Who Benefits from Citizenship? Heterogeneous Effects of Naturalization Decisions in the Early Twentieth-Century United States

Peter Catron (Department of Sociology, University of Washington)

Nathan I. Hoffmann (Department of Sociology, Emory University)

Mingze Li (Department of Sociology, University of Washington)

Abstract

While existing research highlights the occupational advantages of citizenship for immigrants, little is known about how such advantages might vary, and previous studies have not fully dealt with issues of selection. We study the heterogeneous effects of naturalization by leveraging a novel historical dataset of naturalization records from New York City's Southern District in the early 1900s. We link these to U.S. census data from 1920, 1930, and 1940, tracking 1,947 immigrants who declared their intent to naturalize, analyzing occupational trajectories among them, and modeling selection into naturalization using random forests. Our findings reveal that immigrants who completed the naturalization process attained significantly higher occupational status than those who initiated but did not complete naturalization. In addition, propensity-stratified results support a negative selection hypothesis: Immigrants less likely to naturalize enjoyed greater returns to naturalization, surpassing the unnaturalized in the high-propensity, more advantaged group. We support these results with supplementary analyses of full-count census data, where we find similar trends. These results underscore the role of naturalization as a key driver of socioeconomic mobility while highlighting citizenship as a critical stratifying force.

Introduction

Citizenship acquisition is often promoted as both an endpoint of assimilation and a means for advancing it (OECD 2011). Citizenship represents formal inclusion and provides benefits, including social and residency rights, legal protections, full access to the labor market, and feelings of membership and belonging. Consequently, there is a large literature on who naturalizes and why (Bloemraad 2006; DeVoretz and Irastorza 2017; Waldinger, Hoffmann, and Catron 2023; Yang 1994) as well as whether citizenship yields economic benefits (Bratsberg, Ragan, and Nasir 2002; Catron 2019, 2021; Hainmueller, Hangartner, and Ward 2019; Hainmueller et al. 2019; Peters, Vink, and Schmeets 2018; W. Scott 2008). In these studies, various individual and contextual variables are associated with citizenship acquisition. Citizenship is then assumed to offer greater labor market advantages for those who complete the process of naturalization. Few studies, however, seek to understand if selection into citizenship shapes whether and for whom citizenship yields benefits. In this article, rather than focus on average outcomes of citizenship, we seek to understand the heterogeneous effects of naturalization and how barriers to citizenship impact outcomes of this status.

Who benefits from naturalization? In this study, we return to the Age of Mass Migration (1880-1924) to assess the heterogeneous effects of naturalization. While today's immigrants may enter on a vast array of visa statuses, some of which allow for naturalization while otherst do not, yesterday's immigrants all entered the U.S. with the same legal protections (Abramitzky et al. 2021). Focusing on immigrants of an earlier era allows us to better isolate the probabilities of who becomes a citizen and who benefits from that status. We create a new panel dataset that follows individuals from the moment they declare their intent to naturalize through their working career between 1920 and 1940. We webscraped and transcribed declaration of intention and petition

forms of immigrants in New York City that we then link to U.S. census records. This dataset enables us to understand both who becomes a citizen based on demographic, economic, and geographic characteristics of the individual and whether a person's propensity to complete the naturalization process influences any economic benefits it may yield. We supplement our analysis of this original dataset with examination of linked full-count census data. In studying who benefits most from naturalization, we build on a growing literature examining heterogeneous returns to a variety of sociological experiences (Brand et al. 2021; Xie, Brand, and Jann 2012; Hu 2023), such as obtaining a bachelor's degree (Brand and Xie 2010; Brand and Davis 2011; Cheng et al.), parental divorce (Brand et al. 2019), or social mobility (Luo 2021).

In both our original dataset and the full-count census data, we find that on average all individuals benefit from citizenship, with effects holding in fixed-effects models that control for time-varying characteristics. However, we also find that the returns to naturalization vary greatly by propensity to naturalize. We use a random forest model to estimate propensity to naturalize and stratify the sample by those most and least likely to naturalize, finding much greater returns to naturalization for those least likely to naturalize. The low-propensity group begins in less lucrative occupations, but those who naturalize in this group catch up to high-propensity naturalizers and surpass the high-propensity unnaturalized. Despite being less advantaged on characteristics such as literacy and English proficiency,those with low propensity to naturalize but who ultimately do naturalize enjoy a higher occupational premium over their working careers.

We argue that a negative selection hypothesis helps us understand who benefits most from naturalization. That is, individuals who have characteristics that make them less likely to naturalize are those who benefit most from citizenship acquisition. As a selective good, citizenship is costly, bureaucratic, and dependent on prior markers of integration such as socioeconomic status and

language proficiency. Since not all immigrants meet these requirements, citizenship is harder to obtain. However, because citizenship provides concrete occupational advantages through hiring and promotion (Hainmueller et al. 2019; Peters et al. 2018), the expected benefits of citizenship are highest for those least likely to obtain it. As citizenship is promoted as a means for full integration (OECD 2011), our study highlights just who benefits most from citizenship status.

Positive and Negative Selection Hypotheses of Citizenship

Throughout its history, with a few ethno-racial and gender restrictions, the United States has allowed noncitizens to become citizens through the process of naturalization. For most of this history, eligible immigrants needed to live in the country for five years, and have basic knowledge of English and civics, proof of good moral character, and a willingness to swear an oath of allegiance. In return for overcoming these relatively low barriers, naturalized immigrants are then afforded the same rights and privileges as individuals who have U.S. citizenship by being born on U.S. soil or to U.S.-citizen parents. Naturalization, therefore, has underpinned immigrants' strong intra- and intergenerational economic mobility for at least the last 150 years (Catron 2019; Kasinitz 2019; Smith 2024). Citizenship's benefits, however, are not universal for all individuals and groups, and indeed it does not always lead to advantages for some (Catron 2021; Hainmueller et al. 2019). Thus, understanding who benefits from naturalization can help scholars better understand integration processes over time.

To help understand how citizenship benefits may vary, we build on recent developments in the study of heterogeneous treatment effects. While sociologists have long been concerned with explaining variation in individuals' responses to similar social experiences, recent work has sought to formalize how pre-existing social characteristics may structure these diverse effects (Brand and

Thomas 2013; Xie, Brand, and Jann 2012). This framework for studying "heterogeneous treatment effects" or "heterogeneous returns" to social experiences has yielded a number of insights (Hu 2023). For example, these studies have often focused on the heterogeneous effects of a bachelor's degree, finding that individuals least likely to go to college tend to gain the greatest economic benefit (Brand and Xie 2010; Cheng et al. 2021), enjoy the greatest protection to their health (Schafer, Wilkinson, and Ferraro 2013), and see the greatest decrease in fertility (Brand and Davis 2011). Other applications of this framework have found that single mothers' job displacement (Brand and Simon Thomas 2014) and parental divorce (Brand et al. 2019) are most disruptive to children least likely to experience these. We extend this framework to immigration studies by studying returns to naturalization. In our case, propensity to naturalize may vary with the occupational benefits that immigrants would garner from naturalizing.

Implementing this framework in the context of naturalization leads us to consider who *accesses* citizenship in the first place. . On the one hand, immigrants who are most likely to naturalize may be the ones most likely to benefit from naturalization, an expectation we call the *positive selection hypothesis*. In our case, individuals who are more socioeconomically advantaged may be more likely to become a citizen, and the observed economic gains from citizenship may largely reflect prior advantages rather than the causal effect of naturalization itself. On the other hand, immigrants who are least likely to naturalize may experience the most transformative outcomes when they acquire citizenship, an expectation we call the *negative selection hypothesis*. Here, those least likely to naturalize face greater barriers to integration and becoming a citizen may produce more substantial changes in their social, economic, and political incorporation.

There is support in the literature for both of these possibilities. Supporting the positive selection hypothesis, common accounts of who becomes a citizen often focus on the characteristics

of individuals and the expected benefits they may achieve from becoming a citizen (Jasso and Rosenzweig 1986; Yang 1994; Gritti et al. 2025). In particular, personal resources, education, and English-language ability have been shown to positively influence both the propensity to naturalize and the speed with which it happens (Jasso and Rosenzweig 1986). In these accounts, built largely around rational choice theory, immigrants seek to maximize their economic interests, subject to constraints, and then predict future earnings to decide whether they naturalize (DeVoretz and Irastorza 2017; Schneider 2001; Gritti et al. 2025). These frameworks argue that immigrants who have more resources and education are more likely to become a citizen because they are most likely to obtain the benefits that naturalization confers. That is, immigrants who naturalize are positively selected wherein immigrants who share cultural, linguistic, or racial similarities with the native-born population may be more likely to benefit from citizenship because they are perceived to be more easily integrated into society. Naturalized immigrants are expected to do better in the labor market than unnaturalized immigrants due to these characteristics and not because of their citizenship attainment (Gritti et al. 2025).

Conversely, the negative selection hypothesis highlights how naturalization can be most consequential for immigrants with limited resources and greater structural disadvantages. These immigrants often face barriers to naturalization rooted in low levels of human capital, structural constraints, and cultural and language differences. For these migrants, citizenship removes fears of deportation, provides access to state benefits, gives access to occupations that provide stable employment, and enables full participation in civic life. In contrast, highly educated, economically secure immigrants who already have a short social distance to the native-born may only gain marginal additional benefits from formal naturalization. Thus, the expected utility of citizenship may be highest precisely for those least likely to obtain it (Hainmueller et al. 2017). A number of

sociological studies of heterogeneous treatment effects in other scenarios have supported a negative selection hypothesis, finding that individuals least likely to undergo a "treatment" experience the greatest impact when they do (Brand and Davis 2011; Brand and Simon Thomas 2014; Brand and Xie 2010; Brand et al. 2019; Cheng et al. 2021; Schafer, Wilkinson, and Ferraro 2013).

Current scholarship on citizenship shows substantial evidence that naturalization offers occupational advantages across many contexts (Bevelander and Pendakur 2014; Fougère and Safi 2009; Hoxhaj, Vink, and Breuer 2020), though some studies show no effect (Bevelander and DeVoretz 2008; Hainmueller et al. 2023) or even a negative effect (Scott 2008). Importantly, research is even more mixed on who benefits from citizenship. On the one hand, studies suggest that individuals who are most likely to benefit from citizenship are individuals who are further in social distance from the native-born. For instance, in Western Europe, immigrants who arrive with low levels of human capital, few transferable skills, and from countries that are culturally different have a harder time in the local labor market (Heath and Cheung 2007). For these migrants, citizenship helps labor market integration (Fougère and Safi 2009; Hainmueller, Hangartner, and Pietrantuono 2017; Hainmueller et al. 2019; Peters, Schmeets, and Vink 2020). By contrast, other studies suggest that personal resources matter in naturalization decisions (Bloemraad 2006). Those with more resources show higher levels of naturalization rates, leading some to argue that these individuals have weighed the benefits and penalties of naturalization and have determined that benefits outweigh the costs (DeVoretz and Irastorza 2017; Jones-Correa 1998).

Uncovering the heterogeneous effects of citizenship has been difficult in these studies because the penalties and costs vary significantly across contexts and time periods. Elucidating the value of citizenship in the U.S. has been even more challenging because the U.S. rarely provides

quasi-random assignment opportunities for citizens — as has happened in Europe — and there have been declining legal distinctions between permanent residents and citizens (Hoxjah et al. 2020; Schuck 2018; He 2025). Additionally, the many legal and visa statuses makes analyzing the benefits of citizenship difficult given the complexity of the current immigration system (He 2025). We therefore turn to early-twentieth-century migration flows where naturalization and legal procedures were more straightforward. This historical period allows us to more clearly adjudicate between positive and negative selection. If naturalization primarily advantaged those who were already relatively well positioned, then we should expect to see gains concentrated among those with higher propensities to naturalize. By contrast, if citizenship was more transformative for those facing greater barriers to incorporation, then the largest benefits should appear among immigrants with lower propensities to naturalize.

Naturalization in the Age of Mass Migration

Naturalization in the United States has been a relatively straightforward procedure for most of its history. In the age of mass migration, all free white men could declare their intent to naturalize after two years of residence. This process involved a $1 fee where U.S. court clerks would review the applicant to ensure they would qualify for full citizenship (Motomura 2007). Then, after at least five years of residence and two years after declaring intent, intending citizens could petition for naturalization. This step involved a $4 fee, proof that they could speak English, have two character witness statements by citizens, and taking an oath of allegiance. Once immigrants completed and passed both steps, they were given US citizenship and offered all the rights that status conferred.

Despite the relatively straightforward procedure, political status stratified immigrants in the labor market. Laws and policies blocked noncitizens from some occupations such as lawyers and accountants, restricted noncitizen employment on public works projects, and developed "Americans First" policies that would only promote citizens (Catron 2019). Noncitizens often worked in secondary sectors in laborer positions within firms that had little chance for better attainment (Catron 2016). Only after becoming a citizen would firms allow for promotions into craft and managerial positions for immigrants. Similarly, state laws across the country blocked noncitizens from many white-collar jobs and many cities did not allow noncitizens to work on public works projects such as building roads and bridges (Konvitz 1946). By 1920, every state, including Washington D.C., had some restriction on what occupations noncitizens could have. Cities and counties also imposed their own restrictions on some occupations. The number of occupations restricted was positively correlated with the number of aliens in a given state (Fields 1933). Although these laws and policies were not strictly enforced in all cases, the preference given to citizens over noncitizens generated inequalities within European groups that persisted across generations (Catron 2019, 2021).

Public works projects were especially consequential. By the early twentieth century, twenty-three states did not allow noncitizens to work on publicly financed projects such as building bridges, roads, schools, and transit systems. Because of immigrants spatial distribution in the country, 72 percent of all immigrants were subject to these laws (Fields 1933). Courts upheld such exclusions on the grounds that the presence of unemployed Americans justified barring aliens from public employment.

New York City was one of the most important places in which alienage restrictions could directly shape naturalization decisions. New York was the primary receiving city for immigrants

due to its proximity to Ellis Island, and consequently had the largest foreign-born population in the country: In 1910, 40 percent of the city was born in another country (Foner 2008). The city's political economy was dominated by political machines, which channelled government contracts and jobs disproportionately toward citizens, especially within Irish dominated networks (Erie 1990). Immigrants who could vote and hold public employment benefited most from this system, while noncitizens were excluded. Alienage laws further reinforced this divide. When New York began enforcing restrictions on alien labor during the construction of the subway system, the state Supreme Court ruled that "[publicly funded jobs] do not belong to aliens" (People v Crane 1915). The abrupt enforcement triggered a rush of Italian subway workers to secure naturalization papers, overwhelming the Naturalization Bureaus in New York and Bronx counties (NYT 1914). Although the law was amended to permit employment of aliens when citizens were not available, the preference given to citizens in hiring and promotion decisions to citizens over noncitizens created enduring economic disparities. These laws and employer practices led to reactive citizenship where those who were further in social distance to the native-born ultimately naturalized faster than those who were closer in social distance (Waldinger et al. 2023).

While New York was significant given its treatment of aliens, similar alienage laws and employer practices structured opportunities across the country. Restrictions across the country limited occupational opportunities for noncitizens, although they varied in scope and enforcement. Together, however, they created a nationwide system in which citizenship was a critical threshold for economic mobility. The consistency of these patterns showed that citizenship was a practical necessity for full participation in American labor markets (Catron 2019). The Age of Mass Migration, therefore, is a particularly powerful context to assess who stood to gain the most from

becoming a citizen and how the institution of citizenship shaped immigrant incorporation on a national scale.

Data

Hand-Coded Naturalization Documents

Our study uses two samples. The main sample consists of intention and petition naturalization records from New York City's Southern District, sourced from National Archives and Records Administration (NARA) and hosted on FamilySearch.org. These records include 342,895 declarations of intent to naturalize filed between 1907 and 1924 and 154,849 petitions for naturalization filed between 1910 and 1924. We refer to this focal sample as the "hand-coded documents," and we reproduce examples of the digitized records in the Appendix**.**

The digitized petition and intention data from NARA include minimally transcribed information including the person's name, age, and place of birth. We use this information to connect intention forms to petition forms in order to find who did and did not naturalize. We employed a strict record linkage algorithm (Abramitzky et al. 2021) that first standardized first and last names and corrected nicknames. We then matched individuals using exact criteria: first and last names, age, and birthplace. For records that could not be matched using these strict parameters, we expanded our search to include matches within a 1-year age band (both older and younger), followed by a 2-year age band if necessary. We discarded the sample as unmatched if there were multiple matches or no match. Through this iterative matching process, we linked 41,665 declarations to their corresponding petitions. We then used the same method to first link our intention sample to the 1940 full-count census. Lastly, we utilized the HISTID variable from IPUMS and crosswalk files from the Census Linking Project (Abramitzky et al. 2021) to connect

the 1930 census to the 1920 and 1930 censuses. In all, we were able to link 1,327 naturalized and 620 unnaturalized cases to the three censuses, resulting in a final sample of 1,947 cases.

Because the digitized naturalization records had not been fully transcribed, we then hand-transcribed our full sample of declarations. We manually extracted additional demographic details from declarations that were not accessible through web scraping. These included occupation, skin complexion, height, weight, hair color, eye color, distinctive physical marks, New York address, previous residence before immigration, ports of departure and arrival, vessel name, and the date of arrival in the United States. These hand-coded characteristics — especially occupation at the time of declaring intention to naturalize and the person's height — bring in more information of individuals that are not available in full-count census data that have been used in other studies of historic naturalization (Bloemraad 2006; Catron 2019). Table 1 presents the descriptive statistics of the naturalization records.

[Table 1 about here]

Incorporating naturalization records into census records provides many advantages as opposed to only analyzing trajectories across censuses. First, including intention records provides detailed information about individuals not found in censuses including physical descriptions and more detailed economic indicators such as occupation the moment naturalization decisions are made. Second, it reduces measurement error in the naturalization variable in the census because it documents the point in the process that the decision is made. That is, evidence suggests that immigrants were likely to tell census enumerators that they had first papers or were citizens, even when they had not started the process (Schneider 2001). By incorporating these administrative

records, we can better measure selection into citizenship because we have detailed information about the individual at the moment of naturalization decisions and more detailed demographic information that influence these decisions.

Full-Count Census Data

While our first dataset provides detailed information about individuals in New York, we also supplement our analyses with full-count census data to analyze the heterogenous effects of citizenship as a whole.. We use a linked-census dataset of of foreign-born men without U.S. citizen parents, living anywhere in the U.S., who are not naturalized in 1920, and who are successfully linked to the 1930 and 1940 censuses. For this dataset, we use the MLP process produced by IPUMS (Helgertz et al. 2021). To correspond to our dataset mentioned above, we further restrict our sample to immigrants who either naturalized by 1930 or were not marked as naturalized in the 1930 or 1940 censuses. This yields a final sample of 32,268 individuals, which refer to as the "full-count censuses."

The IPUMS *citizen* variable differentiates between two kinds of non-citizens: those who had "received first papers" — i.e., had filed a declaration of intention to naturalize — and those who had not begun the naturalization process, who are labeled as "not a citizen." In our sample, 12,030 individuals had first papers in 1920 and naturalized by 1930; 2,727 had first papers in 1920 but did not naturalize by 1940; 11,982 were not a citizen in 1920 and naturalized by 1930; and 8,529 were not a citizen in 1920 and did not naturalize by 1940. Descriptive statistics for the full sample and these subsamples are shown in Table 2, along with non-naturalized immigrants in the full 1920 Census before linking. Distributions of variables in the full census and the linked sample

are fairly similar, suggesting that linking has not introduced significant bias, at least on observed characteristics

[Table 2 about here]

Methods

Heterogeneous Treatment Effects

To study how returns to naturalization may vary, we implement a heterogeneous treatment effect framework (Brand and Thomas 2013; Xie, Brand, and Jann 2012). This framework draws on the potential outcomes model for causal inference (Rubin 1974). Let us assume we are interested in a binary treatment D and an outcome Y. As is common in causal inference, we use the language of "treatment," but in reality this may be any kind of social experience. This framework defines the outcome if individual i receives the treatment, y_i^1, and the outcome when individual i does not receive the treatment, y_i^0. However, because in reality each individual either receives the treatment or does not, we only observe y_i^1 or y_i^0, but not both; the unobserved y_i^d is the counterfactual outcome. This is the "fundamental problem of causal inference": We cannot observe both potential outcomes for the same individual (Holland 1986). Causal inference allows us to overcome this problem by finding comparable treated and untreated groups or using statistical adjustment to make these groups comparable, so that the group effect average approximates the unestimable individual effect average (Morgan and Winship 2015). These statistical adjustments often rely on modeling selection into treatment, or the propensity to receive treatment, π. We estimate $\hat{\pi}(x_i)$, the individual propensity to receive treatment, as a function of covariates $X = x_i$ for individual i.

The heterogeneous treatment effect framework extends this causal model to directly examining the propensity to receive treatment (Brand and Thomas 2013; Xie, Brand, and Jann 2012). If selection into treatment is meaningfully related to the outcome of interest, then their potential outcomes may also relate to their propensity to receive treatment. By modeling selection processes such as the propensity to attend college or the propensity for a child's parents to divorce, researchers are able to separate individuals into "propensity strata" – defined by discrete ranges of $\hat{\pi}(x_i)$ – in which the effect of a social experience may vary. We define the conditional average treatment effect for individual i with covariates x_i as $\tau(x_i) = E(y_i^1 - y_i^0 \mid X = x_i)$. Individuals with low values of $\hat{\pi}(x_i)$ – i.e., those in a low-propensity stratum – may see different treatment effects than those with high values of $\hat{\pi}(x_i)$ – those in a high-propensity stratum.

In addition to extending this framework to the heterogeneous benefits that immigrants may gain from naturalizing, we harness recent advances in machine learning to nonparametrically model the propensity to naturalize (Brand, Zhou, and Xie 2023; Lee, Lessler, and Stuart 2010; Brand et al. 2021). While propensity-score methods often use logistic regression to estimate $\hat{\pi}(x_i)$, this choice entails functional form assumptions that may limit the extent to which these methods accurately model the selection process. Since we do not need to interpret individual coefficients, we make use of the highly flexible, data-driven method of random forests to estimate $\hat{\pi}(x_i)$, making no parametric assumptions and providing highly accurate propensity score estimates for rich data (Breiman 2001).

Analytic Strategy

Our analysis proceeds in three steps. First, we descriptively examine the trajectories of immigrants who successfully naturalize compared to those who begin the naturalization process but do not

finish. We use three time points. For the hand-coded documents, these include the time of declaration filing (between 1907 and 1924), 1930, and 1940. For the first time point, we calculate the occupational income score (OCCSCORE) value from occupations described in the the declaration forms. We then track their OCCSCORE across time using the 1930 and 1940 census records. Based on average incomes for particular occupations in 1950, OCCSCORE reflects the occupational attainment of immigrants rather than their direct income and is commonly used as a measure of occupational prestige (Abramitzky, Boustan, and Eriksson 2014). Our variable of interest is whether someone successfully naturalizes, a binary treatment variable. At the time of declaration filing, we set this value as false for everyone, while in 1930 or 1940, it is true for those who naturalize. Cases who naturalize do so by 1924 at the latest.

Second, we use OLS regressions with individual and year fixed effects to control for possible confounders. While the fixed effects account for time-invariant person-level confounders as well as year-specific factors, we additionally control for the following time-variant individual variables from the 1920, 1930, and 1940 censuses: whether the person resides in New York City, resides in an urban location, has children, or is married. For these regressions, we cluster standard errors at the person level.

Third, we examine how the returns to naturalization vary by the propensity to naturalize. To do this, we estimate a random forest model for selection into completing the naturalization process using the *grf* package in R (Athey, Tibshirani, and Wager 2019). Random forests are a flexible machine learning method that allow for interactions and higher-order terms in a data-driven way, overcoming the functional form limitations of methods such as logistic regression.[1]

[1] In the Appendix, we present the same models estimated using logistic regression. Results are substantively unchanged. Due to missing data on some variables, the sample for the logistic regression is smaller, at 1,914, while the random forest model does not drop incomplete cases.

They also incorporate covariate missingness as meaningful information, retaining incomplete cases in the analysis. These models control for a number of hand-coded variables from the declarations of intent, including age at declaration, birth region (based on detailed place of birth), complexion ("dark or brown" or "fair or light"), eye color ("blue or grey" or "brown or other"), and OCCSCORE. As a measure of childhood materials conditions, we also include the difference in the immigrant's height from the country average at the time, using data from Baten and Blum (2012). We also control for variables from the 1920 census, namely having children, being literate, being married, the ability to speak English. We stratify the sample by estimated propensity scores, calculating the OCCSCORE trajectory over time for those who do and do not naturalize in high- and low-propensity groups. We also assess variable importance in the random forest models, which reflects the relative number of times a variable was used to split trees in the random forest; variables used more frequently are considered more important to the model and hence better predictors of naturalization. For descriptive statistics of the variables used in both the fixed-effect and random-forest models, see Table 1 above.

In modeling selection into completing the naturalization process, we miss a "true" control group: noncitizens who do not even begin the naturalization process. To overcome this limitation, we compare results from our original dataset to those from full-count census data. While the census data do not include the rich set of hand-coded covariates we garnered from the archival documents — namely occupation at time of declaration of intention to naturalize, height, eye color, complexion, and detailed place of birth — they do allow us to distinguish three groups of immigrants: those who have not begun the naturalization process, those with "first papers" (i.e., those who have filed declarations of intention to naturalize), and naturalized citizens. We replicate the three-part analysis above for these census samples, with separate results for immigrants who

have first papers in 1920 and those that have not begun the naturalization process by 1920. Because this analysis involves immigrants from across the U.S., we include state fixed effects in an additional fixed effect model, and in our random forests models we cluster effects at the state level.

Results

Hand-Coded Naturalization Documents

Figure 1 presents mean OCCSCORE trajectories for our two groups of interest in the hand-coded documents: those who successfully naturalize, and those who begin the process but do not complete it.[2] We see that at the time of declaration filing, the average OCCSCORE for the eventual naturalizers is slightly higher (26.0) than those who do not naturalize (24.9). By 1930, however, the gap between these two groups has widened to over 2 points (29.5 vs. 27.2). This gap slightly narrows in 1940. This is consistent with prior research showing thatnaturalization had positive impacts on economic trajectories of immigrants in the past (Catron 2019). Overall, however, the greater advantage enjoyed by naturalizers even before completing naturalization suggests that the higher OCCSCORE trajectories shown in Figure 1 could simply reflect selection into naturalization rather than returns to naturalization itself.

[Figure 1 about here]

Table 3 presents coefficients from our OLS models. Model 1 includes no fixed effects, and we see a positive, significant coefficient of 3.2: Naturalizers have higher OCCSCORE values after the time of naturalization than non-naturalizers. This coefficient slightly increases in Model 2

[2] In the Appendix, we present tables with the numerical estimates used to create all figures in the main paper.

when we add individual fixed effects, but this effect is slightly reduced to 1.2 as we add a year fixed effect in Model 3. Model 4 includes individual controls and the effect remains significant at 1.4 . These preliminary results suggest that immigrants who complete naturalization are able to enter higher-paid occupations over time.

[Table 3 about here]

[Table 4 about here]

We next examine returns to naturalization for those most and least likely to complete the naturalization process. We use a random forest model to predict whether an individual completes the naturalization process or not. Table 4 presents variable importance for the random forest model estimating propensity to naturalize. At the top of the table are the variables used to split the most trees in the random forest — i.e., those that are most predictive of naturalization. We find that both achieved and ascribed characteristics carry high levels of importance when predicting naturalization. Ascribed characteristics such as age at declaration and the difference in height relative to one's sending country. Height corresponds to the selectivity of immigrants before arrival and is correlated with a number of outcomes including higher earnings, health, physical strength, and other features associated with socioeconomic attainment over time (Steckel 2009). Additionally, achieved characteristics are also important such as marriage, OCCSCCORE at the time of declaration, literacy, and English ability. These results are consistent with prior research (Bloemraad 2006; Yang 1994), and to supplement these outcomes, we present alogistic regression in the Appendix that corroborates our findings.

To disentangle selection from the effects of naturalization, we split the sample by the median estimated propensity to naturalize (0.69) as calculated in the random forest models.[3] In the Appendix, we present summary statistics for characteristics for the low- and high-propensity groups; in addition to having lower OCCSCOREs, those in the low-propensity group are less likely to be literate or speak English and more likely to come from Southern Europe. Figure 2 plots OCCSCORE trajectories for these low- and high-propensity groups, stratified by whether or not individuals complete the naturalization process. In the left panel, in the low-propensity group, OCCSCOREs begin at low levels, around 24 for both those who eventually naturalize and those who do not. In 1930, both groups are in occupations with higher OCCSCOREs than at the time of declaration filing. But naturalizers have made more progress, rising to nearly 29, while the unnaturalized attain an average OCCSCORE of 26.5. This gap remains steady in the 1940 census.

The right panel of Figure 2 presents OCCSCORE trajectories for the group with a high propensity to naturalize. This group has higher OCCSCOREs at the time of declaration than the low-propensity group, with an average of 27 for both those who do and do not eventually naturalize. By 1930, naturalizers have overtaken the unnaturalized, but the gap is smaller than in the low-propensity group, at 28 for the unnaturalized and 30 for the naturalized. This gap remains steady into 1940, though absolute OCCSCOREs decrease slightly. Notable in Figure 2 is the different returns to naturalization for the low- and high-propensity groups. Between the time of declaration filing and the 1930 census, naturalizers in the low-propensity group gain over 4 OCCSCORE points, while naturalizers in the high-propensity group gain about 3. This suggests that immigrants who were *least* likely to complete the naturalization process benefited the most from naturalization.

[3] In the Appendix, we also split the sample in three by propensity terciles. Substantive results remain unchanged.

[Figure 2 about here]

Full-Count Census Data

We next compare results from our hand-coded naturalization documents to full-count census data. Figure 3 presents descriptive trajectories for the two groups in the census data: those who had filed declarations of intention by 1920 ("first papers"), and those who had not begun the naturalization process by 1920 ("not a citizen").. The trajectories are similar to those presented in Figure 1, with greater gaps for the sample who had not begun the citizenship process in 1920. OCCSCOREs are generally lower in the census sample than in the hand-coded sample, and the gap between naturalizers and non-naturalizers is also larger, representing both the higher wage occupations in New York City compared to the rest of the nation and the city's strict penalties on noncitizens for employment.

[Figure 3 about here]

Using these two samples, Tables 5 and 6 fit the same fixed-effects models as in Table 3, but with the addition of a Model 5 that also adds state fixed effects. Results are substantively similar to those in Table 3, with the coefficient for naturalization significant in all models; naturalizers see greater gains in OCCSCORE over time than non-naturalizers. The sample of hand-coded documents sees the greatest gains in OCCSCORE, followed by the first papers sample in the census, with the non-citizen sample without first papers seeing slightly lower gains; however, these differences are not statistically significant.

[Table 5 about here]

[Table 6 about here]

Next, we use random forest models to estimate the propensity to naturalize. Tables 7 and 8 present variable importance for our two full-count census samples. For both those who had first papers in 1920 and those who did not, 1920 OCCSCORE, age, English proficiency, and especially literacy are important predictors of selection into naturalization by 1930. As shown in the Appendix, low-propensity naturalizers in both groups have lower rates of literacy and English proficiency, similarly to the hand-coded sample. Figure 4 presents OCCSCORE trajectories stratified by propensity to naturalize, as estimated by these models. Results are again presented separately for the first-paper and non-citizen samples. In both cases, immigrants with a low propensity to naturalize who do end up naturalizing see greater gains in OCCSCORE than those in the high-propensity group, mirroring the trends presented for the hand-coded sample in Figure 2. However, the low- and high-propensity groups are further apart in the census data. Similarly to the hand-coded sample, the low-propensity group in the sample without first papers in 1920 overtakes the high-propensity non-naturalizers, but the low-propensity individuals in the sample with first papers only reaches parity with the high-propensity non-naturalizers.

Our analyses of full-count census data support the results from our hand-coded intention dataset. The larger sample shows occupational trajectories across the country, giving clear evidence of our two main points: Naturalization benefits immigrants regardless of their starting point, and those who are least likely to naturalize see greater gains from naturalization than high-propensity naturalizers.

[Table 7 about here]

[Table 8 about here]

[Figure 4 about here]

Discussion and Conclusion

Citizenship acquisition has long been used to achieve economic assimilation. In the early twentieth century, citizenship became harder to obtain while the benefits to this status started to expand. This had large consequences whereby individuals who completed the naturalization process achieved better occupational trajectories than those who did not complete the process (Catron 2019). However, not all individuals benefited equally from this status. To date, little research has attempted to understand who benefited most. Indeed, the potential benefits to citizenship are related to the process of becoming a citizen in the first place. On the one hand, individuals who are most likely to naturalize may also be the ones most likely to benefit from this status (DeVortez and Irastorza 2017; Schneider 2001), while on the other hand, those least likely to become a citizen may be most likely gain from this status (Hainmueller, Hangartner, and Pietrantuono 2017). Understanding the selection process into citizenship helps us understand whether and for whom citizenship will provide benefits.

This article fills this research gap by directly analyzing who benefited most from citizenship. To do this, we created a novel hand-coded dataset of naturalization records in New York City that we then linked to complete count census records in order to follow citizens and noncitizens over their working lives. We supplemented these analyses with a linked full-count census dataset to show that trends in New York follow similar patterns in the U.S. as a whole. In all cases, we find that on average citizens achieved higher occupational outcomes over time

compared to those who did not complete the process of naturalization. However, the benefits of citizenship were not the same for all individuals. In particular, individuals who were least likely to become a citizen enjoyed a larger citizenship advantage compared to individuals who were most likely to become a citizen. In fact, individuals with a low propensity to naturalize who ultimately did so achieved similar occupational outcomes as high propensity naturalizers despite starting at a much lower level. By analyzing the heterogenous effects of naturalization, we find ample evidence for a negative selection hypothesis, wherein those least likely to gain access to citizenship are those with the most to gain from this status.

Our findings contribute to the growing literature that examines heterogenous effects of various social outcomes (Brand et al. 2021; Xie, Brand, and Jann 2012). Similar to our results, this research has shown that individuals who are least likely to have a social experience are also the most likely to have the greatest impact if they did (Brand and Simon Thomas 2014; Cheng et al. 2021; Schafer, Wilkinson, and Ferraro 2013). For these individuals, engaging in a social experience that they are less likely to do has a larger transformational impact because although they may be disadvantaged in one domain, they may derive greater benefits from an intervention in another domain. However, opportunity constraints often prevent some individuals from obtaining a social experience despite its potential transformative power. Our research extends this finding by analyzing a different social outcome in a different historical time.

In showing the heterogenous effects of citizenship acquisition, our research also contributes to the economic assimilation of immigrants more broadly (e.g., Abramitzky et al. 2023, 2014; Catron 2020; Catron, Vignau Loria, and Farr 2024). Research on immigrant's economic outcomes generally focuses on the impact of background differences in human capital or how various visa and legal categories shift trajectories (e.g., Ichou 2014; Potochnick and Hall 2021; Villarreal and

Tamborini 2018). While this research has produced important insights into our understanding of economic assimilation, it focuses on average effects of these various statuses. By shifting the focus to understanding who is affected most by a treatment, we can better understand for whom and how different processes work. This reframing is potentially useful for studies on contemporary immigration. The proliferation of different legal and visa statuses has made the political situation for immigrants more complex and stratified (Menjivar 2006; Delgado 2022). But while the political differences between an undocumented migrant and naturalized citizen are large, there are smaller differences between other statuses such as LPR status and naturalization. Reframing the question as who benefits most may help future research understand immigrant decisions.

The above study provides important insight into the processes of immigrant economic incorporation. While research has generally been concerned with citizenships' effects of integration in the labor market (Scott 2008; Peters, Vink, and Schmeets 2017), our study focuses on for whom citizenship matters most. In the early twentieth century, a citizenship premium was created due to employer practices and various alienage laws that widened formal inequalities. At the same time, access to citizenship became harder for many. Yet it was the individuals who struggled to obtain citizenship that would have benefited most. By turning our attention from average outcomes to differential impacts, we can better understand how exclusion from a political status can mask the transformative power of inclusion of this status.

References

Abramitzky, Ran, Leah Boustan, Peter Catron, Dylan Connor, and Rob Voigt. 2023. "The Refugee Advantage: English-Language Attainment in the Early Twentieth Century." *Sociological Science* 10:769–805. https://doi.org/10.15195/v10.a27.

Abramitzky, Ran, Leah Boustan, Elisa Jacome, and Santiago Perez. 2021. "Intergenerational Mobility of Immigrants in the United States over Two Centuries." *American Economic Review* 111(2):580–608. https://doi.org/10.1257/aer.20191586.

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. https://doi.org/10.1086/675805.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *The Annals of Statistics* 47(2):1148–78. https://doi.org/10.1214/18-AOS1709

Baten, J., & Blum, M. (2012). Growing Tall but Unequal: New Findings and New Background Evidence on Anthropometric Welfare in 156 Countries, 1810–1989. *Economic History of Developing Regions*, 27(sup1), S66–S85. https://doi.org/10.1080/20780389.2012.657489

Bean, Frank D., Susan K. Brown, and James D. Bachmeier. 2015. *Parents Without Papers: The Progress and Pitfalls of Mexican American Integration*. Russell Sage Foundation.

Bevelander, Pieter, and Don DeVoretz. 2008. *The Economics of Citizenship*. Malmö: Malmö Institute for Studies of Migration, Diversity and Welfare (MIM), Malmö University.

Bevelander, Pieter, and Ravi Pendakur. 2014. "The Labour Market Integration of Refugee and Family Reunion Immigrants: A Comparison of Outcomes in Canada and Sweden." *Journal of Ethnic and Migration Studies*, 40(5), 689–709. https://doi.org/10.1080/1369183X.2013.849569

Bloemraad, Irene. 2006. "Citizenship Lessons from the Past: The Contours of Immigrant

Naturalization in the Early 20th Century." *Social Science Quarterly* 87(5):927–53. https://doi.org/10.1111/j.1540-6237.2006.00409.x

Brand, Jennie E., Ravaris Moore, Xi Song, and Yu Xie. 2019. "Parental Divorce Is Not Uniformly Disruptive to Children's Educational Attainment." *Proceedings of the National Academy of Sciences* 116(15):7266–71. https://doi.org/10.1073/pnas.1813049116.

Brand, Jennie E., and Dwight Davis. 2011. "The Impact of College Education on Fertility: Evidence for Heterogeneous Effects." *Demography* 48 (3): 863–87. https://doi.org/10.1007/s13524-011-0034-3.

Brand, Jennie E., and Juli Simon Thomas. 2014. "Job Displacement Among Single Mothers: Effects on Children's Outcomes in Young Adulthood." *American Journal of Sociology* 119 (4): 955–1001. https://doi.org/10.1086/675409.

Brand, Jennie E., and Juli Simon Thomas. 2013. "Causal Effect Heterogeneity." In *Handbook of Causal Analysis for Social Research*, edited by Stephen L. Morgan, 189–213. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3_11.

Brand, Jennie E., and Yu Xie. 2010. "Who Benefits Most from College?: Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education." *American Sociological Review* 75(2):273–302. https://doi.org/10.1177/0003122410363567.

Brand, Jennie E., Jiahui Xu, Bernard Koch, and Pablo Geraldo. 2021. "Uncovering Sociological Effect Heterogeneity Using Tree-Based Machine Learning." *Sociological Methodology* 51(2):189–223. https://doi.org/10.1177/0081175021993503.

Brand, Jennie E., Xiang Zhou, and Yu Xie. 2023. "Recent Developments in Causal Inference and Machine Learning." *Annual Review of Sociology* 49 (1): 81–110. https://doi.org/10.1146/annurev-soc-030420-015345.

Bratsberg, Bernt, Jr. Ragan James F., and Zafar M. Nasir. 2002. "The Effect of Naturalization on Wage Growth: A Panel Study of Young Male Immigrants." *Journal of Labor Economics* 20(3):568–97. https://doi.org/10.1086/339616.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Catron, Peter. 2016. "Made in America? Immigrant Occupational Mobility in the First Half of the Twentieth Century." *American Journal of Sociology* 122(2):325–78. https://doi.org/10.1086/688043.

Catron, Peter. 2019. "The Citizenship Advantage: Immigrant Socioeconomic Attainment in the Age of Mass Migration." *American Journal of Sociology* 124(4):999–1042. https://doi.org/10.1086/701297.

Catron, Peter. 2020. "The Melting-Pot Problem? The Persistence and Convergence of Premigration Socioeconomic Status During the Age of Mass Migration." *Social Forces* 99(1):366–97. https://doi.org/10.1093/sf/soz146.

Catron, Peter. 2021. "The Alien Citizen: Social Distance and the Economic Returns to Naturalization in the Southwest." *Social Problems* 70(2):396–415. https://doi.org/10.1093/socpro/spab048.

Catron, Peter, María Vignau Loría, and Sarah Farr. 2024. "Contextual Boundaries: Skin Tone Stratification and Skill Transferability Among Mexicans in the Age of Mass Migration." *Demography* 61(5):1377–1402. https://doi.org/10.1215/00703370-11547756.

Cheng, Siwei, Jennie E. Brand, Xiang Zhou, Yu Xie, and Michael Hout. 2021. "Heterogeneous Returns to College over the Life Course." *Science Advances*, December. https://doi.org/10.1126/sciadv.abg7641.

Delgado, Vanessa. 2022. "Family Formation under the Law: How Immigration Laws Construct Contemporary Latino/a Immigrant Families in the U.S." *Sociology Compass* 16(9):e13027. https://doi.org/10.1111/soc4.13027.

DeVoretz, Don J., and Nahikari Irastorza. 2017. "Economic Theories of Citizenship?" *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2903117.

Erie, Steven P. 1990. *Rainbow's End: Irish-Americans and the Dilemmas of Urban Machine Politics, 1840-1985.* Berkeley: University of California Press.

Fields, Harold. 1933. "Where Shall the Alien Work?" *Social Forces* 12(2):213–21. https://doi.org/10.2307/2569745.

Foner, Nancy. 2008. *From Ellis Island to JFK: New York's Two Great Waves of Immigration*. Yale University Press

Fougère, Denis, and Mirna Safi. 2009. "Naturalization and Employment of Immigrants in France (1968-1999)" edited by A. F. Constant, M. Kahanec, and K. F. Zimmermann. *International Journal of Manpower* 30(1/2):83–96. https://doi.org/10.1108/01437720910948410.

Hainmueller, Jens, Elisa Cascardi, Michael Hotard, Rey Koslowski, Duncan Lawrence, Vasil I. Yasenov, and David Laitin. 2023. "Does Access to Citizenship Confer Socio-Economic Returns? Evidence from a Randomized Control Design." IZA Discussion Paper No. 16173.

Hainmueller, Jens, Dominik Hangartner, and Giuseppe Pietrantuono. 2017. "Catalyst or Crown:

Does Naturalization Promote the Long-Term Social Integration of Immigrants?" *American Political Science Review* 111(2):256–76. https://doi.org/10.1017/S0003055416000745.

Hainmueller, Jens, Dominik Hangartner, and Dalston Ward. 2019. "The Effect of Citizenship on the Long-Term Earnings of Marginalized Immigrants: Quasi-Experimental Evidence from Switzerland." *Science Advances* 5(12):eaay1610. https://doi.org/10.1126/sciadv.aay1610.

He, Qian. 2025. "Less Than Citizens: Varieties of Workplace Marginalization of Immigrants to the United States." *Social Forces* https://doi.org/10.1093/sf/soaf145

Heath, Anthony F., and Sin Yi Cheung, eds. 2007. *Unequal Chances: Ethnic Minorities in Western Labour Markets*. British Academy.

Helgertz, J., Price, J., Wellington, J., Thompson, K. J., Ruggles, S., & Fitch, C. A. (2021). A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 55(1), 12–29. https://doi.org/10.1080/01615440.2021.1985027

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60. https://doi.org/10.2307/2289064.

Hoxhaj, Rezart, Maarten Vink, and Tijana Breuer. 2020. "Immigrant Naturalisation, Employment and Occupational Status in Western Europe." *Frontiers in Sociology* 5:70. https://doi.org/10.3389/fsoc.2020.00070.

Hu, Anning. 2023. "Heterogeneous Treatment Effects Analysis for Social Scientists: A Review." *Social Science Research* 109 (January): 102810. https://doi.org/10.1016/j.ssresearch.2022.102810.

Jasso, G., and M. R. Rosenzweig. 1986. "Family Reunification and the Immigration Multiplier: U.S. Immigration Law, Origin-Country Conditions, and the Reproduction of Immigrants." *Demography* 23(3):291–311.

Jones-Correa, Michael. 1998. *Between Two Nations: The Political Predicament of Latinos in New York City*. Cornell University Press.

Kasinitz, Philip. 2019. "'Immigrants! We Get the Job Done!': Newcomers Remaking America on Broadway." *Ethnic and Racial Studies* 42(6):883–900. https://doi.org/10.1080/01419870.2019.1560109.

Konvitz, Milton R. 1946. *The Alien and the Asiatic in American Law*. Cornell Studies in Civil Liberty. Ithaca, N.Y: Cornell University Press.

Lee, Brian K., Justin Lessler, and Elizabeth A. Stuart. 2010. "Improving Propensity Score Weighting Using Machine Learning." *Statistics in Medicine* 29 (3): 337–46. https://doi.org/10.1002/sim.3782.

Luo, L. (2021). Heterogeneous Effects of Intergenerational Social Mobility: An Improved Method and New Evidence. American Sociological Review, 87(1), 143-173. https://doi.org/10.1177/00031224211052028.

Menjívar, Cecilia. 2006. "Liminal Legality: Salvadoran and Guatemalan Immigrants' Lives in the United States." *American Journal of Sociology* 111(4):999–1037. https://doi.org/10.1086/499509.

Morgan, Stephen L., and Christopher Winship. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. Analytical Methods for Social Research. New York, NY: Cambridge University Press.

Motomura, Hiroshi. 2007. *Americans in Waiting: The Lost Story of Immigration and Citizenship in the United States*. Oxford, UNITED STATES: Oxford University Press, Incorporated.

New York Times. 1914. "Will Dig Subways, But Fight Alien Act." November 19.

OECD. 2011. "Naturalisation: A Passport for the Better Integration of Immigrants?" https://www.oecd.org/en/publications/naturalisation-a-passport-for-the-better-integration-of-immigrants_9789264099104-en.html.

People v. Crane. 1915. https://case-law.vlex.com/vid/people-v-crane-886911175.

Peters, Floris, Hans Schmeets, and Maarten Vink. 2020. "Naturalisation and Immigrant Earnings: Why and to Whom Citizenship Matters." *European Journal of Population* 36(3):511–45. https://doi.org/10.1007/s10680-019-09540-1.

Peters, Floris, Maarten Vink, and Hans Schmeets. 2018. "Anticipating the Citizenship Premium: Before and after Effects of Immigrant Naturalisation on Employment." *Journal of Ethnic and Migration Studies* 44(7):1051–80. https://doi.org/10.1080/1369183X.2017.1367650.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701. https://doi.org/10.1037/h0037350.

Schafer, Markus H., Lindsay R. Wilkinson, and Kenneth F. Ferraro. 2013. "Childhood (Mis)fortune, Educational Attainment, and Adult Health: Contingent Benefits of a College Degree?" *Social Forces* 91 (3): 1007–34. https://doi.org/10.1093/sf/sos192.

Schneider, Dorothee. 2001. "Naturalization and United States Citizenship in Two Periods of

Mass Migration: 1894-1930, 1965-2000." *Journal of American Ethnic History* 21(1):50–82.

Schuck, Peter. 2018. *Citizens, Strangers, And In-Betweens: Essays On Immigration And Citizenship*. New York: Routledge.

Scott, Kirk. 2008. "The Economics of Citizenship: Is There a Naturalization Effect?" in *The Economic Consequences of Citizenship Ascension. A Cross Country Comparison*.

Scott, W. 2008. "Institutions and Organizations: Ideas and Interests." *Institutions and Organizations: Ideas and Interests*. https://digitalcommons.usu.edu/unf_research/55.

Smith, Robert C. 2024. *Dreams Achieved and Denied: Mexican Intergenerational Mobility*. American Sociological Association's Rose Series in Sociology. New York, New York: Russell Sage Foundation.

Waldinger, Roger, Nathan I. Hoffmann, and Peter Catron. 2023. "Impeding Access: Immigrants and Naturalization in Early Twentieth-Century America."

Xie, Yu, Jennie E. Brand, and Ben Jann. 2012. "Estimating Heterogeneous Treatment Effects with Observational Data." *Sociological Methodology* 42(1):314–47. https://doi.org/10.1177/0081175012452652.

Yang, Philip Q. 1994. "Explaining Immigrant Naturalization." *The International Migration Review* 28(3):449–77. https://doi.org/10.2307/2546816.

Tables

Table 1: Hand-coded documents: Summary statistics of variables used in the analysis, by naturalization. Variables with (years) come from the corresponding census, while those with (declaration) are hand-coded from archival declarations of intention to naturalize.

		Naturalized	
Characteristic	**Overall**, N = 1,944[1]	**Naturalized**, N = 1,326[1]	**Not naturalized**, N = 618[1]
Age (declaration)	28 (22, 34)	27 (22, 33)	29 (23, 34)
Birth region (declaration)			
Central/Southeastern Europe	234 (12%)	163 (12%)	71 (11%)
Eastern Europe	472 (24%)	335 (25%)	137 (22%)
Germany	224 (12%)	150 (11%)	74 (12%)
Great Britain	202 (10%)	133 (10%)	69 (11%)
Ireland	248 (13%)	191 (14%)	57 (9.2%)
Missing	62 (3.2%)	44 (3.3%)	18 (2.9%)
Other	70 (3.6%)	45 (3.4%)	25 (4.0%)
Southern Europe	307 (16%)	191 (14%)	116 (19%)
Western Hemisphere	29 (1.5%)	14 (1.1%)	15 (2.4%)
Western/Northern Europe	96 (4.9%)	60 (4.5%)	36 (5.8%)
Complexion (declaration)			
dark or brown	809 (42%)	548 (41%)	261 (42%)
fair or light	1,135 (58%)	778 (59%)	357 (58%)
Difference from mean country height in cm (declaration)	-1 (-5, 3)	-1 (-5, 3)	-1 (-5, 4)
Eye color (declaration)			
blue or grey	894 (46%)	611 (46%)	283 (46%)
brown or other	1,050 (54%)	715 (54%)	335 (54%)
Has children (1920)	1,137 (58%)	807 (61%)	330 (53%)

Has children (1930)	1,469 (76%)	1,023 (77%)	446 (73%)
Has children (1940)	1,363 (70%)	956 (72%)	407 (66%)
Literate (1920)	1,846 (95%)	1,279 (96%)	567 (92%)
Lives in NYC (1920)	1,606 (83%)	1,161 (88%)	445 (72%)
Lives in NYC (1930)	1,557 (80%)	1,127 (85%)	430 (70%)
Lives in NYC (1940)	1,507 (78%)	1,100 (83%)	407 (66%)
Married (1920)	1,450 (75%)	1,037 (78%)	413 (67%)
Married (1930)	1,745 (90%)	1,225 (93%)	520 (85%)
Married (1940)	1,685 (87%)	1,184 (89%)	501 (81%)
OCCSCORE (declaration)	25 (22, 28)	25 (22, 29)	24 (21, 28)
OCCSCORE (1930)	26 (23, 38)	26 (23, 40)	24 (20, 34)
OCCSCORE (1940)	25 (23, 36)	26 (23, 37)	25 (20, 34)
Speaks English (1920)	1,868 (96%)	1,292 (97%)	576 (93%)
Urban area (1920)	1,846 (95%)	1,283 (97%)	563 (91%)
Urban area (1930)	1,809 (93%)	1,254 (95%)	555 (90%)
Urban area (1940)	1,747 (90%)	1,219 (92%)	528 (85%)

[1] Median (IQR); n (%)

Table 2: Full-count censuses: Summary statistics of variables used in the analysis, by naturalization and 1920 status. "First papers in 1920" had filed a declaration of intention to naturalize, while "Not a citizen in 1920" had not begun the process. The "1920 Census before linking" column shows statistics for all immigrants in the 1920 census who were not yet naturalized, before linking to the 1930 and 1940 censuses.

Characteristic	**1920 Census before linking** N = 4,264,0241[1]	**Overall linked sample** N = 35,268[1]	**First papers in 1920, naturalized by 1930** N = 12,030[1]	**First papers in 1920, never naturalized** N = 2,727[1]	**Not a citizen in 1920, naturalized by 1930** N = 11,982[1]	**Not a citizen in 1920, never naturalized** N = 8,529[1]
Age (1920)	35 (27, 45)	33 (27, 40)	34 (29, 39)	36 (31, 43)	30 (23, 38)	34 (26, 41)
Has children (1920)	1,843,992 (43%)	23,036 (65%)	8,909 (74%)	2,154 (79%)	6,502 (54%)	5,471 (64%)
Has children (1930)		27,362 (78%)	10,156 (84%)	2,259 (83%)	8,570 (72%)	6,377 (75%)
Has children (1940)		26,618 (75%)	9,267 (77%)	2,057 (75%)	8,908 (74%)	6,386 (75%)
Literate (1920)	3,404,121 (80%)	30,002 (85%)	11,332 (94%)	2,330 (85%)	10,423 (87%)	5,917 (69%)
Married (1920)	2,500,005 (59%)	26,603 (75%)	10,325 (86%)	2,406 (88%)	7,687 (64%)	6,185 (73%)
Married (1930)		30,530 (87%)	11,278 (94%)	2,483 (91%)	9,828 (82%)	6,941 (81%)
Married (1940)		30,763 (87%)	10,916 (91%)	2,351 (86%)	10,433 (87%)	7,063 (83%)
OCCSCORE (1920)	23 (20, 25)	23 (20, 28)	24 (20, 32)	23 (20, 28)	23 (20, 28)	20 (20, 24)
OCCSCORE (1930)		24 (20, 30)	24 (20, 32)	23 (20, 27)	24 (20, 32)	23 (20, 25)
OCCSCORE (1940)		24 (20, 30)	24 (20, 32)	23 (20, 27)	24 (20, 32)	23 (19, 26)
Speaks English (1920)	3,350,526 (79%)	30,272 (86%)	11,330 (94%)	2,406 (88%)	10,389 (87%)	6,147 (72%)
Urban area (1920)	3,178,001 (75%)	27,165 (77%)	9,669 (80%)	2,056 (75%)	9,367 (78%)	6,073 (71%)
Urban area (1930)		27,725 (79%)	9,800 (81%)	2,020 (74%)	9,608 (80%)	6,297 (74%)
Urban area (1940)		26,834 (76%)	9,548 (79%)	1,925 (71%)	9,345 (78%)	6,016 (71%)
Birth region (1920)						
Central/	609,067 (14%)	4,515 (13%)	1,886 (16%)	387 (14%)	1,527 (13%)	715 (8.4%)

Southeastern Europe						
Eastern Europe	973,469 (23%)	8,652 (25%)	3,045 (25%)	739 (27%)	2,913 (24%)	1,955 (23%)
Germany	233,677 (5.5%)	2,389 (6.8%)	1,157 (9.6%)	148 (5.4%)	921 (7.7%)	163 (1.9%)
Great Britain	209,353 (4.9%)	2,547 (7.2%)	1,018 (8.5%)	165 (6.1%)	1,011 (8.4%)	353 (4.1%)
Ireland	117,128 (2.7%)	913 (2.6%)	454 (3.8%)	40 (1.5%)	344 (2.9%)	75 (0.9%)
Other	215,061 (5.0%)	971 (2.8%)	219 (1.8%)	64 (2.3%)	246 (2.1%)	442 (5.2%)
Southern Europe	825,245 (19%)	7,198 (20%)	1,920 (16%)	466 (17%)	2,709 (23%)	2,103 (25%)
Western Hemisphere	664,606 (16%)	3,944 (11%)	696 (5.8%)	281 (10%)	1,079 (9.0%)	1,888 (22%)
Western/ Northern Europe	416,418 (9.8%)	4,139 (12%)	1,635 (14%)	437 (16%)	1,232 (10%)	835 (9.8%)

[1] Median (Q1, Q3); n (%)

Table 3: Hand-coded documents: OLS models predicting OCCSCORE in the year of declaration, 1930, and 1940. Variables come from the 1920, 1930, and 1940 censuses. Standard errors clustered within person are shown in parentheses.

	(1)	(2)	(3)	(4)
(Intercept)	26.126 ***			
	(0.189)			
Naturalized	3.189 ***	3.405 ***	1.255 *	1.397 *
	(0.282)	(0.320)	(0.551)	(0.549)
Lives in NYC				-0.700
				(0.607)
Urban area				1.583
				(0.858)
Has children				0.632
				(0.457)
Married				2.360 ***
				(0.558)
Person fixed effects	No	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes
Person-years	5409	5409	5409	5409

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 4: Hand-coded documents: Variable importance for propensity to naturalize (based on how often a variable was used to split a tree in the random forest model)

Variable	Importance
Age (declaration)	0.177
Married (1920)	0.17
OCCSCORE (declaration)	0.121
Difference from mean country height in cm (declaration)	0.106
Literate (1920)	0.1
Speaks English (1920)	0.0745
Has children (1920)	0.0543
Birth region: Ireland	0.0499
Birth region: Southern Europe	0.0389
Birth region: Eastern Europe	0.022
Eye color: brown or other	0.019
Complexion: fair or light	0.0182
Birth region: Germany	0.0145
Birth region: Great Britain	0.0141
Birth region: Western/Northern Europe	0.0136
Birth region: Other	0.00408
Birth region: Missing	0.00275
Birth region: Western Hemisphere	0.000545

Table 5: Full-count censuses, first papers in 1920: Fixed effect models predicting OCCSCORE in census data for 1920, 1930, and 1940. Standard errors are clustered within person.

	(1)	(2)	(3)	(4)	(5)
(Intercept)	24.944 ***				
	(0.065)				
Naturalized	1.811 ***	0.886 ***	1.118 ***	0.971 ***	0.973 ***
	(0.088)	(0.076)	(0.169)	(0.168)	(0.168)
Urban area				2.154 ***	2.106 ***
				(0.202)	(0.200)
Has children				0.357 *	0.355 *
				(0.145)	(0.145)
Married				0.905 ***	0.895 ***
				(0.191)	(0.190)
Person fixed effects	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes
State fixed effects	No	No	No	No	Yes
Person-years	40832	40832	40832	40832	40832

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 6: Full-count censuses, noncitizens in 1920: Fixed effect models predicting OCCSCORE in census data for 1920, 1930, and 1940. Standard errors are clustered within person.

	(1)	(2)	(3)	(4)	(5)
(Intercept)	23.132 ***				
	(0.050)				
Naturalized	3.495 ***	1.653 ***	1.111 ***	0.883 ***	0.874 ***
	(0.078)	(0.085)	(0.124)	(0.122)	(0.122)
Urban area				2.188 ***	2.263 ***
				(0.165)	(0.165)
Has children				0.718 ***	0.701 ***
				(0.130)	(0.130)
Married				1.741 ***	1.744 ***
				(0.150)	(0.150)
Person fixed effects	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes
State fixed effects	No	No	No	No	Yes
Person-years	53572	53572	53572	53572	53572

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 7: Full-count censuses, first papers in 1920. Variable importance for propensity to naturalize (based on how often a variable was used to split a tree in the GRF model)

Variable	***Importance***
Literate (1920)	*0.466*
Age (1920)	*0.234*
OCCSCORE (1920)	*0.137*
Speaks English (1920)	*0.0721*
Birth region (1920): Western Hemisphere	*0.0351*
Birth region (1920): Germany	*0.022*
Birth region (1920): Eastern Europe	*0.00847*
Birth region (1920): Southern Europe	*0.00796*
Birth region (1920): Western/Northern Europe	*0.00723*
Has children (1920)	*0.00464*
Married (1920)	*0.0024*
Birth region (1920): Great Britain	*0.0014*
Birth region (1920): Other	*0.00101*
birth_region_1920Other	*0.00017*

Table 8: Full-count censuses, noncitizens in 1920. Variable importance for propensity to naturalize (based on how often a variable was used to split a tree in the GRF model)

Variable	***Importance***
Literate (1920)	*0.399*
Birth region (1920): Western Hemisphere	*0.192*
OCCSCORE (1920)	*0.119*
Age (1920)	*0.105*
Speaks English (1920)	*0.102*
birth_region_1920Other	*0.0264*
Birth region (1920): Germany	*0.0218*
Has children (1920)	*0.0134*
Married (1920)	*0.0113*
Birth region (1920): Great Britain	*0.003*
Birth region (1920): Southern Europe	*0.00248*
Birth region (1920): Eastern Europe	*0.00188*
Birth region (1920): Western/Northern Europe	*0.00172*
Birth region (1920): Other	*0.00013*

Figures

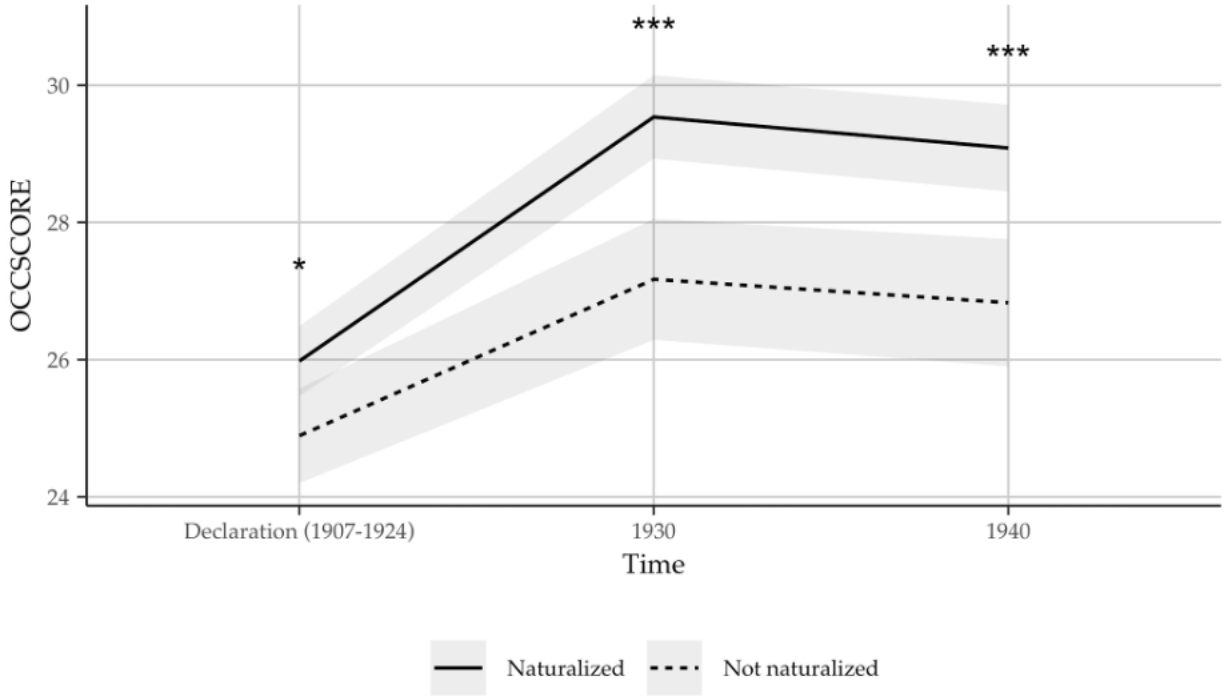


*Figure 1: Hand-coded documents: Mean OCCSCORE trajectories for naturalized (N = 1,326) and unnaturalized (UN = 618) cases, with 95-percent confidence intervals. All cases filed a declaration of intention to naturalize, but not all successfully complete the naturalization process. Stars indicate p-values from Z-tests of the difference in OCCSCORE between the naturalized and non- naturalized groups. ***p < .001; **p < .01; *p < .05; †p < .1.*

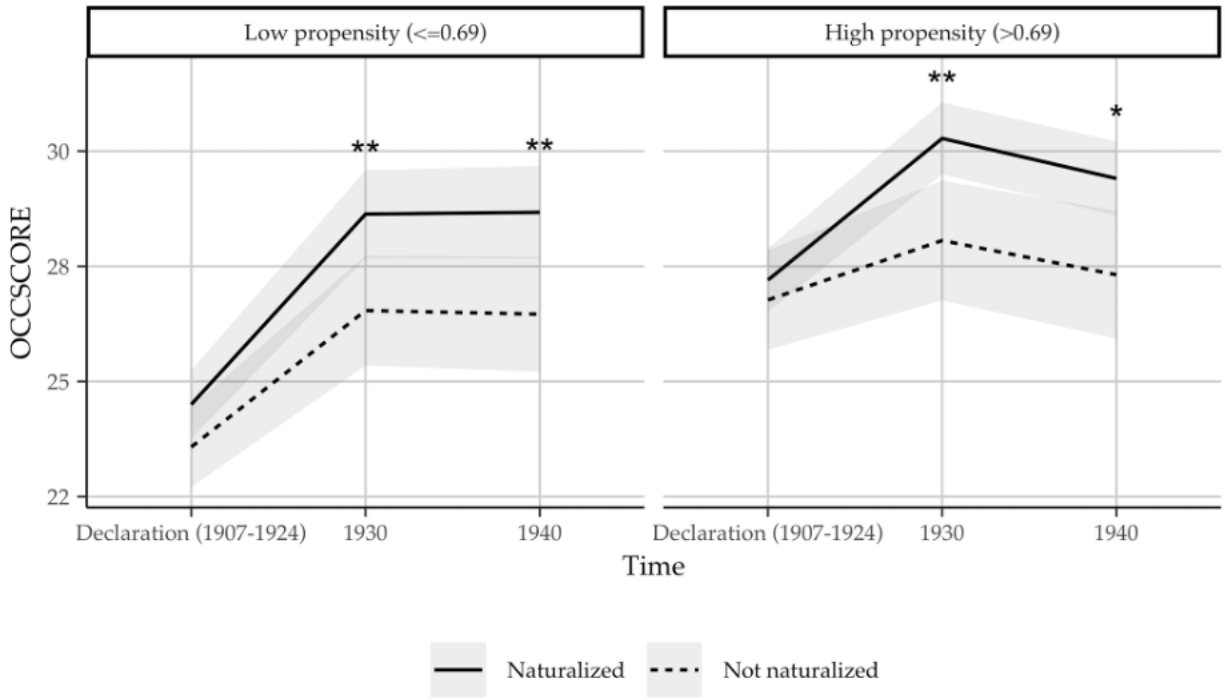


*Figure 2: Hand-coded documents: Observed OCCSCORE values, stratified by estimated propensity to naturalize (based a random forest model) and observed naturalization outcomes. Stars indicate p-values from Z-tests of the difference in OCCSCORE between the naturalized and non naturalized groups. ***p < .001; **p < .01; *p < .05; †p < .1.*

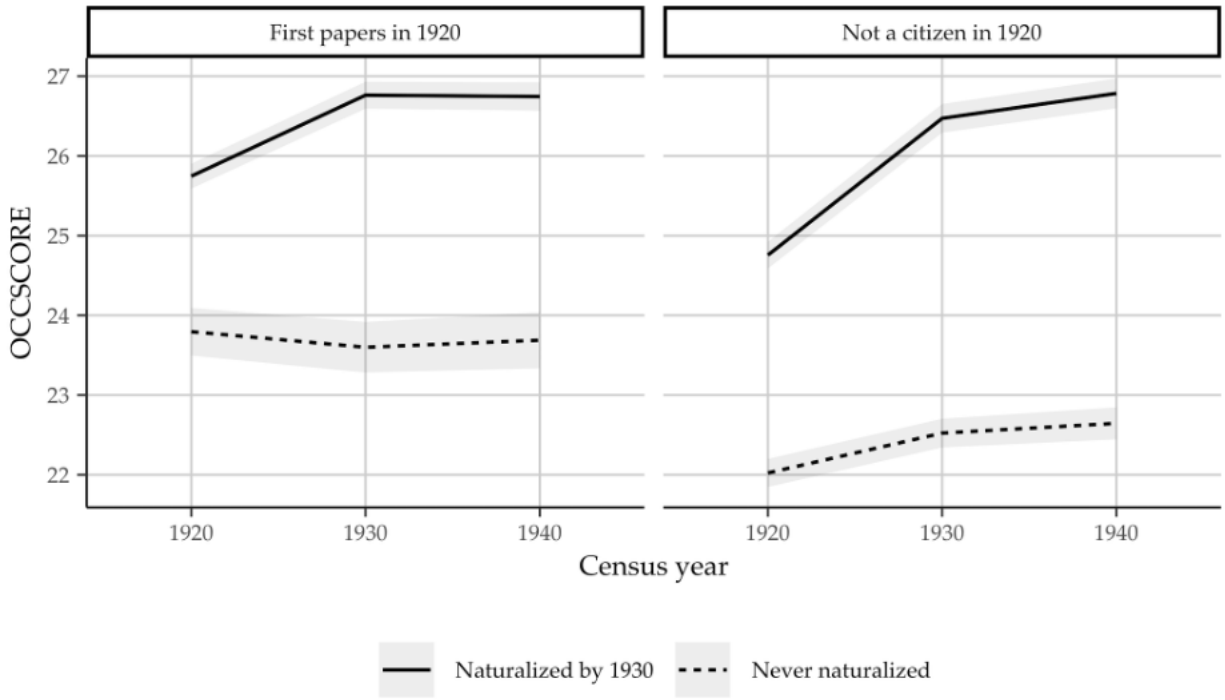


Figure 3: Full-count censuses: Mean OCCSCORE trajectories for immigrant men who are not naturalized in the 1920 census and who either naturalize by 1930 or are not marked as naturalized in any of the three censuses. "First papers" refers to having filed a declaration of intention to naturalize but not yet a petition to naturalize. Ribbons represent 95-percent confidence intervals. All Z-tests for the difference in OCCSCORE between the naturalized and not naturalized groups are significant at the $\alpha = 0.001$ level.

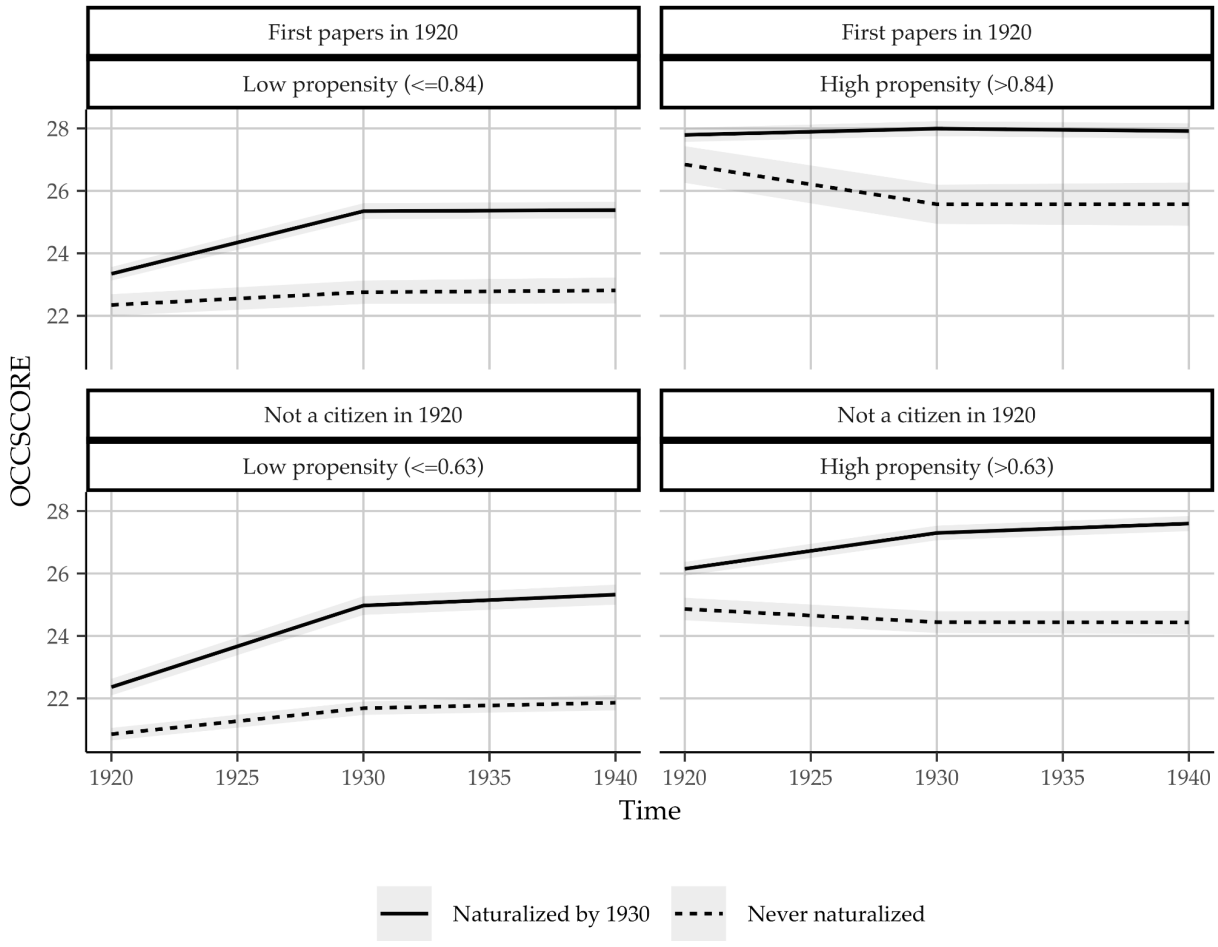
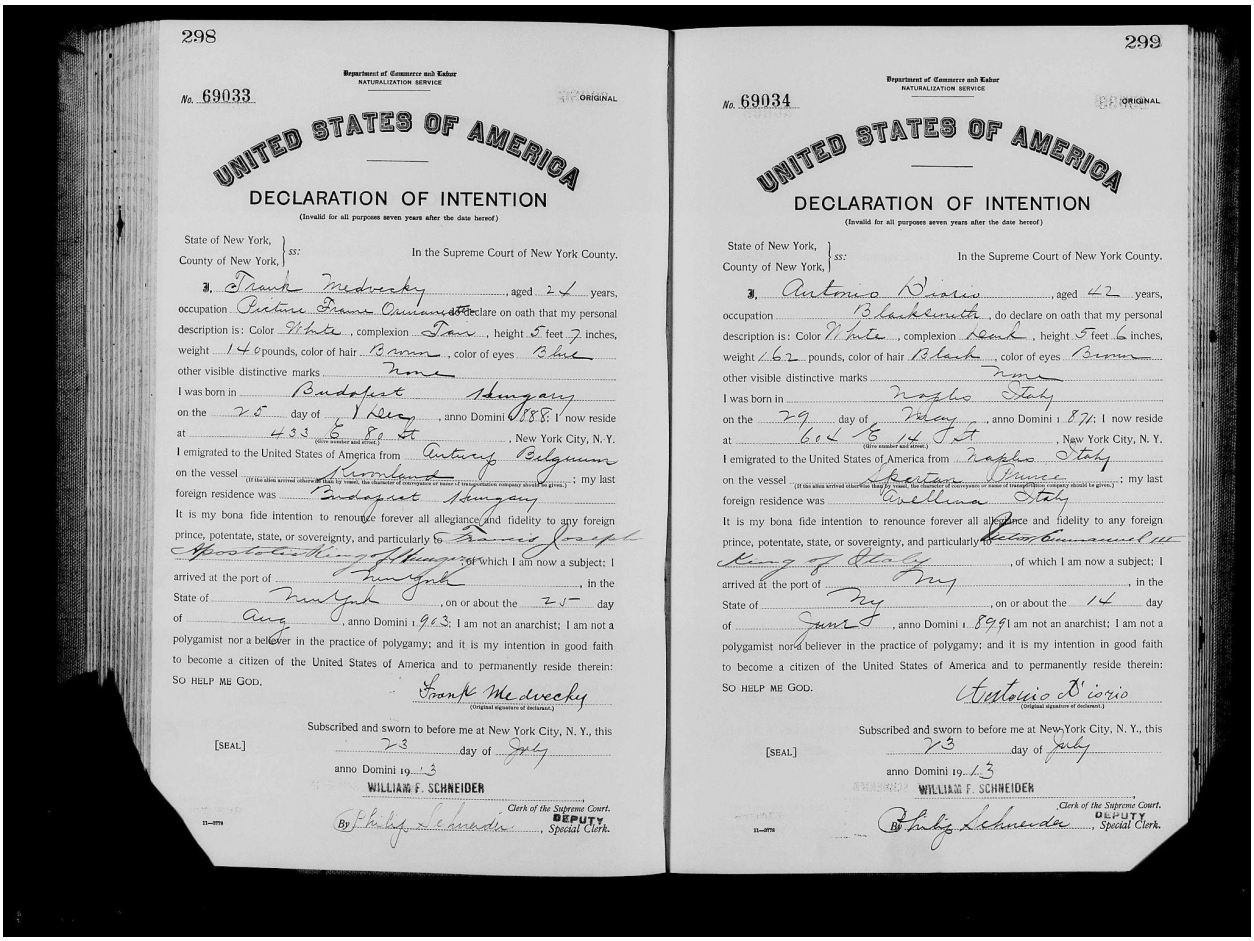


Figure 4: Full-count censuses: Observed OCCSCORE values, stratified by estimated propensity to naturalize (based on a GRF model) and observed naturalization outcomes. All Z-tests for the difference in OCCSCORE between the naturalized and not naturalized groups are significant at the $\alpha = 0.001$ level.

Appendix



298

Department of Commerce and Labor
NATURALIZATION SERVICE

No. 69033 ORIGINAL

UNITED STATES OF AMERICA

DECLARATION OF INTENTION

(Invalid for all purposes seven years after the date hereof)

State of New York, County of New York, } ss: In the Supreme Court of New York County.

I, Frank Medvecky, aged 24 years, occupation Picture Frame Ornamenter, do declare on oath that my personal description is: Color White, complexion Dark, height 5 feet 7 inches, weight 140 pounds, color of hair Brown, color of eyes Blue, other visible distinctive marks None

I was born in Budapest Hungary on the 25 day of Dec, anno Domini 1888; I now reside at 433 E 82 St (Give number and street.), New York City, N. Y. I emigrated to the United States of America from Antwerp Belgium on the vessel Kroonland (If the alien arrived otherwise than by vessel, the character of conveyance or name of transportation company should be given.); my last foreign residence was Budapest Hungary

It is my bona fide intention to renounce forever all allegiance and fidelity to any foreign prince, potentate, state, or sovereignty, and particularly to Francis Joseph Apostolic King of Hungary, of which I am now a subject; I arrived at the port of New York, in the State of New York, on or about the 25 day of Aug, anno Domini 1913; I am not an anarchist; I am not a polygamist nor a believer in the practice of polygamy; and it is my intention in good faith to become a citizen of the United States of America and to permanently reside therein: SO HELP ME GOD.

Frank Medvecky
(Original signature of declarant.)

Subscribed and sworn to before me at New York City, N. Y., this [SEAL] 23 day of July anno Domini 1913

WILLIAM F. SCHNEIDER, Clerk of the Supreme Court.
By Philip Schneider, DEPUTY, Special Clerk.

11—3778

299

Department of Commerce and Labor
NATURALIZATION SERVICE

No. 69034 ORIGINAL

UNITED STATES OF AMERICA

DECLARATION OF INTENTION

(Invalid for all purposes seven years after the date hereof)

State of New York, County of New York, } ss: In the Supreme Court of New York County.

I, Antonio Diorio, aged 42 years, occupation Blacksmith, do declare on oath that my personal description is: Color White, complexion Dark, height 5 feet 6 inches, weight 162 pounds, color of hair Black, color of eyes Brown, other visible distinctive marks None

I was born in Naples Italy on the 29 day of May, anno Domini 1871; I now reside at 604 E 14 St (Give number and street.), New York City, N. Y. I emigrated to the United States of America from Naples Italy on the vessel Spartan Prince (If the alien arrived otherwise than by vessel, the character of conveyance or name of transportation company should be given.); my last foreign residence was Avellina Italy

It is my bona fide intention to renounce forever all allegiance and fidelity to any foreign prince, potentate, state, or sovereignty, and particularly to ~~Victor Emmanuel III~~ King of Italy, of which I am now a subject; I arrived at the port of NY, in the State of NY, on or about the 14 day of June, anno Domini 1899; I am not an anarchist; I am not a polygamist nor a believer in the practice of polygamy; and it is my intention in good faith to become a citizen of the United States of America and to permanently reside therein: SO HELP ME GOD.

Antonio Diorio
(Original signature of declarant.)

Subscribed and sworn to before me at New York City, N. Y., this [SEAL] 23 day of July anno Domini 1913

WILLIAM F. SCHNEIDER, Clerk of the Supreme Court.
By Philip Schneider, DEPUTY, Special Clerk.

11—3778

Figure A.1: Declaration of Intention of Naturalization

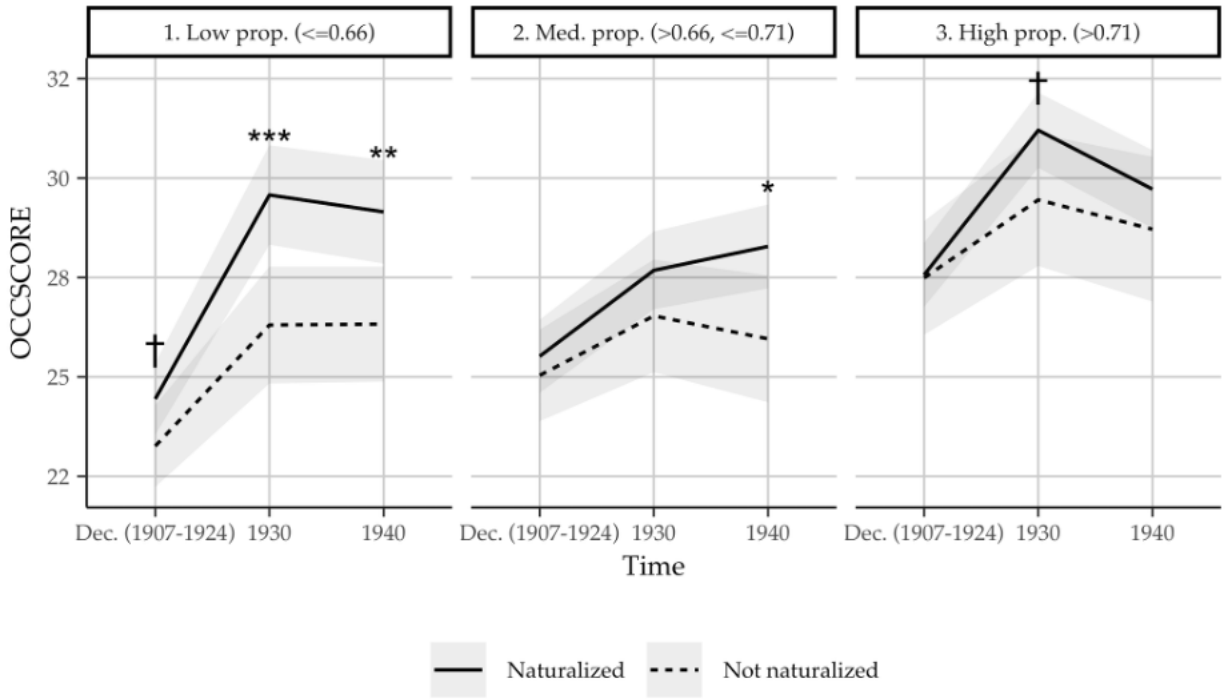


*Figure A.2: Hand-coded documents: Observed OCCSCORE values, stratified by estimated propensity to naturalize (based a random forest model) and observed naturalization outcomes. Stars indicate p-values from Z-tests of the difference in OCCSCORE between the naturalized and non naturalized groups. ***p < .001; **p < .01; *p < .05; †p < .1.*

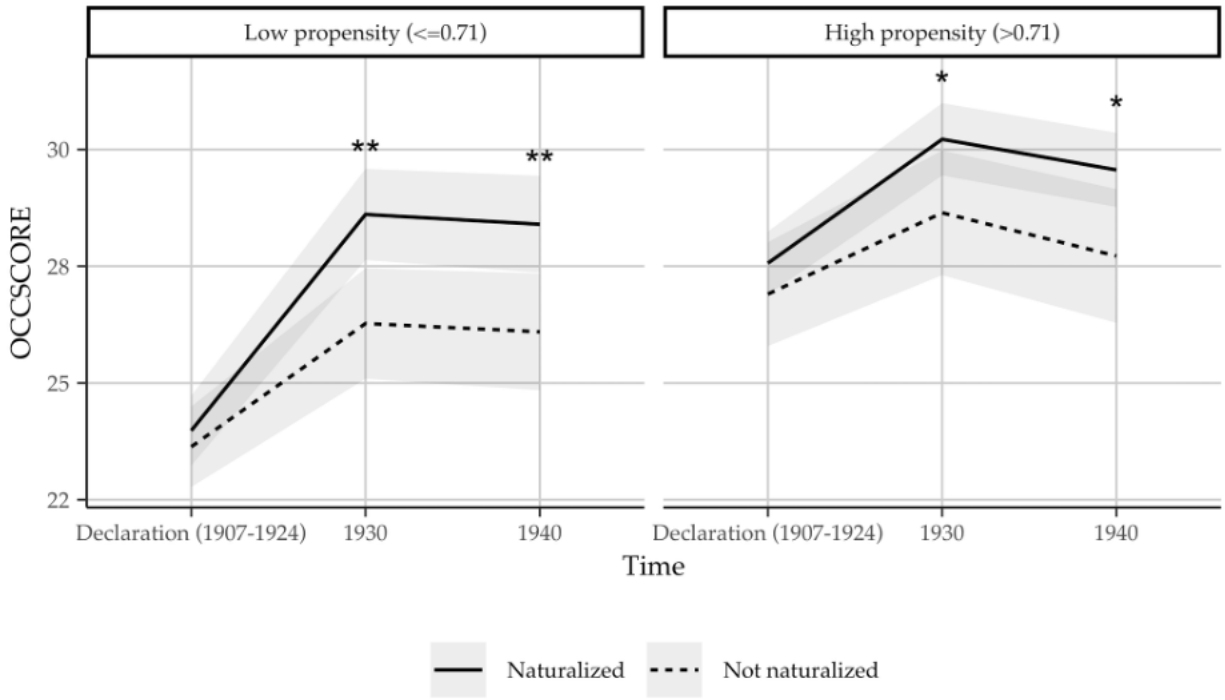


*Figure A.3: Hand-coded documents: Observed OCCSCORE values, stratified by estimated propensity to naturalize (based on logistic regression) and observed naturalization outcomes. Stars indicate p-values from Z-tests of the difference in OCCSCORE between the naturalized and non naturalized groups. ***p < .001; **p < .01; *p < .05; †p < .1.*

Table A.1: Hand-coded documents: Propensity score model for naturalization, fit with logistic regression. Variables come from declarations of intention and the 1920, 1930, and 1940 censuses.

	(1)
Age (declaration)	-0.033 ***
	(0.007)
Married (1920)	0.794 ***
	(0.158)
OCCSCORE (declaration)	0.011
	(0.006)
Difference from mean country height in cm (declaration)	-0.009
	(0.008)
Speaks English (1920)	0.654 *
	(0.256)
Literate (1920)	0.733 **
	(0.232)
Has children (1920)	0.060
	(0.143)
Birth region: Ireland	0.492 *
	(0.219)
Birth region: Southern Europe	-0.206
	(0.199)
Birth region: Western Hemisphere	-0.805
	(0.418)
Birth region: Eastern Europe	-0.025
	(0.183)
Birth region: Western/Northern Europe	-0.271
	(0.265)
Birth region: Great Britain	-0.046
	(0.215)
Complexion: fair or light	-0.050
	(0.117)
Eye color: brown or other	0.150
	(0.117)
Birth region: Germany	-0.004
	(0.212)
Birth region: Missing	0.127
	(0.326)
Birth region: Other	-0.333
	(0.298)
N	1914
logLik	-1138.515
AIC	2315.030

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table A.2: Hand-coded documents: Summary statistics of variables used in the analysis, by propensity to naturalize. Variables with (years) come from the corresponding census, while those with (declaration) are hand-coded from historic naturalization documents.

Characteristic	**Low propensity (<=0.69)** N = 972[1]	**High propensity (>0.69)** N = 972[1]
Age (declaration)	29 (22, 37)	26 (23, 32)
Birth region (declaration)		
Central/Southeastern Europe	113 (12%)	121 (12%)
Eastern Europe	178 (18%)	294 (30%)
Germany	118 (12%)	106 (11%)
Great Britain	105 (11%)	97 (10.0%)
Ireland	90 (9.3%)	158 (16%)
Missing	29 (3.0%)	33 (3.4%)
Other	38 (3.9%)	32 (3.3%)
Southern Europe	232 (24%)	75 (7.7%)
Western Hemisphere	16 (1.6%)	13 (1.3%)
Western/Northern Europe	53 (5.5%)	43 (4.4%)
Complexion (declaration)		
dark or brown	447 (46%)	362 (37%)
fair or light	525 (54%)	610 (63%)
Difference from mean country height in cm (declaration)	-1 (-5, 4)	-1 (-5, 3)
Eye color (declaration)		
blue or grey	430 (44%)	464 (48%)
brown or other	542 (56%)	508 (52%)
Has children (1920)	382 (39%)	755 (78%)
Has children (1930)	650 (67%)	819 (85%)
Has children (1940)	615 (63%)	748 (77%)
Literate (1920)	875 (90%)	971 (100%)
Lives in NYC (1920)	773 (80%)	833 (86%)
Lives in NYC (1930)	763 (79%)	794 (82%)
Lives in NYC (1940)	739 (76%)	768 (79%)
Married (1920)	478 (49%)	972 (100%)
Married (1930)	815 (84%)	930 (96%)
Married (1940)	799 (82%)	886 (91%)
OCCSCORE (declaration)	24 (20, 26)	25 (23, 30)
OCCSCORE (1930)	24 (22, 36)	27 (23, 40)
OCCSCORE (1940)	25 (20, 36)	26 (23, 36)

Speaks English (1920)	899 (92%)	969 (100%)
Urban area (1920)	910 (94%)	936 (96%)
Urban area (1930)	892 (92%)	917 (95%)
Urban area (1940)	858 (88%)	889 (92%)

[1] Median (Q1, Q3); n (%)

Table A.3: Full-count census: Summary statistics of variables used in the analysis, by propensity to naturalize. Variables with (years) come from the corresponding census, while those with (declaration) are hand-coded from historic naturalization documents.

Characteristic	**First papers in 1920, Low propensity (<=0.82)** N = 7,379[1]	**First papers in 1920, High propensity (>0.82)** N = 7,378[1]	**Not a citizen in 1920, Low propensity (<=0.61)** N = 10,256[1]	**Not a citizen in 1920, High propensity (>0.61)** N = 10,255[1]
Age (1920)	37 (32, 43)	32 (28, 37)	35 (28, 42)	29 (21, 36)
Literate (1920)	6,289 (85%)	7,373 (100%)	6,377 (62%)	9,963 (97%)
Birth region (1920)				
Central/Southeastern Europe	1,029 (14%)	1,244 (17%)	767 (7.5%)	1,475 (14%)
Eastern Europe	2,192 (30%)	1,592 (22%)	2,617 (26%)	2,251 (22%)
Germany	235 (3.2%)	1,070 (15%)	42 (0.4%)	1,042 (10%)
Great Britain	357 (4.8%)	826 (11%)	87 (0.8%)	1,277 (12%)
Ireland	103 (1.4%)	391 (5.3%)	33 (0.3%)	386 (3.8%)
Other	156 (2.1%)	127 (1.7%)	502 (4.9%)	186 (1.8%)
Southern Europe	1,253 (17%)	1,133 (15%)	2,765 (27%)	2,047 (20%)
Western Hemisphere	818 (11%)	159 (2.2%)	2,638 (26%)	329 (3.2%)
Western/Northern Europe	1,236 (17%)	836 (11%)	805 (7.8%)	1,262 (12%)
Has children (1920)	6,045 (82%)	5,018 (68%)	7,103 (69%)	4,870 (47%)
Has children (1930)	6,243 (85%)	6,172 (84%)	8,010 (78%)	6,937 (68%)
Has children (1940)	5,599 (76%)	5,725 (78%)	7,825 (76%)	7,469 (73%)
Married (1920)	6,676 (90%)	6,055 (82%)	7,920 (77%)	5,952 (58%)
Married (1930)	6,857 (93%)	6,904 (94%)	8,639 (84%)	8,130 (79%)
Married (1940)	6,512 (88%)	6,755 (92%)	8,701 (85%)	8,795 (86%)
OCCSCORE (1920)	23 (20, 25)	25 (23, 33)	20 (20, 24)	24 (20, 32)
OCCSCORE (1930)	23 (20, 28)	25 (23, 34)	23 (20, 25)	24 (20, 32)
OCCSCORE (1940)	23 (20, 29)	25 (23, 33)	23 (20, 26)	24 (20, 32)
Speaks English (1920)	6,564 (89%)	7,172 (97%)	6,891 (67%)	9,645 (94%)
Urban area (1920)	5,600 (76%)	6,125 (83%)	7,481 (73%)	7,959 (78%)
Urban area (1930)	5,661 (77%)	6,159 (83%)	7,789 (76%)	8,116 (79%)
Urban area (1940)	5,482 (74%)	5,991 (81%)	7,511 (73%)	7,850 (77%)

[1] Median (Q1, Q3); n (%)

Table A.4: Hand-coded documents: OCCSCORE trajectories, as presented in Figure 1

Year	Status by 1924	OCCSCORE	SE	CI Lower Bound	CI Upper Bound
Declaration (1907-1924)	Naturalized	26	0.26	25.5	26.5
Declaration (1907-1924)	Not naturalized	24.9	0.351	24.2	25.6
1930	Naturalized	29.5	0.31	28.9	30.1
1930	Not naturalized	27.2	0.45	26.3	28.1
1940	Naturalized	29.1	0.322	28.5	29.7
1940	Not naturalized	26.8	0.473	25.9	27.8

Table A.5: Hand-coded documents: Stratified trajectories, as presented in Figure 2

Propensity category	Year	Status by 1924	OCCSCORE	SE	CI Lower Bound	CI Upper Bound
Low propensity (<=0.69)	Declaration (1907-1924)	Naturalized	24.5	0.379	23.8	25.2
Low propensity (<=0.69)	1930	Naturalized	28.6	0.485	27.7	29.6
Low propensity (<=0.69)	1940	Naturalized	28.7	0.513	27.7	29.7
Low propensity (<=0.69)	Declaration (1907-1924)	Not naturalized	23.6	0.443	22.7	24.4
Low propensity (<=0.69)	1930	Not naturalized	26.5	0.607	25.4	27.7
Low propensity (<=0.69)	1940	Not naturalized	26.5	0.636	25.2	27.7
High propensity (>0.69)	Declaration (1907-1924)	Naturalized	27.2	0.35	26.5	27.9
High propensity (>0.69)	1930	Naturalized	30.3	0.397	29.5	31.1
High propensity (>0.69)	1940	Naturalized	29.4	0.41	28.6	30.2
High propensity (>0.69)	Declaration (1907-1924)	Not naturalized	26.8	0.55	25.7	27.8
High propensity (>0.69)	1930	Not naturalized	28.1	0.662	26.8	29.4
High propensity (>0.69)	1940	Not naturalized	27.3	0.706	25.9	28.7

Table A.6: Full-count censuses: OCCSCORE trajectories, as presented in Figure 3

Starting point	Year	Status by 1930	OCCSCORE	SE	CI Lower Bound	CI Upper Bound
First papers in 1920	1920	Naturalized by 1930	25.7	0.0795	25.6	25.9
First papers in 1920	1930	Naturalized by 1930	26.8	0.0861	26.6	26.9
First papers in 1920	1940	Naturalized by 1930	26.7	0.0903	26.6	26.9
First papers in 1920	1920	Never naturalized	23.8	0.152	23.5	24.1
First papers in 1920	1930	Never naturalized	23.6	0.161	23.3	23.9
First papers in 1920	1940	Never naturalized	23.7	0.179	23.3	24
Not a citizen in 1920	1920	Naturalized by 1930	24.8	0.0854	24.6	24.9
Not a citizen in 1920	1930	Naturalized by 1930	26.5	0.0913	26.3	26.7
Not a citizen in 1920	1940	Naturalized by 1930	26.8	0.0949	26.6	27
Not a citizen in 1920	1920	Never naturalized	22	0.0894	21.8	22.2
Not a citizen in 1920	1930	Never naturalized	22.5	0.0916	22.3	22.7
Not a citizen in 1920	1940	Never naturalized	22.6	0.102	22.4	22.8

Table A.7: Full-count censuses: Stratified trajectories, as presented in Figure 4

Starting point	Propensity category	Year	Status by 1924	OCCSCORE	SE	CI Lower Bound	CI Upper Bound
First papers in 1920	High propensity (>0.82)	1920	Naturalized by 1930	28	0.114	27.7	28.2
First papers in 1920	High propensity (>0.82)	1930	Naturalized by 1930	28	0.12	27.8	28.3
First papers in 1920	High propensity (>0.82)	1940	Naturalized by 1930	27.9	0.123	27.6	28.1
First papers in 1920	High propensity (>0.82)	1920	Never naturalized	26.8	0.27	26.3	27.3
First papers in 1920	High propensity (>0.82)	1930	Never naturalized	25.2	0.28	24.7	25.8
First papers in 1920	High propensity (>0.82)	1940	Never naturalized	25.3	0.292	24.8	25.9
First papers in 1920	Low propensity (<=0.82)	1920	Naturalized by 1930	23.3	0.1	23.1	23.5
First papers in 1920	Low propensity (<=0.82)	1930	Naturalized by 1930	25.3	0.12	25.1	25.5
First papers in 1920	Low propensity (<=0.82)	1940	Naturalized by 1930	25.3	0.13	25.1	25.6
First papers in 1920	Low propensity (<=0.82)	1920	Never naturalized	22	0.167	21.7	22.3
First papers in 1920	Low propensity (<=0.82)	1930	Never naturalized	22.6	0.191	22.2	23
First papers in 1920	Low propensity (<=0.82)	1940	Never naturalized	22.5	0.22	22.1	22.9
Not a citizen in 1920	Low propensity (<=0.61)	1920	Naturalized by 1930	21.9	0.11	21.7	22.1
Not a citizen in 1920	Low propensity (<=0.61)	1930	Naturalized by 1930	24.8	0.138	24.5	25
Not a citizen in 1920	Low propensity (<=0.61)	1940	Naturalized by 1930	25	0.147	24.7	25.3
Not a citizen in 1920	Low propensity (<=0.61)	1920	Never naturalized	20.6	0.0917	20.4	20.8
Not a citizen in 1920	Low propensity (<=0.61)	1930	Never naturalized	21.6	0.104	21.4	21.8
Not a citizen in 1920	Low propensity (<=0.61)	1940	Never naturalized	21.7	0.12	21.4	21.9

Not a citizen in 1920	High propensity (>0.61)	1920	Naturalized by 1930	26.6	0.115	26.3	26.8
Not a citizen in 1920	High propensity (>0.61)	1930	Naturalized by 1930	27.5	0.118	27.2	27.7
Not a citizen in 1920	High propensity (>0.61)	1940	Naturalized by 1930	27.8	0.122	27.5	28
Not a citizen in 1920	High propensity (>0.61)	1920	Never naturalized	25.3	0.187	24.9	25.6
Not a citizen in 1920	High propensity (>0.61)	1930	Never naturalized	24.5	0.175	24.1	24.8
Not a citizen in 1920	High propensity (>0.61)	1940	Never naturalized	24.5	0.182	24.1	24.8