Ethnic Endogamy after Settling Down for Several Generations: Evidence from the 1930 U.S. Census*

Dafeng Xu†

Abstract

Many studies argue that ethnic endogamy declines by generation; on the other hand, does ethnicity still relate to native-immigrant marriage after several generations of settlement? Based on linguistic origins of surnames, I identify ten (non-British-related) major ethnic origins among native-born men whose parents were also native-born in the 1930 U.S. census. Thus, I study the sample in which individuals' families had settled down in the U.S. for at least three generations. Results suggest ethnic endogamy still existed among some—although not all—ethnic groups: native-born men of Italian, Hispanic, German, Polish, and Russian ancestry were significantly more likely to marry first-generation immigrants who had the same origin with them.

Keywords: ethnic endogamy, marriage, generation, immigration, surname

*I have benefited from the comments of Francine Blau, Nancy Brooks, Jack Dewaard, Dave Hacker, Lawrence Kahn, Dan Lichter, Matt Nelson, Evan Roberts, Tony Smith, Matt Sobek, Rob Warren, and seminar participants at Urban Economic Association annual meeting, North American Regional Science Congress meeting, European Regional Science Association annual meeting, Cornell University, and the University of Minnesota.

†225 19th Ave. S., University of Minnesota, Minneapolis, MN 55455. Email: dafengxu@umn.edu.

1 Introduction

Scholars have long discussed immigrants' marital assimilation (e.g., Gordon, 1964; Pagnini and Morgan, 1990; Kalmijn, 1998; Bisin and Verdier, 2000; Qian and Lichter, 2007; Bleakley and Chin, 2010; Abramitzky et al., 2016) as marriage outcomes further relate to social and economic outcomes (e.g., Meng and Gregory, 2005; Zimmerman, 2007). Immigrants are generally more likely to marry immigrants, and recent immigrants—many of whom are from Latin America, the Caribbean, and Asia—are considered to be less assimilated than European immigrants in the age of mass migration in terms of intermarriage (Waters and Jiménez, 2005).

Scholars propose the three-generation model to explain language assimilation by generation (Alba et al., 2002). Similarly, ethnic endogamy declines by generation (Kalmijn, 1998). But does ethnicity still matter after immigrant families have arrived for many years? Using the 1930 U.S. census, I study marital choices of native-born men whose families had settled down in the U.S. for at least three generations and examine the likelihood of marrying first-generation immigrants of their origin.

Most studies use the birthplace or language to identify ethnicity (e.g., Pagnini and Morgan, 1990; Sassler, 2005), which is, however, impossible in this paper, as both individuals and their parents were native-born. Instead, I identify ethnicity based on linguistic origins of surnames. This is based on the idea from human biology that surnames reveal information about genetic and cultural transmission in the population (Guglielmino et al., 2000).

In a similar context, Logan and Shin (2012) use the IPUMS 1880 - 1910 linked census and show that most "native ethnics" had lower rates of marital assimilation than natives who had arrived for more than three generations. Compared with their data, I use a much large sample—full-count census—to identify third-generation ethnics. One methodological contribution of this paper is that I use surname information to identify ethnicity in the census. Some economic and sociological studies use surnames to determine ethnicity and analyze, e.g., productivity by ethnicity (Foley and Kerr, 2013) and labor market discrimi-

nation (Oreopoulous, 2011; Widner and Chicione, 2011). This paper employs the idea of surname-based identified ethnicity in historical demography, as many surveys do not have questions about ethnicity, or ethnicity is not perfectly documented.

This paper finds some evidence of ethnic endogamy among native ethnics whose families had settled down in the U.S. for several generations. In particular, native-born men with typical Italian, Hispanic, German, Polish, and Russian surnames were significantly more likely to marry female immigrants born in Italy, Hispanic countries, Germany, Poland, and Russia, respectively. The magnitude of endogamy was especially large among Hispanic, German, and Polish ethnics. Therefore, ethnicity could still play a crucial role in determining marital choices for some native ethnics even after their families had arrived in the country for many generations.

In the rest of the paper, Section 2 briefly introduces the historical background. Section 3 presents data, empirical models, and results. Section 4 concludes the paper.

2 Historical Background

During the age of mass migration (1850s - 1920s), the U.S. absorbed over 20 million immigrants from Europe, and first-generation immigrants generally constituted more than 10% of the population, but the countries of origin varied substantially across cohorts.

Table 1 presents the number of immigrants by country of birth, retrieved from the 1850, 1880, 1910, 1920, and 1930 full-count U.S. census.[1] The 1930 census shows that Germany, Poland, Italy, Russia, and Canada were top five sending countries of U.S. immigrants; four of them were non-English-speaking countries. Germany and Canada constantly sent large immigrant populations to the U.S.; although there were many first-generation German-born and Canadian-born immigrants in 1930, many natives of German and Canadian ancestry had settled down in the U.S. for several generations. There were very few Italian and Rus-

[1]The 1890 census documents were destroyed in a fire in 1921; the 1860, 1870, and 1900 full-count census are not yet available (Ruggles et al., 2015).

Table 1: Immigrants by Country of Birth

	1850	1880	1910	1920	1930
Italy	3,981	44,466	1,351,570	1,609,343	1,788,943
Germany	603,043	1,938,236	2,505,833	1,609,910	1,632,840
Canada	148,850	716,178	1,254,426	1,217,330	1,398,965
Poland	2,757	65,643	34,541	1,134,782	1,258,099
Russia	891	32,432	1,562,134	1,451,717	1,197,729
Ireland	998,625	1,853,361	1,356,439	1,050,633	929,511
England	287,769	664,743	889,389	813,325	830,711
Mexico	13,378	68,619	227,172	496,650	650,974
Japan	1	296	133,255	82,433	72,065
% foreign-born	9.7%	13.3%	14.7%	13.2%	11.6%

Source: 1850, 1880, 1910, 1920, and 1930 full-count U.S. census.

sian immigrants in 1880, but Italy and Russia became top sending countries of immigrants in 1910. Hence, many immigrants arriving during 1880 and 1910 were originally from Italy and Russia, and there should be many third-generation Italian and Russian immigrants (as well as second- and first-generation immigrants from these countries) in 1930. This is similar for the Mexican and Japanese population in the U.S. Note that although most Polish immigrants appeared to arrived in the 1910s, it could be due to that Poland regained independence only in 1918, and many earlier Polish immigrants reported either Germany or Russia as the country of birth in earlier censuses.

Many studies compare recent U.S. immigrants with European immigrants in history and argue that immigrants from Europe in the 19th and early 20th century probably assimilated into the U.S. society faster (e.g., Waters and Jiménez, 2005). But assimilation trajectories varied across countries of birth in the era of mass European immigration as well. In general, immigrants from Western Europe—such as Germany and Belgium—were most assimilated, while Southern and Eastern European immigrants assimilated at substantially lower rates, and such patterns of economic and social assimilation persisted even in the second generation (Abramitzky et al., 2014, 2016).

When immigrant families have settled down in the U.S. for more generations, however, they are likely to be nearly fully assimilated in terms of social and economic outcomes.

This could be especially true for European immigrants in the 19th and early 20th century. First, second-generation immigrants—who were "white ethnics" (Logan and Shin, 2012)—became ethnic majorities in the U.S. Second, the U.S. once was highly socially mobile (Ferrie, 2005), hence occupational disadvantages might diminish by generation. Third, second- and third-generation immigrants received education in the U.S. and thus had adequate language skills in the labor market (Waters and Jiménez, 2005).

This paper studies a question related to marital assimilation: ethnic endogamy among native-born men whose families have settled down in the U.S. for many generations. Ethnic endogamy appears to be common among minorities (Kalmijn, 1998; Qian and Lichter, 2007), such as the Asian, Black, and Hispanic population. In 1930, however, most third-generation immigrants were non-Hispanic whites, which were also the majority population in the U.S. Were native-born men of, say, German ancestry more likely to marry female immigrants born in Germany? If so, then ethnicity could still play a crucial role in determining marital choices of U.S. majorities whose ancestors had long arrived in the country.

3 Data and Analysis

This section introduces data, methods, and the empirical analysis. I first discuss the ethnicity identification method. I then present data and descriptive statistics. I finally conduct the empirical analysis of ethnic endogamy.

3.1 Ethnicity Identification

Most prior studies use birthplace or language to identify ethnicity (Pagnini and Morgan, 1990; Sassler, 2005), which is not applicable in this paper. I propose an alternative way to identify ethnicity: linking "training data" of typical surnames by language origin to census. For example, a native-born man *Napolitano*—a typical Italian name—should be of Italian descent even if his parents are also native-born.

A simple way to identify ethnicity based on surnames is to match surnames in census with training data of surnames by language (Mateos, 2007). I construct training data using Wikipedia language-specific surname categories (each category contains several hundred and thousand surnames). There are various record linkage algorithms for statistical software (Wasi and Flaaen, 2015), such as *reclink* (Blasnik, 2010). Such algorithms employ fuzzy matching strategies by comparing string distances (e.g., Jaro-Winkler distance: Jaro, 1989, and Winkler, 1990), which solve two potential problems in identification: (a) a non-Anglicized surname is Anglicized at moderate degrees but still keeps some ethnic-linguistic properties (e.g., *Eisenhauer* to *Eisenhower*); (b) a surname is misspelled as census data are digitized from images (e.g., *Schmidt* to *Schnndt*).

The limitation of linkage algorithms is that ethnicity cannot be identified if an Anglicized ethnic name loses its linguistic properties, as name localization is common among immigrants. Hence, only typical language-specific surnames can be identified, and thus individuals with identified ethnicity only constitute a sub-sample of the full population. Still, the empirical analysis of this paper could suggest ethnic endogamy for those who keep typical ethnic-sounding surnames.

3.2 Data and Descriptive Statistics

This paper uses the 1930 U.S. census male sample. I do not include women because they might change surnames after marriage. In the male sample I only include those (a) who were married and had spousal information in 1930, and (b) whose parents were both native-born. Their families thus had settled down in the U.S. for at least three generations.

I do not identify English-language surnames as many non-English-speaking immigrants Anglicized their surnames, and there were various ethnic origins associated with English (e.g., British, Irish, Canada, Australian). I only identify typical surnames in: Danish, Norwegian, Swedish, French, Italian, Spanish, German, Polish, Russian, and Japanese. This particularly excludes natives of British, Irish, and Canadian ancestry, which constituted the

Table 2: Descriptive Statistics

	Sample size	Age	Age, first marriage	Occupational score	Immigrant spouse	Same-origin spouse
Full sample	14,780,341	42.048	24.339	19.218	0.022	—
		(13.804)	(6.510)	(13.442)	(0.146)	
Unidentified ethnicity	13,474,883	42.136	24.342	19.201	0.022	—
		(13.837)	(6.520)	(13.441)	(0.146)	
Danish	140,892	41.153	24.043	18.184	0.015	0.0003
		(13.692)	(6.762)	(12.301)	(0.122)	(0.0162)
Norwegian	31,029	40.812	24.230	19.067	0.024	0.0013
		(13.566)	(6.504)	(13.289)	(0.152)	(0.0372)
Swedish	39,677	41.159	24.226	19.026	0.020	0.0012
		(13.655)	(6.537)	(13.181)	(0.143)	(0.0344)
French	219,930	41.787	24.234	19.099	0.024	0.0005
		(13.702)	(6.475)	(13.311)	(0.152)	(0.0222)
Italian	137,179	41.866	24.300	19.010	0.023	0.0009
		(13.808)	(6.620)	(13.322)	(0.151)	(0.0303)
Hispanic	52,181	40.423	23.959	17.471	0.046	0.0099
		(13.482)	(6.537)	(11.945)	(0.210)	(0.0164)
German	649,685	40.856	24.416	20.036	0.024	0.0052
		(13.132)	(6.215)	(13.831)	(0.155)	(0.0720)
Polish	8,398	39.040	24.130	19.566	0.035	0.0079
		(13.149)	(6.353)	(13.613)	(0.185)	(0.0883)
Russian	19,022	42.103	24.426	19.528	0.025	0.0007
		(13.745)	(6.453)	(13.970)	(0.156)	(0.0271)
Japanese	7,465	40.694	24.134	18.933	0.024	0.0001
		(13.393)	(6.713)	(13.212)	(0.153)	(0.0116)

Standard deviations are in parentheses.

majority of the population in 1930. Indeed, only more than 10% of individuals in the sample were ethnically identified. That said, there were still over 1 million individuals in the sample whose surnames kept ethnic-sounding.

The average age in the full sample was 42 years old. Individuals with identified ethnicity were slightly younger. The average age of first marriage was 24.3 years old, and there was almost no difference between individuals with and without identified ethnicity. Similarly, the differences in occupational score—the primary economic measure in the 1930 census—across groups were small. The rate of marrying the foreign-born spouse was also similar across groups, although Hispanic and Polish ethnics were significantly more likely to marry immigrants. Finally, among individuals with identified ethnicity, Hispanic, German, and Polish ethnics were more likely to marry immigrants of their origin.

3.3 Empirical Analysis

I now examine the marital choice of individual i of ethnicity j who lived in enumeration district k. The enumeration district was an area covered by one enumerator and, on average, contained less than 2,000 inhabitants and was the smallest geographic unit in the 1930 census. I run the following linear probability model (LPM):

$$I_j = \alpha + \beta' \mathbf{E}_{ijk} + \gamma' \mathbf{X}_{ijk} + \tau_k + \varepsilon \tag{1}$$

where I_j is an indicator of marriage with an immigrant born in country j. $\mathbf{E}$ is the vector of ethnicity, while those of unidentified ethnicity were in the "unidentified group". $\mathbf{X}$ is the vector of individual characteristics such as age and age of first marriage. I control for geographic factors as immigrants have unique settlement patterns (e.g., Massey and Denton, 1985; Bartel, 1989): here τ_k is the enumeration district dummy. I also cluster the standard errors at enumeration district level.

Table 3: Ethnic Endogamy: Full Sample (1)

Spouse birthplace:	Denmark	Norway	Sweden	France	Italy	Hispanic
Danish	−0.0001	0.0001	0.0002*	1.35e-05	6.68e-06	−3.10e-06
	[−1.42]	[1.71]	[1.98]	[0.26]	[0.20]	[−0.55]
Norwegian	−1.31e-06	0.0004	0.0004*	−0.0002	−9.92e-06	−0.0002
	[−0.01]	[1.78]	[1.96]	[−1.71]	[−0.11]	[−1.33]
Swedish	0.0001	0.0002	0.0003	−0.0001	0.0001	−0.0001
	[0.59]	[1.54]	[1.71]	[−0.79]	[1.43]	[−1.10]
French	−3.72e-05	−1.29e-05	−0.0001	−4.56e-05	0.0001	−1.72e-05
	[−1.04]	[−0.29]	[−1.09]	[−0.93]	[1.56]	[−0.36]
Italian	−4.03e-05	−1.37e-05	5.12e-06	0.0001	0.0006***	0.0006***
	[−0.88]	[−0.23]	[0.06]	[1.48]	[7.73]	[5.76]
Hispanic	−1.93e-05	3.04e-05	−0.0002*	3.60e-05	0.0003*	0.0201***
	[−0.23]	[0.32]	[−2.12]	[0.30]	[2.43]	[21.88]
German	−1.13e-05	−7.74e-06	−0.0001*	−3.03e-06	−4.04e-05*	−0.0001***
	[−0.48]	[−0.26]	[−2.03]	[−0.10]	[−2.08]	[−6.27]
Polish	−0.0003*	−0.0003	−0.0001	−0.0004*	−0.0001	0.0004
	[−2.00]	[−1.45]	[−0.41]	[−2.26]	[−0.07]	[1.03]
Russian	0.0001	−2.64e-05	−4.73e-05	2.34e-05	0.0001	−0.0001
	[0.66]	[−0.17]	[−0.21]	[0.13]	[0.53]	[0.65]
Japanese	−0.0003***	−0.0001	−0.0003	0.0003	0.0003	0.0015*
	[−12.06]	[−0.22]	[−0.95]	[0.86]	[0.97]	[2.59]
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.018	0.029	0.019	0.012	0.022	0.071

Enumeration district fixed effects are included. Observations: 14,780,341.
t-statistics are in brackets. Standard errors are clustered at enumeration district level.

Table 4: Ethnic Endogamy: Full Sample (2)

Spouse birthplace:	Germany	Poland	Russia	Japan	Foreign-born
Danish	−0.0001	−2.44e-05	−7.88e-05*	−1.55e-06	−0.0006*
	[−1.08]	[−0.74]	[−2.46]	[−0.21]	[−2.02]
Norwegian	−0.0004	−0.0001*	0.0006**	1.92e-05	−0.0007
	[−1.45]	[−2.12]	[3.26]	[0.58]	[−0.80]
Swedish	0.0001	4.74e-05	0.0003	1.37e-05	−0.0001
	[0.22]	[0.56]	[1.26]	[0.54]	[−1.19]
French	0.0001	2.81e-06	1.40e-05	5.87e-06	0.0010**
	[1.08]	[0.08]	[0.37]	[0.63]	[2.98]
Italian	0.0001	0.0002*	−3.82e-05	9.31e-06	0.0009*
	[0.45]	[2.80]	[−0.83]	[0.74]	[2.25]
Hispanic	−0.0006**	0.0001	−0.0001	1.10e-05	0.0183***
	[−2.92]	[0.85]	[−1.73]	[0.57]	[16.93]
German	0.0011***	4.75e-05	0.0003***	−2.12e-06	0.0002
	[12.29]	[1.86]	[8.95]	[−0.51]	[0.82]
Polish	0.0005	0.0058***	0.0002	−1.05e-05**	0.0050*
	[0.63]	[6.19]	[0.55]	[−2.60]	[2.53]
Russian	−0.0009*	0.0001	0.0004*	−1.11e-05***	0.0005
	[2.42]	[0.78]	[1.97]	[−5.24]	[0.42]
Japanese	−0.0014**	−2.23e-05	0.0002	8.51e-05	0.0015
	[−3.19]	[−0.14]	[0.78]	[0.91]	[0.90]
Other controls	Yes	Yes	Yes	Yes	Yes
R^2	0.019	0.024	0.026	0.014	0.057

Enumeration district fixed effects are included. Observations: 14,780,341.
t-statistics are in brackets. Standard errors are clustered at enumeration district level.

Table 3 and 4 present empirical results of marital choices in the full sample, in which individuals were native-born and their parents were also native-born. I find no evidence that individuals with typical Danish, Norwegian, Swedish, French, and Japanese surnames—hence were possibly associated with these ethnic origins—were more likely to marry immigrants of the same origin. However, I do find evidence of endogamy among native-born men associated with Italian, Hispanic, German, Polish, and Russian ethnicity. Specifically, the LPM results suggest third-(or-more)-generation Hispanic Americans were 2% more likely to marry first-generation immigrants born in Hispanic countries. This magnitude was large given that Mexico and other Latin American countries, compared with many European countries, were far from the top sending countries of U.S. immigrants in the early 20th century (see Table 1). The magnitude of endogamy was also fairly large among German and Polish ethnics.

I rerun the above models in Table 5, using the sub-sample of individuals that had identified ethnicity. The results of Table 3 and 4—including both the qualitative pattern and en-

Table 5: Ethnic Endogamy: Sample of Identified Ethnicity

Ethnicity:	Danish	Norwegian	Swedish	French	Italian
Spouse of the	−0.0001	0.0004	0.0003	−0.0001	0.0006***
same origin	[−1.38]	[1.76]	[1.85]	[−1.12]	[6.51]
R^2	0.140	0.173	0.138	0.096	0.145
Ethnicity:	Hispanic	German	Polish	Russian	Japanese
Spouse of the	0.0183***	0.0011***	0.0054***	0.0002	1.263-05
same origin	[21.06]	[8.23]	[6.09]	[0.91]	[0.246]
R^2	0.191	0.108	0.174	0.164	0.111
Other controls	Yes	Yes	Yes	Yes	Yes

Enumeration district fixed effects are included. Observations: 1,305,458.

t-statistics are in brackets. Standard errors are clustered at enumeration district level.

dogamy magnitudes—are again presented, except that Russian endogamy is not observed. Table 5 suggests the empirical conclusion of this paper is robust to sample change, and is not driven by results of ethnicity identification.

4 Conclusion

Many studies hypothesize the large behavioral differences across generations among immigrants. Second-generation immigrants are less likely to marry within ethnic groups (e.g., Kalmijn, 1998). On the other hand, does ethnicity still matter after several generations of settlement? Using the 1930 U.S. census, I focus on native-born men whose parents were also native-born, and identify their ethnicity based on linguistic origins of surnames. Their families thus had settled down in the U.S. for at least three generations. I then examine whether these natives with identified ethnicity were more likely to marry first-generation immigrant women of the same origin.

Results show that ethnic endogamy did exist among some—although not all—ethnic groups in the U.S. Third-generation ethnics of Italian, Hispanic, German, Polish, and Russian origin were significantly more likely to marry immigrants of their origin. The magnitude of endogamy was particularly large among the Hispanic, German, and Polish group. This suggests that ethnicity might still play an important role in determining marital choices even if individuals' ancestors had long arrived in the country.

Conflicts of Interest

The author declares no conflict of interest in this paper.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy*, 122(3), 467 - 506.

[2] Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2016. "Cultural Assimilation during the Age of Mass Migration." NBER Working Paper No. 22381.

[3] Alba, Richard, John Logan, Amy Lutz and, Brian Stults. 2002. "Only English by the Third Generation? Loss and Preservation of the Mother Tongue among the Grandchildren of Contemporary Immigrants." *Demography*, 39(3), 467 - 484.

[4] Bartel, Ann P. 1989. "Where Do the New U.S. Immigrants Live?" *Journal of Labor Economics*, 7(4), 371 - 391.

[5] Blasnik, Michael. 2010. "RECLINK: Stata Module to Probabilistically Match Records". Statistical Software Components.

[6] Bleakley, Hoyt, and Aimie Chin. 2010. "Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants." *American Economic Journal: Applied Economics*, 2(1), 165 - 192.

[7] Ferrie, Joseph P. 2005. "History Lessons: The End of American Exceptionalism? Mobility in the United States since 1850." *Journal of Economic Perspectives*, 19(3), 199 - 215.

[8] Foley, C. Fritz, and William R. Kerr. 2013. "Ethnic Innovation and U.S. Multinational Firm Activity." *Management Science*, 59(7), 1529 - 1544.

[9] Gordon, Milton M. 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origins*. New York: Oxford University Press.

[10] Guglielmino, C.R., G., Zei, and L.L. Cavalli-Sforza. 2000. "Genetic and Cultural Transmission in Sicily as Revealed by Names and Surnames." *Human Biology*, 63(5), 607 - 627.

[11] Jaro, Matthew A. 1989. "Advances in Record Linking Methodology as Applied to Matching the 1985 Census of Tampa, Florida." *Journal of the American Statistical Association*, 84, 414 - 420.

[12] Kalmijn, Matthijs. 1998. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology*, 24, 395 - 421.

[13] Logan, John, and Hyoung-jin Shin. 2012. "Assimilation by the Third Generation? Marital Choices of White Ethnics at the Dawn of the Twentieth Century." *Social Science Research*, 41(5), 1116 - 1125.

[14] Massey, Douglas S., and Nancy A. Denton. 1985. "Spatial Assimilation as a Socioeconomic Outcome." *American Sociological Review*, 50(1), 94 - 106.

[15] Mateos, Pablo. 2007. "A Review of Name-Based Ethnicity Classification Methods and their Potential in Population Studies." *Proceedings of Australasian Language Technology Association Workshop*, 145 - 149.

[16] Meng, Xin, and Robert G. Gregory. 2005. "Intermarriage and the Economic Assimilation of Immigrants." *Journal of Labor Economics*, 23(1), 135 - 174.

[17] Oreopoulos, Philip. 2011. "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes." *American Economic Journal: Economic Policy*, 3(4), 148 - 171.

[18] Pagnini, Deanna L., and S. Philip Morgan. 1990. "Intermarriage and Social Distance Among U.S. Immigrants at the Turn of the Century." *American Journal of Sociology*, 96(2), 405 - 432.

[19] Qian, Zhenchao, and Daniel T. Lichter. 2007. "Social Boundaries and Marital Assimilation: Interpreting Trends in Racial and Ethnic Intermarriage." *American Sociological Review*, 72(1), 68 - 94.

[20] Sassler, Sharon. 2005. "Gender and Ethnic Differences in Marital Assimilation in the Early Twentieth Century." *International Migration Review*, 39(3), 608 634.

[21] Wasi, Nada, and Aaron Flaaen. 2015. "Record Linkage using STATA: Pre-processing, Linking and Reviewing Utilities." *Stata Journal*, 15(3), 672 - 697.

[22] Waters, Mary C., and Tomás R. Jiménez. 2005. "Assessing Immigrant Assimilation: New Empirical and Theoretical Challenges." *Annual Review of Sociology*, 31, 105 - 125.

[23] Widner, Daniel, and Stephen Chicoine. 2011. "Its All in the Name: Employment Discrimination Against Arab Americans." *Sociological Forum*, 26(4), 806 823.

[24] Winkler, William E. 1990. "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage." *American Statistical Association, 1990, Proceedings of the Section of Survey Research Methods*, 354 - 359.

[25] Zimmerman, Klaus, F. 2007. "The Economics of Migrant Ethnicity." *Journal of Population Economics*, 20(3), 487 - 494.