The Making of Modern America:
Migratory Flows in the Age of Mass Migration[*]

Oriana Bandiera[†]   Imran Rasul[‡]   Martina Viarengo[§]

June 2011

Abstract

We provide new estimates of migrant flows into and out of America during the Age of Mass Migration at the turn of the twentieth century. Our analysis is based on a novel data set of administrative records covering the universe of 24 million migrants who entered Ellis Island, New York between 1892 and 1924. We first use these records to measure inflows into New York, and then scale-up these figures to estimate migrant inflows into America as a whole. Combining these *flow* estimates with census data on the *stock* of foreign-born in America in 1900, 1910 and 1920, we conduct a demographic accounting exercise to estimate out-migration rates in aggregate and for each nationality-age-gender cohort. The accounting exercise overturns common wisdom on two fronts. First, we estimate flows into the US to be 20% and 170% higher than stated in official statistics for the 1900-10 and 1910-20 decades, respectively. Second, we estimate the rate of out-migration from the US to be 76% during 1900-10 and close to 100% during the turbulent 1910-20 decade. These figures are between two and three times larger the official statistics estimates of around 35% in each decade. That migration was effectively a two-way flow between the US and the sending countries has major implications for understanding the potential selection of immigrants that chose to permanently reside in the US at the turn of the twentieth century, their impact on Americans in labor markets, and institutional change both in America and sending country economies.

Keywords: Ellis Island, migration accounting, migratory inflows and outflows.

JEL Classification: F22, N31, N32, O15.

[†]Department of Economics, London School of Economics and Political Science. E-mail: o.bandiera@lse.ac.uk.

[‡]Department of Economics, University College London. E-mail: i.rasul@ucl.ac.uk.

[§]JFK School of Government, Harvard University. Email: viarengo@hks.harvard.edu, m.g.viarengo@lse.ac.uk.

1 Introduction

The US is regarded by many social scientists as the ultimate multicultural society: 31 million individuals were recorded as foreign-born in the 2000 Census, corresponding to 12% of the population, and the US remains the primary destination for immigrants from developing countries [Hanson 2009]. The sheer magnitude of migratory flows into the US, the diversity of immigrants in terms of countries of origin, and the long period over which substantial migratory flows have taken place, marks out the US as an almost unique host country to study in terms of the economic impacts of migration on migrants themselves, natives in the host country, and sending country economies.[1]

This paper provides new evidence on the extent and nature of mass migration into *and out of* the US at the turn of the twentieth century, a period often referred to as the 'Age of Mass Migration', reflecting the fact that tens of millions of individuals migrated to the US from Europe and further afield at this time. Indeed, there are few if any comparable episodes of history in which such large-scale voluntary migration has occurred.[2] A key innovation in our analysis is to use novel data based on administrative records from passenger ship manifests on the universe of individuals that entered the US during this period through its main port of entry, Ellis Island, New York. These records provide details on over 24 million migrants that entered between 1892 and 1924. These individuals are very much those who laid the foundations for modern America – there are an estimated 100 million living descendants of these Ellis Island immigrants today.[3,4]

Despite the economic and historical importance of migration at the turn of the twentieth century, the official statistics commonly used in academic research are known to be measured with considerable error [Jerome 1926, Kuznets and Rubin 1954, Hutchinson 1958, Thomas 1973, McClelland and Zeckhauser 1983, Schaefer 1994, Carter *et al.* 2006] for reasons discussed below. The detailed administrative records we exploit allow us to present alternative estimates of migrant inflows and outflows. In doing so, we draw a very different picture of transatlantic migratory flows than has previously been recognized by scholars.

Our analysis proceeds in three stages. First, we present descriptive evidence on patterns of

---

[1]There are some smaller countries in which the percentage of foreign-born individuals is higher. Among the EU15 countries, none had a foreign-born population larger than 9% in 2006 with the exception of Luxembourg.

[2]In terms of contemporary mass migrations, it is estimated that 8 million Latinos entered the US during the 1980s, and .6 million individuals entered the UK from the A8 countries between 2004-6. Historically, similar quantitative migratory flows have been observed during episodes of either forced migration or when binding constraints are relaxed. For example, in the four years after the India-Pakistan-Bangladesh partition 18 million individuals are estimated to have migrated out of India [Bharadwaj *et al.* 2008]; Russian migrants into Israel corresponded to around a 12% population increase between 1990-4 [Friedberg 2001]. Finally, there were an estimated .6 million migrants from East to West Germany over the period 1989-90 [Heiland 2004].

[3]The immigration station at Ellis Island opened in 1892, and grew dramatically over the subsequent three decades. It lost the principal function for which it was established in 1924, when the second Quota Law was introduced. This legislative act provided for the examination of prospective migrants at American Consulates overseas. The Ellis Island station closed in 1954 [Pitkin 1975].

[4]There are few economics studies that have made use of passenger ship manifests from other ports: (i) Barde and Bobonis [2006] present evidence on detention by nationality at Angel Island, San Francisco, the main port of entry for Asians, for the period 1913-19; (ii) Puerta [2005] presents evidence on chain migration of one million Italians to Buenos Aires from 1882 to 1920; (iii) Ganguli [2010] presents evidence from Ellis Island ship manifests on the effect of literacy tests on inflows.

entry into Ellis Island by year of arrival. We do so in aggregate and for cohorts defined along three dimensions: nationality, age and gender.

Second, we then combine these measures based on administrative records with other data sources to provide a scaled-up estimate of the total inflow of immigrants into America as a whole, that accounts for: (i) missing data in administrative records and other corrections for potentially mis-coded nationalities; (ii) expulsion or death at Ellis Island; (iii) inflows from other sea ports of entry into America; (iv) inflows over land via Canada and Mexico. Each adjustment is described in detail in the Appendix.

Third, to estimate the number of individuals that have out-migrated from the US during this period we use demographic accounting techniques that relate changes in population stocks and migratory flows [Warren and Peck 1980, Jasso and Rosenzweig 1982]. This involved combining the *flow* estimates of immigrant entry into the US between 1900 and 1920 with the *stock* of foreign-born resident in the US at census date 1900-1910-1920, again by nationality-gender-age cohort. Taking the ratio of out-migrants to in-migrants from the same cohort and over the same decade, we then provide novel estimates for out-migration rates for each decade, 1900-10 and 1910-20. These can be directly compared to official statistics compiled at the time [Ferenczi-Willcox 1929], that have been the foundation for earlier academic work.[5]

Our main results are as follows. First, using only the administrative records on entry in Ellis Island, we find the number of immigrant arrivals into New York to be 18% higher than official statistics record between 1900-10, and around 50% higher than is officially recorded during the more turbulent decade of 1910-20. After scaling-up these figures to estimate migrant inflows into America as a whole, we find that our preferred estimate implies the number of immigrant arrivals is 20% higher than official statistics between 1900-10, and more than 170% higher than is officially recorded during 1910-20.

Second, our demographic accounting exercise implies that the number of *out*-migrants from America is mis-measured with even more severe error that are the migrant inflow estimates: we document the number of out-migrants to be three times higher than is officially recorded for the 1900-10 decade, and more than seven times higher for 1910-20. Combining our estimated inflows and outflows by decade, we then find that the rate of out-migration from America, namely the ratio of the number of out-migrants to the number in-migrants from the same cohort and over the same decade to be 76% for the 1900-10 decade, and close to 100% over the 1910-20 decade. Although these results do *not* imply that it is the *same* individuals who enter and exit America within each decade, the results do suggest there is a considerable degree of turnover in the foreign-born population in America at the turn of the twentieth century, and potentially large migrant flows back to European sending countries.

Our estimated rates of out-migration differ substantially from the conventional wisdom among scholars, that have previously thought out-migration rates to be closer to 35% in each decade [Hatton and Williamson 2005]. Given the discrepancy between what we document based on

---

[5]We refer to out-migration rather than return migration because we cannot infer whether individuals return to their original country of origin, or whether they use America as a stepping stone to a third country.

administrative records and what has been officially recorded, it is important to validate our findings using another approach. One way to do so is to examine how the stock of foreign-born varies between census dates. More precisely, US census data suggests the absolute increase in the foreign-born population was from 13.5 in 1910 to 13.9 million in 1920, corresponding to only a 3% increase. In short, *regardless* of the actual migrant inflow estimate – be it from official statistics or administrative records – census data alone suggests that for this decade the undisputed inflow of millions of migrants *must* have been matched by a similarly sized out-flow of foreign-born individuals. Indeed, for this decade our accounting exercise suggests an out-migration rate of just under 100%.

At the core of this paper is an attempt to accurately *measure* migratory inflows and outflows from America during the age of Mass Migration. A natural next step is to try to *explain* these rates of out-flow in relation to economic conditions in sending and receiving countries. While it is certainly the case that for the time period we study, the sheer magnitude of migratory flows into the US is sufficiently large to likely have large and permanent effects on the US economy and sending country economies, such research questions cannot be addressed without first establishing rates of out-migration, as we do in this paper. As the extent of out-migration underpins nearly all aspects of the economics of migration – such as migrants' investments into their human capital and their consequent substitutability with native labor, we view this analysis as the foundation upon which to build a broader future research agenda. We sketch this agenda in the final section.

The paper is organized as follows. Section 2 describes our data sources and presents descriptive evidence. Section 3 presents estimates of inflows both to New York and to the US as a whole, comparing them with official statistics. Section 4 describes the migration accounting exercise and presents the results on out-migration in aggregate, and by nationality-gender-age cohorts. Section 5 concludes. The Appendix places the time period under study into a wider historical context, discusses the established evidence on whether travel times and travel costs relative to wages, are likely to have been sufficiently high to deter out-migration during the period we study, and provides further details the assumptions and additional data underlying the accounting exercise.

2 Data Sources

2.1 Ellis Island Administrative Records

Core to our analysis are the administrative records of passenger ship manifests from Ellis Island, obtained from the American Family Immigration History Centre. The database contains the universe of 24 million individuals whose names appear on the original ship passenger manifests for the Port of New York between 1892 and 1924. The Manifest of Immigrants Act 1819 required that from 1820, the master of every vessel entering a US port list *each* passenger taken on at any foreign port by name, gender, age, occupation, whether or not they *intend* to become a US citizen, and country of origin [Carter and Sutch 1998]. Passenger lists are complete in that *all* ships and *all* passengers are recorded including foreign nationals, US nationals, foreign born nationals that have

acquired US citizenship, and those intending to stay temporarily.[6] Passenger lists were prepared by the ship's captain before the ship departed. The list was created from passenger documents that were required for entry into America. In many cases, these documents would have been created at the time the passenger purchased their ticket. Passengers with incomplete documentation were sent back at the shipping companies expense, thus incentivizing shipping companies to ensure passenger documentation and ship manifests were accurate and complete.[7]

The electronic version of the administrative records from ship manifests include information on the passenger's first name, surname, age in years, gender, place (town and country) of last residence outside the US, date of arrival, whether the individual is a new arrival to the US or a returnee, marital status, ethnicity, the name of vessel, the original port of departure and other ports at which the ship stopped. Over time passenger lists expanded to systematically encompass information on beliefs about politics, marriage, health, literacy, and final destination in the US, although this information is not electronically recorded in the manifests we exploit. Figure 1 shows an example of a ship manifest. The information available in electronic format is in solid boxes; the information we do not have is in dashed boxes. Most relevant for this study is that the available information allows us to measure aggregate inflows of individuals in cohorts defined along four dimensions: nationality, gender, age and year of arrival. Throughout we define nationalities using pre-1918 country borders.

2.2 Census Data

We use US census data from 1880 to 1920 in our analysis. To estimate the stock of foreign-born in America prior to the great wave of migration captured in the Ellis Island administrative records, we use the 100% sample 1880 IPUMS census [Ruggles *et al.* 2009]. This shows there to be 50 million individuals resident in the US on census day 1880, 6.7 million (13%) of which are foreign-born. There are 90 nationalities among the foreign-born population resident in the US in 1880. Recall that the size of the immigrant inflow into the US through Ellis Island alone was over 24 million between 1892-1924, corresponding to almost half the total US population in 1880.

The other IPUMS census samples we use are 1900 (5%), 1910 (1.4%), and 1920 (1%) [Ruggles *et al.* 2009]. In each of these years, around 14% of the total population is foreign-born. Nationality of birth and year of arrival into the US are recorded in these censuses. The nationality of birth is recorded even if individual has obtained US citizenship by census day, which applies to 31% of individuals, with a further 5% recorded to be in the process of acquiring citizenship. As the accounting exercise we conduct applies to those individuals that might have left America after

---

[6] The data was entered by 12,000 volunteers of the Church of Jesus Christ of Latter-day Saints. We have checked for duplicate records, defined to be those in which an individual is recorded to have the same first and second name, age, gender, marital status, place of residence, and to have arrived on the same ship on the same date. Far fewer than one tenth of one percent of records are duplicates so defined. Members of the ship's crew were included on the manifests from 1918 onwards.

[7] Shipping companies were also liable for fines of $200 per alien if they were deemed to have been excludable based on the literacy test introduced in 1917. Indeed, as Goldin [1994] reports, in 1917 fines were levied for only 192 excludable illiterate alien arrivals out of the hundreds of thousands that arrived.

having entered, we do not use information on children of immigrants born in the US and so we make no assumptions on immigrant fertility in the US. Over census years 1900-20, this gives a total of 916,773 foreign-born individuals.

2.3 Descriptive Evidence

To begin with, we present some broad descriptive evidence on the migration patterns from the Ellis Island administrative records. As these have never been previously used for academic research, it is useful to establish whether the migration patterns are broadly consistent with other data sources and previous research. Figure 2Ai shows the total number of arrivals into Ellis Island over the span of data: 1892 to 1924, for the ten countries from which the greatest number of immigrants originate over this time period. As a point of comparison, Figure 2Ai also shows for each nationality, the size of the foreign-born population from that country recorded in the 100% sample of the 1880 US census.

Figure 2Ai highlights the sheer scale of migration during the study period. In 1880 the *total* foreign-born population in the US was 6.7 million; between 1892 and 1924 almost that many individuals arrived from Italy into Ellis Island alone, and 24 million arrived into Ellis Island overall. Moreover, inflows into Ellis Island by nationality for the three decades after 1892 do not closely mirror the initial stocks of these nationalities in the US in 1880. The largest group of immigrants into Ellis Island are Italians, over five million of whom entered over this period. Yet there were approximately zero Italians resident in the US in 1880. This point is further emphasized by comparing Figure 2Ai to Figure 2Aii that shows the ten largest nationalities in the US 1880 census. Some nationalities that were highly prominent in the US in 1880, such as those from France, Bohemia and Switzerland went into relative decline in the decades after 1890. As well recognized by historians, these nationalities were gradually replaced by migrants from Southern and Eastern Europe such as Italy, Russia0, Greece and Spain.[8]

Although Figure 2A focuses on the inflow of foreign-born individuals, there is also a large inflow of US citizens into Ellis Island, actually comprising the fourth largest national group in our data. The vast majority of these individuals are return migrants that have previously acquired US citizenship. For reasons explained below and in the Appendix, the accounting exercise we conduct is based on individuals that arrive to America for the *first* time, rather than returnees. This distinction is recorded in the administrative records.[9]

Figure 2Bi shows migrants' age distribution by gender. To see how the age distribution of new arrivals differs from those already settled in the US, Figure 2Bii shows the age distribution by gender, for the foreign-born population in the US in 1880. As expected male migrants are slightly

---

[8]The Figure shows that some highly prominent nationalities in the US in 1880 such as Canadians and Chinese, do not show up as having significant migrant inflows through Ellis Island, New York. When we later attempt to scale-up immigrant inflows from Ellis Island to those for the US as a whole, we will need to account for the differential patterns, by nationality, of migrant inflows into the US by land and sea.

[9]Dupont *et al.* [2009] use a variety of data sources to show the number of Americans going abroad rose dramatically after the first world war, but that for most of our sample period, the number of American-born travellers was negligible relative to the total flow.

older than females. Both are younger than the foreign-born population of the same gender already resident in the US in 1880, and the age distribution among new migrants is more compressed than among the foreign-born in 1880. We also note that around two thirds of arrivals are men, and the majority of migrants are single. However, around 20% of migrants are single women with females being more likely to be single than married.[10]

Figure 2Ci provides time series evidence on the total number of immigrants into Ellis Island each year, as indicated on the left hand axis. There are large year-to-year fluctuations in migrant numbers, including a collapse in migrant numbers in 1908. This is also evident in official statistics from the time and has been attributed to the spike in US unemployment that year caused by the 1907 financial crisis [Deltas *et al.* 2008]. These large year-to-year fluctuations are hard to reconcile with migration decisions being motivated by *lifetime* differences in utility between sending and receiving country. Such volatility might however be more in line with models of return migration in which individuals only intend to stay in the US temporarily and so the short run economic conditions can impact upon the choice of when to migrate. We return to discuss models of out-migration in the conclusion.[11]

The right hand axis in Figure 2Ci shows the ratio of male to female migrants by arrival year. Until World War One this is relatively constant with around twice as many male migrants arriving annually. This ratio spikes during the later years of data and then appears to revert to a permanently higher level in the 1920s. This long run change might of course in part be related to the 1917 Immigration Act that excluded illiterates and raised the head tax for migrants, leading to relatively higher barriers to entry for women in the short run, and potentially, a change in the composition of female migrants in the longer term.[12]

Figure 2Cii shows how immigrant ages evolved over the study period. In 1892 the average male migrant was just over 26 years old, and the average female migrant was just under 24. Over time, migrants of both genders became older. Remarkably, the mean age of female migrants converges to that of men by the end of the study period. This is correlated with two compositional changes over time, namely the increasing share of married women, who were presumably joining their husbands in America, and migrants originating from Southern and Eastern Europe.

Taken together, these descriptives provide a flavour of the general migration patterns into Ellis Island at the turn of the twentieth century. We now move to the core of our analysis: to account

---

[10]In 1882 the US enacted legislation denying entrance to people who might become public charges, say because they were unable to support themselves. While in theory this policy was gender-neutral, in practice any unaccompanied woman of any age, marital status, or background might be questioned as a potential public charge because she appeared to lack a male provider. Indeed, Ellis Island officials regularly detained women travelling on pre-paid tickets to join husbands in New York if the man failed to show up in person to claim their 'dependents'. Unaccompanied women could only leave Ellis Island with a male accompanying them. This does not imply single women had to get married at Ellis Island however, as is widely perceived [Gabaccia 1984].

[11]As described in Reinhart and Rogoff [2008], the financial crisis of 1907 relates to a six-week stretch of runs on banks predominantly in New York in October and early November of 1907. It was triggered by a failed speculation that caused the bankruptcy of two brokerage firms. As a result, real US GNP declined 12% between the second quarter of 1907 and the first quarter of 1908. Between 1907 and 1909 the value of migrant remittances also plummeted and this might have restricted migrant inflows over this period [Esteves and Khoudour-Casteras 2009].

[12]If marriages are endogamous, such large changes in sex ratios might alter marriage market competition and subsequently affect out-migration [Angrist 2002].

for how out-migration rates change over time. To do so we first describe official statistics related to historic migration inflows and outflows, and highlight some of the known concerns with these official series.

2.4 Official Migration Statistics

Official statistics on migration exist as a result of the Passenger Act 1819, that required the captain of each vessel arriving from abroad to deliver a manifest of all passengers taken on board in a foreign port, as described previously. Copies of these manifests were to be transmitted to the Secretary of State by the shipping companies. These steamship companies were relied upon to provide unofficial data about numbers of arriving and departing passengers from ship manifests. From 1892 onwards, these reports were collected and abstracted by the Office or Bureau of Immigration, which is today part of the Immigration and Naturalization Services (INS). Academics rely on the collation of these statistics by Imre Ferenczi, whose work was conducted under the auspices of the International Labor Office in the early 1920s and published in 1929 by the NBER [Willcox 1929]. These official statistics, referred to as Ferenczi-Willcox [1929], bring together all official data series on post-1820 international migration then available in published form or supplied by governments.

While these official statistics have been used by policy makers and academics for nearly eighty years now, potential defects in these series have also long been recognized [Jerome 1926, Kuznets and Rubin 1954, Hutchinson 1958, Thomas 1973, McClelland and Zeckhauser 1983, Schaefer 1994]. Following these discussions, we describe the main sources of measurement error in migrant flows over the period we study, and how the Ellis Island administrative records allow us to deal with each concern.

First, Hutchinson [1958] and others have argued there was careless collection of ship manifests by port officials. Such errors were compounded by a failure of customs collectors to forward Passenger Abstracts quarterly to the Department of State, and further compounded by a failure of State Department clerks to include all Passenger Abstracts in their annual statistical reports. Together, these errors could aggregate up to cause severe *underestimates* of migrant arrivals. The Ellis Island administrative records contain manifests for all ships that arrived between 1892 and 1924. As discussed above, ship captains faced strong incentives to collect documents from their passengers and compile accurate ship manifests, as failure to do so implied that the shipping companies would have to take passengers with incomplete records back at their own expense.

Second, the official statistics supplied by shipping firms often excluded first-class passengers for the early decades. Moreover cabin-class (second-class) passengers were only first included with migrants in steerage-class (third-class) from 1904 onwards. Furthermore, prior to 1904, not only were statistics from ship manifests restricted to passengers in steerage class, but they were further restricted to *aliens* traveling as steerage passengers, rather than also including US citizens, or returning foreign-born nationals who had earlier acquired US citizenship. As a result, this leads to further under-counting of migrant flows by nationality and in aggregate.[13] In contrast, ship

---

[13]Precisely the same points are made in Barde and Bobonis [2006] who compare the coverage of passengers

manifests contain a complete list of *all* passengers, regardless of class of travel and nationality.

To get a sense of the potential bias induced, we note that Hyde [1975] reports that by 1880 the largest transatlantic liner could carry 300 cabin and 1200 steerage passengers. Keeling [2008] reports a similar division of passengers by class based on records of the Cunard Steamship Company, one of the largest carriers, and Coyne and Murphy [2007] present evidence from steamship arrivals into New York in 1913 from eight leading steamship firms showing that 11% of passengers travelled first class and 17% were in second class. As carriers faced intense competition, to cover fixed costs, large numbers of passengers had to be carried. This led to a steady increase in ship size that averaged 3000 gross tons in each succeeding decade. As cabin rates were approximately double those for steerage, shipping firms might have had incentives to provide relatively more cabin-class places over time. Indeed, Keeling [1999] argues that after 1900 the fastest growing segment of travel was for second-class routes from both Southern and Northern Europe.[14]

This source of measurement error in migrant inflows would also bias official estimates of outflows if those in steerage and other classes had different propensities to out-migrate, as is plausible under many models of out-migration.[15] Our estimates address such concerns because they are based on comprehensive ship manifests that cover passengers in *all* travelling classes, and they contain precise information on place of last residence, US citizenship, and whether the migrant is newly arriving to the US or is a returnee. We therefore shed light on the cumulative degree of mismeasurement arising from these two concerns, by comparing official statistics on entry specifically into New York vis-à-vis the implied inflows from the administrative records from Ellis Island. This comparison is obviously not sensitive to any assumptions on our part on how to scale-up migrant numbers from Ellis Island to the US as whole.

Other concerns also exist with official migration records from this period. Foremost among these are concerns relating to the accurate recording of immigrants via all sea and land ports of entry. When scaling-up the migrant inflows based on administrative records into Ellis Island to those for the US as a whole, we use a variety of data sources to take account of inflows from other sea ports and land crossings from Canada and Mexico, as described in the Appendix.

The two primary sources of measurement error described above arise for measuring both migrant inflow and outflows. On inflows, Thomas [1973] has previously provided evidence on the extent of the shortfall in US immigration statistics caused by these combined sources of error for the mid nineteenth century. He shows that during that period, major port cities such as New York made their own yearly tally of immigrant arrivals, and remarkably, these figures often exceeded federal totals for *all* east coast ports. This is precisely what we will later document to be the case for the 1910-20 decade. Over both decades, our results suggest even more severe errors in officially

---

recorded on ship manifests into Angel Island to those recorded in official statistics. They conclude that the former included passengers in first and cabin class, as well as US citizens.

[14]Frey *et al.* [2011] present evidence that of the 1316 passengers (including maids) on the Titanic, 325 travelled in first class, 285 in second class, and 706 in third class.

[15]Such models include those that emphasize: (i) return migration being planned as part of an optimal life cycle residential location sequence [Borjas and Bratsberg 1996, Dustmann and Weiss 2007]; (ii) target income earner models [Yang 2006]; (iii) erroneous beliefs of migrants, or negative shocks in the US [Pessino 1991, Borjas and Bratsberg 1996]; (iv) differential returns to multi-dimensional skills [Gould and Moav 2009, Dustmann *et al.* 2010].

recorded number of *out*-migrants. This is not altogether surprising given that governments have even weaker incentives to accurately record migrant outflows, a situation that continues today [Warren and Peck 1980, Keeling 2006].

Given these concerns, there have of course been previous attempts to amend the series in Ferenczi-Willcox [1929], although none of these have been able to use the detailed administrative records we exploit. One of the most important attempts to do so was conducted by Kuznets and Rubin [1954], who re-estimated inflows and outflows.[16] As we discuss later in more detail, they find rates of out-migration to be around 50%, rather than the 35% rate implied by Ferenczi-Willcox [1929] over our study period. Our estimates suggest out-migration rates are even higher than Kuznets and Rubin [1954] concluded, being closer to 76% and 100% in the two decades studied.

3   Estimated Inflows

3.1   Inflows into Ellis Island, New York

We first focus on arrivals into Ellis Island New York so that we can make a direct comparison between our estimated immigrant inflows based on administrative records and those from official statistics. We provide two estimates of migrant inflows: the first is based on the raw administrative statistics from which *none* of the adjustments described in the Appendix are made. The second estimate, which is our preferred measure, corrects for missing data, other potential errors in recorded nationalities, and exclusions.

Three other points are important for the comparison between our estimates and official records. First, in the official records in Ferenczi-Willcox [1929] no account is taken of survival probabilities. Hence we simply sum the total number of migrant arrivals into New York over a decade, with no regard for whether these individuals are still alive at the end of the decade. Second, the official statistics make no distinction between new and returnee arrivals. For the purposes of comparison we sum across new and returning arrivals from the administrative records. Finally, we note that official statistics on foreign-born arrivals are actually broken down into: (i) immigrant foreign-born arrivals, or those who intended to settle in the US; (ii) non-immigrant foreign-born arrivals, or admitted aliens who declared an intention not to settle in the US. We can therefore compare our estimates with those for immigrant foreign-born arrivals as well as total foreign-born arrivals.

---

[16]Kuznets and Rubin [1954] developed "an approximation to net immigration," (series AD22), by calculating the difference between arrivals and departures of alien passengers. They use the official data on arrivals and departures for the period beginning in 1908 and make estimates for the period before this year. For the period 1870 to 1890, they use the official data on arrivals (series AD23). They estimate departures by starting with the official number of Americans returning from abroad (assuming that these Americans stay abroad for one year), some assumed mortality rate, and a guess that permanently departing citizens constituted 0.5 percent of all citizen departures. For 1901 to 1907, they estimated emigration by extrapolating the 1908-1914 ratio of departures to arrivals using the official data on arrivals for 1900-1907. For the period 1908-1945, Kuznets and Rubin [1954] used the official data but reported the result rounded to the nearest thousand.

3.1.1 By Decade

The top half of Table 1 refers to 1900-10, and the bottom half to 1910-20. Column 1 shows the official statistics from Ferenczi-Willcox [1929], Columns 2 and 3 show our preferred and raw estimates respectively. To make the comparisons as detailed as possible, the rows correspond to breakdowns by: (i) immigrant foreign-born arrivals; (ii) non-immigrant foreign-born arrivals; (iii) foreign-born arrivals, summed across immigrant and non-immigrants; (iv) US-citizen arrivals; (v) arrivals summed across foreign-born and US-citizens.

The first row shows that for the 1900-10 decade, official statistics state there were $7,431,670$ foreign-born immigrant arrivals into New York. In contrast, our preferred estimate in Column 2 based on administrative records is $8,968,628$, that is 21% higher than Ferenczi-Willcox [1929] as shown in Column 4.

The next row shows that in this decade $713,749$ foreign-born arrivals are officially recorded as having the intention of temporarily remaining in the US, i.e. they are recorded as non-immigrant arrivals. This inflow corresponds to around 9.6% of the magnitude of immigrant foreign-born arrivals. Summing across immigrants and non-immigrants, the next row shows that official statistics suggest in total there were $8,145,419$ foreign-born arrivals into New York. In contrast, our estimate based on administrative records is between 8 and 10% higher, depending on whether the raw or preferred measure is used.

The next row focuses on inflows of US-citizens into New York. Here we see that official statistics actually record significantly *more* US citizen arrivals than do administrative records. Indeed official statistic estimates are around a third higher than those based on administrative records. The discrepancy might arise from the different times at which nationality is recorded in the two data sources. In the ship manifests on which our estimates are based, nationality is recorded *before* entry to the US and is likely to refer to nationality of birth. If official statistics confound nationality of birth with individual's reported *intention* to obtain US citizenship, this can lead to an over-estimate of US-citizen arrivals. As previously noted, US census data suggests that indeed around one third of foreign-born migrants to the US obtain US citizenship by census date. This would almost perfectly account for the approximate one third larger recorded inflows of US citizens in official records relative to the administrative records.

The final row in the upper panel then sums across both categories to give an overall indication of mis-measurement of migrant inflows across temporary and permanent arrivals, and across all nationalities. In this decade our preferred estimate is 5% higher than the official statistics as recorded in Ferenczi-Willcox [1929], implying that while some of the differences might be due to different classifications of US citizens, over half a million migrants who are recorded coming into Ellis Island are missing from the official statistics.

The lower panel of Table 1 repeats the analysis for 1910-20. In this decade the discrepancy between officially recorded foreign-born arrivals and those inferred from the administrative records is more severe. Our preferred estimate of all foreign-born arrivals (immigrants and non-immigrants) is 34% higher than the official statistics. In contrast to the previous decade, official statistics and

administrative records closely correspond in terms of the numbers of US-citizen arrivals over the decade. Finally, the last row shows that summing over the types of migrant and nationalities, total arrival numbers are underestimated by 20 to 28% depending on the estimate derived from administrative records.

To summarize, we find official statistics to generally underestimate migrant inflows, as has long been suggested among scholars. We provide novel evidence on the extent of this mis-measurement that presumably arises predominantly from incomplete collection of ships manifests, the exclusion of first and cabin-class passengers, and not accounting for deported arrivees. The extent of under-counting we uncover is large especially for the more turbulent decade of 1910-1920. While there might be slightly alternative methods of making the adjustments to derive our preferred estimate, it is unlikely that such alternatives would suggest official estimates are indeed accurate.

3.1.2   By Year of Arrival

To get a better sense of how discrepancies between our preferred estimate and official statistics vary by year of arrival, we provide the time series for both measures over the entire period that can be compared, 1892-1924. Figure 3A shows the time series for total arrivals (foreign-born and US-citizen arrivals) from both sources. Up until 1916 the two aggregate series track each other relatively closely. From 1917 onwards, official statistics far underestimate total arrivals from administrative records. Hence the main sources of bias – incomplete collection of ships manifests, the exclusion of first and cabin-class passengers, and not accounting for exclusions, are likely to have become more severe from the end of the first world war onwards.

A similar turning point is highlighted in Figure 3B that shows the time series for US-born arrivals by year. Prior to 1917 official statistics tend to record more US-citizen arrivals into Ellis Island than suggested by administrative records. However the situation is reversed from nearly every year from 1917 onwards. The administrative records therefore suggest that from 1917 onwards, the recording of US-citizen arrivals into Ellis Island was subject to the same sorts of mis-measurement as for foreign-born immigrant arrivals.

3.2   Inflows into America

Table 2 presents an analogous set of comparisons between our estimates based on the Ellis Island administrative records and official statistic estimates of inflows into the US from Ferenczi-Willcox [1929]. We again provide two estimates: the first is based on the raw administrative statistics from which *none* of the adjustments described above and in the Appendix are made. Hence this essentially assumes that New York is the *only* entry point into the US; the second is our preferred estimate that corrects for the following factors: (i) missing data and other potentially mis-coded nationalities; (ii) expulsion or death; (iii) inflows from other sea ports; (iv) inflows over land via Canada and Mexico.

Inflows from other sea ports and over land are quantitatively the two most important adjustments. To correct for the former we use official statistics on yearly inflows by port, and assume

that, while the levels might be underestimated, the ratio between immigrants into Ellis Island and into other ports is correct. Our implicit assumption is that the main sources of errors in official statistics (careless collection of ship manifests, exclusion of first and second class passengers) are proportional to the true number of arrivals in each port. To refine our adjustment we use historical sources to identify the nationalities which were more likely to use other ports, and derive nationality specific adjustment factors as described in the Appendix. To account for inflows over land we use Canadian official statistics by nationality and year and, using estimates from the literature, assume that 40% of migrants arriving in Canada are US bound. On inflows from Mexico, we make the conservative choice of assuming no inflows (either legal or illegal) so as to likely underestimate total immigrant inflows into America.

Table 2 presents the results following a similar formatting as Table 1. For the 1900-10 decade, the first row shows official statistics record 9, 719, 358 foreign-born immigrant arrivals into America. In contrast, the preferred estimate based on administrative records is 41% higher, at over 13.7 million.[17] The next two rows demonstrate that accounting for non-immigrant arrivals, we still find that the number of total foreign-born migrants in official statistics of Ferenczi-Willcox [1929] is 28% lower than implied by our preferred estimate. On US-citizen arrivals, we note a similar pattern as that in Table 1 for this decade: official statistics *over-report* such arrivals relative to administrative records based on nationalities as recorded in ship manifests. Taking into account all arrivals of all nationalities, the fifth row shows that total arrivals are around 20% higher using our data than is officially recorded over this decade.

The lower panel of Table 2 then repeats the analysis for the 1910-20 decade. In this period the discrepancy between officially recorded foreign-born arrivals and those inferred from administrative records is far more severe than for the previous decade. Remarkably, our preferred estimate of all foreign-born immigrant arrivals is 178% higher than the official statistics. The true scale of mis-measurement becomes apparent when we note that in this decade, our derived estimate for immigrant arrivals into New York alone is almost of equal size as is officially recorded for the US as a whole. We are however certainly not the first to note that mis-measurement of historic inflows into America might be of such orders of magnitude. For example, Thomas [1973] shows that major port cities such as New York made their own yearly tally of immigrant arrivals in the mid 19th century and these figures often exceeded federal totals for *all* east coast ports.

Moreover, minor ports gain importance over this decade, so that the ratio of entries through Ellis Island falls from 76% during the 1900-1910 period to 55% during 1910-1920 (and below 50% after 1916). Taken together, the fact that official statistics grossly underestimate migrant flows into Ellis Island, and that our estimates suggest a smaller percentage of all inflows into America went through Ellis Island than the official statistics imply, helps explain the large discrepancy between the official statistics and our estimates.

---

[17]To get a sense of how the total inflows into America are affected by each of the adjustments made we note the following. For the 1900-10 decade, the raw Ellis Island records suggest there are 8.8 million arrivals, as shown in Column 3 of Table 2. Correcting for missing values, US citizenship and expulsion and death, raises the estimate to 8.97 million. Adjusting for other ports of entry raises it to 12.3 million, adjusting for inflows from Canada and Mexico raises it to 13.7 million, as shown in the preferred estimate in Column 2 of Table 2.

As the next two rows demonstrate, accounting for non-immigrant arrivals, we still find that the number of total foreign-born migrants into the US in official statistics to be 126% lower than our preferred estimate. Unlike for the previous decade, US-citizen arrivals are also under recorded by around 15% in official statistics relative to our estimates based on administrative records. Finally, taking into account all arrivals of all nationalities, the fifth row shows that the our preferred measure of the number of total arrivals is slightly more than double what is officially recorded for the 1910-20 decade.

4 Estimated Outflows

4.1 Migration Accounting

The results so far show that many more individuals migrated to America at the turn of the century than previously recognized. We now turn to understand how many of them chose to remain there. To estimate migrant outflows we conduct a demographic accounting exercise that relates changes in population stocks and migratory flows [Warren and Peck 1980, Jasso and Rosenzweig 1982].[18] This procedure allows us to infer the number of individuals that must have out-migrated from America between any two census dates, when the stock of foreign-borns is measured. Taking the ratio of out-migrants to in-migrants allows us to provide novel estimates of out-migration rates by decade. The aggregate rates of out-migration rates can be directly compared to official statistics compiled at the time [Ferenczi-Willcox 1929], as well as other amendments of the official series such as Kuznets and Rubin [1954].

To begin with we define a cohort along three dimensions: gender, age and nationality. Such information is contained both within the Ellis Island administrative records and US Censuses in 1900, 1910 and 1920. In the administrative records we observe 118 different nationalities, namely there is at least one entrant from nationality n between 1900 and 1920. To match with detailed mortality rate data for the study period, we split ages into eleven age groups: (0, 1-4, 5-14, 25-34,...,85+). Hence in total we have 2596 *potential* nationality-gender-age (nga) cohorts of entry (118 nationalities $\times$ 11 age groups $\times$ 2 genders).

Consider first the stock of foreign-born individuals in cohort nga resident in the US in year t, denoted P_t^{nga}. The accounting exercise is based on the fact that the foreign-born population in the US among this nga cohort in the next year $t+1$ is given by,

$$P_{t+1}^{ng,a+1} = (1 - D_t^{ga})P_t^{nga} + I_{t+1}^{ng,a+1} - E_{t+1}^{ng,a+1}, \tag{1}$$

where D_t^{ga} is the mortality rate of foreign-born individuals of gender g and age a in year t, $I_{t+1}^{ng,a+1}$ is the migrant inflow of cohort nga between years t and $t+1$, and $E_{t+1}^{ng,a+1}$ is the outflow of migrants

---

[18]Studies based on more recent data have to distinguish between temporary visitors, such as students or extended business travellers, and permanent migrants. To do so, researchers have typically used INS data for a single entry cohort and then followed these individuals over time. No such issues arise for the time period we study when immigrants always had the possibility to permanently reside in the US.

from the US of those in cohort nga between years t and $t + 1$. By repeated substitution between two census years t and $t + 10$ we derive that the stock of foreign-born individuals in entry cohort nga that reside in the US in census year $t + 10$ is,

$$
\begin{aligned}
P_{t+10}^{ng,a+10} &= \prod_{j=0}^{j=9}(1 - D_{t+j}^{g,a+j})P_t^{nga} \\
&+ \sum_{k=1}^{k=9}\prod_{j=k}^{j=9}(1 - D_{t+j}^{g,a+j})(I_{t+k}^{ng,a+k} - E_{t+k}^{ng,a+k}) + \left(I_{t+10}^{ng,a+10} - E_{t+10}^{ng,a+10}\right).
\end{aligned}
\tag{2}
$$

Hence the cohort population stock in census year $t + 10$ is a function of survivors in this cohort from the previous census in year t, and survivors on census date in year $t + 10$ of the net inflows of migrants for each and every year since the previous census. Rearranging (2) then gives the total number of out-migrants from cohort nga between two census dates,

$$
\begin{aligned}
&E_{t+10}^{ng,a+10} + \\
&\sum_{k=1}^{k=9}\prod_{j=k}^{j=9}(1 - D_{t+j}^{g,a+j})(E_{t+k}^{ng,a+k})
\end{aligned}
=
\frac{\prod_{j=0}^{j=9}(1 - D_{t+j}^{g,a+j})P_t^{nga} - P_{t+10}^{ng,a+10}}{I_{t+10}^{ng,a+10} + \sum_{k=1}^{k=9}\prod_{j=k}^{j=9}(1 - D_{t+j}^{g,a+j})(I_{t+k}^{ng,a+k})}.
\tag{3}
$$

To be clear, the left hand side is the number of out-migrants in cohort nga in census year $t + 10$ that: (i) have out-migrated since the previous census in year t; (ii) are still alive *somewhere*. On this second point we note that we cannot identify whether individuals have returned to their home country, or use the US as a stepping stone before moving to some third country.

The administrative records distinguish between new and returnee migrants. In the Appendix we discuss in more detail why it is preferable to use data only on *new* immigrant arrivals to measure $I_{t+1}^{ng,a+1}$. We do so by considering various scenarios that differ in the year of arrival, departure and return relative to census dates t and $t + 10$. In all cases, the out-migration measure based on new immigrant arrivals is shown to be preferred as it avoids double counting migrants engaged in repeat migrations. In consequence, we mostly do not need to consider whether and when foreign-borns obtain US citizenship, as only information on new migrants, who should all be non-US citizens, is used for the accounting exercise.[19]

4.2 Census Cohorts

To implement (3) we need to measure the stock of foreign-born in a given cohort on two consecutive census dates: P_t^{nga} and $P_{t+10}^{ng,a+10}$. Census data from 1900, 1910 and 1920 allows us to also define cohorts by nationality-gender-age. To calculate the stock of foreign-borns in a cohort we have to take account of the clustered design of the census samples. Denote clusters as $c = 1, ..., L$, household (c, h) is the hth household in cluster c, n_c is number of sampled households in cluster c, and m_{ch} is the number of individuals in household (c, h). Hence the estimated population in

---

[19]We note that between 1900-20, 12.7% of migrants into Ellis Island are returnees. 87% of returnees are US citizens. Given the vast majority of returnees are likely to be foreign-born rather than American-born travellers [Dupont *et al.* 2009], this suggests those engaged in repeat migration obtained US citizenship at some point and is indicative of planned repeat migration.

cohort nga in census year t is,

$$\widehat{P}_t^{nga} = \sum_{c=1}^{L} \sum_{h=1}^{n_c} \sum_{i=1}^{m_{ch}} w_{ihct}^{nga}, \qquad (4)$$

where w_{ihct}^{nga} is individual sampling weight of person i in cohort nga census year t. Out of the 2596 potential cohorts described above, 1287 have at least one individual in them resident in the US between 1900 and 1920.[20]

4.3   Mortality Rates

The final component in (3) is the mortality rate, D_t^a. For years 1900 to 1939, mortality rates specific to race-gender-age-year are available for the US from US Vital Statistics Special Reports [1956], where races are white and other, and age groups are 0, 1-4, 5-14, 15-24,...,75-84, and 85+.[21] These age groups are those we use to define cohort dimension a. The accounting exercise assumes mortality rates among migrants are the same in America and the sending country. This is unlikely to be true so it is useful to check the sensitivity of our results to alternative choices of mortality rate. Two obvious choices are that migrants have the same mortality rate as whites in America, as assumed by Warren and Peck [1980] and Jasso and Rosenzweig [1982], or that migrants have mortality rates similar to those of other races.[22]

A third alternative is based on nationality specific mortality rates, that we have collected from a variety of sources [Tizzano 1965, Nugent 1992, Mitchell 2007]. They cover all the major European sending countries in our study period for census years 1900, 1910 and 1920. The obvious drawback

---

[20]Three further points are of note. First, there are no missing values for age or gender. Only 418 out of 916,773 observations having missing nationality of birth data in these IPUMS census samples. The individual sampling weight is the variable `perwtdet` in the IPUMS census samples. Second, given the data is a sample from the census, we can also construct $var(\widehat{P}_t^{nga})$ accounting for the census sampling frame as follows [Cochran 1977],

$$var(\widehat{P}_t^{nga}) = \sum_{c=1}^{L} (1 - \frac{n_c}{N_c}) \frac{n_c}{n_c - 1} \sum_{h=1}^{n_c} (w_{ihct}^{nga} - \bar{w}_{ct}^{nga})^2,$$

where N_c is the total number of households in the cluster population. $(1 - \frac{n_c}{N_c})$ is referred to as the finite population correction, and is equal to .95, .986, and .99 for the 1900, 1910 and 1920 censuses respectively. When constructing these variances we implicitly assume that within a census year t, for cohorts k and k', $cov(\widehat{P}_t^k, \widehat{P}_t^{k'}) \neq 0$ and that across census years t and t', $cov(\widehat{P}_t^k, \widehat{P}_{t'}^k) = 0$. Treating the census population as the only random variable in (3) we can then construct a confidence interval for the number of out-migrants in cohort k between t and $t+10$. Third, we assume the census takes place on July 1st each census year and so use mid-year inflows of immigrant numbers, I_t^{nga} and $I_{t+10}^{ng,a+10}$, for census years 1900, 1910 and 1920. We make the corresponding adjustment to official statistics to compare these series with our estimates.

[21]For this time period, these mortality data are derived for the Death Registration Area (DRA) consisting of ten states and the District of Columbia that was only successfully established in 1900 [Haines 2001]. These states are Connecticut, Indiana, Maine, Massachusetts, Michigan, New Hampshire, New Jersey, New York, Rhode Island and Vermont: in 1900 43% of the foreign-born population resided in these states. The rates are likely to be upward biased for any group that predominantly migrated to rural locations. The Census Bureau defines White individuals as those 'having origins in any of the original peoples of Europe, the Middle East, or North Africa.

[22]In historic data, foreign-borns had higher mortality rates than whites [Ward 1971, Haines 1977]. The foreign-born population was twice as likely as native population to reside in urban areas in 1900 [Klein 2004], and urban mortality rates were higher [Higgs 1973, Haines 1977]. There is no evidence that the foreign-born had greater mortality from the Spanish Flu outbreak in 1918 conditional on age and urban-rural residence [Taubenberger and Morens 2006]. It is estimated that 500,000 to 675,000 individuals died during the outbreak. By the 1930s, mortality differentials by nativity were converging and had largely disappeared [Haines and Steckel 2000].

to using these mortality rates is that they are not broken down by gender-age-inter-censal years. This would not matter for the fact that for a number of Eastern and Southern European countries, such as Russia and Spain, average mortality rates at the turn of the twentieth century lie *outside* the bounds given by white and other race mortality rates in the US. Hence we later also present estimates of out-migration based on these nationality specific mortality rates.[23]

4.4 Out-migration Rates in Aggregate

We first present estimate of out-migration rates for the US as a whole, aggregating out-migrants across all *nga* cohorts. As our initial aim is to compare the derived rates of out-migration and compare these to official statistics, we continue to take no account of survival probabilities ($D_t^{ga} = 0$), assume half-years within census year, and include both new and returnee arrivals. We therefore re-arrange (3) to estimate the following out-migration rate,

$$\frac{\sum_{k=1}^{k=10} E_{t+k}}{\sum_{k=1}^{k=10} I_{t+k}} = 1 + \frac{P_t - P_{t+10}}{\sum_{k=1}^{k=10} I_{t+k}}, \tag{5}$$

where the number of immigrant arrivals (I), population stock (P) and number of out-migrants (E) are all aggregated across all *nga* cohorts. The results are shown in Table 3A, again split by decade, and following a similar formatting to the earlier Tables.[24]

The first row in the upper panel of Table 3A reiterates that for the 1900-10 decade, our preferred estimate of immigrant inflows is 41% higher than is officially recorded. Combining with information on the foreign-born population stock, the implied number of out-migrants is $10,429,231$, that is more than three times the officially recorded number of out-migrants of $3,377,618$ as Column 4 shows. Using our most preferred estimate then implies an out-migration rate during 1900-10 of .761. Remarkably, this is more than double the official estimate for the decade based on Ferenczi-Willcox [1929], that is .348. This official rate is widely cited among economic historians and demographers although as emphasized throughout, it has long been recognized that the official statistics might severely mis-measure migratory flows to and from America.

Even taking an extreme position and using the lower bound estimate of immigrant arrivals that effectively assumes New York is the *only* entry point into America, the result in Column 5 shows an implied out-migration rate that is 2.35 times the official statistic. To summarize, although the previous results in Tables 1 and 2 have shown inflows to be recorded with error in official records, Table 3A emphasizes that the measurement error in *out-migration* statistics is orders of

---

[23]There are pros and cons of each choice of mortality rate. For example, there evidence in favor of Barker's fetal origin hypothesis of there being long lasting health effects of conditions experienced in utero [Almond 2006]. If so then country of birth mortality rates might be the most relevant. On the other hand, predominant causes of death in the US relate to contagious diseases, and extrapolating average mortality rates to age specific mortality rates might lead to sever biases at tails of the age distribution [Anderson and Ray 2010].

[24]To be clear, this exercise *cannot* be conducted using census data alone. Although in the census we observe year of arrival, we can only construct I_{t+k} for those migrants that choose to reside in the US on census date $t + 10$. The administrative records we exploit are crucial in being able to estimate I_{t+k} for all years and among migrants that no longer reside in the US on census date $t + 10$.

magnitudes larger than for the inflow measures. This is unsurprising given the historic lack of incentives for governments to accurately record outflows, a situation that persists today [Thomas 1973, Warren and Peck 1980, Keeling 2006].[25]

The lower panel of Table 3A repeats the analysis for the more turbulent 1910-20 decade. The first row reiterates that our preferred estimate of immigrant inflows is more than double what is officially recorded. Combining with information on the foreign-born population stock, we find the implied number of out-migrants to be 7.61 times larger than is officially recorded: official estimates suggest that around 2.4 million individuals out-migrated from America over this decade. Our accounting exercise reveals this number to be closer to 18 million. Again, even taking an extreme position and taking the lower bound estimate of immigrant arrivals and out-migrant flows, we find an implied out-migration rate that is 3.72 times the official statistic.

As with the earlier decade, mis-measured migrant inflows are considerably magnified by the errors in number of migrant outflows that lead to vastly different implied rates of out-migration than previously documented. More precisely, using our most preferred estimates, our method implies an out-migration rate of .975, almost three times the official estimate for the decade based on Ferenczi-Willcox [1929], that is .356. Moreover, this out-migration rate of almost one implies the magnitude of flows to and from America during the 1910-20 decade were of comparable size.[26]

Given the vast discrepancy between the out-migration rates we derive and the established wisdom based on official statistics, we present a simple way to validate our estimates. To do so, we compare estimates of the total *stock* of the foreign-born population in the US on each census date. These figures are aggregated from the IPUMS census samples and there is little reason to expect them to be widely mis-measured in aggregate across all nga cohorts.

A special case in which to calculate the out-migration rate from (5) is when the stock of foreign-born is not much changing, so $P_t - P_{t+10}$ is close to zero and the out-migration rate must be close to one. This is precisely the case for the 1910-20 decade. We then note that US census data suggests that during this decade, the increase in the *number* of foreign-born individuals was only marginal between 1910 and 1920, from 13.5 million to 13.9 million, an increase of only 3%. This corresponds closely to our implied out-migration rate of .975, as shown in the lower half of Table 3. In short, the simple fact that the *number* of foreign-borns in the US changed so little between 1910 and 1920 is highly indicative of there being an almost equal number of out-migrants as migrant inflows, so the out-migration rates ought to be close to one. This fact remains true *irrespective* of whichever estimate one wishes to take of migrant inflows over this decade – be it

[25]The US abandoned conducting a systematic review of migrants at the point of their exit from the US in 1957. The US is not alone in keeping poor data on out-migration. For example the UK has no official mechanism to record out-migration, relying instead on estimates based on limited surveys such as the *International Passenger Survey* or the *Quarterly National Household Survey*. It has also long been recognized that sending country records are of lower quality than receiving country data [Willcox 1979]. For example Ferenzi-Willcox [1929] show that prior to 1900 each sending country with the exception of England, reports far fewer migrants to the US than those recorded by US port officials.

[26]The official estimates of out-migration rates are based only on foreign-born immigrant arrivals into the US, excluding non-immigrant arrivals that are reported to have declared an intention not to settle in the US. Including both immigrant and non-immigrant arrivals obviously increases the denominator on the left hand side of (5), reducing the official out-migration rate further to .315 for the 1900-10 decade, and to .289 for the 1910-20 decade.

the $6,659,210$ recorded immigrant arrivals in Ferenczi-Willcox [1929] or the $18,511,266$ based on our preferred estimates and accounting exercise.

Our analysis is based on the assumption that the census accurately reports the number of foreign-born in aggregate (P_t, P_{t+10}). This might not be the case if the census were poorly targeted in rural areas where many migrants chose to reside. Hence an alternative thought experiment that sheds light on the plausibility of our estimates is to ask what would the error have to be in these census estimates if the official Ferenczi-Wilcox series were to be correct. From Table 3A we see that for 1900-10, we document there to be just over seven million more out-migrants than officially recorded. If the official record is correct, this implies census data must mis-measure the change in the foreign-born population by a similar magnitude. Given that in 1900 the US census reports there to 10.6 million foreign-borns in the US, this magnitude of measurement error in census data appears unlikely to drive our findings.

As emphasized throughout, it has long been argued that the official series in Ferenczi-Willcox [1929] are likely measured with error. In Table 3B we compare our estimates with one of the most detailed earlier attempts to correct these official statistics, conducted by Kuznets and Rubin [1954]. A few points are of note. First, Kuznets-Rubin [1954] generally estimate higher rates of out-migration, at .448 for the 1900-10 decade, rising to .536 for the 1910-20 decade. This difference with Ferenczi-Willcox [1929] is mostly driven by higher estimated numbers of out-migrants for each decade. However, the estimates provided by Kuznets-Rubin [1954] still remain far below those we derive using our detailed administrative records.[27]

To properly interpret the high rates of out-migration we document, it is important to stress that the accounting exercise we conduct does not imply that it is the *same* individuals that arrive and depart America within each decade. Rather, the evidence highlights that the magnitude of flows to and from America are far more similar in each decade than has been previously recognized. We find that for every four migrants that entered between 1900-10, three left over the period; over the 1910-20 decade there was an almost equal flow of individuals to and from America. In short, the notion that the American economy simply absorbed such large numbers of migrant inflows is incorrect. That migration was effectively a two-way flow between the US and the sending countries has major implications for understanding the potential selection of immigrants that chose to permanently reside in the US at the turn of the twentieth century, their impact on Americans in labor markets, and institutional change both in America and sending country economies.

4.5 Out-migration Rates By Cohort

Table 4 presents, for each decade, a disaggregated analysis of out-migration rates by cohort. The first row shows our previous most preferred estimate of out-migration defined in (5). Given that we are no longer focussed on comparisons with official statistics, the next few rows move to a more

---

[27]Another notable attempt to re-estimate out-migration rates is Keeling [2006] who uses shipping records (not passenger ship manifests) to estimate that between 1908-1914, 5.1 out of 10.4 million crossings out-migrated, again suggesting an out-migration rate close to .5.

meaningful out-migration rate defined from (3),

$$\frac{E_{t+10}^{ng,a+10} + \sum_{k=1}^{k=9} \prod_{j=k}^{j=9} (1 - D_{t+j}^{a+j})(E_{t+k}^{ng,a+k})}{I_{t+10}^{ng,a+10} + \sum_{k=1}^{k=9} \prod_{j=k}^{j=9} (1 - D_{t+j}^{a+j})(I_{t+k}^{ng,a+k})} = 1 + \frac{\prod_{j=0}^{j=9} (1 - D_{t+j}^{a+j})P_t^{nga} - P_{t+10}^{ng,a+10}}{I_{t+10}^{ng,a+10} + \sum_{k=1}^{k=9} \prod_{j=k}^{j=9} (1 - D_{t+j}^{a+j})(I_{t+k}^{ng,a+k})}. \quad (6)$$

The second row in Table 4 then bases the out-migration rate estimate only on *new* migrant arrivals. As described in the Appendix this is the more preferred economic measure of out-migration as it avoids double-counting migration spells. We see that the out-migration falls slightly in each decade with this adjustment.

The next batch of adjustments additionally account for survival probabilities of migrants between dates of arrival and census date. Hence the out-migration rate is the ratio of out-migrants to in-migrants in cohort nga in census year $t + 10$ that in-migrated since t, and out-migrated before census $t + 10$ and are *still alive somewhere* at $t + 10$. This reflects potentially different age distributions of migrant inflows and outflows. As previously discussed, various mortality rates can be used. Table 4 shows how the inferred out-migration rate varies as we adopt mortality rates of whites in the US, other races in the US, and nationality specific mortality rates. Aggregating across cohorts, we find that out-migration rates are highest assuming white mortality rates and lowest assuming other race mortality rates as expected.

The remaining rows break down out-migration rates by the three dimensions along which cohorts are defined, assuming white mortality rates. In the decade 1900-10 we find that men are far more likely to out migrate than women, with out-migration rates of .865 and .558 respectively. However in the later decade 1910-20 there are no differences by gender. This might be because of the changing composition of migrants by gender suggested by Figure 2 that shows that the ages of women migrants converged to those of men, and that a greater share of female migrants are married over time. Breaking out-migration rates into those of children (defined as being aged 14 or less at entry) and adults (those aged at least 15 at entry) we see that adults have lower out-migration rates than those that arrive as children.

The lower part of Table 4 then breaks down rates of out-migration by nationality. We do so for the ten countries that send the most migrants over the period, as shown in Figure 2. Taking these countries as a whole, out-migration rates in 1900-10 are .599, slightly lower than the rate among all countries (.632). This changes in the second decade when out-migration from these major sending countries (.978) is far higher than the average across all countries (.812). Examining out-migration rates by country reveals with this is so.

In the 1900-10 decade, there is little correlation between the number of migrant inflows from the sending country and rates of out-migration. Rather we generally find higher out-migration rates among Southern European countries of Italy, Greece and Spain, and the lowest out-migration rate is among Russians. A similar ranking across countries has previously been documented by Hatton and Williamson [2005] based on official statistics. These suggest out-migration rates to be highest (at around .5) for Spanish and Italian migrants, and lowest among Russians, Irish and Scandinavian migrants.

In the next decade, out-migration rates rise tremendously across all countries. Importantly, among those countries not allied to the US during first world war, Germany and Austria-Hungary, out-migration rates exceed one. In other words more individuals left the US during this period than entered (and were still alive in 1920). This presumably reflects anti-German discrimination in the US that has been documented during this period [Moser 2010].[28]

As a check on the underlying accounting methodology, the final row shows estimated out-migration rates among Canadians to the US. Reassuringly, for this group for whom the underlying decision to migrate to the neighboring country might be very different to migrants from Europe, we find out-migration rates that are considerably lower at around .4 in each decade. Hence the demographic accounting exercise that we conduct does not necessarily imply out-migration rates to be far higher than previously thought for this era for all countries.

To appreciate how the nature of international migration might have changed from the Age of Mass Migration we study, we compare our estimates with those based on contemporary data. For example, the official US Bureau of the Census estimates an out-migration rate for 1981-90 as 22%, although this figure does not distinguish between those with permanent or temporary visas. Using Census and INS data that establishes whether migrants are permanent or not, Warren and Peck [1980] estimate out-migration rates for 1960-70 to be 18%, and on average, higher for women and older age groups. Using similar data Borjas and Bratsburg [1996] estimate an out-migration rate of 17.5% within a 10 year period, much of it within the first five years. Finally, Jasso and Rosenzweig [1982] combine INS administrative records at entry for the 1971 cohort of legal permanent immigrants with their subsequent naturalization to estimate 10 year out-migration rates of between 30 and 50% by nationality.

This comparison reinforces the notion mentioned earlier, that historic rates of out-migration during the Age of Mass Migration we study are significantly higher than is observed today. As a consequence many of the choices and outcomes economists seek to understand related to migration, such as migrant investments into their human capital, the substitutability of migrants and natives in the labor market, and the gains to sending country economies, might have been historically very different than is the case from international migration today.

5   Conclusions

We use a novel data set of administrative records on 24 million migrants that entered Ellis Island, New York between 1892 and 1924 to estimate migrant flows into and out of America at the turn of the twentieth century. Combining immigrant *flow* estimates from these administrative records with census data on the *stock* of foreign-born in America in 1900, 1910 and 1920, we estimate out-migration rates in aggregate and by nationality-gender-age cohort. The accounting exercise

---

[28]At the outbreak of the first world war in 1914, the US remained neutral. In 1917, the US joined the Allied powers of the triple entente (the United Kingdom, France and Russia). The opposing Central Powers comprised of Germany, Austria-Hungary, the Ottoman Empire and Bulgaria from 1915 onwards. The other country that both had significant migratory flows to America and was involved in the war is Italy, who from 1915 onwards were among the Allied countries.

reveals the rate of out-migration from America between 1900-10 to be 76%, and the rate to be close to 100% during the turbulent decade of 1910-20. In comparison, official statistics from the time, that have long been suspected of being measured with error, have previously led scholars to report out-migration rates of around 35% in each decade.

In this paper, our primary aim has been to accurately measure migratory flows to and from the US at the turn of the twentieth century. In future research we plan to build on the findings presented to provide an analysis of *why* individuals out-migrate. Clearly, such behavior is hard to explain in simple income-maximizing models in the presence of large wage differentials [Sjastaad 1962, Harris and Todaro 1970, Gibson and McKenzie 2009]. Indeed there is a vast literature estimating substantial gains from migration [Hanson 2009, Abramitzky *et al.* 2010].[29]

A number of models however do seek to explain return migration using explanations based on: (i) return migration being planned as part of an optimal life cycle residential location sequence to first migrate, accumulate capital, and then return home [Borjas and Bratsberg 1996, Dustmann and Weiss 2007]; (ii) time varying complementarities between consumption and location [Hill 1987, Djajic and Milbourne 1988]; (iii) target income models [Yang 2006]; (iv) erroneous beliefs or negative shocks [Pessino 1991, Borjas and Bratsberg 1996]; (v) differential returns to multi-dimensional skills [Gould and Moav 2009, Dustmann *et al.* 2010]. A first strand of future research is therefore to exploit variation in the out-migration rates across nationality-age-gender and by year of arrival cohorts to help distinguish between these causes of out-migration.

The current analysis also has significant implications for understanding the behavior of migrants in America, their impact on Americans in labor markets, and the longer term impacts on sending country economies. This forms the basis of a second broad strand of future research.

On the link between out-migration and immigrant behavior in the US, in most theoretical frameworks the likelihood to out-migrate affects investment into own human capital, children's human capital, savings, and social networks [Galor and Stark 1990, Dustmann 1997, 1999, 2008, Cortes 2004]. In turn, many of these investments determine the substitutability of immigrant and native labor and this has implications for the causal impact immigrants have on natives in host country labor markets [Angrist and Kugler 2003, Ottaviano and Peri 2006, Borjas *et al.* 2008].[30]

On the longer term impacts on sending country economies, establishing rates of out-migration sheds light on the benefits that accrue to sending countries through remittances, brain gain [Mountford 1997, Dustmann and Weiss 2007, Docquier and Rapoport 2008, Mayr and Peri 2008], and the establishment of information networks between sending and host nations [Kerr 2008].

The evidence presented in this paper suggests the extent of migratory flows into and out of America at the turn of the twentieth century are far larger than previously recognized by scholars

---

[29]Abramitzky *et al.* [2010] show that among Norwegian migrants between 1865 and 1900, comparing within-brother pairs of stayer and leavers, the mean rate of return to migration is 120% (90%) for those originating from urban (rural) areas in Norway. This estimates accounts for self-selection by using birth order as an instrument for the propensity to migrate.

[30]This remains a hotly debated topic with mixed evidence on wage outcomes on natives of immigrants: negative effects are found in Borjas *et al.* [1996], Borjas [1999, 2003], and little impact is found in Card [1990, 2001, 2005], Dustmann *et al.* [2005]. Our analysis suggests that differing propensities to out-migrate across cohorts might be important for explaining such impacts and reconciling different research findings.

of many disciplines. Building on this insight through the agenda described above that seeks to understand the process of out-migration, and the consequent impact of migration on migrants, natives, and the economic and institutional development of both America and sending countries. We might therefore help reconcile the views that historic mass migration had large beneficial effects on origin countries [Hatton and Williamson 2005] but that this is no longer true today. Given the period we study is mostly characterized by a complete absence of administrative barriers to migration, studying these historic behaviors might be informative on the effects of easing migration barriers today.[31]

References

[1] ABRAMITZKY.R, L.P.BOUSTAN AND K.ERIKSSON (2010) Europe's Tired, Poor, Huddled Masses: Self-selection and Economic Outcomes in the Age of Mass Migration, NBER Working Paper 15648.

[2] ALMOND.D (2006) "Is the 1918 Pandemic Influenza Over? Long-term Effects of *in Utero* Influenza Exposure in the Post-1940 U.S. Population," *Journal of Political Economy* 114: 672-712.

[3] ANDERSON.S AND D.RAY (2010) "Missing Women: Age and Disease," *Review of Economic Studies* 77: 1262-300.

[4] ANGRIST.J (2002) "How do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation," *Quarterly Journal of Economics* 117: 997-1038.

[5] ANGRIST.J AND A.D.KUGLER (2003) "Protective or Counter-Productive? Labour Market Institutions and the Effect of Immigration on EU Natives," *Economic Journal* 113: 302-31.

[6] BARDE.R AND G.J.BOBONIS (2006) "Detention at Angel Island: First Empirical Evidence," *Social Science History* 30: 103-36.

[7] BHARADWAJ.P, A.I.KHWAJA AND A.R.MIAN (2008) "The Big March: Migratory Flows after the Partition of India," *Economic and Political Weekly* 43: 39-49.

[8] BORJAS.G.J AND B.BRATSBERG (1996) "Who Leaves? The Outmigration of the Foreign-Born," *Review of Economics and Statistics* 78: 165-76.

---

[31] As Goldin [1994] and Hatton and Williamson [2005] describe, pre-World War One, migration took place without restrictions such as visas or quotas. From 1917 onwards, significant legislative changes began to restrict entry into the US, through the imposition of quotas and tougher eligibility criteria, and post World War Two, nearly all international migration has taken place subject to some form of binding constraint. This has led to the view that a return to an era of free movement of labor, as with trade, would raise world welfare [Rodrik 2002, Rosenzweig 2005, Kremer and Watt 2006]. Indeed, there are suggestions that a 3% increase in labor migration would result in half the gains associated with complete trade liberalization [Winters *et al.* 2003]. The removal of all barriers to migration between OECD and non-OECD countries would boost world output between 92 and 172% [Klein and Ventura 2009].

[9] BORJAS.G.J (1999) "The Economic Analysis of Immigration," *Handbook of Labor Economics* 3: 1697-760.

[10] BORJAS.G.J (2003) "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics* 118: 1335-74.

[11] BORJAS.R, J.GROGGER AND G.H.HANSON (2008) Imperfect Substitution Between Immigrants and Natives: A Reappraisal, NBER Working Paper 13887.

[12] CARD.D (1990) "The Impact of Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review* 43: 245-57.

[13] CARD.D (2001) "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher immigration," *Journal of Labor Economics* 19: 22-64.

[14] CARD.D (2005) "Is the New Immigration Really so Bad?," *Economic Journal* 115: 300-23.

[15] CARTER.S.B AND R.SUTCH (1998) "Historical Background to Current Immigration Issues," in J.P. Smith and B. Edmonston (eds.), *The Immigration Debate: Studies in the Economic, Demographic and Fiscal Effects of Immigration*, Washington DC: National Academy Press.

[16] CARTER.S.B, S.GARTNER, M.R.HAINES, A.L.OLMSTEAD, R.SUTCH AND G.WRIGHT (2006) *Historical Statistics of the United States: Millennial Edition*, New York: Cambridge University Press.

[17] CORTES.K.E (2004) "Are Refugees Different from Economic Immigrants? Some Empirical Evidence on the Heterogeneity of Immigrant Groups in the Unites States," *Review of Economics and Statistics* 86: 465-80.

[18] COYNE.R.W AND P.J.MURPHY (2007) "The Golden Age: Service Management on Transatlantic Ocean Liners", *Journal of Management History* 13: 172-91.

[19] DELTAS.G, R.SICOTTE AND P.TOMCZAK (2008) "Passenger Shipping Cartels and Their Effect on Trans-Atlantic Migration," *Review of Economics and Statistics* 90: 119-33.

[20] DJAJI.S AND R.MILBOURNE (1988) "A General Equilibrium Model of Guest-Worker Migration: The Source-Country Perspective," *Journal of International Economics* 25: 335-51.

[21] DOCQUIER.F AND H.RAPOPORT (2008) "Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers," *Economic Journal* 118: 631-52.

[22] DUPONT.B, A.GHANDI AND T.J.WEISS (2009) Fluctuations in Overseas Travel by Americans, 1820-2000, NBER Working Paper 14847.

[23] DUSTMANN.C (1997) "Return Migration, Uncertainty and Precautionary Savings," *Journal of Development Economics* 52: 295-316.

[24] DUSTMANN.C (1999) "Temporary Migration, Human Capital, and Language Fluency of Migrants," *Scandinavian Journal of Economics* 101: 297-314.

[25] DUSTMANN.C (2008) "Return Migration, Investment in Children, and Intergenerational Mobility," *Journal of Human Resources* 43: 299-324.

[26] DUSTMANN.C, I.FADLON AND Y.WEISS (2010) Return Migration, Human Capital Accumulation and the Brain Drain, CREAM Discussion Paper 13/10.

[27] DUSTMANN.C, F.FABBRI AND I.PRESTON (2005) "The Impact of Immigration on the British Labour Market," *Economic Journal* 115: 324-41.

[28] DUSTMANN.C AND Y.WEISS (2007) "Return Migration: Theory and Empirical Evidence from the UK," *British Journal of Industrial Relations* 45: 236-56.

[29] ESTEVES.R AND D.KHOUDOUR-CASTERAS (2009) Remittances, Capital Flows and Financial Development during the Mass Migration Period, 1870-1913, CEPII Working Paper 2009-12.

[30] FERENCZI.I AND W.F.WILLCOX (1929) *International Migrations*, Cambridge, MA: NBER Publication.

[31] FREY.B.S, D.A.SAVAGE AND B.TORGLER (2011) "Behavior under Extreme Conditions: The Titanic Disaster", *Journal of Economic Perspectives* 25: 209-22.

[32] FRIEDBERG.R (2001) "The Impact of Migration on the Israeli Labor Market," *Quarterly Journal of Economics* 116: 1373-408.

[33] GABACCIA.D (1984) *From Sicily to Elizabeth Street: Housing and Social Change Among Italian Immigrants*, Albany: SUNY Press.

[34] GALOR.O AND O.STARK (1990) "Migrants' Savings, the Probability of Return Migration and Migrants' Performance," *International Economic Review* 31: 463-67.

[35] GANGULI.I (2010) Immigration at the Turn of the 20th Century: the Literacy Test and Ethnic-Industry Wages, mimeo, Harvard University.

[36] GIBSON.J AND D.MCKENZIE (2009) "The Microeconomic Determinants of Emigration and Return Migration of the Best and Brightest: Evidence from the Pacific," *Journal of Development Economics,* forthcoming.

[37] GOLDIN.C (1994) "The Political Economy of Immigration Restriction in the United States, 1890 to 1921," The Regulated Economy: *A Historical Approach to Political Economy*, C.Goldin and G.D.Libecap (eds.), University of Chicago Press.

[38] HAINES.M.R (1977) "Mortality in Nineteenth Century America: Estimates from New York and Pennsylvania Census Data, 1865 and 1900," *Demography* 14: 311-31.

[39] HAINES.M.R AND R.H.STECKEL (2000) *A Population History of North America*, New York: Cambridge University Press.

[40] HAINES.M.R (2001) "The Urban Mortality Transition in the United States, 1800-1940", *Annales de Demographie Historique* No. 1, pp.33-64 (Table A-2).

[41] HANSON.G (2009) "The Economic Consequences of the International Migration of Labor," *Annual Review of Economics* 1: 179-208.

[42] HARRIS.J.R AND M.P.TODARO (1970) "Migration, Unemployment and Development: a Two-Sector Analysis," *American Economic Review* 60: 126-42.

[43] HATTON.T.J AND J.G.WILLIAMSON (2005) *Global Migration and the World Economy: Two Centuries of Policy and Performance*, Boston: MIT Press.

[44] HEILAND.F (2004) "Trends in East-West Germany Migration From 1989 to 2002," *Demographic Research* 11: 173-94.

[45] HIGGS.R (1973) "Mortality in Rural America " *Explorations in Economic History* 10: 177-95.

[46] HILL.J.K (1987) "Immigrant Decisions Concerning Duration of Stay and Migratory Frequency," *Journal of Development Economics* 25: 221-34.

[47] HUTCHINSON.E.P (1958) "Notes on Immigration Statistics of the United States," *Journal of the American Statistical Association* 53: 963-1025.

[48] HYDE.F.E (1975) *Cunard and the North Atlantic 1840-1973: A History of Shipping and Financial Management*, London: MacMillan.

[49] JASSO.G AND M.R.ROSENZWEIG (1982) "Estimating the Emigration Rates of Legal Immigrants Using Administrative and Survey Data: The 1971 Cohort of Immigrants to the United States," *Demography* 19: 279-90.

[50] JEROME.H (1926) *Migration and Business Cycles*, Cambridge, MA: NBER Publication.

[51] KEELING.D (2006) "The Business of Transatlantic Migration Between Europe and the USA, 1900-1914," *Journal of Economic History* 66: 476-80.

[52] KEELING.D (2008) Shipping Companies and Transatlantic Migration Costs: The Case of Cunard, 1880-1914, mimeo, University of Zurich.

[53] KERR.W.R (2008) "Ethnic Scientific Communities and International Technology Diffusion," *Review of Economics and Statistics* 90: 518-37.

[54] KLEIN.H.S (2004) *A Population History of the United States*, New York: Cambridge University Press.

[55] KLEIN.P AND G.VENTURA (2009) "Productivity Differences and the Dynamic Effects of Labor Movements," *Journal of Monetary Economics* 56: 1059-73.

[56] KREMER.M AND S.WATT (2006) The Globalization of Household Production, mimeo, Harvard University.

[57] KUZNETS.S AND E.RUBIN (1954) *Immigration and the Foreign Born*, Cambridge, MA: NBER Publication.

[58] MAYR.K AND G.PERI (2008) Return Migration as a Channel of Brain Gain, NBER Working Paper 14039.

[59] MCCLELLAND.P.D AND R.J.ZECKHAUSER (1983) *Demographic Dimensions of the New Republic: American Interregional Migration, Vital Statistics, and Manumissions, 1800-1860*, New York: Cambridge University Press.

[60] MITCHELL.B.R (2007) *International Historical Statistics: Europe, 1750-2005*, London: Palgrave Macmillan.

[61] MOSER.P (2010) Taste-Based Discrimination-Empirical Evidence From a Shock to Preferences after WWI, NBER Working Paper 14003.

[62] MOUNTFORD.A (1997) "Can a Brain Drain be Good for Growth in the Source Economy?," *Journal of Development Economics* 53: 287-303.

[63] NUGENT.W (1992) *Crossings: The Great Transatlantic Migrations, 1870-1914*, Bloomington and Indianapolis: Indiana University Press.

[64] OTTAVIANO.G AND G.PERI (2006) Rethinking the Effects of Immigration on Wages, NBER Working Paper 12497.

[65] PESSINO.C (1991) "Sequential Migration Theory and Evidence From Peru," *Journal of Development Economics* 36: 55-87.

[66] PITKIN.T.M (1975) *Keepers of the Gate: A History of Ellis Island*, New York: New York University Press.

[67] PUERTA.J.M (2005) Hold on the Chain! Italian Immigrants and Their Networks in Buenos Aires 1882-1920, mimeo, Pompeu Fabra University.

[68] RAMIREZ.D (2001) *Crossing the 49th Parallel: Migration from Canada to the United States, 1900-1930*, Ithaca, NY: Cornell University Press.

[69] REINHART.C.M AND K.ROGOFF (2008) Is the 2007 US Sub-prime Financial Crisis So Different? An International Historical Comparison, NBER Working Paper 13761.

[70] RODRIK.D (2002) Feasible Globalizations, NBER Working Paper 9129.

[71] RUGGLES.S, M.SOBEK, A.TRENT, C.A.FITCH, R.GOEKEN,P.K.HALL, M.KING AND C.RONNANDER (2009) *Integrated Public Use Microdata Series: Version 4.0 [Machine-readable database]*, Minneapolis, MN: Minnesota Population Center.

[72] SCHAEFER.D (1994) "U.S. Migration, 1850-59," in D. Schaefer and T. Weiss (eds.), *American Economic Development in Historical Perspective*, Stanford, CA: Stanford University Press.

[73] SJAASTAD.L.A (1962) "The Costs and Returns of Human Migration," *Journal of Political Economy* 70: 80-93.

[74] TAUBENBERGER.J.K AND D.M.MORENS (2006) "1918 Influenza: The Mother of All Pandemics," *Emerging Infectious Diseases* 12: 15-22.

[75] THOMAS.B (1973) *Migration and Economic Growth; a Study of Great Britain and the Atlantic Economy*, New York: Cambridge University Press.

[76] TIZZANO.A (1965) *Mortalita' Generale - Sviluppo della Popolazione Italiana dal 1861 al 1961*, Roma: ISTAT.

[77] WARD.D (1971) *Cities and Immigrants: A Geography of Change in Nineteenth-Century America*, New York: Oxford University Press.

[78] WARREN.R AND J.M.PECK (1980) "Foreign-Born Emigration From the United States: 1960 to 1970," *Demography* 17: 71-84.

[79] WILLCOX.W.F (1929) *International Migrations, Volume I: Statistics*, Cambridge, MA: NBER Publication.

[80] WINTERS.L.A, T.L.WALMSLEY, Z.K.WANG AND R.GRYNBERG (2003) "Liberalising Temporary Movement of Natural Persons: An Agenda for the Development Round," *The World Economy* 26: 1137-61.

[81] YANG.D, (2006) "Why do Migrants Return to Poor Countries? Evidence from Philippine Migrants' Responses to Exchange Rate Shocks," *Review of Economics and Statistics* 88: 715-35.

A Appendix

A.1 Historic Context

Figure A1A shows the time series for transatlantic immigrant numbers in five year spans from the mid-1800s to the mid-1900s, based on the official Ferenczi-Willcox [1929] statistics. While our analysis ultimately shows these figures likely significantly underestimate inflows, they still remain the main source from which to describe very long run trends in migration to America. While for most of the nineteenth century annual inflows were around 300,000, they rose dramatically from 1880 onwards. The time period we study from 1892 to 1924 covers the peak years of migration into America, many of which witnessed over a million annual migrant arrivals. Towards the end of our study period migrant numbers dramatically reduced during World War One, recovered somewhat in the early 1920s and dwindled again during the depression era. Finally, the figure highlights the gradual change in national composition of European migrants. In the mid 1800s the vast majority originated from Britain and Germany, but by the end of our study period, the majority originated from Southern and Eastern Europe.[32]

Migration for much of the period we study was characterized by few *administrative* constraints. Pre-1917, migration took place without restrictions such as visas or quotas. Post-1917, significant legislative changes began to restrict entry into the US, through the imposition of quotas and tougher eligibility criteria, and post World War Two, nearly all international migration has taken place subject to some form of binding constraint [Goldin 1994].

Although the Age of Mass Migration was characterized by few administrative constraints, there remain two other constraints of potential first order: (i) non-monetary costs such as travel times; (ii) monetary travel costs relative to wages. On time costs, the move from sail ships to steamships for passenger travel began in the 1860s [Cameron 1989], reducing voyage times from Europe to the US from five weeks to two weeks [Hyde 1975, Cohn 1984]. Moreover, shipping firms began publishing departure schedules which further reduced passenger uncertainty, waiting times and the opportunity cost of foregone employment [Keeling 1999]. Complementary changes in transportation technology – such as the growth of railroad networks, canals and steamship travel within Europe and the US, also contributed to the overall surge in transatlantic migration flows from the mid to late 1800s [O'Rourke and Williamson 1999].[33]

---

[32]Hatton [2003] documents the relative importance of America as a destination for British migrants during the decades either side of 1900. As Figure A1B shows, in nearly all years the US remained the modal destination choice. By 1910 although more migrants left for Canada as their first destination, many did so with the intention of onward travel to the US, as we later account for. For migrants from other countries, the overwhelming majority had the US as their destination [Hatton and Williamson 2005].

[33]Hurd [1975] provides data on the growth of railroad mileages by country from 1850 to 1910. In the most important sending countries, the UK and Germany, railroad mileage increased 3 and 13 times respectively over this period. The US witnessed an explosion in railroad mileage from 9021 miles in 1850 to 249,902 miles in 1910. Apart from the precipitous decline in transportation costs, other important factors suggested by Hatton and Williamson [2005] driving the Age of Mass Migration include the use of government subsidies, the elimination of any remaining restrictions (such as Britain removing restrictions on migration in 1825 and 1827, Germany in the 1920s and Sweden in the 1840s), and the Irish famine – the Irish accounted for the majority of migrants to the US between 1846 and 1850 [Ó Gráda and O'Rourke 1997].

On ship mortality rates were relatively low. Even during the period of longer sail ship voyages these were estimated to be around 1% between 1820-60 [Cohn 1984]. Many ships stopped *en route* between their original port in Europe and their final port destination in America. In such cases, passenger lodgings were often arranged [Herson 2008]. In summary, time and opportunity costs should not be considered as significant barriers to migration or out-migration during the period we study.

The monetary costs of passage plummeted well before 1880. Between 1816 and 1860 the British-US route index fell by 80% [Hatton and Williamson 2005]. In absolute terms, by 1860 the average voyage cost is estimated to be $20 by Hatton and Williamson [1998] and Keeling [1999]. Ó Gráda and O'Rourke [1997] suggest it was just a few dollars even by the late 1840s. In relative terms, we note that the second half of the nineteenth century was a period of generally rising real wages throughout Europe, with wages converging to those of the US [Williamson 1995, O'Rourke and Williamson 1997]. Abramitzky *et al.* [2010] document the monetary cost of voyage represented 18% of annual earnings of a Norwegian farm labourer. There is also evidence that many immigrant tickets were pre-paid by previous immigrants, strengthening chain migration. For example, this is estimated to be the case for 30% of Finnish immigrants between 1891 and 1914, 50% of Swedes in the 1880s, and 40% of Norwegians in 1870s [Hvidt 1975, Kero 1991].[34]

A.2   New Arrivals and Returnees

The administrative records distinguish between individuals arriving to America for the first time, and returnees. To see which of these is most relevant for the accounting exercise we consider various scenarios in which we vary the years of arrival, departure and return relative to census dates t and $t+10$. These scenarios are represented in Figure A2. Each row represents a different scenario, and towards the right hand side of the figure, the resulting accounting exercise is conducted in two cases: when immigrant inflows are measured using all arrivals (new and returnee) and when only new arrivals (I_{t+10}^{new}) are counted. Of course these scenarios are not an exhaustive list, but other cases are derivatives of those considered. Throughout we show that whenever $I_{t+10} \neq I_{t+10}^{new}$, using I_{t+10} would lead us to overestimate the number of out-migrants.

In scenario one the individual arrives before census date t and departs before census date $t+10$. The individual is therefore measured as resident in the US on census date t ($P_t = 1$), the individual is not observed arriving to the US between census dates ($I_{t+10} = I_{t+10}^{new} = 0$), and is not recorded as being in the US on census date $t+10$ ($P_{t+10} = 0$). Hence as shown in Columns 1 and 2, accounting for out-migration using information on either new or returnee migrants leads to the

---

[34]In terms of entry costs into America, a series of legislative changes increased the nominal monetary costs of entry. In particular, the Immigration Act of 1894 doubled the head tax on entry to $1. The Immigration Act of 1907 increased this to $4 and gave wider terms of exclusion. The Immigration Act of 1917 raised the head tax to $8 and also excluded illiterates from entry [Reisler 1976, Scruggs 1988]. Subsequent legislation attempted to favor migrants from the more long established nationalities in the US. The first quantitative immigration law was the Quota Law of 1921 that limited the number of aliens of any nationality entering the US to 3% of the foreign-born persons of that nationality who lived in the US in 1910. The Immigration Act 1924 then established the 'national origins quota system', whereby an annual quota of immigrants by nationality was set to 2% of the number of foreign-born persons of such nationality resident in the US in 1890.

same conclusion that out-migration between census dates is one ($E_{t+10} = 1$). This is so because the individual is never recorded as an immigrant between census dates and so it is irrelevant whether information based on new or returnee arrivals is used. This is also true in the second scenario considered, in which the individual arrives and departs between census dates t and $t + 10$. Again, as the migrant is only observed entering the US once, $I_{t+10} = I_{t+10}^{new} = 1$ and $E_{t+10} = 1$ using either immigrant measure. The remaining scenarios are more interesting because in each $I_{t+10} \neq I_{t+10}^{new}$ and so E_{t+10} is sensitive to the choice of whether we use information on returnees.

In scenario three the individual arrives before census date t, departs and then returns to the US before $t + 10$. Hence the individual is observed on both census dates ($P_t = P_{t+10} = 1$). The individual is observed arriving into the US between census dates so $I_{t+10} = 1$, but because this arrivee is a returnee, $I_{t+10}^{new} = 0$. Hence the measure of out-migration that uses data on all arrivals $E_{t+10} = 1$, but if only information on new arrivals is used, $E_{t+10} = 0$. The measure based on *new* arrivals is preferred – the individual out-migrated only temporarily from the US between census dates. E_{t+10}^{new} reflects this but E_{t+10} does not.

Scenario four is the same as scenario three except the individual leaves America after having returned once, and this departure occurs before $t + 10$. As before $I_{t+10} = 1$, but because this arrivee is a returnee, $I_{t+10}^{new} = 0$. In contrast to the previous scenario the individual is not recorded as being in the US on census date $t + 10$, so $P_t = 1$ and $P_{t+10} = 0$. As a result, $E_{t+10} = 2$ which is misleading given the spirit of the accounting exercise. In contrast $E_{t+10}^{new} = 1$ as is more intuitive so that out-migrations are only counted once. However this scenario emphasizes that because out-migration rates necessarily have to be defined over some time period (t to $t + 10$) it can be the case that individuals that out-migrate just before census date $t + 10$ and plan to return after $t + 10$, will be counted as out-migrants for the accounting exercise for the period t to $t + 10$.

In the remaining two scenarios the individual arrives after census date t. In scenario five the individual departs and then returns to the US before $t+10$. As the individual enters twice between census dates, $I_{t+10} = 2$ and $I_{t+10}^{new} = 1$ because the individual is only recorded as a new arrival on first entry. As a result, $E_{t+10} = 1$ and $E_{t+10}^{new} = 0$. Given the temporary nature of the individual's out-migration, E_{t+10}^{new} is again the preferred measure.

Finally, scenario six is the same as scenario five except the individual leaves the US before census date $t + 10$ and so is recorded in neither US census, $P_t = P_{t+10} = 0$. It is still the case that $I_{t+10} = 2$ and $I_{t+10}^{new} = 1$ and so $E_{t+10} = 2$ and $E_{t+10}^{new} = 1$. As in all other cases, the out-migration measure based on new immigrant arrivals is then preferred.

A.3 Scaling-Up

The administrative records from Ellis Island are the basis from which we measure immigrant inflows into the Port of New York. As discussed above, when conducting the accounting exercise we restrict attention throughout to new immigrant arrivals. Aggregating this across all nationality-gender-age (nga) cohorts that arrived into New York in year t gives $I_{t+1}^{ng,a+1}$. However, in the accounting exercise described in (3) the figures for the population stock, migrant inflows, and

migrant out-flows refer to the *total* numbers for the US as a whole. Conceptually, it makes little sense to define *out-migration* at the level of a port of entry (unlike for migration inflows). Rather numbers and rates of out-migration need to be defined for the US in aggregate. Hence to conduct the accounting exercise, we first need to scale-up the immigrant inflows measured from Ellis Island records to those for the US. We proceed in five steps. In each, our approach is to base adjustments on conservative assumptions so that our preferred immigrant estimate is likely to underestimate true immigrant inflows, and hence underestimate numbers of out-migrants.

A.3.1 Missing Data

Our first adjustment corrects for missing nationality data in the administrative records. Between 1900-20 there are 14,917,859 individuals arriving into the US for the first time. Nationality is missing for 4.6% of records, age is missing for 1.6%, gender is missing for .7%. Arrival date is available for virtually all records. We replace missing values of nationality assuming nationality takes the modal value among individuals on the same ship and arrival date, place of residence, and surname. Doing so, the percentage of records with missing nationality information falls from 4.6% to 1.2%. To fill in the remaining missing values, we assume values for nationality, gender and age are missing at random, and assign missing values equally across all nga cohorts for each arrival year t. Hence we re-scale immigration numbers as follows,

$$\widetilde{I}_t^{nga} = \left[\frac{\#\text{total obs}_t}{\#\text{total obs}_t - \#\text{missing obs}_t} \right] \times I_t^{nga}, \tag{7}$$

where $\#\text{total obs}_t$ is the total number of new arrivals into Ellis Island in year t, and $\#\text{missing obs}_t$ is the number of new arrivals with missing information in nationality, gender or age.

A.3.2 Other Potential Mis-coding of Nationalities

As explained above the accounting exercise is conducted using information on *new* migrants to the US, not returnees. Although in theory all such new migrants should be recorded as non-US citizens, we find that 2.8% of then have their nationality recorded as US in the administrative records. There are a number of possible explanations. First, they might be US-born citizens that are actually returning to the US after having travelled abroad. Alternatively, they might be foreign-born citizens genuinely entering the US for the first time but with the intention of obtaining US citizenship. We assume all such individuals are foreign-born nationals and then assign them to a nationality-gender-age cohort. To conduct this assignment we combine the censuses from 1900, 1910 and 1920 and focus on the 284,000 foreign-born individuals that are recorded to be US naturalized citizens on census date. For each individual, the US census also records their year of entry into the US. We use this to calculate the share of naturalized US citizens in each cohort nga by year of arrival t. In the Ellis Island administrative records we then assign these shares to the 2.8% of new immigrant arrivals recorded to have US nationality.

A.3.3 Deaths and Expulsions

The third adjustment corrects for death in voyage and expulsion at the port of entry. As ship manifests recorded passengers close to the point of departure, those that died *en route* to the US need to be removed and should not be included in immigrant inflows. Reassuringly, on ship mortality rates were relatively low. Even during the earlier period of sail ship voyages that were two to three times longer than steam ship voyages, on ship mortality rates were estimated to be around 1% between 1820-60 [Cohn 1984]. We note that during its entire years of operation from 1892 to 1954, there were only 3500 deaths at Ellis Island, 350 births, and three suicides.[35]

On expulsion, nine out of every 100,000 immigrants were detained for mental examination and further questioning, and 2% were denied access to the US because of them suffering from a chronic contagious disease, having a criminal background, or being declared insane. From 1903 passenger ship companies were fined $100 for every excluded passenger, discouraging them taking on board ill, disabled or impoverished passengers.[36] In short, death and expulsion is likely to lead to there being a small difference between those recorded on ship manifests as travelling to the US and the actual number of immigrant arrivals. We take a conservative approach and assume 2% of all immigrants recorded on ship manifests either died or were expelled, and we assume US-citizens are not expelled.

A.3.4 Other Sea Ports

Prior to 1892 arrivals were recorded only for sea ports of entry of the continental US and Alaska. In 1894 immigrants to the US who arrived via Canadian sea ports began to be included. On land ports, counting arrivals at the land borders was not required by the early immigration acts. Complete reporting was attempted in 1855 with partial success, was interrupted by the Civil War, and was discontinued in 1885. Beginning in 1894, European immigrants who arrived at Canadian ports declaring an intention to proceed to the US were included in immigration statistics. In 1904 land border entry posts were finally established on the Mexican and Canadian borders. More stations were opened over time, but reporting of land border arrivals was not fully established until 1908 [Ramirez 2001]. However, not all migrants entering via the Canadian and Mexican borders were counted for inclusion in the immigration statistics. Before 1930, no count was made of residents of Canada, Newfoundland, or Mexico who had been living there for a year or longer and who self-reported planning to remain in the US for less than six months. Hence if individuals out-migrated due to, for example, negative shocks [Pessino 1991, Borjas and Bratsberg 1996], they would not be captured in official statistics.[37]

---

[35] As Cohn [1984] notes, given the length of the trip and taking into account the ages of the immigrants, this mortality rate is approximately four times higher than that experienced by non-migrants. Mortality was especially high among children and the elderly. There appears to have been little trend over time in mortality or differences in the loss rate by nationality.

[36] Office of Immigration statistics on arrivals mention 2419 aliens being debarred in 1893, 2799 being debarred in 1896, and 1880 being debarred in 1897.

[37] A more minor concern is that migrant arrivals from US territories were not always accurately recorded either. Arrivals in Alaska were first reported in 1871, but only irregularly thereafter until 1904, after which Alaska was

To account for other sea ports of entry we note first that for the period we study, the other major ports of entry are Baltimore, Boston, Philadelphia, and San Francisco. Official statistics record inflows by year into each port. Between 1900 and 1920, according to these official statistics Ellis Island accounted for 75% of all inflows into the US in the median year. Between 1900 and 1910, the median share is 76% and this falls to 55% for the 1910-20 decade. This fall is entirely due to the years during and just after the first world war – Ellis Island drops below 50% of all immigrant arrivals from 1916 to 1919.

We assume that the *ratio* of immigrants into Ellis Island relative to all sea ports as recorded in official statistics is correct, even though the number of immigrant inflows at each port is likely measured with error. This is equivalent to assuming the ratio of true arrivals into Ellis Island and the US as a whole is the same as the ratio of the errors with which each is measured in official statistics. The main sources of error in official statistics for sea ports are: (i) careless collection of ship manifests; (ii) the exclusion of first and cabin-class passengers; (iii) the exclusion of US citizens, or returning foreign-born nationals who had earlier acquired US citizenship. Hence our underlying assumption is justified as long as these combined sources of error are proportionate to the true number of arrivals into each port. Taking this to be the case we therefore re-scale immigration inflows for each year of arrival as follows,

$$\widetilde{\widetilde{I}}_t^{nga} = \left[\frac{\#\text{arrivals into USA}_t}{\#\text{arrivals into NY}_t} \right] \times \widetilde{I}_t^{nga}, \tag{8}$$

where the ratio $\frac{\#\text{arrivals into USA}_t}{\#\text{arrivals into NY}_t}$ for each year of arrival t is taken from official statistics.

It might however be reasonable to assume some nationalities specialized in arriving into some sea ports especially given the importance of chain migration [Stark 1988, Stolarik 1988]. Hence we also derive an alternative adjustment that allows this weight to be nationality-year specific.

We have combined details from a variety of sources, including Stolarik [1988], Filby and Meyer [1981], Tepper [1993] and Cohn [2009] to understand which nationalities might have gone to other ports during the time period we study.[38] Based on this literature, we assume the following nationalities are likely to have relatively more entrants via other sea ports of entry than the average nationality as assumed in (8): Britain, Ireland, Germany, Austria-Hungary, Italy, Greece, Portugal, Poland, Russia, Armenia, Ukraine, Belgium, as well as migrants from the following

regularly included among the places of entry. Arrivals in Hawaii were first included in 1901, Puerto Rico in 1902, Guam in 1929, Samoa in 1932 and the Virgin Islands in 1942. Arrivals in and departures from the Philippines were recorded in the port tables for 1910-24, but were not included in the total immigration data. For 1925-31, such arrivals and departures were obtained annually from the Bureau of Insular Affairs, War Department, and published in separate tables. Since 1932, the Immigration Service has kept no records of arrivals in the Philippines or departures from the Philippines to foreign countries.

[38]Boston was the terminus for Britain's Cunard steamship line and rates were subsidized by the British government. This port was predominantly used by Irish, Italian, Greek, Portuguese, Polish and Russian Jews, and Armenians. A key advantage of Baltimore was its railroad links to the American West, strengthened by the 1867 agreement between the Baltimore and Ohio Railroad and the North German Lloyd Steamship Line. Immigrant groups were predominantly German, Irish and English. Philadelphia was the port for The American Line and Red Star Line shipping companies. They had weekly sailings from Liverpool, Antwerp and Hamburg. These brought large numbers of migrants from Poland, Russia and Austria-Hungary. San Francisco became a major port for immigration from southern and eastern Europe, and it remained the major entry point for the Chinese.

Pacific rim countries that would of course have predominantly entered through San Francisco: China, Australia, Japan, Korea and the Philippines. For these countries, the scaling-up factor is set to be the 90th percentile value of $\left[\frac{\#\text{arrivals into USA}_t}{\#\text{arrivals into NY}_t}\right]$ for each decade (1900-1910, 1910-1920). For all other countries, the adjustment in (8) is still used.

A.3.5 Land Borders

Our fifth adjustment aims to conservatively account for inflows into the US from its land borders. To deal first with inflows from Canada, we note that from 1894 onwards, US-bound overseas immigrants disembarking in Canadian ports had to undergo inspection by US officers stationed there before being allowed to proceed to a US destination. According to Smith [2000], the steamship lines agreed to treat all passengers destined to the United States as if they would be landing at a US port of entry. This meant completing a US ship passenger manifest form and selling tickets only to those who appeared admissible under US law. Canadian railroads agreed to carry only those immigrants who were legally admitted to the United States to US destinations.

By the time the first immigration stations opened on the US-Canada border, around 40% of migrants arriving in Canada were thought to be heading for the US, and might have accounted for up to 22% of immigrants into US [Hatton and Williamson 2005]. To measure these inflows, we use official statistics on immigrant inflows into Canada by nationality and year from 1900 to 1920, as reported by the Division of Immigration, Department of Manpower and Immigration of Canada [Anderson and Frideres 1981]. We assume the share of individuals in any gender-age cohort within a nationality is the same as for inflows into Ellis Island for the same nationality and year of arrival, and throughout we assume 40% of arrivals into Canada are *en route* to the US.[39]

On inflows from Mexico, we focus on measuring inflows of Mexicans. In contrast to the land border with Canada, there is little evidence that non-Mexicans, especially Europeans, entered the US via Mexico in large numbers. Focusing then on Mexican migrants, we note that large-scale migration began in 1900 when US financed railroads penetrated the Mexican interior [Cardoso 1980, Hart 1987]. A generally accepted figure is that around 50,000 Mexican immigrants from Mexico were arriving into the US annually by 1908-10, but there are reasons to suspect this is an underestimate [Briggs 1984, Gonzales and Fernandez 2003]. US industrialists intensified recruitment of Mexicans when World War One broke out [Driscoll 1999]. To get a sense of the scale of Mexicans officially recorded in the US, there are 68,000 Mexicans recorded in the 1880 census, and this rises to 500,000 in 1920 census. However, Mexico never features as a top ten source country in terms of population stocks in the US. Hence given the paucity of evidence on Mexican inflows, we choose to follow the most conservative route and assume these are zero rather than potentially over estimate their number at least in some years.

---

[39]The statistics reported in Anderson and Frideres [1981] have been extracted from the Division of Immigration, Department of Manpower and Immigration of Canada. The data includes information on 46 immigrant nationalities arriving to Canada over 1900-1920. It is based on the administrative records collected by the Canadian port officials and transferred to the government that had the responsibility to fill immigration reports. The data includes arrivals to all Canadian ports.

On illegals, at the start of our sample period given the lack of legislation related to immigration it is unclear how to even define an illegal migrant. Of course this changes over time and there is a view that flows of illegals increased in response to tougher immigration controls embodied in the 1921 Quota Act [Briggs 1984, Gemery 1994, Hatton and Williamson 2005]. Indeed, Briggs [1984] estimates hundreds of thousands, and perhaps millions, of illegal immigrants entered the US in the 1920s via Canada and Mexico after the 1921 Quota Act was passed. As a result, the US border patrol was established in 1924, the same year as the Immigration Act passed to tighten borders and significantly improve migration statistics [Massey *et al.* 2002]. Hence given the paucity of evidence on illegal inflows, we again follow the most conservative route and assume these are zero.

A.3.6 Remaining Error in Out-migrant Numbers

Having made the five adjustments above we obtain estimates for immigration inflows into the US as whole by nationality-gender-age-year of entry cohorts. We then combine this information with census data and set mortality rates to zero to calculate (3) for each cohort and decade. We then aggregate across all cohorts of the same nationality by decade to examine nationalities for which the accounting exercise produces negative estimates of out-migration. Such estimates are obviously incorrect and can help provide insight into likely further adjustments that are required.

The results are shown in Table A1 and Figure A2. Columns A in Table A1 shows that of the 118 nationalities represented in the Ellis Island administrative records, 66 (56%) have at least one individual recorded to be resident in the US in the 1900 census, with the remaining 52 nationalities recorded to have a zero population in 1900 ($P_{1900}^n = 0$). Calculating (3) for the 1900-10 decade, Column B then shows the number of nationalities that are implied to have positive, zero, or negative numbers of out-migrants. The rows in Table A1 correspond to estimates of out-migrant numbers based on alternative scaling-up adjustments in immigrant numbers.

For example, the first row shows that if we take the raw data from Ellis Island and make none of the adjustments described above, then for the 1900-10 decade: (i) 85 out of 118 nationalities are found to have a strictly positive number of out-migrants; (ii) 17 nationalities are found to have zero out-migrants; (iii) 16 nationalities are inferred to have negative numbers of out-migrants. The remaining rows in Column B show that these numbers are relatively stable as we make sequentially more of the adjustments to scale-up immigrant numbers described above. Even with all our preferred adjustments, the bottom row shows there remain 14 nationalities for which a negative number of out-migrants is implied.[40]

Figure A2a then shows for these 14 nations for whom total out-migration over 1900-10, $\sum_{k=1}^{k=10} E_{1900+k}^n < 0$, the actual stock of foreign-borns of each nationality observed in the US census in 1880, 1900, 1910 and 1920. Two points are of note.

First, the country for which negative out-migration is estimated but that has the largest population resident in the US is Mexico. This is not surprising given the difficulties described

---

[40]The number of nationalities with implied negative numbers of out-migrants declines if we exploit information on $var(\widehat{P}_t^{nga})$ as described in Section 4.2 and check the number of nationalities for which the number of out-migrants is significantly below zero.

above in obtaining reliable information on inflows across US-Mexico land borders. Indeed we have followed a conservative approach and set these inflows to zero.[41]

Second, many of the countries listed are in the Pacific rim. Hence we expect the vast majority of such migrants to enter the US through Angel Island, San Francisco or other West Coast ports. This suggests the adjustment embodied in (8) is likely too conservative for such nationalities. For the remaining countries, the foreign born population in the US is very small in all census dates. Hence any small sampling variation could lead us to find a negative number of out-migrants.

We now repeat the analysis for the more turbulent decade of 1910-20. Returning to Table A1 we see that at the start of this decade on census date in 1910, individuals from 76 of the 118 nationalities represented in the Ellis Island demonstrative records are identified to be resident in the US. Using the raw unadjusted data from Ellis Island we find that for 1910-20 decade: (i) 57 out of 118 nationalities are found to have a strictly positive number of out-migrants; (ii) 18 nationalities are found to have zero out-migrants; (iii) 43 nationalities are inferred to have negative numbers of out-migrants. The remaining rows in Column B show that as we make each adjustment described previously, our preferred estimate then has 31 nationalities with implied negative numbers of out-migrants.

Figure A2b shows for these 31 nations the actual stock of foreign-borns of each nationality observed in the US census in 1880, 1900, 1910 and 1920.[42] Two points are of note. First, Mexico again ranks as a problematic case. The other problematic cases of Poland and Bohemia relate to countries that experienced border conflicts post-1917, and Galicia that experienced jurisdictional changes over the study period. These changes might have led to discrepancies between reported nationalities in administrative records at time of entry, and reported nationalities on later census dates. Second, the actual population stocks resident in the US from most of these countries are very small, and this is true across the four census dates.

Hence our final adjustment is to set negative out-migrant numbers to zero for any nga cohort for which $\sum_{k=1}^{k=10} E_{t+k}^{ng,a+k} < 0$, as is done in Borjas and Bratsberg [1996].

References

[1] ANDERSON.A.B AND J.S.FRIDERES (1981) *Ethnicity in Canada: Theoretical Perspectives*, Toronto: Butterworths.

[2] BRIGGS.V.M (1984) *Immigration and the American Labor Force*, Baltimore: John Hopkins University Press.

---

[41]Indeed, the problem might be even more severe as the 1910 and 1920 censuses are known to have been conducted at times of the year when the Mexican migratory population was at its lowest point during the year [Cardosa 1980].

[42]We can use the 1880 census, in which 100% and 1% samples are available to given an indication of the potential mis-measurement of population stocks. From the 1880 census, we find the median ratio of $\widehat{P}_t$ based on a 100% sample to that based on a 1% sample to be 1.15. Hence as the 1920 census is the smallest sample we use (1%) we might well under-count $P_{t+10}^{ng,a+10}$ in 1920, making it more likely that we find a negative number of out-migrants for 1910-20. This might in part explain the larger number of problematic countries for this decade.

[3] CAMERON.R (1989) *A Concise Economic History of the World from Paleolothic to the Present*, Oxford: Oxford University Press.

[4] CARDOSO.L (1980) *Mexican Emigration to the United States, 1897-1931*, Tucson: University of Arizona Press.

[5] COHN.R.L (1984) "Mortality on Immigrant Voyages to New York, 1836-1853," *Journal of Economic History* 44: 289-300.

[6] COHN.R.L (2009) *Mass Migration Under Sail: European Immigration to the Antebellum United States*, New York: Cambridge University Press.

[7] DRISCOLL.A.K (1999) "Risk of High School Dropout Among Immigrant and Native Hispanic Youth," *International Migration Review* 33: 857-75.

[8] FILBY.W.P AND M.K.MEYER (1981) *Passengers and Immigration Lists Index*, Detroit, MI: Book Tower.

[9] GEMERY.H.A (1994) "Immigrants and Emigrants: International Migration and the U.S. Labor Market in the Great Depression," in T.J. Hatton and J.G. Williamson (eds.), *Migration and the International Labor Market, 1850-1939*, London: Routledge.

[10] GOLDIN.C (1994) "The Political Economy of Immigration Restriction in the United States, 1890 to 1921," The Regulated Economy: *A Historical Approach to Political Economy*, C.Goldin and G.D.Libecap (eds.), University of Chicago Press.

[11] GONZALES.G.G AND R.A.FERNANDEZ (1980) *Mexican Emigration to the United States 1897-1931*, Tucson: University of Arizona Press.

[12] HART.J (1987) *Revolutionary Mexico: The Coming and Process of the Mexican Revolution*, Berkeley and Los Angeles: University of California Press.

[13] HATTON.T.J (2003) Emigration from the UK, 1870-1913 and 1950-1998, IZA Discussion Papers 830.

[14] HATTON.T.J AND J.G.WILLIAMSON (1998) *The Age of Mass Migration: Causes and Economic Impact*, New York: Oxford University Press.

[15] HERSON.J (2008) Liverpool as a Diasporic City, Conference Paper, Economic History Society Annual Conference.

[16] HVIDT.K (1975) *Flight to America: the Social Background of 300,000 Danish Emigrants*, New York: Academic Press.

[17] KEELING.D (1999) "The Transportation Revolution and Transatlantic Migration, 1850-1914," *Research in Economic History* 25: 39-74.

[18] KERO.R (1991) "Migration Traditions From Finland to North America," *in R.J. Vecoli and S. Sinke (eds.), A Century of European Migrations, 1830-1930*, Urbana, IL: University of Illinois Press.

[19] MASSEY.D, J.DURAND, AND N.J.MALONE (2002) *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Economic Integration*, New York: Russell Sage Foundation.

[20] O'GRADA.C AND K.O'ROURKE (1997) "Mass Migration as Disaster Relief: Lessons from the Great Irish Famine," *European Review of Economic History* 1: 3-25.

[21] O'ROURKE.K AND J.G.WILLIAMSON (1999) *Globalization and History, the Evolution of a 19th Century Atlantic Economy*, Cambridge, MA: MIT Press.

[22] REISLER.M (1976) *By the Sweat of their Brows: Mexican Immigrant Labor in the United States, 1900-1940*, Westport, CN: Greenwood Press.

[23] SCRUGGS.O.M (1988) *Wetbacks and the Farm Labor Problem: Mexican Agricultural Labor in the United States 1942-1954*, New York: Garland Publishing, Inc.

[24] SMITH.M.L (2000) "By Way of Canada: U.S. Records of Immigration Across the U.S.-Canadian Border, 1895-1954 (St. Albans Lists)", Prologue, Montreal: National Archives, Vol. 32, No. 3.

[25] STARK.O (1988) "On Marriage and Migration," *European Journal of Population* 4: 23-37.

[26] STOLARIK.M.M (1988) *Forgotten Doors: The Other Ports of Entry to the United States*, Philadelphia: Balch Institute Press.

[27] TEPPER.M (1993) *American Passenger Arrival Records: A Guide to the Records of Immigrants Arriving at American Ports by Sail and Steam*, Baltimore, MD: Genealogical Publishing Company.

[28] WILLIAMSON.J.G (1995) "The Evolution of Global Labor Markets since 1830: Background Evidence and Hypotheses," *Explorations in Economic History* 32: 141-96.

Table 1: Official Statistics and Administrative Record Measures of Migrant Inflows into New York, by Decade

	(1) Official Statistic Ferenzi-Willcox [1929]	(2) Preferred Estimate	(3) Raw Data	(4) Ratio: Preferred Estimate to Official Statistic (Col 2/Col 1)	(5) Ratio: Raw Data Estimate to Official Statistic (Col 3/Col 1)
1900-1910					
Total immigrant foreign-born arrivals	7431670	8968628	8792771	1.21	1.18
Total non-immigrant foreign-born arrivals	713749	-	-	-	-
Total foreign-born arrivals	8145419	8968628	8792771	1.10	1.08
Total US citizen arrivals	1546237	1200336	1150045	0.78	0.74
Total arrivals	9691656	10168964	9942816	1.05	1.03
1910-1920					
Total immigrant foreign-born arrivals	4416448	7054163	6624076	1.60	1.50
Total non-immigrant foreign-born arrivals	856931	-	-	-	-
Total foreign-born arrivals	5273379	7054163	6624076	1.34	1.26
Total US citizen arrivals	1214658	1249759	1147865	1.03	0.95
Total arrivals	6488037	8303922	7771941	1.28	1.20

Table 2: Official Statistics and Administrative Record Measures of Migrant Inflows for the US, by Decade

	(1) Official Statistic Ferenzi-Willcox [1929]	(2) Preferred Estimate	(3) Raw Data	(4) Ratio: Preferred Estimate to Official Statistic (Col 2/Col 1)	(5) Ratio: Raw Data Estimate to Official Statistic (Col 3/Col 1)
1900-1910					
Total immigrant foreign-born arrivals	9719358	13712006	8792771	1.41	0.90
Total non-immigrant foreign-born arrivals	994168				
Total foreign-born arrivals	10713526	13712006	8792771	1.28	0.82
Total US citizen arrivals	2040674	1570248	1150045	0.77	0.56
Total arrivals	12754200	15282254	9942816	1.20	0.78
1910-1920					
Total immigrant foreign-born arrivals	6659210	18511266	6624076	2.78	0.99
Total non-immigrant foreign-born arrivals	1540972				
Total foreign-born arrivals	8200182	18511266	6624076	2.26	0.81
Total US citizen arrivals	2111460	2426712	1147865	1.15	0.54
Total arrivals	10311642	20937978	7771941	2.03	0.75

Notes: The official statistics in Column 1 of Tables 1 and 2 are from Ferenzi-Willcox [1929]. For all other statistics derived from Ellis Island Administrative records, these are based on the total number of immigrant arrivals (new and returnee). For statistics related to arrivals into New York City in Table 1, the preferred estimate figure in Column 2 is based on corrections for missing data, other potential errors in recorded nationalities, and exclusions. The estimate in Column 3 is based on the raw administrative statistics from which no adjustments are made. For statistics related to arrivals and departures into the US in Table 2, the preferred estimate in Column 2 is based on corrections for missing data and other potentially mis-coded nationalities, expulsion or death, inflows from other sea ports, and inflows over land via Canada and Mexico. The estimate in Column 3 is based on the raw administrative statistics from which no adjustments are made. We assume the census takes place on July 1st each census year and so use mid-year inflows of immigrant numbers, for census years 1900, 1910 and 1920. We make the corresponding adjustment to official statistics to compare these series with our estimates.

Table 3A: Official Statistics and Administrative Record Measures of Migrant Flows for the US, by Decade

	(1) Official Statistic Ferenzi-Willcox [1929]	(2) Preferred Estimate	(3) Raw Data	(4) Ratio: Preferred Estimate to Official Statistic (Col 2/Col 1)	(5) Ratio: Raw Data Estimate to Official Statistic (Col 3/Col 1)
1900-1910					
Total immigrant foreign-born arrivals	9719358	13712006	8792771	1.41	0.90
Total migrant departures from US	3377618	10429231	7191956	3.09	2.13
Implied out-migration rate for US	0.348	0.761	0.818	2.19	2.35
1910-1920					
Total immigrant foreign-born arrivals	6659210	18511266	6624076	2.78	0.99
Total migrant departures from US	2372071	18048715	8828942	7.61	3.72
Implied out-migration rate for US	0.356	0.975	1.333	2.74	3.74

Table 3B: Kuznets-Rubin [1954] and Administrative Record Measures of Migrant Flows for the US, by Decade

	(1) Kuznets-Rubin [1954]	(2) Preferred Estimate	(3) Raw Data	(4) Ratio: Preferred Estimate to Official Statistic (Col 2/Col 1)	(5) Ratio: Raw Data Estimate to Official Statistic (Col 3/Col 1)
1900-1910					
Total immigrant foreign-born arrivals	9447500	13712006	8792771	1.45	0.93
Total migrant departures from US	4230000	10429231	7191956	2.47	1.70
Implied out-migration rate for US	0.448	0.761	0.818	1.70	1.83
1910-1920					
Total immigrant foreign-born arrivals	7400000	18511266	6624076	2.50	0.90
Total migrant departures from US	3963000	18048715	8828942	4.55	2.23
Implied out-migration rate for US	0.536	0.975	1.333	1.82	2.49

Notes: The official statistics in Column 1 of Table 3A are from Ferenzi-Willcox [1929]. In Table 3B the comparison is made to the Kuznets-Rubin [1954] correction of these official statistics. For all other statistics derived from Ellis Island Administrative records, these are based on the total number of immigrant arrivals (new and returnee). The preferred estimate in Column 2 is based on corrections for missing data and other potentially mis-coded nationalities, expulsion or death, inflows from other sea ports, and inflows over land via Canada and Mexico. The lower bound estimate in Column 3 is based on the raw administrative statistics from which no adjustments are made. The implied out-migration rate is the total number of migrant departures divided by the total number of immigrant arrivals into the US in the same decade. We assume the census takes place on July 1st each census year and so use mid-year inflows of immigrant numbers, for census years 1900, 1910 and 1920. We make the corresponding adjustment to official statistics to compare these series with our estimates.

Table 4: Out-migration Rate Estimates by Cohort and Decade

Cohort	1900-1910			1910-1920		
	(1) Total immigrant arrivals	(2) Total immigrant departures	(3) Implied out-migration rate	(4) Total immigrant arrivals	(5) Total immigrant departures	(6) Implied out-migration rate
Aggregate based on total immigrant arrivals	13712006	10429231	0.761	18511266	18048715	0.975
Aggregate based on first time immigrant arrivals	13317559	10034791	0.754	13863483	13400968	0.967
White mortality rate	14154747	8951276	0.632	14582711	11836158	0.812
Other mortality rate	14107050	8239115	0.584	14621274	10900747	0.746
Nationality specific mortality rate	14211636	8750683	0.616	14631129	11595469	0.793
Men	8472316	7332065	0.865	10857499	10419854	0.960
Women	4846656	2704150	0.558	3018955	2994091	0.992
Aged 0-14 at time of arrival	1561221	1305334	0.836	1421342	1615365	1.137
Aged 15+ at time of arrival	12651111	7726132	0.611	13211091	10302094	0.780

Nationality	1900-1910			1910-1920		
	(1) Total immigrant arrivals	(2) Total immigrant departures	(3) Implied out-migration rate	(4) Total immigrant arrivals	(5) Total immigrant departures	(6) Implied out-migration rate
Top Ten Nationalities Based on Immigrant Arrivals into NYC Between 1892-1924	11996584	7181314	0.599	11002964	10765310	0.978
Rank 1: Italy	3372036	2438093	0.723	2721625	2281362	0.838
Rank 2: Austria-Hungary	2869037	1584087	0.552	878582	1338729	1.524
Rank 3: Russia	2024757	825060	0.407	1116179	1085628	0.973
Rank 4: Great Britain	964993	645387	0.669	2914890	2772513	0.951
Rank 5: Germany	1164191	608361	0.523	646595	787111	1.217
Rank 6: Ireland	644574	477324	0.741	556334	522350	0.939
Rank 7: Sweden	397799	236088	0.593	442348	398894	0.902
Rank 8: Greece	352056	263177	0.748	462087	393909	0.852
Rank 9: Norway	277015	159524	0.576	479516	453094	0.945
Rank 10: Spain	144674	121729	0.841	628530	562680	0.895
Other: Canada	307064	120745	0.393	363390	157123	0.432

Notes: All statistics derived from Ellis Island Administrative records are based on the number of new immigrant arrivals, except in the first row that is based on the total number of immigrant arrivals (new and returnee). These preferred estimates are based on corrections for missing data and other potentially mis-coded nationalities, expulsion or death, inflows from other sea ports, and inflows over land via Canada and Mexico. In the first two rows, a survival rate of one is assumed. In the third and fourth rows, survival rates of whites and "other" race are used. In the fifth row, nationality specific mortality rates are used for Italy, Austria-Hungary, Great Britain, Russia, Germany, Ireland, Sweden, Spain, France, Norway, Denmark, Finland, Belgium, Romania and Switzerland. For all other nationalities, white mortality rates are assumed. For gender and age specific cohorts, white mortality rates are assumed. In the lower panel, for the country specific cohorts, country specific mortality rates are used. The ten countries chosen (plus Canada) are those from which the most immigrant arrivals originate from into Ellis Island over the period 1892-1924.

Table A1: Sign of Estimates of Emigrant Numbers, by Decade

| | Decade: 1900-10 | | | | | Decade: 1910-20 | | | | |
| | A. Number (%) of 1900 Census Values That Are: | | B. Number (%) of Emigration Estimates That Are: | | | C. Number (%) of 1910 Census Values That Are: | | C. Number (%) of Emigration Estimates That Are: | | |
	Zero	Positive	Negative	Zero	Positive	Zero	Positive	Negative	Zero	Positive
Nationalities (118)										
Census population	52 (44.1)	66 (55.9)				42 (35.6)	76 (64.4)			
Raw data from Ellis Island (lower bound)			16 (13.6)	17 (14.4)	85 (72.0)			43 (36.4)	18 (15.3)	57 (48.3)
Adjusting for missing values and exclusions			17 (14.4)	20 (17.0)	81 (68.6)			43 (36.4)	19 (16.1)	56 (47.5)
Adjusting for other ports of entry			16 (13.6)	21 (17.8)	81 (68.6)			37 (31.4)	19 (16.1)	62 (52.5)
Adjusting for other ports of entry and US citizens			15 (12.7)	21 (17.8)	82 (69.5)			32 (27.1)	19 (16.1)	67 (56.8)
Adjusting for other ports of entry including Canada and US citizens			14 (11.9)	21 (17.8)	83 (70.3)			31 (26.3)	19 (16.1)	68 (57.6)

Notes: The unit of observation is nationality. There are 118 nationalities in the sample, from which at least one migrant entered the US via Ellis Island from 1900-20. The adjustment for other ports of entry scales up estimates of immigrant and out-migrant numbers using the ratio of arrivals into the US to those into New York each year. The correction for inflows from Canada assumes 40% of immigrant arrivals into Canada arrive in the US and that their age distribution is the same as into Ellis Island in the same nationality-year of arrival cohort. The correction including US citizens corrects for some foreign born immigrants entering the US after having obtained US citizenship and therefore having US nationality.

Figure 1: Passenger Ship Manifest from March 3rd, 1903

Full Name

Age, Gender, Marital Status

Occupation/skills

Nationality

Race

Last Residence

Links to Others in US: Final Destination etc.

Political views

Health

Notes: The passenger ship manifest shown was accessed from http://www.ellisisland.org/search on April 24th 2010. Fields indicated in solid (dashed) boxes are available (are not available) in the electronic format of the administrative records.

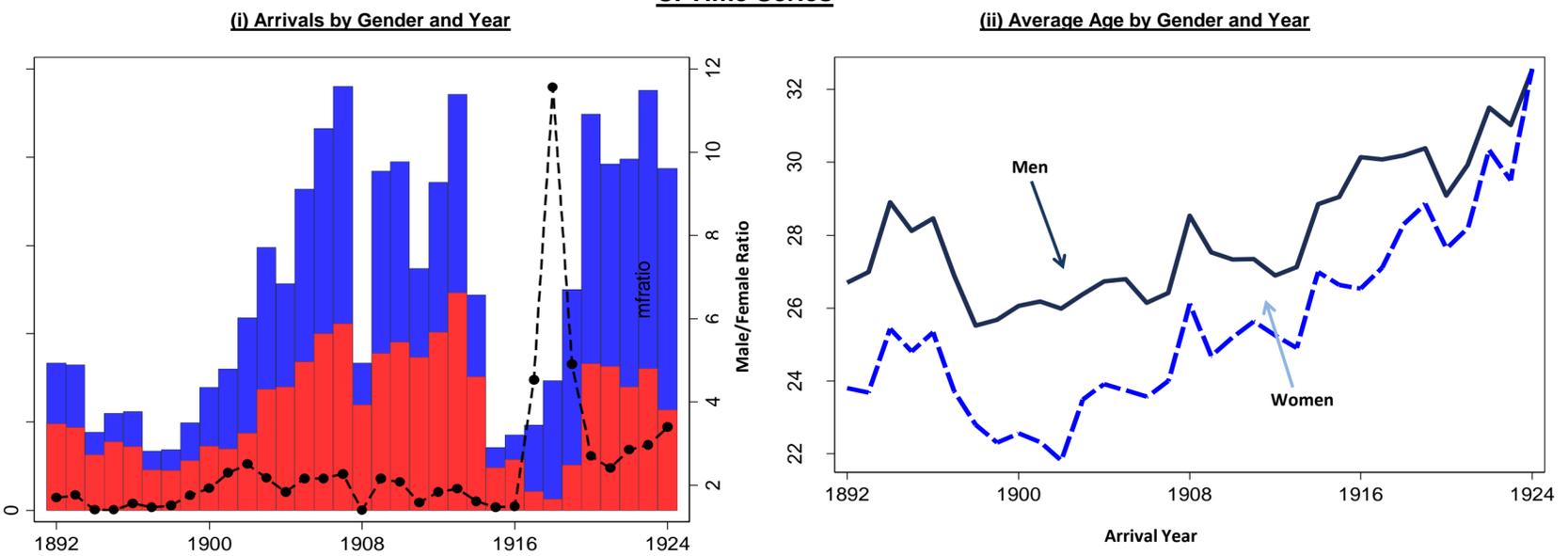# Figure 2: Descriptive Evidence from Administrative Records

A. Total Immigrants, by Nation of Birth

(i) Ellis Island Arrivals 1892-1924, and Population in 1880 US Census

EI Arrivals

1880 Census

Countries (top to bottom): Italy, Austria Hungary, Russia, Britain, Germany, Ireland, Sweden, Greece, Norway, Spain

Million

(ii) Foreign Born Population in 1880 US Census

Countries (top to bottom): Germany, Ireland, Britain, Canada, Sweden, Norway, France, China, Bohemia, Switzerland

Million

B. Age Distribution, by Gender

(i) Ellis Island Arrivals 1892-1924

Age>=85 & age<=100
Age>=75 & age<= 84
Age>=65 & age<= 74
Age>=55 & age<= 64
Age>=45 & age<= 54
Age>=35 & age<= 44
Age>=25 & age<= 34
Age>=15 & age<= 24
Age>= 5 & age<= 14
Age>= 1 & age<= 4
Age>= 0 & age<= 0

Women — Men

Percentage of Total Population of the Same Gender

(ii) Foreign Born Population in 1880 US Census

Age>=85 & age<=100
Age>=75 & age<= 84
Age>=65 & age<= 74
Age>=55 & age<= 64
Age>=45 & age<= 54
Age>=35 & age<= 44
Age>=25 & age<= 34
Age>=15 & age<= 24
Age>= 5 & age<= 14
Age>= 1 & age<= 4
Age>= 0 & age<= 0

Women — Men

Percentage of Total Population of the Same Gender

C. Time Series

(i) Arrivals by Gender and Year

mfratio

Male/Female Ratio

(ii) Average Age by Gender and Year

Men

Women

Arrival Year
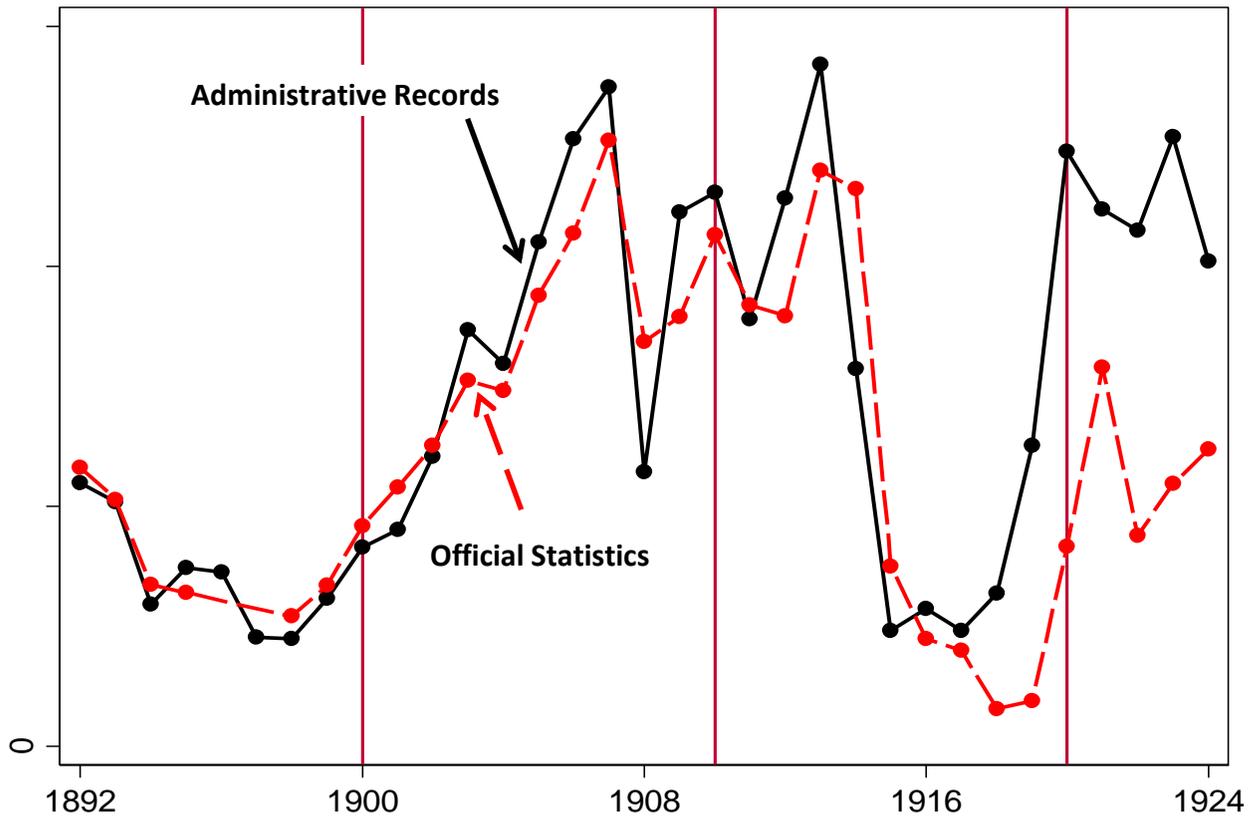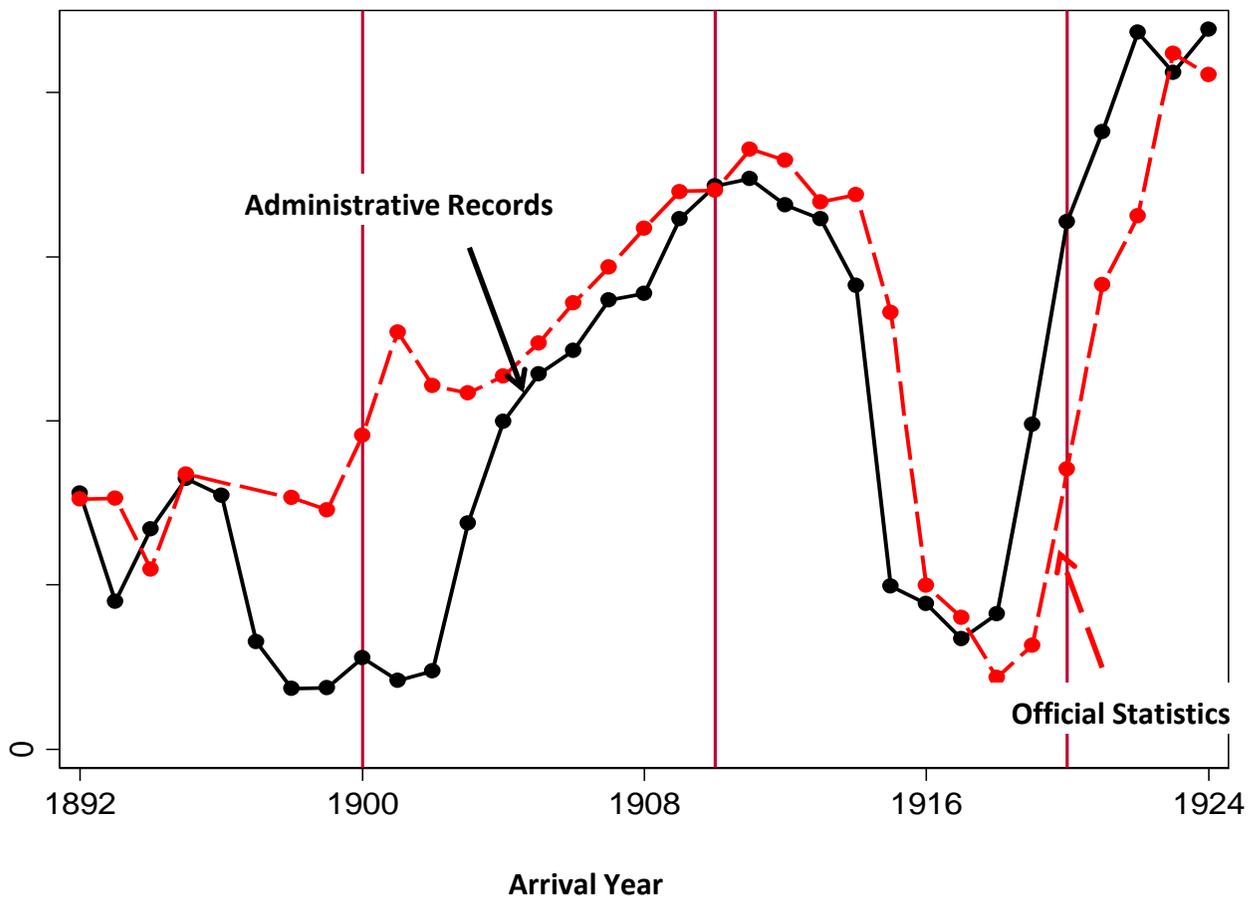
Notes: All graphs are based on the administrative data from Ellis Island records, without any adjustments. The figures for the 1880 Census are based on the 100% IPUMS sample. Figure 2Ai shows the total number of arrivals into Ellis Island from 1892 to 1924, for the ten countries from which the greatest number of immigrants originate over this time period. Figure 2Ai also shows for each nationality, the size of the foreign-born population from that country recorded in the 1880 US census. Figure 2Bi shows migrants' age distribution by gender. Figure 2Bii shows the age distribution by gender, for the foreign-born population in the US in 1880. These age pyramids show the proportion of the population of the same gender that is within a given age group. Figure 2Ci provides time series evidence on the total number of immigrants into Ellis Island each year, as indicated on the left hand axis. The right hand axis in Figure 2Ci shows the ratio of male to female migrants by arrival year. Figure 2Cii shows the average age of immigrants by year of arrival.

Figure 3: Official Statistics and Administrative Records on Arrivals into Ellis Island New York

A. Total Passenger Arrivals (Foreign and US Born) by Year

B. American Arrivals by Year

Arrival Year

Notes: All figures refer to arrivals into New York City. Official statistics figures are from Ferenzi-Willcox [1929]. All graphs are based on the administrative data from Ellis Island records with corrections for missing data, other potential errors in recorded nationalities, and exclusions. The vertical lines in each Figure correspond to 1900, 1910 and 1920. Figure 3A shows the time series for total arrivals (foreign-born and US-citizen arrivals) from both sources. Figure 3B shows the time series for US-born arrivals by year, from each source.

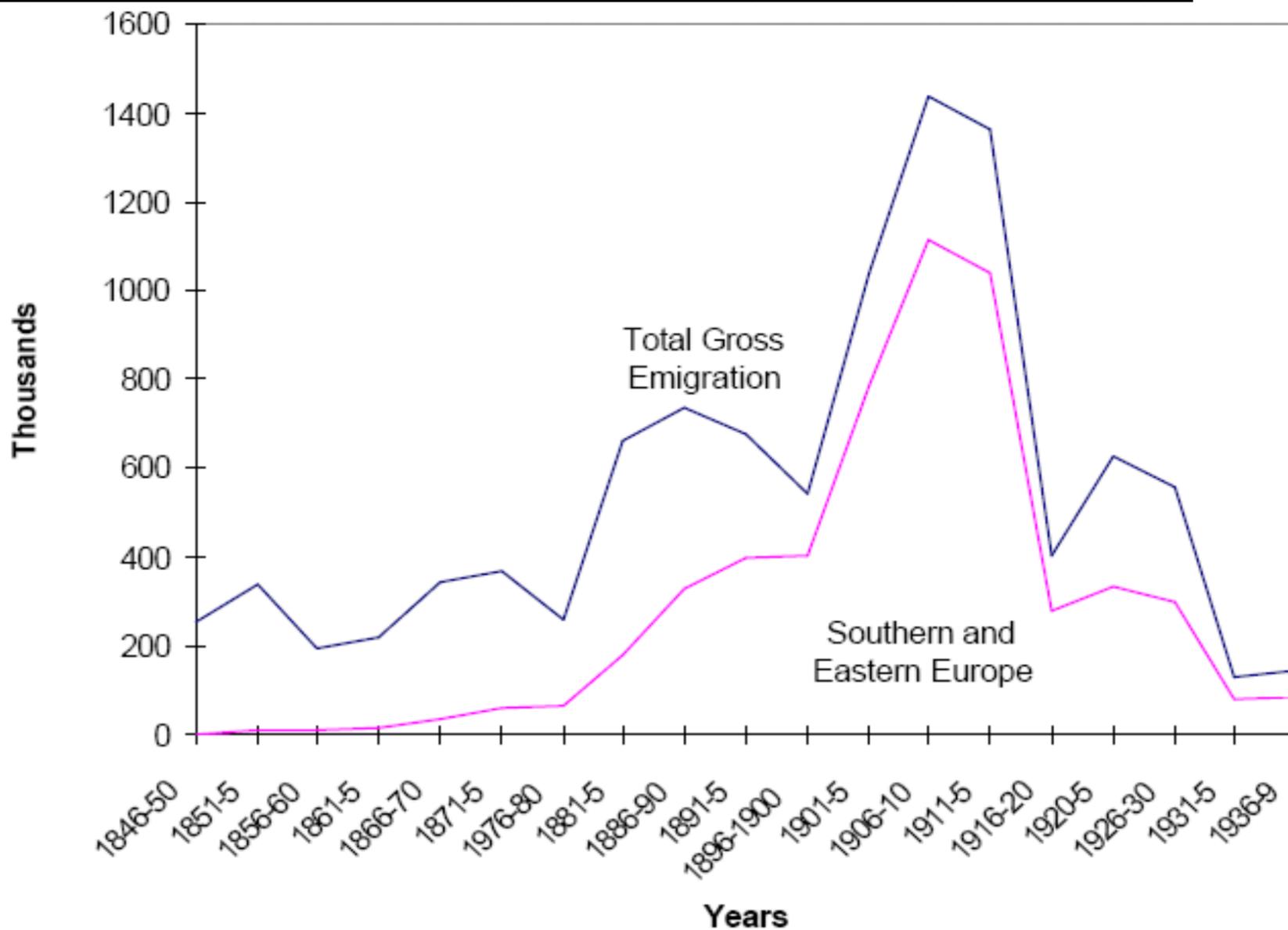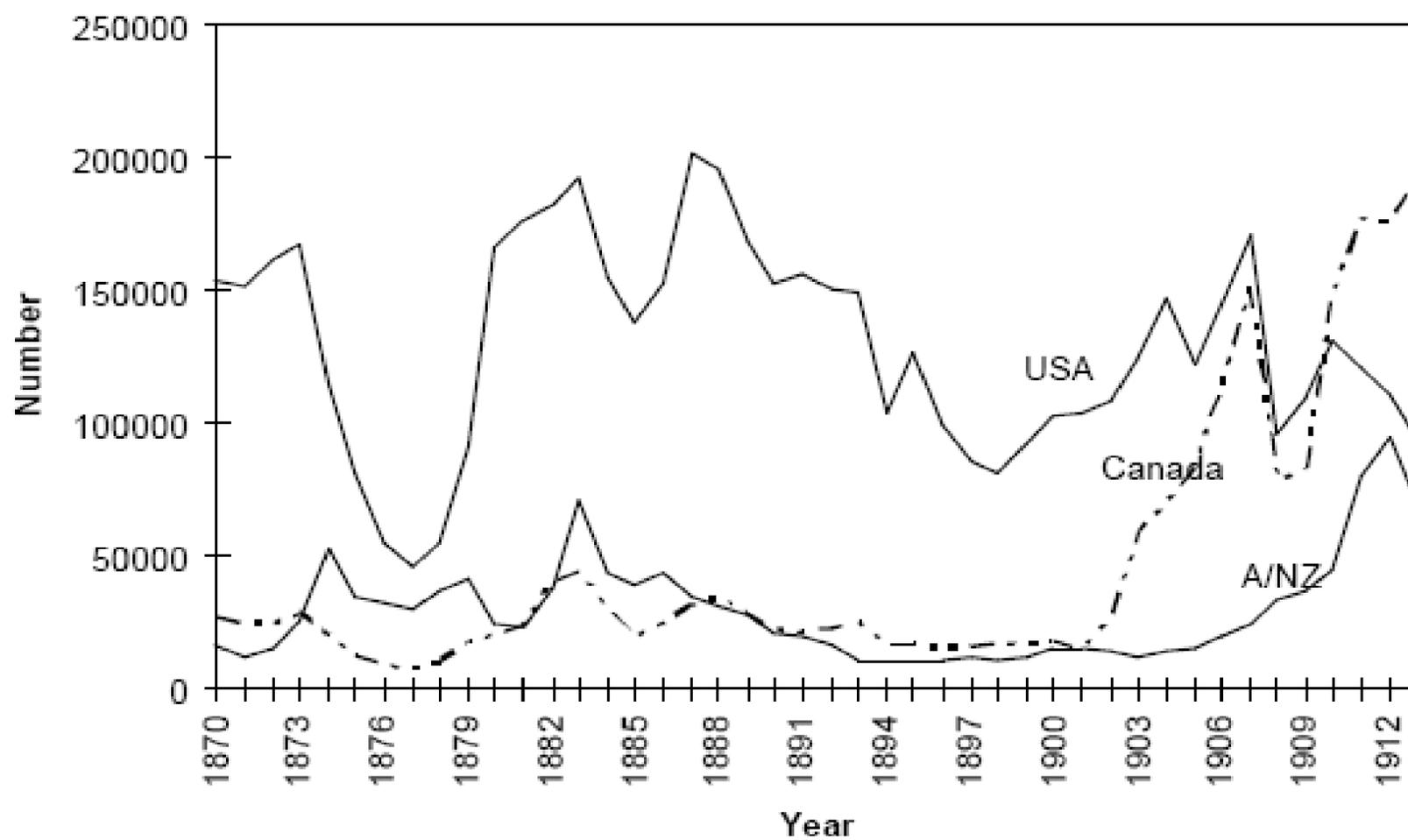**Figure A1A: Gross Intercontinental Migration From Europe: 1846-1939 (annual averages)**

Figure A1B: Gross UK Migration to United States, Canada, Australia
Figure 1B: and New Zealand, 1870-1913

Notes: Source for Figure A1A: Ferenci and Wilcox (1929). Source for Figure A1B: Hatton (2003), A/NZ refers to Australia and New Zealand as destination.

Figure A2: Accounting for Returnees and New Arrivals

Scenario	Timeline events	P_t	I_{t+1-}	$Inew_{t+10}$	P_{t+10}	Column 1 $E_{t+10}=P_t+I_{t+10}-P_{t+10}$	Column 2 $E_{t+10}=P_t+Inew_{t+10}-P_{t+10}$
1	New Arrival, Departed	1	0	0	0	1	1
2	New Arrival, Departed	0	1	1	0	1	1
3	New Arrival, Departed, Return Arrival	1	1	0	1	1	0
4	New Arrival, Departed, Return Arrival, Departed	1	1	0	0	2	1
5	New Arrival, Departed, Return Arrival	0	2	1	1	1	0
6	New Arrival, Departed, Return Arrival, Departed	0	2	1	0	2	1

Census Date: P_t Census Date: P_{t+10} Time

Figure A3a: Census Populations in 1880, 1900, 1910 and 1920, for Countries With Implied Negative Emigrant Numbers 1900-1910
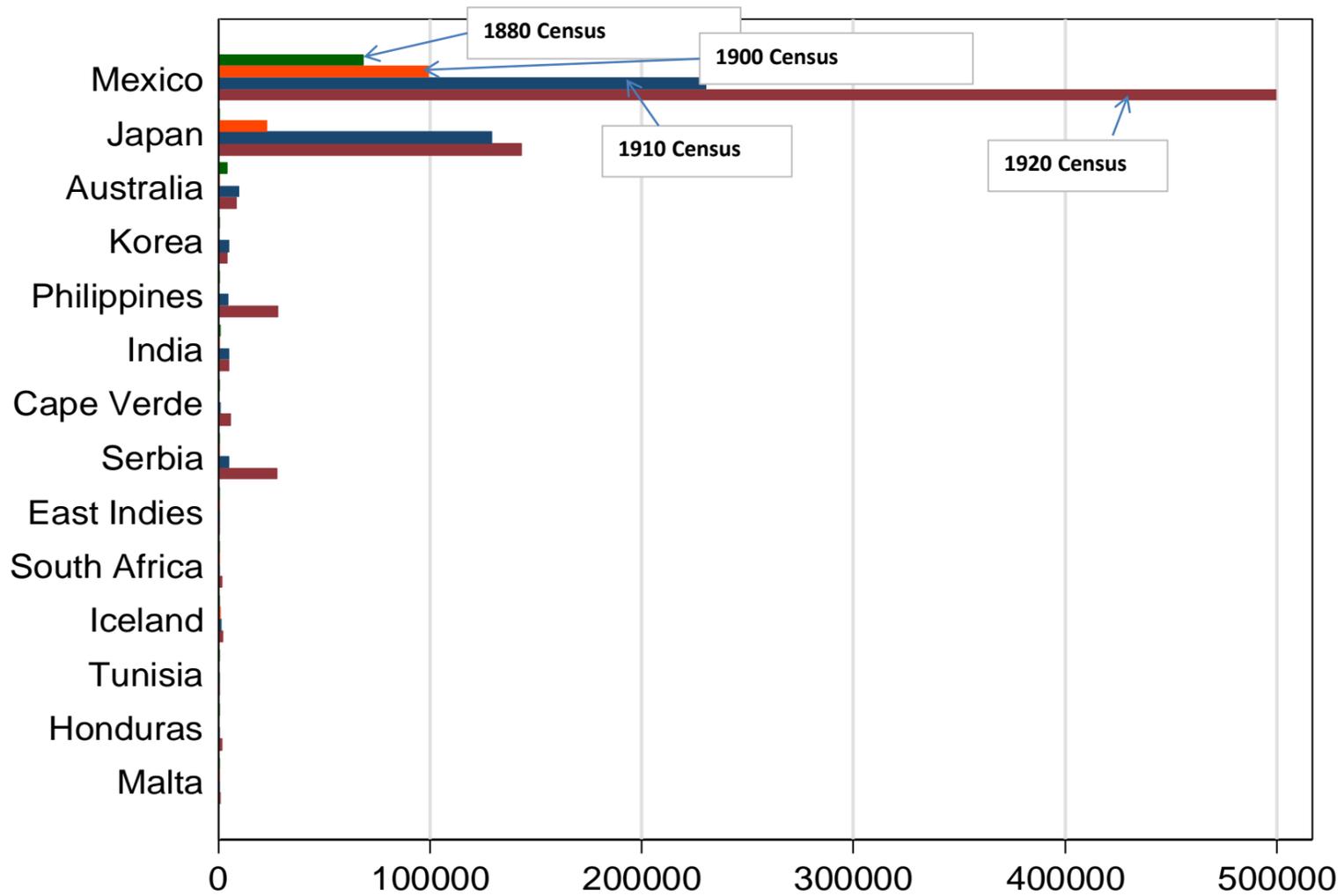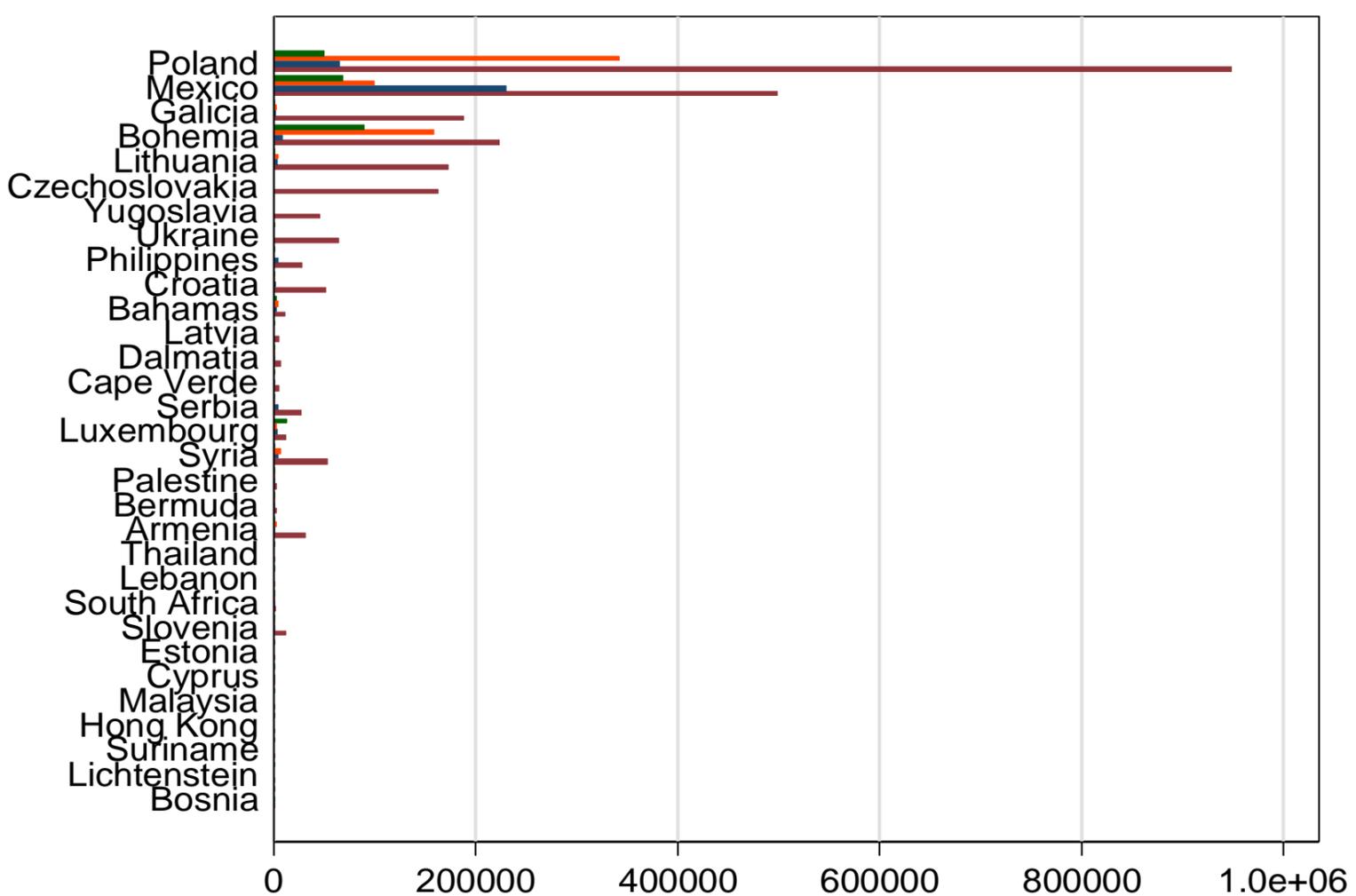

Figure A3b: Census Populations in 1880, 1900, 1910 and 1920, for Countries With Implied Negative Emigrant Numbers 1910-1920

Notes: The sample of countries for the Figure in each decade is those countries for whom the total estimated number of emigrant departures is found to be negative after making corrections for missing values, excluded immigrants, weighting for the ratio of official arrivals into the US to those into New York each year, inflows from Canada, and foreign born individuals with US citizenship. for these countries, the figure then shows the population in the US based on Census data from 1880, 1900, 1910 and 1920.