Hatton, Timothy J.

Working Paper
The cliometrics of international migration: A survey

IZA Discussion Papers, No. 4900

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Hatton, Timothy J. (2010) : The cliometrics of international migration: A survey, IZA Discussion Papers, No. 4900, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
http://hdl.handle.net/10419/36811

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.
IZA DP No. 4900

The Cliometrics of International Migration: A Survey

Timothy J. Hatton

April 2010
The Cliometrics of International Migration: A Survey

Timothy J. Hatton
University of Essex,
Australian National University and IZA

Discussion Paper No. 4900
April 2010

IZA
P.O. Box 7240
53072 Bonn
Germany
Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.
ABSTRACT

The Cliometrics of International Migration: A Survey*

This is a survey of some of the key studies in the literature on international migration in history that may be described as cliometric. This literature uses the concepts and approaches of applied economics to investigate a range of historical issues and there are strong parallels with the questions that have been addressed in the literature on contemporary migrations. Here I focus on the period 1850 to 1940 and chiefly on migration from Europe to the New World. The survey is organised around six themes that include: the forces driving migration, over time and across space; the assimilation of migrants and their effects on wages and income distribution in source and destination countries; and the evolution of immigration policy. While this literature has drawn heavily on the tool kit of applied economists it also provides a wider perspective on many of the issues that concern migration today.

JEL Classification: F22, N30, J61

Keywords: international migration, economic history

Corresponding author:

Timothy J. Hatton
Department of Economics
University of Essex
Wivenhoe Park
Colchester CO4 3SQ
United Kingdom
E-mail: hatton@essex.ac.uk

* This paper was prepared for a special issue of the Journal of Economic Surveys on Cliometrics. I am grateful for useful comments from Blanca Sánchez Alonso and Chris Minns as well as an anonymous referee.
Introduction

This survey provides an overview of the quantitative historical literature on international migration that has developed over the last 30 years or so. I focus on those contributions that are classified as ‘cliometric’ in the sense that they use explicit theoretical models and/or use econometric methods, or are to some degree quantitative. This historical literature has flourished side by side with a burgeoning literature on contemporary migration issues. As I will show, the historical literature has borrowed from, and has often been stimulated by, the work of economists concerned with current policy issues. But the traffic is not all one way.

I concentrate on six key themes that have been the subject of both historical and contemporary analysis. The first is to account for the trends and fluctuations in established migration streams. This was the focus of much of the early quantitative literature that followed from debates over the links between migration and business cycles, but which has not featured quite so prominently in the analysis of economists. The key issues there are whether push or pull forces dominate in migration and what other variables matter. The second theme asks who migrated and why? Why does the propensity to migrate differ across countries, regions, towns and villages and how does it vary between individuals? This literature seeks not only to understand what motivates migrants but also the forces that initiate migration streams, and what mechanisms cause them to persist.

While the first two themes concentrate on the causes of migration, the next two focus on the results of migration, particularly in the receiving countries. The third theme is immigrant assimilation. This has been one of the most fiercely debated issues among economists and it is the one where economic historians have borrowed most from the contemporary literature. The fourth theme is the effects of migration on the economy at large and in particular on the wages and living standards of those in the receiving country with whom migrants compete. Did mass migration foster wage convergence between sending and receiving countries? And how did it affect income distributions and skill premia? These too are highly contentious issues in the literature on recent migrations and no less so in the historical context.

The last two themes are areas where the cliometric literature is still relatively thin but which have attracted growing interest in recent years. The fifth theme is the political economy of immigration policy. This literature seeks to understand the forces that led to the adoption increasingly restrictive immigration policies in the half century after 1890. Can this be interpreted as a backlash to growing immigration pressures? And if so, what political-economy mechanisms were at work? The final theme is the study of international migration outside the traditional focus of European migration to the New World. As has often been noted, migration from and within Asia was even larger than emigration from Europe. While there is a rich historical literature that documents and discusses Asian migration, until recently, it has not featured widely in cliometric research.

As with any survey article, much has to be omitted. I focus chiefly on migration from Europe to the New World, from the mid-nineteenth century to the middle of the twentieth century, which is where the bulk of the cliometric literature has concentrated. As a result I ignore the
lively, but somewhat independent literature in the slave trade and I omit the significant literature on migration under indentured servitude before the mid-nineteenth century. I concentrate largely on intercontinental migration, leaving aside the literature on rural-urban, inter-regional and cross-border migration, a literature so rich and extensive as to deserve separate treatment.

**Push and Pull**

The earliest cliometric analyses focused on whether fluctuations in international migration streams were driven chiefly by ‘push’ factors operating at home or ‘pull’ effects from abroad. It originated in the debate between Jerome (1926) who argued that the timing of US immigration before 1914 was dominated by the American business cycle and Thomas (1941) who, in the context of Swedish migration, stressed the effect of home country factors. The literature that followed was facilitated by the monumental work of Ferenczi and Willcox (1929), which provided a rich time series database on emigration from Europe and immigration to the New World from the middle of the nineteenth century to the 1920s. The literature that subsequently developed in the 1960s and 1970s was critically reviewed as part of an excellent series of articles by Gould (1979).

The typical study in this genre specified a regression for the aggregate migration flow (or rate) from a source country to a destination country as a function of business cycle indicators (such as industrial output gaps or unemployment rates) at home and abroad and some measure of wage rates or incomes at home and abroad. These models are often estimated with lags to capture adjustment dynamics and sometimes also include a measure of the stock of previous migrants. In his critique of the earlier literature Gould pointed out that, where both home and foreign business cycle indicators are included, the latter typically dominate in terms of size and significance of the coefficients. He also noted that indicators of wage or incomes often fail to achieve significance in the present of business cycle indicators and typically fail to support the view that the incentive to migrate is positively related to incomes abroad and negatively to incomes at home. He concluded that: “By and large, reaction to this literature is one of some disappointment, for not only has it failed to generate important new insights...... it has had only limited success in confirming or denying old interpretations” (Gould, 1979, p. 668).

An important shortcoming in the studies reviewed by Gould is that they lack a coherent economic model of the emigration decision, which makes the results difficult to interpret. It is hard to believe that the decision to emigrate depends only cyclical conditions at the destination or, for that matter, exclusively on conditions at home (except perhaps in the case of war or famine). Similarly, emigration decisions must have been based on some comparison, however approximate, between wage rates or expected incomes at home and abroad, especially in the long run. Accordingly, Hatton (1995) developed a model of the migration decision using a simple microeconomic framework in which potential migrants base their decision on the comparison of future expected incomes at home and abroad.
Following Todaro (1969) expected income at a given location depends on the probability of employment and the wage rate. Because migrants are risk averse and because greater uncertainty attaches to the employment probabilities than to the wage rates, the former takes a larger ‘weight’ in the migration function. Dynamics are accounted for by an adaptive expectations process and the model also includes the migrant stock to represent the so-called friends and relatives effect.¹

This model provides a benchmark for evaluating the relationship among the coefficients attached to different variables in an econometric model of emigration. Estimates on annual time series for gross emigration from the UK between 1870 and 1913 broadly support these priors. Specifically they show that wage rates and employment rates, both at home and abroad all matter, with the coefficients on the employment rates exceeding those on wage rates. While short run fluctuations are largely driven by employment or activity rates (which typically are not trended), long run trends are determined by relative wage rates and the trend in the migrant stock, which is itself determined by previous emigration. Applying the same model to emigration from Ireland, Hatton and Williamson (1998, p. 83) found that the 17 percent fall in the (foreign to home) wage ratio between 1876/80 and 1909/13 accounted for a decline of 4 per thousand in the emigration rate. Over the same period the fall in the ratio of the migrant stock to home population contributed a similar amount to the fall in the emigration rate.

When it has been included in time series models, the migrant stock almost always proves to be very powerful. This is often interpreted as evidence of migrant networks generating strong persistence through chain migration. But as the migrant stock is essentially the cumulative flow it is sometimes difficult to distinguish network effects from the equation dynamics. In particular, if it is more recent immigrants that matter, then lags of the dependent variable may be picking up these effects. One important mechanism is the flow of remittances, a significant part of which took the form of pre-paid tickets and other assistance to new migrants. Magee and Thompson (2006a) provide new estimates of the flow of remittances to the UK from 1875 to 1913, which amounted to 3.4 percent of the value of exports in the latter year. Magee and Thompson (2006b) study the determinants of the flow of money orders, which they find to depend on the migrant stock and income in the migrant-receiving country. This would be

¹ The model is derived from a logarithmic utility function and assumes that future expected values of the wage and of employment probabilities are related with geometric lags to past values in an adaptive process. This gives the following model for migration (see Hatton and Williamson, 1998, pp. 61-63 for the full derivation):

\[ M_t = (1-\lambda)b_1\ln(w_t/w_{th}) + (1-\lambda)b_2\ln(e_t) - (1-\lambda)b_3\ln(e_{th}) + (1-\lambda)b_{5}\ln(M_{st}) + (1-\lambda)b_{6}t + \lambda M_{t-1} \]

where \( M \) is the emigration rate, \( w \) denotes wage rates and, \( e \), employment rates abroad (f) and at home (h). The terms in the lower line of the equation (with parameters \( \epsilon \)) are the costs of migration as reflected by the migrant stock (MST) and a time trend for falling transport costs. The last term is the lagged dependent variable arising from the adaptive expectations process with parameter \( \lambda \). In the upper line of the equation, the fraction \( 3/2 \) reflects the greater weight given to the employment terms as a result of employment uncertainty (arising from the concavity of the utility function). The coefficient \( \gamma \) allows for lower uncertainty to attach to home employment as compared with abroad. Thus it could range between 1 (equal uncertainty with abroad) and \( 2/3 \) (no uncertainty). Further dynamics are added to the estimating equation to reflect the option value of waiting—see Hatton (1995).
consistent with the idea that the effect of these variables on the flow of new migrants works partly through remittances.

A number of studies have identified other variables that determine trends and variations in time series migration rates. One important variable is the age structure of the population. Since the potential gains from migration are greater at younger ages demographic structure should matter independently of other variables in the model. Quigley (1972) found that demographic forces mattered for emigration from Sweden as did Larsen (1982) for Denmark. Using the rate of natural increase 25 years earlier as a proxy for the size of the birth cohort, Hatton and Williamson (1998, p. 72) find similar results for three Scandinavian countries. They estimate that between 1895 and 1905 lagged natural increase raised the emigration rate by 1 per thousand in Sweden, by 1.3 per thousand in Norway and by 0.3 per thousand in Denmark. However, these effects were weaker for countries such as Spain and Italy (Sánchez Alonso 2000b; Hatton and Williamson, 1998, p. 104). Interestingly, Greenwood (2007) finds that a higher current birth rate tends to reduce emigration of those at parenting age, probably reflecting the higher cost of family emigration. Overall, these studies offer some support for the hypothesis that demographic structure was an important determinant of emigration.

Any sensible economic model of migration must take account of the costs as well as the benefits. The costs of migration have been absent from most studies, because of the lack of suitable time series data. Yet there is reason to think they are important, and a full accounting of the costs would include not only the costs of an ocean passage, but also the overland costs at each end of the journey, and the living costs (or foregone income) while in transit. The cost of passage fell after 1850 with the transition from sail to steam, a decline that occurred later on the longer routes. And effective prices fell rather more steeply because of shorter passage times and better quality on-board accommodation (Keeling, 1999a; Sánchez Alonso 2007), not to mention overland travel. On the routes to North America there were sharp swings in passage prices associated with the effectiveness of shipping cartels (Keeling 1999b). Deltas et al. (2008) estimate the effects of cartels on quarterly data for the volume of migration on different shipping routes to the US and Canada between 1899 and 1911. Periods when cartels were in operation were associated with lower emigration, implying a price elasticity of about -0.7.

Policy related variables also influenced the costs of migration, either directly or indirectly. Some receiving countries such as Australia and Brazil directly encouraged immigration by providing assisted passages to selected groups of migrants. The programme of assisted passages in Australia substantially reduced the costs for UK migrants in 1911-13, and Pope

---

2 To illustrate, let the wage difference (destination minus source country) per year of working life be a constant D. If the age range of potential working-age migrants, a, runs from 20 to 65, and the discount rate is r, then the present value of the gains will be: $PV(a) = \frac{D}{r} \left[ 1 - (1+r)^{-65-a} \right],$ which is a decreasing function of a.

3 Sánchez-Alonso (2007) shows that fares from the UK to the USA fell from the 1850s, while those from Spain to Argentina, Brazil and Cuba declined from the 1870s. McDonald and Shlomowitz (1991) find no downward trend in the contract price of passages to Australia between 1847 and 1885.
(1981) found that this accounted for much of the surge in immigration in those years. For emigrants from Spain, the depreciation of the Peseta after 1892 sharply increased the costs of migration, which were denominated in gold currencies. Sánchez-Alonso (2000a) finds that the exchange rate had a negative effect on emigration from Spain, such that emigration would have been as much as 40 percent higher between 1892 and 1905 in the absence of the depreciation. By contrast, tariff protection increased emigration by more than 20 percent. On a more speculative level, Khoudor Castéras (2008) argues that, for Germany, the rise of social insurance and the fall in working hours contributed to the decline in the emigration rate from the 1880s.

Overall the push-pull model of international migration has worked pretty well on a variety of time series for different migration streams. But one issue that remains is why migration flows were so volatile. Emigration rates often increased or fell by a quarter or even a half in a year or two, only to recover again a few years later. In the time series models much of the volatility is explained by the cyclical employment variables and the equation dynamics. But this is itself a puzzle. Given that migration decisions were based on comparing future expected lifetime earning profiles, one might expect that short run changes, quickly reversed, would have little effect. One reason for the surprising short run volatility is the option value of waiting. While it might be worth emigrating today even though unemployment was high in the destination, it would be better still to wait a year or two if conditions were expected to improve. While it is tempting to think that emigrants timed their moves in order to maximise the life cycle benefits overall, we have little direct evidence for this.

Source Countries, Regions and Localities.

A greater challenge is to explain the differences in the levels and the trends in emigration rates across countries. There was certainly a wide variety of emigration rates from European countries in the late nineteenth century. The highest overall was Ireland where the annual gross emigration rate averaged 12 per thousand of the population between 1850 and 1913. Countries such as Sweden and Norway had rates approaching 5 per thousand between 1870 and 1913; the rates for Germany and Belgium were under two per thousand, while that for France was less than one per thousand. Furthermore the long run trends in emigration differed widely: the Irish emigration rate declined from the 1860s and the German and Norwegian rates declined from the 1880s. But at almost the same time the Italian and Spanish emigration rates began a steep ascent which was halted only by the outbreak of War, and a similar pattern is observed for a number of Eastern European countries.

One important stylised fact is that during the onset of modern economic growth in Europe, national emigration rates often rose, steeply at first from very low levels, rising more gently.

---

4 For example, Taylor (1994) finds that similar results emerge for migration to Argentina (largely from Italy and Spain) as for migration to Australia (from the U.K.).

5 For example, Deltas et al. (2008) found little evidence of inter-temporal substitution in response to shifts in the cost of migration.
to a peak, and then gradually falling. This evolution, often seen as a multi stage process, has sometimes been called the ‘mobility transition’ (Zelinski, 1971). Though not universal, such patterns have been identified in studies of aggregate emigration rates from a number of countries (Akerman, 1976, p. 25; Hatton and Williamson, 1998, p. 47). In his pioneering study of European emigration rates, Easterlin (1961) stressed the effects of population growth spilling over into emigration. On the other hand Tomaske (1971) found little effect for source country per capita income on cross country emigration rates. Thus income or real wage differences alone can explain little of the cross country variation in emigration rates—their effects can only be observed in the presence of other variables. But the early studies of emigration across countries and over time were constrained by lack of suitable data. The development of internationally comparable data such as Williamson’s (1995) database on real wages in Europe and the New World gave a fillip to further research.

Hatton and Williamson (1998) studied decade-average emigration rates from 12 European countries between 1860 and 1913 using a limited set of explanatory variables. In this analysis the real wage ratio (source to destination) has the expected negative effect. A ten percent increase in the wage ratio reduced emigration in the long run by 1.27 per thousand per annum—a substantial effect that is broadly consistent with the results from annual time series. But the share of the labour force in agriculture had only a weak negative effect—suggesting that, on balance, agricultural populations were less mobile than urban/industrial populations. The effect of lagged natural increase strongly supports Easterlin’s view that the demographic transition drove emigration. This estimate suggests that half of excess births spilled over into emigration—a very large effect indeed and one deserving further investigation.

It has often been observed that emigration gathered momentum first in the richer countries of Europe and, as it spread to the countries of Southern and Eastern Europe, it often coincided with a quickening in the pace of development. It has been argued that those with the greatest incentive to migrate were initially too poor to afford the costs of long-distance migration but economic development served to ease this ‘poverty trap’. Using annual time series, Faini and Venturini (1994) find for Italy, and Sánchez Alonso (2000b) finds for Spain, that the rise in emigration after 1880 was positively influenced by per capita income at home. For Italy, the gradual increase in the migrant stock also helped to ease the poverty trap as previous migrants helped to finance the passage of subsequent cohorts of migrants (Moretti, 1999). Thus the poverty trap would have been less binding the higher was the pre-existing migrant stock (Hatton and Williamson 2005, p. 65). In Ireland, the large migrant stock inherited from the Famine era meant that the poverty trap was much less binding from mid-century than it was for countries with fewer past emigrants. In contrast to Italy, the Irish emigration rate fell as conditions at home improved because the declining incentive to emigrate dominated the growing capacity to emigrate.

Looking at panel data for the same 12 countries Greenwood (2007) finds that the larger the manufacturing and agricultural shares in total employment the younger were the emigrants and the more they responded to job opportunities.
The wide diversity of regional emigration rates within any given country and their apparent convergence over time presents a similar puzzle. Gould (1980b) argued that this was due to a process of information diffusion, often originating from ports or trade routes, but as information is not directly measurable, this is rather hard to test formally. Baines (1994, p. 532) has argued that convergence was not universal among regional emigration rates in different European countries. Clearly, in some countries where emigration was relatively well established, the information diffusion hypothesis would have less force. Thus, across the 32 Irish counties, the coefficient of variation in emigration rates rose from 0.31 in 1881 to 0.71 in 1901, falling back to 0.41 in 1911. By contrast in Italy and Spain there was strong convergence. While greater literacy might have been expected to foster the flow of information, evidence for Italian provinces and for Irish counties indicates that emigration rates were not positively associated with literacy (Hatton and Williamson, 1998, Chs. 5 and 6). But, looking across Spanish provinces in 1888 and 1911, Sánchez Alonso (2000b) finds that the growth in literacy had positive effects on the emigration rate.

Studies of emigration across regions find that the degree of urbanisation has mixed effects (e.g. Lowell, 1987, pp. 48-52, 217-8). While urban populations were typically more mobile, they also enjoyed higher incomes than did rural dwellers. Key factors in rural areas include local demographic growth, which is almost always positive, and the patterns of landholding, which seems to have had different effects in different countries. For Norway and Sweden, Lowell (1987, pp. 212-16, 221) found that emigration was negatively related to local wage rates, but also positively related to the number of landless labourers and the share of land occupied by large estates. For Ireland, Hatton and Williamson (1998) found that the prevalence of smallholdings had a negative effect on emigration while measures of poverty had positive effects. Thus attachment to small parcels of land kept the Irish at home, but in Italy emigration increased with the incidence of owner occupation and sharecropping. The differing effects of variables reflecting small holding are puzzling. A partial explanation might be that in countries like Italy, with high rates of return migration, emigration was often undertaken with the objective of acquiring land on returning home.

It is difficult to advance much further without getting down to the local level, and in series of important papers Wegge (1998, 1999, 2002) studied emigration from the Hesse-Cassel region of Germany between 1832 and 1857. While those with resources such as land, which could be sold or mortgaged, would not face liquidity constraints they also had less incentive to move; by contrast unskilled labourers were more often constrained. As a result the highest emigration rates were among artisans—those with transferrable skills and enough resources to emigrate (Wegge, 2002). 7 Consistent with this, village level data indicates that higher wages and greater scarcity of land led to higher emigration. It also provides strong evidence that network effects were powerful; current emigration rates were strongly related to past emigration from the same village and from the surrounding county. In addition, those

---

7 Summarizing occupational data for ships lists Cohn (2009, Ch. 5) confirms this pattern for German immigrants but finds that the English, and especially the Irish, were rather less skilled than the populations from which they were drawn. The evidence also suggests that from around 1830 the skills of immigrants to the US declined as migration costs fell and mass migration accelerated.
emigrants that could be identified as networked carried less cash with them (Wegge, 1998). This further supports the idea that network effects helped to unlock the poverty trap. Another interesting finding is that emigration was higher from villages where the custom was for one son to inherit rather than for the previous generation’s assets (particularly land) to be evenly divided among the male siblings (Wegge, 1999).

There are three interrelated issues that remain under-explored in cliometric work on the causes of international migration. One is the role of politics and oppression. Some migrants left Europe for reasons of religious or political persecution—a parallel with today’s asylum seekers. Perhaps the most prominent example is the persecution of Jews that is often associated with the steep rise in emigration from the Russian empire after 1880. In her study of annual time series Platt-Boustan (2007) finds that the key events of 1891, 1903 and 1905-6 had substantial effects on emigration in the short run but that most of the upward trend in emigration was due to the standard economic variables.\(^8\) The political climate in host countries (such as the ebb and flow of nativist sentiment in the US) may also, at times, have had some influence (Cohn, 2000). Bertocchi and Strozzi (2008) examine the effects of host country political variables across 14 countries in Europe and the New World. They find that immigration was positively related to political participation (democracy and suffrage) and also to rights for immigrants (access to citizenship, land and education). Nevertheless the dominant forces were the usual economic and demographic variables.

A second issue is the choice of destination. Migrants from a given country went to a relatively narrow range of destinations. While the US was the main destination for most migrants from Europe, large numbers of British and Irish went to the British dominions and a large share of southern Europeans went to South American countries. It is often suggested that colonial ties, common language and cultural affinity meant that different destinations were poor substitutes (Taylor 1994), a feature that was reinforced by the friends and relatives effect. Yet there were potential substitutes among the English speaking countries and within South America. As Green et al. (2002) show, British migrants moved between Canada and the US, to where their skills were best rewarded. But while migrants from northern Italy (more often skilled and urban) went to South America, those from the Italian South (more often unskilled and rural) went increasingly to the United States. This apparent paradox may be explained by the fact that unskilled labour commanded higher wages in the US (and language was less of an impediment for unskilled employment), while the more skilled northern Italians moved easily into business and trade in Argentina (see Klein, 1983; Baily, 1983).

However, attempts to identify substitution between destinations in migration equations have not been very successful. Analysing the destinations of emigrants across 69 Italian provinces, Hatton and Williamson (1998, p. 119) found that the choice between Latin America and the United States was driven mainly by the share and type of agricultural employment and by the destination choice of past emigrants. In a recent paper, Balderas and Greenwood (2010)\(^8\) These events are the expulsion of Jews from Moscow in 1891; the Kishinev massacre of 1903; and the pogroms that followed the riots of 1905.
analyse times series for emigration from 12 source countries to three destinations: Argentina, Brazil and the USA. Like other studies they find relative wages and the migrant stock mattered, with mixed results for economic activity. Using instrumental variables they examine the effect of migration to one destination on migration to other destinations in order to measure the substitution effect. But they find little evidence of substitution between Argentina and Brazil or between either of these and the United States.

A third issue is return migration. As Gould (1980a, p. 50) remarked: “If the immigrants came, as so many models assert, because of higher wages and better job opportunities in the USA than in Europe, why did so many go back? As obvious a question as this has been totally ignored by the majority of econometric studies on Pre world War One migration.” Gould’s question is still as apposite now as it was 30 years ago. Some migrants moved with the seasons⁹ others timed their stay according to essentially the same variables that drove outward migration. Over time, return migration increased as transport costs fell and as the share of Southern and Eastern Europeans rose. A key indicator of the intention to return is the rising share of males in the outward flow (Gould, 1980a, p. 60; Hatton and Williamson 2005, p. 80). While there has been some analysis of the impact of return migrants (e.g. Cinel, 1991), the data is somewhat limited and the question of why return migration varied so much across different source countries (and across localities) remains under-researched.

*Migrants in the Labour Market*

There has been a vigorous debate in the economics literature about how well or badly immigrants assimilate after arrival in the host country, and the main focus has been on earnings. On arrival, immigrants have earnings significantly lower than the native born but the gap narrows as they acquire host country skills and experience. In his pioneering study of US immigrants in 1970 Chiswick (1978) found that male immigrant earnings converged on those of the native born, and even overtook them after 10-15 years of US experience. But others such as Borjas (1985, 1994) argued that cross sectional estimates overestimate the true assimilation effect if the ‘quality’ of cohorts declined over time.

This is illustrated in Figure 1. The solid lines represent the ‘true’ age-earning profiles for different cohorts of immigrants and the native-born. Here the most recent arrivals, cohort 3, have the lowest earnings profile and the earliest arrivals, cohort 1, have the highest profile. At a given point in time, the most recent cohort is observed at age A3 while the oldest cohort is observed around age A1. The dotted line shows what the estimated age earning profile for immigrants might look like when estimated from a single cross-section. The cross-sectional estimate generates a (somewhat misleading) steeper upward slope as the earlier cohorts have both higher labour market quality and longer experience in the host country. Because a cross-section cannot distinguish between host country experience effects and cohort effects, it tends to overestimate the assimilation effect for any given cohort. Analysing wage data from

⁹ Those from southern Europe were known as *golondrinas* (swallows).
different US censuses Borjas (1992) found that cohort quality declined by about 20 percent between those who arrived in 1955-60 those who arrived in 1975-80.

There is a striking parallel between the debates over US immigrant assimilation in the 30 years after 1970 and in the 30 years after 1880. Some late nineteenth century observers argued that the waves of ‘new immigrants’ arriving from Southern and Eastern Europe assimilated less well and had lower labour market quality than those who came earlier from Northwestern Europe. This view was expressed by the US Immigration Commission (the Dillingham Commission), which reported in 1911 after deliberating this issue for four years. In the recommendations of the (majority) Report it was argued that “while the American people, as in the past, welcome the oppressed of other lands, care should be taken that immigration be such, both in quality and quantity as not to make too difficult the process of assimilation” (Vol. 1 p. 45). Similarly in a widely circulated book, Jenks and Lauck argued of the new immigrants that “their general as well as their industrial progress and assimilation are retarded by segregation in colonies and communities where they have very little contact with American life and small opportunity to acquire the English language” (1926, p. 78).

In the light of Figure 1 one might expect that cross sectional estimates of earnings function for natives and immigrants would give an (over) optimistic picture of immigrant assimilation. But several studies that exploit microdata collected by State Labor Bureaus, have come to the opposite conclusion. For workers in Michigan’s copper mines and in agricultural implements and ironworking industries around 1890, Hannon (1982a; 1982b) found that immigrant wage profiles were relatively flat, with at best very little convergence on the earnings of the native-born. Eichengreen and Gemery (1986) obtained similar results for workers in Iowa in 1988/5 and concluded that assimilation was slow, especially for those who acquired their skills before migration. And in his study of earnings profiles in Michigan and California, Hanes (1996) found that immigrant earnings growth was well below that of natives up to the age of about 40. However these findings seem to be the result of using a quadratic age-earnings function. Using a more flexible functional form on the same data Hatton (1997) finds that immigrant earnings gradually converge on those of the native-born.10

Surveys from around 1890 contain relatively few new immigrants, making it difficult to measure their assimilation. Using grouped data on wages in 1909 reported by the Dillingham Commission, Blau (1980) found that male immigrants from Northwest Europe earned 12.2 percent less than natives on arrival but caught up after 11.4 years while those from Southern and Eastern Europe started with a 17.8 percent disadvantage but caught up after 16.6 years. Using the same data source, this more optimistic picture is confirmed by Chiswick (1992) and Hatton (2000). Allowing for individual source-country effects, the latter finds that immigrant wages converged on those of the native-born by about one percent per year over

---

10 The typical age-earning profile slopes steeply upwards from the age of about 16 to 25 and then follows a much flatter curve. In these datasets the native-born are relatively young and so the quadratic function picks up the steep rise at the younger ages. By contrast, immigrants are typically older and so the quadratic estimated for them is much flatter. From the estimated profiles it thus appears that the earnings of natives grow faster up to about age 40, and then decline faster at higher ages.
the 20 years since migration. After 20 years, those from northwest Europe earned 4.7 percent more than the native-born while those from Southern and Eastern Europe earned 2.9 percent less. As a result of the change in the source-country composition, immigrant quality (based on earnings) declined by about 5 percent between 1873 and 1913 as compared with around 25 percent between 1935-40 and 1975-80 (Hatton, 2000). Thus the shift in source composition before World War 1 had much less effect on the labour market performance of immigrants than has been the case post-World War 2.

Evidence from the census produces a similar account for a wider range of occupations. Minns (2000) found that immigrant cohorts tracked across the US censuses of 1900 and 1910 experienced the same level of upgrading as in the cross section. But particularly notable was the rate of movement up the occupational ladder from blue collar to white collar, especially among the new immigrants. This paints a rather more optimistic picture than do studies that concentrate only on blue collar occupations, and hence ignore a key part of upward mobility. But there are differences by nationality and by host country. Thus Green and MacKinnon (2001) find slower assimilation for British immigrants observed in the 1901 Canadian census than for Jews or non-whites. Clearly, occupational mobility was a key element in the immigrants’ assimilation process, and nowhere more so than among Russian Jews in the US (Chiswick, 1991). For Jews, upward movement into self-employment and small business was an important route and it seems to have operated more strongly for those who migrated to New York than for those who migrated to London (Godley 2001). Evidence for Canada indicates that Jews were much more likely to become self-employed than non-Jewish immigrants (Minns and Rizov, 2005).

It is often argued that upward mobility is a partly a result of positive self-selection. It is widely believed that, on average, immigrants were healthier, more energetic and enterprising and had more human capital than the populations from which they were drawn. Certainly they had more education than non-emigrants, particularly the southern Europeans (Sánchez Alonso (2007, p. 414-6). But other characteristics such as ability are much harder to measure. Mokyr and Ferrie (1994) suggest that immigrants supplied a disproportionate number of entrepreneurs and businessmen in the US. One piece of circumstantial evidence comes from the labour market performance of second generation immigrants. Their outcomes were better than for first generation immigrants, and the evidence from wage surveys and from the census suggests that they often outperformed the native-born (Hatton, 1997; Minns, 2000). An interpretation of this is that second generation immigrants inherited some of the characteristics of their positively selected parents but suffered less of the disadvantage faced by first generation immigrants.

---

11 Even more striking is the positive educational selection of immigrants from Mexico to the United States in 1910 and 1940; despite this, Mexican immigrants still had much less education than the US-born (Feliciano, 2001).
12 There is some suggestion that male immigrants from Southern Italy to the US were taller than average, but women were shorter (Danubio et al., 2004).
Ferrie (1999) examined upward mobility among antebellum immigrants by matching individual-level data recorded in ships passenger lists and in the censuses of 1850 and 1860. He found that there was some downward occupational mobility on arrival, but this was followed by steep upward mobility, particularly for the young and literate British and German immigrants. Ferrie (1999) also studied the links between the occupational and geographic mobility of immigrants. More than two thirds of immigrants arriving in 1840-1950 moved county in the subsequent decade and these moves were associated with both upward and downward occupational mobility. Relative to non-movers, labourers who moved location increased their wealth—the more so the further they moved. The high rates of mobility observed among immigrants raises the further question of what determined their initial location.

A number of studies have analysed the intended destination of immigrants on arrival in the United States at the turn of the twentieth century (Dunlevy and Gemery, 1977; Dunlevy, 1980; 1983; Dunlevy and Saba, 1992). Regression analysis indicates that immigrants migrated towards states on the eastern seaboard (close to New York), towards those with relatively high incomes, and towards those with high population densities. But the most important single influence is the location of previous immigrants from the same source country. The evidence indicates that geographic dispersion increased somewhat with onward moves (Dunlevy, 1980), but that immigrants avoided the US South (Dunlevy, 1983). A study of the intended destinations of Canadian immigrants in 1912 found, in addition, that they selected destinations and sectors that made best use of their skills (Green and Green, 1993). While migrants often moved to opportunity, there were nevertheless distinct regional and local concentrations, typically in urban areas and gateway cities and often differing by ethnicity.

As immigrants chose specific locations, this raises the question of whether they crowded out or displaced the native-born population. Hatton and Williamson (1998, p. 168) found that foreign born in-migration to the Northeast States in 1880-1910 led to out-migration of the native born. Furthermore the effect is large: for every 100 foreign born in-migrants to a state, native out-migration increased by 40. Thus the westward movement of the native-born was not just the pull of opportunity in the west but partly an immigrant-induced push from the east. When the immigration to the Northeast slowed dramatically in the 1920s this process slowed down. Collins (1997) finds that it also quickened the movement of blacks from the South to the Northern States. One implication of findings like these is that, even though immigrants were highly concentrated, their effects on the labour market percolated throughout the economy.

The Impact of Migration

One of the most contentious issues in the immigration debate has been what are the economy-wide effects of immigration and one reason it has been so contentious is because of the link with policy. The issues that have been debated in the historical literature can best be
illustrated in with a textbook diagram (Figure 2). Here there are two countries (R and S) with combined labour force measured as the width of the box. The receiving country’s labour demand curve, DR, slopes down from the left and the sending country’s labour demand, DS, curve slopes down from the right. The initial allocation of labour in the receiving country is LR$_1$ and its wage, WR$_1$, exceeds that of the sending country, WS$_1$. Migration from S to R increases the labour force in country R to LR$_2$, so that the wage falls to WR$_2$; in country S the labour force shrinks and the wage increases to WS$_2$.

The research questions that stem from this are as follows: First, did immigration depress real wages in the New World and increase them in the Old World, and if so, by how much? Second, how far did migration lead to convergence of wage rates between the New and Old Worlds? Third, are there other forces associated with migration—or with pre-1914 globalisation more generally—that shifted labour demand curves such that they either offset or reinforced the partial equilibrium effects depicted in Figure 1? And finally, how far did immigration affect the distribution of income in the New and Old Worlds? This is illustrated in the diagram by recognising that the income of other (fixed) factors of production in country R is measured in the initial situation as the area of the triangle A,B,WR$_1$, which after the migration becomes A,C,WR$_2$.

What was the impact of mass migration on real wage rates and on wage convergence between sending and receiving countries before 1914? Several different approaches have been employed. Using regression analysis on annual time series for Australia for 1860-1913, Pope and Withers (1994) found little evidence of a negative effect of immigration on wages. However they used the immigrant flow to explain the wage level, whereas theory suggests immigrant flows should explain wage change. A similar point applies to their finding that the immigrant flow had no effect on the unemployment rate (Pope and Withers, 1993). Other studies have correlated immigration with changes in wages across locations within a country (the so-called spatial correlations approach). Ljungberg (1997) finds that emigration explained about half of the rise in wages across Swedish counties between in 1870 and 1910. And Goldin (1994) found a negative relationship between wage growth and immigration across US cities between 1890 and 1923. While this approach captures the local effects of immigration it may not be a good guide to the overall national effects if (as suggested above) there is significant interregional labour mobility in response to migration inflows.

In their multi-country study, Taylor and Williamson (1997) first calculate the cumulative effect of migration on total labour supply from 1870 to 1910 in order to estimate the counterfactual labour force in 1910 in the absence of migration. They then use labour demand elasticities to estimate the effects of migration on real wages for 12 Old World countries and 5 New World countries. The effects are roughly proportional to the scale of immigration so

---

13 In these calculations the labour demand elasticities are based on the assumption that aggregate production functions are of Cobb-Douglas form (i.e. where the elasticity of factor substitution is -1). It is worth stressing that in this and most of the other studies noted below the effect of immigration wages in sending or receiving countries is based on the assumption that labour demand curves are downward sloping, and so the migration effect is not tested directly.
that in the absence of mass migration after 1870 real wages in 1910 would have been higher by 27 percent in Argentina, by 17 percent in Australia and by 9 percent in the United States; conversely real wages would have been lower by 24 percent in Ireland, by 22 percent in Italy but by only 5 percent in Great Britain and 2 percent in Germany. Overall the real wage ratio between the New and Old Worlds fell by 11 percent, whereas under the no-migration counterfactual it would have increased by 10 percent (Taylor and Williamson, 1997, p. 41). Thus, international migration more than accounts for the observed wage convergence between 1970 and 1910.

These assessments can be enriched by using computable general equilibrium (CGE) models to allow for a variety of economy-wide adjustments, although usually for only one country at a time. For Sweden, Karlstrom (1985) estimated that emigration up to 1890 raised the real wage by 9 percent, while for Norway Riis and Thonstad (1989) estimated that emigration up to 1910 raised income per capita by 6 percent. O’Rourke and Williamson (1995) found that emigration served to raise urban wages by in Sweden by 12.3 percent above what they would have been in its absence. Boyer et al. (1994) find that, in the absence of emigration after 1851, Irish real wages would have been lower by between 19 and 34 percent. For Argentina, Taylor (1997) finds that real wages would have been 25 percent higher by 1914 in the absence of immigration from 1870. O’Rourke et al. (1994) estimated that in the absence of emigration urban real wages would have been higher in 1910 by as much as 34 percent in the United States and lower by 12 percent in Britain. In terms of broad magnitudes, most of these figures are consistent with those estimated using the much cruder method of applying labour demand elasticities to counterfactual labour force estimates, but two points are worthy of note.

The first is that the CGE models allow a rich array of adjustments. Thus in the absence of emigration, labour intensive sectors would have shrunk less in the Old World and would have expanded less in the New World. In countries like Ireland and Sweden the share of agriculture would have declined more slowly, especially the most labour intensive sectors. And since these are open economy models, adjustments also come through international trade. To the extent that factor intensities become more similar, migration tends to be a substitute for trade, and so trade grows less rapidly than it would have in the absence of migration. On the other hand falling transport costs stimulated trade, which served to magnify the real wage effects of differing labour intensities. This latter effect tended to cause divergence rather than convergence in real wages.

The second is that the size of the wage effects depends critically on the assumption that is made about capital mobility. The results quoted above are based on the assumption that the capital stock in the end year in the counterfactual is the same as the observed level. The models show that, in the absence of the mass migrations, rates of return on capital would have been much lower in the New World and much higher in the Old World. It seems likely that much of the capital that flowed from the Old World to the New was effectively chasing the higher returns brought about by migration-induced labour force growth. If those returns were effectively arbitraged by the international capital market, then in the absence of
migration, capital would have retreated to Europe and rates of return would not have diverged. Thus with perfectly mobile capital, the wage effects of migration are much smaller. To give one illustration: in the absence of emigration since 1870 the British wage in 1910 would have been 6.6 percent lower with capital mobility as compared with 12.2 percent lower with no capital mobility (O’Rourke et al., 1994, p. 209). This has important implications for wage convergence in the Atlantic economy. Taylor and Williamson (1997) estimate that whereas migration explains 125 percent of the observed wage convergence in the absence of capital adjustment, it explains ‘only’ 70 percent when capital is perfectly mobile.

That there are any wage effects under capital mobility reflects the fact that there are other fixed factors, most importantly land. As its relative scarcity increased, land prices and rental rates tended to increase in the New World relative to those the Old World (O’Rourke and Williamson 1999, p. 179). And as land scarcity increased relative to labour in the New World the wage/rental ratios fell while they tended to increase in the Old World. Since the average landowner was richer than the average labourer these trends contributed to increasing inequality in the New World and decreasing inequality in the Old World. It is important to stress, however, that the focus is on distribution rather than on income per capita or overall efficiency. This is illustrated by Figure 1, where capital (which we could think of here as land) is assumed immobile. Considering the initial population only, the overall “immigration surplus” is the area of the triangle BCD, which is typically small. By contrast, the transfer from wage earners to land owners is the rectangle, WR₁,B,D,WR₂ which is much larger.

As O’Rourke and Williamson (1999) show, the magnitude of these effects differed between countries, according to scale of migration, the structure of the economy and the reaction to globalisation. Thus the rise in the wage-rental ratio was more muted in the European countries that resorted to agricultural protection (France, Germany, Spain) than in those that maintained free trade. This has led to a lively debate on the distributional effects of globalisation before 1914. A number of studies have sought to refine the original findings sometimes by revising the data series. Bohlin and Larsson (2007) find that the wage rental/ratio in Sweden increased more slowly than previously thought, which they argue is due to Sweden’s turn to protectionism from 1888. Others stress regional diversity within the New World. Emery et al. (2007) find that the wage rental ratio fell strongly from 1880 in land abundant Western Canada but not in the East. This convergence within Canada was reversed in the period of de-globalisation from 1914. For the Australian colonies Shanahan and Wilson (2007) find that while New South Wales, Victoria and Queensland exhibited declining wage

14 For other countries the results for 1870-1910 (capital immobile versus capital mobile) are: USA: 34.0, 9.2 (O’Rourke et al., 1994, p. 209); Argentina: 27.0, 25.0 (Taylor 1997, p.121); Ireland (from 1850): -19, -6 (Boyer et al., 1994, p. 235).

15 To give an illustration, for the immigration to the United States up to 1990, Borjas (1995, p. 8) calculated the loss to native-born workers would be 1.9 percent of GDP and the gain to native capital 2.0 percent of GDP, so that the immigration surplus is just 0.1 percent of GDP. But the overall surplus could be larger for skilled migration in a model that distinguishes between skilled and unskilled workers and where capital and skills are complements.
rental ratios. But South Australia bucked the trend until the 1880s due to differing colonial land policies.

In the debate on inequality and globalisation since the 1970s most of the focus has been on trends in the wage distribution. If skilled and unskilled workers are imperfect substitutes then unskilled immigration should widen the wage distribution.16 Were the effects of migration on wage-rental ratios and wage-income ratios before 1914 also reflected in the distribution of wages? The evidence suggests that immigration before 1914 increased the share of unskilled labour in the US and Canada where the wage distribution widened and reduced it for Britain, where the wage distribution narrowed. For other countries (both sending and receiving) the impact of migration on the skill mix and hence on the wage distribution is less clear. In his counterfactual analysis Anderson (2001) finds that the wage effects were generally in the expected direction but that they correlate poorly with the actual ex-post changes in wage inequality. Clearly other factors were at work. Using regression analysis Betrán and Pons (2004) find that net migration increased the skill premium for the United States but narrowed it for France, the UK, Spain and Italy. But this effect is only observed in the presence of a range of other variables that represent skill biased technical change, capital intensity, structural change, and a range of variables reflecting labour market institutions. These results are qualitatively similar to those obtained from a CGE analysis of the trend in the skill premium for UK economy from 1880 to 1913 (Betran and Pons 2010).

**Immigration Policy**

As emigration to the New World ascended to ever greater heights in the two decades before the First World War pressures for restricting immigration mounted. Nowhere more so than in the United States, which eventually imposed a literacy test in 1917 followed by the first immigration quota in 1921. Some observers have seen this as a policy backlash in response to the rise in immigration itself (Williamson, 1998); thus mass migration sowed the seeds of its own destruction. But what were the mechanisms involved? As Foreman Peck puts it: “The two key questions of any political economy of international migration are: (1) who gains and who loses from migration? And (2) who is in a position to do something about it? The answer to the first question identifies a demand for policies and that to the second reveals a policy supply” (1992, p. 360).

The answer to the first question is discussed in the previous section and is illustrated in Figure 1. Those who competed most directly with immigrants, such as low-skilled blue collar workers, had the most to lose and are likely to complain the most loudly. The answer to the second question depends on who has the vote and what particular interest they would vote

---

16 In fact there are two issues: (a) substitutability between skill (or education) groups and (b) substitutability between immigrants and non-immigrants within skill groups. The latter point is reflected in the recent debate following Borjas (2003). For United States manufacturing in 1890, Foreman Peck (1992) found evidence of less than perfect substitutability between immigrants and non-immigrants, some of which may reflect differences in skills.
for. As the franchise widened it typically percolated down the hierarchy of class and income diluting the political weight of landowners and capitalists and giving a stronger voice to urban unskilled labour. As Engerman and Sokoloff (2005) show, voting rates at the turn of the century were about a third for adults in North America and but less than ten percent in South America where the *Latifundia* retained its grip on power. Immigration restrictions came earlier in the US than in the (comparably democratic) land abundant dominions partly due to their membership of the British Empire. In labour abundant Europe labour’s interest was to encourage emigration and hence the wider the franchise the less emigration was discouraged (Foreman Peck, 1992, p. 366-7).

So what underlies public attitudes towards immigration? Two candidates stand out. The first is that, as Figure 1 illustrates, those who compete most directly with immigrants have the most to lose and are likely to complain the most loudly. Low skilled workers in blue collar occupations, who were gaining the franchise and joining unions, would be the natural constituency to support restrictive immigration policy. The alternative is the cultural hypothesis: that the growing numbers of southern and eastern Europeans were seen as challenging existing cultural norms and therefore as less acceptable.\textsuperscript{17} As the southern and eastern Europeans whose numbers were increasing steeply tended also to be less skilled and less educated, these hypotheses are (almost) observationally equivalent. Both the literacy test of 1917 and the quota Acts of 1921, and 1924 deliberately discriminated against these ‘new immigrants’.\textsuperscript{18}

Following the literature on immigration opinion for more recent times\textsuperscript{19}, Richardson (2005) provides an interesting test of these hypotheses using data on opinion towards immigration collected by the Kansas State Labor Bureau in 1895-7. He finds that the overwhelming majority of workers in this blue collar sample wanted immigration to be restricted or to be suppressed altogether. Union members were more strongly opposed to immigration than non-unionists and those with incomes in the middle of the range were more opposed than the poorest. Although increasing immigration in a locality leads to stronger opposition to immigration, the source-country composition of those immigrants matters much less. Hence the results favour the labour market competition hypothesis over the ethnic differences hypothesis. There is also evidence that opinion was more negative in years when unemployment was high. This is consistent with much of the qualitative literature that links the ebb and flow of anti-immigrant sentiment with business cycle conditions.\textsuperscript{20}

\textsuperscript{17} Just as the Irish had been regarded in the two decades after the Famine.
\textsuperscript{18} The Emergency Quota Act (also known as the Johnson Quota Act) limited annual immigration from each national origin to 3 percent of the stock in the 1910 census. The Immigration Act of 1924 (the Johnson-Reed Act) limited the number of migrants to 2 percent of the number in the 1890 census (effective 1929). Asians remained excluded. For the background to legislation in this period see Daniels, 2004, Ch. 2.
\textsuperscript{19} See for example: Mayda (2006); O’Rourke and Sinnot (2005); Dustmann and Preston (2007).
\textsuperscript{20} It may also explain the imposition of the Emergency quota in 1921, as the unemployment rate increased from 5.2 percent in 1920 to 11.7 in 1921. An alternative hypothesis is that with rapid growth in education in southern and Eastern Europe the literacy test was proving less effective than anticipated (Goldin, 1994, p. 226).
How did anti-immigrant sentiment get translated into restrictive immigration policy? In a highly influential paper Goldin (1994) examines voting patterns in the US Congress. Beginning in 1897 a series of bills incorporating a literacy test failed to pass into law until, in 1917, the House of Representatives and the Senate both overrode Presidents Wilson’s veto.\(^{21}\) Labour market effects seem to have been an important factor. Analysing vote for an override in the House in 1915 Goldin found that representatives were more likely to vote for restriction the more rapid the growth of the foreign-born population and slower the growth of wages rates in their district during the preceding years. But the higher the district’s immigrant density (the level, not the change) the less likely was a vote for restriction. This is consistent with evidence that first- and second-generation immigrants were less anti-immigration that the native-born (Richardson, 2005, p. 16).

While the run up to restriction in the US has gained most of the attention there has also been a search for more general patterns. In other New World countries the door remained ajar for longer, but there is evidence that in Canada, Australia and New Zealand policy was also tightening. That meant reducing or eliminating inducements to immigrants, and as in the US, eventually imposing tests (such as the dictation test of the white Australia policy) that screened out migrants from less favoured sources. Timmer and Williamson (1998) developed an index of immigration policy in five New World countries from 1870 up to 1930. Using these data they found that for Argentina, Brazil, Canada and the United States tougher policies were associated with declining unskilled relative wages. In Argentina and Australia with relatively homogenous immigration streams the rising share of foreign born also led to increasing restriction. In Canada and the United States where immigrant origins were more diverse it was the rise in immigration from low wage countries, differing in ethnicity and religion from earlier immigrants, that helped to close the door.

While immigration policies typically get tighter in recessions (Shughart et al., 1986), recessions seem to have more decisive effects when they are preceded by a gradual accumulation of forces that shift opinion against immigration.\(^{22}\) This may help explain the imposition of the emergency quota in the US, as the unemployment rate increased from 5.2 percent in 1920 to 11.7 in 1921. While the US led in the early 1920s, the door slammed shut in other New World countries during the Great Depression.\(^{23}\) But two other factors may also have been important. The first is that the interwar period saw a dramatic decline in international capital flows. So the effect of capital mobility in muting the wage effects of immigration would have been smaller than before 1914 and thus we might suppose that

---

\(^{21}\) The Dillingham Commission had previously recommended using a literacy test to restrict immigration.

\(^{22}\) A good example from a later era is the sudden abandonment of guestworker programmes (arbeitstopp) in Germany and other European countries at the time of the first oil price shock in 1973-4.

\(^{23}\) In Canada, the mildly restrictive Immigration Acts of 1906, 1910 and 1923 were followed much tougher regulation in 1930 and 1931, when Orders in Council banned all new immigration except for British and Americans with sufficient capital or assured employment. In Australia, the White Australia Policy was introduced in 1901 and tightened in 1924 and 1928 and then in 1930 a £50 immigration fee was introduced. In New Zealand, government assistance to immigration was abandoned in 1930 and the Department of Immigration was closed in 1931. In newly created South Africa, the immigration laws of 1902 and 1906 (framed along similar lines to those of Australia) were followed in 1930 by national origins quota based on the American model. See Daniels (1995).
workers would have been more opposed to immigration than in the era when capital chased labour (Hatton and Williamson, 2008). Second, the continuing expansion of the franchise gave greater voice to those at the bottom of the income distribution (Hatton and Williamson, 2007).

Not surprisingly international migration was dramatically lower in the interwar period as compared with the decades before 1914. Some of this was clearly the result of policy. Thus Pope (1981) found that lower immigration to Australia after 1920 could be accounted for by the reduction in subsidies and the tightening of eligibility. Taylor (1994) found shifts in the migration equations for both Australia and Argentina and Gemery (1994) found that the push-pull model no longer explained immigration to the United States. In the US where the quotas bore heavily on the ‘new immigrant’ countries, the fall was the most dramatic of all. However immigration declined even for source countries for which the quotas were not binding. During the Great Depression immigration fell even further as policy tightened more and as migrants were deterred by the high rates of unemployment in the New World. It possible also that the declining young adult cohorts and expanding welfare states caused Europeans to stay at home, but such hypotheses await more formal investigation.

The deglobalisation of the interwar period reversed many of the trends that had been underway since the middle of the nineteenth century. What migration there was shifted back from ‘new’ to ‘old’ sources; immigrant ‘quality’ increased, and the skill premium in the US declined as low skilled immigration plummeted (Hatton and Williamson, 2005, pp. 193-5). And although the skill premium also declined in European countries the decline was less dramatic than in the US and Canada (Anderson, 2001, p.94; Betrán and Pons, 2004, p. 39). More generally, wage convergence ceased and wage rental ratio stopped rising in the New World and falling in the Old World. One implication is that the conditions that gave rise to the immigration backlash were gradually reversed. These conditions ultimately provided the backdrop to the easing of immigration policies in the postwar period (Hatton and Williamson, 2005, Ch. 10). In the US where the (long delayed) policy backlash was sharpest the resumption took longer than in other New World countries where the backlash was less intense.

**International Migration in the Third World**

The preceding sections focus exclusively on emigration from Europe to the New World, within what might be called the greater Atlantic economy. This has received the vast bulk of the attention in cliometric work on migration. Yet as McKeown (2004) points out, it accounts for only a third to 40 percent of long-distance international migration in the era up to 1940. Most notable were the fifty million or more migrants from labour-abundant India and South

---

24 Two caveats are appropriate here. One is that a far larger proportion of Third World transoceanic migration was temporary or short-term. The other is that a significant share of McKeown’s total for Asia is migration within the same country or landmass—notably the migrations (mainly of Chinese and Russians) to Manchuria and Siberia.
China to labour-scarce regions such as Burma, Ceylon, parts of Southeast Asia, and the Pacific Islands as well as more distant locations on the coast of Africa, South America and the Caribbean. They were largely shut out of the greater Atlantic economy, first by the costs of migration and, from the 1880s, by anti-Asian immigration policies. The bulk of intercontinental migrants from Asia travelled as indentured labourers (reflecting the high costs of migration), until this too was curtailed by policy. Hence the largest numbers moved to the primary product exporters within Asia. Not surprisingly, migration within Asia has received growing attention, although the cliometric literature is still relatively small.

Were these migrations driven by the same forces that were observed in the Atlantic migrations? Huff and Caggiano (2007) focus on emigrants moving from India and south China to work on plantations in Burma, Thailand and Malaya. Using the standard push-pull model outlined above, they find evidence that relative wages were important, just as for transatlantic migration, but that proxies for labour demand are generally insignificant. This possibly reflects that migrants were contracted to specific sectors rather than responding to economy-wide labour market conditions. Except in one case the migrant stock was found to exert a powerful influence, something that Huff and Caggiano (2007, p. 49) attribute to the spread of information about opportunities for migration.25 Just as in Europe, the regional origins of migrants were very unevenly distributed, with notably high rates of emigration from Madras in India and Fukien and Kwangtung in China. Although high-emigration localities were often on or near the coast, we lack detailed studies of the other factors involved. One interesting study of migration from different Japanese prefectures to the Northwest United States stresses the importance of information flows set in train by pioneer migrants (Murayama, 1991). It seems likely that information flows were (even) more important in the context of Asian emigration that they were in Europe.

In Asia the volume of migration was often large relative to the receiving country population but small relative to that of the source country. Taken together, Burma, Thailand and Malaya received 15 million immigrants between 1881 and 1939, a figure in excess of their combined population in 1881. However, about fourth fifths returned, typically after a stay of three to five years (Huff and Caggiano, 2007, pp.38-9). Given the vast populations of India and China, some observers have followed W. A. Lewis in characterising them as highly elastic sources of labour. As a result, the wage in the Southeast Asian destinations should be pinned down by labour market conditions in India and China. Using cointegration and error correction models, Huff and Caggiano (2007; 2008) find that Indian wages in Burma and Malaya and Chinese wages in Thailand and Malaya were strongly related to the source wage in the respective source countries, with little trend in the wage gap and little evidence of reverse causality.26 Thus the long-run supply of labour to these booming Asian economies seems to have been more elastic than was the supply of European labour to the New World.

---

25 Interestingly they find that the migrant stock is not significant for Indian migration to Burma, the majority of whom were recruited by a kangany (head man, or gang leader).
26 Nevertheless the absolute real wages in the destination were two to three times those at the source. Huff and Caggiano (2008, p. 305-6) attribute some of this to the costs of migration and relocation, but the largest
Longer distance migrations to the plantation enclaves of the Caribbean, the Pacific and Africa were much smaller in magnitude and were conducted through contract or indenture (more so for Indians than Chinese). One reason is that the costs were much higher. Relative to the home wage, the cost of passage for these Indian and Chinese migrants were of the order of ten times those facing Europeans migrating to the New World (Hatton and Williamson, 2005, p. 140). The key element in migration under indentured servitude in the middle of the nineteenth century was the length of the contract (Northrup, 1995, pp. 115-6). The further the distance, the longer it took to recoup the costs of passage and recruitment, and hence the longer was the contract. But if the cost were high relative to the home wage so were the rewards. Wage ratios between destination and source in the range of five to nine for intercontinental migration were two to three times those for migration to Southeast Asia and were vastly higher than they were around the greater Atlantic economy (Hatton and Williamson, 2005, p. 137).

The fall in transport costs and the export-led boom that characterised Southeast Asia pervaded much of the resource abundant periphery. That raises two questions. First, did it also lead to convergence in real wages and in factor price ratios? And second, how important a part did migration play? Huff and Caggiano (2007) find some evidence of real wage convergence within Asia. However, Williamson (2002) has developed a real wage database for a wider range of countries of the periphery, which shows the opposite: real wage divergence (Hatton and Williamson, 2005, p. 146). On the other hand there is evidence of convergence in the wage-rental ratios from 1870 to 1940, which is largely accounted for by the trends in land rents. Regression evidence indicates that wage-rental ratios were driven chiefly by trends in the terms of trade, as Heckscher-Ohlin trade theory would predict (Williamson, 2002, p. 78). Thus “either the terms of trade shocks were simply too big and favoured the land- and resource-abundant regions, or the net migrations were too small or both” (Hatton and Williamson, 2005, p. 146).

Was there a policy backlash comparable with that in the Atlantic economy? Migration under indentured servitude, which had partially filled the labour supply gap left by the abolition of slavery, came under fire from the middle of the nineteenth century. On the fringes of the Atlantic economy, where there was potential competition with workers of European origin, it was fiercely opposed and severely restricted. But it survived longer in island economies like those of the West Indies, Mauritius, Reunion and Fiji, until the Indian trade was finally abolished by Britain in 1916 and in India itself a few years later. But by that time it has been in decline for two or three decades mainly because the deteriorating terms of trade reduced the demand for migrant labour (Hatton and Williamson, 2005, p. 150). Closer to home, a dramatic tightening of policy took place in Southeast Asia in the 1930s as export markets

---

component was due to the ‘universal stipulation’ among migrants to accumulate savings and to send back remittances.

27 In addition to Southeast Asia (Burma, Java, Malaya, the Philippines, Thailand, and the Straits settlements), it embraced South Asia (Assam, the Punjab, and Ceylon), tropical and semitropical Latin America (the Caribbean, the Brazilian northeast, British and French Guyana, and coastal Peru), North Africa, East Africa and the Indian ocean (Egypt, Kenya Mauritius, Natal and Reunion).
collapsed. For example, Thailand introduced a literacy test and costly residence permits in 1932. And in Malaya, where unemployment fell disproportionately on immigrants, the government embarked on a massive policy of subsidised repatriation (Huff, 2001). Thus although there are parallels with the Atlantic economy in the timing of restriction, it does not seem to have been driven by quite the same long run forces.

**Conclusion**

The cliometrics of international migration has been a vibrant research area for the last three decades. And while it has borrowed extensively from the parallel literature in economics it has examined a range of experience that has placed some of the issues studied by economists in sharper relief would otherwise be possible. For example it offers the opportunity to assess the determinants of migration largely free of restrictive immigration policies and it offers a valuable testing ground for the political economy of immigration policy. But the cliometric literature is patchy; the debate is most lively, and the literature is densest, where the available data is richest. For that reason, if not for others, the bulk of research has focused on the rise and fall of mass migration in the greater Atlantic economy, and especially on immigration to the United States. Countries in regions such as Latin America and Eastern Europe have received rather less attention. But one of the triumphs of recent research is to provide an integrated, multi-country view of the causes and effects of mass migration in the Atlantic economy as a whole. By contrast migration in Asia and elsewhere in the periphery has remained underdeveloped, notwithstanding the efforts of a number of pioneering scholars. No doubt we shall see more on that topic in the future.
References


Figure 1: Immigrant and Native Age Earning Profiles

Estimated
Native wage
Imm cohort 1
Imm cohort 2
Imm cohort 3

Wage

A3  A2  A1  Age
Figure 2: The Economic Effects of Migration