Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No. 1980
February 2024

Should I stay or should I go? Return migration from the United States

Alan Manning
Graham Mazeine

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Economic and Social Research Council

Abstract
Return migration is important, but how many migrants leave and who is poorly understood. This paper proposes a new method for estimating return migration rates using aggregated repeated cross-sectional data, treating the number of migrants in a group who arrived in a particular year as an unobserved fixed effect, and the observed number (including, importantly, observed zeroes) in the arrival or subsequent years as observations from a Poisson distribution. Compared to existing methods, this allows us to estimate return rates for many more migrant groups, allowing more in-depth analysis of the factors that influence return migration rates. We apply this method to US data and find a decreasing hazard, with most returns occurring by eight years after arrival, when about 13% of migrants have left. The return rate is significantly lower for women, those who arrive at a young age, and those from poorer; it is higher for those on non-immigrant visas for work or study. We also provide suggestive evidence that, conditional on their country of origin, those with lower education are more likely to return.

Keywords: return migration
JEL Codes: F22; J15; J61; O15

This paper was produced as part of the Centre's Community Wellbeing Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We are grateful to seminar participants at CEP for their comments, and for funding from the ERC under project LPIGMANN. Special thanks to Monica Langella for help with the country-level data.

Alan Manning, London School of Economics and Centre for Economic Performance at LSE. Graham Mazeine, University College London.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

A sizeable fraction of immigrants subsequently leave their destination country, in what is known as temporary or return migration (as most leave for their country of birth). According to OECD figures, the immigration inflow as a share of the population averaged 0.6 percentage points per annum for a set of immigrant-receiving OECD countries in the period 2000-2018[1]. But, the change in the migrant stock was only 0.3 percentage points per annum, the gap being largely explained by return migration. Some existing estimates suggest that at least a quarter of global migration flows are emigrations back to the country of birth (Azose and Raftery, 2019). These statistics are suggestive but subject to a lot of uncertainty; according to a recent survey (Dustmann and Goerlach, 2016), "the fact that migration temporariness has been so persistently ignored in the empirical literature is largely related to the inability to measure it". Many governments collect and publish voluminous statistics on immigration, but little on emigration (LaLonde and Topel, 1993), so estimates of the immigration flow and the migrant stock typically come from different sources (as in the OECD figures quoted above). It is also likely that return migration rates vary systematically among migrant groups, but we know little about that (see the review in (OECD, 2008)).

Differential return migration rates may have important implications for immigration policy. The size and mix of the stock of migrants is probably what matters for the impact of immigration, and the stock is the result of the interaction between inflows and outflows. While some aspects of the outflow are influenced by policies like whether visas are temporary or permanent or enforcement of immigration laws, immigration policies mainly regulate the inflow of migrants by determining who can enter. Most of the outflow is determined by individual choices to leave when they do not have to, i.e. voluntary return migration. Without an estimate of the level of return migration, it is hard to assess the impact on the stock of migrants of policies to change the number of visas issued.

Perhaps the best sources of data on return migration are the few longitudinal surveys that try to follow migrants as they cross borders. Most longitudinal surveys are designed so that people leave the survey when they leave the country (an exception is the Mexican Migration Project (Durand and Massey, 2019), though that is more interested in circular migration), but some try to measure whether the reason for survey exit is emigration (Constant and Massey, 2003; Nekby, 2006; Dustmann and Weiss, 2007; Gundel and Peters, 2008; Bijwaard and Wahba, 2014; Dustmann and Goerlach, 2016; Kuhlenkasper and Steinhardt, 2017; Adda et al., 2022). However, the success of this approach depends on assuming that the sizable non-response rates are not correlated with emigration. Van Hook et al. (2006) and Zakharenko (2008) use the panel element of the US Current Population Survey (CPS) and find sizable return migration, but are constrained

[1] This is for the EU15 countries with non-missing data plus Australia, Norway, Switzerland, and the United States, taken from https://stats.oecd.org/Index.aspx?DataSetCode=MIG)

by a limited sample size and the emigration-attrition independence assumption discussed above. Lubotsky (2007) uses Social Security data to follow migrants over time, but a given migrant not appearing in the data in a given year can mean they emigrated, did not work that year, or worked illegally. Cross-country matching of censuses is sometimes possible, but then it is "generally not possible to control for the date of arrival in the destination country" (OECD (2008),p168), while (Azose and Raftery, 2019)'s "pseudo-Bayesian" method is limited to the estimation of return migration in five-year windows, and further constrained by the unreliability of UN migrant stock statistics.

Given these problems, the most common method for estimating return migration rates is what is known as the "residual" method, which uses repeated cross-section data such as censuses or household surveys, sometimes supplemented by immigration statistics (see, for example, Warren and Peck (1980); Jasso and Rosenzweig (1982); Ahmed and Robinson (1994); Borjas and Bratsberg (1996); Dustmann and Weiss (2007); Bhaskar et al. (2013)). Denote by $M_g(0)$ the number of migrants in some group g who migrate in the year of arrival, and denote by $M_g(t)$ the number in the country at some later date t. The group needs to be defined in such a way that there are only two ways to leave it: emigration or death. If the fraction alive at 0 who survive to t (one minus the mortality rate) is $S_g(t)$ then the remain or retention rate (known as the survival rate in survival analysis) at time t is:

$$R_g(t) = \frac{M_g(t)}{S_g(t)\, M_g(0)} \tag{1}$$

and the return rate is one minus the retention/remain rate (OECD, 2008). To estimate return migration rates, the (unobserved) true values of $M_g(t)$ and $S_g(t)$ are replaced by estimates from, for example, censuses and life tables.

Although the residual method has the advantage of using readily available data, it does have problems. If both numerator and denominator are estimated from surveys and so contain sampling variation, there is no guarantee that the estimated remain rate (1) will be between zero and one and decline over time, as it must do. The ratio of two random variables is also typically not an unbiased estimate of the return migration rate. It is even possible that no one in the group is observed in the year of entry, in which case the estimate of $M_g(0)$ is zero and the return rate is undefined. The existing literature attempts to avoid these problems in a number of ways. One is to use administrative rather than survey data on the number of immigrants admitted (Jasso and Rosenzweig, 1982; Borjas and Bratsberg, 1996), so that the denominator is no longer a random variable and cannot be zero. However, there may be biases caused by using different data sources for the numerator and the denominator (e.g. some of the migrants may be unauthorised so have no record of admission in official statistics); a lack of availability of administrative data can also be constraining. A second approach is to define the groups very broadly so that sampling variation is reduced and the chance of estimating invalid estimated return rates is minimized. However, this makes the groups much more heterogeneous, and unobserved heterogeneity is known to be a source of bias in estimates of survival models (see, for example, Lancaster (1979)). In addition, broadly defined groups means small numbers of observations when doing a statistical analysis of the factors influencing return migration. Finally, the sometimes long gaps between data releases means that short-term (e.g.within five years of arrival) return rates are often impossible to estimate, as those who arrive and/or leave during the intervening period are never measured in the data.

In this paper, we propose a simple modification of the residual method that allows for the estimation of return migration rates for small groups, and then apply it to US data from the Census and the ACS. We treat the original size of a group in the year of arrival, $M_g(0)$, as an unobserved fixed effect. We then view the number observed in the survey in any year as a random variable, modeled as a count variable in which zero is a possible outcome. This approach allows us to define groups in much narrower ways, leading to more observations than in other studies. More narrowly-defined groups are also likely to be more homogeneous than broadly-defined groups, reducing the extent of bias from unobserved heterogeneity. Having more groups also allows a more detailed investigation of the factors that influence return rates. In our application, our groups are defined as migrants of a particular gender, from a particular country of birth, who came to the US in a particular year, in a particular age group at arrival.

The plan of the paper is as follows. Section 2 describes the data we use, while Section 3 outlines our approach in more detail. Section 4 presents our results. We find that a Weibull model is a good empirical approximation to the return rate where the return rate falls sharply over time. We estimate that about 13% of immigrants have left the US eight years after arrival, with very little emigration after that; this is similar to the figures for the US reported in OECD (2008), but somewhat smaller than most other estimates in the literature. There is variation across groups; female immigrants are more likely to remain in the US, as are those who arrive at younger ages, those from poorer countries and those on immigrant visas. Distance between the US and the origin country, the volume of trade between the two countries, and the level of democracy in the origin do not appear to be related to migrants' propensity to return. We also investigate whether return migrants are positively or negatively selected with respect to educational attainment and income; though our methodology may be biased by educational upgrading within the relevant migrant population, we present suggestive evidence that, for a given country, those with lower education are more likely to return.

2 Data

2.1 Data on the Number of Migrants

We use data from the 5% sample of the 2000 US Census and from the American Community Survey (ACS) for the years 2001-2020, which is approximately a .4% sample of the US population for the years 2001-2004 and a 1% population sample from 2005 on. Respondents to these surveys are asked their country of birth and their most recent year of arrival in the United States[2]. Our sample is the foreign-born who answer both these questions and also report their age and gender. After dropping non-institutional group quarters observations, we have 1.6 million foreign-born respondents in 2000, 110,000-120,000 per year between 2001 and 2004, and 325,000-405,000 per year up to 2020.

Our sample is constructed as follows. We restrict attention to those who arrived between 1998 and 2018, as we do not have some variables we use for earlier arrivals and 2018 is the latest possible arrival date in our data. One of the variables we use to define a group is single year of arrival, so there are 21 possible values. We drop migrants who were over the age of 65 at arrival, as mortality is likely to be large relative to return migration rates for this group, and use 4 age of arrival bands (0-15, 16-24, 25-44, and 45-64) as part of our group definition. After the sample restrictions we have 148 countries of birth (which we also use to define the group) with at least one migrant at some point during the sample period. We also distinguish between male and female arrivals. A group in our application is therefore defined as a country of origin-sex-arrival age-arrival year cell. We choose these variables to define groups (and not, for example, educational attainment) because group membership cannot change subsequent to migration, so the only way to leave the group is death or emigration from the US. Given this disaggregation (far more than in other studies), many cells have no observations in any particular year. We fill in the data set and record these as zeroes; in our approach, these are important data, not a problem for the estimation method. The maximum possible number of groups is 24,864=21*148*4*2. However, because our estimation method has a group fixed effect, only those groups with observations in at least two calendar years will contribute to the estimation. Most of the data is retained: the number of groups used in estimation is 18,508, with most of the dropped groups being dropped due to missing covariates (which are discussed in Section 2.2 below), and 94.0% of migrants in the data for our 1998-2018 arrival period are in a cell that is used for estimation.

Table 1 provides some demographic information on the characteristics of migrants. The first column shows some information about the stock of migrants, averaged over all years in the sample. The second column shows the same characteristics for the inflow, defined as those who report arriving in the year of the sample.

[2] Our estimates are thus our estimates of the duration of each stay and cannot say anything about repeat or circular migrants to the US.

There are three reasons the characteristics of stock and flow can differ: differential mortality and return migration rates, and a changing inflow over time (the stock reflects historic as well as current inflows). To give some indication of changes in the flow, we report characteristics of the inflow for four arrival-year bins: 1998-2002, 2003-2007, 2008-2012, and 2013-2018. Perhaps the most noteworthy changes over time in the flow of migrants is the rising proportion of new migrants aged 45+ from about 12% in 1998-2002 to 21% in 2013-18, most likely the result of waiting lists for some immigrant visas that can be decades-long, leading to much later ages of arrival (Carr and Tienda, 2013). There is also a noticeable shift in the region of origin distribution, with more recent arrivals about half as likely to come from Central/North America and the Caribbean as in earlier years, while the proportion of immigrants coming from Asia has nearly doubled to become the largest source of incoming migrants.

2.2 Other Explanatory Variables

Some of the explanatory variables we use are suggested by a number of theories of return migration which are usefully surveyed in Dustmann and Goerlach (2016). Many of these theories are based around the comparison of life in the destination and origin countries. As a measure of of the gap in material living standards between the two countries, we use purchasing power parity-adjusted home country GDP per capita relative to the US (from World Bank data). GDP per capita has been found to play an important role in determining the destination (both desired and actual) of migrants (Borjas, 1987; Hatton, 2005; Mayda, 2010; Ortega and Peri, 2013; Clemens, 2014; Borjas et al., 2019; Langella and Manning, 2021), so it would not be surprising if it also influenced the rate of return migration. A simple-minded view would be that the lower the income in the origin country, the less attractive it is to return (for example, Lessem (2018) finds that increases in Mexican wages reduce the length of stay of Mexican immigrants in the US). However, there are a number of reasons why the relationship might be more complicated; if savings targets in the host country can be met quicker (Galor and Stark, 1990), or capital accumulated in the destination country can be used to finance a better lifestyle in the origin country (Dustmann, 2003), a bigger income gap between the US and the origin may lead to shorter stays; Dustmann presents some evidence for non-monotonicity from German panel microdata. Adda et al. (2022) argue that those who expect to stay longer will invest more in destination-specific human capital, producing a correlation between wage growth and duration of stay. We do not have the microdata to test many of these hypotheses but it is important to investigate a possible non-monotonic relationship

Table 1: Share of US immigrant population, 1998-2018

	Stock	Flow (all years)	1998-2002	2003-2007	2008-2012	2013-2018
Sex:						
Male	0.502	0.512	0.531	0.537	0.503	0.491
	(0.500)	(0.500)	(0.499)	(0.499)	(0.500)	(0.500)
Female	0.498	0.488	0.469	0.463	0.497	0.509
	(0.500)	(0.500)	(0.499)	(0.499)	(0.500)	(0.500)
Age at arrival:						
0-15	0.243	0.220	0.260	0.213	0.205	0.213
	(0.429)	(0.414)	(0.439)	(0.409)	(0.404)	(0.409)
16-24	0.280	0.238	0.260	0.258	0.241	0.213
	(0.449)	(0.426)	(0.438)	(0.437)	(0.428)	(0.409)
25-44	0.370	0.368	0.360	0.384	0.360	0.366
	(0.483)	(0.482)	(0.480)	(0.486)	(0.480)	(0.482)
45-64	0.085	0.116	0.086	0.108	0.129	0.130
	(0.279)	(0.321)	(0.281)	(0.310)	(0.335)	(0.337)
65+	0.022	0.057	0.034	0.038	0.065	0.078
	(0.145)	(0.232)	(0.182)	(0.191)	(0.246)	(0.268)
Region of origin:						
North/Central America	0.402	0.329	0.465	0.418	0.262	0.239
	(0.490)	(0.470)	(0.499)	(0.493)	(0.440)	(0.427)
Caribbean	0.059	0.052	0.040	0.049	0.063	0.054
	(0.235)	(0.222)	(0.196)	(0.216)	(0.242)	(0.225)
South America	0.076	0.073	0.103	0.061	0.059	0.073
	(0.265)	(0.259)	(0.304)	(0.239)	(0.236)	(0.260)
Europe	0.094	0.119	0.124	0.119	0.123	0.113
	(0.292)	(0.323)	(0.330)	(0.324)	(0.328)	(0.317)
Asia	0.287	0.338	0.213	0.282	0.386	0.411
	(0.453)	(0.473)	(0.410)	(0.450)	(0.487)	(0.492)
Middle East	0.023	0.033	0.017	0.023	0.048	0.039
	(0.149)	(0.179)	(0.130)	(0.151)	(0.214)	(0.193)
Africa	0.052	0.046	0.027	0.039	0.048	0.060
	(0.222)	(0.210)	(0.163)	(0.194)	(0.214)	(0.237)
Pacific	0.007	0.010	0.010	0.009	0.011	0.011
	(0.0836)	(0.101)	(0.0998)	(0.0929)	(0.103)	(0.105)
Sum of weights	280058201	26633179	4968159	6421120	5846693	9397207
Total cells	182766	13655	1995	3106	3613	4941

Each column displays the proportion of migrants in that row's category, with standard deviations in parentheses. 'Stock' column gives the proportions of all migrants in all years of our sample, 'Flow' gives the proportions of migrants that arrive during our sample period, and the remaining columns give the proportions of arrivals in the listed years.

between income gaps and return migration.[3].

Table 2 presents evidence on PPP-adjusted GDP per capita relative to the United States in migrants' countries of origin. Note that those from richer countries are a bigger share of the inflow than the stock, suggesting that they are more likely to return home. The flow does not appear to have changed greatly, with over 75% of US immigrants coming from countries with GDP per capita less than 25% of the US level. It may also be not just the current level but also the prospects for growth that affects return migration (Dustmann et al. (2011); Adda et al. (2022)), so we also experiment with also using the growth rate of GDP per capita in our regressions.

One concern may be that relative GDP per capita may not be a good measure of relative income levels in the two countries that can be expected by migrants. For example, Hendricks and Schoellman (2018) shows that immigrants from poorer countries tend to be located lower on the US income distribution than those from richer ones. To address these concerns, some of our specifications include country fixed effects, which control for any time-invariant factors. Our results on the role of income are robust to this.

The quality of life in the home country relative to the US is likely to depend not just on the material standard of living but also other factors. We use data from Polity5 on the level of democracy in the origin country[4], and data on conflict occurrence from the Uppsala Conflict Data Project. Table 2 shows how these variables have been changing in the stock and flow of migrants.

The proportion of an immigrant cohort remaining in the US after t years is likely to depend not just on origin conditions in year t, but in every year after leaving the origin, since immigrants decide whether or not to return in each year after arrival. To reflect this, we use a running average for each of the time-varying country of origin variables, which captures the average magnitude of the variable in the period since the immigrant cohort's arrival in the US. This can be thought of as derived from a specification in which the hazard rate in any year depends on the current characteristics and the log return rate after t years is the sum of the hazard rate. These three variables can be interpreted as the average relative GDP/capita in the origin country since arrival, the average level of democracy, and the proportion of years since arrival in which the origin country experienced armed conflict[5]. We also use the distance (in kilometers) from each country of origin's most populous city to the closer of New York City and Los Angeles, data taken from CEPII. Distance has been found to be very important in explaining the destination of migrants (Lewer and

[3]Zelinsky (1971) proposed the "migration transition" hypothesis in which there is a non-monotonic relationship between income and emigration: rising income per capita in the origin country first allows more potential migrants to incur fixed costs of migration, but, after a certain threshold, rises in the standard of living in the home country reduces emigration. (Clemens, 2014, 2017, 2020; Docquier et al., 2014; Djajic et al., 2016)claim support for this, though others (Ortega and Peri, 2013; Bencek and Schneiderheinze, 2020; Langella and Manning, 2021) are more skeptical. There could conceivably be a hump-shape in return migration, for the same or different reasons.

[4]Polity's scale runs from -10 to 10, with -10 indicating autocracy and 10 indicating full democracy.

[5]The source variable from the UCDP is a dummy indicating conflict occurrence

Van den Berg, 2008; Gallardo-Sejas et al., 2006); it may be less important in determining return migration once a journey has been made, but we investigate this. On the whole, immigrants entering the US later in our sample period come from more distant, less democratic, and more conflict-prone origin countries than earlier arrivals.

Finally, we control for visa type on entry to the US. Those who enter on non-immigrant visas, with no path to permanent residency without changing visa, are expected to leave the US with a much greater frequency than those who come on immigrant visas, with the accompanying permanent residency rights. There is also likely to be heterogeneity across visa categories not directly tied to immigrant or non-immigrant status: for example, those who come on family-related visas may be more or less likely to return than those who come on work visas[6]. We use Department of Homeland Security (DHS) data and the the Immigration and Naturalization Service (INS) Yearbook of Immigration Statistics to construct, for each country and year of admission, the proportion of migrants on immigrant visas, and the proportion of entry visas in the categories of family, work, student, humanitarian (including refugees), and diversity. All of our data sources and the construction of our country-level variables are described in more detail in Appendix A.

Table 2: Share of US immigrant population, 1998-2018 (continued)

	Stock	Flow (all years)	1998-2002	2003-2007	2008-2012	2013-2018
Relative GDP/capita, origin country:						
0-25%	0.504	0.495	0.388	0.426	0.519	0.584
	(0.500)	(0.500)	(0.487)	(0.494)	(0.500)	(0.493)
26-50%	0.373	0.311	0.459	0.384	0.254	0.219
	(0.484)	(0.463)	(0.498)	(0.486)	(0.435)	(0.413)
51-75%	0.057	0.080	0.065	0.080	0.085	0.085
	(0.232)	(0.271)	(0.247)	(0.272)	(0.279)	(0.278)
75%+	0.066	0.114	0.088	0.110	0.142	0.112
	(0.248)	(0.317)	(0.283)	(0.312)	(0.350)	(0.316)
Origin conflict	0.579	0.499	0.324	0.545	0.532	0.538
	(0.494)	(0.500)	(0.468)	(0.498)	(0.499)	(0.499)
Polity IV score	6.980	6.926	7.401	7.302	6.777	6.511
	(2.974)	(3.234)	(2.644)	(2.858)	(3.391)	(3.580)
Distance to US (1,000 mi.)	6.264	6.762	5.594	6.109	7.246	7.525
	(3.904)	(3.908)	(3.635)	(3.854)	(3.921)	(3.852)
Sum of weights	280058201	26633179	4968159	6421120	5846693	9397207
Total cells	182766	13655	1995	3106	3613	4941

Each column displays the proportion of migrants in that row's category, with standard deviations in parentheses. 'Stock' column gives the proportions of all migrants in all years of our sample, 'Flow' gives the proportions of migrants that arrive during our sample period, and the remaining columns give the proportions of arrivals in the listed years.

[6]Type of visa has been found to be very important in predicting return migration in UK administrative data - see (Hall et al., 2023)

3 Methodology

We now derive our regression specification from the general formula for the return rate. Re-arrangement of (1) means that the number of migrants who remain after t years can be written as:

$$M_g(t) = R_g(t)\, S_g(t)\, M_g(0) \tag{2}$$

where $M_g(0)$ is the number of migrants in some group g who migrate in the year of arrival, $M_g(t)$ the number in the country at some later date t, $S_g(t)$ is the fraction alive at 0 who survive to t (one minus the mortality rate) and $R_g(t)$ is the fraction who remain in the country. If the sampling probability of someone being observed in the survey is $p_g(t)$ then the expected number of migrants of group g observed in the sample after t years is:

$$E[N_g(t)] = p_g(t)\, M_g(t) = p_g(t)\, R_g(t)\, S_g(t)\, M_g(0) \tag{3}$$

Taking logs, this can be written as:

$$ln(E[N_g(t)]) = ln(p_g(t)) + ln(R_g(t)) + ln(S_g(t)) + ln(M_g(0)) \tag{4}$$

The first term on the right-hand side of (4) $ln(p_g(t))$ is the log probability of the individual being in the sample. In a truly random sample, this would be a constant. In our surveys, there are weights which are the inverse of probability of the individual appearing in the sample, so $ln(p_g(t))$ is known. We also make an adjustment to the weights for the undersampling of immigrants who arrive in the year of the survey. The ACS samples continuously through the year, but interviews in January can only find migrants arriving that year who arrive in January. By December all migrants that year will potentially be sampled. This leads to an under-count of the number of migrants arriving in the current year; a December arrival is only $\frac{1}{12}$ as likely to be sampled as a January arrival. Month of arrival is unobservable in our dataset, but we assume arrivals are uniformly distributed through the year, in which case we can multiply the weights on all contemporaneous arrivals by the reciprocal of $\frac{1}{12}\sum_{i=1}^{12} i$, or 1.846, to account for the undersampling. In practice, this adjustment affects the estimated return rate between the year of arrival and the following year, but not between later years. For the 2000 Census data, we multiply the survey weights by 4 to reflect the fact that the survey is designed to capture the US population on April 1st of the census year, thereby missing foreign arrivals in the remaining $\frac{3}{4}$ of the year.

Using these adjustments, each observed migrant has a weight attached to them. To measure the total number of migrants in a group (however defined), we simply add up the weights over all migrants within the group. This gives us an estimate for the total number of arrivals in the US in a given year, and the time path of that stock over the following years. These form our dependent variable $N_g(t)$.

The third term $ln(S_g(t))$ is the log probability of living from the year of arrival to the year of observation. Following previous practice (Warren and Peck, 1980; Ahmed and Robinson, 1994; Bhaskar et al., 2013), we compute the survival rate from the life tables produced by the US Social Security Administration, which uses data from the National Center for Health Statistics[7], with mortality rates varying by age and gender. The published mortality tables do not distinguish between natives and migrants, so the assumption here is that mortality rates are the same for both groups. In practice, the survival adjustment makes little difference for younger migrants when mortality rates are small relative to return rates, but is more problematic for older migrants, where the reverse is the case. For this reason, we restrict our main analysis to migrants who are younger than 65 years at arrival.

The last term in (4) $ln(M_g(0))$ is the log of the number of the group that arrived; this is where we innovate, treating this as an unobserved fixed effect in estimation. The second term in (4) $ln(R_g(t))$ is what we are interested in, the log of the retention rate (which is one minus the return rate). To estimate the parameters of (4) we use a fixed effects Poisson model, where the dependent variable is the number of migrants from group g observed at time t. This empirical model allows us to include years where no migrants from a particular group are observed; in fact, it is important not to exclude them to avoid bias. The weights and the adjustment for mortality rates can either be included as offsets or used to adjust the dependent variable; the results are identical for both methods. In the next section, we present the results of a set of progressively more complicated models for the return rate.

4 Results

4.1 Baseline Models

Our first model simply includes a dummy variable for each year since arrival assuming that all groups have the same return rates. The excluded dummy variable is for the year of arrival. The coefficient on the dummy variable for year t then estimates $ln(R(t)) - ln(R(0)) = ln(R(t))$, as it must be the case that $ln(R(0)) = 0$. Figure 1 presents the results from this specification, plotting the estimated fixed effects (along with confidence intervals which are very small) to visually demonstrate the time path of the stock of immigrants in the period after arrival. The estimated coefficients generally become smaller over time which

[7] Death probabilities can be downloaded from https://www.ssa.gov/oact/HistEst/Death/2020/DeathProbabilities2020.html.

is reassuring as they are survival rates. However there are clear spikes at years divisible by 10, suggestive of "heaping", where some migrants approximate to round numbers how long they have been in the US. The time path of the migrant stock is relatively smooth apart from those spikes.

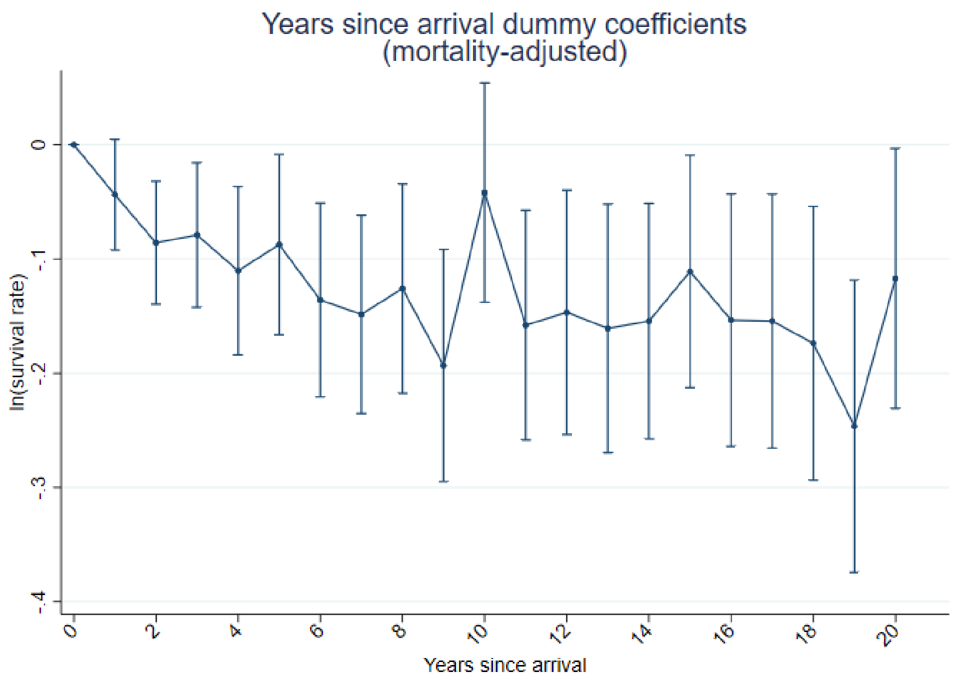

Figure 1: Mortality-adjusted survival rates from the United States

Our estimates imply that after eight years the log retention rate is about -0.13, implying that 13% of migrants have left the US eight years after arrival. Few migrants are estimated to leave after this point. It is hard to compare directly our estimates of return migration rates with existing estimates using the residual approach in the literature, as most do not report single-year hazard rates, and most also pool many arrival years in a single group. Warren and Peck (1980) defines a group as all immigrants who arrived in the US between 1960 and 1970, finding that 18% had left by the latter year. Jasso and Rosenzweig (1982) defines a group as all immigrants admitted to the US in 1971, estimating that up to 50% may have left by 1979. Borjas and Bratsberg (1996) defines the group as 1970-1974 arrivals, estimating a 10-year return rate of 21.5%. Ahmed and Robinson (1994) use ten-year arrival cohorts, as does Bhaskar et al. (2013), which finds that 14% of those who arrived in the US during the 1990s but were still there in 2000 had left by 2010. (OECD (2008)) finds that around 19% of 1999 entries to the United States had left by 2004, excluding first-year exits, lower retention rates than in other OECD countries. Thus, our estimates suggest a lower propensity to return than most previous estimates for the US. Our finding that very few migrants leave after eight years or so is consistent with (Dustmann and Weiss, 2007; Hall et al., 2023)'s findings for the UK and Bijwaard and

Wahba (2014)'s findings for the Netherlands, where the return intensity fades after 5 or 6 years.

This baseline model makes the implausible assumption that all migrant groups have the same return rate. Although, one could, in principle, interact the dummy variable for each year since arrival with group characteristics, the results would be hard to interpret, and estimation would be computationally expensive. For that reason, we seek a more parsimonious model of time that gives a good fit for the retention rate. The model we find fits well is a Weibull model, in which the return rate has the form:

$$ln\left(R_g\left(t\right)\right) = -\beta\frac{5}{100\lambda}\left(\frac{t}{5}\right)^{\lambda}, t \geq 1, \lambda \leq 1 \tag{5}$$

with β and λ parameters to be estimated. We always also include dummy variables for 10 and 20 years since arrival, to account for the heaping effect in retrospective year of arrival reporting discussed above. The parameterisation of the Weibull model in (5) is chosen to make reported estimates readily interpretable; the derivative of the log retention rate with respect to t is minus the hazard rate of leaving which, given (5) takes the form:

$$100 * h_g\left(t\right) = \frac{\partial ln\left(R_g\left(t\right)\right)}{\partial t} = \beta\left(\frac{t}{5}\right)^{\lambda-1} \tag{6}$$

This implies that the estimated β is the (positive) annual return (or hazard rate) in percentage points after five years. The estimate of λ tells us about the time dependence in the annual return or hazard rate. With λ<1, the annual return rate is declining over time (i.e. negative duration dependence), leading to the shape of the log return rate in Figure 1. We use a grid search to compute the value of λ that best fits the data; in the aggregated data, the value is 0.365 (S.E.=.069), with an associated estimate of β of 0.811 (S.E.=.314). These estimates imply that the 0.81% of migrants who have been in the US for five years will leave in the next year. The estimated value of λ implies that the hazard rate after one year is about three times that after five years ($= \left(\frac{1}{5}\right)^{0.365-1}$) and the hazard rate after ten years is two-thirds that after five years ($= \left(\frac{10}{5}\right)^{0.365-1}$).

The Weibull model implies that the log retention rate should be linear in log time since arrival: Figure 2 shows that this provides a reasonable fit to the data.

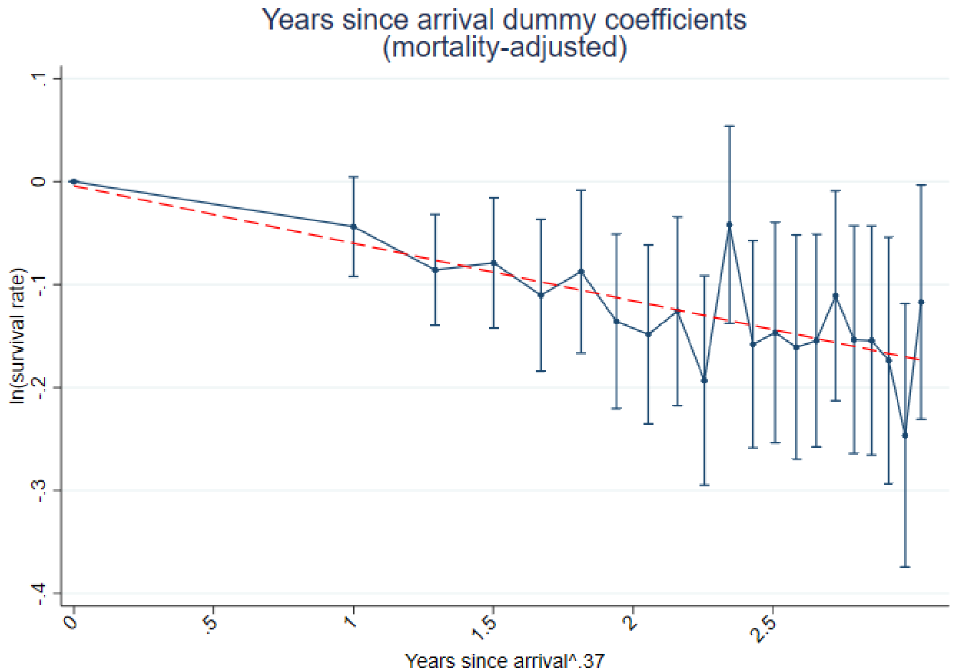

Figure 2: Mortality-adjusted survival rates from the United States

λ is also precisely estimated; Figure 3 shows the log likelihood of the model for different values, and illustrates that the optimal $\widehat{\lambda}$ occurs at a clear maximum of the log likelihood function.

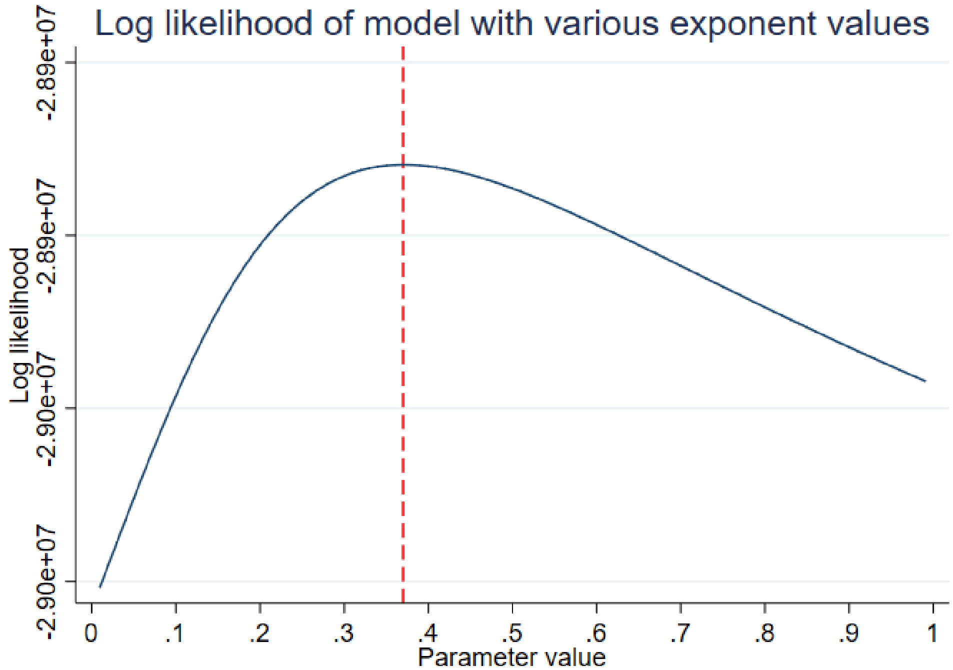

Figure 3: Log likelihood of aggregate model

4.2 Including covariates

Our next model introduces group characteristics on the right hand side and takes the form:

$$ln\left(R_g\left(t\right)\right) = -\left[x'_{gt}\beta\right]\frac{5}{100\lambda}\left(\frac{t}{5}\right)^{\lambda}, t \geq 1 \tag{7}$$

i.e. it interacts group characteristics with a Weibull function for the time dependence of the log retention rate. Note that group characteristics only enter (7) interacted with time; there is no "main effect". This is because $R_g\left(0\right) = 1$ by definition, i.e. the retention rate must be equal to 1 in the initial period for all groups; it is the interaction with years since arrival that is important. As before, λ tells us about the time dependence in the hazard rate while the coefficient on a group characteristic tells us about how the one-year hazard rate after 5 years varies with that characteristic. To see this, note that:

$$100 * \frac{\partial ln\left(h_g\left(t\right)\right)}{\partial x'_{gt}} = \frac{\partial^2 ln\left(R_g\left(t\right)\right)}{\partial t \partial x'_{gt}} = \beta\left(\frac{t}{5}\right)^{\lambda-1} \tag{8}$$

We center the continuous independent variables at their sample means, so the estimated constant remains interpretable as the return rate after 5 years at the mean value of the regressors. The first column of Table 3 presents estimates where we control for sex, age at arrival, the proportion of migrants from that country in that year of migration who entered the US in our five categories of visa, as well as the proportion on immigrant visas and year fixed effects. We include the country of birth characteristics discussed above in Section 2: the log of relative income per capita between the US and the origin country, measures of democracy and conflict in the origin, and the distance from the US. To compute standard errors, we use a bootstrap procedure. We draw (with replacement) 500 random samples from the dataset, where the sampling unit is the migrants' country of origin. For each bootstrap sample, we estimate the parameters of the return rate model, using a grid search to estimate λ, the exponent on the time trend. Table 3 (and Table 4 below) report the mean estimated parameter vector over the 500 bootstrap sample estimates and use the standard deviation of those estimates as an estimate of the standard error of each coefficient.

The estimate of λ in column 1 of Table 3 is 0.40, similar to the 0.37 found when we did not include any covariates[8]. Turning to the coefficients on the covariates, recall that there are no main effects: in line with (7), they are all interactions with the function of years since arrival. The estimated constant implies that the hazard rate after 5 years is 2.1% for our base category (which is male, arrived aged 25-44, and all on family non-immigrant visas with average levels of the continuous covariates), The coefficient of -1.07 on the female dummy means that the return rate is lower for women than for men. This is similar to the findings in OECD (2008), who find a 5-year return rate of 22% for women and 16% for men[9]. We also find that younger age groups have lower return rates. Those who come from richer countries are less likely to remain in the US; the difference in the annual hazard rate after 5 years between immigrants that come from a country at the 10th percentile of relative GDP/capita and immigrants from a country at the 90th percentile is estimated to be 5.91 percentage points. We find that those from democracies are less likely to remain, though the impact is very small: a one-point increase in the origin country's Polity5 democracy index in the 5 years since arrival is associated with a .08pp higher return hazard rate. We find that those from origin countries with ongoing conflict are more likely to return; this is opposite to what might be expected but turns out not to be robust to the inclusion of country fixed effects. The effect of physical distance between the US and the origin country is effectively zero; this contrasts with most studies of flows that find a strong chilling effect of distance. However, the migrants we are studying have already made the journey, so it is perhaps not surprising that the likelihood of a return journey is not strongly related to how far they have traveled. Those on immigrant visas are much more likely to remain, which is to be expected, as these visas offer permanent residence–this estimate, however, is quite noisy, and not statistically significant. The non-significance of the visa type coefficients indicate that conditional on immigrant vs. non-immigrant status, visa type largely does not influence return propensities, with the exception of Diversity visas, possessors of which are significantly

[8]The similarity in the two estimates of λ might be surprising given the general conclusion in duration models that the greater the amount of unmodeled heterogeneity in hazard rates, the more negative the estimated duration dependence is likely to be; see, for example, Lancaster (1979), Lancaster and Nickell (1980), and Elbers and Ridder (1982). The intuition is that those with long durations are disproportionately drawn from those with low exit rates. However, this logic does not directly apply to our analysis, as we estimate at the group level, and the weights on different groups do not change with duration.

[9]The authors note that while this result is robust for the US, it is not true in Europe, where male and female immigrants exhibit similar return propensities.

more likely to remain in the US.

Columns 2 through 4 test the robustness of our results to the inclusion of other controls and fixed effects. Column 2 contains results from the main specification with the inclusion of origin country fixed effects. The impact of immigrant demographics are robust to using within-country variation only, and the effect of relative GDP/capita in the origin is actually larger than without country-fixed effects. One possible reason for this is that migrants from low-income countries tend to have lower earnings in the US Hendricks and Schoellman (2018)[10], so our estimate of income differentials without country fixed effects understates the true effect of income. The inclusion of country fixed effects controls for any possible time-invariant difference between our relative income measure and the true one. With country fixed effects the sign of the impact of conflict in the origin country is now flipped to be negative. The immigrant visa share is now insignificant, but still negative: statistical insignificance in this specification is not surprising given that the majority of the variation in immigrant visa share is across countries rather than within them.

Column 3 adds controls for two more economic determinants of migration: total US-origin trade as a share of GDP in the origin, and the natural log of the exchange rate, denominated in foreign currency. While theory suggests a positive relationship between trade ties and migrant flows, it is less clear what impact trade should have on return migration; the trade/GDP coefficient is greater than zero but not significant. Exchange rates have been shown to affect immigrants' decisions about reservation wages Dustmann et al. (2021), remittances (Yang, 2008), and labour supply (Nekoei, 2013). Exchange rates have an a priori ambiguous relationship with return rates: a depreciated currency in the origin means USD-denominated remittances are now more valuable in the origin, and so would increase the relative value of remaining in the US, but also might allow migrants to return home sooner, if they have a target level of remittances.Yang (2006) finds that host currency appreciation increases the return rate. However, we find that the exchange rate coefficient is a quite precise zero.

We also investigate possible non-monotonicity in the impact of relative income. We replace the relative GDP/capita term with ten dummies for being from a country with 0-10%, 10-20%, 20-30%, and so on of the US's GDP/capita. The estimated coefficients rise monotonically with home GDP/capita, suggesting that migrants from richer countries are more likely to go home than those from poorer ones. These estimates are not necessarily inconsistent with the Zelinsky (1971) "hump-shape" hypothesis for emigration; one of the arguments for the hump-shape is liquidity constraints faced by migrants in the poorest countries. All the migrants in our sample have overcome any such constraints, so it may not be that surprising that there is only a monotonic relationship for return migration.

[10]We show in Figures A.1 and A.2 of Appendix A that this is true in the ACS data, as well.

Table 3: FE Poisson coefficient estimates

	(1)	(2)	(3)	(4)	(5)
Return rate after 5 years	2.13***	3.21***	2.10***	6.18**	2.21**
	(0.526)	(0.811)	(0.510)	(1.950)	(0.835)
Exponent on time trend	0.40***	0.46***	0.40***	0.58***	0.41***
	(0.029)	(0.032)	(0.029)	(0.120)	(0.032)
Female	-1.07*	-1.16*	-1.06*	-1.90	-1.02
	(0.504)	(0.518)	(0.508)	(0.991)	(0.531)
Age at arrival:					
0-15	-0.88*	-0.92*	-0.88*	-1.44	-0.85*
	(0.355)	(0.400)	(0.345)	(0.817)	(0.380)
16-24	-0.40	-0.39	-0.36	-0.86	-0.37
	(0.292)	(0.307)	(0.303)	(0.621)	(0.308)
45-64	0.98***	0.99***	0.97***	1.75*	0.99***
	(0.151)	(0.163)	(0.155)	(0.683)	(0.160)
relative GDP/capita	0.73***	1.04**	0.74***	0.85***	.
	(0.098)	(0.355)	(0.111)	(0.230)	(.)
PIV Democracy score	0.08	0.07	0.09	-0.03	0.05
	(0.064)	(0.094)	(0.061)	(0.219)	(0.050)
Conflict indicator	0.50	-0.84	0.38	-0.64	0.14
	(0.348)	(0.448)	(0.340)	(1.142)	(0.314)
Distance from origin (1,000 k.m.)	0.03		0.04	-0.04	0.04
	(0.057)		(0.063)	(0.126)	(0.049)
US Trade/GDP in origin			0.58		
			(1.894)		
ln(exchange rate, 1/USD)			0.06		
			(0.078)		
Immigrant visa proportion	-3.42	-0.19	-3.41	-5.20	-3.16
	(1.878)	(1.504)	(1.929)	(3.180)	(1.895)
Visa type proportion:					
Humanitarian	0.92	3.63	1.01	3.63	0.94
	(1.530)	(2.332)	(1.606)	(2.881)	(1.558)
Work	0.86	1.26	1.10	1.54	1.54
	(2.126)	(2.788)	(2.144)	(3.946)	(2.048)
Diversity	-3.08	0.91	-2.33	2.71	-3.62
	(1.878)	(3.221)	(2.065)	(4.921)	(1.879)
Study	2.07	4.00	1.99	1.92	2.31
	(2.245)	(3.996)	(2.451)	(5.372)	(2.441)
N	208836	208836	205349	208836	208836
Country FEs	No	Yes	No	No	No
Year FEs	Yes	Yes	Yes	Yes	Yes

All coefficients apart from the exponentiated time trend (the 5-year return rate at the top of the table) are from interaction terms with this exponentiated time trend. The first column contains the main specification, while (2) adds country fixed effects. (3) adds two more measures of distance, total trade/GDP in the origin and the natural log of the exchange rate in foreign currency. (4) uses the main specification, weighted by inflow in the arrival year. (5) replaces the relative GDP/capita term with 10 bins for 0-10%, 10-20%, etc. Standard errors in parentheses, clustered by country of origin. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

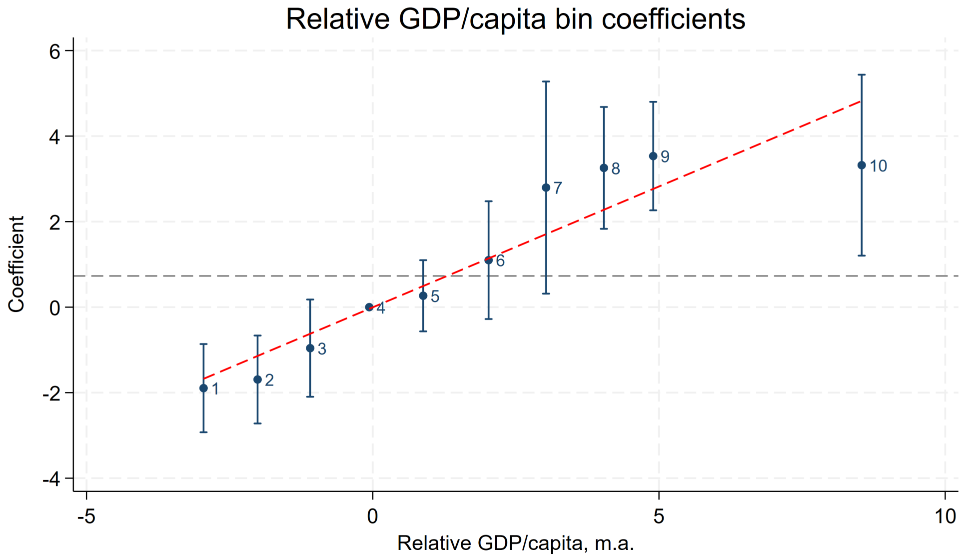

Figure 4: Relative GDP/capita coefficients, Column 5 specification

As migrants from poorer countries are more likely to remain in the US but also tend to have lower income, this is a form of selection that the literature normally calls positive selection of return migrants; those who are, on average, less successful in the US are more likely to remain. But our methodology cannot say anything about selective return migration for people from the same country. Ideally one would use longitudinal individual data to investigate this,as in Adda et al. (2022), Constant and Massey (2003), and Abramitzky et al. (2014). The existing evidence on this form of selection is somewhat conflicting; Constant and Massey (2003) finds negative selection of returning migrants from Germany using GSOEP data, while for the US Jasso and Rosenzweig (1988) finds that the high-skilled are less likely to naturalize and therefore more likely to leave, while Abramitzky et al. (2014) finds evidence of negative selection of return migrants from the US using data from the "Age of Migration" between 1850 and 1913. On the other hand, Reagan and Olsen (2000) finds little evidence of any emigrant selection at all in the US, and Dostie et al. (2023) finds the same lack of selection for Canada during the early 2000s. As Borjas and Bratsberg (1996) points out, different patterns of selection can be explained by a Roy-type model, with emigration patterns depending on whether the original in-migrants were positively or negatively selected–if they were positively selected, the "worst of the best" leave, and if negatively selected, the "best of the worst" (Constant and Massey, 2003) .

4.3 Further Robustness Checks

The Online Appendix contains a set of different robustness checks. Table B.1 of Appendix B shows that results about how the return rate varies with covariates are robust to using different values for duration dependence in return rates, λ; we show results for $\lambda = 0, 0.5, 1$. Table B.2 of Appendix B investigate the role of GDP/capita in more detail. First, we adjust the relative GDP/capita term to reflect the fact that immigrants from different countries have different levels of income when in the US (the last section of Appendix A shows that, on average, immigrants from poorer countries tend to have lower wages in the US than those from richer countries though there is considerable residual variation, and some countries' migrants do better than average in the US despite the origin being poorer than average). We replace the relative GDP/capita term in our main regression specification with $\sqrt{(\frac{GDPC_{ot}}{GDPC_{US,t}})(\frac{w_{US,t}}{w_{ot}})}$, where $GDPC_{ct}$ is GDP/capita in country c at time t, w_{ct} is the median wage of immigrants from origin o, and $w_{US,t}$ is the median wage of all US workers in year t. This term is the geometric mean of relative GDP/capita and the inverse of the relative median wage for immigrants from the given country once in the US; we use the ACS measure of wage income and compute the yearly median among those with positive income. Results from the regression with this term are very similar to those with the original relative GDP/capita term. We also investigate whether the growth rate of relative living standards matters for determining return propensities in addition to the level; we find income growth to be unimportant but the level of relative income remains important. The level of relative living standards seems to matter more than its growth rate in determining return migration rates.

Table B.2 also shows that weighting each observation in the sample average inflow from that country of origin over our sample period[11] makes little difference to the results.

(7) assumes that λ and β are the same for all groups. Appendix C presents separate estimates for sub-groups to see whether this is a reasonable approximation to the data, reporting estimates for men and women (Table C.1), by age group (Table C.2) and by region of origin (Tables C.3 and C.4). The estimated value of λ varies between 0.32 and 0.45, all suggesting marked negative duration dependence. There are some differences in the estimates of β (as would be expected even by chance) but, overall, the conclusions about the factors that influence return rates discussed below hold up for the sub-groups.

Appendix D present estimates from a different estimation approach that does not use a bootstrap. We use the whole sample, estimating $\widehat{\lambda}$ by choosing the value that maximizes the likelihood function in a grid search and then reporting coefficient and standard errors conditional on the maximized valued of

[11]The weight on the observation for group g after t years is calculated as $w_{gt} = \frac{\bar{n}_{g,0}}{\sum_{g'=1}^{G} \bar{n}_{g',0}}$, where $\bar{n}_{g,0}$ is the mean inflow from origin country o over the course of the sample.

λ (standard errors are computed by clustering at the country of birth level). Compared to the preferred bootstrap estimates this method does not produce a standard error for $\widehat{\lambda}$. The coefficient point estimates are very similar to the bootstrapped ones, while the non-bootstrapped standard errors are a bit smaller than the non-bootstrapped SEs as one would expect given they treat $\widehat{\lambda}$ as fixed, not estimated.

4.4 Education and return migration

We have shown that immigrants from poorer countries are more likely to remain in the US. As migrants from these countries have lower average levels of earnings, this is a form of negative selection. But there may also be important selection effects among migrants from the same country. This section investigates one aspect of this using education as an additional characteristic to define migrant groups. One difficulty with this is that education can increase over time for an individual, decreasing the number of immigrants measured in low education groups and increasing the number in the higher education group. Conditioning on education might causing us to detect a spurious negative impact of education on return rates. However, we also know that education rarely increases for older individuals, so one could perhaps use education to define groups if one restricts to those who were older on arrival. For this reason, we restrict our sample to arriving migrants in the top two age groups (25-44 and 45-64), to minimize the chance of educational upgrading contaminating our results, and estimate the full model with and without dummies for education level.

The results are presented in Table 4 below. Column 1 contains results from estimating the main specification (without the age group dummies for the excluded arrival groups); the estimates are similar in magnitude to those for the full sample. Column 2 includes dummies for "No high school", "Some college", and "College degree", with the excluded dummy being "High school", to assess the relationship between education and return rates. Migrants without high school degrees are most likely to return, followed by those who finished high school, and then those with some college. The coefficient for possessing a college degree is insignificant, but greater than zero, indicating a possible non-monotonicity in return propensity by education level. This is in line with other evidence on the impact of education on return rates, with low- and high-education migrants most likely to leave (OECD, 2008). It suggests that within countries the lowest educated (who are likely to have lower earnings) are less likely to remain; this is a form of positive selection of stayers to be set against the negative between-country selection described earlier.

Table 4: FE Poisson coefficient estimates

	(1)	(2)
Return rate after 5 years	2.26***	2.07**
	(0.650)	(0.740)
Exponent on time trend	0.42***	0.43***
	(0.036)	(0.034)
Female	-1.41*	-1.36*
	(0.673)	(0.681)
45-64	1.05***	1.05***
	(0.157)	(0.203)
Education level:		
No high school		0.60
		(0.349)
Some college		-1.56***
		(0.319)
College degree		0.63
		(0.428)
relative GDP/capita	0.72***	0.73***
	(0.115)	(0.118)
PIV Democracy score	0.06	0.07
	(0.077)	(0.080)
Conflict indicator	0.75	0.63
	(0.432)	(0.403)
Distance from origin (1,000 k.m.)	-0.01	-0.00
	(0.077)	(0.072)
Immigrant visa proportion	-3.88	-3.93
	(2.648)	(2.664)
Visa type proportion:		
Humanitarian	0.37	0.30
	(1.879)	(1.853)
Work	1.77	1.40
	(2.780)	(2.969)
Diversity	-2.96	-2.30
	(1.870)	(1.899)
Study	1.97	1.95
	(2.953)	(2.759)
N	346937	346937
Country FEs	No	No
Year FEs	Yes	Yes

All coefficients apart from the exponentiated time trend (the constant at the top of the table) are from interaction terms with this exponentiated time trend. In these specifications, we restrict the sample to migrants who arrived between the ages of 25 and 64, to test the impact of education on return propensities while minimizing the possibility that migrants change arrival group by obtaining further qualifications. (1) contains the main specification that we run on the full sample, while (2) adds the education variable to the right hand side. Standard errors in parentheses, clustered by country of origin. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

5 Conclusion

This paper has proposed a new method for estimating return migration rates using aggregated data, treating the number of migrants in a group who arrived in a particular year as an unobserved fixed effect, and the observed number (including, importantly, observed zeroes) in the arrival or subsequent years as observations from a Poisson distribution. Compared to existing methods, this allows us to estimate return rates for more granular migrant groups, allowing more in-depth analysis of the factors that influence return migration rates. We apply this method to US data and find a decreasing return hazard, with most returns occurring by eight years after arrival, when about 13% of migrants have left. It would be interesting to estimate similar models for Europe where retention rates seem to be much lower than in the US (OECD, 2008). The return rate is significantly lower for women, those who arrive at a young age, those from poorer countries, and is higher for migrants on non-immigrant visas for work or study. Our results also suggest the possibility of positive selection of remainers conditional on their country of origin. There are limitations to our approach: we are unable to say anything about how returns rates vary within our groups, e.g. whether it is those who are more or less successful within-groups who are more likely to leave the US (Dustmann and Weiss, 2007). For that question, longitudinal individual data is needed with clear information about when someone leaves the country. Because this type of data is rare, we think that by making use of more readily-available datasets, our approach can provide useful insights into the factors that influence return migration, an important yet under-researched aspect of the migration experience.

References

Abramitzky, R., L. Boustan, and K. Eriksson (2014): "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 122, 467 – 506.

Adda, J., C. Dustmann, and J.-S. Görlach (2022): "The dynamics of return migration, human capital accumulation, and wage assimilation," *The Review of Economic Studies*.

Ahmed, B. and J. G. Robinson (1994): "Estimates of Emigration of the Foreign-Born Population: 1980-1990," Working paper, U.S. Census Bureau Population Division.

Azose, J. J. and A. E. Raftery (2019): "Estimation of emigration, return migration, and transit migration between all pairs of countries," *Proceedings of the National Academy of Sciences*, 116, 116–122.

Bencek, D. and C. Schneiderheinze (2020): "Higher economic growth in poor countries, lower migration flows to the OECD: Revisiting the migration hump with panel data," Tech. rep.

Bhaskar, R., B. Arenas-Germosen, and C. Dick (2013): "Demographic Analysis 2010: Sensitivity Analysis of the Foreign-Born Migration Component," Working Paper 98, U.S. Census Bureau Population Division.

Bijwaard, G. E. and J. Wahba (2014): "Do high-income or low-income immigrants leave faster?" *Journal of Development Economics*, 108, 54–68.

Borjas, G. and B. Bratsberg (1996): "Who Leaves? The Outmigration of the Foreign-Born," *The Review of Economics and Statistics*, 78, 165–76.

Borjas, G. J. (1987): "Self-selection and the earnings of immigrants," *American Economic Review*, 531–553.

Borjas, G. J., I. Kauppinen, and P. Poutvaara (2019): "Self-selection of emigrants: Theory and evidence on stochastic dominance in observable and unobservable characteristics," *The Economic Journal*, 129, 143–171.

Carr, S. and M. Tienda (2013): "Family sponsorship and late-age immigration in aging America: Revised and expanded estimates of chained migration," *Population Research and Policy Review*, 32, 825–849.

Clemens, M. (2017): "Migration is a Form of Development: The Need for Innovation to Regulate Migration for Mutual Benefit," Tech. Rep. 2017/18, UN Population Division.

Clemens, M. A. (2014): "Does development reduce migration?" in *International Handbook on Migration and Economic Development*, Edward Elgar Publishing.

——— (2020): “The emigration life cycle: How development shapes emigration from poor countries,” .

Constant, A. and D. S. Massey (2003): “Self-selection, earnings, and out-migration: A longitudinal study of immigrants to Germany,” *Journal of population Economics*, 16, 631–653.

Djajic, S., M. G. Kirdar, and A. Vinogradova (2016): “Source-country earnings and emigration,” *Journal of International Economics*, 99, 46–67.

Docquier, F., G. Peri, and I. Ruyssen (2014): “The Cross-country Determinants of Potential and Actual Migration,” *International Migration Review*, 48, 37–99.

Dostie, B., J. Li, D. Card, and D. Parent (2023): “Employer policies and the immigrantânative earnings gap,” *Journal of Econometrics*, 233, 544–567.

Durand, J. and D. S. Massey (2019): “Evolution of the Mexico-U.S. Migration System: Insights from the Mexican Migration Project,” *The Annals of the American Academy of Political and Social Science*, 684, 21–42.

Dustmann, C. (2003): “Return Migration, Wage Differentials, and the Optimal Migration Duration,” *European Economic Review*, 47, 353–369.

Dustmann, C., I. Fadlon, and Y. Weiss (2011): “Return migration, human capital accumulation and the brain drain,” *Journal of Development Economics*, 58–67.

Dustmann, C. and J.-S. Goerlach (2016): “The Economics of Temporary Migrations,” *Journal of Economic Literature*, 54, 98–136.

Dustmann, C., H. Ku, T. Surovtseva, et al. (2021): “Real exchange rates and the earnings of immigrants,” .

Dustmann, C. and Y. Weiss (2007): “Return Migration: Theory and Empirical Evidence from the UK,” *British Journal of Industrial Relations*, 45, 236–256.

Elbers, C. and G. Ridder (1982): “True and spurious duration dependence: The identifiability of the proportional hazard model,” *The Review of Economic Studies*, 49, 403–409.

Gallardo-Sejas, H., S.-G. Pareja, R. Llorca-Vivero, and J. A. Martínez-Serrano (2006): “Determinants of European immigration: a cross-country analysis,” *Applied Economics Letters*, 13, 769–773.

Galor, O. and O. Stark (1990): “Migrants’ savings, the probability of return migration and migrants’ performance,” *International Economic Review*, 463–467.

Gundel, S. and H. Peters (2008): "What Determines the Duration of Stay of Immigrants in Germany?: Evidence from a Longitudinal Duration Analysis," SOEPpapers on Multidisciplinary Panel Data Research 79, DIW Berlin, The German Socio-Economic Panel (SOEP).

Hall, T., A. Manning, and M. Sumption (2023): "Why are the latest net migration figures not a reliable guide to future trends?" Tech. rep., Migration Observatory, University of Oxford.

Hatton, T. J. (2005): "Explaining trends in UK immigration," *Journal of Population Economics*, 18, 719–740.

Hendricks, L. and T. Schoellman (2018): "Human capital and development accounting: New evidence from wage gains at migration," *Quarterly Journal of Economics*, 133, 665–700.

Jasso, G. and M. 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, 279–290.

——— (1988): "How well do US immigrants do? Vintage effects, emigration selectivity, and occupational mobility," *Research in Population Economics*, 6, 229–253.

Kuhlenkasper, T. and M. F. Steinhardt (2017): "Who leaves and when? Selective outmigration of immigrants from Germany," *Economic Systems*, 41, 610–621.

LaLonde, R. and R. Topel (1993): "Economic impact of international migration and the economic performance of migrants," in *Handbook of Population and Family Economics*, ed. by M. R. Rosenzweig and O. Stark, Elsevier, vol. 1, Part B, chap. 14, 799–850, 1 ed.

Lancaster, T. (1979): "Econometric methods for the duration of unemployment," *Econometrica: Journal of the Econometric Society*, 939–956.

Lancaster, T. and S. Nickell (1980): "The analysis of re-employment probabilities for the unemployed," *Journal of the Royal Statistical Society: Series A (General)*, 143, 141–152.

Langella, M. and A. Manning (2021): "Income and the Desire to Migrate," *CEP Discussion Paper no. 1794*.

Lessem, R. (2018): "Mexico–US immigration: effects of wages and border enforcement," *The Review of Economic Studies*, 85, 2353–2388.

Lewer, J. J. and H. Van den Berg (2008): "A gravity model of immigration," *Economics Letters*, 99, 164–167.

Lubotsky, D. (2007): "Chutes or ladders? A longitudinal analysis of immigrant earnings," *Journal of Political Economy*, 115, 820–867.

Mayda, A. M. (2010): "International migration: A panel data analysis of the determinants of bilateral flows," *Journal of Population Economics*, 23, 1249–1274.

Nekby, L. (2006): "The emigration of immigrants, return vs onward migration: evidence from Sweden," *Journal of Population Economics*, 19, 197–226.

Nekoei, A. (2013): "Immigrants' labor supply and exchange rate volatility," *American Economic Journal: Applied Economics*, 5, 144–64.

OECD (2008): *International Migration Outlook 2008*, chap. Return Migration: A New Perspective.

Ortega, F. and G. Peri (2013): "The effect of income and immigration policies on international migration," *Migration Studies*, 1, 47–74.

Reagan, P. B. and R. J. Olsen (2000): "You can go home again: Evidence from longitudinal data," *Demography*, 37, 339–350.

Van Hook, J., W. Zhang, F. D Bean, and J. S Passel (2006): "Foreign-born emigration: A new approach and estimates based on matched CPS files," *Demography*, 43, 361–382.

Warren, R. and J. Peck (1980): "Foreign-Born Emigration From The United States: 1960 To 1970," *Demography*, 17, 71–84.

Yang, D. (2006): "Why do migrants return to poor countries? Evidence from Philippine migrants' responses to exchange rate shocks," *The Review of Economics and Statistics*, 88, 715–735.

——— (2008): "International migration, remittances and household investment: Evidence from Philippine migrantsâ exchange rate shocks," *The Economic Journal*, 118, 591–630.

Zakharenko, R. (2008): "Return migration: an empirical investigation," *Available at SSRN 1350375.*

Zelinsky, W. (1971): "The hypothesis of the mobility transition," *Geographical Review*, 219–249.

Appendix A: Data Description

In this appendix, we describe in detail the various data sources we use in our empirical analysis.

American Community Survey (ACS) and 2000 US Census

Data on the current number of migrants in each arrival cohort (starting with 1998 arrivals) that are still in the United States during the years 2001-2018 come from the American Community Survey (ACS), while data for the year 2000 comes from the 5% sample from the US Census of that year.[12] The ACS is a yearly survey of approximately .4% of the US population. We use the variable *bpld* (birthplace) to assign respondents to their country of origin, and *yrimmig* (year of arrival in the US) to assign them to an arrival cohort defined by arrival year. We construct the number of migrants in each arrival cohort-year cell by summing the weight variable *perwt* for each individual observation in that cell. We drop all respondents that do not report an arrival year or country of origin, as well as those in group quarters, which leaves us with 1.6 million foreign-born respondents in 2000, 110,000-120,000 per year between 2001 and 2004, and 325,000-405,000 per year up to 2018. Table 1 of the main text provides descriptive statistics about the demographics of our sample, split into four bins for their period of arrival in the US.

Democracy, conflict, and distance data

For data on the political institutions of the migrant's origin country, we use the annual Polity5 series from the Center for Systemic Peace.[13] For a given year, the Polity5 index measures separately the level of democracy and autocracy in a given state, or polity. Both scores go from 0 to 10, with both scores being based on the competitiveness and openness of elections, the constraints placed on the chief executive, and the regulation of and competitiveness of political participation by the population. The autocracy score is subtracted from the democracy score to create an index that ranges from -10 to 10, where -10 is the most autocratic, and 10 is the most democratic. The full sample gives Polity5 scores for every country from 1946 to 2018; we use the subset of the sample from 1998 to 2018.

For data on the prevalence of conflict in the country of origin, we use a dataset from the Uppsala Conflict Data Program.[14] The UCDP collects a wealth of data on organized violence and armed conflict every year, including the incidence of conflict, the number of casualties, and the level of violence involved in

[12]https://www.census.gov/programs-surveys/acs
[13]https://www.systemicpeace.org/inscrdata.html
[14]https://ucdp.uu.se/encyclopedia

each conflict incident. The main dataset runs from 1946 to 2022. We construct an indicator variable for the existence of any armed conflict involving the country in question for a given year.

We take our measure of distance between the US and the country of origin from data provided by the Centre d'Etudes Prospectives et d'Informations Internationales ('Centre for Prospective Studies and International Information', CEPII).[15] Their GeoDist dataset provides inter-city distances (in kilometers) for various city dyads, including between a country's capital city and Los Angeles, as well as the capital and New York City. For our measure of distance, we take the smaller of the distance from the capital to LA and to New York.

Entry visa data

The data on the distribution of entry visas by country of origin and arrival year comes from a combination of datasets from the Department of Homeland Security (DHS) and the Immigration and Naturalization Service (INS). The DHS provides data on the number of non-immigrant visas granted by issuing office (country of origin) in a given year, for each specific type of non-immigrant visa (e.g., H1B).[16] We categorize each visa according to whether they are a work, family, humanitarian non-refugee, study, or short-term visitor visa, using the State Department's Directory of Visa Categories.[17] We exclude from our non-immigrant visa count any short-term visas issued (because they are likely to leave the US within a year of entry).

For immigrant visas (with the prospect of eventual permanent settlement), we use the yearly Yearbook of Immigration Statistics put together by the Office of Homeland Security Statistics (OHSS), which contains statistics on approved Lawful Permanent Residents (LPR) visas by country of origin, split up into family, work, diversity, and humanitarian visa categories.[18] We also use the Yearbooks for refugee admission statistics.

To deal with countries that do not appear in a given year of the data (e.g. Serbia and Montenegro in the late 1990s), we divide the visas issued for the country that contains the non-appearing countries (here, Yugoslavia) proportionally to their respective populations in 2000. Finally, we construct the proportion of each type of visa issued (exclusive categories for immigrant, non-immigrant, and refugee visas, and another separate set of exclusive categories for study, work, family, humanitarian, diversity, and refugee visas) by summing the relevant visa types, and dividing by the total visas issued for that country for that year, excluding short-term visitor visas.

[15] http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=6
[16] https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics.html
[17] https://travel.state.gov/content/travel/en/us-visas/visa-information-resources/all-visa-categories.html
[18] https://www.dhs.gov/ohss/topics/immigration/yearbook

Other data

We use the World Bank's open database for data on PPP GDP per capita in the country of origin.[19] Yearly data is available for each country of origin in our sample. In our empirical analysis, we construct GDP/capita of the origin country relative to the United States by dividing the former by the latter. We also use the World Bank's exchange rate data, which measures the average official price of the US dollar in local currency for each year.

Our trade flows data comes from the US Census Bureau, which measures bilateral imports and exports between the US and each of its trading partners, for every calendar year.[20] We index these flows to 2010 using the US CPI measure of inflation, and combine these yearly data with World Bank GDP per capita data (in 2010 constant USD) to construct each country of origin's trade volume with the US as a percentage of their GDP/capita in that year.

Origin GDP/capita and income of US immigrants

This section contains two graphs that illustrate the stylized fact that immigrants from countries with relatively lower living standards are also lower in the US income distribution. Figure A.1 plots the median wage income for immigrants from each of the 10 relative GDP/capita (PPP-adjusted) bins we use in some of our regressions in the main text, with the first bin containing immigrants from countries with 0-10% of US GDP/capita, the second containing those from countries with 10-20%, etc. The results are displayed separately by year to avoid issues with inflation adjustment. In all years there is a positive relationship between relative GDP/capita in the origin and the median wage for those immigrants in the US. Figure A.2 plots median wage income in the US versus the (PPP-adjusted) level of GDP/capita in the origin for each country of origin, again by year. Both figures suggest a correlation between the average living standard in the origin country and the median wage of immigrants from that country in the US–though obviously we are not controlling for positive or negative selection of migrants into the US.

[19] https://data.worldbank.org/

[20] https://www.census.gov/foreign-trade/balance/index.html

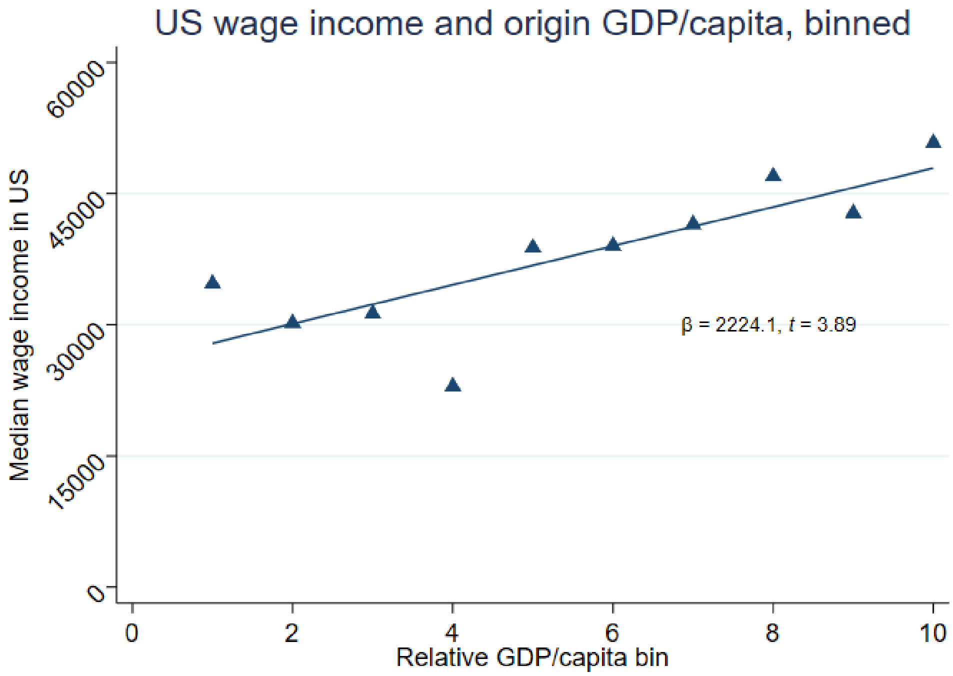

Figure A.1: US wage income and relative GDP/capita in origin, binned

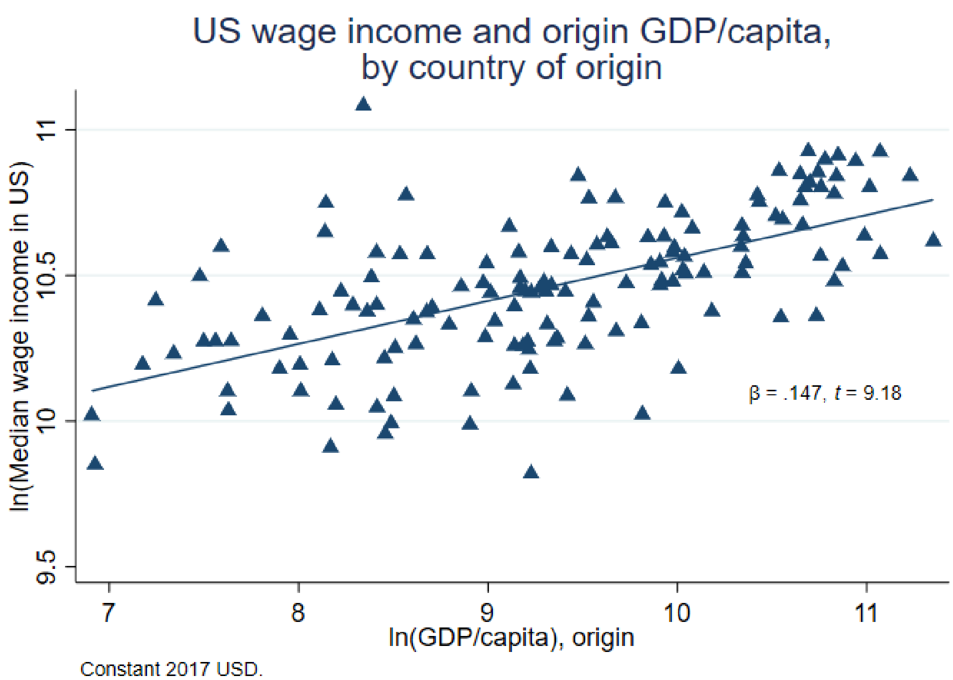

Figure A.2: US wage income and relative GDP/capita in origin, by country

Appendix B: Robustness of main specification

This appendix contains estimation results from various alternative specifications discussed in the main text, in order to assess the robustness of our main results and explore alternative parameterisations of our return rate function. Table B.1 contains estimates from the main specification with our standard set of covariates, with the exponent set to the likelihood-maximizing value (.40) in Column 1, and set to 0 (making the function of time ln[years since arrival]), .5, and 1 in Columns 2 - 4, respectively. It is reassuring that although the exponent is precisely estimated in the main specification, the estimated coefficients and return propensity orderings are not sensitive to the value of the exponent: younger arrivals are more likely to stay in the US than immigrants in older arrival groups, as are female arrivals and arrivals from relatively poorer countries. Conflict in the origin is positively associated with returning, while the level of democracy in the origin and the distance between the origin and the US are insignificant and close to 0 in all specifications.

In Table B.2, we control for the modified relative GDP/capita term discussed in the main text (Column 1), the growth rate of relative living standards with (Column 2) and without (Column 3) a control for the level. Results in the first column are very similar to those from the specification with the original relative GDP/capita term; the latter term's coefficient increases to .82 and remains significant, while the magnitudes and signs of the other coefficients are mostly unchanged. From the last two columns, it is apparent that the impact of the growth rate on the return propensity is very small, with an estimated coefficient of -0.11. Adding a control for the level moves this coefficient back towards zero, equal to a now-insignificant -0.03. Relative to the impact of ln(relative GDP/capita) ($\hat{\beta} = 0.74$) and to the magnitudes of the other coefficients in the model, the growth rate of relative living standards has a negligible association with a migrant's probability of returning to their country of origin.

Table B.1: FE Poisson coefficient estimates

	(1)	(2)	(3)	(4)
Return rate after 5 years	2.13***	2.61**	2.50***	3.51***
	(0.526)	(0.795)	(0.563)	(0.918)
Exponent on time trend	0.40***	0.00***	0.50***	1.00***
	(0.029)	(.)	(.)	(.)
Female	-1.07*	-1.22*	-1.15*	-1.13*
	(0.504)	(0.575)	(0.551)	(0.470)
Age at arrival:				
0-15	-0.88*	-1.09*	-0.96*	-0.99**
	(0.355)	(0.471)	(0.416)	(0.361)
16-24	-0.40	-0.33	-0.40	-0.35
	(0.292)	(0.214)	(0.311)	(0.227)
45-64	0.98***	-0.02	0.95***	0.49**
	(0.151)	(0.192)	(0.161)	(0.153)
relative GDP/capita	0.73***	0.76***	0.82***	0.80***
	(0.098)	(0.223)	(0.098)	(0.105)
PIV Democracy score	0.08	0.09	0.09	0.08
	(0.064)	(0.090)	(0.071)	(0.080)
Conflict indicator	0.50	0.70	0.62	0.72
	(0.348)	(0.461)	(0.375)	(0.395)
Distance from origin (1,000 k.m.)	0.03	0.10	0.04	0.04
	(0.057)	(0.068)	(0.066)	(0.066)
Immigrant visa proportion	-3.42	-6.03	-3.98*	-3.87
	(1.878)	(3.747)	(2.024)	(2.104)
Visa type proportion:				
Humanitarian	0.92	-0.96	0.92	0.95
	(1.530)	(2.472)	(1.669)	(1.791)
Work	0.86	-2.82	0.38	-0.26
	(2.126)	(3.941)	(2.392)	(2.570)
Diversity	-3.08	-2.82	-3.00	-2.11
	(1.878)	(2.101)	(1.969)	(1.958)
Study	2.07	0.42	1.84	1.36
	(2.245)	(3.616)	(2.735)	(2.688)
N	208836	191643	208836	208836
Country FEs	No	No	No	No
Year FEs	Yes	Yes	Yes	Yes

All coefficients apart from the exponentiated time trend (the 5-year return rate at the top of the table) are from interaction terms with this exponentiated time trend. Column (1) contains the results from the main specification found in the main text for comparison. (2) estimates the main specification with the exponent equal in the limit to 0, which implies the function of time is ln(years since arrival). (3) and and (4) estimate the main specification with the exponent equal to .5 and 1, respectively. Standard errors, clustered by country of origin, in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table B.2: FE Poisson coefficient estimates

	(1)	(2)	(3)
Return rate after 5 years	2.05***	2.89***	3.30***
	(0.4639)	(0.6439)	(0.5425)
Exponent on time trent	0.40***	0.40***	0.40***
	(0.0290)	(0.0000)	(0.0000)
Female	-1.07*	-1.35**	-1.41**
	(0.5111)	(0.5032)	(0.4978)
Age at arrival:			
0-15	-0.86*	-1.37**	-1.47***
	(0.3760)	(0.4308)	(0.4149)
16-24	-0.43	-0.57**	-0.46*
	(0.2938)	(0.2108)	(0.2124)
45-64	0.96***	-0.11	-0.15
	(0.1650)	(0.2437)	(0.2264)
Growth rate of relative GDP/capita		-0.11*	-0.03
		(0.0502)	(0.0411)
relative GDP/capita			0.74***
			(0.1282)
adjusted relative GDP/capita	0.82***		
	(0.1573)		
PIV Democracy score	0.09	0.10	0.10
	(0.0691)	(0.0945)	(0.0767)
Conflict indicator	-0.21	-0.12	0.63
	(0.3202)	(0.6768)	(0.3575)
Distance from origin (1,000 k.m.)	0.12	-0.06	0.08
	(0.0739)	(0.1240)	(0.0702)
Immigrant visa proportion	-2.79	-7.45	-6.21
	(3.2288)	(4.7521)	(3.2854)
Visa type proportion:			
Humanitarian	3.06	0.71	-0.32
	(2.1321)	(3.5045)	(2.2917)
Work	3.72	1.70	-2.03
	(3.4276)	(5.4859)	(3.4426)
Diversity	-1.42	-3.49	-2.77
	(1.7817)	(2.7774)	(2.2528)
Study	3.78	3.39	0.13
	(3.2818)	(5.0962)	(3.3843)
N	200787	189705	189705
Country FEs	No	No	No
Year FEs	Yes	Yes	Yes

All coefficients apart from the exponentiated time trend (the 5-year return rate at the top of the table) are from interaction terms with this exponentiated time trend. Column (1) replaces the relative GDP/capita term with the geometric mean of relative GDP/capita and the inverse of the relative median wage in the US. Column (2) contains estimates from the main specification with the relative GDP/capita term replaced by its growth rate, while (3) re-adds the level as an additional control. Standard errors, clustered by country of origin, in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Appendix C: Subgroup return rates

This appendix presents the full results for regressions that restrict the sample to subgroups of interest, denoted by the column label. Estimating the model separately by group does not drastically change our conclusions about the ordering of return propensities: younger arrivals are more likely to stay, as are females and arrivals from relatively poorer countries; as in our full sample results, the impacts of distance and democracy on return propensities are very small. There is some variation in the estimated coefficients for different types of visa; the impact of coming to the US on a humanitarian or diversity visa, for example, changes depending on the subsample. However, the impact of coming on an immigrant category visa on the return rate is always negative. There is also some variation in the estimated exponent on the time trend, with estimates ranging between .32 and .45, all indicative of negative duration dependence.

Table C.1: FE Poisson coefficient estimates by sex

	Male	Female
Return rate after 5 years	2.29***	0.88***
	(0.5999)	(0.2385)
Exponent on time trend	0.43***	0.36***
	(0.0361)	(0.0290)
Age at arrival:		
0-15	-1.90**	0.14
	(0.6755)	(0.1969)
16-24	-0.68	-0.11
	(0.3579)	(0.2509)
45-64	0.28	1.60***
	(0.1962)	(0.2211)
relative GDP/capita	0.76***	0.67***
	(0.1135)	(0.1184)
PIV Democracy score	0.06	0.10
	(0.0810)	(0.0608)
Conflict indicator	0.74	0.23
	(0.5522)	(0.2893)
Distance from origin (1,000 k.m.)	-0.04	0.09
	(0.1007)	(0.0522)
Immigrant visa share	-4.10	-3.16
	(2.3379)	(2.0879)
Visa type proportion:		
Humanitarian	0.71	0.85
	(1.8624)	(1.5678)
Work	0.21	0.93
	(2.9099)	(2.0178)
Diversity	-5.23**	-0.64
	(1.9847)	(1.8096)
Study	1.55	2.31
	(2.5707)	(2.5400)
N	104257	104579
Country FEs	No	No
Year FEs	Yes	Yes

All coefficients apart from the exponentiated time trend (the 5-year return rate at the top of the table) are from interaction terms with this exponentiated time trend. Each column contains results from the main specification, with the sample restricted to the sex indicated. Standard errors, clustered by country of origin, in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table C.2: FE Poisson coefficient estimates by age at arrival

	0-15	16-24	25-44	45-64
Return rate after 5 years	1.09***	1.69***	2.36**	2.82***
	(0.2965)	(0.4864)	(0.7306)	(0.8499)
Exponent on time trend	0.44***	0.37***	0.45***	0.32***
	(0.0462)	(0.0374)	(0.0343)	(0.0582)
Female	0.07	-1.35	-1.79*	-0.19
	(0.1263)	(0.7518)	(0.7074)	(0.5687)
relative GDP/capita	0.52***	1.00***	0.78***	0.48**
	(0.1273)	(0.1480)	(0.1220)	(0.1844)
PIV Democracy score	0.12	0.12	0.04	0.10
	(0.0717)	(0.0852)	(0.0866)	(0.0837)
Conflict indicator	0.01	0.32	0.79	0.60
	(0.4233)	(0.3798)	(0.4205)	(0.4021)
Distance from origin (1,000 k.m.)	0.13	0.04	0.01	-0.04
	(0.0743)	(0.0723)	(0.0779)	(0.0809)
Immigrant visa share	-2.67	-4.71*	-4.52	-2.19
	(2.6651)	(1.9325)	(2.6206)	(3.0987)
Visa type proportion:				
Humanitarian	1.62	0.95	0.23	-0.02
	(1.8027)	(1.7966)	(2.0072)	(2.0881)
Work	-0.50	-0.58	1.17	2.51
	(2.7984)	(2.4523)	(2.8853)	(3.0840)
Diversity	-4.19	-2.34	-3.98	-0.22
	(2.5397)	(2.4345)	(2.1059)	(2.5644)
Study	-0.54	2.21	1.16	3.96
	(3.2273)	(2.9676)	(2.9581)	(3.7594)
N	52887	53215	53524	49210
Country FEs	No	No	No	No
Year FEs	Yes	Yes	Yes	Yes

All coefficients apart from the exponentiated time trend (the 5-year return rate at the top of the table) are from interaction terms with this exponentiated time trend. Each column contains results from the main specification, with the sample restricted to the arrival age group indicated. Standard errors, clustered by country of origin, in parentheses. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

Table C.3: FE Poisson coefficient estimates by region of origin

	North/Central America	Caribbean	South America
Return rate after 5 years	-9.28	-46.46	0.32
	(35.1106)	(135.4008)	(4.0090)
Exponent on time trend	0.42**	0.37	0.40***
	(0.1288)	(0.2229)	(0.1001)
Female	-2.09***	-0.33*	-0.67***
	(0.6289)	(0.1408)	(0.1839)
Age at arrival:			
0-15	-1.43*	1.25	0.05
	(0.5536)	(0.6588)	(0.2697)
16-24	-0.96**	0.39	-0.06
	(0.3043)	(0.5022)	(0.1602)
45-64	1.78**	0.72*	1.02**
	(0.5693)	(0.3281)	(0.3569)
relative GDP/capita	-0.09	1.10	0.34
	(1.4911)	(4.8028)	(1.3588)
PIV Democracy score	0.26	1.14	-0.05
	(1.0295)	(3.6284)	(0.7356)
Conflict indicator	-1.03	3.48	0.35
	(3.8469)	(43.2180)	(2.4245)
Distance from origin (1,000 k.m.)	-2.20	10.99	0.14
	(7.1437)	(89.1860)	(1.0384)
Immigrant visa share	0.57	38.98	-7.52
	(5.1959)	(135.7181)	(7.3532)
Visa type proportion:			
Humanitarian	-13.71	13.73	12.46
	(28.1144)	(131.7627)	(26.5074)
Work	4.48	18.64	-1.47
	(23.6718)	(100.0036)	(5.0570)
Diversity	-17.87	216.58	-69.11
	(221.8056)	(386.0690)	(58.9870)
Study	37.18	-21.01	-4.44
	(21.0565)	(34.9861)	(10.3281)
N	14521	7016	17981
Country FEs	No	No	No
Year FEs	Yes	Yes	Yes

All coefficients apart from the exponentiated time trend (the 5-year return rate at the top of the table) are from interaction terms with this exponentiated time trend. Each column contains results from the main specification, with the sample restricted to the region of origin indicated. Standard errors, clustered by country of origin, in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table C.4: FE Poisson coefficient estimates by region of origin (cont.)

	Europe	Asia	Illegal immigration sample
Return rate after 5 years	0.76	2.34	11.30
	(0.8256)	(2.5612)	(9.9021)
Exponent on time trend	0.38***	0.42***	0.39***
	(0.0237)	(0.1064)	(0.0450)
Female	-0.10	-0.49	-1.42
	(0.1622)	(0.4103)	(0.7751)
Age at arrival:			
0-15	-1.28*	-0.59	-1.05
	(0.5288)	(0.5510)	(0.5923)
16-24	1.45*	-0.45	-1.01***
	(0.6304)	(0.5524)	(0.2227)
45-64	0.44	0.79**	1.25***
	(0.4348)	(0.2815)	(0.2385)
relative GDP/capita	0.86***	0.74	1.65*
	(0.2140)	(0.4989)	(0.8121)
PIV Democracy score	0.14	0.12	0.25
	(0.1601)	(0.1816)	(0.9853)
Conflict indicator	0.36	-0.35	-0.51
	(0.9518)	(1.0525)	(0.8368)
Distance from origin (1,000 k.m.)	0.10	-0.03	0.22
	(0.4312)	(0.4399)	(0.2511)
Immigrant visa share	-1.93	-4.25	-3.07
	(5.0110)	(6.1436)	(2.6738)
Visa type proportion:			
Humanitarian	5.52	-1.83	9.86
	(4.7907)	(3.7318)	(8.3766)
Work	3.50	-0.81	0.31
	(6.8957)	(4.2725)	(3.1869)
Diversity	2.82	-3.69	136.92
	(4.2563)	(8.8547)	(215.2964)
Study	5.11	-0.35	6.24
	(7.5797)	(8.1498)	(7.3999)
N	71474	35995	16344
Country FEs	No	No	No
Year FEs	Yes	Yes	Yes

All coefficients apart from the exponentiated time trend (the 5-year return rate at the top of the table) are from interaction terms with this exponentiated time trend. Each column contains results from the main specification, with the sample restricted to the arrival age group indicated. Illegal immigration sample origin countries are Mexico, El Salvador, Guatemala, India, Honduras, China, the Philippines, Colombia, Brazil, and Venezuela. Standard errors, clustered by country of origin, in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Appendix D: Non-Bootstrapped Results

This appendix contains two tables that present the results from the same specifications as Tables 3 and 4 of the main text, but without bootstrapping the coefficient or SE estimates. The coefficients are extremely close to the bootstrapped ones, but the SEs are slightly smaller. This is to be expected, because the non-bootstrapped SEs do not account for error clustering by country of origin, and because the non-bootstrap estimation procedure treats $\widehat{\lambda}$ as fixed when estimating the other parameters, when in fact it is itself an estimated parameter, leading to an understatement of the other standard errors.

Table D.1: FE Poisson coefficient estimates

	(1)	(2)	(3)	(4)	(5)
Return rate after 5 years	2.31***	3.38***	2.27***	3.50***	2.60***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Exponent on time trend	0.41***	0.46***	0.41***	0.41***	0.42***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-1.18*	-1.29*	-1.19*	-2.22***	-1.20*
	(0.518)	(0.526)	(0.520)	(0.453)	(0.522)
Age at arrival:					
0-15	-0.96**	-1.01**	-0.96**	-1.70***	-0.98**
	(0.348)	(0.370)	(0.348)	(0.186)	(0.351)
16-24	-0.44	-0.45	-0.45	-1.04***	-0.45
	(0.277)	(0.289)	(0.277)	(0.122)	(0.280)
45-64	0.96***	0.97***	0.97***	1.05***	0.97***
	(0.152)	(0.149)	(0.150)	(0.098)	(0.147)
relative GDP/capita	0.75***	0.98***	0.74***	0.73***	
	(0.071)	(0.277)	(0.073)	(0.119)	
PIV Democracy score	0.06	0.07	0.06	0.14	0.03
	(0.051)	(0.079)	(0.049)	(0.090)	(0.039)
Conflict indicator	0.62**	-0.74	0.50*	0.37	0.28
	(0.230)	(0.393)	(0.234)	(0.523)	(0.240)
Distance from origin (1,000 k.m.)	0.01		0.03	-0.04	0.02
	(0.038)		(0.046)	(0.048)	(0.033)
US Trade/GDP in origin			1.33		
			(1.281)		
ln(exchange rate, 1/USD)			0.04		
			(0.058)		
Immigrant visa proportion	-3.70*	-0.12	-3.53*	-2.78	-3.49*
	(1.702)	(1.499)	(1.600)	(1.996)	(1.581)
Visa type proportion:					
Humanitarian	0.79	4.41*	1.14	4.45	0.95
	(1.371)	(2.064)	(1.383)	(2.475)	(1.342)
Work	0.69	2.28	1.37	1.10	1.40
	(1.963)	(2.103)	(1.774)	(1.636)	(1.639)
Diversity	-3.18*	2.34	-2.27	-5.94***	-3.68*
	(1.544)	(2.759)	(1.650)	(1.591)	(1.640)
Study	1.89	5.39	2.33	4.84	2.20
	(1.887)	(3.705)	(1.921)	(3.186)	(1.902)
N	208836	208836	205349	208836	208836
Country FEs	No	Yes	No	No	No
Year FEs	Yes	Yes	Yes	Yes	Yes

This table presents the non-bootstrapped version of the estimates in Table 3 of the main text. All coefficients apart from the exponentiated time trend (the 5-year return rate at the top of the table) are from interaction terms with this exponentiated time trend. The first column contains the main specification, while (2) adds country fixed effects. (3) adds two more measures of distance, total trade/GDP in the origin and the natural log of the exchange rate in foreign currency. (4) uses the main specification, weighted by inflow in the arrival year. (5) replaces the relative GDP/capita term with 10 bins for 0-10%, 10-20%, etc. Standard errors in parentheses, clustered by country of origin. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table D.2: FE Poisson coefficient estimates

	(1)	(2)
Return rate after 5 years	2.47***	2.28***
	(0.000)	(0.000)
Exponent on time trend	0.43***	0.43***
	(0.000)	(0.000)
Female	-1.57*	-1.52*
	(0.681)	(0.696)
45-64	1.02***	1.01***
	(0.148)	(0.178)
Education level:		
No high school		0.72***
		(0.218)
Some college		-1.58***
		(0.302)
College degree		0.61
		(0.436)
relative GDP/capita	0.74***	0.75***
	(0.082)	(0.083)
PIV Democracy score	0.03	0.04
	(0.064)	(0.063)
Conflict indicator	0.97***	0.84***
	(0.259)	(0.254)
Distance from origin (1,000 k.m.)	-0.03	-0.03
	(0.049)	(0.053)
Immigrant visa proportion	-4.24	-4.11
	(2.180)	(2.228)
Visa type proportion:		
Humanitarian	0.03	0.17
	(1.651)	(1.665)
Work	1.51	1.42
	(2.514)	(2.651)
Diversity	-3.33*	-2.51
	(1.673)	(1.623)
Study	1.56	1.67
	(2.290)	(2.218)
N	346937	346937
Country FEs	No	No
Year FEs	Yes	Yes

This table presents the non-bootstrapped version of the estimates in Table 3 of the main text. All coefficients apart from the exponentiated time trend (the constant at the top of the table) are from interaction terms with this exponentiated time trend. In these specifications, we restrict the sample to migrants who arrived between the ages of 25 and 64, to test the impact of education on return propensities while minimizing the possibility that migrants change arrival group by obtaining further qualifications. (1) contains the main specification that we run on the full sample, while (2) adds the education variable to the right hand side. Standard errors in parentheses, clustered by country of origin. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1979	Rosa Sanchis-Guarner José Montalbán Felix Weinhardt	Home broadband and human capital formation
1978	Luca Macedoni John Morrow Vladimir Tyazhelnikov	Firms in product space: Adoption, growth and competition
1977	Gonzalo Nunez-Chaim Henry G. Overman Capucine Riom	Does subsidising business advice improve firm performance? Evidence from a large RCT
1976	Julian Alves Bruno Serra Jason Greenberg Yaxin Guo Ravija Harjai John Van Reenen	Labour market power: New evidence on Non-Compete Agreements and the effects of M&A in the UK
1975	Virginia Minni	Global managers, local workers: Wage setting inside a multinational firm
1974	Jonathan Colmer Suvy Qin John Voorheis Reed Walker	The changing nature of pollution, income and environmental inequality in the United States
1973	Italo Colantone Gianmarco I.P. Ottaviano Piero Stanig	The social footprint of globalization: Towards the introduction of strategic industries in quantitative trade models
1972	Viet Nguyen-Tien Robert J.R. Elliott Eric Strobl Chengyu Zhang	Estimating the longevity of electric vehicles: What do 300 million MOT test results tell us?
1971	Michael Amior Jan Stuhler	Immigration, monopsony and the distribution of firm pay

1970	Jonathan Colmer David Lagakos Martin Shu	Is the electricity sector a weak link in development?
1969	Gaia Dossi Marta Morando	Political ideology and innovation
1968	Natalie Irmert Jan Bietenbeck Linn Mattisson Felix Weinhardt	Autonomous schools, achievement and segregation
1967	Stephen B. Billings Adam Soliman	The erosion of homeownership and minority wealth
1966	Pawel Bukowski Pawel Chrostek Filip Novokmet Marek Skawiński	Income inequality in the 21st century Poland
1965	Benny Kleinman Ernest Liu Stephen J. Redding Motohiro Yogo	Neoclassical growth in an interdependent world
1964	Hanwei Huang Gianmarco I.P. Ottaviano	Rethinking revealed comparative advantage with micro and macro data
1963	Natalie Chen Dennis Novy Carlo Perroni Horng Chern Wong	Urban-biased structural change
1962	Robin Kaiji Gong Yao Amber Li Kalina Manova Stephen Teng Sun	Tickets to the global market: First US patent awards and Chinese firm exports
1961	Antonin Bergeaud Arthur Guillouzouic	Proximity of firms to scientific production

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE