The Emigration Life Cycle: How Development Shapes Emigration from Poor Countries

Michael A. Clemens

DISCUSSION PAPER SERIES

The Emigration Life Cycle: How Development Shapes Emigration from Poor Countries

Michael A. Clemens
Center for Global Development and IZA

AUGUST 2020

ABSTRACT

The Emigration Life Cycle: How Development Shapes Emigration from Poor Countries[*]

Many governments seek to reduce emigration from low-income countries by encouraging economic development there. A large literature, however, observes that average emigration rates are higher in countries with sustained increases in GDP per capita than in either chronically poor countries or established rich countries. This suggests an emigration life cycle in which average emigration first rises, then falls with development. But this hypothesis has not been tested with global datasets controlling for unobserved heterogeneity between countries.

This paper finds that emigration rises on average as GDP per capita initially rises in poor countries, slowing after roughly US$5,000 at purchasing power parity, and reversing after roughly $10,000. Before this reversal, the within-country elasticity of rising emigration prevalence to rising GDP per capita is +0.35 to all destinations, and +0.74 to rich destinations. This relationship between emigration flows and economic growth is highly robust to country and time effects (fixed or random), specification (linear, log, nonparametric), emigration measure (stock or flow), country subsamples (rich destinations, large origins), and historical period (1960–2019 or 1850–1914). Decomposition of channels for this relationship highlight the joint importance of demographic transition, education investment, and structural change, but question a large role for transportation costs or policy barriers.

Corresponding author:
Michael A. Clemens
Center for Global Development
2055 L Street NW
Washington, DC 20036
USA
E-mail: mclemens@cgdev.org

In fundamental economic models of emigration, workers choose to live abroad if their earnings would rise enough to exceed the costs (Sjaastad 1962). This carries an intuitive, testable implication: When earnings rise in their country of origin, more workers should choose to live there rather than live abroad. A large literature has suggested, however, that the effects of economic development on emigration are more complex—that rising incomes coincide with structural changes that can initially cause more workers to emigrate. Thus typical poor countries exhibit an *emigration life cycle* (Hatton and Williamson 1998) in which early stages of economic development *raise* emigration rates. That research has new relevance in an age when many governments encourage economic development in poor countries with the explicit goal of *reducing* emigration (e.g. Caselli 2019).

But despite recent major advances in the availability of long-term datasets on migration patterns, those datasets have not been fully investigated to test the hypothesis of the emigration life cycle. Across developing countries at a moment in time, relatively richer developing countries have much higher emigration rates than the poorest developing countries in cross section (e.g. Clemens 2014; Dao et al. 2018; Idu 2019). Yet it has remained unclear if these correlations accurately represent the paths that developing countries take over time (Lucas 2019, 18), whether they are driven by migration to relatively rich countries or to other relatively poor countries, and whether they are driven by very small migrant-origin countries such as island nations with high emigration rates (Bade and de Kemp 2018, 43).

This paper estimates and decomposes the relationship between real gross domestic product (GDP) per capita and net emigration rates in developing countries. The results confirm the existence of the emigration life cycle, not only across but within typical developing countries. As poor countries have developed toward US$10,000 per capita at purchasing power parity (PPP) from 1960 to the present, controlling for country fixed effects, a 100 percent increase in real GDP per capita has been associated with a 35 percent rise in emigration from developing countries to all destination countries, and a 74 percent rise in emigration to high-income destination countries. The magnitude of this relationship is about one-third smaller than in cross-section. The rise in emigration slows on average between a GDP per capita PPP$5,000 and PPP$10,000, and reverses thereafter. This has been the average experience of economic development rather than a universal law.

This result is highly robust. The positive sign on the income-emigration relationship for relatively poor countries holds under country and time effects; under fixed or random effects; under linear, log, or nonparametric regression; under measurement of emigration as stocks (prevalence) or flows (incidence); and under restriction to rich destination countries or large origin countries. The analysis includes an illustration of how incorrect regression specifications can, with the same data, yield spurious estimates. Overall, the emigration life cycle is a strong and typical feature of economic development within countries over time.[1]

The paper also examines selected channels for this reduced-form relationship. It does this by decomposing the life cycle both within and between two very different historical eras. It compares in a single empirical framework all emigration worldwide 1960–2019, and emigration to the Western Hemisphere 1850–1914. The latter setting exhibits few policy barriers and much higher transportation costs. The literature posits several channels by which the life cycle could arise (e.g. Hatton and Williamson 1998; Dao et al. 2018). Rising income typically raises education and urbanization, and the associated demographic transition can produce large cohorts of youths seeking work, all of which can encourage emigration. Other channels could arise abroad: Richer potential emigrants might be better able to finance international travel, or to overcome migration policy barriers erected overseas. The evidence from both eras is consistent with a complex mix of channels related to demographic transition, rising education, and structural change—to broadly comparable degrees. It is not consistent with theories positing a large explanatory role for transportation costs and policy barriers.

The economic history literature first documented that poor countries exhibited such an "emigration life cycle" as they developed in the 19th and early 20th centuries (Williamson 1974, 371; Akerman 1976, 25–32; Gould 1979; Hatton and Williamson 1994a, 1994b, 1994c, 1998, 2005a; Hatton 2001, 2010; Faini and Venturini 1994; Chiswick and Hatton 2003; O'Rourke 2009). A more recent strand of work has found that developing countries in the late 20th and early 21st centuries have exhibited the same pattern: emigration rates are typically much higher in middle-income developing countries than in low-income developing countries (Massey 1988; Skeldon

---

[1]The term "emigration life cycle" is closely related but not identical to other terms in the literature. The "mobility transition" (Zelinsky 1971) originated in the geography literature to describe long-term generational associations between rural development and rural-urban migration. The term "migration hump" was coined to describe greater emigration flows from Mexico associated with reduced barriers to trade and investment (Martin 1993).

1997; Hatton and Williamson 2002, 2003, 2005b, 2011; Clark et al. 2004; de Haas 2005, 2007; Faini and Venturini 2010; Clemens 2014; Williamson 2015; Ferrie and Hatton 2015; Djajic et al. 2016; Dao et al. 2018; Ruhe et al. 2020). In early stages of development, origin-country income per capita correlates positively with emigration rates to a single destination country—the United States (Clark et al. 2007; Hatton and Williamson 2011) or Germany (Vogler and Rotte 2000)—and from a single origin country, Mexico (Hanson and McIntosh 2010). Studies of migration to OECD countries collectively have found this relationship only among low-income countries (Pedersen et al. 2008; Mayda 2010)—incidental to a different research focus and without exploring the result, which is contradicted by some linear regression specifications (Ortega and Peri 2013).

A different strand of work explores the relationship between income and emigration at the household level. It finds that individual families with exogenously higher income exhibit greater tendency to emigrate, from Mexico (Orrenius and Zavodny 2005; McKenzie and Rapoport 2007; Angelucci 2015; Görlach 2019), rural Bangladesh (Bryan, Chowdhury and Mobarak 2014), Indonesia (Bazzi 2017), Comoros (Gazeaud et al. 2019), and Honduras (Millán et al. 2020)—a literature surveyed by Adhikari and Gentilini (2018). Clemens and Mendola (2020) find the same pattern using observational differences in income across households in 99 developing countries.

The contribution of this work is to provide transparent, comprehensive, and global country-level empirical tests of a long-standing hypothesis in the development, labor, and economic history literatures, controlling for unobserved heterogeneity using more complete and recently available data (Özden et al. 2011; United Nations 2019). The present work is unique in its global scope, its array of robustness tests, its analysis of both modern and historical migration in the same empirical framework, and its decomposition of the within-country relationship. It reconciles discordant results in the prior literature by highlighting errors of empirical method.

The paper begins with a discussion of basic theory and the data to test it. It proceeds to an analysis of the data from 1960 to present. It then explores selected channels for the emigration life cycle, first by comparing post-1960 estimates to pre-1914 estimates, and then by decomposing the relationship within each of those eras.

1 Economic development and the decision to emigrate

Why might the relationship between income and emigration start out positive, then reverse as incomes rise? Perhaps the most parsimonious model that can generate the life-cycle pattern is a special case of the model of emigration scale due to Grogger and Hanson (2011, formalizing Roy 1951).

1.1 A minimal economic model of location choice

Suppose that a worker in low-income country 0 decides whether or not to move for work in the higher-income country 1. In country $k \in \{0, 1\}$, the Mincerian (1958) wage is

$$w_k = \exp\left(\mu_k + \delta_k s\right), \tag{1}$$

where μ sets the base wage for the unskilled, and s is schooling with return δ. Suppose that the cost of migrating from country 0 to 1 is

$$c_1 = \bar{c} + \theta s, \tag{2}$$

where $\bar{c}$ captures a fixed cost for all workers and θ captures costs that vary with education. This cost could vary inversely with education ($\theta < 0$) for a variety of reasons, notably that relatively poor people could be capital-constrained to pay for migration and more educated people tend to have more capital (Hanson 2006). The cost of not moving (c_0) is zero.

Finally, let the worker's utility in each country be $U_k \equiv \alpha(w_k - c_k) + \varepsilon$, where ε is unobserved and idiosyncratic to each person. Assuming that ε follows the Gumbel (Extreme Value Type I) distribution across individuals (McFadden 1974), the prevalence of emigration is given by

$$\ln \frac{E_1}{E_0} = \alpha\left(w_1 - w_0\right) - \alpha\bar{c} - \alpha\theta s, \tag{3}$$

where E_1 people emigrate and E_0 do not. The relationship in (3) captures the basic intuition that, all else equal, rising incomes in the origin country (w_0) should reduce the prevalence of emigration.

But over the course of economic development, all else need not remain equal. Suppose that the cause of the rise in origin-country incomes is a rise in average education s. This raises not just workers' earnings at home, $w_0(s)$, but also their potential earnings abroad, $w_1(s)$. Depending on the parameters μ and δ in (1), the wage gain $w_1 - w_0$ could rise with s, so the prevalence of emigration could rise with $w_0(s)$ even without capital constraints (Hanson 2010, 4380). Moreover, and separately, higher s would tend to raise migration prevalence in the location choice equation (3) by reducing costs for the capital-constrained ($\theta < 0$).

Even in this skeletal model, for poor countries, where the initial gap $w_1 - w_0$ is very large, rising fundamentals like s could raise emigration prevalence at early stages of development. This pattern could reverse at later stages when w_0 gets much higher and the importance of capital constraints diminishes in a more affluent population. In brief: At low w_0 it can be simultaneously true that the partial derivative $\frac{\partial \ln E_1/E_0}{\partial w_0} < 0$ and the total derivative $\frac{d \ln E_1/E_0}{d w_0(s)} > 0$, though the latter can change sign at higher w_0. Clemens and Mendola (2020) point out that this is an instance of Simpson's Paradox.

1.2   Richer models

The emigration life cycle can arise, then, even if we model emigration and development as nothing but location choice under rising home-country income. But development, of course, is far more than rising incomes for each individual (Massey et al. 1993). It is strongly associated with a transition out of agriculture and into urban employment, rising human capital, and greater interpersonal connections abroad, all of which can reduce the cost of emigration (Dao et al. 2018).

A richer model would naturally capture several other features of the development process that can encourage emigration (Clemens 2014). For example, the demographic transition that usually accompanies early economic development can produce a large surge in the number of young people just entering the labor force, who face lower costs of relocating (Hatton and Williamson 1998, 2011; Hanson and McIntosh 2016). This initial upward pressure on emigration, too, would dissipate as development proceeds.

Models incorporating uncertainty in earnings abroad could likewise accentuate the positive ef-

fects of rising incomes on emigration prevalence (Batista and McKenzie 2019; Bah and Batista 2018). Families with more disposable income could have greater ability to self-insure against that risk and thus be more likely to make the attempt (Gazeaud et al. 2019).

Models incorporating relative deprivation can moreover predict a nonmonotonic relationship between overall development and emigration (Stark and Yitzhaki 1988). In the minimal location choice model (3), the potential migrant's utility is exclusively a function of her own income abroad relative to counterfactual income at home. Relative deprivation models place some weight on income at home relative to that of others at home. Economic development can initially raise the number of higher-income individuals in the average potential migrant's comparison group, raising her relative deprivation and thus incentive to emigrate.

Emigration can be not only an effect of development in the country of migrant origin, but also a cause. It can cause the development of more inclusive political institutions (Barsbai et al. 2017; Karadja and Prawitz 2019), technological advance (Kerr 2008; Bahar and Rapoport 2018), and international trade (Felbermayr et al. 2015). These and other mechanisms would tend to produce a positive relationship between early-stage economic development and emigration rates.

What this entire class of models shares is the prediction that the relationship between rising incomes and emigration can be complex, can be positive at low levels of income, and can change in magnitude or even sign at higher levels of income. Taken as a whole, this literature suggests a reasonable theoretical prior that the relationship between rising incomes and the tendency to emigrate is nonmonotonic, as observed in cross section. But whether the emigration life cycle is the typical experience of developing countries over time remains an empirical question.

2 Data on emigration prevalence and incidence, 1960–2019

Until recently it was difficult to conduct even the simplest systematic tests of the relationship between rising incomes and emigration prevalence in (3). Available datasets had to cobble together emigration flows from each developing country by aggregating *immigration* flows of various kinds recorded by certain destination countries.

Such datasets have numerous disadvantages and cannot be used to test the theory above. They cover gross annual flows to a limited number of destinations and combine incommensurable types of movement reported by different destination-country governments (permanent versus temporary, regular versus irregular, foreign-born versus foreign nationals versus natives) (e.g. de Beer et al. 2010). Immigration flows in the main compilation of this kind, by the United Nations (2015) are not commensurable between countries and do not measure net migration. For example, the immigration data compiled for the United States and Spain report the issuance of residency permits, a large share of whose recipients arrived years or even decades beforehand, adjusting from a prior nonimmigrant visa or from irregular status. This can make such measures correlate *inversely* with physical migration. That database reports a rise of 55 percent in Mexico-US migration between 1995 and 2010, as measured by the issuance of residency permits, despite the fact that true Mexico-US migration *fell* by roughly half during that period (Passel and Cohn 2009).

Many countries also report the same person moving multiple times in one year as multiple migrants, or include their own nationals returning from abroad in their measures of immigration from each country. Some studies have addressed the comparability of flows across destination countries by focusing on a single destination country (Clark et al. 2007; Vogler and Rotte 2000), but this sacrifices a comprehensive picture of emigration from each origin country.

Moreover, such flow data also do not measure *net* migration from poor countries because several governments, such as those of the United States and Greece, do not report emigration flows at all. Economic models of location choice cannot be tested with gross flows alone, because even a rising gross migrant flow is compatible with a rising or falling net flow of people choosing to live in one country or another.

All of this changed with the global compilation of census, labor force survey, and population register data by Özden, Parsons, Schiff and Walmsley (2011), updated by the United Nations (2019). These datasets created full bilateral matrices of migrant stocks in every destination country on earth, from every origin country on earth, once each decade over a long period (1960–2019). They are based on a standardized definition of migrants (permanent migrants, defined by country of birth, whether regular or irregular).

This allows a comprehensive measure of emigrant stocks (prevalence) by summing the people born in each country living in all other countries, or any subset of them. Combined with origin-country population, this allows estimates of static migration prevalence as specified in equation (3). It also allows a comprehensive measure of sustained net emigration flows (incidence), simply by differencing emigrant stocks between decades. This has the advantage of being by far the best existing method of measuring sustained changes in net emigration flow rates out of all poor countries, with commensurable data.

The principal disadvantage of inferring net flows from changes in stocks is minor for the present purpose: It abstracts from emigrant deaths as a source of change in emigrant stocks over a 10-year period. Bias from this source is unlikely to affect the analysis to follow because it would exert a predominantly negative bias on changes in emigrant prevalence as a measure of net emigration flows. For example, the number of India-born people in Pakistan or the number of Algeria-born people in France could fall between 1980 and 1990, not due to migration in that decade but simply because some of those displaced at the time of decolonization died during that decade. True net emigration from India or Algeria would then be higher than that estimated by the change in emigrant prevalence. Thus for countries whose GDP per capita grew overall since 1960—the vast majority of developing countries—any positive relationship between changing emigrant prevalence and economic development would understate the true relationship. Nevertheless, in the Appendix the results in the main text are tested and found to be robust to using a measure of emigration (Abel 2018) that has been adjusted by demographic modeling to remove the effect of deaths.

3 The emigration life cycle, 1960–present

This section presents simple bivariate tests of the country-level relationship between emigration prevalence and average incomes in the migrant-origin country.

The empirics include both regressions using a linear specification of emigration E_1/E_0 and the log specification $\ln\left(E_1/E_0\right)$ from equation (3). The linear specification is included for transparency and because the relationship between economic growth and absolute migration prevalence might

be of interest. But the preferred estimates in the tables use the log specification grounded in the location choice equation. The log specification has the additional advantage that it implicitly controls for network effects on migration: Any initial condition that raises emigration from a given country with a fixed elasticity is absorbed into fixed effects. This includes the size of a country's prexisting diaspora, a major determinant of migration patterns (Munshi 2020). Finally, emigration prevalence is specified as $\ln\left(E_1/E_0\right)$ rather than the also-intuitive specification $\ln\left(E_1/(E_0 + E_1)\right)$ because the latter does not rest on the location choice equation (3), though in practice the two measures are highly correlated.
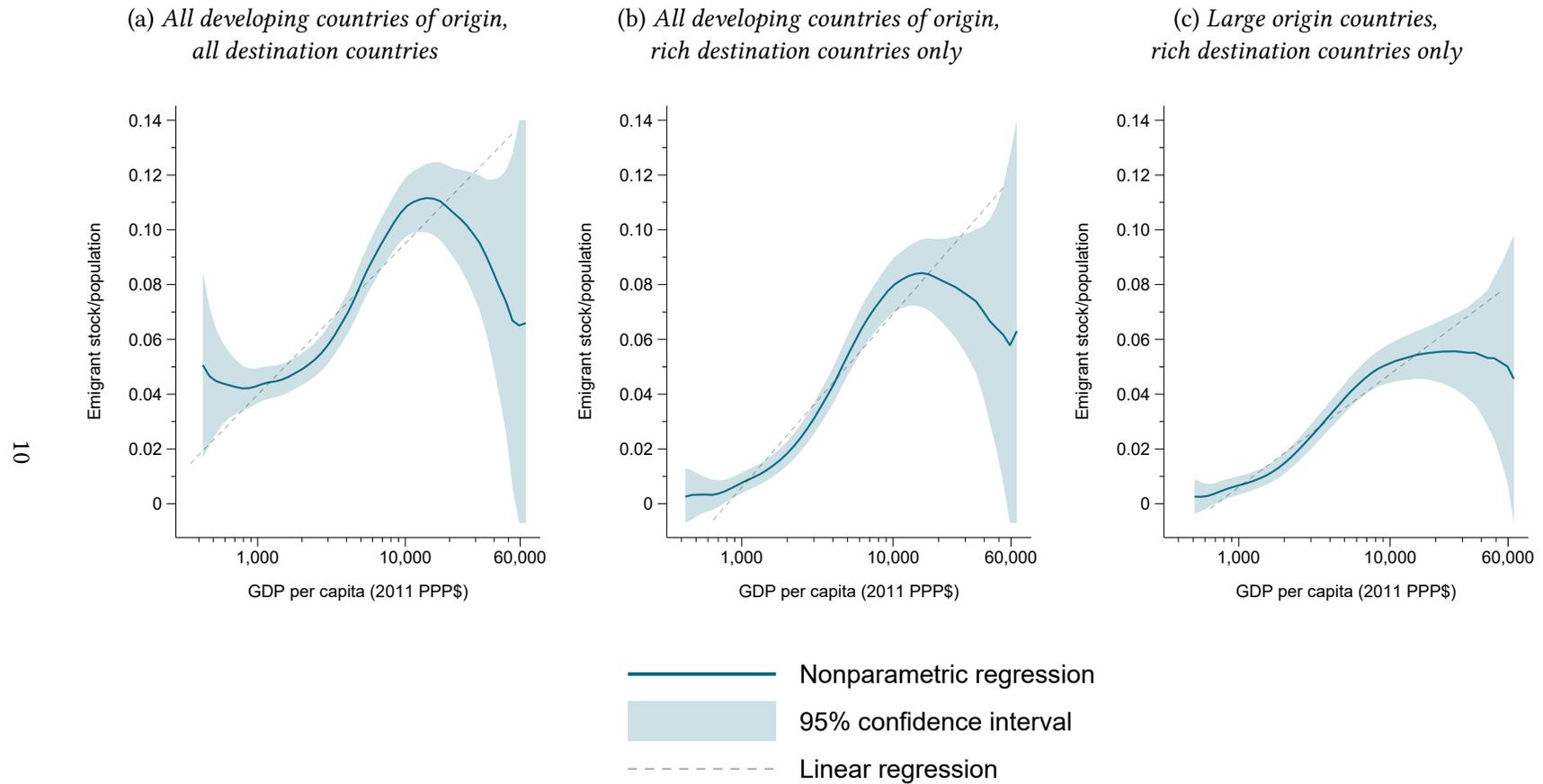
The first step is to pool the data on emigrant stocks, for all developing countries and all years, and conduct a nonparametric regression of emigrant stock on real GDP per capita in constant PPP-adjusted dollars.

3.1 Emigration prevalence and the level of GDP per capita

This core regression is presented in Figure 1a. This nonparametric Fan (1992) regression includes all countries of migrant origin that were developing countries for most of the period of analysis—that is, all countries that were not classified as "high income" by the World Bank in 1990. Thus France is excluded because it was a "high income" country for the entire period, but the Republic of Korea is included because it was not a "high income" country in 1990, though it is today. The emigration prevalence measures includes emigrants to all destination countries on earth. This pooled analysis combines variation in income and emigration within origin countries over time, and across origin countries within time-periods.

The emigration life cycle is striking in Figure 1a. In countries and years where GDP per capita is PPP\$1,000 (around today's level of Niger or Congo-Kinshasa), emigration prevalence has averaged about 4 percent of the population. When GDP per capita is around PPP\$10,000 (around today's level of Jordan or the Philippines), emigration prevalence has been over 2.5 times higher, averaging about 11 percent of the population. The confidence interval shows that the hypothesis of no such rise, or the hypothesis of lower emigration prevalence at higher levels of development, can be rejected beyond reasonable doubt in this income range.

9

Figure 1: Post-1960 Emigration Prevalence: Emigration versus origin-country income per capita, pooled countries and years 1960–2019

(a) *All developing countries of origin,*
all destination countries

(b) *All developing countries of origin,*
rich destination countries only

(c) *Large origin countries,*
rich destination countries only

Nonparametric regression

95% confidence interval

Linear regression

Observations are by origin country-year, in seven years (1960, 1970, 1980, 1990, 2000, 2010, 2019). "Developing countries" are countries classified as "low income" or ''middle income" by the World Bank in 1990—that is, countries that were not classified as "high income" for most of the period 1960–2019. Solid nonparametric regression line is Fan (1992) local-linear regression, Epanechnikov kernel, optimal bandwidth minimizes conditional weighted mean integrated squared error. Dashed line is linear ordinary least squares fitted to the same data. Figure 1a has $N = 908$ and bandwidth 0.372 log points, Figure 1b has $N = 908$ and bandwidth 0.395 log points, Figure 1c has $N = 708$ and bandwidth 0.465 log points. "Rich" destination countries are those that were classified as "High Income" by the World Bank as of fiscal year 2020 (over \$12,375 gross national income per capita at Atlas exchange rates). "Large" origin countries are those with population above the 25th percentile (that is, greater than 2.49 million) in 2019. 95% confidence interval clipped at graph edge for legibility.

At levels of income above roughly PPP$10,000, emigration prevalence tends to be sharply lower. By the time that once-developing countries have risen to PPP$40,000, emigration prevalence is no longer statistically distinguishable from its average level in the poorest countries. The figure demonstrates the statistically significant nonlinearity of the relationship by superimposing a (dashed) linear fit to the same data.

This relationship is even more pronounced if the emigration prevalence measures are restricted to those residing in countries that are "rich"—that is, those defined by the World Bank as "high income" in 2020. This is shown in Figure 1b. For these destination countries, emigration prevalence from middle-income origin countries is an order of magnitude higher than from low-income origin countries. In countries and years where GDP per capita is PPP$1,000, emigration prevalence to rich countries has averaged 0.8 percent of the population. At PPP$10,000, it has averaged 8.2 percent of the population. At even higher incomes, emigration prevalence to rich destinations is sharply lower.[2]

This result is not driven by small countries of migrant origin. Figure 2c shows the result from Figure 1b after dropping the smallest quartile of countries—that is, dropping all countries with a 2019 populations below 2.49 million. The absolute rise in emigration prevalence with income becomes moderately smaller: Emigration prevalence is 5.3 percent at GDP per capita PPP$10,000, compared with 0.7 percent at GDP per capita PPP$1,000. But the relative change is similar: With or without the smallest quarter of developing nations, the prevalence of emigration to rich countries is an order of magnitude higher at GDP per capita PPP$10,000 than at PPP$1,000, in data pooled across countries and years.

---

[2]How to classify countries presents a conundrum. If all countries are classified as either 'developing' origin countries or 'rich' destinations, but never both, this requires throwing away a great deal of information, such as the relationship between emigration and development that occurred in Korea before it became a high-income country. The approach here is to consider a country a 'developing' country of migrant origin if it had *not* reached the World Bank classification of "high income" by 1990, that is, it was low- or middle-income during the majority of the period under study, and to consider a country a 'rich' country of migrant destination if it had reached the World Bank "high income" classification by fiscal year 2020. This results in Korea and 16 other countries with the relevant data available being included as 'developing' origin countries, and having their inflows included in worldwide emigration to 'rich' countries. This is appropriate, since there appears to be no theoretical reason that the drivers of Korea's emigration to today's rich countries when Korea was a developing country should differ fundamentally from the drivers of now-developing countries to now-rich Korea. Note the the unit of observation in all regressions is the country of origin, so that no country is doubly classified in an econometric sense. The 17 countries thus treated are: Antigua and Barbuda, Bahrain, Barbados, Chile, Greece, Hungary, Republic of Korea, Malta, Oman, Panama, Poland, Portugal, Saudi Arabia, Seychelles, St. Kitts and Nevis, Trinidad and Tobago, and Uruguay.

3.2 Net emigration flows and economic growth

It is possible, in principle, that typical countries do not follow the curves in Figure 1 as they develop. There could be some unobserved trait of countries that reach middle-income status, such as their geographic location or prior diaspora size, that makes them start from a higher emigration rate. It is therefore useful to track within-country changes over time, testing for the emigration life cycle in the relationship between net emigration *flows* (changes in stocks) and economic *growth* (changes in levels of GDP per capita). This approach controls for any unobserved time-invariant country traits.

Figure 2 displays the raw data for within-country changes in emigration prevalence versus GDP per capita for all countries that experienced positive net economic growth during this period. The arrows show the change in emigration prevalence versus GDP per capita for all developing countries, 1970–2019. (The starting year 1970 is chosen to maximize coverage of 98 countries; starting in 1960 allows coverage of only 69 countries, but looks very similar.)

This visualization of the raw data makes it clear that within-country changes in Figure 2 closely follow the curve seen in the pooled data of Figure 1. Consider Figure 2a, covering all developing countries of origin and all destination countries. For countries that grew but remained poor during this period, the changes in emigration exhibit little pattern. Emigration prevalence fell in Burkina Faso and Mozambique, but rose in Uganda and Haiti. This matches the relatively flat part at the left of the pooled-data curve in Figure 1a. For countries that successfully grew toward PPP$10,000 GDP per capita, there is a striking pattern of sharply rising emigration prevalence, matching the steep part of the pooled-data curve in Figure 1a. And for developing countries that grew above PPP$10,000, emigration prevalence grew slowly or fell, matching the rightmost part of the emigration life cycle in Figure 1a.

This pattern is even clearer when the data are restricted to cover rich destination countries only (Figure 2b), and again when small migrant-origin countries are dropped (Figure 2c). If emigration prevalence fell as countries develop, we would observe a predominance of arrows pointing downward and to the right in all panels of Figure 2. Instead, we observe typical developing countries roughly following the curves in Figure 1 as they develop over time.

Figure 2: Post-1960 Emigration Flows and Economic Growth: Changes in emigration prevalence vs. increases in origin-country income per capita, within-country changes, 1970–2019

(a) *All developing countries of origin,*
all destination countries

(b) *All developing countries of origin,*
rich destination countries only

(c) *Large origin countries,*
rich destination countries only
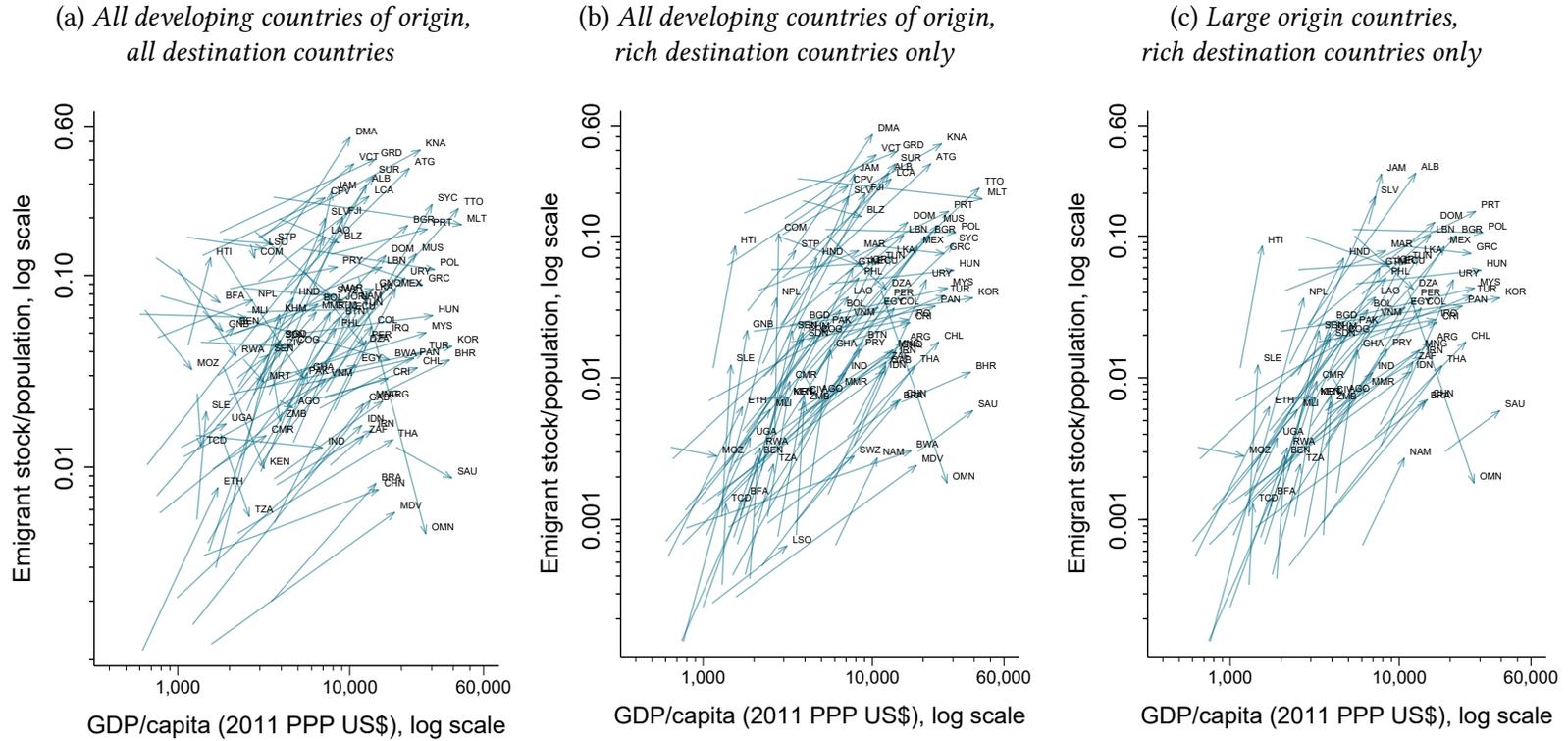
Countries with positive cumulative growth in real GDP per capita between 1970 and 2019. $N = 98$ in (a) and (b), $N = 73$ in (c). Arrows show within-country changes: Start of each arrow is emigrant stock and GDP/capita in 1970, tip of each arrow is same numbers in 2019. Excludes countries classified as "high income" by the World Bank as of calendar year 1990. Begins in 1970 rather than 1960 in order to maximize country coverage; the same data back to 1960 are available for only 69 countries. "Rich" destination countries are those classified as "High Income" by the World Bank as of fiscal year 2020 (over $12,375 gross national income per capita at Atlas exchange rates). "Large" origin countries are those with population above the 25th percentile (that is, greater than 2.49 million) in 2019.

An advantage of presenting the raw data in Figure 2 is transparency. A disadvantage is that this does not present the most rigorous tests. In particular, it is difficult to tell in Figure 2 whether countries with a greater degree of economic growth had a greater degree of rise in emigration prevalence. Arrows for such countries could point upward and to the right simply because GDP per capita and emigration are rising everywhere due to other causes, not because there is any link at the country level between the magnitude of one and the magnitude of the other.

One way to maintain transparency while making the tests more rigorous is shown in Figure 3a. It controls for unobserved traits at the country level that make a country start out very poor. In the figure, the sample used in the nonparametric regression in Figure 1a is restricted to include only countries that, at some point during 1960–2019, had a GDP per capita under PPP$2,000 in constant 2011 dollars. That is, all observations to the right of the gray area in Figure 3a are countries that were once very poor but had successful, sustained economic growth. The further those countries grew economically, the more their emigration prevalence rose. That regression, in the solid line, is superimposed on the unrestricted regression, in the dotted line. The two lines are not statistically distinguishable. This conclusion is not sensitive to choosing a different, arbitrary GDP per capita cutoff (Figure 3b).[3] This implies that if there is some substantial un-observed difference between the country-years on the right of the curve in Figure 3a and those on the left, that difference is independent of traits that make developing countries start out very poor or start out richer.

It nevertheless remains possible, in principle, that there is some unobserved difference between the countries that did experience sustained economic growth and those that did not—a difference that would have sharply altered the consequent changes in emigration prevalence. In other words, if the countries in Figure 3a that grew little and remained in the gray area had instead grown to the right of the gray area, perhaps emigration prevalence would have fallen in those countries due to some unobserved characteristic that distinguishes them from the countries that did grow.

---

[3]The World Bank defines country income categories by average incomes in Atlas exchange rate dollars, not PPP-adjusted dollars, so there is no strict definition of 'low-income' countries based on PPP dollars. But PPP$2,000 corresponds roughly to the average PPP-adjusted GDP per capita among "low-income" countries, and PPP$4,000 corresponds roughly to the highest levels of GDP per capita observed in "low-income" countries. These numbers where the basis for the arbitrary cutoffs chosen in the figure.

Figure 3: Post-1960 Emigration Prevalence: Once-poor countries moving along the migrant-stock curve as they grow richer

(a) *Only countries starting <$2,000/capita*   (b) *Only countries starting <$4,000/capita*
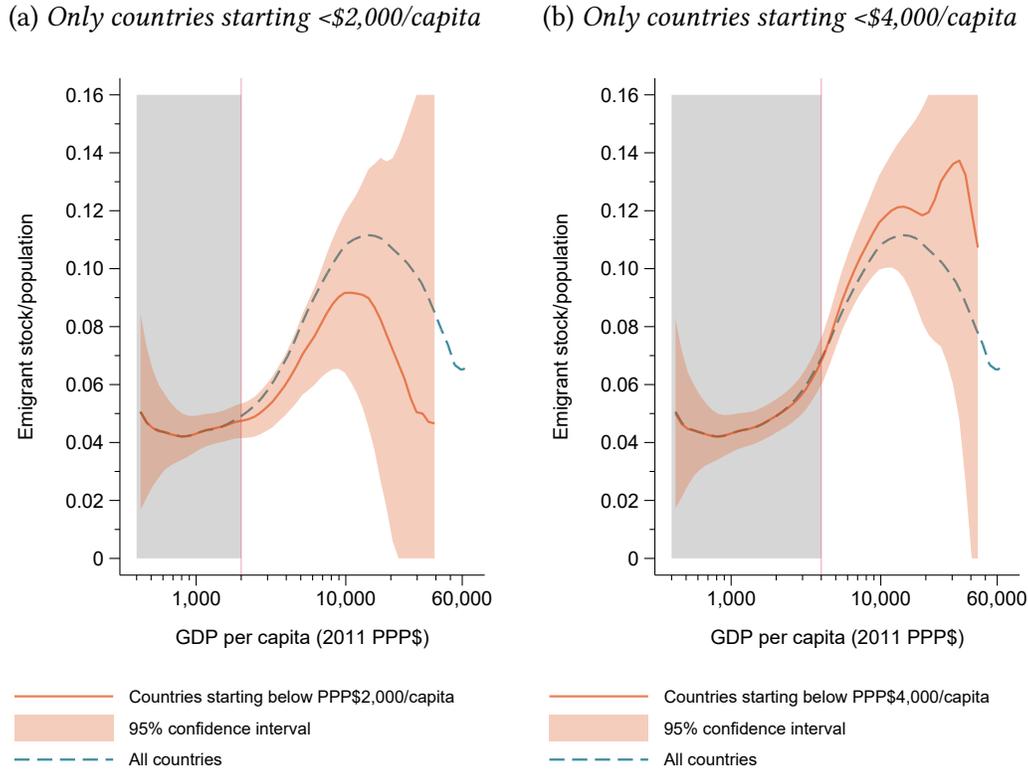
Observations by origin country-year. Solid line in Figure 3a shows only migrant-origin countries whose GDP per capita was below PPP$2,000 at some point during 1960–2019. Thus the line to the right of the shaded area shows countries that grew to higher incomes starting from very low income. The dashed line, for reference, is the regression line from Figure 1a for all origin countries. Figure 3b is identical, but for a PPP$4,000 cutoff. All destination countries. Nonparametric regression line is a Fan (1992) local-linear regression, Epanechnikov kernel; optimal bandwidth minimizes conditional weighted mean integrated squared error (0.471 log points in Figure 3a, 0.416 in Figure 3b).

To address concerns of this type, Figure 4 presents a nonparametric regression of changes in emigration prevalence on changes in GDP per capita, using exclusively within-country changes over time, for countries that started out below PPP$2,000 GDP per capita. The horizontal axis is relative growth in GDP per capita relative to the lowest level of GDP per capita seen during 1960–2019. The vertical axis shows changes in emigration prevalence relative to the emigration prevalence observed in the year of lowest GDP per capita. This curve is similar to the curve in Figure 3a, though it peaks at a somewhat lower emigration prevalence (6 percent versus 9 percent). This indicates that the relationships in Figure 1a and Figure 3a are somewhat biased by time-invariant unobserved heterogeneity between countries, but the broad pattern of the

Figure 4: Post-1960 Emigration Flows and Economic Growth: Changes in emigration prevalence versus increases in origin-country GDP per capita, using only within-country economic growth starting <$2,000/capita
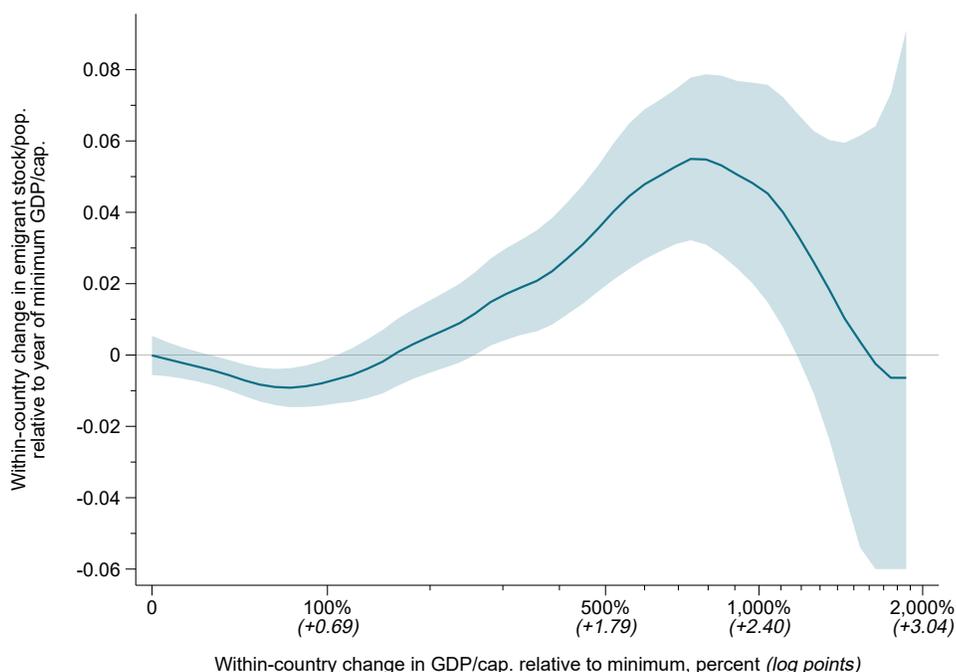
Observations by origin country-year. Includes only migrant-origin countries whose GDP per capita (2011 PPP$) was below $2,000 at some point during 1960–2019. Horizontal axis shows change of GDP per capita from its lowest value in subsequent years, and vertical axis shows change in emigrant stock during the same years. All destination countries. Nonparametric regression line is a Fan (1992) local-linear regression, Epanechnikov kernel; optimal bandwidth minimizes conditional weighted mean integrated squared error (0.260 log points).

emigration life cycle does not arise from such unobserved heterogeneity.

The regression in Figure 4 is a stringent test. It discards all information contained in the trajectories of countries in Figure 1a that started out at higher levels of income, making the strong assumption that these are uninformative about the future trajectories of today's poor countries (e.g. Mummolo and Peterson 2018). That is, it starts from the assumption that nothing can be learned about the expected trajectory of Mali from the experience of countries that started out somewhat richer than Mali. In fact, Figures 1 and 4 and imply that countries starting out very poor have largely followed the same trajectory as those that did not. The strong visual resemblance among all of the Figures 1 through 4 indicates that the emigration life cycle is a process

that occurs within poor countries over time, not an illusion created by snapshots across countries.
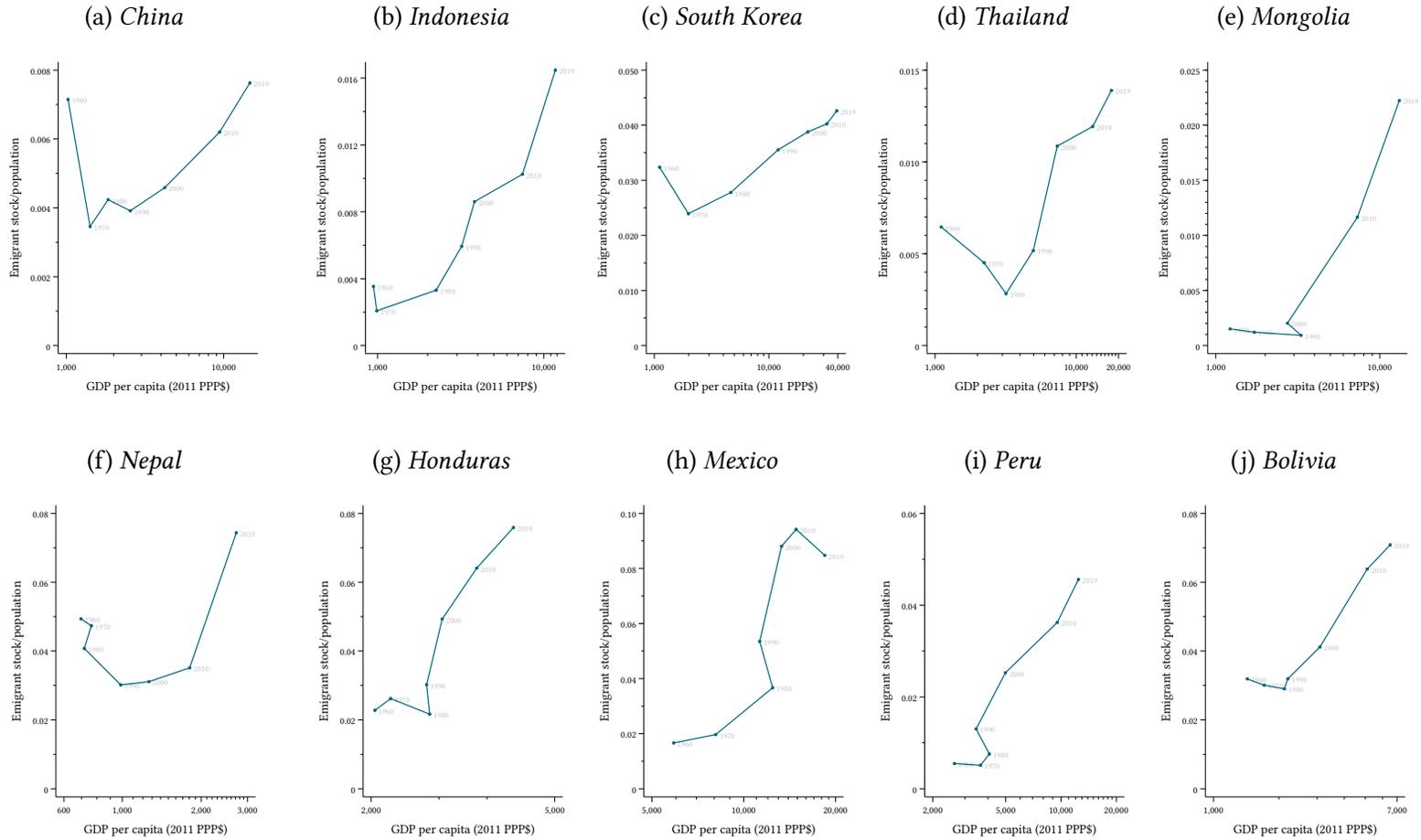
Because emigration prevalence is often higher in microstates, the policy literature has raised the question of whether these average relationships are driven by very small countries (e.g. Bade and de Kemp 2018, 43). They are not, as Figures 1c and 2c show. The absolute height of the life-cycle curve does appear somewhat lower among large developing countries, but the relative rise in emigration prevalence during development is not substantially different in the subsample of relatively large countries. Figure 5 clarifies this by graphing the raw data for several important developing countries individually. The pattern of sustained increases in emigration prevalence during development takeoffs is observed in large countries of migrant origin (China, Mexico). It is observed in some of the major economic development success stories of the past half century (Indonesia, Thailand, Peru). It is not confined to unrepresentative microstates.

3.3 Alternative regression specifications

The preceding regressions embody numerous assumptions in their choice of regression specifications. These choices require trade-offs. By estimating a clearly nonlinear relationship nonparametrically, the preceding regressions preclude the use of traditional linear fixed and random effects to control for time-invariant unobserved heterogeneity among countries or time periods. They also embody the assumption that the relationship between economic growth and rising emigration prevalence will have the same absolute magnitude in countries with low or high emigration prevalence. They embody the assumption, based on the location choice equation (3), that the proper measure of emigration prevalence is relative emigrant stocks. All of these assumptions are relaxed in this section, and the costs and benefits discussed.

Table 1 presents linear regressions corresponding to the nonparametric regressions in Figure 1 and the raw-data graphs in Figure 2. The important cost of using a linear estimator in this case is that it requires an arbitrary choice about which portion of the nonlinear relationship to characterize linearly. Here the linear regressions are arbitrarily restricted to country-years below PPP$10,000, based on the turning point evident in Figure 1. The benefit of this arbitrary choice is that it allows more transparent and familiar methods of controlling for unobserved

17

Figure 5: POST-1960 EMIGRANT STOCKS: Emigration versus GDP/capita, representative countries

(a) *China*

(b) *Indonesia*

(c) *South Korea*

(d) *Thailand*

(e) *Mongolia*

(f) *Nepal*

(g) *Honduras*

(h) *Mexico*

(i) *Peru*

(j) *Bolivia*

All destination countries.

country traits. The coefficients in the table are the coefficient estimates on ln(GDP per capita) in a bivariate regression with a measure of emigration prevalence as the dependent variable, plus a (suppressed) constant term.

The first row begins with all developing countries of origin, the second row restricts emigrant counts to rich destination countries only, and the third row restricts the sample to relatively large origin countries, just as in Figures 1 and 2. The first four columns measure emigration prevalence with emigrants as a fraction of the origin country population; the other four columns use the natural log of that number. This latter measure allows the relationship with GDP per capita to differ at different absolute levels of emigration prevalence. Within each dependent variable, the first column shows an ordinary least squares regression. The second column includes country fixed effects. The third column includes country random effects. The fourth column includes country random effects in both the slope and the intercept. In all regressions, standard errors (in parentheses) and the associated p-values (in gray text) are clustered by country.

The overall pattern in the table is that emigration prevalence is positively associated with GDP per capita below PPP\$10,000, with very high statistical precision, under the strong assumption of linearity. The hypothesis of no association can be rejected in all specifications at levels below $p = 0.001$. The hypothesis that emigration falls in the early stages of development can be rejected even more strongly.

This finding is robust to controlling for country fixed effects, country random effects, and country random effects in both the slope and intercept. When country effects are controlled for, the magnitude of the slope is roughly two-thirds of its value when they are not controlled for. This matches the visual evidence from the nonparametric regressions in Figures 3a and 4 that roughly two-thirds of the rise seen in the pooled data remains when time-invariant country heterogeneity is gradually stripped away to analyze within-country variance only. When emigration prevalence is specified linearly, the slope for rich destination countries and large origin countries is similar to or slightly less than the slope for all destination countries and all developing origin countries, as in the pooled data of Figure 1. When emigration prevalence is specified in logs, the slope for rich destination countries and large origin countries is much steeper than for all destination countries and all developing origin countries, as in the within-country changes of

Table 1: Post-1960 Emigration: Linear regressions with arbitrary income cutoff

	GDP/capita < PPP$10,000							
Subsample:								
Dep. var:	Emigrant stock/pop.				ln(Emigrant stock/pop.)			
Country effects, intercept:	—	Fixed	Random	Random	—	Fixed	Random	Random
Country effects, slope:	—	—	—	Random	—	—	—	Random
Unrestricted, N = 687								
ln(GDP/capita)	0.025	0.016	0.018	0.021	0.383	0.347	0.362	0.347
s.e.	(0.0056)	(0.0047)	(0.0045)	(0.0062)	(0.1039)	(0.0690)	(0.0656)	(0.0701)
p-val.	< 0.0001	0.0008	0.0001	0.0006	0.0003	< 0.0001	< 0.0001	< 0.0001
Rich destination countries only, N = 687								
ln(GDP/capita)	0.031	0.020	0.021	0.023	1.290	0.737	0.856	0.916
s.e.	(0.0049)	(0.0042)	(0.0040)	(0.0032)	(0.1310)	(0.1091)	(0.0985)	(0.0970)
p-val.	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Rich destination countries and large origin countries only, N = 560								
ln(GDP/capita)	0.021	0.013	0.014	0.016	1.218	0.730	0.857	0.896
s.e.	(0.0044)	(0.0035)	(0.0033)	(0.0043)	(0.1382)	(0.1240)	(0.1096)	(0.1104)
p-val.	< 0.0001	0.0003	< 0.0001	0.0002	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Standard errors in parentheses, clustered by country. *p*-values in gray (H_0: coefficient = 0). Constant term omitted in table. "Rich" destination countries are those that were classified as "high income" by the World Bank as of fiscal year 2020 (over $12,375 gross national income per capita at Atlas exchange rates). "Large" origin countries are those with population above the 25th percentile (that is, greater than 2.49 million) in 2019.

Figure 2 (where the vertical axis is in logs).

Table 1 implies that from 1960 to present, average developing countries whose economies have grown more over time have experienced larger increases in emigration prevalence over time. Developing countries with GDP per capita below PPP$10,000 where real GDP per capita rose by 1 log point (a factor of 2.72) saw an average rise of 1.6 percentage points in the fraction of their population that lived in any other country. A rise of 100 percent in average incomes was associated with a rise of 34.7 percent in the size of the emigrant stock relative to the origin-country population. The corresponding results for emigration to rich countries in particular are +2.0 percentage points and +73.7 percent, respectively. These estimates arise entirely from within-country variation, ignoring all information contained in between-country variation.

A natural next question is how much of this pattern arises from global changes that affected all countries between time periods. The uncritical inclusion of time fixed effects has important disadvantages: They absorb rises in average emigration prevalence for all countries collectively, for any reason—including rises in income shared by all countries. But up to half of changes in economic growth in the poorest countries are caused by changing economic conditions at the global level (e.g. Drummond and Ramírez 2009). In the present dataset, time effects explain half of the within-country variance in average incomes.[4] Any emigration effects of that portion of economic growth are absorbed away by time fixed effects, eliminating part of the very effect that the regressions seek to measure. With time fixed effects in the regression, the within-country coefficient on income can no longer be interpreted as the relationship between emigration flows and economic growth. Instead, it is the relationship between emigration flows and *country-specific deviations of economic growth from global average economic growth*. That deviation can be negative even for a country exhibiting positive economic growth.

With this in mind, panel (a) of Table 2 shows the results of the core specification in Table 1 with time fixed effects. The coefficient estimates remain positive but decline in magnitude, by roughly 25–50 percent. Controlling for country-invariant heterogeneity over time, developing countries with GDP per capita below PPP$10,000 where real GDP per capita rose by one log point *more*

---

[4]In a simple regression of ln GDP per capita on country fixed effects, adjusted R^2 is 0.685, and with year fixed effects adjusted R^2 is 0.837. Thus the fraction of within-country variance in ln GDP per capita explained by time effects is $(1 - 0.837)/(1 - 0.685) = 0.520$.

than it rose in the average developing country saw an average rise of 0.9 percentage points in the fraction of their population that lived in any other country, or a rise of 15.2 percent in the size of the emigrant stock relative to the origin-country population (row 2, columns 2 and 6, of Table 2).

Finally, the preceding regressions have followed the theory in (3) in investigating the relationship between emigrant stocks and income levels, or emigration flows and income growth. But another relationship of interest might be that between emigration flows and income levels. It may be useful to test whether the emigration life cycle pattern is robust to that alternative specification.

Figure 6 repeats the analysis of Figure 1a, Figure 3a, and Figure 4 but changes the dependent variable to net emigration flow rates (changes in the emigrant stock per 1,000 people in the origin-country population, *per year* on average during each period 1960–1970, 1970–1980, 1980–1990, 1990–2000, 2000–2010, and 2010–2019). GDP per capita levels are measured in the first year of each period during which flows are measured.

The emigration life cycle remains apparent in emigration flows versus income levels, in Figure 6a. Emigration flows relative to origin-country population are higher than in the poorest countries at levels of GDP per capita up to roughly PPP$5,000. At even higher levels of income, these flow rates are flat, and at levels above PPP$10,000 they begin to fall. The initial rise is slightly less steep for countries that started out below PPP$2,000 GDP per capita (Figure 6b), but the pattern of rise and fall remains highly statistically significant. That is, the life cycle in Figure 6a is not driven primarily by unobserved heterogeneity between developing countries that start out very poor and those that do not. The rise and fall in flows remains statistically significant in exclusively within-country variation for countries starting out below PPP$2,000 (Figure 6c).

Again, these results are not driven by microstates. Figure 7 graphs the raw data on net emigration flow rates for the same countries for which stocks were shown in Figure 5. China, Indonesia, Mexico, the Republic of Korea, and many other countries where sustained economic growth took hold saw decades of large increases in the net emigration outflow rate.
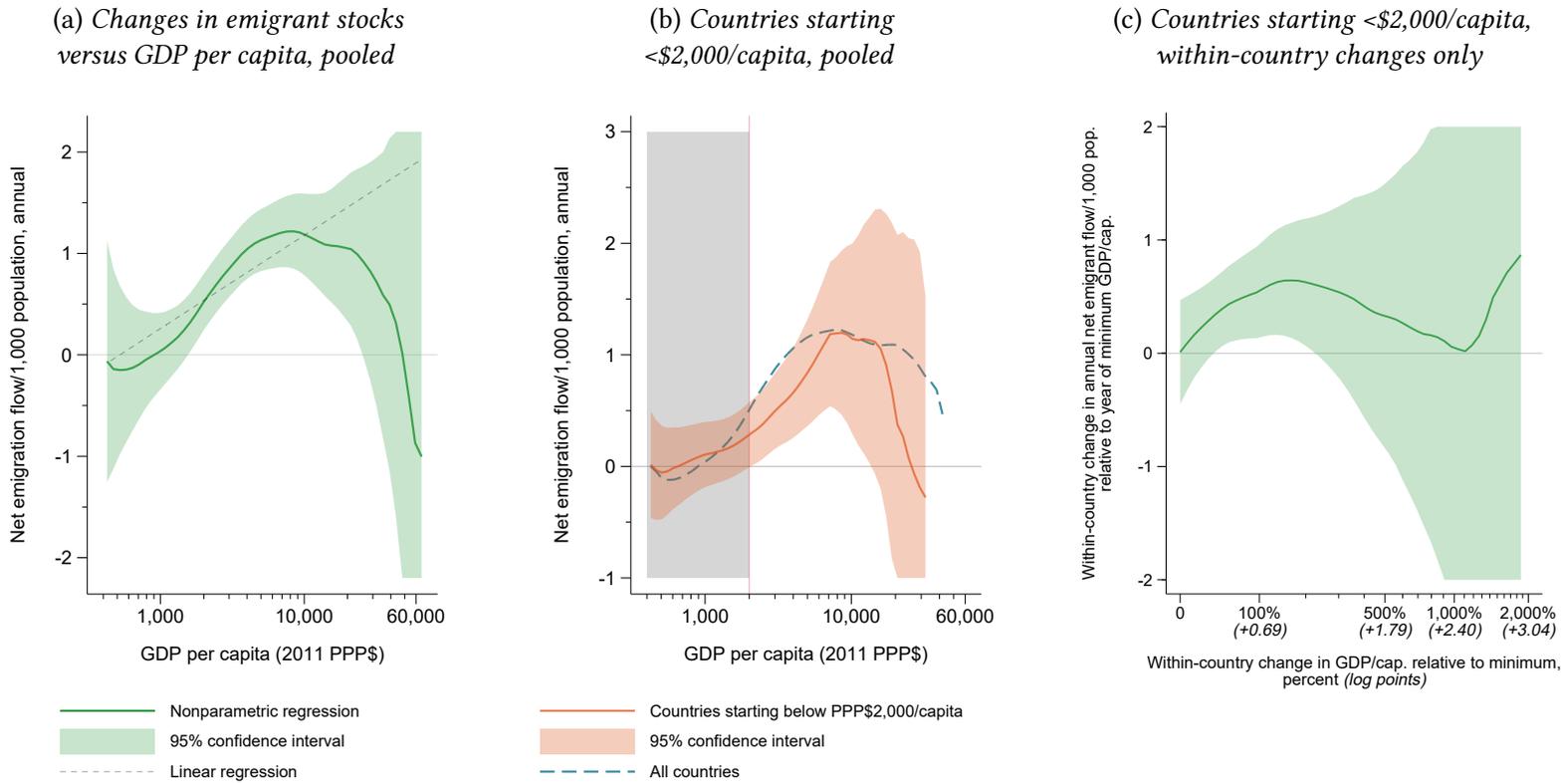
These results on emigration *flows* versus income *levels*, like the earlier results on emigration *flows* versus income *growth*, are highly robust to changes of regression specification. They are

Table 2: Post-1960 Emigration: Linear regressions with stocks or flows as dependent variable, with or without time effects

(a) Stocks

	Subsample:	GDP/capita < PPP$10,000							
	Dep. var:	Emigrant stock/pop.				ln(Emigrant stock/pop.)			
Country effects, intercept:		—	Fixed	Random	Random	—	Fixed	Random	Random
Country effects, slope:		—	—	—	Random	—	—	—	Random
Unrestricted, N = 687									
ln(GDP/capita)		0.025	0.016	0.018	0.021	0.383	0.347	0.362	0.347
	s.e.	(0.0056)	(0.0047)	(0.0045)	(0.0062)	(0.1039)	(0.0690)	(0.0656)	(0.0701)
	p-val.	< 0.0001	0.0008	0.0001	0.0006	0.0003	< 0.0001	< 0.0001	< 0.0001
Unrestricted, with year fixed effects, N = 687									
ln(GDP/capita)		0.024	0.009	0.012	0.015	0.339	0.152	0.201	0.184
	s.e.	(0.0057)	(0.0047)	(0.0043)	(0.0062)	(0.1073)	(0.0769)	(0.0694)	(0.0770)
	p-val.	0.0001	0.0617	0.0051	0.0127	0.0020	0.0500	0.0038	0.0170

(b) Flows

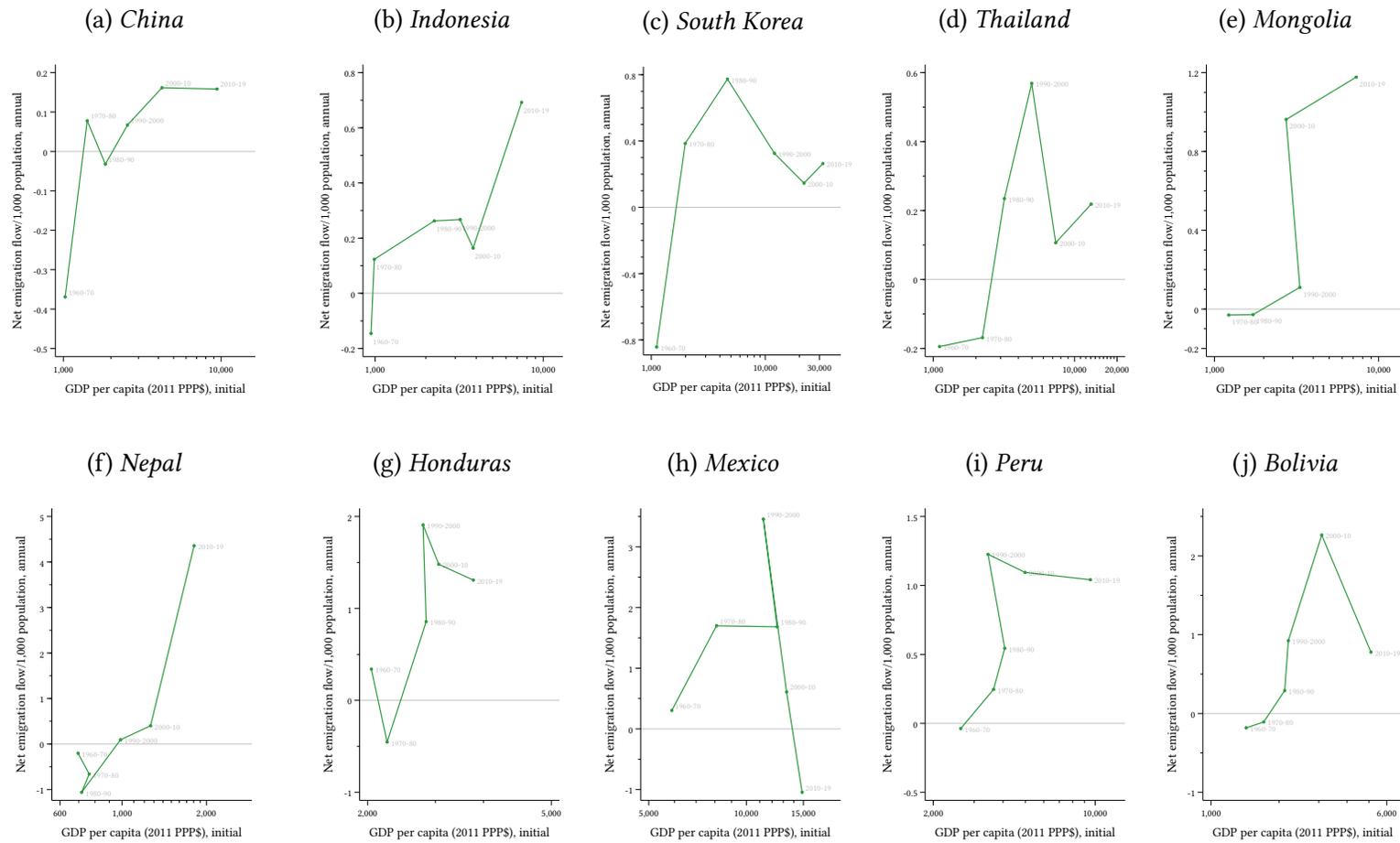|	Subsample:	GDP/capita < PPP$5,000							
	Dep. var:	Net emigrant flow/1,000 pop.				asinh(Net emigrant flow/1,000 pop.)			
Unrestricted, N = 461									
ln(GDP/capita)		0.974	0.864	0.971	0.997	0.483	0.367	0.468	0.468
	s.e.	(0.2439)	(0.3072)	(0.2525)	(0.2411)	(0.0892)	(0.1301)	(0.0920)	(0.0873)
	p-val.	0.0001	0.0058	0.0001	< 0.0001	< 0.0001	0.0057	< 0.0001	< 0.0001
Unrestricted, with year fixed effects, N = 461									
ln(GDP/capita)		0.944	0.666	0.878	0.940	0.473	0.297	0.452	0.469
	s.e.	(0.2389)	(0.3298)	(0.2560)	(0.2422)	(0.0872)	(0.1318)	(0.0891)	(0.0863)
	p-val.	0.0001	0.0458	0.0006	0.0001	< 0.0001	0.0259	< 0.0001	< 0.0001

Standard errors in parentheses, clustered by country. *p*-values in gray (H_0: coefficient = 0). Constant term omitted in table. "asinh" is the inverse hyperbolic sine. Flows are measured as *annual* change in emigrant stock per 1,000 population of the migrant-origin country.

Figure 6: Post-1960 Net Emigrant Flows: Alternative specification with changes in emigrant stocks as dependent variable

(a) *Changes in emigrant stocks versus GDP per capita, pooled*

(b) *Countries starting <$2,000/capita, pooled*

(c) *Countries starting <$2,000/capita, within-country changes only*

Nonparametric regression
95% confidence interval
Linear regression

Countries starting below PPP$2,000/capita
95% confidence interval
All countries

Observations by origin country-year, in six years (GDP per capita measured in 1960, 1970, 1980, 1990, 2010; change in emigrant stock *per year* measured in *subsequent* decade 1960–1970, 1970–1980, 1980–1990, 1990–2000, 2000–2010, and 2010-2019). Solid nonparametric regression line is a Fan (1992) local-linear regression, Epanechnikov kernel; optimal bandwidth minimizes conditional weighted mean integrated squared error. Dashed line in (a) is linear OLS fit to the same data. Dashed curve in (b) is the nonparametric regression from (a) for all countries. In (a) $N = 761$ and bandwidth is 0.562 log points; in (b) $N = 368$ and bandwidth is 0.443 log points; in (c) $N = 283$ and bandwidth is 0.461 log points. 95% confidence intervals clipped at edge of graph for legibility.

Figure 7: POST-1960 NET EMIGRATION FLOWS: Changes in emigrant stocks versus GDP per capita, representative countries

(a) *China* (b) *Indonesia* (c) *South Korea* (d) *Thailand* (e) *Mongolia*

(f) *Nepal* (g) *Honduras* (h) *Mexico* (i) *Peru* (j) *Bolivia*

All destination countries. Flows are annualized average change in emigrant stock during decade; GDP per capita is in initial year of each decade.

not driven by country effects, by the linear specification of the flow rate, or by country-invariant time effects. This is shown in panel (b) of Table 2. The linear regression specifications there again have an important disadvantage, which is that they require an arbitrary assumption about what portion of the manifestly nonlinear relationship to characterize linearly. The table assumes a linear relationship for flows below a GDP per capita of PPP$5,000. This is arbitrarily chosen, but guided by the peak in Figure 6a and the inflection point in Figure 1a, as well as by estimates in the literature (e.g. Djajic et al. 2016).[5]

The regressions show that, after absorbing all time-invariant heterogeneity between countries, developing countries with GDP per capita below PPP$5,000 where real GDP per capita rose by 1 log point (a factor of 2.72) saw an average rise of 0.86 in the annual net flow of emigrants per 1,000 population over the following decade, or a rise of 36.7 percent in the size of the net emigrant flow relative to the origin-country population (row 3, columns 2 and 6, of Table 2). In other words, among low-income countries, the within-country elasticity of emigration to GDP per capita is similar whether the measure of emigration is flow rates (0.367) or stocks (0.347) (Table 2, column 6). Including time fixed effects, a lower bound on the effect of interest is that the same increase in incomes was associated with an average rise of 0.67 in the annual number of emigrants net of immigrants per 1,000 population over the following decade, or a rise of 29.7 percent in the size of the net emigrant flow relative to the origin-country population (Table 2, row 4, columns 2 and 6).

In sum, the results in the prior section are highly robust to specification changes including linear regression with country effects, a logarithmic dependent variable, time-period fixed effects, and a specification that estimates the relationship between the emigration *flow* rate and the *level* of GDP per capita, rather than between the emigration *flow* rate and the *growth* of GDP per capita. Economic development since 1960 has initially been accompanied by large average increases in emigration from developing countries. This conclusion holds even when using exclusively within-country variation (country effects) and flexibly controlling for any changes in global conditions (time effects).

---

[5]Because changes in migrant stocks frequently take values below zero, the nonlinear transformation of the flow rates uses the inverse hyperbolic sine (asinh) rather than the logarithm, but the coefficient estimates can be similarly interpreted as elasticities (Bellemare and Wichman 2020).

4 The emigration life cycle before 1914

Why does the emigration life cycle occur? The literature has proposed a variety of mechanisms, including changes in human capital, demographic structure, and urban-rural structure typical to economic development (e.g. Hatton and Williamson 1998; Dao et al. 2018). This analysis seeks to shed light on such mechanisms by considering similarities and differences in the emigraiton life cycle between the second wave of globalilzation (1960–2019) and the first (1850–1914).
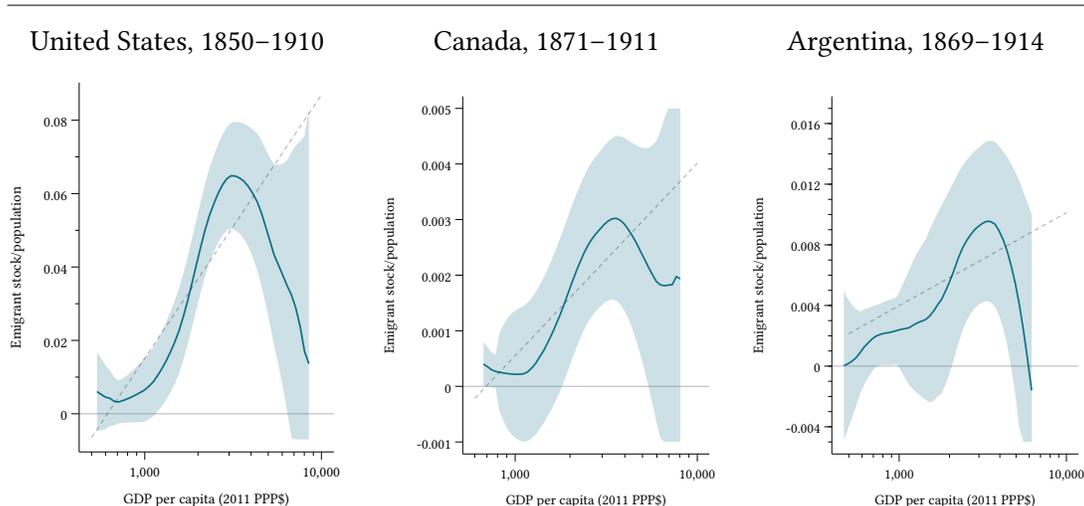
The initial step is simply to test for the existence of the emigration life cycle in the 1850–1914 period using the same empirical framework as above. This is a test of the longstanding hypothesis that the more recent emigration life cycle has recapitulated the older one (e.g. Hatton and Williamson 2002; Ferrie and Hatton 2015; Williamson 2015) even when controlling for unobserved differences between countries. Previous empirical work on the effects of origin-country economic growth on emigration before 1914, back to Thomas (1973) and before, has not directly compared the emigration life cycle during these two historical eras in a single quantitative framework.

There is no *prima facie* reason for the life cycle to proceed similarly in the two different historical eras. They differ in many ways. The sharp decline in overseas transportation costs after 1840 was relatively larger than any such decline after 1960. International migration was more constrained by policy barriers after 1960 than before 1914—with major exceptions such as the United States' tight restrictions on Chinese immigration and Asian naturalization. The absolute and globally relative incomes and education levels of potential emigrants were different among Europeans before 1914 and developing-country residents after 1960. Remittances and the ease of circular migration were quite different in the two eras, surely shaping the broad relationship between development and migration.

This section uses census data from migrant-destination countries to quantify the relationship between economic development and emigration during 1850–1914. It uses the same methods as in the previous analysis of the 1960–2019 period. The sole substantive difference arises from the fact that a complete global matrix of bilateral migrant stocks before 1914 does not exist. The analysis must rely instead on emigrant stocks recorded by censuses in the three most im-

Figure 8: Pre-1914 Emigrant Stocks: European emigrant stocks versus origin-country income per capita, three principal overseas destination countries, 1850−1914

(a) *Pooled samples:*

United States, 1850−1910   Canada, 1871−1911   Argentina, 1869−1914

(b) *Within-country changes only, starting below PPP$1,500/capita:*

United States, 1850−1910   Canada, 1871−1911   Argentina, 1869−1914
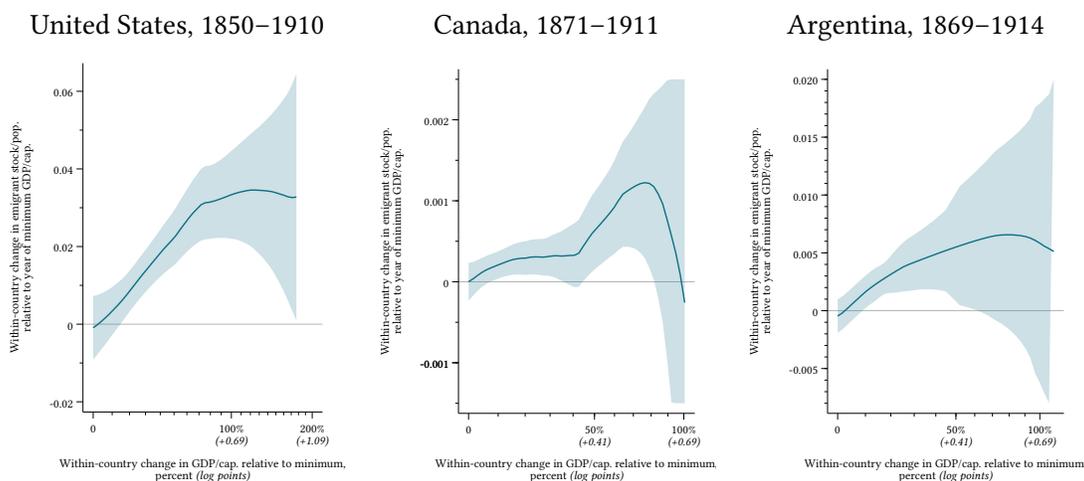
Observations in each destination-country panel are by origin country and year. U.S.: 211 observations, 32 origin countries, seven census years (1850, 1860, 1870, 1880, 1890, 1900, 1910), bandwidth 0.285 in (a). Canada: 156 observations, 35 origin countries, five census years (1871, 1881, 1891, 1901, 1911), bandwidth 0.334 in (a). Argentina: 57 observations, 21 origin countries, three census years (1869, 1895, 1914), bandwidth 0.360 in (a). Solid nonparametric regression line is a Fan (1992) local-linear regression, Epanechnikov kernel; optimal bandwidth minimizes conditional weighted mean integrated squared error. Dashed line in (a) is linear OLS fitted to the same data. Shaded area is 95% confidence interval, clipped for graph legibility. Part (b) shows change in emigrant stock at each destination versus change in GDP per capita at the origin in subsequent years starting from the year of minimum GDP per capita, and includes only migrant-origin countries whose GDP per capita (2011 PPP$) was below $1,500 at some point during the period of observation for each destination country. U.S.: 69 observations, 11 origin countries, bandwidth 0.204 in (b). Canada: 56 observations, 13 origin countries, bandwidth 0.091 in (b). Argentina: 15 observations, 6 origin countries, bandwidth 0.250 in (b).

portant destination countries for extra-European emigration by Europeans during this period. This precludes a truly global comparison of the two eras, omitting in particular pre-1914 Chinese, Japanese, and Indian emigration to Southeast Asia, Africa, and Latin America outside of Argentina. The approach here does, however, have the advantage of comparing two migration flows that have been most prominent in the literature on the two migration eras.

The three destination countries considered here—the United States, Canada, and Argentina—capture the vast majority of mass emigration from the Old World during this period. The United States alone was the destination of roughly two-thirds of these migrants (Caruana 2015), and the three destinations collectively accounted for 81 percent of this mass emigration (Baines 1995, 2). Because emigrant stocks were measured at different times in the three destinations, the analysis here is separated by destination rather than pooled. Historical estimates of GDP per capita (in 2011 PPP-adjusted dollars) are taken from the Maddison Project (Bolt et al. 2018), as is standard in the literature (e.g. Taylor and Williamson 1997).

The first step is to re-create the earlier Figure 1, covering 1960–2019, for the years before 1914 and for each destination. Panel (a) of Figure 8 shows these regressions. The first column shows a pooled nonparametric regression of the emigrant stock in the United States as a fraction of the origin country population, on real GDP per capita in the origin country, with seven decennial observations over the period 1850–1910. The superimposed dashed line shows a linear OLS regression fitted to the same data. The second column shows the same regression using the five Canadian censuses over the period 1871–1911, and the third column uses the three Argentinian censuses over the period 1869–1914.

The emigration life cycle is evident. Moreover, there is a striking resemblance between the pre-1914 curves in Figure 8 and the post-1960 curve in Figure 1. Emigration prevalence in the "middle-income" countries of the pre-1914 era was an order of magnitude higher than emigration prevalence in the poorest countries. But above a turning point, relatively richer countries exhibited relatively lower emigration prevalence. Even the absolute height of the curves is similar across history: Summing the heights of the curve peaks in panel (a) of Figure 8 gives roughly 8 percent of the origin-country population going to the three destinations. Since these destinations cover 81 percent of mass emigration during this period, a fuller accounting might yield a

Figure 9: PRE-1914 EMIGRANT STOCKS: Bilateral stocks versus GDP per capita, 1850–1914
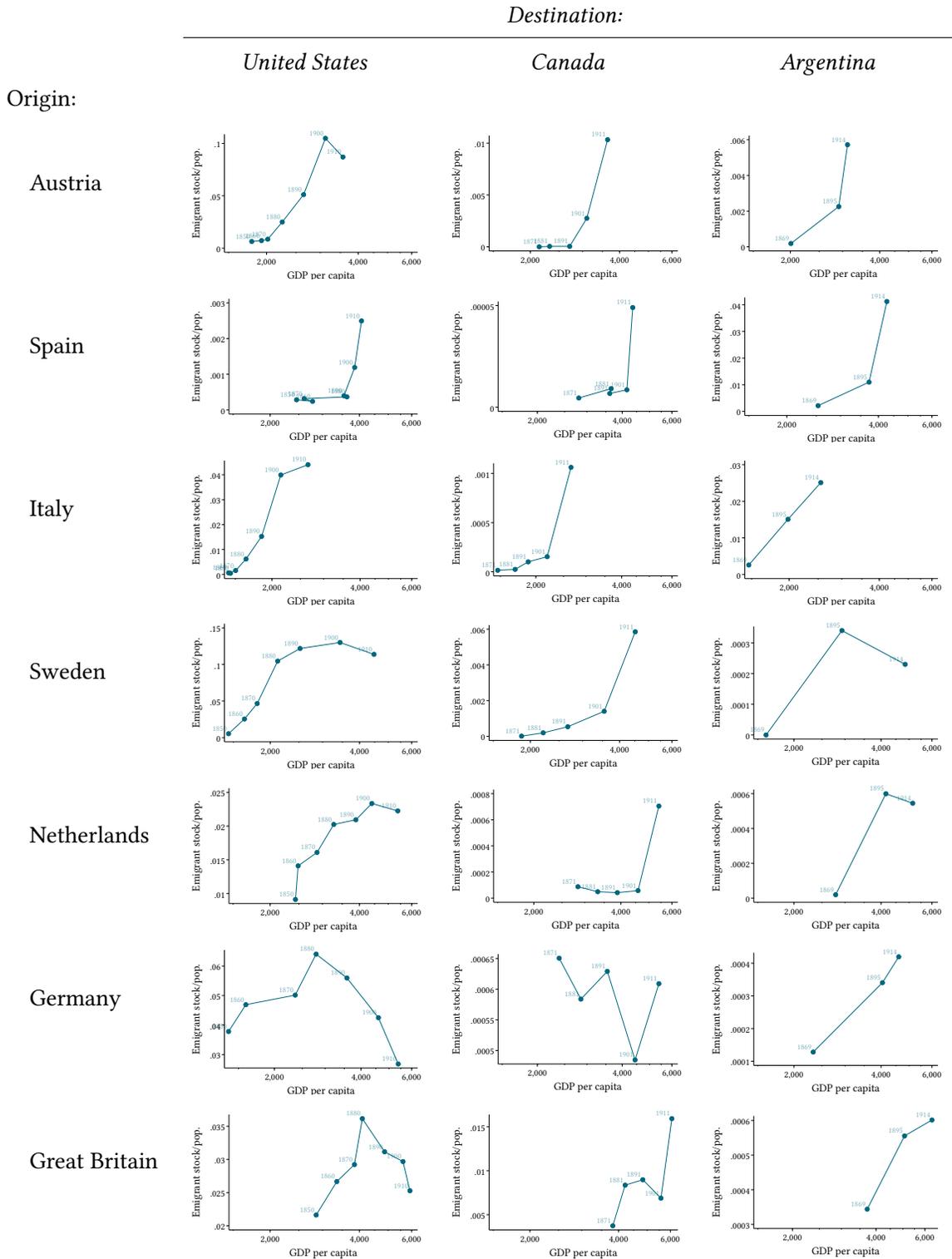
peak of very roughly 10 percent of the origin-country population going to all destinations at the peak of the curve pre-1914. The height of the post-1960 curve peak in Figure 1a is 11 percent.

The principal difference between the life-cycle curves in the two historical eras is that the turning point came at a lower absolute GDP per capita in the pre-1914 data—around PPP$3,000. This is predicted by theory in the location choice equation (3) given that the richest countries in the world at that time were in the range of PPP$5,000–PPP$8,000.

Panel (b) of Figure 8 utilizes exclusively within-origin-country variation in GDP per capita and emigrant stocks, re-creating the post-1960 Figure 4 with pre-1914 data. The sole difference is that the arbitrary definition of a "poor" country is shifted from PPP$2,000 GDP per capita to $1,500, to imperfectly account for the fact that the whole world was poorer before 1914 than after 1960. The figure shows only countries that were "poor" by this definition at some point during the period of observation, so that their minimum GDP per capita fell below PPP$1,500. The horizontal axis shows positive growth from that minimum income level in subsequent years. The vertical axis shows the change in emigrant stock relative to its value in the year of minimum income.

The regressions in panel (b) of Figure 8 show that the general shape of the emigration life cycle before 1914 is robust to controlling for any time-invariant unobserved differences between countries. As in the post-1960 data, the magnitude of the positive relationship early in development is substantially smaller using within-country variation only than when pooling within- and between-country variation. The height of the curves in panel (b) is roughly half of the height in panel (a). Recall that the slope of the emigration life cycle after 1960 was about one-third lower when estimated using country fixed effects. Here, again, the life cycle is similar in the two eras.

The results of these nonparametric regressions are not driven by outlier countries. Figure 9 graphs the raw data for several countries that are representative of mass emigration during the pre-1914 era, corresponding to the post-1960 graphs in the earlier Figure 5. Emigration prevalence rose along with average incomes in Sweden, Italy, and Spain. It did so as well in Great Britain, Germany, the Netherlands, and Austria, before hitting a turning point around PPP$3,000–PPP$4,000, above which emigration prevalence declined as incomes in the migrant

Figure 10: PRE-1914 NET EMIGRATION FLOWS: Changes in European emigrant stocks versus origin-country income per capita, three principal overseas destination countries, 1850–1914

(a) *Pooled samples:*
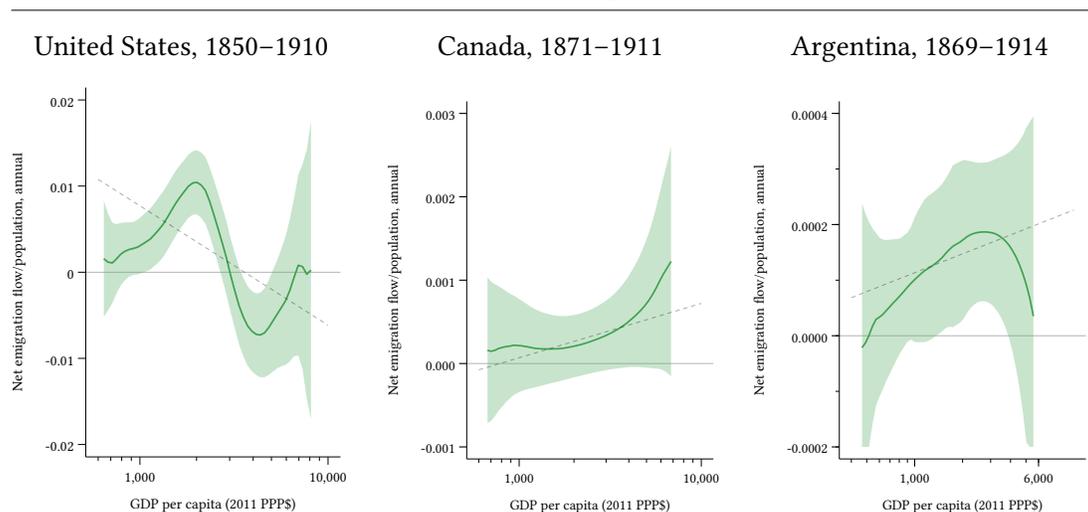
(b) *Within-country changes only, starting below PPP$1,500/capita:*
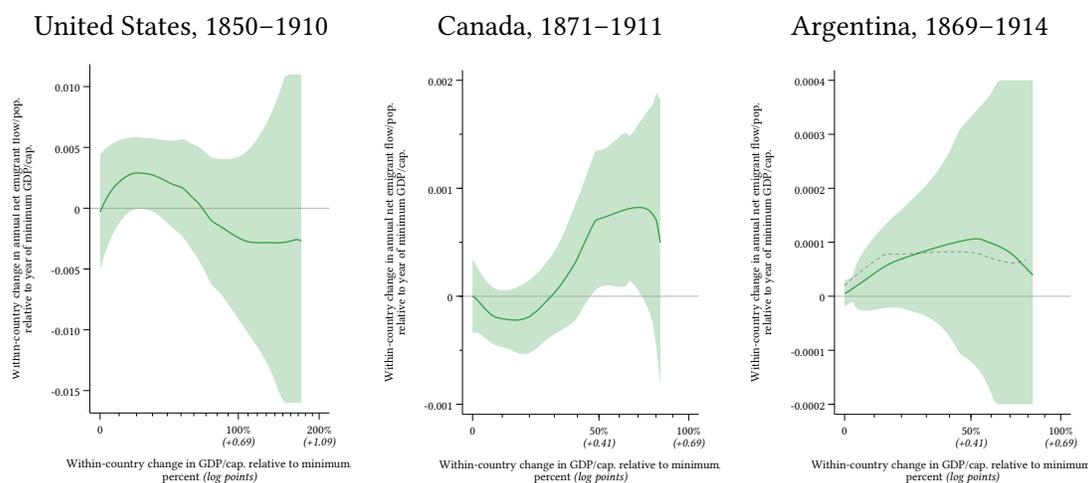
Observations in each destination-country panel are by origin country and year. Horizontal axis shows GDP per capita of origin country, and vertical axis shows annualized change in net emigration flow from each origin to each destination during the *subsequent* period. U.S.: 179 observations, 32 origin countries, six periods (1850–60, 1860–70, 1870–80, 1880–90, 1890–1900, 1900–1910), bandwidth 0.212 in (a). Canada: 118 observations, 34 origin countries, four periods (1871–81, 1881–91, 1891–1901, 1901–11), bandwidth 0.542 in (a). Argentina: 36 observations, 21 origin countries, 2 periods (1869–1895, 1895–1914), bandwidth 0.413 in (a). Part (b) shows change in emigrant flow at each destination vs. change in GDP per capita at the origin, in subsequent years starting from the year of minimum GDP per capita, and includes only migrant-origin countries whose GDP per capita (2011 PPP$) was below $1,500 at some point during the period of observation for each destination country. U.S.: 60 observations, 10 origin countries, bandwidth 0.200 in (a). Canada: 42 observations, 13 origin countries, bandwidth 0.092 in (b). Argentina: 9 observations, 5 origin countries, bandwidth 0.200 in (b). In (b), for Argentina only, sample too small to compute confidence interval for the true regression line (dashed); confidence interval shown is for all countries (not just those starting below $1,500/capita), the solid line, which which follows a similar trajectory.

origin countries converged with incomes at the destination. This nonmonotonicity has been well documented (O'Rourke et al. 1996; Taylor and Williamson 1997). But its close quantitative correspondence to the patterns after 1960 has not previously been documented.

The analysis concludes with two robustness checks using alternative specifications. First, as before, the analysis so far speaks to two relationships: the relationship between emigration *stocks* and GDP per capita *levels* (Figure 8, panel a) both across and within countries, and the relationship between emigration *flows* and economic *growth* (Figure 8, panel b and Figure 9) within countries. An alternative specification that might be of interest would test the relationship between emigration *flows* and the *level* of GDP per capita, as was done with post-1960 data in Figure 6. That analysis is re-created using pre-1914 data in Figure 10. Here again, emigration flow rates initially rise at higher and higher levels of economic development, both in pooled data (panel a) and using exclusively within-country growth for once-poor countries (panel b).

The same pattern is evident in graphs of the raw data for several major migrant-origin countries (Figure 11). Emigration flow rates typically rose along with initial economic development in Austria, Spain, Sweden, and Germany. The flow rate declined only as the origin countries became richer than a certain threshold. The only settings where the initial rise is not evident are in corridors (Great Britain–US, Netherlands–US) where emigration rates had already risen to high levels before the time window of observation begins, or where the rise is obscured by low-frequency data collection in Argentina.[6]

The second robustness check is performed by using origin-country data on emigration. It is possible, in principle, that the use of a limited number of destination countries reduces the comparability of the results before 1914 with the results after 1960. Perhaps the relationship between origin-country development and emigration to other destinations was somehow different in the two historical eras. So it is important to check whether, in limited settings where measures of net emigration to *all* destinations are available, quantitatively similar results hold. Figure 12 shows the relationship between origin-country income per capita and net emigration flows for the four

---

[6]For example, Swedish and Dutch emigration to Argentina was extremely low at the beginning of the period 1869–1895, so if a census had been conducted in Argentina around 1880, the figures would reveal a rising-then-falling pattern. But because the entire 1869–1885 period is aggregated, and there was no census before 1869, the initial rise of the emigration rate with GDP per capita is obscured.

Figure 11: PRE-1914 NET EMIGRATION FLOWS: Changes in bilateral stocks, 1850–1914
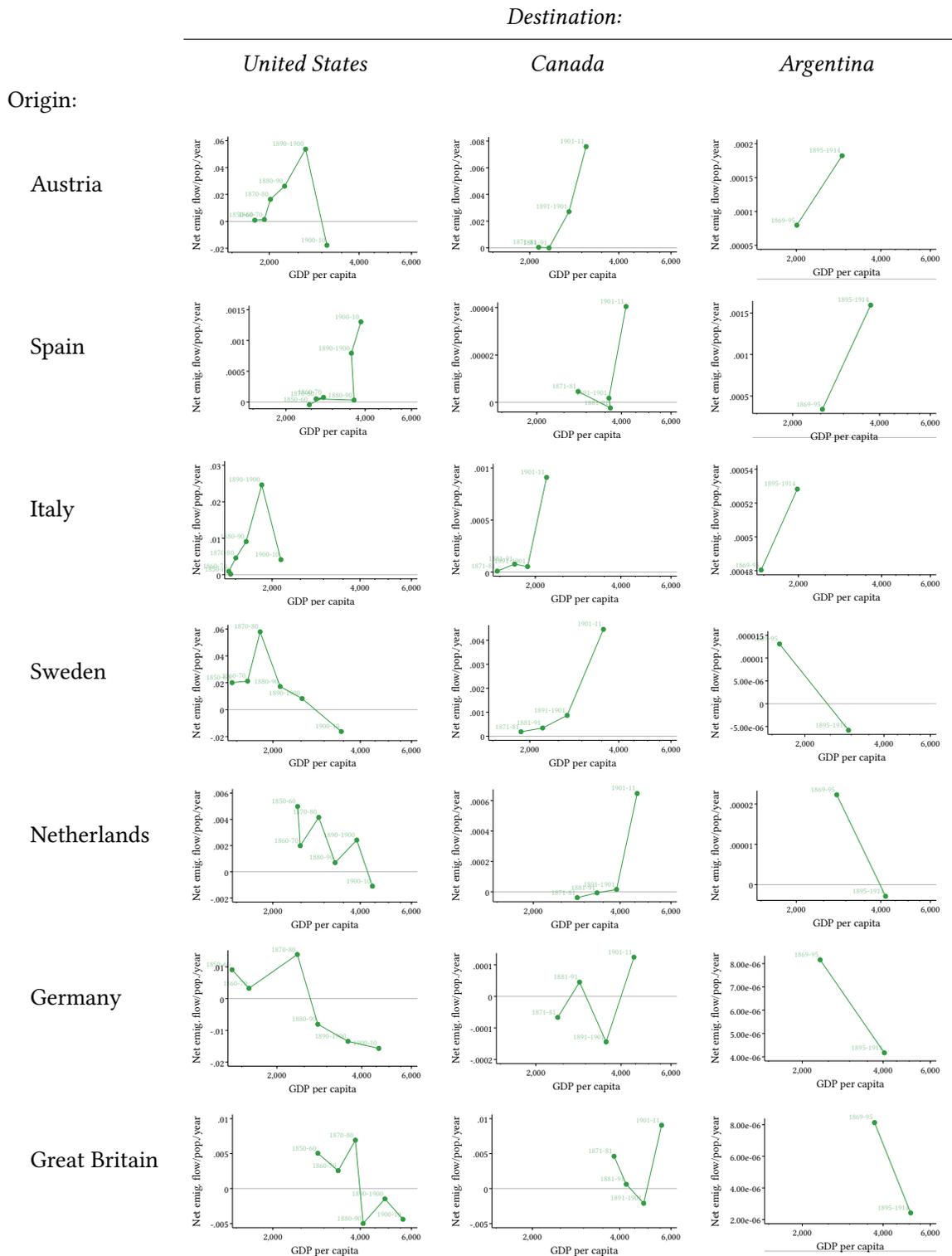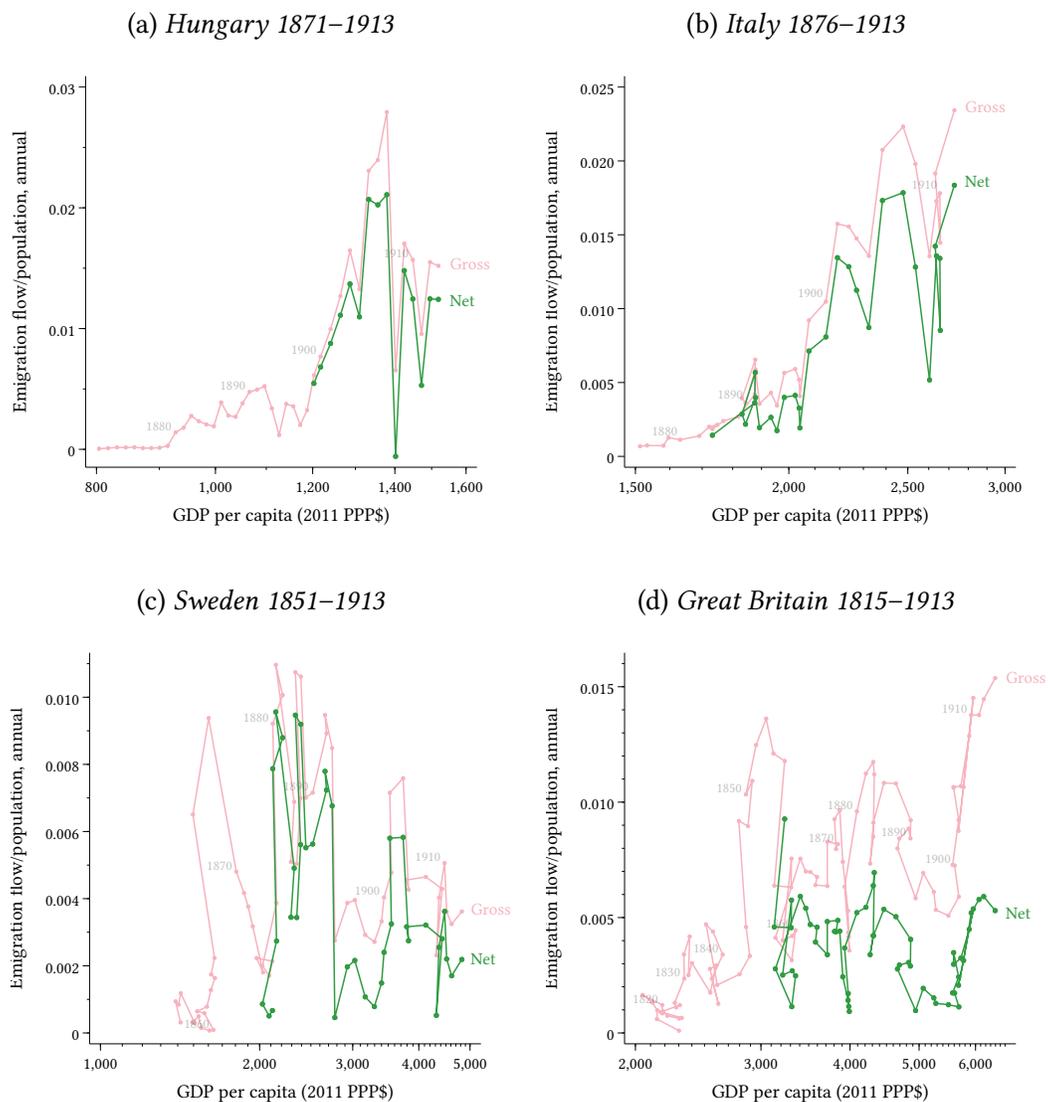
Figure 12: PRE-1914 NET EMIGRATION FLOWS: Net emigration flows versus origin-country income per capita, four origin countries to all extra-European destinations

(a) *Hungary 1871–1913*

(b) *Italy 1876–1913*

(c) *Sweden 1851–1913*

(d) *Great Britain 1815–1913*

Emigration flows to all extra-European destinations from Ferenczi (1929), details in Appendix.

European countries where the principal source of historical statistics (Ferenczi 1929) records *net* emigration flows for a substantial portion of the pre-1914 years: Hungary, Italy, Sweden, and Great Britain. In all other European countries during this period, either reliable emigration statistics were not collected by Ferenczi (1929), or they were collected without information on reverse flows that would allow calculation of net emigration. The figure shows net emigration flows to all extra-European destinations collectively.

The prior results are robust to this further change of specification. Emigration to all extra-European destinations rose sharply along with economic development in countries that started out relatively poor: Hungary and Italy in the late 19th century, as Faini and Venturini (1994) described for Italy. Emigration first rose, and then fell with economic development in countries that became relatively rich during this period: Great Britain 1815–1913, and Sweden 1851–1913. For three countries (Sweden, Germany, and Great Britain) it is possible to directly compare the evolution of the destination census-based flow rates to the United States (Figure 11) with the origin port-based flow rates to all extra-European destinations (Figure 12), and the results are very similar in all three.

This exercise suggests that the emigration life cycle is an empirical regularity of developing economies that is robust across centuries. The cycle between 1850 and 1914 was qualitatively and quantitatively similar to the cycle between 1960 and 2019. The principal difference between the two eras is that the turning point, past which emigration fell with further economic development, was lower in the earlier era when all nations were poorer than in the more recent era.

5   Decomposing mechanisms for the life cycle

The tools are now in place to shed light on suggestive mechanisms for the emigration life cycle. Comparing the two eras contains information about the relative role of different mechanisms that have been important in the literature: human capital accumulation, demographic change, structural change, technological change, and policy barriers.

The evidence so far is not compatible with a model implying that the emigration life cycle has been largely determined by advances in transportation technology. These helped transportation

costs plummet in the late 19th century. In principle, this alone could have led emigration rates to rise as then-poor countries developed. But although transportation costs did fall in the late 20th century as well, the magnitude of that decline was nowhere near comparably large (e.g. Shah Mohammed and Williamson 2004; Hummels 2007). So if transportation costs were a first-order mechanism for the life cycle historically, the pattern might have dissipated after World War Two. It did not.

Beyond this, the post-1960 life cycle is robust to controlling for year fixed effects, which should absorb the effects of global advances in transportation technology. Including year effects does dampen the magnitude of the life cycle, but as discussed above, that reduced magnitude should be seen as a lower bound on the true magnitude because year effects absorb roughly half the variance in within-country economic growth. The evidence is compatible with a nonzero role for changing transportation costs in generating the life cycle pattern, but not compatible with a predominant role.

The evidence is also not compatible with models implying a large role for policy barriers to migration after 1960. Before 1914, none of the countries of destination in Figure 8 imposed substantial policy restrictions on migrants from any of the countries of origin considered there. If policy barriers against low-income countries were a first-order determinant of the emigration life cycle after 1960, the pattern should have been weak or absent before 1914 (Hatton and Williamson 2011, 21). It was not.

Some of the other posited channels allow more direct tests, both between and within the two historical eras. Table 3 estimates the relative importance of three other channels using a Gelbach (2016) decomposition of the coefficient on ln GDP per capita, in both eras. It quantifies how much of the coefficient on ln GDP per capita in a linear regression with country fixed effects can be explained by differences in human capital accumulation, demographic structure, or rural-urban structure.

The most basic indicators for each of these changes are chosen so that comparable measures are available for many countries over the entire period 1850–2019. Human capital accumulation by working-age people is proxied by the net secondary-school enrollment rate lagged by one

37

Table 3: POST-1960 VS. PRE-1914 EMIGRATION: Decomposing the Life Cycle in the two eras

Dep. var:	Emigrant stock/pop.				ln(Emigrant stock/pop.)			
Country effects:	Fixed		Fixed		Fixed		Fixed	
	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
Emigration to all countries, 1960–2019, origin-country GDP/capita < PPP$10,000, constant sample, 81 countries, N = 445								
ln(GDP/capita), t	0.0126	(0.00563)	−0.00893	(0.00362)	0.361	(0.0959)	−0.134	(0.106)
Net secondary enrollment, $t-10$			0.000538	(0.000243)			0.00771	(0.00374)
Total fertility rate, $t-20$			−0.00675	(0.00343)			−0.140	(0.0459)
Child mortality rate, $t-20$			0.0000251	(0.0000557)			−0.000600	(0.00108)
Urbanization rate, t			0.000198	(0.000461)			0.00427	(0.00839)
Gelbach decomposition								
Δ ln(GDP/capita) coeff.			0.0248	(0.00742)			0.446	(0.108)
…Education			0.00617	(0.00648)			0.125	(0.0908)
…Youth			0.0105	(0.00858)			0.188	(0.152)
…Urbanization			0.00807	(0.00784)			0.133	(0.123)
R^2	0.039		0.197		0.086		0.275	
Emigration to the United States, 1850–1910, origin-country GDP/capita < PPP$3,000, constant sample, 30 countries, N = 141								
ln(GDP/capita), t	0.0606	(0.0137)	0.0286	(0.0121)	4.35	(0.711)	3.11	(0.974)
Net secondary enrollment, $t-10$			−0.00216	(0.00173)			0.106	(0.0914)
Total fertility rate, $t-20$			0.000182	(0.00228)			−0.942	(0.200)
Child mortality rate, $t-20$			−0.000102	(0.0000613)			0.000397	(0.00330)
Urbanization rate, t			0.00367	(0.00110)			0.0467	(0.0722)
Gelbach decomposition								
Δ ln(GDP/capita) coeff.			0.0320	(0.00801)			1.24	(0.828)
…Education			−0.00758	(0.00859)			0.373	(0.353)
…Youth			0.0042	(0.00276)			0.417	(0.224)
…Urbanization			0.0354	(0.0128)			0.450	(0.790)
R^2	0.376		0.518		0.506		0.589	

Standard errors in parentheses, clustered by country, to the right of each coefficient estimate. Constant term omitted in table. Coefficient estimates in this table (restricted to constant sample with observations for all covariates) need not match estimates in Table 1 (full sample). R^2 is for within-country variance over time. School enrollment and urbanization are measured in percent, fertility is measured in number of children per woman, under-five child mortality is measured per 1,000 live births. For the 2019 observation only, the one-decade lag is proxied by the 2010 value ($t-9$) and the two-decade lag is proxied by the 2000 value in 2000 ($t-19$).

decade (Lee and Lee 2016). Changes in the size of the working-age youth cohort are proxied by the total fertility rate and the child mortality rate, each lagged by two decades (following Horiuchi and Preston 1988; Macunovich 1999, 2000; Hatton 2001; Fargues 2011, 599). The rural-urban structure of the economy is proxied by the percent of the population living in urban areas. The structural transformation from a rural agricultural economy toward a more urban economy is one of the strongest and most universal features of economic development (Timmer 2009).[7]

The estimates in Table 3 imply that all three of these channels shaped the emigration life cycle in both eras, with suggestive evidence that the demographic channel was somewhat more important than the other two channels. In the 1960–2019 regressions at the top of the table, including the covariates in the regression is sufficient to reverse the sign of the coefficient on ln GDP per capita, whether the emigration rate is specified linearly or logarithmically. That is, the three channels collectively can fully account for rising emigration rates early in the process of economic development.

The signs on all covariates in the multivariate regressions are as posited by the literature: There is a statistically significant positive relationship with schooling, observed in cross-section as a major channel by Dao et al. (2018), in these panel fixed-effects regressions. There is a negative and highly significant coefficient on the fertility rate, as suggested by Macunovich (1999, 2000): Initially high fertility is absorbed by the country fixed effect, and the negative coefficient reflects the rise in youth cohort size that typically occurs two decades after the sustained negative *change* in fertility begins. Finally, there is a positive and substantial but statistically imprecise coefficient on urbanization.

In either the linear or log specification, the magnitude of the demographic channel is slightly larger, but similar in size to and statistically indistinguishable from the other two channels. As discussed above, the logarithmic results are preferred because they allow for a fixed-elasticity effect of prior diaspora size on subsequent migration flows to be absorbed into the country fixed effect.

---

[7]The regressions are restricted to a constant sample for each historical period, so that changes in the coefficient on ln GDP per capita arise exclusively from changing the regression specification. This has the side effect that the bivariate regressions in Table 1 on an unrestricted sample are not identical to the bivariate regressions in Table 3, though the magnitude of the coefficient estimates is similar.

The results are remarkably similar in the 1850–1914 regressions, with the United States as migrant-destination country, at the bottom of the table. Again, the three channels explain a large portion of the pre-1914 life cycle, though not the entire life cycle as seen after 1960. Before 1914, the education, demographic, and urbanization channels explain roughly half (linear) or one third (log) of the positive coefficient on ln GDP per capita. In the preferred logarithmic specification, the relative magnitudes of the three channels are comparable, though the estimates for education and urbanization are far from statistically precise.

The absolute elasticity of emigration to rising income 1850–1914 was much larger than after 1960–2019. This could arise in part from the fact that the richest destination countries after 1960 were far richer than the richest destination countries before 1914. The ln GDP per capita gap between poor origins and rich destinations expanded from a factor of roughly 7 in the late 19th century (2.0 natural log points) to a factor of roughly 50 in the late 19th (3.9 natural log points). All else equal, it took a migrant-origin country less economic growth in the earlier era—in relative terms and in absolute dollars—to arrive at a level where the payoff to migration diminished. This would tend to compress the entire life cycle pattern horizontally in the earlier era relative to the later era, raising the relative magnitude of the coefficient on ln GDP per capita.

Overall, the evidence in Table 3 suggests striking similarities between the emigration life cycle across historical eras. In both eras, the life cycle is strongly present using within-country variance only. In both eras, a large portion of the life cycle is explained by changes in demographic structure, the structure of national production, and human capital accumulation over the course of economic development—in roughly comparable measures. In both eras, there is suggestive evidence that the strongest channel is demographic change: A surge in the size of the youth cohort can increase competition for employment (the Easterlin (1978) hypothesis), driving emigration for employment. This is consistent with the importance of demographic change to the global relationship between economic growth and emigration, earlier documented for the pre-1914 era by Hatton (2001) and for Latin America in more recent years by Hanson and McIntosh (2010, 2012, 2016). It is also consistent with the importance of the education channel in time series that is observed in cross section by Dao et al. (2018).[8] These factors along with shifts in the structure

---

[8]This decomposition can only be considered an indicative decomposition of partial mechanisms, rather than a precise decomposition of total mechanisms, due to strong causal relationships among the tested components. For example, Hatton (2001) do not find a strong role for education—proxied by adult literacy—in the emigration life cycle.

of produciton have helped generate the emigration life cycle for the past two centuries.

6 Flawed approaches

The evidence in section 3 indicates that after 1960, emigration rates rose as GDP per capita rose in the average, relatively poor developing country, with an elasticity around +0.3. A recent study nevertheless estimated this elasticity as −0.5 (Benček and Schneiderheinze 2019, Table 3), for developing countries in general and the poorest in particular. The study arrives at this figure using annual estimates of gross, regular emigration flows in a three-decade panel (for most countries, running from the mid-1980s to 2015). It describes this negative coefficient as "the true relationship between economic development and emigration" and attributes positive coefficients elsewhere in the literature to a lack of "rigorous fixed effects panel estimations that exploit the variation over time" (Benček and Schneiderheinze 2019, 1, 17).

This finding is difficult to reconcile with another finding of the same study: that average emigration rates in the poorest countries are already close to zero. It finds that in countries around PPP$500 GDP per capita, observed migration is just 0.3 percent of the population per decade, or 0.3 emigrants per 1,000 population per year (Benček and Schneiderheinze 2019, Figure 3). If correct, this finding would imply that in today's poorest countries (e.g. GDP per capita PPP$500)—if they grew into into middle-income countries (e.g. GDP per capita PPP$7,000)—the average emigration rate would only fall even closer to zero, collapsing to just 0.08 percent of the population per decade.[9] But the observed emigration rate in countries that actually reach PPP$7,000 per capita is 37 times higher than this, the same study finds: about 3 percent per decade. A disconnect of this magnitude requires one of two things. It requires either an explanation of how it is that today's poor countries differ so radically from the countries that have preceded them in economic development—which the study does not offer—or scrutiny of the methods underlying the estimate.

But since other components in the analysis are both cause and effect of rising education, such as urbanization and demographic transition, a decomposition of the 'total' effect of education rather than the partial effect is not possible in this empirical framework. Thus it would be incorrect to infer that the present findings contradict other findings of no *partial* role for education pre-1914. Rather, this study finds a substantial role for education above and beyond the (substantial) portion of the variance in education that is explained by demographic change and urbanization, while Hatton (2001) do not. But both studies are compatible with a substantial role for education in total causation.

[9] $\exp(\ln 0.3 + (-0.5) \times (\ln 7,000 - \ln 500)) = 0.08$.

In fact, the methods used in the Benček and Schneiderheinze study are incorrect, and generate the negative coefficient spuriously. The methods do not and cannot measure the effect of economic development on emigration, but instead measure the effect of short-run income shocks on emigration. In brief, the error in the study is a special case of the "spurious regressions" problem, first described by Yule (1926) and named by Granger and Newbold (1974): The study's method fails to account for the fact that more than one variable in its regressions is nonstationary, giving a spurious result.

The study runs various versions of the regression

$$\ln M_t = \alpha + \beta \ln Y_t + \delta \ln L_t + \varepsilon_t, \tag{4}$$

where M_t is the absolute number of gross regular emigrants from a country of origin in year t; Y_t is real GDP; L_t is population; α, β, and δ are coefficients to be estimated, and ε is an error term. In annual panel data on countries, the study estimates that $\hat{\beta} = -0.5$ (Benček and Schneiderheinze 2019, Table 3), and interprets this as the effect of "economic development" on emigration.

But this coefficient estimate represents exclusively the effect of short-run shocks, not "economic development", which requires sustained long-term rises in income. Intuitively, this is because the coefficient on the logarithm of GDP is estimated after partialing out the effect of the logarithm of population, shifted by a constant, on GDP. But sustained growth in GDP per capita (that is, economic development) means that on average, increases in GDP *are* a multiple of increases in population. Partialing out the effect of log population and a constant means absorbing away all long-run changes in GDP per capita. The method eliminates the entire effect that it seeks to measure.

More formally, the problem may be summarized by observing that because both population and GDP are nonstationary, controlling for the nonstationary component of population absorbs the nonstationary component of GDP—that is, development. Population is well described by $\ln L_t = \ln L_0 + \pi t$, where L_0 is initial population and π is the population growth rate. GDP is exactly described by $\ln Y_t = \ln Y_0 + (\pi + \gamma)t + u_t$, where Y_0 is initial GDP, γ is the long-run growth rate of GDP *per capita*, and u_t is a short-run shock. Combining these two expressions gives $\ln Y_t = \Theta + \left(1 + \frac{\gamma}{\pi}\right)\ln L_t + u_t$, where $\Theta \equiv \ln Y_0 - \left(1 + \frac{\gamma}{\pi}\right)\ln L_0$. Substituting this into

regression (4) gives

$$\ln M_t = \left[\alpha + \beta\Theta\right] + \beta \cdot u_t + \left[\beta\left(1 + \frac{\gamma}{\pi}\right) + \delta\right] \cdot \ln L_t + \varepsilon_t. \tag{5}$$

The coefficient on ln GDP in regression (4) is identical to the coefficient one would obtain by regressing the same migration flows on short-run *shocks* to ln GDP (that is, u_t) in regression (5), which controls for ln population and includes a constant term. But such annual deviations from the long-term trend of GDP per capita contain no information about the overall level of economic development, because the time trend in GDP has been removed. This is a special case of Granger and Newbold's spurious regressions problem because, as the above derivation makes clear, the problem would not arise if either GDP or population were stationary.

The Appendix shows how correcting this error in the Benček and Schneiderheinze estimate yields a positive relationship between emigration flows and sustained growth GDP per capita similar in magnitude to the fixed-effects estimates in Table 1 above. It also shows how to reproduce the spurious negative coefficient with real data from a quintessential case of rising emigration in early stages of economic development: Mexico-US migration after 1960.

7   Conclusion

These results confirm the existence of the emigration life cycle in developing countries. Emigration rises on average as low-income countries develop into middle-income countries, then falls as they further develop into high-income countries. This is a process that occurs in average developing countries over time, not simply a pattern across countries in cross-section, though about one-third of the height of the life-cycle curve in pooled data is explained by unobserved country heterogeneity.

The emigration life cycle is observed in the relationship between income levels and emigrant stocks, the relationship between income growth and emigration flows, and the relationship between income levels and emigration flows. It is evident in emigration to all destinations as well as rich-country destinations specifically. It is evident in large developing countries as well as small ones. It is robust to controlling for country-invariant heterogeneity between time periods.

And it has been robust across centuries.

The analysis confirms a striking quantitative correspondence between the shape of the emigration life cycle in the first era of globalization (1850–1914) and the second (1960–present). In both eras, key channels for the life cycle were broadly similar: demographic change was central in both eras, with important contributions from rising human capital investment and urbanization. Emigration from developing countries has been encouraged by the fundamental forces that are central to development itself.

Put differently, the data reveal an Engel curve for emigration at the country level—and that curve is nonmonotonic, as the economic history literature has long proposed. At early stages of development, potential emigrants in low-income countries on average treat emigration as a normal good: As incomes rise, they spend more on the investments that complement migration. Only at later stages of development do they begin to treat emigration as an inferior good, reducing those investments as incomes rise further.

These correlations have a strong claim to be interpreted as causal, in one sense but not another. Specifically, they have some claim to be interpreted as *necessary but not necessarily sufficient* causal relationships between economic growth and emigration. That is, the emigration life cycle pattern is so robust that it seems to reflect very common and important features of the development process. The empirics discussed here do not allow the relationship to be decomposed into a *sufficient* causal component (*for poor countries on average, an increase in growth all else equal causes an increase in emigration*) and a *necessary* component (*third factors that are on average necessary to growth takeoffs in poor countries, such as urbanization, also cause an increase in emigration*). Either of these means that for average poor countries, emigration goes hand in hand with greater economic growth, but neither can be ruled out as the principal causal channel.

What can plausibly be ruled out is that the relationship is driven by reverse causation—that a principal driver of economic growth in the average poor country is emigration itself. To be sure, the literature suggests important channels of reverse causation from emigration to the determinants of growth in the country of migrant origin (e.g. Barsbai et al. 2017; Karadja and Prawitz 2019). But among the channels in the literature, none would imply a magnitude of reverse cau-

sation anywhere near capable of generating the magnitude of the correlations estimated here. Migrant remittances, for example, do not have a clearly established, systematic effect on economic growth (e.g. Clemens and McKenzie 2018).

The existence of the emigration life cycle represents an uncommon instance of strong agreement among the various social science disciplines that have considered migration and development. The life cycle pattern, as reviewed above, has been found in development economics and economic history (summarized by Williamson 2015). Beyond that, in sociology, Massey (1989, 1) concludes that "emigration is a normal by-product of economic development," and de Haas et al. (2019, 893–894) find that "higher levels of economic and human development …are initially associated to higher levels of emigration" while "[o]nly when countries achieve higher development levels does emigration decrease." In geography, Skeldon (2008, 8) concludes, "Migration is primarily a consequence of development, no matter how defined." In history, Gozzini (2006, 330) concludes, "International migrants do not hail from poor and isolated circumstances with no links to the world's market places, but rather from regions and countries experiencing rapid change and development on account of their integration into the global trade, information, and production networks." Clemens (2014) reviews four decades of similar findings from these and other disciplines including urban studies and anthropology. The present analysis corroborates that large literature from a global, quantitative standpoint spanning modern history and tracing within-country evolution over time.

Fruitful further investigation of the emigration life cycle might involve detailed country studies clarifying the channels by which sustained economic development, at first, unlocked larger emigration flows in South Korea, Mexico, Indonesia, and other countries where sustained growth has occurred in tandem with large waves of emigration. The mechanisms for household-level Engel curves are only starting to be understood (e.g. de Vreyer, Lambert and Ravallion 2020). The determinants of variance around the average life cycle—why the curve is steeper or shallower for different countries—are poorly understood. The disaggregated channels for the national-level Engel curve for emigration likewise require more inquiry (e.g. Dao et al. 2018), clarifying how the shape of the curve depends on country traits and overseas policy.

References

Abel, Guy J, "Estimates of global bilateral migration flows by gender between 1960 and 2015," *International Migration Review*, 2018, *52* (3), 809–852.

Adhikari, Samik and Ugo Gentilini, "Should I Stay or Should I Go: Do Cash Transfers Affect Migration?," Policy Research Working Paper 8525. Washington, DC: World Bank 2018.

Akerman, S, "Theories of migration research," in Harald Runblom and Hans Norman, eds., *From Sweden to America: A History of the Migration*, Minneapolis: University of Minnesota Press, 1976.

Angelucci, Manuela, "Migration and financial constraints: Evidence from Mexico," *Review of Economics and Statistics*, 2015, *97* (1), 224–228.

Bade, Jan and Antonie de Kemp, "Development and Migration," Policy and Operations Evaluation Department (IOB), The Hague: Ministerie van Buitenlandse Zaken 2018.

Bah, Tijan L and Catia Batista, "Understanding Willingness to Migrate Illegally: Evidence from a Lab in the Field Experiment," NOVAfrica Working Paper 1803. Lisbon: Universidade Nova de Lisboa 2018.

Bahar, Dany and Hillel Rapoport, "Migration, knowledge diffusion and the comparative advantage of nations," *Economic Journal*, 2018, *128* (612), F273–F305.

Baines, Dudley, *Emigration from Europe, 1815–1930*, Cambridge: Cambridge University Press, 1995.

Bairoch, Paul and Gary Goertz, "Factors of urbanisation in the nineteenth century developed countries: a descriptive and econometric analysis," *Urban Studies*, 1986, *23* (4), 285–305.

Barsbai, Toman, Hillel Rapoport, Andreas Steinmayr, and Christoph Trebesch, "The Effect of Labor Migration on the Diffusion of Democracy: Evidence from a Former Soviet Republic," *American Economic Journal: Applied Economics*, July 2017, *9* (3), 36–69.

Batista, Catia and David McKenzie, "Testing Classic Theories of Migration in the Lab," Working Paper, Universidade Nova de Lisboa and World Bank 2019.

Bazzi, Samuel, "Wealth heterogeneity and the income elasticity of migration," *American Economic Journal: Applied Economics*, 2017, *9* (2), 219–255.

Bellemare, Marc F and Casey J Wichman, "Elasticities and the inverse hyperbolic sine transformation," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 50–61.

Benček, David and Claas Schneiderheinze, "More Development, Less Emigration to OECD Countries: Identifying inconsistencies between cross-sectional and time-series estimates of the migration hump," Kiel Working Paper, No. 2145. Kiel Institute for the World Economy 2019.

Bertoli, Simone, J Fernández-Huertas Moraga, and Francesc Ortega, "Crossing the border: Self-selection, earnings and individual migration decisions," *Journal of Development Economics*, 2013, *101*, 75–91.

Bolt, Jutta, Robert Inklaar, Herman de Jong, and Jan Luiten van Zanden, "Rebasing 'Maddison': New income comparisons and the shape of long-run economic development," Maddison Project Working paper 10. Groningen Growth and Development Centre, Faculty of Economics and Business. University of Groningen 2018.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh," *Econometrica*, 2014, *82* (5), 1671–1748.

Caruana Galizia, Paul, "Explaining Mediterranean Emigration," in "Mediterranean Labor Markets in the

First Age of Globalization: An Economic History of Real Wages and Market Integration," New York: Palgrave Macmillan, 2015, pp. 31–66.

Caselli, Marco, "'Let Us Help Them at Home': Policies and Misunderstandings on Migrant Flows Across the Mediterranean Border," *Journal of International Migration and Integration*, 2019, *20* (4), 983–993.

Chiswick, Barry and Timothy J Hatton, "International migration and the integration of labor markets," in "Globalization in historical perspective," University of Chicago Press, 2003, pp. 65–120.

Clark, Ximena, Timothy J Hatton, and Jeffrey G Williamson, "What Explains Emigration Out of Latin America?," *World Development*, 2004, *32* (11), 1871–1890.

__ , __ , **and** __ , "Explaining U.S. Immigration, 1971–1998," *Review of Economics and Statistics*, 2007, *89* (2), 359–373.

Clemens, Michael A, "Does development reduce migration?," in Robert E B Lucas, ed., *International Handbook on Migration and Economic Development*, Cheltenham: Edward Elgar Publishing, 2014, pp. 152–185.

__ **and David McKenzie**, "Why don't remittances appear to affect growth?," *Economic Journal*, 2018, *128* (612), F179–F209.

__ **and Mariapia Mendola**, "Migration from developing countries: Selection, income elasticity, and Simpson's paradox," CGD Working Paper 539. Washington, DC: Center for Global Development 2020.

Dao, Thu Hien, Frédéric Docquier, Chris Parsons, and Giovanni Peri, "Migration and development: Dissecting the anatomy of the mobility transition," *Journal of Development Economics*, 2018, *132*, 88–101.

de Beer, Joop, James Raymer, Rob van der Erf, and Leo van Wissen, "Overcoming the problems of inconsistent international migration data: A new method applied to flows in Europe," *European Journal of Population/Revue européenne de Démographie*, 2010, *26* (4), 459–481.

de Haas, Hein, "International migration, remittances and development: myths and facts," *Third World Quarterly*, 2005, *26* (8), 1269–1284.

__ , "Turning the tide? Why development will not stop migration," *Development and Change*, 2007, *38* (5), 819–841.

__ , **Mathias Czaika, Marie-Laurence Flahaux, Edo Mahendra, Katharina Natter, Simona Vezzoli, and María Villares-Varela**, "International Migration: Trends, Determinants, and Policy Effects," *Population and Development Review*, 2019, *45* (4), 885–922.

de Vreyer, Philippe, Sylvie Lambert, and Martin Ravallion, "Unpacking Household Engel Curves," Working Paper 26850, National Bureau of Economic Research March 2020.

Djajic, Slobodan, Murat G. Kirdar, and Alexandra Vinogradova, "Source-country earnings and emigration," *Journal of International Economics*, 2016, *99*, 46 – 67.

Drummond, Paulo Flavio Nacif and Gustavo Pérez Ramírez, "Spillovers from the rest of the world into Sub-Saharan African countries," IMF Working Paper WP/09/155. Washington, DC: International Monetary Fund 2009.

Easterlin, Richard A, "What will 1984 be like? Socioeconomic implications of recent twists in age structure," *Demography*, 1978, *15* (4), 397–432.

Faini, Riccardo and Alessandra Venturini, "Italian Emigration in the Pre-War Period," in Timothy J Hatton and Jeffrey G Williamson, eds., *Migration and the International Labor Market 1850–1939*, London: Routledge, 1994, pp. 72–90.

__ **and** __ , "Development and Migration: Lessons from Southern Europe," in Gil S Epstein and Ira N Gang,

eds., *Migration and Culture (Frontiers of Economics and Globalization, Volume 8*, Bingley: Emerald Group Publishing, 2010, pp. 105–136.

Fan, Jianqing, "Design-adaptive nonparametric regression," *Journal of the American Statistical Association*, 1992, *87* (420), 998–1004.

Fargues, Philippe, "International migration and the demographic transition: A two-way interaction," *International Migration Review*, 2011, *45* (3), 588–614.

Feenstra, Robert C, Robert Inklaar, and Marcel P Timmer, "The next generation of the Penn World Table," *American Economic Review*, 2015, *105* (10), 3150–3182.

Felbermayr, Gabriel, Volker Grossmann, and Wilhelm Kohler, "Migration, International Trade, and Capital Formation: Cause or Effect?," in Barry R. Chiswick and Paul W. Miller, eds., *Handbook of the Economics of International Migration*, Vol. 1, North-Holland, 2015, pp. 913–1025.

Ferenczi, Imre, "International migration statistics," in Walter F Willcox, ed., *International Migrations, Volume I: Statistics*, New York: National Bureau of Economic Research, 1929.

Ferrie, Joseph P and Timothy J Hatton, "Two Centuries of International Migration," in Barry R Chiswick and Paul W Miller, eds., *Handbook of the Economics of International Migration*, Vol. 1, Elsevier, 2015, pp. 53–88.

Fink-Jensen, Jonathan, "Urbanization Ratio, Version 1," Clio Infra database, IISH Dataverse. Amsterdam: International Institute of Social History, Royal Netherlands Academy of Arts and Sciences 2015.

Gapminder, "Child Mortality Estimates, Version 11," Stockholm: Gapminder 2020.

— , "Total Fertility Rate Estimates, Version 12," Stockholm: Gapminder 2020.

Gazeaud, Jules, Eric Mvukiyehe, and Olivier Sterck, "Cash transfers and migration: Theory and evidence from a randomized controlled trial," CSAE Working Paper WPS/2019-16, Centre for the Study of African Economies. Oxford: Oxford University 2019.

Gelbach, Jonah B, "When do covariates matter? And which ones, and how much?," *Journal of Labor Economics*, 2016, *34* (2), 509–543.

Gibson, Campbell and Kay Jung, "Historical Census Statistics on the Foreign-Born Population of the United States: 1850–2000," Population Division Working Paper No. 81. Washington, DC: US Census Bureau 2006.

Görlach, Joseph-Simon, "Borrowing Constraints, Migrant Selection, and the Dynamics of Return and Repeat Migration," Working Paper, Dept. of Economics. Milan: Bocconi University 2019.

Gould, John D, "European inter-continental emigration 1815-1914: patterns and causes," *Journal of European Economic History*, 1979, *8* (3), 593–679.

Gozzini, Giovanni, "The global system of international migrations, 1900 and 2000: A comparative approach," *Journal of Global History*, 2006, *1* (3), 321–341.

Granger, Clive and P Newbold, "Spurious regressions in econometrics," *Journal of Econometrics*, 1974, *2* (2), 111–120.

Grogger, Jeffrey and Gordon H Hanson, "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 2011, *95* (1), 42–57.

Hanson, Gordon and Craig McIntosh, "Is the Mediterranean the New Rio Grande? US and EU immigration pressures in the long run," *Journal of Economic Perspectives*, 2016, *30* (4), 57–82.

Hanson, Gordon H, "Illegal Migration from Mexico to the United States," *Journal of Economic Literature*,

December 2006, *44* (4), 869–924.

__ , "International migration and the developing world," in Dani Rodrik and Mark Rosenzweig, eds., *Handbook of Development Economics*, Vol. 5, Amsterdam: North-Holland, 2010, pp. 4363–4414.

__ **and Craig McIntosh**, "The Great Mexican Emigration," *Review of Economics and Statistics*, 2010, *92* (4), 798–810.

__ **and** __ , "Birth rates and border crossings: Latin American migration to the US, Canada, Spain and the UK," *Economic Journal*, 2012, *122* (561), 707–726.

Hatton, Timothy J, "The Age of Mass Migration: What We Can and Can't Explain," in Subrata Ghatak and Anne Showstack Sassoon, eds., *Migration and Mobility: The European Context*, London: Palgrave Macmillan UK, 2001, pp. 11–29.

__ , "The cliometrics of international migration: a survey," *Journal of Economic Surveys*, 2010, *24* (5), 941–969.

__ **and Jeffrey G Williamson**, "International migration and world development: a historical perspective," in "Economic Aspects of International Migration," Springer, 1994, pp. 3–56.

__ **and** __ , "Late-Comers to Mass Emigration: The Latin Experience," in Timothy J Hatton and Jeffrey G Williamson, eds., *Migration and the International Labor Market 1850–1939*, London: Routledge, 1994, pp. 55–71.

__ **and** __ , "What Drove the Mass Migrations from Europe in the Late Nineteenth Century?," *Population and Development Review*, 1994, *20* (3), 533–559.

__ **and** __ , *The Age of Mass Migration: Causes and Economic Impact*, Oxford: Oxford University Press, 1998.

__ **and** __ , "Out of Africa? Using the past to project African emigration pressure in the future," *Review of International Economics*, 2002, *10* (3), 556–573.

__ **and** __ , "Demographic and economic pressure on emigration out of Africa," *Scandinavian Journal of Economics*, 2003, *105* (3), 465–486.

__ **and** __ , *Global Migration and the World Economy: Two centuries of policy and performance*, Cambridge, MA: MIT Press, 2005.

__ **and** __ , "What fundamentals drive world migration?," in George J Borjas and Jeff Crisp, eds., *Poverty, International Migration and Asylum*, Springer, 2005, pp. 15–38.

__ **and** __ , "Are third world emigration forces abating?," *World Development*, 2011, *39* (1), 20–32.

Horiuchi, Shiro and Samuel H Preston, "Age-specific growth rates: The legacy of past population dynamics," *Demography*, 1988, *25* (3), 429–441.

Hummels, David, "Transportation costs and international trade in the second era of globalization," *Journal of Economic Perspectives*, 2007, *21* (3), 131–154.

Idu, Roxana, "Source Country Economic Development and Dynamics of the Skill Composition of Emigration," *Economies*, 2019, *7* (1), 18.

Karadja, Mounir and Erik Prawitz, "Exit, Voice, and Political Change: Evidence from Swedish Mass Migration to the United States," *Journal of Political Economy*, 2019, *127* (4), 1864–1925.

Kerr, William R., "Ethnic scientific communities and international technology diffusion," *The Review of Economics and Statistics*, 2008, *90* (3), 518–537.

Lee, Jong-Wha and Hanol Lee, "Human capital in the long run," *Journal of Development Economics*,

2016, *122*, 147 – 169.

Lucas, Robert E B, "Migration and development The role for development aid," Delmi Research overview 2019:5. Stockholm: Migration Studies Delegation 2019.

Maccio, Guillermo and Diva Elizalde, *La población no nativa de la Argentina: 1869–1991*, República Argentina, Ministerio de Economía y Obras y Servicios Públicos, 1996.

Macunovich, Diane J, "The role of relative cohort size and relative income in the demographic transition," *Population and Environment*, 1999, *21* (2), 155–192.

— , "Relative cohort size: source of a unifying theory of global fertility transition?," *Population and Development Review*, 2000, *26* (2), 235–261.

Martin, Philip L, *Trade and Migration: NAFTA and Agriculture.*, Washington, DC: Institute for International Economics, 1993.

Massey, Douglas S, "Economic Development and International Migration in Comparative Perspective," *Population and Development Review*, 1988, *14* (3), 383–414.

— , *Economic Development and International Migration In Comparative Perspective*, Washington, D.C.: Commission for the Study of International Migration and Cooperative Economic Development, 1989.

— , **Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J Edward Taylor**, "Theories of international migration: A review and appraisal," *Population and Development Review*, 1993, *19* (3), 431–466.

Mayda, Anna Maria, "International migration: A panel data analysis of the determinants of bilateral flows," *Journal of Population Economics*, 2010, *23* (4), 1249–1274.

McFadden, Daniel, "The measurement of urban travel demand," *Journal of Public Economics*, 1974, *3* (4), 303–328.

McKenzie, David and Hillel Rapoport, "Network effects and the dynamics of migration and inequality: theory and evidence from Mexico," *Journal of Development Economics*, 2007, *84* (1), 1–24.

Millán, Teresa Molina, Karen Macours, John A Maluccio, and Luis Tejerina, "Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program," *Journal of Development Economics*, 2020, *143*, 102385.

Mincer, Jacob, "Investment in human capital and personal income distribution," *Journal of Political Economy*, 1958, *66* (4), 281–302.

Minnesota Population Center, "Integrated Public Use Microdata Series, International: Version 7.2 [dataset]," Minneapolis, MN: IPUMS 2019.

Mohammed, Saif I Shah and Jeffrey G Williamson, "Freight rates and productivity gains in British tramp shipping 1869–1950," *Explorations in Economic History*, 2004, *41* (2), 172–203.

Mummolo, Jonathan and Erik Peterson, "Improving the interpretation of fixed effects regression results," *Political Science Research and Methods*, 2018, *6* (4), 829–835.

Munshi, Kaivan, "Social Networks and Migration," *Annual Review of Economics*, 2020, *12* (1), 503–524.

O'Rourke, Kevin H, "The era of free migration: Lessons for today," in Andrew Sobel, ed., *Challenges of Globalization*, Routledge, 2009, pp. 70–89.

— , **Alan M Taylor, and Jeffrey G Williamson**, "Factor price convergence in the late nineteenth century," *International Economic Review*, 1996, *37* (3), 499–530.

Orrenius, Pia M and Madeline Zavodny, "Self-selection among undocumented immigrants from Mex-

ico," *Journal of Development Economics*, 2005, *78* (1), 215–240.

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

Özden, Çağlar, Christopher R Parsons, Maurice Schiff, and Terrie L Walmsley, "Where on earth is everybody? The evolution of global bilateral migration 1960–2000," *World Bank Economic Review*, 2011, *25* (1), 12–56.

Passel, Jeffrey S and D'Vera Cohn, *Mexican immigrants: How many come? How many leave?*, Washington, DC: Pew Hispanic Center, 2009.

__ **and Roberto Suro**, *Rise, peak, and decline: Trends in US immigration 1992–2004*, Washington, DC: Pew Hispanic Center, 2005.

Pedersen, Peder J, Mariola Pytlikova, and Nina Smith, "Selection and network effects—Migration flows into OECD countries 1990–2000," *European Economic Review*, 2008, *52* (7), 1160–1186.

Roy, Andrew Donald, "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 1951, *3* (2), 135–146.

Ruhe, Constantin, Charles Martin-Shields, and Lisa Maria Groß, "The Asylum Hump: Why Country Income Level Predicts New Asylum Seekers, But Not New Refugees," *Journal of Refugee Studies*, 2020, *forthcoming*.

Sjaastad, Larry A, "The costs and returns of human migration," *Journal of Political Economy*, 1962, *70* (5, Part 2), 80–93.

Skeldon, Ronald, *Migration and Development: A Global Perspective*, Harlow: Longman, 1997.

__ , "Migration and development," Presented at UN Expert Group Meeting on International Migration and Development in Asia and the Pacific, Bangkok, Thailand, UN/POP/EGM-MIG/2008/4, 20–21 Sep., 2008.

Stark, O. and S. Yitzhaki, "Labour Migration as a Response to Relative Deprivation," *Journal of Population Economics*, 1988, *1* (1), 57–70.

Taylor, Alan M and Jeffrey G Williamson, "Convergence in the age of mass migration," *European Review of Economic History*, 1997, *1* (1), 27–63.

Thomas, Brinley, *Migration and economic growth: A study of Great Britain and the Atlantic economy*, 2nd ed., Cambridge: Cambridge University Press, 1973.

Timmer, C Peter, *A World without Agriculture: The structural transformation in historical perspective*, Washington, DC: American Enterprise Institute, 2009.

United Nations, "International Migration Flows to and from Selected Countries: The 2015 Revision," UN DESA, Population Division, POP/DB/MIG/Flow/Rev.2015. New York: United Nations 2015.

__ , "International Migrant Stock 2019: Documentation," UN DESA, Population Division, POP/DB/MIG/Stock/Rev.2019. New York: United Nations 2019.

Vogler, Michael and Ralph Rotte, "The effects of development on migration: Theoretical issues and new empirical evidence," *Journal of Population Economics*, 2000, *13* (3), 485–508.

Williamson, Jeffrey G, "Migration to the new world: long term influences and impact," *Explorations in Economic History*, 1974, *11* (4), 357–389.

__ , "World migration in historical perspective: Four big issues," in Barry R Chiswick and Paul W Miller, eds., *Handbook of the Economics of International Migration*, Vol. 1, Elsevier, 2015, pp. 89–101.

Yule, G. Udny, "Why do We Sometimes Get Nonsense Correlations between Time Series? A Study in

Sampling and the Nature of Time Series," *Journal of the Royal Statistical Society*, 1926, *89* (1), 1–63.

Zelinsky, Wilbur, "The hypothesis of the mobility transition," *Geographical Review*, 1971, pp. 219–249.

Online Appendix

"The Emigration Life Cycle: How Development Shapes Emigration from Poor Countries"

Michael A. Clemens — July 2020

A1 Data Sources

A1.1 World, 1960–2019

Post-1960 migrant stocks and flows: Bilateral migrant stocks by country of birth from 1960 to 2000 are from the World Bank *Global Bilateral Migration Database* created by Özden, Parsons, Schiff and Walmsley (2011) (https://datacatalog.worldbank.org/dataset/global-bilateral-migration-database, accessed October 19, 2018). Bilateral stocks for 2010 are from the World Bank's *Bilateral Migration Matrix 2010* (https://datacatalog.worldbank.org/dataset/bilateral-migration-2010, accessed October 19, 2018). Bilateral stocks for 2019 are from the United Nations Department of Economic and Social Affairs, Population Division, as described in United Nations (2019) (https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates19.asp, accessed January 13, 2020). Net migration flows between census rounds are approximated by changes in migrant stocks. In all sources, overseas departments and territories of high-income countries (such as the Falkland Islands, Martinique, and Tokelau) are likewise considered to be high-income countries.

Post-1960 income and population: Real GDP per capita in Purchasing Power Parity-adjusted 2011 US dollars is taken from the Penn World Table 9.1 (Feenstra et al. 2015) at DOI: 10.15141/S50T0R. The GDP measure used is expenditure-side real GDP at chained PPPs, suitable for comparing relative living standards across countries and over time. This release of the Penn World Table ends in 2017, so the GDP per capita series is extended to 2019 using the real GDP and population growth rates in the International Monetary Fund *World Economic Outlook* from October 2019. GDP per capita estimates for Afghanistan, Cuba, Democratic People's Republic of Korea, and Libya are from the Maddison Project Database (Bolt et al. 2018), likewise in 2011 PPP$. The Maddison series for Afghanistan and Libya end in 2016, so the GDP per capita series for these two countries are extended to 2019 using the real GDP and population growth rates in the IMF *World Economic Outlook* from October 2019. The 2019 observations for Cuba and North Korea are extrapolated from 2015 assuming continuation of the four-year country-specific growth rate 2011–2015. Population is from the World Bank *World Development Indicators* (data code SP.POP.TOTL), accessed January 24, 2020. At the time of data access the population series ended in 2018, so the population series is extended to 2019 using the 2018–2019 population growth rate in the IMF *World Economic Outlook* from October 2019.

Historical World Bank country classifications by income are from the World Bank's *World Development Indicators* database, https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries, accessed December 19, 2019. *School enrollment* data are from Lee and Lee (2016). *Child mortality* (under age 5) estimates are from Gapminder (2020a). Total Fertility Rate (average number of babies born per woman across all childbearing years) are from Gapminder (2020b). *Urbanization* rates (Urban population, % of total population) are from the World Bank's *World Development Indicators* (SP.URB.TOTL.IN.ZS, accessed July 31, 2020).

A1.2 World, 1850–1914

Pre-1914 migrant stocks and flows to the UNITED STATES: Full-count census stocks of the foreign-born by country of birth for the United States 1850–1920 are from Gibson and Jung (2006, Table 4). I corrected one

error in this secondary source: Gibson and Jung give the Cuba-born population of the United States as 6,917 in 1870, and 5,319 in 1860. In fact the Cuba-born population was 6,917 in 1880 (*1880 Census: Volume 1. Statistics of the Population of the United States*, p. 470), and 5,319 in 1870 (*1870 Census: Volume 1. The Statistics of the Population of the United States*, p. 338).

Pre-1914 migrant stocks and flows to Canada: Census public use microdata sample-based estimates of the foreign-born by country of birth in Canada 1871-1911 are from Minnesota Population Center (2019). I acknowledge the following researchers and agencies that made the underlying data available to IPUMS. **Canada 1871**: Gordon Darroch and Michael Ornstein, Canadian Historical Social Mobility Project. National Sample of the 1871 Census of Canada [computer file]. Toronto, Ontario: York Institute for Social Research and Department of Sociology, York University, 1979. **Canada 1881**: Lisa Dillon, 1881 Canadian Census Project, North Atlantic Population Project, and Minnesota Population Center. National Sample of the 1881 Census of Canada (version 2.0). Montréal, QC: Département de Démographie, Université de Montréal [distributor], 2008. **Canada 1891**: Kris Inwood and Chelsea Jack. National Sample of the 1891 Census of Canada. Guelph, Canada: University of Guelph, 2011. **Canada 1901**: Canadian Families Project. National Sample of the 1901 Census of Canada. Victoria, Canada: University of Victoria, 2002. **Canada 1911**: Chad Gaffield, Peter Baskerville, Sean Cadigan, Marc St-Hilaire, Claude Bellavance, France Normand, Gordon Darroch, Carl Amhrein, Lorne Tepperman, Charles Jones, and Eric Sager. Canadian Century Research Infrastructure Project. National Sample of the 1911 Census of Canada [dataset]. Edmonton, Canada: University of Alberta [distributor], 2009.

Pre-1914 migrant stocks and flows to Argentina: Full-count census stocks of the foreign-born in Argentina for several countries of birth are given in Maccio and Elizalde (1996, p. 16). This is supplemented with additional countries of birth and years by referring to the original full-count census reports: Great Britain and Switzerland in 1869 are from Diego G. de la Fuente (1872), *Primer Censo de la República Argentina*, p. XXXII. Great Britain, Switzerland, Belgium, Greece, Denmark, Netherlands, and Sweden in 1895 are from Diego G. de la Fuente (1898), *Segundo Censo de la República Argentina*, Tomo II, p. XLIV. Great Britain, Switzerland, Belgium, Greece, Denmark, Netherlands, and Sweden in 1914 are from Alberto B. Martínez (1916), *República Argentina, Tercer Censo Nacional*, Tomo I, pp. 205–206. An estimate of zero for Sweden in 1869 is made based on the lack of any recorded migration from Sweden to Argentina prior to 1869 in Ferenczi (1929, p. 545). Similarly, the stock of Netherlands-born in Argentina in 1869 is approximated as 70, as this is the cumulative pre-1869 migration from the Netherlands recorded in the same table by Ferenczi.

Pre-1914 emigrant flows to All Extra-European Destinations: For **Italy**, gross emigrant flows of Italian citizens 1876–1913 are from Ferenczi (1929, p. 820), Italy Table VII, extra-European destinations only. Immigrant flows in 1884 and 1887–1913 are from Ferenczi (1929, p. 839) Italy Table XXI, intercontinental third-class passengers. The adjacent tables indicate that third-class passengers are roughly 90 percent of the total inflow, and that the vast majority of these are Italian citizens. Net migration is the difference between these emigrant and immigrant flows. For **Sweden**, gross emigration 1851–1913 is all emigrants to all destinations, from Ferenczi (1929, p. 757) Sweden Table II. Net migration to extra-European destinations is the difference between emigration to extra-European countries 1876–1913 from Ferenczi (1929, p. 756) Sweden Table I, and immigration from extra-European countries 1876–1913 from Ferenczi (1929, p. 760) Sweden Table IV. For **Great Britain**, gross emigration 1815–1913 is passengers to extra-European countries, any nationality, from Ferenczi (1929, p. 627) British Isles Table IV. Net emigration is the difference between an emigration series and an immigration series. That emigration series is citizen passengers to extra-European destinations 1853–1913 from Ferenczi (1929, p. 627) British Isles Table IV; the immigration series 1854–1869 is inward movement of passengers (including transmigrants) from extra-European countries, from Ferenczi (1929, p. 637) British Isles Table IX, and from 1870 to 1913 it is inward movement of passenger citizens from extra-European countries, from Ferenczi (1929, p. 640) British Isles Table XII. For **Hungary**, gross emigration is intercontinental emigration of citizens, from Ferenczi (1929, p. 716) Hungary Table V. Immigration is immigration of citizens, from Ferenczi (1929, p. 720) Hungary Table XV.

Pre-1914 income and population: GDP per capita and population estimates in each year are from the Maddison Project Database (Bolt et al. 2018), linearly interpolated for occasional gaps in the annual estimates. *School enrollment* data are from Lee and Lee (2016). *Child mortality* (under age 5) estimates are from Gapminder (2020a). Total Fertility Rate (average number of babies born per woman across all childbearing years) are from Gapminder (2020b). *Urbanization* rates (Urban population, % of total population) are from the Clio Infra Project (Fink-Jensen 2015) linearly interpolated between missing years. That source contains no estimates for Russia or Serbia; these are taken from Bairoch and Goertz (1986, 288).

A1.3 Mexico: 1960–2017

Annual emigration from Mexico to the United States: This is estimated by extending the method of Passel and Suro (2005) and Passel and Cohn (2009). They count prior-year US arrivals from Mexico in the US Current Population Survey, making various adjustments and comparing with other data sources. This paper creates similar estimates for a longer period using the US American Community Survey (ACS). It begins by counting the (weighted) number of Mexican-born people who report each year 1960–2017 as their year of immigration to the US, in each annual round of the ACS 2000–2018. Arrivals in the same year of the survey (e.g., 2018 arrivals reported in the 2018 ACS) are omitted, since the ACS is conducted in the middle of the year; thus the most recent flow estimate is for 2017. For the years 1999–2012, annual inflows *net of departures within five years* are estimated as the number of Mexico-born in the five subsequent rounds of the ACS reporting arrival in that year, averaged. For example, the estimated net inflow for 2013 is the average (weighted) number of Mexico-born reporting 2013 as their year of arrival in the five ACS rounds 2014–2018. Comparable estimates for years-of-arrival after 2013 or before 1999 require estimates of attrition (emigration and death). The *decadal* attrition rate is estimated for each year of arrival 1960–1998 as the difference between the average number of Mexico-born reporting arrival in that year in ACS rounds 2010–2018 and the same average in ACS rounds 2000-2008. This rate is close to 10 percent per decade for all years of arrival 1960–1998. Based on this, estimated arrivals in the years 2014–2017 are slightly scaled *down* to account for attrition that would have happened if they had been measured over the full five-year retrospective period: Each year's arrivals in year t 2014–2017 are multiplied by the annualized attrition factor $(1 - 0.1)^{(t-2013)/10}$. Similarly, estimated arrivals before 1999 must be scaled *up* to account for attrition between the year five years after the inflow and the earliest date of surveying (the 2000 ACS). Each year's arrivals in year t 1960–1998 are multiplied by the annual *inverse* attrition factor $(1/(1 - 0.1))^{(1999-t)/10}$. These latter estimates for the period 1960–1998 are compared against inflow rates measured closer to the year of arrival in earlier full-count censuses (e.g. arrivals 1980–1986 reported in the 1990 full-count census, arrivals 1970–1974 reported in the 1980 full-count census, and arrivals 1960–1964 reported in the 1970 full-count census). These independent checks match closely, validating the method of adjusting the ACS estimates for attrition. The resulting flow measure should be interpreted as a rough measure of *net* emigration from Mexico, given that 1) the vast majority of Mexican emigration during this period was to the United States, and 2) the outflow rate from Mexico to the United States is constructed to omit people who depart in the year of arrival or in any of the subsequent five years. Mexico's population and real GDP in PPP-adjusted 2011 US dollars is taken from the Penn World Table 9.1 (Feenstra et al. 2015) at DOI: 10.15141/S50T0R. The GDP measure used is expenditure-side real GDP at chained PPPs, suitable for comparing relative living standards across countries and over time.

# A2 Alternative migration flow data adjusted for mortality

The main text uses net changes in emigrant stocks, defined by country of birth, as a transparent proxy for net emigration flows. This has many advantages, including comparability: Suppose one Mexican person arrived in the United States on a permanent resident visa in 1965, and another arrived as an irregular migrant in 1965 but received a permanent resident visa due to a regularization program in 1988. In both cases the physical act of migration was identical, and would appear identically in measures of a change in the Mexican migrant population of the United States in the 1960s as measured by correct census data.

But if measured by administrative data on the issuance of permanent resident visas, one would appear to have moved decades after the other.

But measurement with net changes in emigrant stocks has disadvantages as well. Important among these is that emigrant stocks can decline simply because emigrants' lives end, not due to any net return migration to the origin country. For this reason, researchers have combined stock data with demographic modeling to estimate true migration flows, as in Abel (2018).

Figure A1, Figure A2, Figure A3, and Figure A4 repeat the regressions in main-text Figure 6 and Figure 7 using the migration flow estimates of Abel (2018) in place of net changes in emigrant stocks. The patterns observed in the main text are qualitatively identical to those shown here. Emigration *flows* rise with the *level* of GDP per capita among poorer countries. There is a turning point around PPP$5,000. The within-country paths followed by once-poor countries differ little from the path implied by the correlation in data pooling countries and time periods. These patterns are not confined to microstates but are observed in China, Indonesia, Mexico, Peru, and other large developing countries.

Figure A1: Net emigration flows estimated by Abel (2018) versus origin-country income per capita, pooled sample 1960–2010
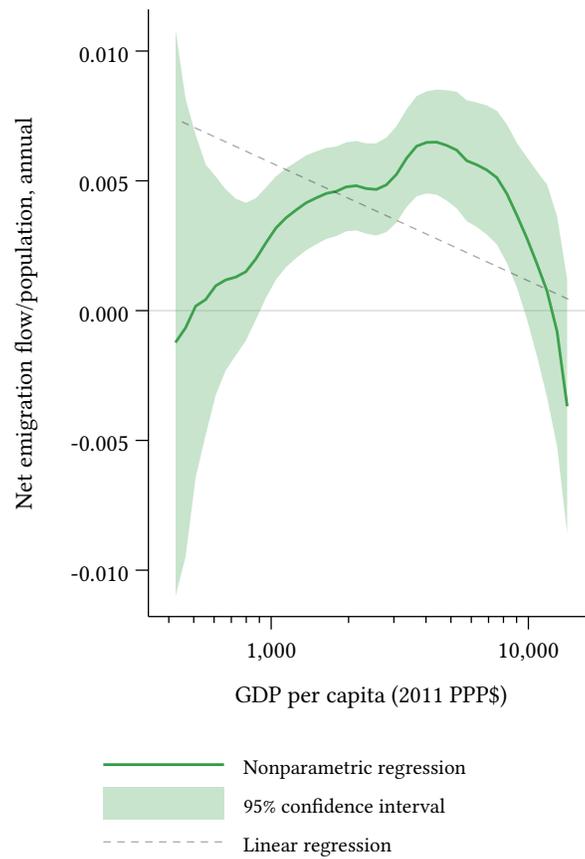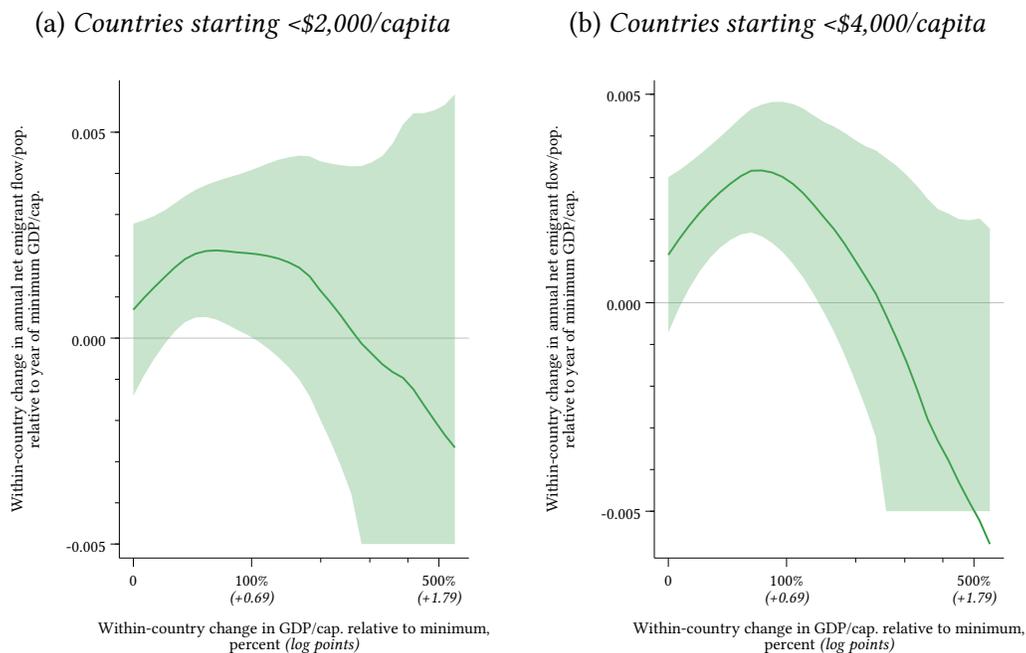

Flows are annualized average net emigration flow during decade estimated by Abel (2018), all destinations. GDP per capita is in initial year of each decade. Nonparametric regression line is a Fan (1992) local-linear regression, Epanechnikov kernel; optimal bandwidth minimizes conditional weighted mean integrated squared error.

Figure A2: Once-poor countries moving along the emigration-flow curve as they grow richer, net emigration flows estimated by Abel (2018), 1960−2010

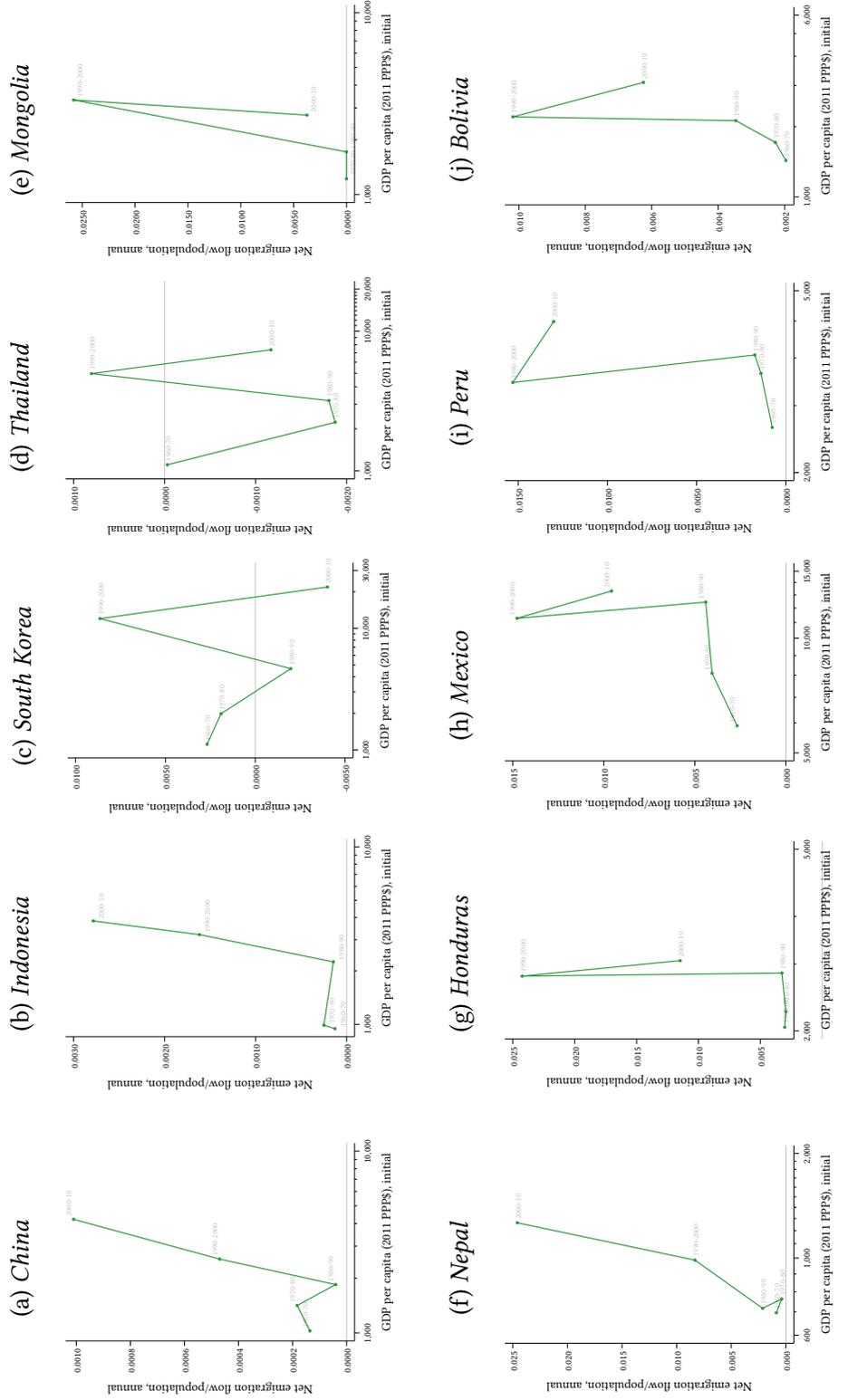(a) *Countries starting <$2,000/capita*   (b) *Countries starting <$4,000/capita*
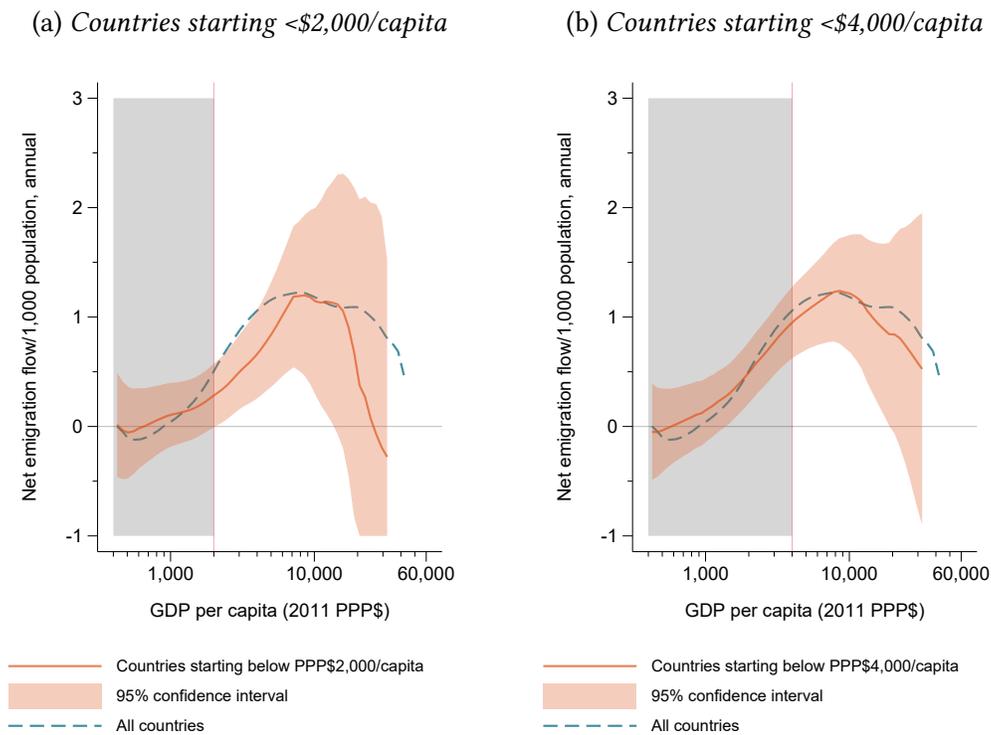
Solid line in Figure A2a shows only migrant-origin countries whose GDP per capita was below PPP$2,000 at some point during 1960−2010. Thus the line outside the shaded area shows countries that grew to higher incomes starting from very low incomes. The dashed line, for reference, is the regression line from Figure 6a for all origin countries. Figure A2b is identical but for a PPP$4,000 cutoff. All destination countries. Nonparametric regression line is a Fan (1992) local-linear regression, Epanechnikov kernel; optimal bandwidth minimizes conditional weighted mean integrated squared error (0.471 log points in Figure A2a, 0.416 in Figure A2b). 95% confidence interval clipped for legibility.

Figure A3: Net emigration flows estimated by Abel (2018) 1960–2010, using only within-country economic growth

(a) *Countries starting <$2,000/capita*   (b) *Countries starting <$4,000/capita*

Includes only migrant-origin countries whose GDP per capita (2011 PPP$) was below $2,000 at some point during 1960–2010. Horizontal axis shows change of GDP per capita from its lowest value in subsequent years, and vertical axis shows change in emigrant stock during the same years. All destination countries. Nonparametric regression line is a Fan (1992) local-linear regression, Epanechnikov kernel; optimal bandwidth minimizes conditional weighted mean integrated squared error (0.260 log points). Flows are annualized average change in emigrant stock during decade. GDP per capita is in initial year of each decade.

Figure A4: Net emigration flow estimated by Abel (2018) versus GDP per capita, representative countries

(a) *China*   (b) *Indonesia*   (c) *South Korea*   (d) *Thailand*   (e) *Mongolia*

(f) *Nepal*   (g) *Honduras*   (h) *Mexico*   (i) *Peru*   (j) *Bolivia*

All destination countries. Flows are annualized average net emigration during each decade 1960–2010 estimated by Abel (2018); GDP per capita is in initial year of each decade.

A-8

A3 Reproducing and reconciling spurious results

The error in the analysis of Benček and Schneiderheinze (2019) can be illustrated by reproducing the spurious negative coefficient with annual data on the emigration of Mexicans to the United States.

Mexico exemplifies the emigration life cycle, as shown earlier in long-run data (Figures 5h and 7h in the main text). The same emigration life cycle can be seen in high-frequency annual data, where the annual emigration rate from Mexico first rose by a factor of five, then dropped back to its original levels (Figure A5a), as Mexico underwent sustained economic development after 1960 (Figure A5b). Graphing the annual emigration rate against GDP per capita reveals the unmistakable inverse-U relationship of the emigration life cycle (Figure A5c). Any empirical method that cannot reveal the life cycle in this quintessential case cannot be informative about empirical patterns of development and emigration more generally.

But in the same data, the method used in the Benček and Schneiderheinze study generates a negative and highly significant coefficient estimate during the period covered by that study. Table A1, column 1, shows the regression (4) run on annual emigration from Mexico. Column 2 of the table regresses the number of emigrants on population, which generates the residuals shown in Figure A6a. Column 3 of the table regresses GDP on population, which generates the residuals shown in Figure A6b. Regressing the first residuals on the second gives a coefficient, in column 4, identical to the partial coefficient on GDP in column 1. That regression is shown graphically in Figure A6c.

This result is spurious. The residuals in Figure A6b contain no information about long-run economic development in Mexico. Those residuals are what remains when the long-run economic development of Mexico—the line in Figure A6b—has been *removed* from the quantity used to calculate the coefficient.
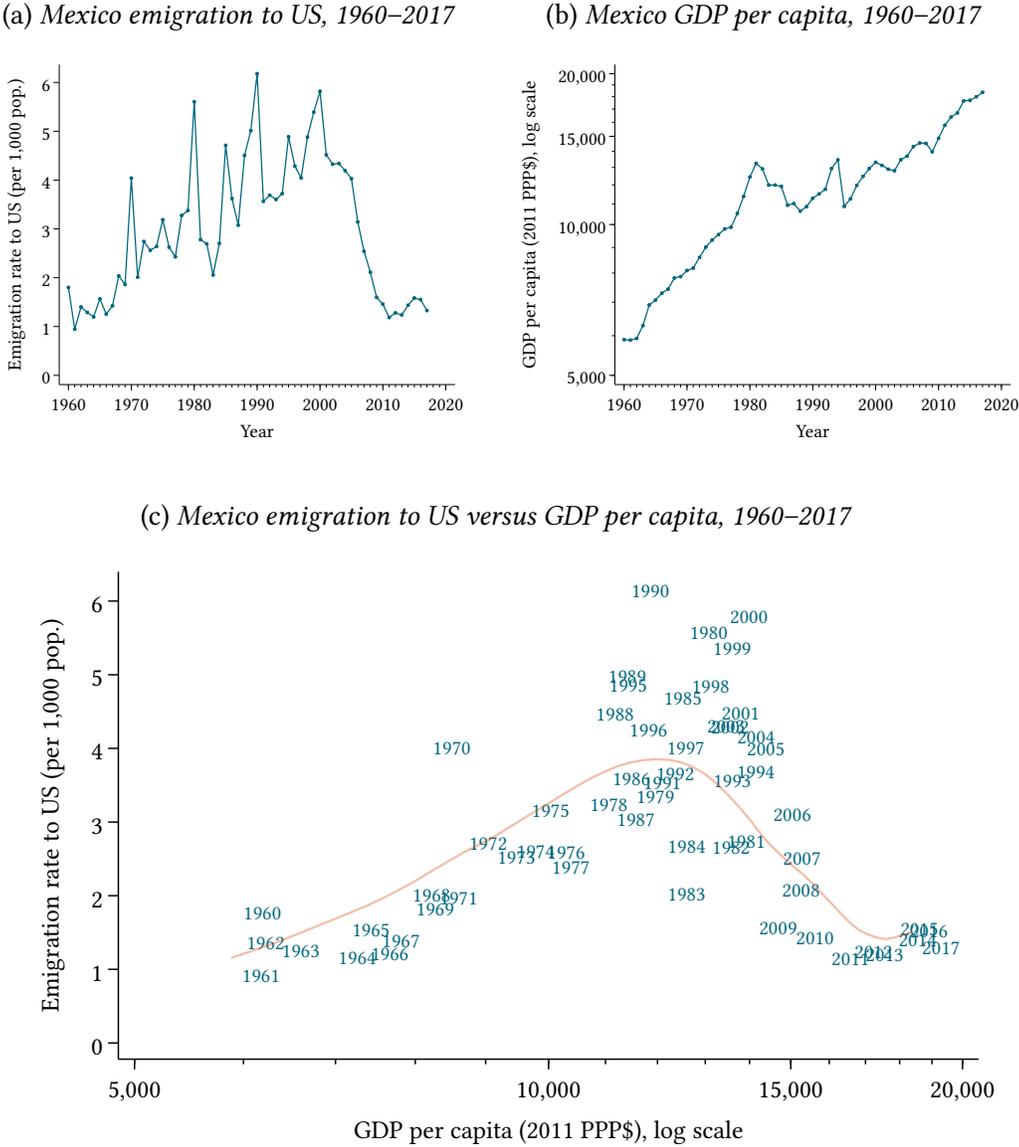
The regression in Figure A6c, the negative quantity estimated by Benček and Schneiderheinze, simply shows the effects of short-run booms and busts. It shows that when Mexico experienced a severe recession in the mid-1990s (the "peso crisis"), emigration spiked. Thus the points for 1995 and 1996 in the upper left of the figure show years when GDP per capita was experiencing a negative short-run shock, with a positive short-run emigration response. Conversely, in the bottom right of the figure, 1981 and 1982 (just before the Latin American debt crisis) were years of a positive shock to the Mexican economy (peak oil prices) that reduced emigration in the short run. None of this bears any relation to the long-run rise in Mexico's GDP per capita—Mexico's *economic development*—the effects of which have been removed by this empirical method.

At the global level, the regression used by Benček and Schneiderheinze would be informative about the effect of transitory shocks on emigration, if this were the subject under study. The elasticity of −0.5 implies that a short-run fall of GDP per capita of 10 percent, such as in a Venezuela-type economic crisis, is associated with a 5 percent increase in emigration the following year. That would be a sensible estimate of the short-run effects of shocks, corroborating prior results on the effects of shocks (such as Bertoli et al. 2013). But it is uninformative about the relationship between economic development and emigration.

The flaws in this method run deeper than the odd regression specification in (4). The obvious fix would be to simply regress the emigration *rate* on GDP *per capita*—the approach of the present paper. But even this would leave two major problems.

First, as discussed by Clemens (2014, 165–166), it would mean modeling a strikingly nonlinear relationship with a linear regression. When the true data-generating process follows an inverse U, a linear fit can give a positive, zero, or negative coefficient estimate depending on which portion of the curve happens to be observed. This is shown for Mexico in Table A1, column 5, using annual data during the same post-1980 period studied by Benček and Schneiderheinze. This column regresses the emigration *rate* on GDP *per capita*. Again, the coefficient estimate is negative, large in absolute value, and highly statistically significant. But this result arises only because the linear regression coefficient for a portion of the curve
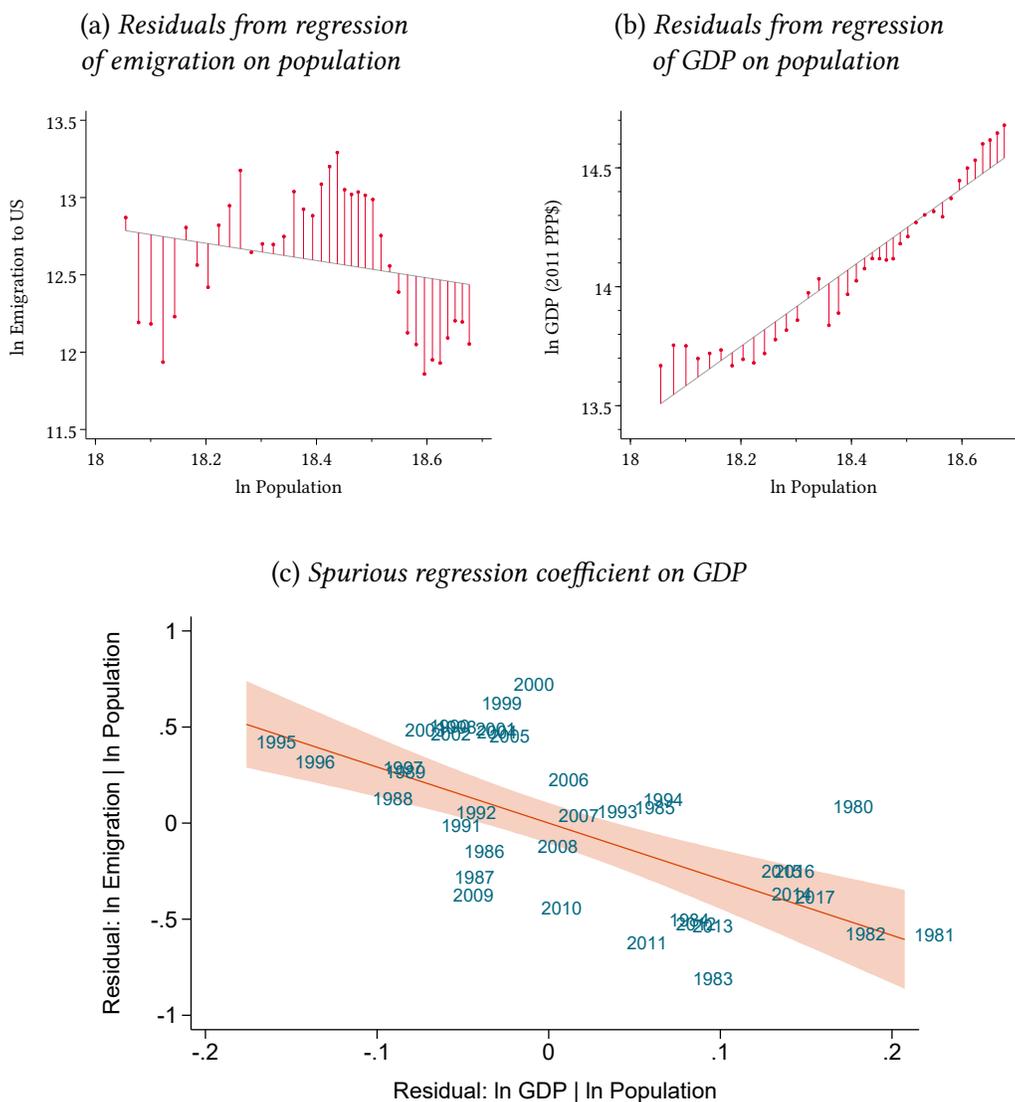
Figure A5: Post-1960 Emigration from Mexico: Annual Mexico-US migration and Mexican economic development

(a) *Mexico emigration to US, 1960–2017*

(b) *Mexico GDP per capita, 1960–2017*

(c) *Mexico emigration to US versus GDP per capita, 1960–2017*

Emigration from Mexico to the US is measured as number of Mexican-born people moving to the US each year per 1,000 population of Mexico, net of US departures within five years of arrival. Details in the Data Sources section of the Appendix. In Figure A5c, line shows a Fan (1992) local-linear regression, Epanechnikov kernel, bandwidth $1,500.

cannot describe the manifest nonlinear emigration life cycle in Figure A5c.

Second, even a correct regression specification would still measure spurious relationships if it considered

Figure A6: POST-1980 EMIGRATION—MEXICO: Illustration of spurious regressions using short-run annual data

(a) *Residuals from regression*
of emigration on population

(b) *Residuals from regression*
of GDP on population

(c) *Spurious regression coefficient on GDP*

Figures A6a, A6b, and A6c, respectively, depict the regressions in columns 2, 3, and 4 of Table A1. In Figure A6c, the solid line shows linear fit and shaded area shows 95% confidence interval on the local predicted value.

the *annual* relationship between GDP per capita and the emigration rate. Year-to-year measurements of GDP per capita do not contain any information about *economic development*—sustained trends in GDP per capita—beyond the information contained in averages over longer periods. Using higher-frequency data only raises the portion of the income-emigration relationship that is explained by short-lived shocks. Short-lived shocks above trend do not constitute economic development by any meaningful definition of that term. High-frequency data are, then, inferior to low-frequency data for this purpose. This is why

Appendix Table A1: Post-1980 Emigration—Mexico: Spurious regressions using annual data

Dep. var.:	ln Emigrants		ln GDP	Residual: ln Emigrants | ln Pop	ln (Emigrants per 1,000 population)
ln GDP	-2.919^{***} (0.566)				
ln Population	4.290^{***} (0.984)	-0.560 (0.379)	1.661^{***} (0.085)		
Residual: ln GDP | ln Pop				-2.919^{***} (0.566)	
ln GDP per capita					-2.568^{***} (0.345)
Adj. R^2	0.434	0.031	0.911	0.399	0.596
N	38	38	38	38	38

Standard errors in parentheses. Constant term omitted in table. $^{***}\ p < 0.001$. Degrees of freedom in column (4) adjusted to 35 rather than 36, for strict comparability with column (1).

the present work uses decadal averages: because higher-frequency data contain no additional relevant information, and much additional irrelevant information.

Correcting the error in Benček and Schneiderheinze (2019) shows that those estimates imply a positive relationship between GDP per capita and emigration similar to the estimates in Table 1 in the main text. If we were to straightforwardly regress the emigration *rate* on GDP *per capita*, this would equate to the constraint that in regression (4), $\beta = 1 - \delta$.[10] This constraint allows one to back out the relationship between the long-term trend in GDP and emigration contained in the coefficient estimate on population ($\ln L_t$) in regression (5). In the core fixed-effects result of Benček and Schneiderheinze (2019, Table 1, Model 3), the coefficient on log population is 1.59. Letting $\beta \left(1 + \frac{\gamma}{\pi}\right) + \delta = 1.59$ and constraining $\beta = 1 - \delta$ gives an equation for the coefficient of interest, the association between the *trend* in GDP and the amount of emigration:

$$\beta^* = 0.59 \times \frac{\pi}{\gamma}. \tag{A.1}$$

This can be calibrated with recent data from low-income countries. According to the World Bank, between 1990 and 2019 the population of low-income countries grew at an annual average of $\pi = 2.76$ percent, while real GDP per capita at PPP grew at an annual average of $\gamma = 1.42$ percent.[11] This implies an estimate of 1.15 for the coefficient of interest in equation (A.1), the elasticity of gross emigration to the within-country trend in real GDP during 1990–2019, holding population constant. In main-text Table 1, columns 5–8, in the second row (high-income destination countries), the elasticity of the net emigration rate to growth in GDP per capita was estimated for the period 1960–2019 as falling in the range of 0.73–1.29. The sign and rough magnitude of the two sets of estimates are the same once the error is corrected.[12]

---

[10]Subtract $\ln L_t$ from both sides of equation (4) in the main text and set $\beta = 1 - \delta$ to get $\ln \frac{M_t}{L_t} = \alpha + \beta \ln \frac{Y_t}{L_t} + \varepsilon_t$.

[11]The World Bank does not estimate GDP per capita at PPP for low-income countries as a whole prior to 1990.

[12]These quantities are equal in sign and comparable in magnitude, but should not be identical because they measure slightly different things. The corrected estimate here applies to gross annual emigration from low-income developing countries to OECD countries, during 1990–2019; the results in Table 1 apply to net decadal emigration to high-income countries from all developing countries with income per capita below PPP$10,000/year, during 1960–2019.