Clemens, Michael A.

Working Paper
Does Development Reduce Migration?

IZA Discussion Papers, No. 8592

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Clemens, Michael A. (2014) : Does Development Reduce Migration?, IZA Discussion Papers, No. 8592, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
http://hdl.handle.net/10419/106535

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Does Development Reduce Migration?

Michael A. Clemens

October 2014
ABSTRACT

Does Development Reduce Migration?*

The most basic economic theory suggests that rising incomes in developing countries will deter emigration from those countries, an idea that captivates policymakers in international aid and trade diplomacy. A lengthy literature and recent data suggest something quite different: that over the course of a “mobility transition”, emigration generally rises with economic development until countries reach upper-middle income, and only thereafter falls. This note quantifies the shape of the mobility transition in every decade since 1960. It then briefly surveys 45 years of research, which has yielded six classes of theory to explain the mobility transition and numerous tests of its existence and characteristics in both macro- and micro-level data. The note concludes by suggesting five questions that require further study.

NON-TECHNICAL SUMMARY

Many aid packages and trade deals for poor countries seek to reduce migration pressure from those countries. But the latest data and a half-century of research suggest that as typical poor countries grow, emigration pressure actually rises. Migration pressure only typically falls when they grow past upper-middle-income status. In poor countries, more development means more migration, not less. Policies to foster development may thus be most effective when paired with policies to accommodate mobility.

JEL Classification: F22, J61, O15

Keywords: migration, mobility, development, growth, transition, hump, lifecycle, inequality, poverty, aid, emigration, demand, pressure

Corresponding author:

Michael A. Clemens
Center for Global Development
2055 L Street NW, 5th floor
Washington, DC 20036
USA
E-mail: mclemens@cgdev.org

1 Introduction

Will economic development in poor countries cause less migration, or more? This simple question has tremendous relevance to the future of migration policy as the world develops, and to development assistance policy as many donor countries seek ways to deter migration. It has spawned 45 years of theoretical and empirical research, in economics as well as in sociology, demography, and geography.

This note reviews what economists know and what they do not know about the effect of development on international migration. I begin by summarizing the conventional wisdom, in the policy worlds of aid and trade diplomacy, that economic development swiftly creates conditions that deter emigration. I then summarize the basic facts about the observed association between economic development and emigration, in the best and most recent data we have. These show a clear and pronounced inverted-U relationship, across countries, between economic development and both stocks and flows of emigrants. This inverted-U goes by many names, but here I refer to it by the earliest name: the “mobility transition”.

I proceed to sketch six classes of theory that the literature has offered to explain the mobility transition. I then summarize the macro- and micro-level quantitative evidence we have testing the existence and shape of the mobility transition in different settings. I conclude by suggesting five questions that require further research, all of which relate to sorting out the mechanisms that produce the transition.

2 Conventional wisdom in policy circles

Two generations have passed since Isaac (1947: 98) suggested that developed countries could substitute for migration with overseas aid and trade policy, reducing migration via “the establishment of new industries in under-capitalized, densely populated countries.” The International Labor Organization (1976) formally recommended that “multilateral and bilateral cooperation” should encourage “appropriate
intensified capital movements and transfers of technical knowledge … designed to avoid the need for workers to migrate.” A decade later the ILO (1984) again recommended “bilateral and multilateral agreements” to “create more employment opportunities … in countries of emigration so as to reduce the need to migrate.” Böhning and Schloeter-Paredes (1994) review some of this history.

The idea that development assistance can reduce migration has continued to captivate leaders in aid and diplomacy circles. In the United States, the development effects of aid and trade diplomacy have been explicitly marketed as a migration deterrent. The U.S. Commission for the Study of International Migration and Cooperative Economic Development (CSIMCED 1990) recommended a range of aid and trade policies for the Western Hemisphere to the George H. W. Bush administration, because “[d]evelopment…is the only way to diminish migratory pressures over time.” As he signed side agreements to the North American Free Trade Agreement, US president Bill Clinton predicted that “as the benefits of economic growth are spread in Mexico … there will be less illegal immigration because more Mexicans will be able to support their children by staying home” (White House 1993). Mexican president Carlos Salinas agreed that “higher wages in Mexico … will mean fewer migrants to the United States and Canada” (State Dept. 1993).

Similar beliefs have driven aid and diplomacy discussions in Europe. Unveiling a massive aid package for Africa, European Commission president José Manuel Barroso stated that “the problem of immigration, the dramatic consequences of which we are witnessing, can only be addressed effectively in the long term through an ambitious and coordinated Development cooperation to fight its root causes” (EC 2005). The European Commission (2008) recommends that EU migration policy should “shift from a primarily security-centered approach focused on reducing migratory pressures” and should “focus much more on … improving the … socio-economic situation in low-income and middle-income countries, and in regions characterized by high emigration pressures.” France built its “Co-development” policy, the basis for aid agreements with numerous African countries starting in 2006, on the idea that “development should
contribute, in the shorter or longer term, to reducing emigration pressures in Third World countries” (Khoudour-Castéras 2009).

This policy consensus is not without nuance. Many policymakers—especially those in migration policy circles—acknowledge that the shorter-term relationship between development and migration may not be simple. The U.S. commission above (CSIMCED 1990) found that “while job-creating economic growth is the ultimate solution to reducing these migratory pressures, the economic development process itself tends in the short term to stimulate migration by raising expectations and enhancing people’s ability to migrate.” This idea is widely accepted among migration researchers and policymakers (Martin and Taylor 1996, Lucas 2005: 49), and finds expression in the qualifiers “over time” and “longer term” that appear in many policy claims that development must eventually reduce migration.

That said, such statements rarely define the length of the “short term”, and are often made in the context of ongoing crises of high migration flows under pressure for short term action. This leaves them open to interpretation as suggesting that, at least after a few years, development policies will indeed serve to reduce migration. In short, aid and diplomacy discussions are dominated by the idea that the relationship between development and “migratory pressure” must be negative, though they often allow that the relationship may be ambiguous in some settings and for an unspecified period. Few such statements have an empirical basis to posit any part of this pattern—the dominant negative relationship between development and migration, the possibility of a temporary positive relationship, or the duration of ‘temporary’.

3 The mobility transition in recent data

Data on international migration flows are generally poor. Migration data easily available for research stand at a level of coverage and accuracy attained several decades ago by data on international trade flows
Recent years, however, have seen tremendous advances. Though most countries lack effective means to measure permanent emigration at the border, researchers have had success in measuring emigration by counting the foreign-born in laborious compilations of censuses, surveys, and population registers of destination countries—then inferring emigration from the origin (Parsons et al. 2005, Dumont and Lemaître 2005, Docquier and Marfouk 2007).

Recently a team based at the World Bank created a set of five global migration matrices spanning the last half of the 20th century (Özden et al. 2011). Built using over a thousand censuses and population registers, the five matrices count the number of people born in each country on earth residing in each other country, once per decade from 1960 to 2000. Summing across all countries of destination allows unprecedentedly accurate estimation of total emigrant stocks for each country at each point in time, as well as decade-by-decade net emigration flows. The United Nations Population Division has recently created similar matrices, using similar methods, for the years 1990, 2000, and 2010. These data allow us for the first time to inquire whether richer countries exhibit broadly lower or higher rates of emigration, whether measured as emigrant stocks or net medium-term emigration flows.

There is a marked inverted-U relationship between emigrant stocks and real income per capita throughout the late 20th century, and the inverted-U became more pronounced as years passed. Figure 1 shows nonparametric cross-country regressions of emigrant stock on income per capita at Purchasing Power Parity (PPP). Here, each country’s emigrant stock is measured as the total number of people born in that country residing outside that country, divided by that country’s population. The left-hand side of the figure shows the five timepoints in the World Bank data (1960–2000); the right-hand side shows the three timepoints in the UN data (1990–2010).

In all years, there is no hint of a negative relationship between income and emigrant stock between PPP income per capita of roughly $600 (that of today’s Niger or Ethiopia) and about $7,500 (today’s Albania
or Colombia). In this range of income, in fact, the relationship is positive. The rise in emigrant stocks associated with higher income levels in this range is statistically significant at the 5% level (see Appendix). The magnitude of the positive relationship is substantial. Early in the second half of the 20th century, richer countries in this range on average had emigrant stocks about three percentage points larger than the poorest countries. By the end of the century, this difference grew to nine percentage points, and seems to have continued growing since then.

At higher levels of income the pattern reverses. For countries with per-capita incomes above roughly PPP$7,000–8,000, there is a clear negative relationship in all years. This negative relationship, likewise, is statistically significant at the 5% level. All countries in this range of income are defined by the World Bank as “upper-middle-income countries” or “high income countries”. The richest countries on earth on average had emigrant stocks about three to five percentage points lower than the middle-income countries with the greatest emigrant stocks.

We see similar patterns for net emigration flows. Figure 2 shows non-parametric cross-country regressions of the decade-by-decade net emigration flows from each country on the level of PPP income per capita. Net emigration flow rates are measured as the difference between emigration stocks at the end of the decade and stocks at the beginning of the decade, divided by population at the beginning of the decade. The flow rates are per decade, not per annum. Income per capita is likewise measured at the beginning of each decade. Again, the World Bank data for four flow periods are on the left (1960–70, 1970–80, 1980–90, 1990–2000), and the UN data for two flow periods on the right (1990–2000, 2000–10).

In all of these decades there is no sign of a negative relationship between income and net emigration flows below an income per capita of roughly PPP$5,000–6,000 (today’s Jordan or Jamaica). For countries with higher incomes there is a marked negative association between rising incomes and net emigration
flows. Almost all such countries, again, are defined as “upper-middle-income” or “high income” by the World Bank. The positive relationship at low income levels is statistically significant in all four decades; the negative relationship at higher income levels is statistically significant in all decades except 1970–80 (see Appendix). The rise and fall are substantial in magnitude: countries at the turning point have net emigration rates two or three percentage points higher than the world’s poorest countries—measured in percentage of the population per decade. The world’s richest countries exhibit emigration rates two or three percentage points lower than countries at the turning point, in all decades except 1970–80.

In short, the best-available data show an unmistakable inverted-U pattern across countries in the relationship between overall economic development and emigration. The data offer no sign that among low-income or lower-middle-income countries, rising incomes are associated with smaller emigrant stocks or lower net emigration rates. To the contrary, typical countries in this group show a positive, significant association between average incomes and emigration. The relationship changes markedly somewhere around PPP$6,000–8,000. Among countries above this level of income—upper-middle-income countries or richer—higher incomes are associated with reduced emigration. But even the very richest countries do not systematically exhibit lower emigration rates than the poorest countries.

These data do not necessarily demonstrate a transition path that any given country must take, for the same reason that mortality rates in an observed population do not determine the life expectancy of any given child. But if the observed inverted-U does approximate a transition path even roughly, how long might such a transition take? Figures 1 and 2 suggest a rough timescale—under the strong assumption that these cross-sectional snapshots of stocks and flows are informative about the path that given groups of countries followed.
Figure 1: Cross-section nonparametric regressions of emigrant stocks on real income per capita, 1960–2010

Lines show Nadaraya-Watson kernel-weighted local mean, Epanechnikov kernel, bandwidth 0.5 natural log points. Emigrant stocks are the number of people born in each country residing outside that country, divided by national population. Observations are country-years. In World Bank data, \( N \) is 106 for 1960, 140 for 1970, 140 for 1980, 164 for 1990, and 164 for 2000. Macau and Brunei omitted. In UN data, \( N \) is 155 in 1990–2010.
Figure 2: Nonparametric regressions of decadal emigrant flows on initial real income per capita, 1960–2010

Lines are Nadaraya-Watson kernel-weighted local mean, Epanechnikov kernel, bandwidth 0.5 natural log points. Emigrant flow is the decadal change in the number people born in each country who reside outside it, divided by decade-initial national population. The flow rate is thus per decade, not per annum. ‘Initial’ GDP/capita means in the first year of each decade. Observations are country-decades. In World Bank data, $N$ is 106 for 1960–70, 140 for 1970–80, 140 for 1980–90, 164 for 1990–2000. Macau and Brunei omitted. In UN data, $N$ is 155 for both periods.
We can approximate the transition time as the time for a given country to reach income per capita of PPP$7,000. At a healthy real per capita growth rate of 2% per year, it would take 133 years for a country starting from $500 per capita (today’s Niger or Burundi) and 63 years for a country starting from $2,000 per capita (today’s Cambodia or Zambia). At a strong growth rate of 3% per year, these durations would be 89 years and 42 years, respectively. These are optimistic growth scenarios, given that during 1960–2000 the average country experienced real growth in per capita income of 1.8% per year (Barro and Sala-i-Martin 2003: 4). And most poor countries grew more slowly. If the path followed by given groups of developing countries over time resembles these cross-sectional relationships, then, the migration transition is a process of generations. For some it may take on the order of a century. This duration has implications for macro-empirics on the mobility transition, to which I will return below.

4 Theories of a mobility transition

Why would the migration-development relationship trace out an inverted-U? Intuition and the simplest models of economic migration suggest something quite different: that higher earnings in origin countries should be accompanied by monotonic declines in the propensity to emigrate. The founding economic theories of migration—sometimes collectively called the ‘neoclassical’ model—posit migration as a costly move from worse economic circumstances to better ones (Ravenstein 1885, Isaac 1947: 23, Lewis 1954). Migration is, then, a form of investment in human capital (Schultz 1961, Sjaastad 1962, Schultz 1972: 4). Higher reservation earnings strictly reduce the incentive to make such an investment. Even as these models have evolved to include risk (Todaro 1969, Harris and Todaro 1970) and income distribution (Roy 1951, Borjas 1987), they still predict that rising average incomes at the origin will tend to strictly reduce migration.

In the 1970s, a series of researchers began to sketch a different class of theories about the relationship between rising average incomes and emigration. These theories predict an initial period, starting from low
levels of income, during which rising incomes are accompanied by rising rates of emigration. Only after some turning point are further increases in income accompanied by falling rates of emigration. This inverted-U relationship has been called the ‘mobility transition’ (Zelinsky 1971), ‘migration curve’ (Akerman 1976), ‘migration transition’ (Gould 1979), ‘migration hump’ (Martin 1993), and ‘emigration lifecycle’ (Hatton and Williamson 1994).

All theories of the mobility transition posit other forces, accompanying economic growth, that exert a separate, upward pressure on emigration. At low levels of income, this upward pressure tends to dominate the emigration-deterrent effect of a rising reservation wage. For this reason there is no necessary conflict between the ‘neoclassical’ model and the ‘mobility transition’ model. In the language of calculus, the partial derivative of emigration with respect to income can be negative—all else equal—while the total derivative is either positive or negative—because all else is not equal. In the language of economics, observed mobility transitions do not refute the neoclassical model for the same reason that the existence of inferior goods (in the presence of substitutes) does not refute the basic theory consumer demand (in the absence of substitutes).

Different scholars have stressed different theories for a mobility transition. Many of these are reviewed by Massey (1999), Lucas (2005: 50–51), Hatton and Williamson (2005a: 64, 252), and Williamson (2014). Here I briefly but comprehensively survey these theories.

4.1 Demographic transition

Rising incomes can be associated with demographic changes that favor emigration. At early stages of economic development, rising incomes generally cause child mortality to fall (Pritchett and Summers 1996) well before fertility consequently falls (Lee 2003). During this demographic transition, then, a rising population and youth bulge can accompany continued economic growth. If wages are downwardly
rigid this can mean rising unemployment and thus emigration pressure—compounded because younger workers are much more likely to migrate internationally than their older counterparts. This exerts upward pressure on emigration even as average incomes at home rise. As the demographic transition ends, emigration falls with further economic growth. Easterlin (1961) first suggested this explanation for concurrent rises in incomes and emigration, which was developed by Tomaske (1971). It is discussed as a theory of the mobility transition by Zelinsky (1971), Hatton and Williamson (1994), Lucas (2005: 50).

4.2 Credit constraints

Rising incomes can help potential migrants finance the costs of international mobility. Credit markets in developing countries are incomplete and highly distorted. While international migration often raises workers’ incomes by hundreds of percent (Clemens, Montenegro, and Pritchett 2008), these benefits occur well into the future. Prior to migration, potential migrants must finance costly prerequisites that can include overseas search, insurance, transportation, education, visa fees, passports (McKenzie 2007), and smuggling (Salt and Stein 1997).

So first, as development of the origin country proceeds, more and more households may be able to finance these costs simply because growth brings them disposable income (Vanderkamp 1971, Faini and Venturini 1994, Hatton and Williamson 1994, Ghatak and Levine 1994). Second, and compounding this effect, previous migrants often help finance the costs of migration by their family and friends (Gould 1980: 293, Hatton and Williamson 1994). Third, development at the origin brings deeper financial markets, which can be used to finance borrowing for migration. Fourth, credit constraints could produce a mobility transition by another channel. If migrants use overseas work to finance profitable investments in the origin country, development at the origin can raise the returns to those investments. For example, a credit-constrained worker wishing take advantage of home-country growth by opening a business may migrate to earn startup capital (Lucas 2005: 51).
4.3 Information asymmetry

Once a few workers have migrated, they can accelerate migration even as origin-country incomes rise, by providing information to potential migrants (Greenwood 1969, Massey 1988). Prior migrants can provide information about job search, earning potential, cost of living, marriage partners, legal formalities, and extralegal channels of movement. This can lead to an S-shaped diffusion process of cumulative migration stocks, equivalent to an inverted-U timepath of migration flows coincident with economic growth.

Epstein (2008) builds a formal model of how information asymmetries across borders can generate migrant herding and a mobility transition, even as growth is occurring at the origin. Previous migrants can also reduce potential migrants’ need for information, such as by offering in-kind insurance against risk—offering food and housing during unemployment (Martin and Taylor 1996). These mechanisms for the mobility transition are discussed less formally by Gould (1980: 293) and Baines (1994).

4.4 Structural change and worker dislocation

Economic development can be associated with structural change that alters the costs and benefits of emigration. In the development process new sectors rise and old sectors—particularly agriculture—decline. This can encourage geographic mobility as the opportunity cost of departing declining sectors falls, changing the relative costs and benefits of both domestic and international migration. “The onset of modernization … brings with it a great shaking loose of migrants from the countryside” (Zelinsky 1971). The process of development “displaces many people from traditional livelihoods and past ways of life. … Most become internal migrants, … [b]ut some always migrate internationally, seeking wider opportunities in more dynamic economies abroad” (Massey 1988).
Moreover, these changes can encourage specifically international labor mobility when they occur in the context of liberalizing trade, either through lower policy barriers or lower transportation costs. Venables (1999) reviews how this can occur in various workhorse theories of international trade. For example, while in basic trade theory the liberalization of goods trade tends to equalize the price of the most intensively-used factor between home and foreign, with sector-specific factors the prices of less-intensively-used factors may not equalize. A simple case is that trade liberalization of agriculture trade between two countries (specific factor: land) can create pressure for labor mobility from the agriculture importer to the exporter—if the exporter’s initial endowment of labor is relatively low.

Faini and Venturini (1993) and Martin and Taylor (1996) discuss various other mechanisms in this family as theories of the mobility transition. For instance, if infrastructure in the migrant-origin country is extremely poor, this reduces their productivity at home. If trade is liberalized between the migrant-origin and migrant-destination country, firms in the destination country might be able to out-compete firms at the origin by employing migrants made more productive by superior infrastructure. Similar country-specific differentials in worker productivity could arise through economies of scale in production, or barriers to technology diffusion such as intellectual property law. These too would create migration pressure even as trade raises economic growth at the origin.

4.5 *Inequality*

Economic development can be associated with changes in the distribution of income that affect the demand for migration. Because rising incomes are never uniformly distributed, “development is apt to be characterized by a growing gap between expectations and achievements—that is, by an increased awareness of insufficiency and a decreased tolerance of both poverty and privilege” (Heilbroner 1963: 26). As average incomes rise, a group’s reference level of satisfactory income can rise faster than the incomes of some members of the group. Thus rising average incomes can be accompanied by rising
numbers of the relatively deprived, who might seek to join a different reference frame by moving. And even if potential migrants care about their absolute rather than relative gains, the Roy model as elaborated by Borjas (1987) can still yield a mobility transition via inequality dynamics: changing inequality as development proceeds at the origin can mean that the gains from migration accruing to workers at different points in the income distribution can shift over time—raising and lowering migration rates.

The economic theory behind these forces is extensively elaborated (Stark 1984; Stark, Taylor, and Yitzhaki 1986; Stark and Yitzhaki 1988; Stark, Taylor, and Yitzhaki 1988; Stark and Taylor 1991; Stark 2006). A broad prediction of these models is that emigration is positively associated with inequality in the country of origin. There is substantial evidence of patterns in the level of inequality at different stages of development (e.g. Thornton 2001, Frazer 2006), corroborating Kuznets’ (1955) hypothesis that early stages of development tend to cause rising inequality—though the evidence on falling inequality at later stages of development is weaker (Piketty 2014). The effect of growth trajectories on inequality is not settled, but any effect at all could shape the mobility transition.

Feedback effects could further heighten or reduce the effect of inequality on emigration. Migration could raise or lower inequality at the origin, depending on where in the income distribution migrants come from, how much return migration occurs, and how much migrants earn (Massey 1988). This mechanism is discussed as a theory of the mobility transition by Martin and Taylor (1996). Gould (1980: 295) appears to be the first to discuss links between emigration and inequality.

4.6  Immigration barriers abroad

All of the preceding theories generate a mobility transition due to changes in the demand for migration. A mobility transition can also be shaped by changes in the supply of legal migration opportunities (Hatton and Williamson 2005a: 240–1, Hatton and Williamson 2010). All important migrant-destination countries
have visa classes that are more easily available to high-income workers than low-income workers. These include ‘points’-based settler visas, skilled employment-based work visas, temporary business visas, student visas, and investor visas.

Many non-visa policies likewise constitute de-facto migration barriers (Czaika and de Haas 2013). These too can impose constraints on migration from poorer countries that can decrease as they develop. For instance, many of the poorest African countries lack an internationally-accredited nursing and medical schools, making it difficult for health professionals from there to obtain licenses to work abroad—even if they can get a visa on other grounds. To the extent that international accreditation of training facilities tends to expand with economic development, this too could contribute to a mobility transition.

In this short review, it may seem fragmentary to offer six classes of theory. It reflects a fragmentary literature, despite longstanding aspirations of synthesis. “Indeed, we have no theory of migration as of yet!” (Lucas 1978). “At present, there is no single, coherent theory of international migration, only a fragmented set of theories that have developed largely in isolation from one another” (Massey et al. 1993). Many years later, these assessments remain valid.

5 Empirical tests of the income-elasticity of migration

Researchers have studied the quantitative empirical link between income and migration over four decades. Much of this work analyzes ‘macro’ data—aggregate flows out of some broad geographic area such as a nation, province, or state. Some more recent work analyzes ‘micro’ data—propensity to migrate across individuals, households, or villages.

This section points out a few strong patterns that describe the empirical literature as a whole. I identify these patterns with the knowledge that this literature defies concise summary. The studies use a wide
variety of locations, time periods, indicators of ‘development’, indicators of ‘migration’, estimation methods, and functional forms. Many issues thus remain unsettled, as I will discuss in the section to follow.

5.1 Macro studies

The macro literature begins with Vanderkamp’s (1971) first regressions of aggregate area-to-area migration flows on average incomes at the origin. He studies interprovincial migration flows within Canada, in repeated cross-sections 1947–1966. Lucas (1975) is the first to investigate the question at the international level, studying the 1973 cross-sectional association between origin-country income per capita and the flow of nontemporary labor certification applications to the United States from 103 countries. More than a dozen studies have followed.

A coarse summary of this literature is that cross-sectional studies generally find either a positive or inverted-U relationship between income and emigration from developing countries, while time-series studies do not find a consistent relationship between income and emigration. I will detail these descriptions and then propose an explanation for the broad disagreement: that the duration of the mobility transition is much longer than the data used in most time-series studies.

Macro studies in cross-section generally find a positive linear or inverted-U relationship between emigration and income across large numbers of developing countries. That is, studies that test only a linear income term generally find a positive coefficient, and studies that also include a squared income term find that the linear term is positive and the squared term negative. I include in this group pooled panels without fixed effects, in which between-group variance in income is typically much larger than within-group variance. Some of these find a positive relationship between emigration and linear income per capita at the origin and do not test for a nonlinear effect (Vanderkamp 1971; Bertoli and Fernández-
Huertas 2012a [without fixed effects], Ortega and Peri 2013 [without fixed effects]), while others include a squared income term and find an inverted-U relationship (Adams and Page 2003, Letouzé et al. 2009; de Haas 2010; Djajic, Kirdar, and Vinogradova 2013 [for low-skill majority of migrants]). Lucas (2006) finds that African countries above PPP$2,000 per capita income have higher emigration rates to the OECD than those below this line, though this pattern does not hold for intra-African migration.

The emigration-development association remains positive among developing countries in cross-section in studies that proxy for ‘development’ with an indicator other than income per capita. They find a positive linear relationship between emigration and measures of industrialization in output and the labor force (Greenwood and McDowell 1991), the inverse of a poverty measure (Hatton and Williamson 2005b), and the Human Development Index (Czaika and da Haas 2012). Two studies find a positive linear relationship between emigration rates and foreign aid receipts, in sub-Saharan Africa (Belloc 2014) and across developing countries in general (Berthélemy, Beuran, and Maurel 2009).

This consensus finding of the cross-sectional studies accommodates the three studies that do not find a positive relationship, because each of the three is incommensurable with the preceding studies. First, Lucas (1975) finds a negative linear relationship between nontemporary employment visa certification applications to the US and income per capita of the origin country, but this result is not commensurable with those above because 1) this visa covers only 7% of 1973 US permanent visa applications, most of which were related to family reunification, and 2) it reflects visa applications rather than movement. Second, Karemera, Oguledo, and Davis (2000) find a negative linear relationship between migration to the United States and origin-country income per capita (though the same relationship is positive for migration to Canada). But neither is this result commensurable with the above findings; it obtains only when controlling for several origin-country traits that are related to the level of development—including unemployment, inflation, a civil liberties index, financial ratings, and region fixed effects. It is thus unclear how to interpret the coefficient on income. Third, Belot and Ederveen (2012) regress bilateral
migration flows on origin-country income and find a negative linear relationship. But the sample comprises exclusively migration flows between high-income OECD countries. The setting, too, is thus incommensurable with that of the above studies. In high-income OECD countries many of the theoretical mechanisms for the mobility transition—such as the demographic transition and credit constraints—would be much less important.

The pattern is quite different in time-series macro studies, which begin with Wilkinson (1970). Time-series studies do not consistently find a positive (or negative) relationship between emigration and the level of income. Here I use ‘time-series’ to refer to macro studies that discard between-group variation in the level of initial income—either by controlling for country fixed effects, controlling for lagged emigration, or regressing on changes in income rather than levels of income. Some of the same studies that find a positive linear relationship between emigration and income in pooled data across all countries, dominated by between-group variance in income, find a negative relationship when country fixed effects are added (Bertoli and Fernández-Huertas 2012a, Ortega and Peri 2013).

But the findings of other time-series studies are mixed. There is no relationship between linear origin-country income per capita and migration in Mayda’s (2010) study of migration from all countries to the OECD, in Naudé’s (2010) study of emigration from sub-Saharan Africa, or in Hanson and McIntosh’s (2012) study of migration to the US, UK, Canada, and Spain—all of which include origin-country fixed effects. Again with fixed effects, Vogler and Rotte (2000) find an inverted-U relationship between origin-country incomes and migration to Germany from 86 Asian and African countries, and Telli (2014) offers similar findings for migration to the UK, while Clark, Hatton, and Williamson (2004, 2007) find a negative linear relationship for migration to the United States. Hatton and Williamson (2003) find a negative relationship between income growth and emigration from 21 African countries, but do not test the level. Faini and Venturini (1993, 2010) find an inverted-U relationship between time-series emigration from Southern Europe and incomes at the origin, while Hatton and Williamson (2011) find an inverted-U
relationship between time and migration rates from all countries to the United States, controlling for origin-country fixed effects and for lagged migration. Carlos (2002) finds a negative linear relationship between income per capita in the Philippines and emigration from the Philippines alone to 26 destination countries, including destination fixed effects. In pre-1914 data, Wilkinson finds a typically negative relationship between annual emigration rates from Europe and origin-country incomes, but Hatton and Williamson (1994) find a positive, insignificant linear relationship between emigration and origin-country real wages, while finding an inverted-U relationship with time.

Why do the time-series studies show such a wide range of inconsistent findings, while the cross-section studies consistently find a positive linear or inverted-U relationship? One explanation is that the time-series dimension of most of these studies is much shorter than a plausible timescale for the mobility transition. Most of the time-series studies have a time dimension in the range of 15–25 years. If indeed the typical duration of the mobility transition is on the order of two centuries, as estimated earlier—a century to reach the peak, another to fall—each observation of a country over time captures only a small slice of the transition. Such a slice, even if the overall transition follows an inverted-U, can be locally close to linear and either rising or falling monotonically.

Figure 3 illustrates this explanation, showing first a single hypothetical dataset in cross-section and then the same data in time-series. Each line segment between two dots represents one country. In that idealized situation, a cross-section or pooled regression would show an inverted-U relationship. A linear regression would show a positive relationship if the richest countries on the right-hand side were omitted. This is what is found in the cross-section or pooled studies. But the same dataset in time-series would yield no clear pattern in the coefficient—it would be roughly zero, somewhat positive, or somewhat negative. The time-series studies broadly corroborate this pattern.
Another potential reason for the markedly different pattern in time-series studies is that many of them use high-frequency data—often annual (e.g. Wilkinson 1970; Clark, Hatton, and Williamson 2007; Mayda 2010; Ortega and Peri 2013). When between-group variance in incomes is largely or completely controlled away—that is, when the effects of very long-term changes in income are controlled away—the income-migration relationship is more sensitive to annual fluctuations. If a country while under observation experiences one crisis that provokes a sudden fall in income and a sudden spike in emigration, this will tend to generate a negative relationship between emigration and income even if both shortly return to pre-crisis levels. It is theoretically plausible that for a poor country, the fifty years of positive growth can increase migration in the long run (through the mobility transition) but one year of negative growth can increase migration in the short run (as workers temporarily flee the crisis). The balance between these two effects shifts when we move from cross-section to time-series data. Observing countries over time series that are short relative to the transition duration tends to wipe out the longer effect, leaving evidence only of the shorter effect.

The literature to date offers limited evidence on the six theoretical mechanisms discussed earlier. Many of these studies test the income-emigration relationship only incidentally to some other purpose, without setting out to characterize the mobility transition. Some studies finding a positive emigration-income relationship interpret it strictly as evidence of credit constraints (e.g. Vanderkamp 1971; Djajic, Kirdar, and Vinogradova 2013). Carlos (2002: 93) interprets her negative relationship between average earnings in the Philippines and emigration rates as a sign that Filipino migrants are not credit-constrained, and are able to finance their moves via family, informal credit institutions, or their overseas employers. Other studies fit this pattern as well: Lucas (2006) finds that richer African countries have higher emigration rates to the OECD, but not higher emigration rates to nearby African countries, consistent with an important role for credit and information constraints in shaping emigration rates. Belot and Ederveen (2012) find no sign of a mobility transition in bilateral flows amongst the richest countries, also consistent with an important role for credit constraints in shaping the transition in poor countries.
A handful of macro studies do set out to test different mechanisms for the mobility transition against each other. Hatton and Williamson (1994) find that in pre-1913 European emigration, the inverted-U “emigration lifecycle” was shaped by structural transformation and demographic change at the origin and by networks of previous migrants at the destination (helping migrants reduce credit constraints and information asymmetry). Hatton and Williamson (2003) find that “exactly the same forces” shaped emigration rates from Africa in the late 20th century, and predict that “there is a good chance that by 2025 Africa will record far greater mass migrations than did nineteenth-century Europe.” Hatton and Williamson (2005b) find that three of the biggest forces shaping the migration transition worldwide since 1850 are “poverty constraints on the move in poor sending countries; the size of the young adult share in sending and receiving country populations; and the size of the foreign-born migrant stock from the sending counties currently residing in the receiving countries.” But the precise relative contributions of each of these forces to the mobility transition, or their importance relative to that of immigration barriers abroad or inequality shifts at home, is difficult to judge in existing work.

Separately, a substantial literature tests the effects underlying some of the posited mechanisms for a mobility transition, without directly testing them against other possible mechanisms for the transition. For example, various recent studies test the effects of migrant networks and visa policy on migration flows (subsections 4.2, 4.3, and 2.6 above), but do not assess the ability of these effects to explain the mobility transition (e.g. Mayda 2010, Bertoli and Fernández-Huertas 2012b, Ortega and Peri 2013).
Figure 3: Why a long-term migration transition can be difficult to detect in time-series data

The two parts of the figure show the same, hypothetical data, in two ways. The upper part shows a hypothetical transition in emigration stocks or flows over many years, in a cross section of ten curve segments. The lower part shows the same ten curve segments in time series. Imagine a regression of emigration on time, or on anything generally rising over time, such as income per capita. In cross section, the regression would yield an inverse U over 200 years. In time series, the same data over 20 years would yield no relationship. Any curvature would be difficult to detect.
A somewhat smaller literature explores the relationship between economic resources and emigration rates at the micro level—where the unit of analysis is workers, households, or villages. The micro literature begins with Lucas (1985). Of the 24 studies I am aware of, the large majority (17) study cross-sectional data rather than panels, and 10 of the 24 study data from Mexico. The measure of economic resources varies greatly across studies. It includes both stocks (land holdings, land access, education, public infrastructure) and flows (income, consumption, expenditure, cash transfers), and is measured by some studies in absolute terms, by others in relative terms. This heterogeneity makes the studies difficult to compare systematically, though some very broad patterns emerge.

Most of these studies find evidence of a positive relationship between micro-level migration and some measure of wealth or income, in settings where credit constraints are plausibly important. Bilsborrow et al. (1987) find an inverted-U relationship between land assets and rural-urban migration in Ecuador. Stark and Taylor (1989, 1991) find a positive linear relationship between household wealth and migration from a small region of Mexico to the US, and Stark and Taylor (1991) find an inverted-U relationship between nonremittance income and migration. Various studies find inverted-U relationships between emigration and landholdings in El Salvador (Halliday 2006), nondurable household consumption in Mexico (McKenzie and Rapoport 2007), household consumption in India (Czaika 2012), and land holdings and rice income in Indonesia (Bazzi 2013). Beauchemin and Schoumaker (2005) find that in Burkina Faso, “contrary to policy expectations, most components of rural development either have no effect on migration or rather tend to encourage migration to cities.” Chort (2012) finds that household expenditures are positively associated with emigration in a panel of Mexican households. Both Azuara (2009) and Angelucci (2013) find that conditional cash transfers in Mexico caused higher migration rates from poor rural communities, contradicting earlier work by Steckov et al. (2005).
In a few micro-studies the findings are less clear-cut, but nevertheless often show patterns compatible with the importance of credit and information barriers. Lucas (1985) finds a negative relationship between cattle assets and rural-urban migration in Botswana, but no relationship between cattle assets and longer-haul migration to South Africa. The greater travel cost and information barriers associated with long-distance migration are a plausible reason for this pattern. Orrenius and Zavodny (2005) find that unauthorized migration from Mexico is positively associated with agricultural GDP per capita in Mexico for migration trips longer than three years, but negatively associated for shorter trips. McKenzie and Rapoport (2010) find a positive association between human capital stocks and emigration in Mexican households, but only in areas of low migration prevalence—where credit and information barriers are plausibly highest. Kennan and Walker (2011) find a negative association between income and intra-US migration in a long panel, but credit constraints are less likely to bind within a rich country. In two studies using nineteenth-century micro data, Abramitzky, Boustan, and Eriksson (2013) find that household wealth has a negative linear association with emigration from Norway—interpreting this as evidence that credit and information constraints had been reduced by prior migration of many Norwegians; while Groote and Tassenaar (2000) find that the absence-of-hunger tended to encourage emigration from an impoverished portion of the rural Netherlands.

Finally, two of the micro studies investigate the association between exclusively relative economic resources and migration, finding that lower migration rates are associated with greater relative access to land in Nepal (Bhandari 2004) and higher relative wages in Mexico (Quinn 2006). But these studies control for absolute measures of assets (e.g. toilet facilities in Bhandari 2004, household education in Quinn 2006), so these relative findings are compatible with the existence of (for example) a mobility transition due to credit constraints in absolute terms.
Few of these studies seek to explicitly test different mechanisms for a micro-level mobility transition against each other. Many interpret their findings as evidence of the degree to which households are credit-constrained (e.g. McKenzie and Rapoport 2007, Abramitzky, Boustan, and Eriksson 2013), but other mechanisms for a mobility transition are less studied. Amuedo-Dorantes and Mundra (2007) find that stronger social networks among Mexican migrants are associated with higher earnings in the United States, but do not explicitly test this as a mechanism for the transition.

5.3 *Inequality and migration*

A number of macro and micro studies investigate the effect of inequality on emigration—and vice versa. I briefly summarize these because they are relevant to the theoretical channel by which shifts in inequality could shape the mobility transition. But most of these do not explicitly test the degree to which changes in inequality could explain the mobility transition relative to other explanations.

The literature begins with Gould (1980: 295), who is the first to empirically estimate the emigration-inequality link in historical macro data. In other macro work, Adams and Page (2003) find a positive association between the origin-country Gini coefficient and emigration and Czaika and de Haas (2013) likewise find a positive relationship between origin-country indices of relative deprivation and bilateral migration stocks. There is an inverted-U relationship between the relative Gini coefficient at the origin (relative to the destination) and migration in Hatton and Williamson (2005b), Clark, Hatton and Williamson (2007), and Mayda (2010).

The micro studies begin with Stark and Taylor (1989, 1991), who find a positive association between purely relative deprivation and emigration from a small part of rural Mexico, just as Bhandari (2004) and Quinn (2006) find in other settings. Reversing the direction of causation, McKenzie and Rapoport (2007) find that migration tends to reduce inequality in Mexico. Quinn (2006) finds the opposite, though the
results are difficult to compare since the two studies use different surveys with different sampling methods, different measures of migration, and different measures of inequality.

6 What we do not know about the mobility transition

Much remains to be done. The theoretical literature on the mobility transition is somewhat well-developed, though parts could benefit from clearer formulation through formalization. The empirical literature has made progress in detecting the exist of a mobility transition, in macro- and micro-level data, but less progress in definitively sorting out the relative importance of the six classes of mechanisms theoretically capable of generating a transition.

Many important questions lie at this frontier of sorting out the mechanisms of the transition. Here I discuss five questions. All relate to understanding the mechanisms that create the transition.

1. Will the same relationship continue in the future? To the extent that the drivers of the mobility transition change, the transition could have a very different shape in the future than in the past. For example, the relationship between demographic change and economic growth is likely to be different in coming decades than in the late 20th century, and even quite poor countries could proceed through a demographic transition as shifting Preston curves alter past relationships between economic growth and child survival. This could reshape the inverted-U (Lucas 2005: 50–51). A recent literature has shown interest in using the mobility transition to predict, at least in sign and in rough magnitudes, future changes in migration for Mexico (Chiquiar and Salcedo 2013), Morocco (de Haas 2007a), and the world (Hatton and Williamson 2011). These predictions will be more accurate to the extent that we understand the relative contributions of the six theories of the mobility transition, and how those differ among settings.
2. *Is the mobility transition shaped primarily by supply or by demand?* Most theories of the mobility transition express interest in understanding the determinants of ‘migration pressure’ or the willingness of workers in developing countries to supply their labor to other countries. But in nearly all empirical work we observe realized migration flows, not willingness to supply labor. Migration flows are determined by the intersection of labor supply to overseas jobs and overseas demand for foreign labor. That demand can be influenced by many natural forces characteristic to the destination country—demographic change at the destination, language, geography. Demand for overseas labor is also heavily regulated by a worldwide thicket of visas, quotas, licensing restrictions, interdiction ships, and physical barriers. These barriers, as discussed above, can also shape the mobility transition. One approach to isolating the effects of labor supply determinants from labor demand determinants is to explore the relationship between economic development and expressed interest or intent to migrate, either in micro data (Van Dalen, Groenewold, and Schoorl 2005; Arnold, Hamilton, and Moore 2011) or at the national level (Pelham and Torres 2008). But this is suboptimal for various reasons. People’s interests and intentions are shaped by what they perceive as realistic, which in turn can be shaped on overseas restrictions on demand for foreign labor. Expressed intent is not, then, a pure measure of labor supply. Another problem is that expressed interest in migrating can be quite different from actual migration behavior (Chort 2012; Beam, McKenzie, and Yang 2013). Creative empirical strategies are needed to identify supply effects and demand effects on the shape of the mobility transition.

3. *Is there a micro-macro conundrum?* We lack a unified theory that can explain the shape and characteristics of the mobility transition at both the micro and macro scales, accounting for both changes in absolute income and relative income. Observe the following pattern in data from the Gallup World Poll, which asks adults in most countries if they would like to emigrate (“ideally, if you had the opportunity”): “In countries that are at the bottom quartile in GDP worldwide, desired migration rates are about twice as high as they are in countries in the top quartile in GDP.
However, for countries at every level of GDP, it is generally those individuals who are richer rather than poorer (by the standards of their own countries) who are more likely to say they wish to migrate” (Pelham and Torres 2008). It is no doubt possible to construct a single model that coherently displays all of these features—predicting that desired labor supply to overseas work falls when richer respondents are among richer people, but rises when respondents among poorer people get richer. Such a synthesis, coherently explaining all of these features at all scales, has not been carried out. It would generate a rich set of empirically testable predictions.

4. **What determines the relative importance of different mechanisms in different settings?** I have discussed the fact that few empirical studies have set out to test different mechanisms for the transition against each other. A handful of studies do this at the macro level and none at the micro level. None is set up to comprehensively test the relative contributions of all six mechanisms in a single empirical setting. Much less is any set up to assess why the relative contributions of different mechanisms would differ across different empirical settings.

5. **What are welfare-optimal policy responses to the mobility transition?** To state a fact bluntly, much of the research interest in the mobility transition has been spurred by a broad policy interest in finding ways to deter poor people from turning up at airports and beaches in rich countries. “[M]any of the policy statements about migration suggest that investing in development is a means to reduce migration” (Bakewell 2008). There is no empirical economic basis for this view, after decades of testing, at least on a timescale that is meaningful to most politicians. The weight of our current evidence is that this view is generally false. This leaves a hole in policymakers’ toolkit for encouraging welfare-optimal outcomes at the interface between migration and development. What could fill the hole? Martin and Straubhaar (2002) review some of the alternatives, and conclude: “The best practices to manage migration lie primarily in two areas: maximizing the amount and payoff from remittances and encouraging nationals abroad to return
and assist in the development of their areas of origin”. Other policies designed to accommodate greater flows as development proceeds are discussed by Pritchett (2006) for low-skill workers and Clemens (2009) for high-skill workers. Much more and more creative policy development is needed, supported by the best evidence.

7 Summary

45 years of research have made progress on understanding the mobility transition. We have six classes of theory that can explain aspects of the transition, none of which are mutually exclusive. We have recently-compiled data showing a clear bivariate inverted-U pattern for emigration stocks and flows in cross-section, across the full range of observed GDP per capita.

We have a range of macro studies showing strong evidence of a tendency for emigration stocks and flows to be higher among developing countries that are progressively richer in cross-section, though many of these do not test for the decline in the right half of the inverted-U. There is no evidence that emigration rates tend to fall as countries get richer, until they are above an income per capital PPP$7,000–8,000. We also have a growing number of micro studies, from several countries, showing that among poor households the richer ones in absolute terms display a greater tendency to migrate to faraway rich countries.

We have more mixed results from macro studies in time series, but this appears to have an explanation. The cross-section evidence suggests that most of the datasets used in time-series studies do not cover a sufficient length of time to describe the shape of the transition. Very little of the micro evidence is in time series, as household panels that track international migration were very rare until recently.
All of this suggests that today’s typical poor countries are unlikely to exhibit lower emigration rates as they develop—1) until they reach income levels in the upper range of middle-income countries, and 2) unless they follow patterns of demographic and structural transition markedly different from developing countries in the past. If anything they are likely to exhibit somewhat higher rates of emigration as they develop.

We do not know enough about the mechanisms that create this observed pattern. Theories of the transition are well-developed, though they could benefit from more formalization and unification in a single framework that can explain patterns observed at both the macro- and micro-levels. And we know much too little about the relative importance of the six classes of theoretical mechanism that have been identified, what makes their relative importance change in different settings, and how they might alter the shape of the mobility transition in the future. Empirical work would move forward fruitfully with research designs set up to test the full range of mechanisms against each other.

But we do know enough to begin to advise policymakers more confidently on the broad outlines of the mobility transition. 26 years ago, Massey (1988) wrote,

“A realistic appraisal of the relationship between migration and economic development, therefore, suggests that policies to promote additional economic growth in sending nations will not reduce immigration to the United States in the short run; indeed, they may increase it. … If it is in the interest of the United States to promote rapid economic development in Mexico, then it is also in its interest to accept relatively large numbers of Mexican immigrants”.

But policymakers in the aid and development world have not been guided by such research. 20 years forward, de Haas (2007b) laments the continuing prevalence of unsupported ideas:
“In the poorest countries, especially (such as the sub-Saharan African countries which are the target of much international aid), any take-off development is likely to lead to accelerating take-off emigration for the coming decades, which is the opposite of what ‘development instead of migration’ policies implicitly or explicitly aim to achieve” (de Haas 2007b).

In short, there is enough evidence now for researchers to be confident that the mobility transition exists, and for policymakers to move beyond ‘development instead of migration policies’. Further research progress will help us better characterize the transition, especially the mechanisms that blend to create it.

“This might make it easier to avoid the assumption that mobility is normal for the wealthy, international elite, but a symptom of failure among the poor.” (Bakewell 2008)
References


Hanson, G. H., & McIntosh, C. (2012). ‘Birth Rates and Border Crossings: Latin American Migration to the US, Canada, Spain and the UK,’ Economic Journal, 122(561), 707-726.


White House (1993), ‘Remarks by President Clinton, President Bush, President Carter, President Ford, and Vice President Gore in signing of NAFTA side agreements,’ Office of the Press Secretary, September 14.


Appendix: Data Sources and Full Nonparametric Regressions


Total emigrant stock is defined as the number of people born in a country residing in a different country. The emigrant stock rate is the total emigrant stock in a year divided by national population in the same year. The emigrant flow rate is the change in the emigrant stock over a decade, divided by national population at the start of the decade. For these calculations, the ‘World Bank’ numbers use the national population given in the Penn World Table (below); the ‘UN’ numbers use: United Nations, Department of Economic and Social Affairs, Population Division (2013). World Population Prospects: The 2012 Revision, DVD Edition.

Real GDP per capita is in 2005 PPP US$ and is taken from the Penn World Table version 8.0, as expenditure-based real GDP, “to compare relative living standards across countries and over time”, (Feenstra, Inklaar, and Timmer 2013) divided by total population. Accessed Dec. 27, 2013.

Figures A1 and A3 show the full data underlying the kernel-weighted local average curves for each cross-sectional nonparametric regression of emigrant stock on real income per capita in Figure 1, plus the associated 95% confidence interval. Figures A2 and A4 show the same information for the flow regressions in Figure 2. Countries are identified by their three-letter codes as in ISO 3166-1 alpha-3.
Figure A1: Nonparametric regressions of emigrant stocks on real income per capita, cross-sections by decade: World Bank data

Solid lines are Nadaraya-Watson local mean with dashed lines showing 95% confidence interval, Epanechnikov kernel, bandwidth 0.5 natural log points. Observations are country-years; $N$ is 106 for 1960, 140 for 1970, 140 for 1980), 164 for 1990, and 164 for 2000. Macau and Brunei omitted.
Figure A2: Nonparametric regressions of emigrant flows on decade-initial real income per capita: World Bank data

Solid lines are Nadaraya-Watson local mean with dashed lines showing 95% confidence interval, Epanechnikov kernel, bandwidth 0.5 natural log points. Observations are country-decades; $N$ is 106 for 1960–70, 140 for 1970–80, 140 for 1980–90, 164 for 1990–2000. Macau and Brunei omitted.
Figure A3: Nonparametric regressions of emigrant stocks on real income per capita, cross-sections by decade: United Nations data

Solid lines are Nadaraya-Watson local mean with dashed lines showing 95% confidence interval, Epanechnikov kernel, bandwidth 0.5 natural log points. Observations are country-decades; \( N \) is 155 in all periods.
Figure A4: Nonparametric regressions of emigrant flows on decade-initial real income per capita: United Nations data

Solid lines are Nadaraya-Watson local mean with dashed lines showing 95% confidence interval, Epanechnikov kernel, bandwidth 0.5 natural log points. Observations are country-decades; $N$ is 155 in all periods.