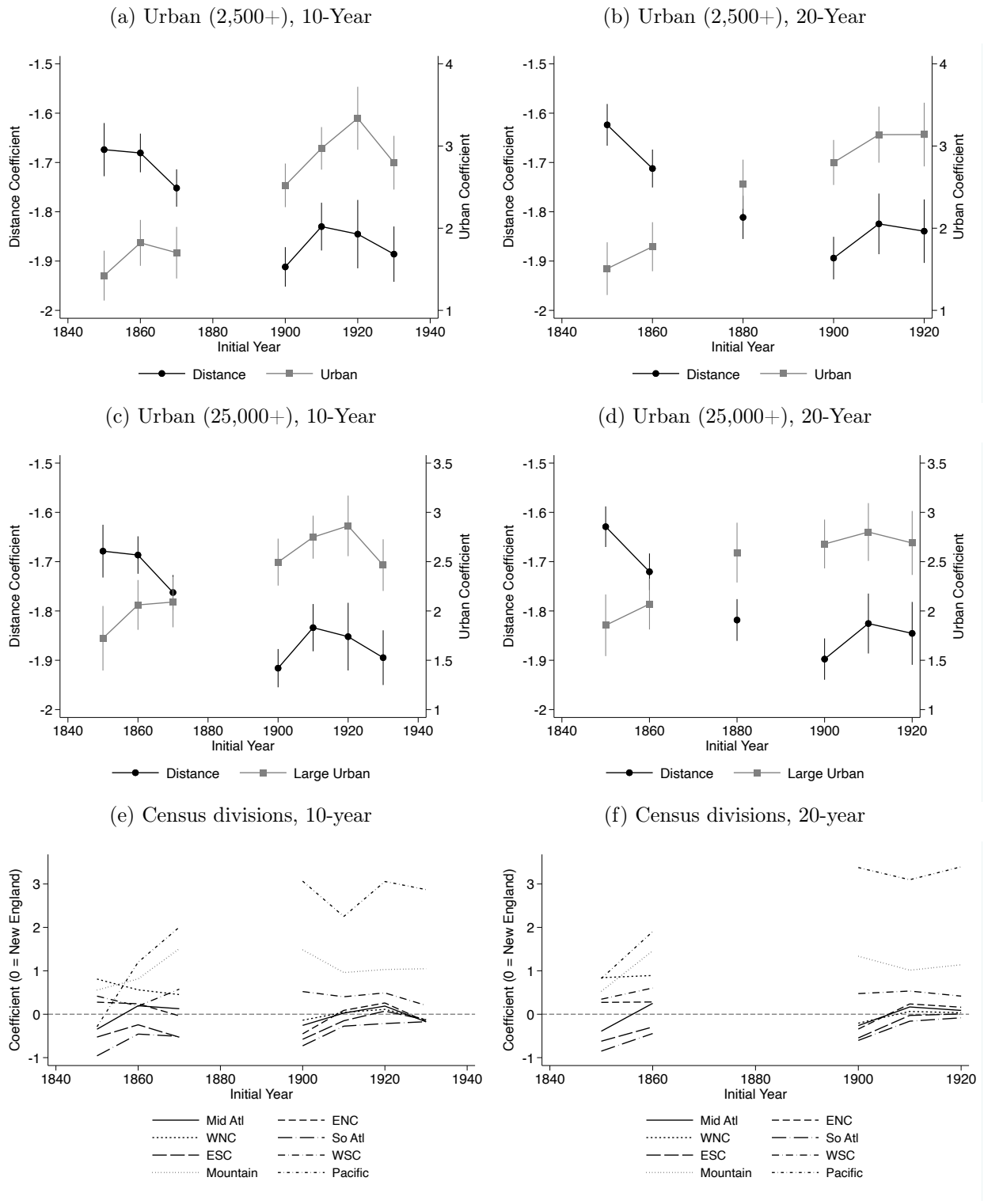


Figure D.3.4: Conditional Logit Results

Note: Panels (a)–(d) present coefficients on distance and the measure of urban residence from the conditional logit estimation. Panels (e) and (f) present the coefficients from the regressions of panels (a) and (b), respectively, on the census division indicators, with New England excluded.

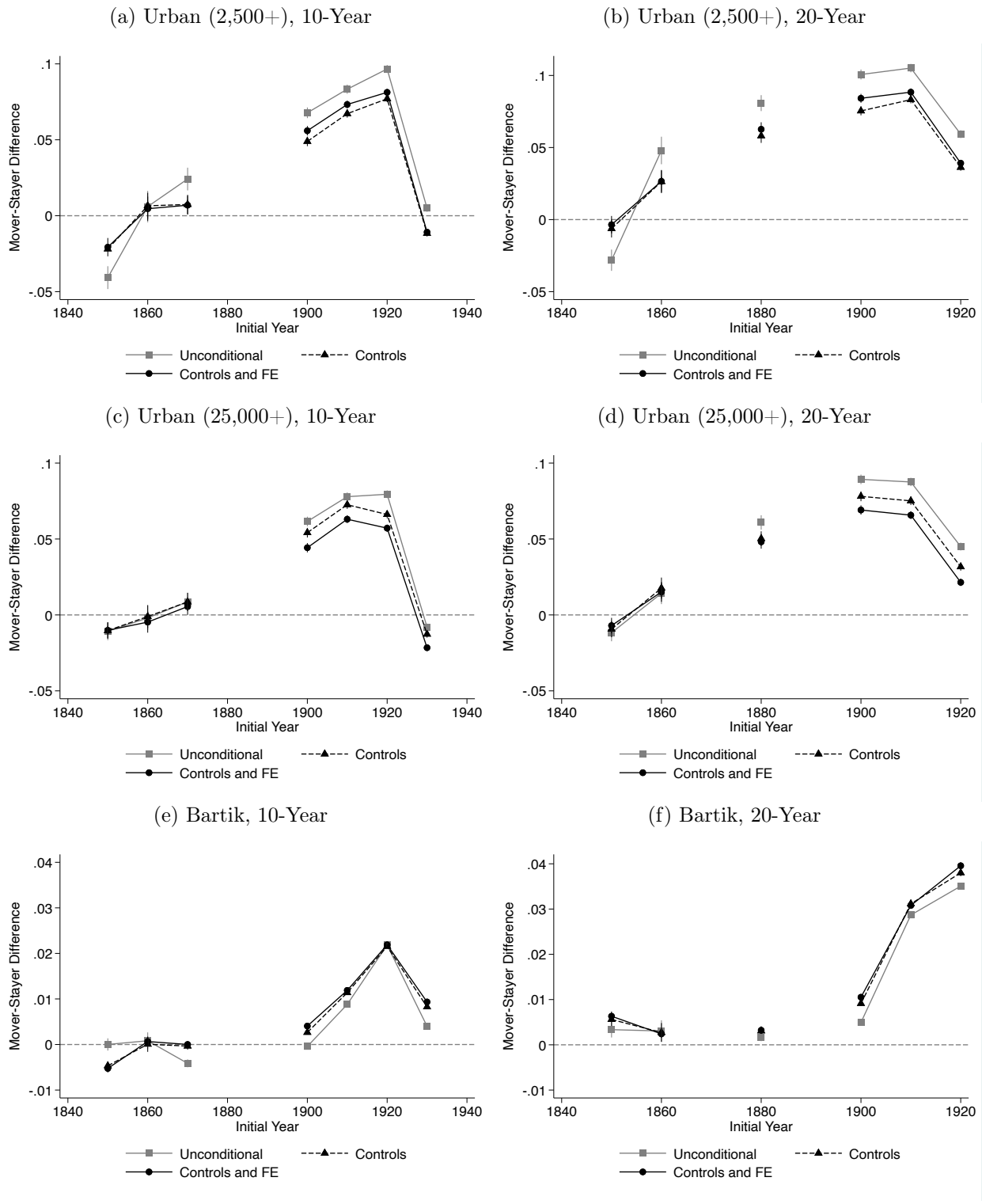


Figure D.3.5: Movers' change in urbanization and Bartik

Note: This figure compares the growth in urbanization or labor demand experienced by movers to that experienced by stayers, with either no controls, controlling for all observables, and controlling for all observables and county fixed effects.

D.4 Intersection of matches plus corroboration

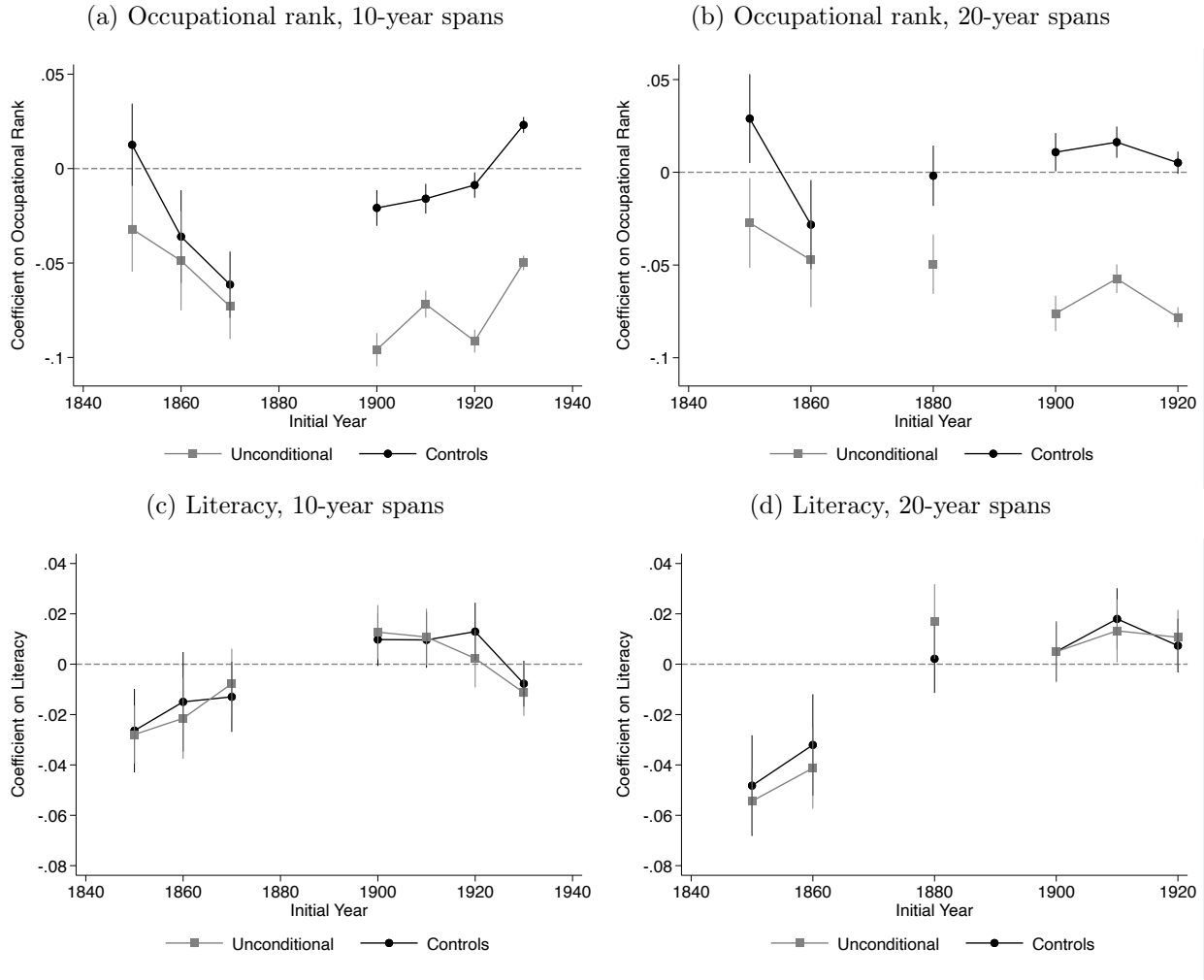


Figure D.4.6: Migrant selection by measure and span, individual characteristics

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as county fixed effects.

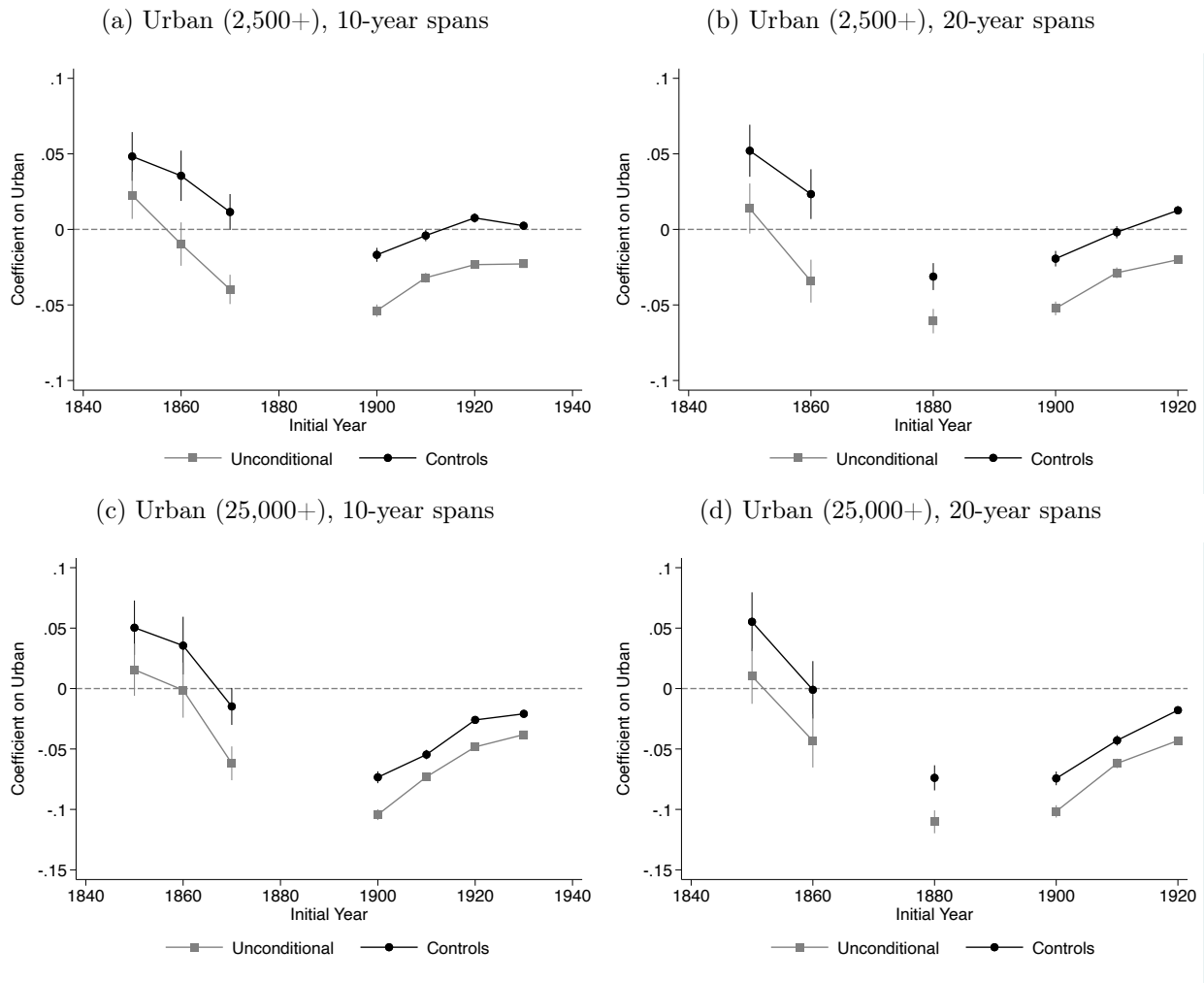


Figure D.4.7: Migrant selection by measure and span, urban residence

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as state fixed effects.

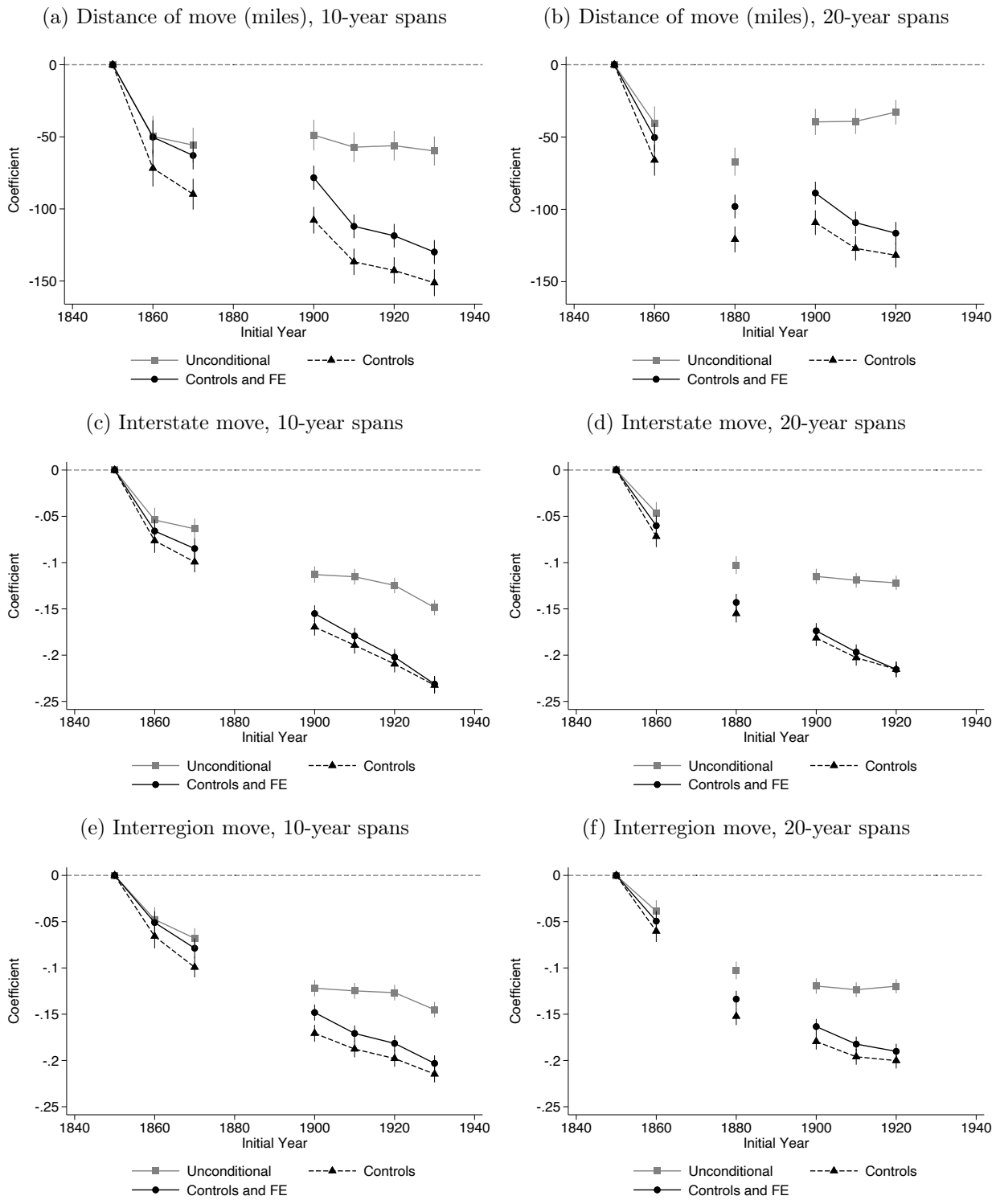


Figure D.4.8: Changes in distance of move

Note: These figures show the coefficients on initial-year fixed effects in regressions of destination characteristics with and without controls, with 1850 as the excluded year. "Controls" includes all controls available in all censuses, including state fixed effects. "Controls with FE' includes also county fixed effects. Vertical lines are 95-percent confidence intervals. Observations weighted to correct for selection into linkage.

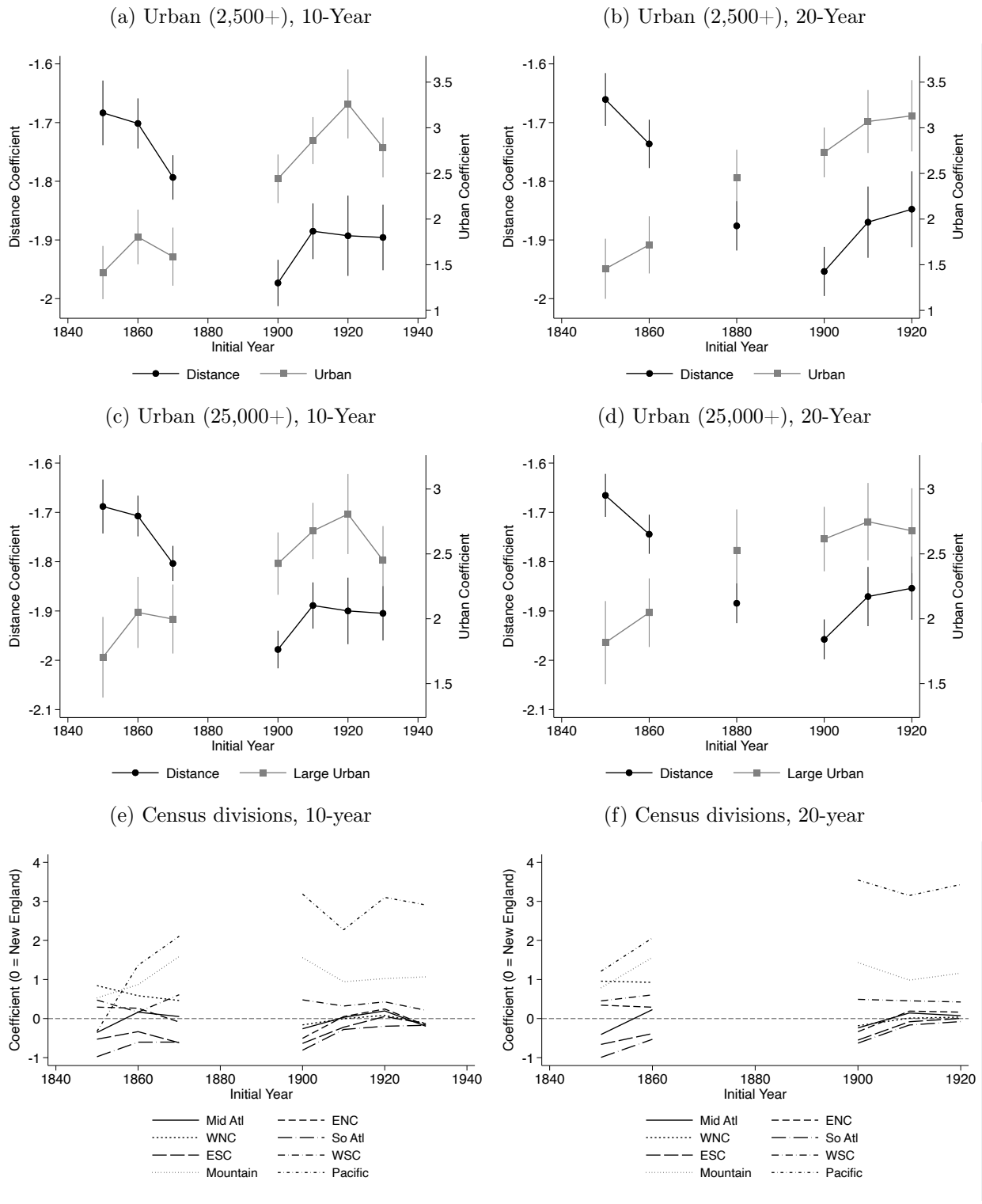


Figure D.4.9: Conditional Logit Results

Note: Panels (a)–(d) present coefficients on distance and the measure of urban residence from the conditional logit estimation. Panels (e) and (f) present the coefficients from the regressions of panels (a) and (b), respectively, on the census division indicators, with New England excluded.

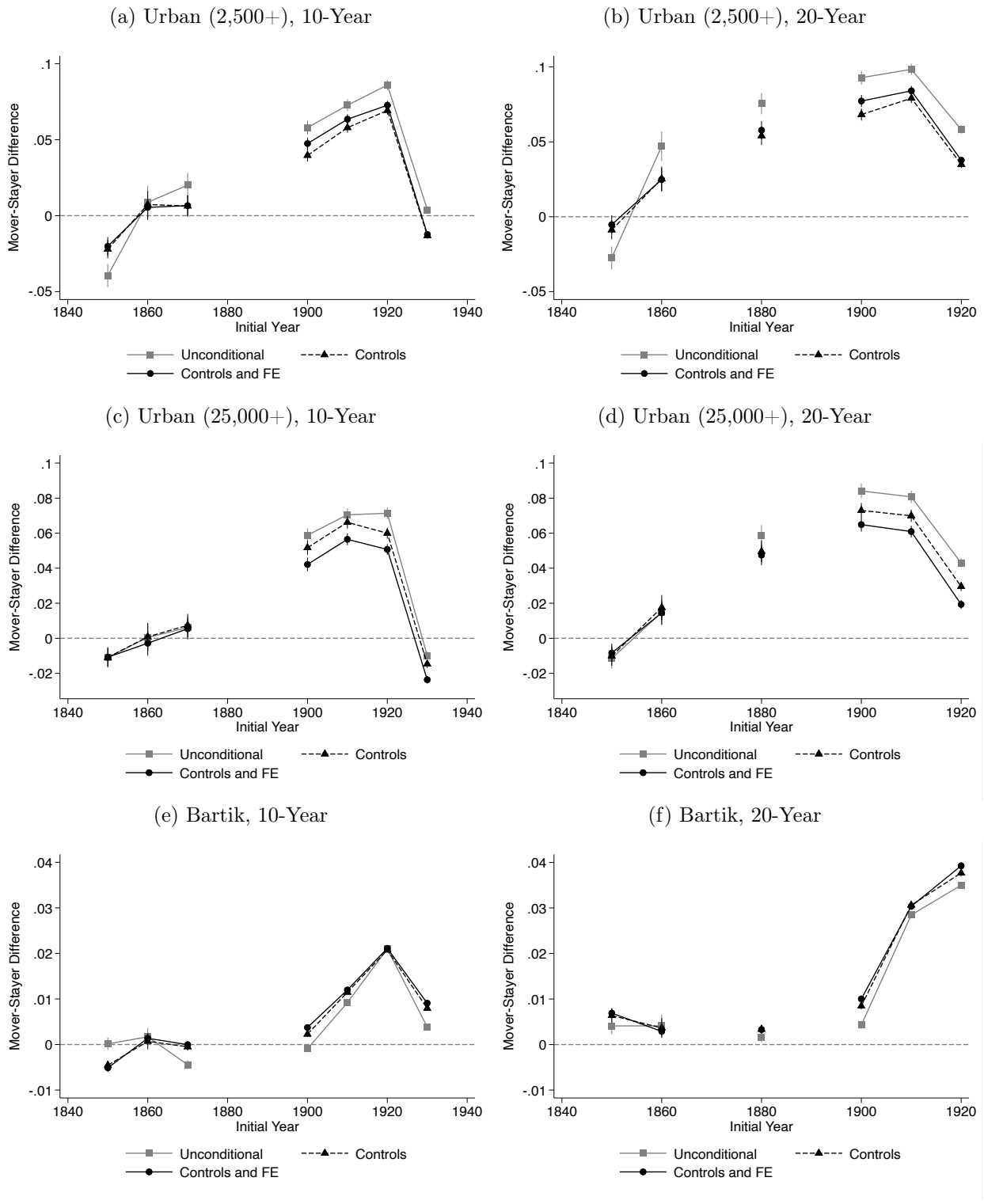


Figure D.4.10: Movers' change in urbanization and Bartik

Note: This figure compares the growth in urbanization or labor demand experienced by movers to that experienced by stayers, with either no controls, controlling for all observables, and controlling for all observables and county fixed effects.

E Results dropping second-generation immigrants

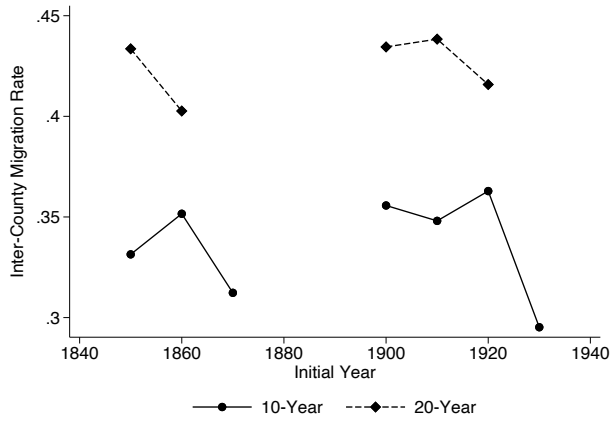


Figure E.1: Inter-county migration rates by span, corrected for false matches

Note: This 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 the main 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.

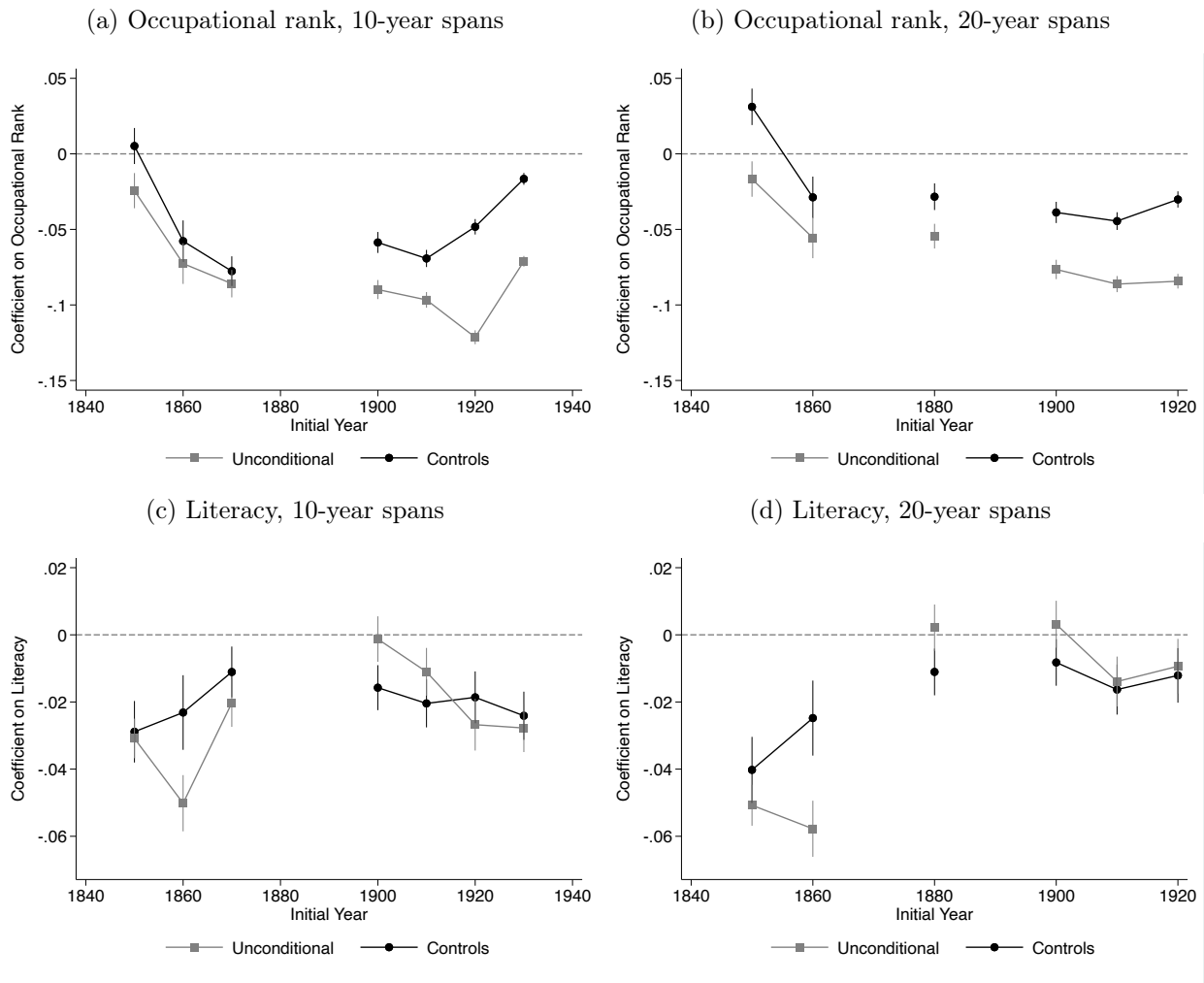


Figure E.2: Migrant selection by measure and span

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as county fixed effects.

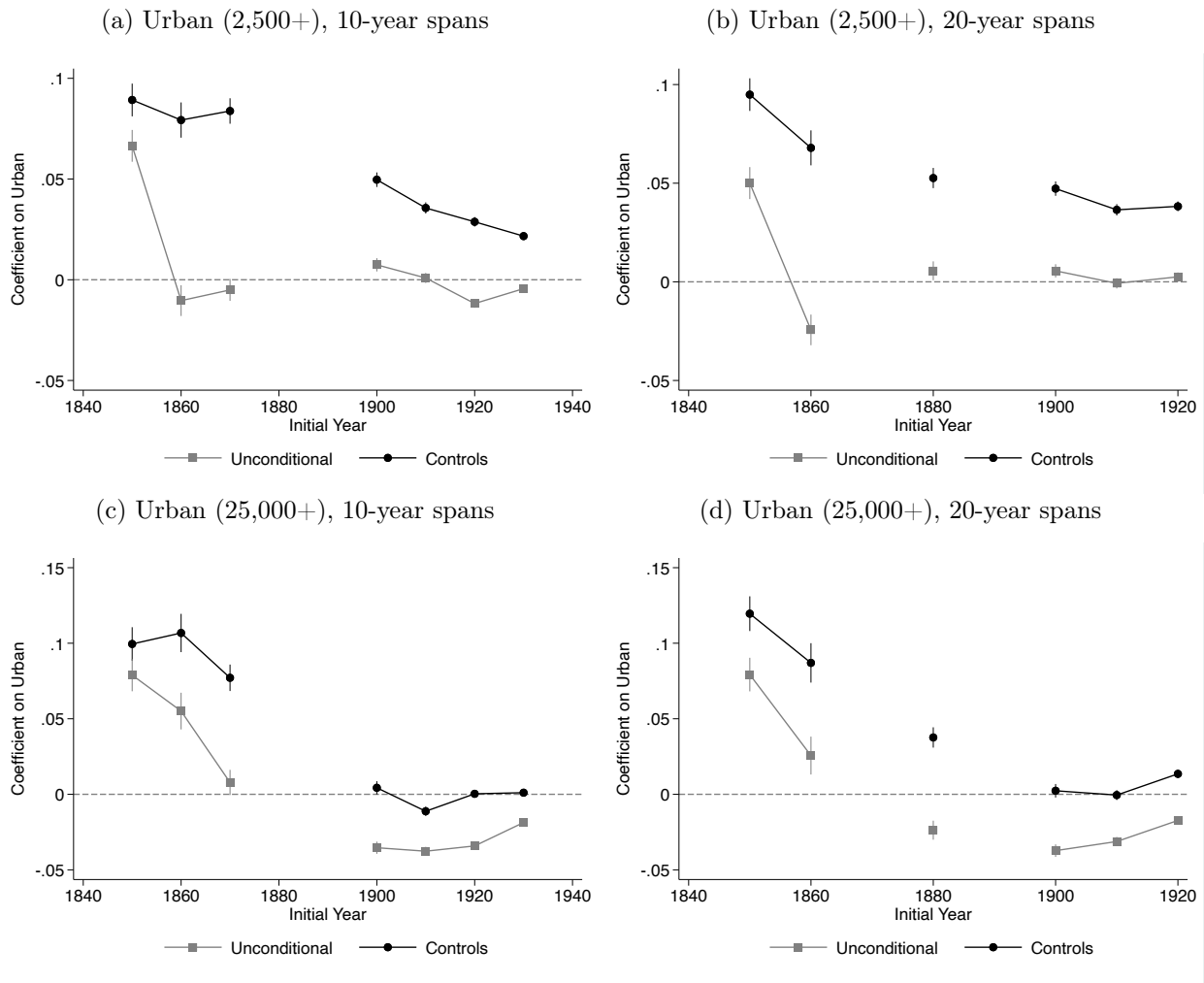


Figure E.3: Migrant selection by measure and span

Note: Each figure shows the coefficient on the variable in question in a selection regression for migration over the stated span. All observations are weighted by inverse linkage probability. Vertical lines indicate 95-percent confidence intervals. Controls are all controls available in all census years, as well as state fixed effects.

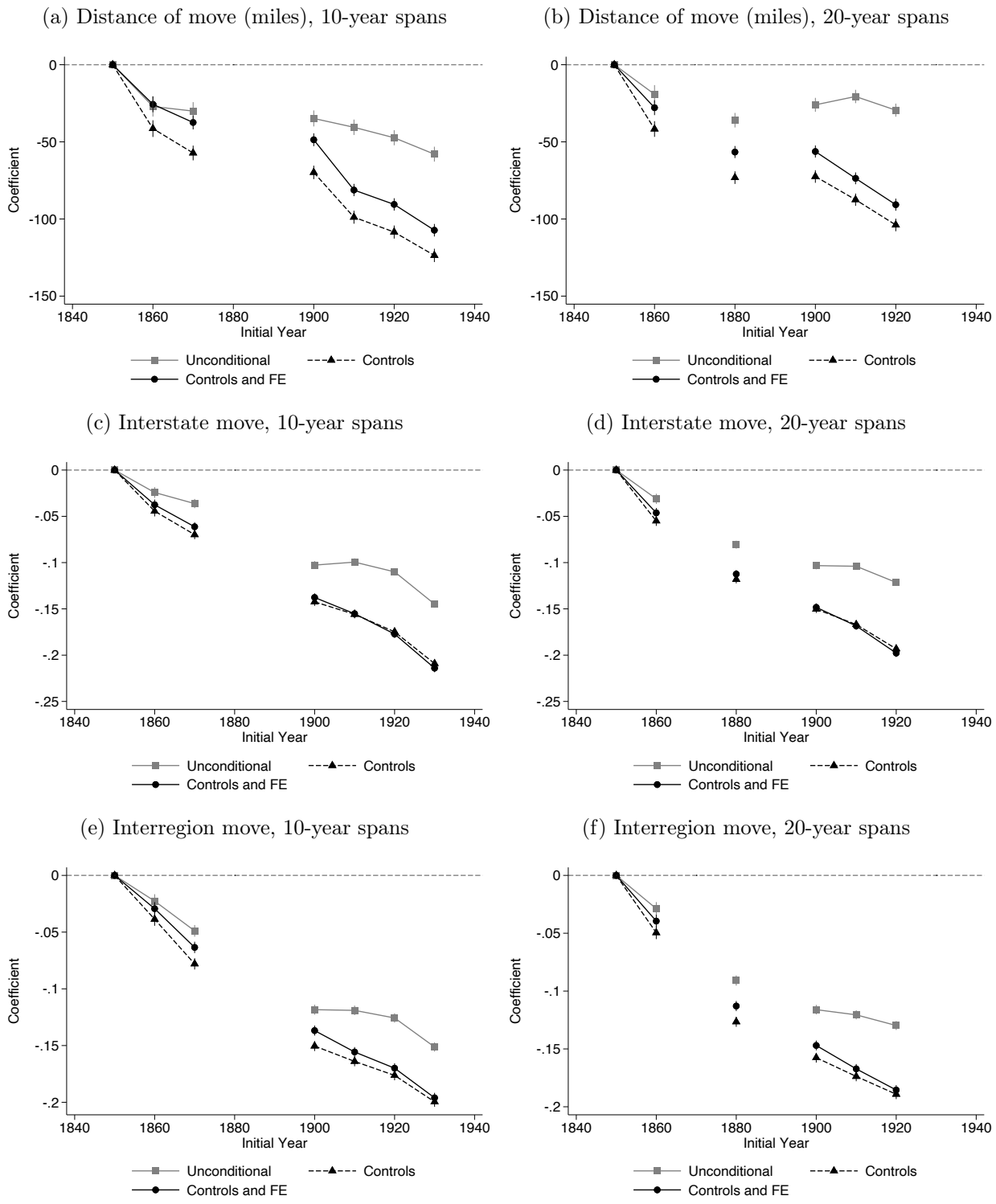


Figure E.4: Changes in distance of move

Note: These figures show the coefficients on initial-year fixed effects in regressions of destination characteristics with and without controls, with 1850 as the excluded year. "Controls" includes all controls available in all censuses, including state fixed effects. "Controls with FE' includes also county fixed effects. Vertical lines are 95-percent confidence intervals. Observations weighted to correct for selection into linkage.

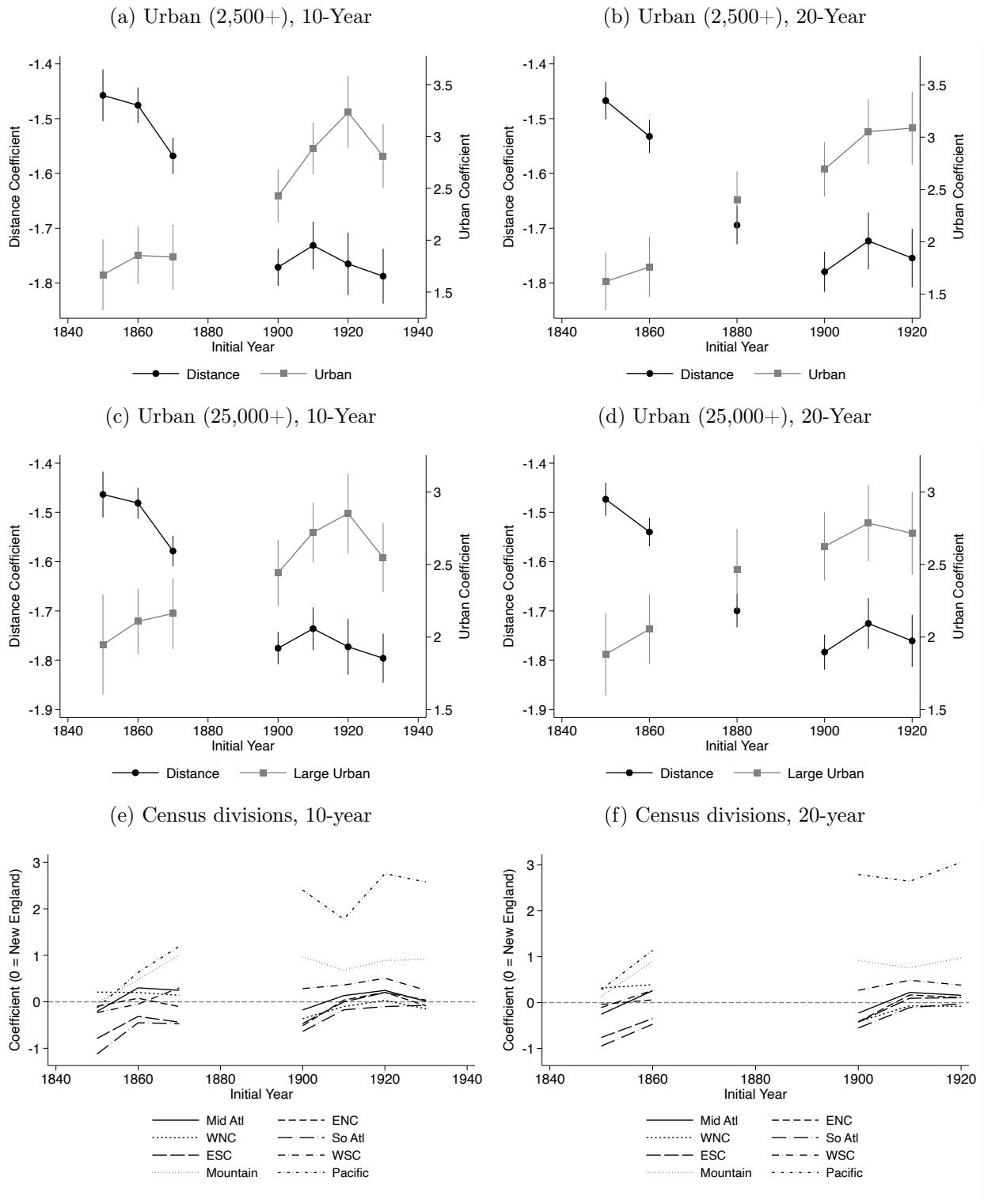


Figure E.5: Conditional Logit Results

Note: Panels (a)–(d) present coefficients on distance and the measure of urban residence from the conditional logit estimation. Panels (e) and (f) present the coefficients from the regressions of panels (a) and (b), respectively, on the census division indicators, with New England excluded.

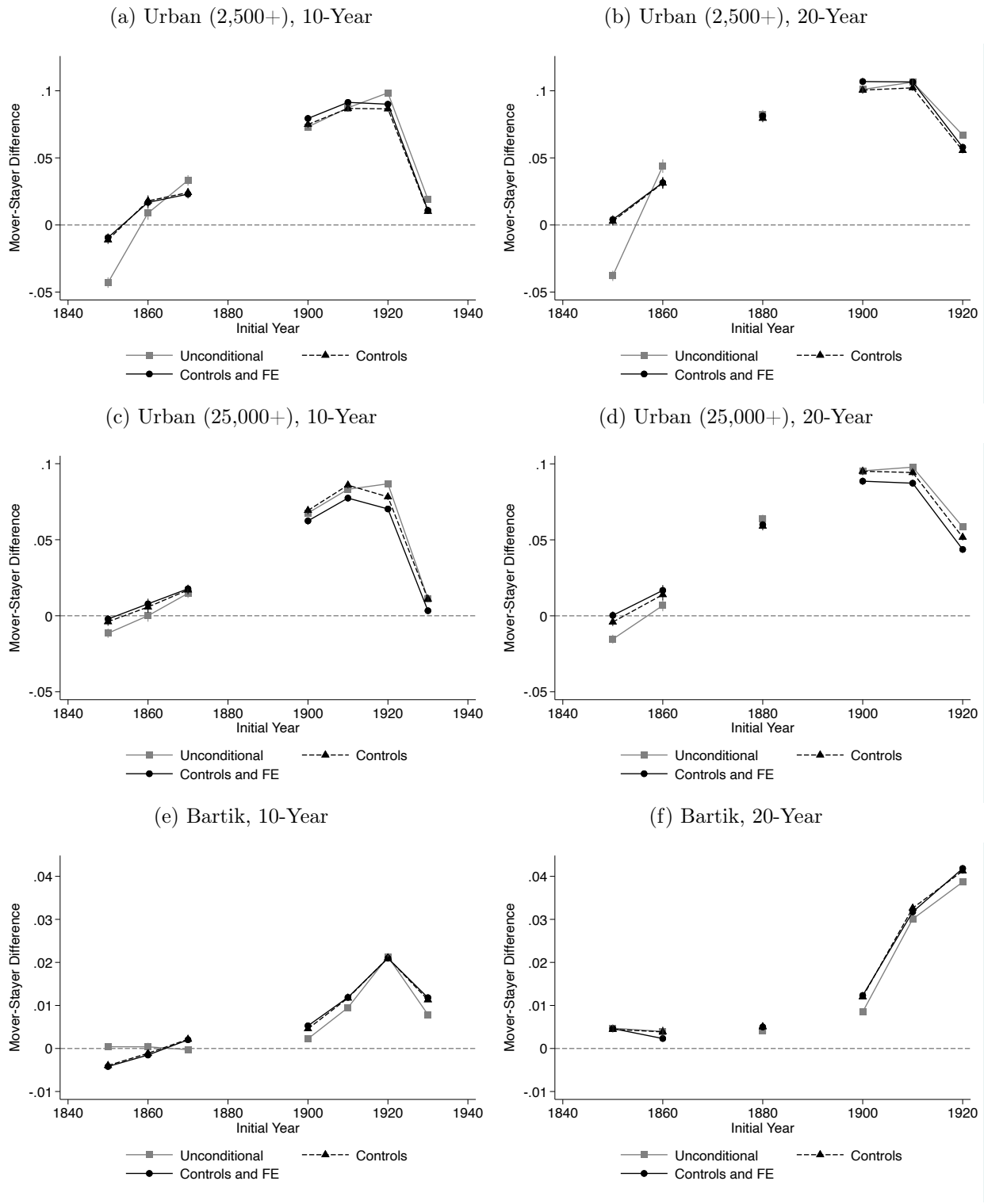


Figure E.6: Movers' change in urbanization and Bartik

Note: This figure compares the growth in urbanization or labor demand experienced by movers to that experienced by stayers, with either no controls, controlling for all observables, and controlling for all observables and county fixed effects.

References

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 (2022). *Miscellaneous US Census Crosswalks, 1850–1930*. Ann Arbor: Inter-university Consortium for Political and Social Research [distributor]. doi:10.3886/E171861V2.